

## CONTENTS

NING WANG, DAOCHUN XU, WENBIN LI, CHEN CHEN, YONGSHENG HUANG: Feasibility study of a new thermoelectric conversion device utilizing the temperature differences in forest soil . . . . .	1-12
QINGYOU YAN, YOUWEI WAN, XU WANG: Competitive evaluation based on integer-valued DEA model with different constraint sets . . .	13-22
HAN PENG: Insulation and monitoring system for pure electric vehicle based on microcontroller unit . . . . .	23-34
YANG YU: Dynamics and simulation analysis of table tennis robot based on independent joint control . . . . .	35-44
XIAOHONG LIU, YU DING, KAIMING YIN: Genetic algorithm based on the crest value of binary function . . . . .	45-56
HAIBO LIU: Seismic performance analysis of frame and frame-shear wall structures based on energy balance . . . . .	57-68
JUMEI ZHANG: Fault diagnosis methods of rotating machinery based on mathematical morphology . . . . .	69-78
SHUNA ZHOU: Design and application of computer network system integration based on network topology . . . . .	79-88
RONGSHENG LIU, MINFANG PENG, XUN WAN, HAIYAN ZHANG, WEI ZHOU: Evaluation of transformer insulation condition based on cloud matter element model . . . . .	89-102
LIJUN XIE, WENHUI YANG: Simulation study on automotive EMB system based on self-tuning fuzzy PID control . . . . .	103-116
LI KE, YIN KE: Application of structural bending differential equation algorithm in bracing composite structure . . . . .	117-130
CHEN YAN: Design and kinematic control simulation of wheeled mobile table tennis manipulator . . . . .	131-144
YANTIAN ZHANG, YUHANG CHEN, JINSONG TAO: Electric braking method in hydroturbine generating unit and transformation scheme .	145-154
LIAN DAI, YULIN LAI, MENGLIANG SHAO: Application of electronic information technology in modern logistics system . . . . .	155-164

JIAN YANG, LINBIN WU: Field programming technology for instrument sensor's correction parameter . . . . .	165–174
ZIHUI ZHENG, XIAOHU ZHOU: Integration management of strategic supply chain based on genetic algorithm . . . . .	175–186
JIA WANG: A computer behavioral analysis algorithm based on image color classification statistics . . . . .	187–196
CHENGBO HU, YONGLING LU, FENGBO TAO, WEIHUA CHENG, WEN-NAN WU: Auxiliary method of transformer intelligent partial discharge detection based on UWB technology . . . . .	197–206
WEI LIU, SHUANG LIU: Design of medical track logistics transmission and simulation system based on internet of things . . . . .	207–220
CHANGXING GENG, PENG WANG, PENGBO WANG: Image segmentation algorithm based on improved fuzzy clustering . . . . .	221–232
CHENGBO HU, XIAOBO WANG, YONGLING LU, WEIHUA CHENG, WEI DAI: Research on an improved UWB space location method . . . . .	233–244
TIAN BOJIN, CHEN KE, AI QIANKE, QIU XIAOPING: Analysis of RFID technology based on technology principles and construction of development model . . . . .	245–254
JUN HU: Application of ZigBee wireless sensor network in gas monitoring system . . . . .	255–264
CHANGXING GENG, PENG WANG, PENGBO WANG: Image registration algorithm based on SIFT feature descriptor . . . . .	265–276
DAN QIU, SHULI GONG: Data acquisition and fusion system based on wireless sensor . . . . .	277–288
RONG TAO, MENG LUO JI: Mining urban active point circle based on spatio-temporal constraint data . . . . .	289–300
NI NA ZHOU, LI WANG: Fuzzy adaptive control for boiler based on nonlinear depth recursion . . . . .	301–312
WU DAFEI: Software test data generation algorithm based on multi-dimensional space-time granularity . . . . .	312–324

BAOYAN ZHANG: Chaotic time series prediction based on the fusion of multi-source collaborative data feature constraints . . . . .	325–336
Q. U. JIE: Practical online learning model based on big data balance . .	337–348
SUN JING: Analysis algorithm for internet of things big data based on multi-granularity functional . . . . .	349–360
XIAOFENG LV: Image analysis of workpiece joints with uncertainty for cooperative feature constraints . . . . .	361–368
YAN ZHENGSHU: Expression of moving object group movement in large scale traffic network . . . . .	369–380
ZHANG JING, GUO XINCHANG: High-dimensional data express model based on tensor . . . . .	381–390
JUFANG GONG: Semantic analysis of English vocabulary based on random feature selection . . . . .	391–402
CAILIN LU: Wireless multimedia network feasible path routing algorithm . . . . .	403–414
HONGNA LI, HUI ZHAO: Fault diagnosis of smart grid based on improved immune optimization algorithm . . . . .	415–426
ZHIXIA HAN: Leakage monitoring algorithm of water supply pipe network based on information entropy difference . . . . .	427–436
YULONG QI, ENYI ZHOU: Human resource data location privacy protection method based on prefix characteristics . . . . .	437–446
DAFEI WU: High-throughput computing optimization of cloud computing platform based on Markov jump . . . . .	447–456
LIN GAN, JUN DU, XINGCONG FU: Virtual reality remote control system based on image . . . . .	457–468
WANG BOHUAI: Research on data acquisition and fusion system based on wireless sensor . . . . .	469–478
ZHOU JUNPING: Analysis of key technologies of distributed file system based on big data . . . . .	479–488

GUO JING: Research on application of DDos attack detection technology based on software defined network . . . . .	489-498
HONGJI CHEN: Research on the application of digital media art in ani- mation control based on Maya MEL language . . . . .	499-508
TANG GUOQIANG: Computer intrusion factor decomposition detection based on biotechnology . . . . .	509-520
JIYUN XU: Optimization design of incineration processing of radioactive solid waste . . . . .	521-530
NING GUANGQING: Application of automatic control technology of drill- ing machine based on PLC . . . . .	531-540
CHANG CHUNYANG: Application of PLC technology in fire-fighting con- trol system of oil field united station . . . . .	541-550
SHUAI MAN: Improvement research of genetic algorithm and particle swarm optimization algorithm based on analytical mathematics . . .	551-560

# Feasibility study of a new thermoelectric conversion device utilizing the temperature differences in forest soil<sup>1</sup>

NING WANG<sup>2</sup>, DAOCHUN XU<sup>2,3</sup>, WENBIN LI<sup>2,3</sup>,  
CHEN CHEN<sup>2</sup>, YONGSHENG HUANG<sup>2</sup>

**Abstract.** The new thermoelectric device can be powered for forest wireless sensors stably. There is a temperature difference between the forest's various soil layers—it is possible to make use of this temperature difference to generate electricity. This device is mainly composed of heat pipes and thermoelectric power generators (TEGs), which can transfer soil heat to electricity. By simulating the forest soil environment, the experimental results show that for a stable soil temperature, the device can generate approximately 298.5 mV, which is superior to other existing device. The device provided a new type of power supply for wireless sensors. The results provide the theoretical and technical basis for a form of power generation that utilizes a forest's soil temperature.

**Key words.** Wireless sensor, soil heat resource, thermoelectric power generation.

## 1. Introduction

Because of geological movement, climate drought, man-made reasons, forest fires and mudflows occur frequently. Early and timely prevention and monitoring of forest fires [1] is essential for reducing the loss of natural resources and mitigating economic losses [2, 3]. It is also important to monitor the forest structure and the mortality of tree populations [4]. Forest wireless sensors utilize meteorological and remote sensing methods (integrated with data mining methods) to monitor the impact of drought on forests [5]. Wireless sensors can be used to uncover a forest's structure and explore the risk of the forest changing and evolving ecologically, monitoring wildlife in its

---

<sup>1</sup>Support for this project from the National Natural Science Foundation of China (31670716), the China Postdoctoral Science Special Foundation (2016T90044), and the China Postdoctoral Science Foundation (2015M570945) is gratefully acknowledged.

<sup>2</sup>School of Technology, Beijing Forestry University, Beijing, 100083, China

<sup>3</sup>Corresponding authors

forest habitat [6], forest residue burning [3], vegetation changes, and precipitation of forest biota [7]. This requires the support of real-time systems related to forest monitoring. Having forest fire detection systems with real-time processing can lead to excessive energy consumption though. The application of wireless sensor network monitoring technology is an aspect of precision monitoring technology [8]. It is integral to the entire forest monitoring system, which has become a key concern in many countries in recent years [9, 10]. Nowadays sustainable natural energy can take the form of solar energy, wind energy, hydro energy, vibrational energy, and geothermal energy; these are examples of sustainable acquisition and conversion of natural pollution-free energy. In forest environment there is little advantage to using solar energy because electricity cannot be generated at night. When it is rainy, while the wind in the forest is unreliable, too. Therefore, relying on solar or wind energy in forest is not feasible. To understand how we might be able to generate power in forests, it should be noted that in general many countries have decided to utilize generators that generate electric power from waste heat via industrial waste heat, garbage incinerators, or waste heat from automobiles.

In recent years, a variety of devices have been developed to recover waste heat: for example, an apparatus was developed to recover thermal waste [11]. Vehicle exhaust can be recovered by heat pipes and thermal batteries [12]. Meanwhile, Andre Moser developed a thermoelectric collector to utilize the natural temperature difference between a building's walls and the surrounding air to essentially turn this energy into electrical energy that is stored in a capacitor. However, no device has been developed to collect energy from the temperature differences in the soil [13]. Additionally, Meydbray built a harvester that could utilize the temperature differences through different degrees of sun irradiation in forests; however, the system did not function very efficiently. In China, most research into thermoelectric power generation aims to produce either power generators or materials that can be used for thermoelectric power generation. Additionally, a thermoelectric generator device has been proposed for automotive exhaust gas conversion [14] and industrial waste heat conversion [15]. At present, solar energy and waste heat generators are used to study the heat transfer efficiency of such system [16]. Deng built a thermoelectric conversion system model and applied it to cars, which verified the energy-saving effects of using thermoelectric collectors in cars [17]. Meanwhile, Zhe invented a thermoelectric conversion device that utilizes solar energy as the main energy source; however, those devices cannot be applied in forest environments [18]. The leaves of trees in the forest are thick so that it is difficult to utilize the solar energy.

Because of the low thermal conductivity of soil and atmospheric effects, there is a temperature difference among forest's soil depth, i.e., there is a temperature difference between the upper and lower layers of soil. Based on the Seebeck effect, any temperature difference can be used to generate electricity [15, 19]. If the temperature difference between the forest and the soil surface were utilized for power generation, it could provide a stable and reliable micro power source for wireless sensors in a forest. The key technical issue is how to transmit the energy generated from the soil's heat to the TEG. Having been invented only in recent years, heat pipes are a new technology that has a superior metallic heat transfer performance compared

with traditional systems; it also provides a way to efficiently transfer heat energy. One successful study has demonstrated the transfer of heat from lower layer of soil to upper layer of soil through heat pipe. The main feature of heat pipe is its small thermal resistance, fast heat transfer characteristics, high efficiency, lightweight, small in size, and reliable. For general applications of heat pipe technology, however, its thermoelectric conversion efficiency needs to be improved.

Studies have shown that heat pipes can be buried in the ground to facilitate heat exchange with the ground. A large number of studies have focused on the use of heat pipes that transfer heat from underground to the surface. A snowmelt system for roads was designed to melt snow and ice using geothermal energy. Thus, previous research can lay the theoretical and technical foundation for a forest soil thermoelectric power generation system as well as laying the foundation for exploring a heat transfer mechanism from the ground to a thermoelectric generator through the heat pipe. Thermoelectric generator connected to a heat pipe can be applied in the forest environment to solve the issue of power supply problem for forest wireless sensor network.

Thermoelectric conversion technology is applied in many areas; however, thermoelectric power supplies for forest wireless sensor applications are still lacking. The thermoelectric conversion device presented here has an increased heat transfer efficiency compared with that demonstrated in the original study. The design of our thermoelectric conversion device depends entirely on heat at the soil's surface, which is more suitable for a forest environment. The particular structure of the device is superior to that of others, which were not specifically designed for forestry applications. The main purpose of our study was: a) the design and fabrication of a thermoelectric conversion system and the improvement of the model for simulating the soil environment and b) to test the performance of the thermoelectric conversion system. We aimed to verifying the feasibility of thermoelectric power generation from soil, to improve the thermoelectric conversion efficiency, and to conduct the research in forest-like conditions.

## 2. Materials and methods

Thermoelectric generators are able to directly convert thermal heat differences into electricity when there is a temperature difference between the two ends of a thermoelectric material. The thermoelectric effect is actually based on either the Seebeck effect, the Peltier effect, the Thomson effect, the Joule effect, or the Fourier effect. Modern thermoelectric power generation devices are based on both the Seebeck effect and the Peltier effect. Thermoelectric power generation has the advantage of being environmentally safe, with neither noise nor pollution generated. To generate electricity, many p-type and n-type semiconductor legs are sandwiched between two electrically insulating materials that have thermoelectric properties. There is a linear relationship between the thermal electromotive force and the temperature difference between the hot and cold faces. The thermoelectric conversion figure of merit  $Z$  is related to the properties of the semiconductor material. The figure of

merit is based on the material's thermoelectric transfer properties:

$$Z = \frac{(N\alpha)^2}{KR}, \quad (1)$$

where  $R$  is the internal electric resistance,  $K$  is thermal conductivity and  $N$  is the number of thermoelectric couples incorporating p-type and n-type semiconductor elements. According to the above definition, the figure of merit  $Z$  is related to the Seebeck coefficient  $\alpha$ , the geometric dimension of the semiconductor galvanic couple, the total resistance of the thermocouple arm, and the thermal conductivity.

The proposed thermoelectric conversion device includes a heat pipe, eight TEGs, two copper sleeves, and copper fins; its structure is depicted in detail in Fig. 1.

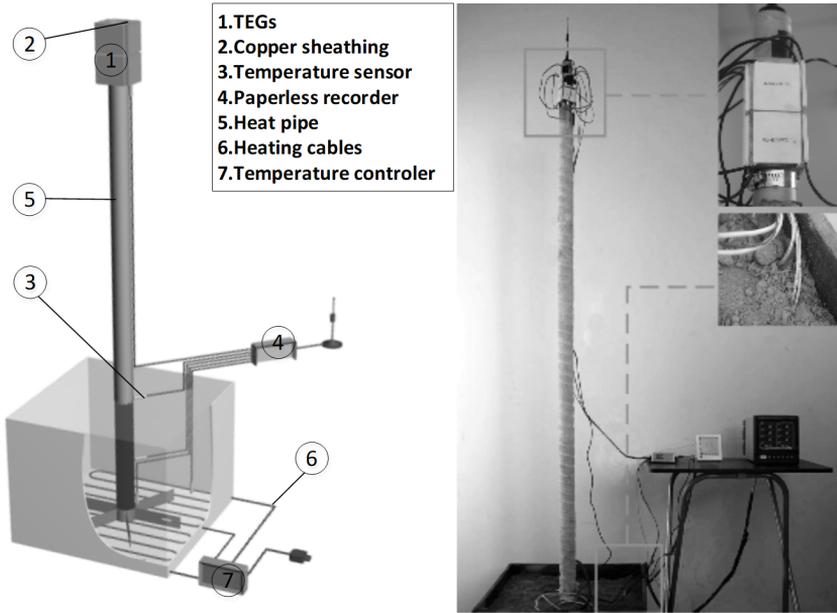


Fig. 1. Thermoelectric conversion device

Three types of heat are relevant to thermoelectric conversion devices: the heat generated by the heat conduction between the soil and the heat pipe,  $W_{sh}$ ; the heat generated by the heat conduction between the soil and the copper sleeves,  $W_{hc}$ ; and the heat generated by the heat conduction between the thermoelectric generator and the copper sleeves,  $W_{ct}$ . The expression for  $W_{sh}$  is as follows:

$$W_{sh} = -hA \frac{\partial T}{\partial n}, \quad (2)$$

where  $h, n$  are the thermal conductivity coefficient and the unit outward normal, respectively. Symbol  $A$  is the contact area between the heat pipe and soil and  $T$  is the temperature of heat pipe in the soil.

Thermal energy absorbed by the heat pipe through the soil is transported from

an evaporator to a condensation section. Then, the heat energy is transmitted to the copper sleeves by heat conduction:

$$W_{hc} = W_{sh} - \Phi_b - q_C - q_l. \quad (3)$$

Here,  $\Phi_b$  is the loss heat of the heat pipe,  $q_C$  is the loss heat of the copper bush and  $q_l$  is the loss heat of other parts.

Soil thermal energy absorbed by the heat pipe is scattered in five directions. Eventually, the energy absorbed by the thermoelectric energy collection module is calculated according to formulae

$$W_{ct} = W_{hc} - \Phi_{Cu} - q_{Cua} - q_l, \quad (4)$$

where

$$\Phi_{Cu} = \varepsilon_c A_e \sigma_b (T_c^4 - T_a^4), \quad q_{Cua} = \lambda_c A_c \Delta T_2, \quad (5)$$

where  $\varepsilon_c$ ,  $\sigma_b$ ,  $A_e$ ,  $A_c$ ,  $T_c$ ,  $T_a$  are the emissivity of thermoelectric energy collection module average surface of energy conversion system, Stefan Boltzmann constant, thermoelectric energy conversion coefficient of the surface area of radiation, convection area between the system surface layer and air in the process of energy collection, the copper bush average surface temperature and average temperature of the environment of thermoelectric energy conversion system, respectively.

The power obtained from the thermoelectric electric generator is given by:

$$q_{TEG} = Q_h - Q_l, \quad (6)$$

where  $q_{TEG}$ ,  $Q_h$ ,  $Q_l$  are power emitted from thermoelectric power generation cell current, heat flows of the hot side, the heat flows of the cold side, the heat generated by the Peltier effect, the heat passing through the semiconductor compose  $Q_h$  and  $Q_l$  together. The heat generated by the Peltier effect, the Joule heat in the semiconductor, and the temperature difference based on the above analysis may be given as

$$\begin{aligned} Q_h &= \alpha_{ab} I T_h + \lambda(T_h + T_l) - \frac{1}{2} R I^2, \\ Q_l &= \alpha_{ab} I T_l + \lambda(T_h - T_l) + \frac{1}{2} R I^2, \end{aligned} \quad (7)$$

where  $\alpha_{ab}$ ,  $T_h$ ,  $T_l$ ,  $\lambda$ ,  $I$ ,  $R$  are the total Seebeck coefficient of thermoelectric power generation sheet, temperature of the cold side of the thermoelectric generator, temperature of the hot side of the thermoelectric generator, total semiconductor chip thermal conductivity of the thermoelectric power generation, current in the closed circuit and sheet resistance of the thermoelectric power generation, respectively.

The heat flow through the TEG is a function of the performance of the thermoelectric materials and thermoelectric component geometries. It can also be described as follows:

$$q_{TEG} = N(\alpha_p - \alpha_n) I T_u + K(T_u - T_d) - \frac{1}{2} R I^2. \quad (8)$$

In (8),  $\alpha_p$  is the P-thermoelectric power generation of TEG,  $\alpha_n$  is the N-thermoelectric power generation of TEG,  $T_u$  is the high temperature of TEG,  $T_d$  is the low temperature of TEG, and  $K$  is the total thermal conductivity of TEG. Finally,  $N$  denotes the number of P-N pairs in TEG.

From the above relationship, we can find that

$$W_{ct} = q_{TEG} \cdot \quad (9)$$

Therefore, the proposed thermoelectric conversion device is theoretically feasible. The proposed thermoelectric conversion device includes a heat pipe, eight TEGs, two copper sleeves, and copper fins. As shown in Fig. 1, the heat pipe's evaporator is located in a steady external soil heat source (simulating a forest's soil environment) in which the temperature is higher than the air temperature, which was controlled by air conditioning. The principle of how a heat pipe works is also shown; the heat pipe's working fluid evaporates and becomes a gas as the soil heat is absorbed at the evaporator side. Under atmospheric pressure the working gas condenses into droplets when it arrives at the condenser and transfers the heat to the copper sleeves. Then, the working liquid travels back to the evaporator end, achieving an energy self-transfer after repeated cycles. A heat pipe's internal processes include two-phase flow and phase change heat, and so the inner heat transfer principle is very complex. The simplified model of heat pipes divides the heat transfer process into three parts: the heat exchange in the condensation section, the heat exchange at the evaporator, and the heat exchange with the insulation of the heat pipe. The adiabatic section of the heat pipe was covered with asbestos to reduce the loss of heat. There were also two copper sleeves set into the condensation section of the heat pipe. On each surface of the copper sleeves four TEGs were inlaid. These kind of heat pipes do not have wicks and they depend on the gravity principle. Therefore, the inner pressure of the heat pipes is determined by the vapor pressure of the evaporating working liquid. The working liquid will evaporate when the surface of the heat pipe is warm. The heat pipe produces a pressure difference because the vapor temperature and pressure of the heat pipe's evaporator is slightly higher than its other parts, which promotes steam flow to the condensation section of the heat pipe. When the steam condenses on the heat pipe's wall, it releases heat, which is transferred to the condensation section. Then, the condensed liquid returns to the evaporator under gravity. This process will loop as long as the heat source continues to exist.

To build an experimental power system platform with a steady external heat source (simulation of the soil environment) we used a heat pipe, eight TEGs, two copper sleeves, and copper fins. There were four temperature sensors on the heat pipe. The first temperature sensor was placed on the evaporation section. The second was placed on the adiabatic section, while the third was placed on the evaporation section. The last one was used to record the surrounding air temperature. The temperature sensors were used to monitor and record the temperature of the various parts of the heat pipe. There were two parallel copper sleeves mounted on the evaporator section of the heat pipe; one TEG was placed on each side of it. Therefore, there were eight TEGs on the copper sleeves in total, which were used to connect the TEGs and the heat pipe. The soil's heat can be transferred through the

heat pipe and the copper sleeves to the TEGs. Then, the heat can be converted to electricity through the TEGs based on the Seebeck effect. Our measurement results verify the mathematical model. We proposed an optimum design for the thermoelectric conversion device, which provides the technical basis for improvements to electricity generation systems.

### 3. Results

As shown in Fig. 1, there are three temperature measurement points distributed on the heat pipe that gauge the evaporator temperature B, the section temperature C, and the adiabatic section temperature D. The air temperature has the label A. The length of the heat pipe is 2 m, its diameter is 38 mm, and its wall thickness is 3 mm. The working liquid volume takes up 1/40th of its entire volume. The condensation section of the heat pipe in this experiment was covered with asbestos. There were two sets of control experiments. The thermoelectric conversion device was placed in both soil and water. The air temperature was 14 °C, which was controlled by air conditioning. The temperature of the soil was changed by heating cables buried in the soil. The temperature difference between the soil and the air ranged from 5 to 20 °C. The temperature change of each of the three parts of the heat pipe was recorded for every 1 °C of change. Although the heat transfer coefficient of the heat pipe is high, the customized heat pipe in this device is larger in diameter than that of a conventional heat pipe and is also longer. In addition, the heat transfer coefficient of water is much higher than that of soil. The results of the measurement are illustrated in Fig. 2. First, the heat transfer occurs much faster for a heat pipe with asbestos than for one without it. The maximum temperature of the evaporation section of the heat pipe without asbestos remains constant, which takes almost 30 min. According to Fig. 2, it takes only 15 min for the pipe to reach its maximum temperature after which its temperature remains constant. Simultaneously, using asbestos significantly reduces heat loss to the surroundings; the temperature of the evaporator is also higher than before. The temperature of the evaporation section of the heat pipe with asbestos was by 2 °C higher than that without asbestos. Second, compared with water, the heat transfer coefficient of soil is much lower. In Fig. 2, it can be seen that it took nearly 70 min for the heat pipe to reach a constant temperature. Based on the data in Fig. 2, we exploited the temperature difference, which was less than 2 °C, between the three parts of the pipe. In particular, the temperature difference between the evaporator and the adiabatic section was only 1 °C. The temperature of the heat pipe was constant after being in the soil for about 70 min. This shows that the heat pipe has good isothermal properties and a low thermal resistance. Nevertheless, the temperature of the top of the heat pipe was a bit lower, because there is a cap on the inner heat pipe to seal the working liquid. Hence, there is only a 1 °C difference between the end of the heat pipe and 1.5 m up the heat pipe. When the heat pipe works stably, the temperature difference between the heat pipe and the heat resource is 6 °C. According to the definition of the Carnot efficiency, the heat transfer efficiency cannot be one hundred percent. Therefore, there is a heat loss of 6 °. After a series of experiments this value remained

constant. The physical parameters of TEGs are given in Table 1.

As can be seen in Figs. 3 and 4, there were some irregular data. First, for a low temperature difference of 5–9 °C, the amount of electricity generated by the device was very small. As the temperature difference increased, the voltage and current noticeably increased. When the resistance in the parallel circuit was as large as the TEG’s inner resistance, the voltage was larger than for other resistance values for the same temperature difference.

Table 1. Physical characteristics of TEG

MPN	Couple	$V_{DC}$ (V)	$R_{TE}$ ( $\Omega$ )	$I_{MAX}$ (A)	$P_{MAX}$ (W)	$T_{MAX}$ ( $^{\circ}$ C)
TEG-12708T237	127	3.4	5	1.81	6.2	250

Second, there are further abnormal data in the chart for temperature differences between 18 and 20 °C. Through repeated measurements and analyses the reason behind the abnormal data was found. One side of the TEG was adhered to the hot section, while the other part was exposed to the relatively cold air. According to the first law of thermodynamics, heat always spontaneous transfers from high to low temperature objects without external force being necessary. Hence, both sides of the TEG are dynamically consistent. We conducted another experiment to compare the difference in electricity that is generated when the temperature rises and falls by 1 °C for each temperature up to 20 °C (with 1 °C intervals). When the air temperature was 14 °C, the temperature difference between the air and the soil was adjusted from 5 °C all the way to 20 °C; the recorded data is presented in Fig. 3. This verifies that no matter what the temperature of the cold side of the TEG is to begin with, once the temperature difference between the cold side and hot side (air and soil temperature difference) reaches a constant value, the value of the generated electricity will also be stable.

As shown in Fig. 5, we varied the number of thermoelectric generators in the device while keeping the other conditions constant. As the temperature difference increased, the voltage in the three different circuits increased correspondingly. There was an approximately linear relationship between the behaviors of the three circuits. The voltage in the circuit with eight TEGs was nearly four times that of the voltage in the circuit with four TEGs and eight times that of the voltage in the circuit with one TEG. The voltage initially increased very slowly for low temperature differences. As the temperature difference increased, the voltage increased more. The tests showed that the more TEGs there are, the greater the voltage is that is generated for the same temperature difference. The voltage was largest for the largest temperature difference.

## 4. Conclusion

This paper investigated a thermoelectric conversion device that can be used as a power supply for forest wireless sensors. The thermoelectric conversion device is based on the Seebeck effect and heat pipe working principle. The main conclusions

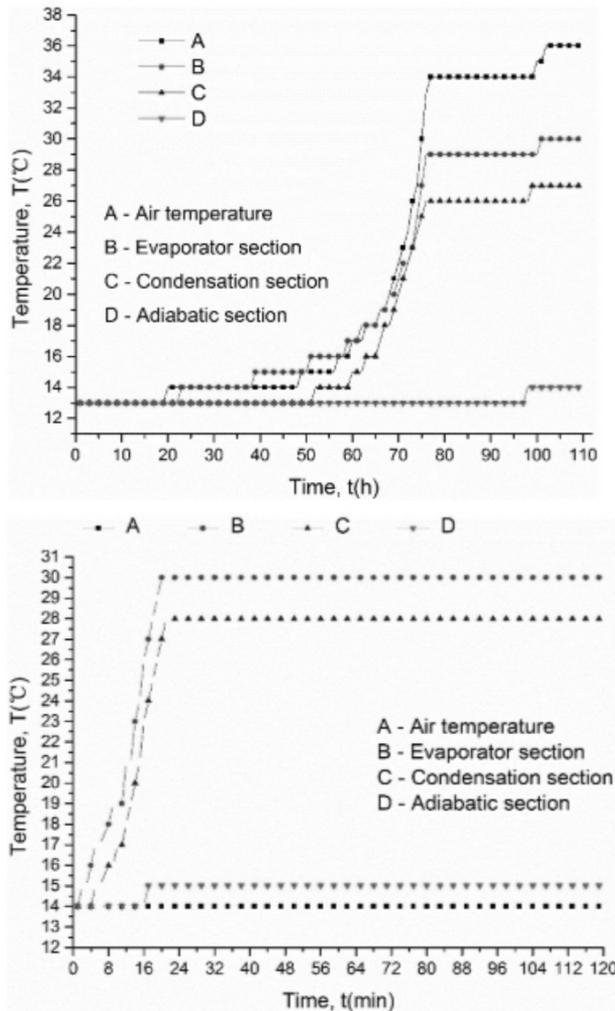


Fig. 2. Startup characteristics of the heat pipe after 110 h and 120 min

of the study are as follows:

The paper theoretically verified the feasibility of the thermoelectric conversion device to power low power wireless sensors.

A thermoelectric conversion device was designed that could be applied in a forest in future.

The thermoelectric conversion device was then manufactured. The device consists of a heat pipe, thermoelectric electronic generator, copper sleeves, and endothermic fins. The components of the device were designed to take on special shape to optimize the functioning of the device.

The thermoelectric conversion device produces 298.5 mV for the temperature difference of 20 °C. This is sufficient meet the demands of a forest wireless sensor. The

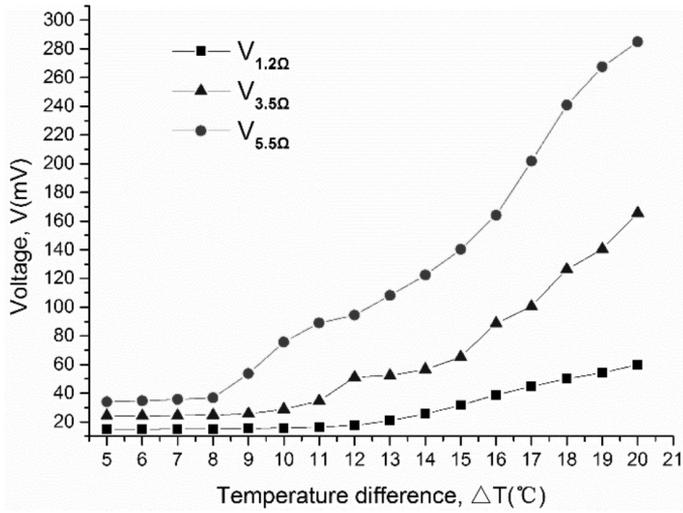


Fig. 3. Voltage as a function of the temperature difference ( $\Delta T \uparrow$ )

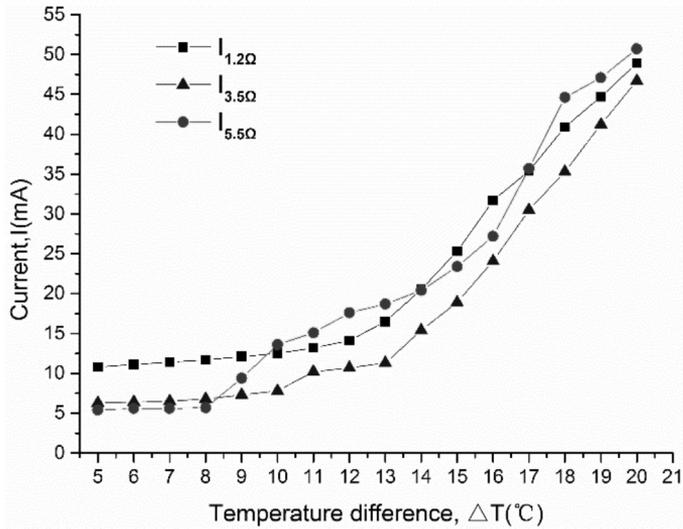


Fig. 4. Current as a function of the temperature difference ( $\Delta T \uparrow$ )

results of this study will thus contribute towards future application of thermoelectric generators in forests.

## References

- [1] N. E. RIKLI, A. ALNASSER: *Lightweight trust model for the detection of concealed malicious nodes in sparse wireless ad hoc networks*. International Journal of Distributed Sensor Networks 12 (2016), No. 7, 1–16.

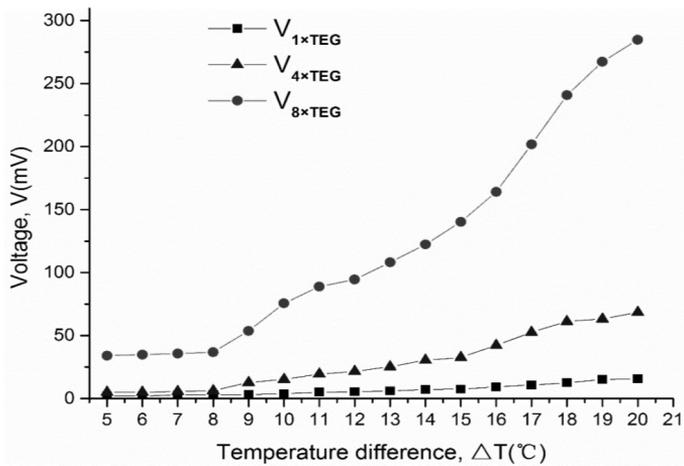


Fig. 5. Voltage changes with different TEGs

- [2] W. C. WU, Z. L. DU, J. L. CUI, Z. SHI, Y. DENG: *Thermoelectric generator used in fire-alarm temperature sensing*. Journal of Electronic Materials 44 (2015), No. 6, 1851–1857.
- [3] X. F. YAN, H. CHENG, Y. D. ZHAO, W. YU, H. HUANG, X. ZHENG: *Real-time identification of smoldering and flaming combustion phases in forest using a wireless sensor network-based multi-sensor system and artificial neural network*. Sensors (Basel) 16 (2016), No. 8, 1228.
- [4] J. A. CALE, J. G. KLUTSCH, N. ERBILGIN, J. F. NEGRÓN, J. D. CASTELLO: *Using structural sustainability for forest health monitoring and triage: Case study of a mountain pine beetle (Dendroctonus ponderosae)-impacted landscape*. Ecological Indicators 70 (2016), 451–459.
- [5] S. P. NORMAN, F. H. KOCH, W. W. HARGROVE: *Review of broad-scale drought monitoring of forests: Toward an integrated data mining approach*. Forest Ecology and Management 380 (2016), 346–358.
- [6] X. J. DING, G. D. SUN, G. X. YANG, X. SHANG: *Link investigation of IEEE 802.15.4 wireless sensor networks in forests*. Sensors (Basel) 16 (2016), No. 7, 987.
- [7] P. J. HANSON, J. F. WELTZIN: *Drought disturbance from climate change: response of United States forests*. Science of the Total Environment 262 (2000), No. 3, 205–220.
- [8] X. XU, W. F. LIANG, X. H. JIA, W. XU: *Network throughput maximization in unreliable wireless sensor networks with minimal remote data transfer cost*. Wireless Communications and Mobile Computing 16 (2016), TOC No. 10, 1176–1191.
- [9] S. CHAMANIAN, S. BAGHAEI, H. ULUSAN, Ö. ZORLU, H. KÜLAH, E. UYSAL-BIYIKOGLU: *Powering-up wireless sensor nodes utilizing rechargeable batteries and an electromagnetic vibration energy harvesting system*. Energies 7 (2014), No. 10, 6323–6339.
- [10] E. E. AKTAKKA, K. NAJAFI: *A micro inertial energy harvesting platform with self-supplied power management circuit for autonomous wireless sensor nodes*. IEEE Journal of Solid-State Circuits 49, (2014), No. 9, 2017–2029.
- [11] J. IBARRA-BAHENA, R. J. ROMERO, J. CEREZO, C. V. VALDEZ-MORALES, Y. R. GALINDO-LUNA, L. VELAZQUEZ-AVELAR: *Experimental assessment of an absorption heat transformer prototype at different temperature levels into generator and into evaporator operating with water/carrol mixture*. Experimental Thermal and Fluid Science 60 (2015), 275–283.
- [12] B. ORR, B. SINGH, L. TAN, A. AKBARZADEH: *Electricity generation from an exhaust*

- heat recovery system utilising thermoelectric cells and heat pipes*. Applied Thermal Engineering 73 (2014), No. 1, 588–597.
- [13] A. MOSER, M. ERD, M. KOSTIC, K. COBRY, M. KROENER, P. WOIAS: *Thermoelectric energy harvesting from transient ambient temperature gradients*. Journal of Electronic Materials 41 (2012), No. 6, 1653–1661.
- [14] B. ORR, A. AKBARZADEH, M. MOCHIZUKI, R. SING: *A review of car waste heat recovery systems utilising thermoelectric generators and heat pipes*. Applied Thermal Engineering 101 (2016), 490–495.
- [15] Q. LUO, P. LI, L. CAI, P. ZHOU, D. TANG, P. ZHAI, Q. ZHANG: *A thermoelectric waste-heat-recovery system for portland cement rotary kilns*. Journal of Electronic Materials 44 (2015), No. 6, 1750–1762.
- [16] P. BALAMURUGAN, A. MANI: *Comparison of compact and tubular generators performance for R134a–DMF*. Experimental Thermal and Fluid Science 45 (2013), 54–62.
- [17] Y. D. DENG, Y. ZHANG, C. Q. SU: *Modular analysis of automobile exhaust thermoelectric power generation system*. Journal of Electronic Materials 44 (2015), No. 6, 1491–1497.
- [18] Z. ZHANG, W. B. LI, J. M. KAN: *Behavior of a thermoelectric power generation device based on solar irradiation and the earth’s surface-air temperature difference*. Energy Conversion and Management 97 (2015), 178–187.
- [19] X. LIU, Y. D. DENG, W. S. WANG, C. Q. SU: *Experimental investigation of exhaust thermoelectric system and application for vehicle*. Journal of Electronic Materials 44 (2015), No. 6, 2203–2210.

Received June 29, 2017

# Competitive evaluation based on integer-valued DEA model with different constraint sets<sup>1</sup>

QINGYOU YAN<sup>2</sup>, YOUWEI WAN<sup>2,3</sup>, XU WANG<sup>2</sup>

**Abstract.** To discuss the integer-valued DEA (data envelopment analysis) model, different nations' values different medals in different ways and the targets for the inefficient nations should be integer-valued in evaluating the performance of participating nations in the competition. It makes some adjustment to the model which considers different constraint sets to make the efficiency score of each nation lie between zero and unity. And it proposes the radial-based integer-valued DEA model (the RDI model) to obtain the integer-valued targets for the inefficient nations. Based on these, it presents an integer-valued DEA model with different constraint sets (the RDID model) to evaluate the participating nations at the competition. In addition, comparisons are made among the RDID model, the integer-valued DEA model with the same constraint set (the RDIS model) and the BCC model. The results demonstrated the feasibility and justice of the RDID model. Based on the above findings, it is concluded that the model can also be used to evaluate some similar problems in an instructive way.

**Key words.** Data envelopment analysis (DEA), different constraint sets, integer values, performance evaluating.

## 1. Introduction

Data envelopment analysis (DEA) has been widely used to evaluate the performance of the participating nations. DEA measures the relative efficiency of a set of decision making units (DMUs) through the programs, which started with the work of Charnes and Cooper. Later, more and more different cases led to the proposition of other models, such as the BCC, additive, hybrid, cross efficiency models, which broadened the application area of DEA. DEA possesses two notable advantages. One is that it does not have any assumptions on the production function, and the

---

<sup>1</sup>The authors acknowledge the National Natural Science Foundation of China (Grant: 51578109) and the National Natural Science Foundation of China (Grant: 51121005).

<sup>2</sup>School of Economics and Management, North China Electric Power University, Beijing, 102206, China

<sup>3</sup>Corresponding author

other is that it does not impose any subjective weights on the multiple inputs and outputs. Therefore, it can be widely used in evaluating the participating nations at the competition.

In some papers, two inputs (i.e., GNP and population) and three outputs (i.e., total numbers of golden, silver, bronze medals each participating nation got) are taken into consideration to appraise the performance of each nation based on a classical DEA model with restricted weights [1]. They found out that all the participating nations showed the positive or negative trends in the five consecutive Olympics Games. They ranked the efficient nations by just counting the times that the efficient nations appeared in the reference sets of the inefficient nations. However, the case that the two efficient nations appeared in the reference sets the same times might occur. A zero-sum game DEA model was proposed and used to evaluate the performance of participating countries, in which the two inputs (GDP and population) were considered and weights were also restricted. But they ignored the improvement of the inefficient nations. Although it was considered it, the linear combination of efficient nations in the referee sets may be the unattainable goals for the inefficient nations [2]. Some scholars used a two-stage method to analyze the achievements of participating nations by linking the self-organizing mappings to DEA model. Here the input indicators were GDP per capita, population, disability adjusted life expectancy and index of equality of child survival. They categorized the participating nations into several groups. But the combination of data mining and ranking based on DEA may not be perfect.

The technique of vote-ranking was combined with the cross-evaluation methods to assess the performance of the participating nations at the competition. It could rank the participating nations effectively, but it could not provide an efficient target for the inefficient nations owing to the ignorance of the difference between the inefficient nation and its frontier target. As a result, the targets provided the proper benchmarks for the inefficient ones. As for the problem, the Context-dependent DEA (CRA-DEA) model which allows multiple constraint sets have to be considered to be employed. But the efficiency scores may not always lie between zero and unity [3]. Also there does not exist a model which considers both the integer problem and different constraint sets for different DMUs.

This paper mainly tackles the above problems. In section 2, an integer-valued DEA model was proposed to evaluate the performance of participating nations while taking into consideration different constraint sets for different DMUs. In section 3, the model was used to evaluate the participating nations of the 2012 Olympics, the integer-valued targets were given for inefficient participating nations and they were ranked. In addition, some comparisons were made to show the justice and feasibility of the model.

## 2. Materials and methods

In this section, we introduce an integer-valued DEA model with different constraint sets under variable returns-to-scale. Supposing there are no decision making units (DMUs), each representing a participating nation.

GDP per capita is a measure for attainable resource to train athletes, build and maintain training facilities, develop better training methods and so on. Compared with GDP, GDP per capita is a better indicator to show the economic power of the nation [4]. So we use GDP per capita as one of the inputs instead of GDP. Population size determines the pool from which potential athletes can be drawn, so it is another significant indicator explaining Olympic achievement. GDP per capita and population are the most important indicators expressing the economic and demographic power of nations. Therefore, we use population and GDP per capita as the two inputs, and the number of golden, silver and bronzes medals as the three outputs.

### *2.1. DEA model with different constraint sets*

Here we firstly introduce the output-oriented DEA model under variant returns-to-scale with the same constraint set. The model makes no exception to all the countries. If we use this model to evaluate the efficiency of the participating country, it seems somewhat improper. We need to refer to Cook and Zhu's CAR-DEA model to make Model 1 more suitable for the evaluation of all the participating nations at the competition.

Before introducing the DEA model with different constraint sets, we divide the participating nations of the competition into 4 groups according to the criteria from the World Bank. The first group includes the nations with the GDP below \$825 per capita. Under-developing countries from Africa and Middle-Asia belong to the group [5]. The second group contains the nations with GDP ranging from \$826 to \$3357 per capita. Some developing nations from Eastern Africa, Southern Africa and Eastern Europe are included in the group. The third group includes some developing and low developed countries from Middle-Europe, America and southern Africa with GDP ranging from \$3358 to \$10461 per capita [6]. The nations in the fourth group are well-developed, and most of them are from Western Europe and northern America with GDP above \$10462 per capita. Each group has its own constraint set in the form

$$c_{rL}^k \mu_r \leq \mu_l \leq c_{rU}^k \mu_r, \quad k = 1, 2, 3, 4, \quad r = 2, 3. \quad (1)$$

Here  $c_{2L}^k$  means that in the  $k$ th group, at least  $c_{2L}^k$  silver medals and not more than  $c_{2U}^k$  silver medals are equivalent to one golden medal. Symbols  $c_{3L}^k$  and  $c_{3U}^k$  can be interpreted in the same way. However, when we put all these different constraint sets for DMUs from different groups together, it may lead to unfeasible solution. As a result, some adjustments are made to tackle the problem. When  $\mu_r' = c_{rL}^k / c_{rL}^l \mu_r$ , the restrictions can be replaced by  $c_{rL}^l \mu_r' \leq \mu_l \leq (c_{rL}^l / c_{rL}^k) c_{rU}^k \mu_r'$  [7].

The restrictions for different nations have the same lower bound in this way. As for the common upper bound, it can be determined by  $\bar{c}_{rU} = \min \{ \bar{c}_{rU}^1, \bar{c}_{rU}^2, \bar{c}_{rU}^3, \bar{c}_{rU}^4 \}$  where  $\bar{c}_{rU}^k = (c_{rL}^1 / c_{rL}^k) c_{rU}^k$ ,  $k = 1, 2, 3, 4$ .

According to the above adjustment, the output-oriented CAR-DEA model was obtained [8].

$$\begin{aligned}
& \text{Min } \sum_{i=1}^2 v_i x_{ij} + \mu_0, \\
& \text{s.t. } \sum_{i=1}^2 v_i x_{ij_k} + \mu_0 - \sum_{r=1}^3 \mu_r \frac{c_{rL}^l}{c_{rL}^k} y_{rj_k} \geq 0, \quad , k = 1, 2, 3, 4, j_k \in J_k, \\
& \sum_{r=1}^3 \mu_r y_{rj} = 1, \\
& c_{rL} \mu_r \leq \mu_l \leq c_{rU} \mu_r, \quad r = 2, 3, \\
& \mu_r, v_i \geq 0 \quad \forall r, i, \mu_0 \text{ free.}
\end{aligned} \tag{2}$$

Here  $x_{ij_k}$  denotes the  $i$ th input of DMU $_{jj}$  from the  $k$ th group and  $y_{ij_k}$  denotes the  $t$ th output of the DMU. But the optimal values of Model 2 do not always exceed unity. Sometimes its optimal value is bigger than unity and sometimes it is smaller than unity. As a result, the efficiency scores which are the inverse of the optimal values do not always lie between zero and unity [9]. The reason for it is that we constrain  $\sum_{r=1}^s \mu_r y_{r0} = 1$ . As a matter of fact, the production frontier varies for DMUs from different group. In order to make the optimal value of the corresponding model bigger than unity, we have to make the weights of outputs consistent. As a result, we substitute  $\sum_{r=1}^s \mu_r y_{r0} = 1$  with  $\sum_{r=1}^3 \mu_r \frac{c_{rL}^l}{c_{rL}^k} y_{rj(k0)}$  to tackle the problem. Here comes our revised model.

$$\begin{aligned}
& \text{Min } \sum_{i=1}^2 v_i x_{ij} + \mu_0, \\
& \text{s.t. } \sum_{i=1}^2 v_i x_{ij_k} + \mu_0 - \sum_{r=1}^3 \mu_r \frac{c_{rL}^l}{c_{rL}^k} y_{rj_k} \geq 0, \quad , k = 1, 2, 3, 4, j_k \in J_k, \\
& \sum_{r=1}^3 \mu_r \frac{c_{rL}^l}{c_{rL}^k} y_{rj_{k0}} = 1, \\
& \bar{c}_{rL} \mu_r \leq \mu_l \leq \bar{c}_{rU} \mu_r, \quad r = 2, 3, \\
& \mu_r, v_i \geq 0 \quad \forall r, i, \mu_0 \text{ free.}
\end{aligned} \tag{3}$$

As for the determination of weights, such as, and so on, we have to conform to the classification of the nations. Then we can get Table 1 to show the range of each weight.

According to Table 1, we find out that countries in the fourth group value the golden medals most. In their eyes, at least 3 silver medals or at least 4 bronzes medals are equivalent to one golden medal. No more than 5 silver medals or 8 bronze medals are equivalent to one golden medal [10]. But nations in the first group even regard one silver medal or one bronze medal equivalent to one golden medal, which shows these nations do not care about whether the medals they obtain are golden or not. In their mind, any achievements at the competition are their pride. All these can

clearly demonstrate the different attitudes of different nations to the medals, which match the development situation of each nation well. As for the common ratio, we can obtain them based on the above adjustment scheme. Therefore, we can insert the ratio into Model 3. In order to elaborate the model, we can get its dual form.

Table 2. Inputs and outputs of participating nations of 2012 Olympics

Ratio	Bound	Group 1: $k = 1$	Group 2: $k = 2$	Group 3: $k = 3$	Group 4: $k = 4$	Common ratio
	Lower	1	1	2	3	1
	Upper	2	2	4	5	1.6667
	Lower	1	1	3	4	1
	Upper	2	2	6	8	2

**2.2. Integer-valued DEA model**

In the above models, the targets are not necessarily the whole numbers. Therefore, when the targets are not the whole numbers, they cannot work as the benchmarks for the inefficient DMU to improve its performance. Just rounding the number to the nearest whole number may lead to the overestimation or underestimation [11]. It does not make any difference to large nations, while it makes much difference to small nations. In order to tackle the problem, we propose an integer-valued DEA model: radial distance-based integer-valued DEA model (hereafter referred to as RDI model). Firstly, we introduce it in its input-oriented form.

We can elaborate it through Fig. 1.

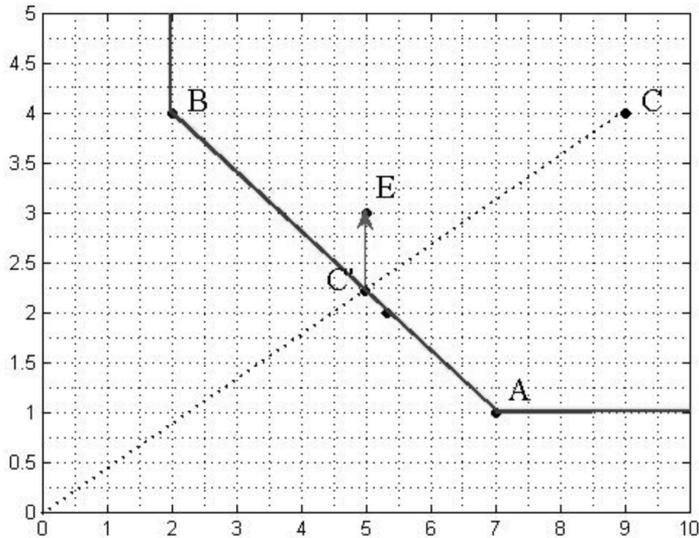


Fig. 1. Ways to find the optimal value based on the RDI model

It is noted that in the objective function of the RDI model, the objective of the slacks is searching for its minimal norm. The reason for it is that the smaller is, the closer to the production frontier the targets are. In the RDI model, the optimal is in the  $R^+$  direction of. In other words, in the RDI model, we first determine the point, and then search for based on (the green path in Fig. 1) [12]. Obviously, the optimal solution of is always obtained in the intersection of PPS and the radial direction line, which means the optimal obtained in the RDI model is the same as the one got in the CCR model.

In Fig. 1, C is the DMU under estimation. We can also describe the process of searching for the optimal point of the RDI model in two steps. First, we search for the intersection of PPS and the radial direction line (i.e., point C' in Fig. 1). Second, we start from the intersection point, search along the  $R^+$  direction for a certain integer-valued point with the minimal norm (i.e., point E in Figure 1), because any point in the  $R^+$  direction of the intersection point with a larger norm is much worse.

### 3. Results

We categorize all the participating nations of the 2012 London Olympics into four groups based on the criteria from the World Bank mentioned above. We use population and GDP per capita as the two inputs, and the number of golden, silver and bronzes medals as the three outputs.

Table 2. Inputs and outputs of participating nations of 2012 Olympics

Nation	GDP per capita (dollar)	Population	Golden medals	Silver medals	Bronze medals	Group
America	51601.37	313232000	46	29	29	4
Britain	41706.77	62698360	29	17	19	4
Russia	14541.39	138739900	24	25	33	4
South Korea	25080.82	48754660	13	8	7	4
Germany	43367.65	61471830	11	19	14	4
France	41136.59	65312250	11	11	12	4
Italy	34281.72	61016800	8	9	11	4
Hungary	12712.91	9976062	8	4	5	4
Australia	70494.16	21766710	7	16	12	4
Japan	47080.01	126475700	7	14	17	4
Republic of Kazakhstan	13111.22	15522370	7	1	5	4
Netherlands	48859.66	16847010	6	6	8	4
New Zealand	39964.46	4290347	5	3	5	4

We collect GDP per capita and population of each participating nation at the 2012 Olympics from the official website of the World Bank. The medals each participating nation obtained are gathered from the official website of the Olympics [13]. We can obtain Table 2.

There are six nations in the first group, seven nations in the second group, twenty-four nations in the third group and forty-eight nations in the fourth group. Therefore,  $J_1$ ,  $J_2$ ,  $J_3$  and  $J_4$  represent 6, 7, 24 and 48, respectively [14]. Then we can evaluate the participating nations at the 2012 Olympics based on the RDID model.

Table 3. Efficiency score and targets of each participating nation based on RDID model

Nation	$\eta$	Efficiency score	Target for golden medals	Target for silver medals	Target for bronze medals	Group
America	1	1	46	29	29	4
Britain	1	1	29	17	19	4
Russia	1	1	24	25	33	4
South Korea	1.47	0.69799	17	11	14	4
Germany	1.57	0.638127	28	18	22	4
France	1.92	0.519108	29	17	19	4
Italy	2.06	0.484438	24	15	18	4
Hungary	1	1	8	4	5	4
Australia	1	1	7	16	12	4
Japan	2.08	0.480042	25	23	30	4
Republic of Kazakhstan	1.39	0.719425	8	5	6	4
Netherlands	1.28	0.780009	8	9	8	4
New Zealand	1	1	5	3	5	4

The optimal values of all participating nations lie above unity. The efficiency scores are the inverse of the optimal scores and they all lie between zero and unity. The participating nations with the high efficiency score behave well at the 2012 Olympics. The targets of participating nations are all the whole numbers, which can be exactly used as the improvement benchmarks for the inefficient participating nations. The efficiency scores of the RDID model and targets of each nation are listed in the Table 4.

In the first group, the nations are under-developed. The GDP per capita of each nation in the group is the lowest in the four groups. They do not value the golden medals too much. There is only one nation whose efficiency score is unity, which denotes that there is no improvement for the nation. So its targets are the same as the medals it got. The efficiency score of Afghanistan is the lowest, which means there is much improvement for Afghanistan. Its targets are 4, 4 and 4, so it needs to

get four more golden medals, four more silver medals and three more silver medals to be efficient.

In the second group, there are no efficient participating nations. The efficiency score of Kenya is the highest and it only needs to get another two golden medals to be efficient based on the medals it obtained. Among the participating nations in the second group, the targets of Morocco are the highest and its medals is the lowest. Therefore, its efficiency score is the lowest.

In the third group, there are four efficient nations, namely China, Jamaica, Ukraine and Grenada. The efficiency score of Turkey is the lowest in the group and it needs to acquire eleven more golden medals to become efficient.

In the fourth group, there are seven efficient nations in the group, namely America, Britain, Russia, Hungary, Australia, New Zealand and Panama. The efficiency score of Argentina is the lowest, and it need to get another eleven golden medals, eight silver medals and nine bronze medals to become efficient. The nations in the group value the golden medals more than three other groups, which can be seen from the AR ranges in Table 1.

Next, we make a comparison between the efficiency scores of the RDID model and the RDI model with the same constraint set (hereafter referred to as the RDIS model). As for the RDIS model, we set and in order to make a clear comparison. And we rank the nations according to the optimal value in the ascending order. As for efficient nations, they are all ranked as the first placers. Nations in the first and second groups are underestimated through the RDID model, compared to the RDIS model. Because in the RDID model, the golden medals of these nations are considered less important than these in the RDIS model [15]. Nations in the third and fourth groups are overestimated through the RDID model, compared to the RDIS model. The reason for it is that the golden medals of these nations are considered more important in the RDID model than these in the RDIS model.

At last, we make a comparison between the targets we obtain and the targets of the BCC model. We find out that the targets we obtain through the RDID model are all the integer-valued, which can work as the benchmarks for the inefficient nations. But the targets through the BCC model are generally fractional, which cannot work as the targets of the inefficient nation effectively. We also find out that the targets provided by the RDID model are not always a rounding up or down of the fractional targets. For example, the targets of South Korea through the BCC model are (21.09, 12.98, 13.75), but its targets through the RDID model are (17, 11, 14). Even if the targets obtained by the BCC model are the whole numbers, they are worse than the targets of the RDID model, such as Azerbaijan. Its targets through the RDID model are (5, 4, 5), while its targets through the BCC model are (2, 2, 6). It is obvious that (5, 4, 5) are better than (2, 2, 6). The comparisons above show that the RDID model not only provide a proper integer-valued target for the inefficient nations but also evaluate the participating nations from different views.

Table 4. Comparison between the RDID and RDIS models

Nations	RDID	Ranking	RDIS	Ranking
America	1	1	1	1
Britain	1	1	1	1
Russia	1	1	1.25	6
South Korea	1.47	9	1.68	13
Germany	1.57	10	1.71	14
France	1.92	17	2.17	21
Italy	2.06	20	2.51	27
Hungary	1	1	1.000944	2
Australia	1	1	1	1
Japan	2.08	22	2.93	29
Republic of Kazakhstan	1.39	8	1.64	12
Netherlands	1.28	6	1.37	9
New Zealand	1	1	1.005449	3

## 4. Conclusion

The RDID model is mainly focused on to evaluate the participating nations at the Olympics. There are two main priorities for the model. Firstly, it can provide a more reasonable target for the inefficient nations because its targets are the whole numbers. People do not need to round the targets to the nearest whole number to get the integer-valued targets. It can provide proper targets for the inefficient nations. Secondly, it considers the different views of different participating nations to golden, silver and bronze medals and makes some adjustment to the original model to make the efficiency score lie between zero and unity. Moreover, the participating nations are divided into four group based on the criteria of the World Bank and give the proper constraint set to each group. Therefore, the RDID model can evaluate the participating nations from different view. And the comparisons between the RDID and RDIS models are made, and the RDID and BCC models again demonstrate the feasibility and justice of the RDID model. In conclusion, the model is instructive because the similar problems can be tackled through it.

## References

- [1] Y. WANG, K. KAKAMU: *Comment on "Measuring the performance of nations at beijing summer olympics using integer-valued DEA model"*. Journal of Sports Economics 17 (2016), No. 4, 418–422.
- [2] M. KHOVEYNI, R. ESLAMI, M. KHODABAKHSHI, G. R. JAHANSHALOO, F. HOSSEINZADEH LOTFI: *Recognizing strong and weak congestion slack based in data envelopment analysis*. Computers & Industrial Engineering 64 (2013), No. 2, 731–738.
- [3] F. MILLER, J. WANG, J. ZHU, Y. CHEN, J. HOCKENBERRY: *Investigation of the im-*

- Impact of the Massachusetts health care reform on hospital costs and quality of care.* Annals of Operations Research 250 (2017), No. 1, 129–146.
- [4] M. ABBASI, M. A. KAVIANI: *Operational efficiency-based ranking framework using uncertain DEA methods: An application to the cement industry in Iran.* Management Decision 54 (2016), No. 4, 902–928.
  - [5] M. TAVANA, A. K. YAZDI, M. SHIRI, J. RAPPAPORT: *An EFQM-Rembrandt excellence model based on the theory of displaced ideal.* Benchmarking: An International Journal 18 (2011), No. 5, 644–667.
  - [6] H. YAN, Q. WEI, G. HAO: *DEA models for resource reallocation and production input/output estimation.* European Journal of Operational Research 136 (2002), No. 1, 19–31.
  - [7] J. WU, J. CHU, Q. ZHU, P. YIN, L. LIANG: *DEA cross-efficiency evaluation based on satisfaction degree: an application to technology selection.* International Journal of Production Research 54 (2016), No. 20, 5990–6007.
  - [8] G. STORTI: *Minimum distance estimation of GARCH (1,1) models.* Computational Statistics & Data Analysis 51 (2006), No. 3, 1803–1821.
  - [9] F. ZHU: *Zero-inflated poisson and negative binomial integer-valued GARCH models.* Journal of Statistical Planning and Inference 142 (2012), No. 4, 826–839.
  - [10] A. HAFEZALKOTOB, E. HAJI-SAMI, H. OMRANI: *Robust DEA under discrete uncertain data: A case study of Iranian electricity distribution companies.* Journal of Industrial Engineering International 11, (2015), No. 2, 199–208.
  - [11] D. GUO, J. WU: *A complete ranking of DMUs with undesirable outputs using restrictions in DEA models.* Mathematical and Computer Modelling 58 (2013), Nos. 5–6, 1102–1109.
  - [12] A. ARJOMANDI, A. VALADKHANI, M. O'BRIEN: *Analysing banks' intermediation and operational performance using the Hicks–Moorsteen TFP index: The case of Iran.* Research in International Business and Finance 30 (2014), 111–125.
  - [13] S. LIM, J. ZHU: *A note on two-stage network DEA model: Frontier projection and duality.* European Journal of Operational Research 248 (2016), No. 1, 342–346.

Received June 29, 2017

# Insulation and monitoring system for pure electric vehicle based on microcontroller unit

HAN PENG<sup>1</sup>

**Abstract.** The purpose of this paper is to study insulation and monitoring system of pure electric vehicle. By insulating the performance decline of insulated wires, it is proved that pure electric vehicle based on micro controller unit will lead to serious consequence. The insulation detecting methods is used in the market to put forward an improved detecting method. Its characteristic refers to utilizing the bias resistance to detect insulation, while the main control chip provides signals to control the bias resistance. In the end, the peripheral circuit was designed centered on Advanced RISC Machine (ARM) micro controller unit (MCU for short). Modular programming is adopted to simplify the programming process. The experimental results show the accuracy and reliability of the insulation detection system by constructing an experimental platform. In the verification experiment, the insulation resistance is measured and its value with theoretical value  $100\text{k}\Omega$  is compared. The relative error is only 0.173%, and the standard deviation is 3.70. The insulation detecting system of pure electric vehicle designed this time is of certain accuracy and feasibility. Based on the the above finding, it is concluded that the design of insulation detection system is suitable for pure electric vehicle based on micro controller unit.

**Key words.** Micro controller unit, pure electric vehicle, insulation detecting system, main control chip.

## 1. Introduction

Environmental pollution caused by the fuel vehicles exhaust is becoming more and more severe owing to the increasingly expanding of automobile market. Faced with the existing energy-environment problems, the research, development and promotion of high-efficient clean electric vehicle are becoming an irresistible trend in automotive field [1]. However, most cars are fuel-powered, so they are free from the disasters caused by insulation. On the contrast, electric vehicle is electric-driven. To ensure enough power, there is always a high-voltage power supply assembled in the vehicle [2]. In the high voltage circuit of electric car, high-voltage and low-resistance

---

<sup>1</sup>School of Mechanical Engineering, North China University of Water Resource and Electric Power, 450045, China

will form heavy current. Therefore, the battery voltage of high voltage side of the electric vehicle should remain high-voltage (about 300 V). In addition, the positive and negative pole of the high-voltage battery are connected on the ends of the insulated wire, whose resistance is very low, so transient current of the high-voltage side loop will be very heavy [3].

As a result, the quality of the insulation system of the electric vehicle is of high importance. Insulation aging of the high-voltage side will pose a threat to drivers even the security of passengers.

## 2. Materials and methods

### 2.1. Design of insulation detecting system

Power system is very important to the electric car. Particular electric devices are schematically shown in Fig. 1. The requirements to voltage is different in different places, and meanwhile the high and low voltage side power system appeared [4]. High-voltage power system consists of four parts and drives the operation of high power devices like motor [5]. While the low-voltage power system supplies electric to electric devices, in order to avoid electric leakage and other issues. We emphasis on the high-voltage electric system in this research.

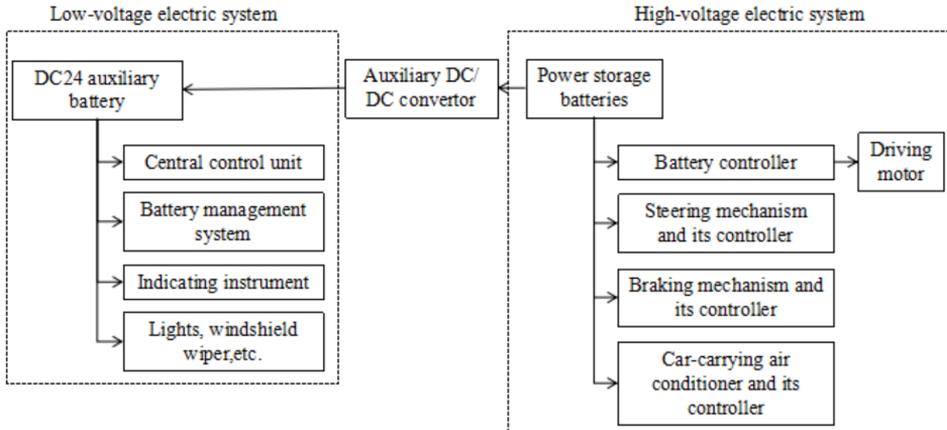


Fig. 1. Electric devices of electric vehicle

At present, the frequently used insulation detecting methods have bus end voltage method, AC signals injection method and DC voltage insulation measurement method [6]. The circuit used in bus end voltage method is simple but of unreliability, cannot influence fault like short circuit to ground. AC signals injection method refers to injecting AC signals in the DC system of the high-voltage side of electric vehicle, comparing, calculating and working out the insulation condition of the monitoring system. DC voltage insulation measurement method is realized by measuring the insulated resistance. Mosfet tube is used in the switch, so it is unavoidable for the

high current signals to produce some disturbing signals during the measuring. This will influence on the calculating accuracy [7].

## 2.2. Insulation detecting method

The insulation condition of electric vehicle is measured by the insulated resistance of DC positive and negative bus to the earth, according to the regulations of BS ISO 6469-1-2009: divide nominal voltage  $U$  of electric vehicle DC system by insulation resistance value, the result greater than  $100 \Omega/V$  conforms with the security requirement, the result lower than this value shows the insulation fault of pure electric vehicle [8].

The principle of measurement is shown in Fig. 2, in which,  $V_b$  is the storage battery voltage,  $R_p$  and  $R_n$  are the insulation resistances of the positive bus and negative bus on the earth, respectively. The exterior of the short dash box is the monitoring circuit model of the insulated resistance of the pure electric vehicle, in which  $R_0$  is the nominal bias resistance,  $R_0$ ,  $S_1$ ,  $S_2$  construct a bias resistance network.  $R_1$  and  $R_2$ ,  $R_3$  and  $R_4$  construct the measurement voltage division circuit,  $V_p$  and  $V_n$  represent the voltages of positive pole and negative pole to the earth, respectively. During measuring, turn off  $S_1$  and  $S_2$ , and draw  $V_p$  and  $V_n$ . And then according to the values of  $V_p$  and  $V_n$ , judge whether  $R_0$  is in parallel with  $R_p$  or in series with  $R_n$ .

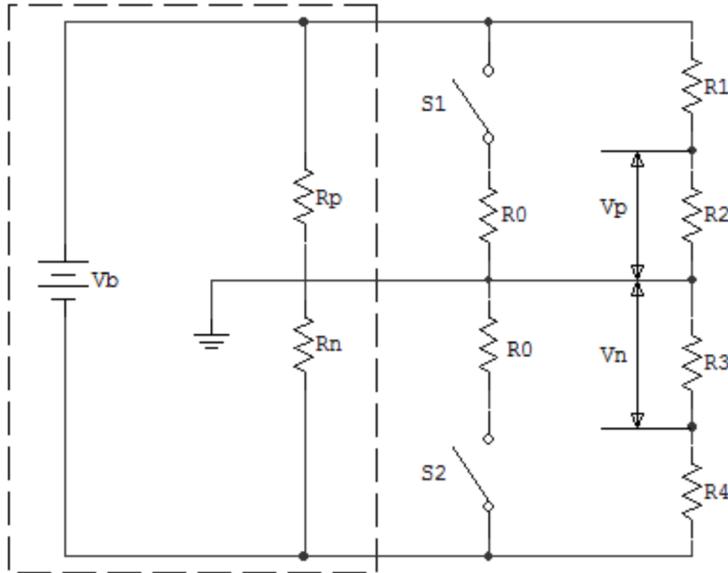


Fig. 2. Schematic circuit diagram of insulation detecting

If  $V_p > V_n$ , turn off  $S_1$  and disconnect  $S_2$ , then we can measure the voltage value of the positive and negative buses to earth  $V'_p$  and  $V'_n$ . According to the circuitous principle, we can work out the calculating formula of insulation resistance  $R_i$  of the

DC high-voltage system:

$$\frac{V_p}{R_p} = \frac{V_n}{R_n}, \quad \frac{V'_p}{R_p//R_0} = \frac{V_n}{R_n}. \quad (1)$$

Now we can draw that

$$R_n = R_0 \left( \frac{V_p V'_n}{V'_p V_n} - 1 \right). \quad (2)$$

As  $V_p > V_n$ , then  $R_p \geq R_n$ . Therefore, choose the lower resistance  $R_n$  as  $R_i$ . For the same reason, when  $V_p < V_n$ , the relative insulation resistance can be worked out.

### 3. Hardware design

We use the intelligent system of the high-voltage fault diagnosis and safety monitoring as the monitoring device of the pure electric vehicle this time, to reach a better effect in monitoring. Fig.3 refers to the systematic principle construction diagram. ARM chip is the core part of the whole controlling system. We apply with STM32F107 chip, whose detecting speed and accuracy are relatively higher. Under the control of the measuring main controller, this system can measure the accuracy of voltages to earth of the positive and negative bus, and is of good anti-interference quality and security [9]. As to the design of alarming system, based on the voltages to earth of positive and negative bus collected by the system, we can make comparative calculation with the standard value of the electric vehicle. When there is fault, it can give an alarm, or the system will turn off the circuit to protect the safety of the device [10].

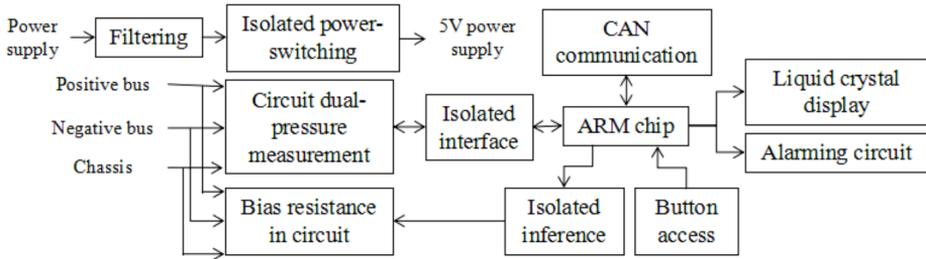


Fig. 3. Hardware principle construction diagram

Relative functions module circuits of the above hardware structure will be introduced in details as follows.

#### 3.1. Power circuit

Seeing at the security and practicability of design, two kinds of working power supply 24 V and 5 V are applied this time. Those power inputted by them should

be under relative protections like filtering and reversed connection-avoidance. Pure electric vehicle uses DC 24V, so we designed the 5V power circuit and switch it with the 24V power circuit [11].

### 3.2. Measuring circuit

When the electric vehicle is moving, the insulation detecting environment will be worse, thus the voltage of insulated end is high and unstable [12]. Voltage measured by STM32 is only several VA, so we apply TE6664N chip to construct the measuring circuit, using two circuits in measuring, one for collecting the voltages between ground wire and positive bus, the other one for collecting voltages between negative bus and ground wire. Moreover, high-pressure optronic relay can be used to control high-voltage circuit and the bias resistance access circuit. The voltages decreased after the signals pass the multilevel voltage division resistor [13].

### 3.3. Bias resistance access circuit

Figure 4 represents the schematic diagram of the access circuit of bias resistance R0. C\_Detect+ and C\_Detect- is from the controlling signals of STM32, and is equal to switch S1 and S2 in the schematic diagram. SGIELD is the ground wire, and U2 and U3 is optronics, playing the role of insulating to raise the stability of circuits. The resistance connected with the right end of optronics is bias resistance R0. Controlling signals output by the STM32 control R0 by controlling the continuity of U2 and U3, it can decide whether connecting R0 with the circuit of positive bus to the ground or with the negative one [14].

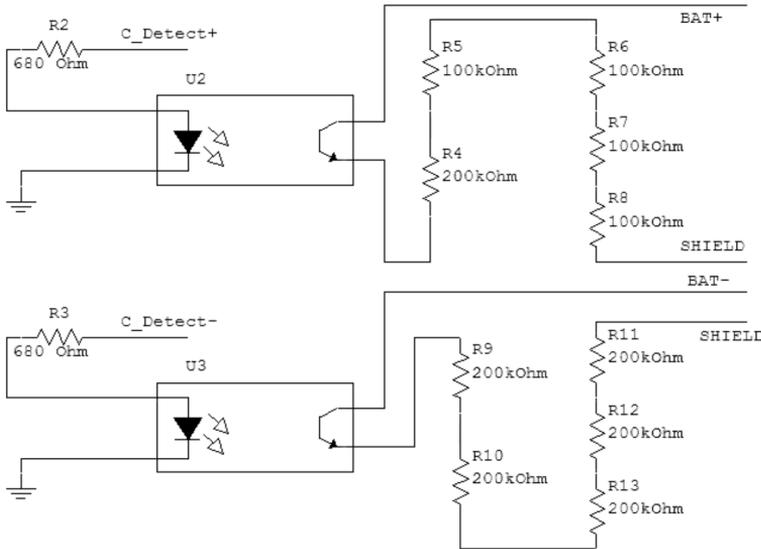


Fig. 4. Schematic diagram of bias resistance access circuit( BAT+ and BAT- refer to the positive bus end and negative bus end of the power storage battery respectively)



(2.8-cun, 16-bits TFT LCD, 320×240 resolution ratio), and CAN communication circuit (applying with module TD301DCAN) (not covered here).

Moreover, we also design the anti-interference quality and operating managing of the hardware. In the field of pure electric vehicle, high-voltage DC system was seen as strong electromagnetic interference. Therefore, its anti-interference quality will be set as a standard, whose value largely decides the reliability of the electric parameter, and also plays a vital role in the operating of the system. During the operating of high-voltage system, there are two levels of fault: minor fault and major fault, which is called class one failure and class two failure in software respectively. Fault in high-voltage loop belongs to class one, and fault in insulation resistance refers to class two. Faults resulted from the high temperature of power storage supply can be defined to class one failure and class two failure according to the extent of damages.

## 4. Software design

The core of control section of this research is STM32F107. Form a complete set of software by invoking firmware library to control each function module, thus to realize the monitoring to the insulation system, and at the same time, realize the functions of relay, PWM signals control, alarm display, CAN communication, sampling measurement controlling, insulated resistance calculating and so on. In this way, the principle of active insulation monitoring is realized. We can draw the resistance values of the positive and negative insulated resistance with relevant calculating method.

We adopts Keil uvision3 as the software development platform, and realize the programming of STM32F107 by invoking firmware library. The operating environment of the pure electric vehicle is complicated, so we add some protective circuits and anti-interference measurements to reinforce the reliability of the system, such as digital filtering technology and “watchdog” technique. The flow chart of main systematic program is shown in Fig. 7.

There are twice sampling of AD needed to be done, sampling two data signals every time, thus to draw the insulated resistance value. The first sampling is bias resistance and will not be involved in the whole system. Collect two data signals at the second time, work out the insulated resistance value of the bus insulated wire. The flow chart of AD sampling is in Fig. 8.

Fig. 9 refers to light alarm system. The positive and negative insulated resistance value calculated above is the condition to judge the light alarm condition. Figure 10 shows the flow chart of operating managing and disconnecting controlling, when in failure.

The pure electric vehicle may suffer from faults no matter in the starting, operating and parking process. When encountering fault, the monitoring system of high-voltage electricity will be access to the control strategy of disconnecting when in failure, thus to turn off the high-voltage loop.

During the reliability design and optimization of the systematic software, concerning with the functions needed to be realized as well as the whole process, the

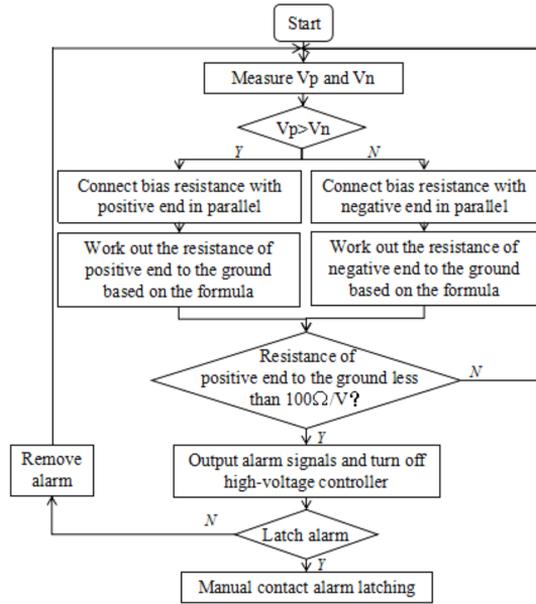


Fig. 7. Flow chart of main systematic program

linkage process of every motion must be fluent and natural during the whole insulation detecting. So we use optional structure “Switch” in the complete system. The connections between every action statement is accomplished by its multi-branch statements. All in all, frequently used modular programming method used in this time makes a higher systematic readability.

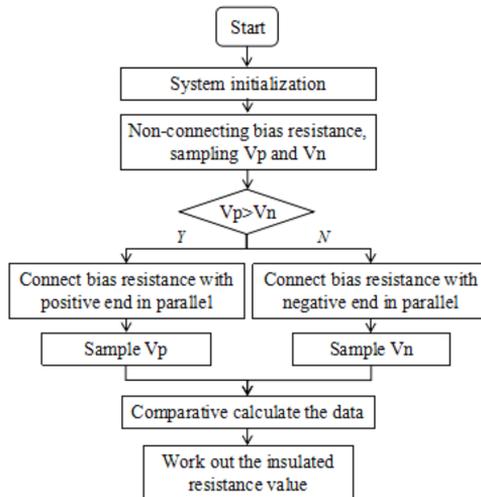


Fig. 8. Flow chart of AD sampling

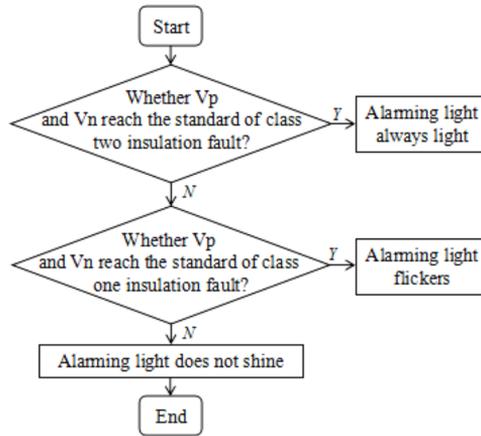


Fig. 9. Light alarm program

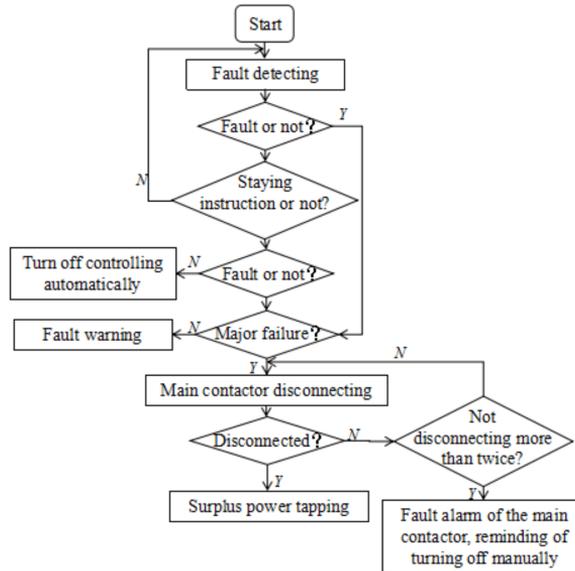


Fig. 10. Flow chart of operating managing and disconnecting controlling when in failure

## 5. Results

To judge the insulation condition of the insulated system, we construct the system test platform. The test platform constructed for the high-voltage safety test system of pure electric vehicle includes: 200 V high-voltage DC supply, 24 V DC supply, oscilloscope, multimeter, safety monitoring system, etc.

To ensure rationality and comprehensiveness of the test, our testing sequence is from partial to the whole, i.e. debugging software first and then the hardware,

debugging the low-voltage environment first and then the high-voltage. these are all the sequences to conduct test. The test results of insulated resistance value are shown in Table 1.

Table 1. Test results of insulated resistance value (kΩ)

	No.1	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10	Average value
Measured value	97	98	98	101	99	100	97	98	102	99.1
Standard value										
Error	-3	-2	-2	1	-1	0	-3	-2	2	

The standard value is:

$$S = \sqrt{\frac{\sum_{i=1}^n (s_i - \bar{s})^2}{n}} =$$

$$= \sqrt{\frac{3^2 + 1 + 2^2 + 2^2 + 1 + 1 + 0 + 3^2 + 2^2 + 2^2}{10}} \text{ k}\Omega = 3.70 \text{ k}\Omega.$$

Seeing the results in Table 1 and comparing the measured value of insulated resistance with theoretical value 100 kΩ, we can draw that the relative error is only 0.173%, and the standard deviation is 3.70. This is persuasive enough to prove the accuracy of the calculating module of the insulated resistance designed in this research.

## 6. Conclusion

We designed a detecting system based on ARM MCU directing at high voltage circuit safety monitoring on pure electric vehicle, and tested the feasibility of it. The main research process and result are as follows:

To put forward an improved insulation detection method concerning with the reasons why the high voltage side of electric vehicle need insulation detecting and also referring with the existing detecting methods in the market.

To design the schematics of many aspects like interposing the bias resistance, insulation detecting, high and low voltage protection, high voltage loop interlocking according to the principle of voltage detection. To show the whole hardware structure of the system. To compile a program that can realize the relative functions of hardware electric circuit based on Keil uVision 3 as the software developing platform, as well as the modular principles.

To construct the testing platform, and testify the feasibility and reliability of this theory and method.

## References

- [1] H. M. VASQUEZ, M. KUTTNER: *System and method for monitoring an electrical device*. Patent US9172233, Assignee: Early Rescue Solutions, LLC, Inventors: M. H. Vasquez, M. Kuttner (2015).
- [2] D. U. YONG, Q. SI: *Application of harmonic electric field method to  $\pm 800\text{kV}$  UHV DC composite insulator on-line detection*. High Voltage Engineering 38 (2012), No. 02, 382–386.
- [3] C. ZHOU, S. HU, W. SHA, Q. LIU, X. YU: *Active detection system of insulation resistance in electric vehicle*. Journal of Electronic Measurement and Instrumentation 27 (2013), No. 05, 409–414.
- [4] Y. DU, R. DING: *Detection of traffic sound for driving safety: Effects of car body sound insulation*. Noise Control Engineering Journal 62 (2014), No. 6, 436–448.
- [5] P. FABIAN, M. HAYNES, H. BABCOCK, M. HOOKER: *Characterization and qualification of cyanate ester/epoxy insulation for NSTX-U fusion magnets*. IEEE Transactions on Applied Superconductivity 23 (2013), No. 3, paper 7700204.
- [6] B. DONG, Y. T. TIAN, C. J. ZHOU: *Fuzzy logic-based optimal control method for energy management of pure electric vehicle*. Journal of Jilin University (Engineering and Technology Edition) 45 (2015), No. 2, 516–525.
- [7] J. YANG, Z. ZENG, Y. TANG, J. YAN, H. HE, Y. WU: *Load frequency control in isolated micro-grid with electrical vehicle based on multivariable generalized predictive theory*. Energies 8 (2015), No. 3, 2145–2164.
- [8] D. T. QIN, M. Y. YAO, S. J. CHEN, S. K. LYU: *Shifting process control for two-speed automated mechanical transmission of pure electric vehicles*. International Journal of Precision Engineering and Manufacturing 17 (2016), No. 5, 623–629.
- [9] P. YU, T. ZHANG, X. H. WANG, R. GUO: *Surge analysis and active-passive control for a central driven pure electric vehicle*. Journal of Vibration and Shock 34 (2015), No. 13, 53–59, 65.
- [10] Y. H. HUNG, Y. X. LIN, C. H. WU, S. Z. CHEN: *Mechatronics design and experimental verification of an electric-vehicle-based hybrid thermal management system*. Advances in Mechanical Engineering 8, (2016), No. 2, 1–9.
- [11] F. MACHADO, J. P. F. TROVÃO, C. H. ANTUNES: *Effectiveness of supercapacitors in pure electric vehicles using a hybrid metaheuristic approach*. IEEE Transactions on Vehicular Technology 65 (2015), No. 1, 29–36.
- [12] D. SAVITSKI, V. IVANOV, B. SHYROKAU, J. D. SMET, J. THEUNISSEN: *Experimental study on continuous abs operation in pure regenerative mode for full electric vehicle*. SAE International Journal of Passenger Cars - Mechanical Systems 8 (2015), No. 1, 364–369.
- [13] F. CHEN, Y. SUN, C. ZHAO, J. YANG, H. YE, J. ZHU, Q. TANG: *Elastic pipeline design of high performance micro-controller YL8MCU for signal processing of digital home appliances*. Open Automation and Control Systems Journal 7 (2015), No. 1, 863–872.
- [14] S. M. KUMAR, V. K. GOBINATH, D. R. P. RAJARATHNAM, D. JAYANTH, P. S. KUMAR, D. NANDAGOPAL: *Automatic LID controller for laptop using micro-controller*. International Journal of Applied Engineering Research 10 (2015), No. 93, 261–263.
- [15] G. M. MA, Z. WU, H. Y. ZHOU, J. JIANG, W. X. CHEN, S. S. ZHENG, C. R. LI, X. LI: *A wireless and passive on-line temperature monitoring system for GIS based on surface acoustic wave sensor*. IEEE Transactions on Power Delivery 31 (2016), No. 3, 1270–1279.

Received June 29, 2017



# Dynamics and simulation analysis of table tennis robot based on independent joint control

YANG YU<sup>1</sup>

**Abstract.** The purpose of this paper is to prove the accuracy and flexibility of the speed of the table tennis robot when it strikes. The method is to analyze table tennis mechanical simulation based on the control strategy of robot independent joint. Independent control strategy is adopted at first, and PID control principle is also used to design and analyze the control system of table tennis robot and mechanical-control co-simulation system. In addition, the virtual scene is simulated when two robots are playing. Finally, performance index of robots is evaluated and optimized to confirm proper parameter of sample robot design and provide theory instructions for the developing and producing of table tennis robot. According to control demands of servo motor, the three loop control system of table tennis is established. The experimental results show that the co-simulation and control system of table tennis prove its correctness. Based on the above finding, it is concluded that the mechanical parameters of hybrid robot can be optimized. Meanwhile, simulation and analyses in the study of table tennis robot based on the control system of independent joint control strategy play an important part in developing and producing hybrid robot 3 robot programming system (RPS) + robot programming (RP).

**Key words.** Table tennis robot, PID control system, mechanical simulation, control strategy of independent joint.

## 1. Introduction

### *1.1. Description of the problem*

Robot is one of the important inventions in the 20 century. Table tennis robot is of wide interest and also a universal technology platform combining machinery, vision, control and other disciplines [1]. In terms of technology, crucial technology in table tennis can directly radiate to all kinds of related fields; in terms of teaching, to develop robots which can play table tennis with people or play with each other is useful for activating students' passion and enthusiasm for exploring scientific knowledge; In terms of recreation and sports, table tennis robot can be used in such fields

---

<sup>1</sup>Department of P. E, North China University of Water Resources and Electric Power, China

as entertainment, exercise, partner training, competition, which can enrich people's free time and help them build body health [2–3]. Thus, the research of table tennis is important in the filed of intelligent robot research.

### ***1.2. State of the art***

Scholars both at home and abroad have made many researches about table tennis robot. Table tennis robot shows rapid development in advanced countries, while China began to do researches about it in a relatively late time. Some scholars make deep researches on vision prediction and servo control system of table tennis [4]. There are some studies on visual tracking, trajectory prediction and control technology. Studies on the trajectory prediction of table tennis are carried out [5].

A great deal of mechanical simulations are made at home and abroad. Some simulation researches on decomposition control of arm accelerating speed of table tennis robot are analyzed [6]. Experts make researches on robust control of operation arm of table tennis with anti flexible joint [7]. In the robot industry in the 21 century, there is trend that product design is completed with the help of computer simulation [8]. As lots of practical experiments are carried out after the bases of a series of virtual simulation experiments, which improve the efficiency and reduce risks and costs, thus design defects of the product are dealt with in time before the shaping of physical objects [9–10]. It also goes to the mechanical simulation of table tennis simulation. The study on mechanical simulation offer important instructions for the developing and producing of real simple machine. For reasons above, table tennis robot is chosen as the experiment platform in the study for simulation and analyses to design and simulate the mechanical system and control system of table tennis robot.

## **2. Materials and methods**

Dynamics of table tennis robot is analyzed and dynamic model for hybrid robot 3-RPS+RP and 6R serial robot is established before the research, and the trajectory of its striking ball is programmed; next, the dynamic model is proven accurate by theory calculation and mutual simulation, which lay a foundation for the simulation in the environment of striking balls of table tennis and its control system.

### ***2.1. The control system of table tennis robot***

Hybrid robot 3-RPS+RP is the study is controlled by PID control with single joint, whose body is connected by parallel mechanism and serial branches. Robot control is realized by driving joint of the motor, and the AC servo motor control model is used to drive the joints of table tennis robot. Compared with DC servo motor, there are many advantages of AC servo motor: it has no brush and commutator which ensures reliable operation and low demands for maintenance; the heat dissipation of the stator winding is more convenient; it has small inertia, which is easier to improve the speed of the system; it has relatively small volume and weight in the

same power [11]. The servo motor adopts the traditional PID controller, which is of good stability, simple procedure for parameter setting and excellent robustness. PID control principle [12] is a reasonable algorithm based on imprecise model and system information estimation. The principle of PID control system is shown in Fig. 1.

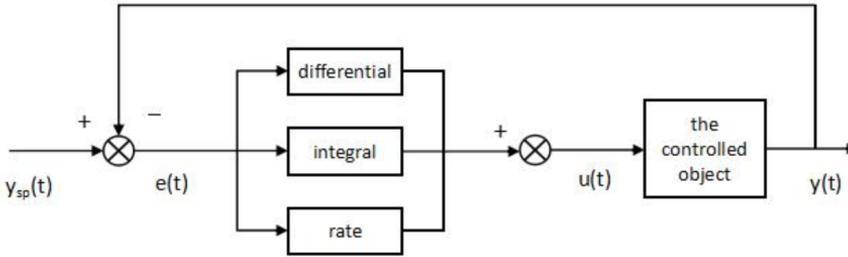


Fig. 1. Principle of PID control system

System shown in Figure 1 mainly consists of PID controller and the controlled device. It composes control deviations according theoretical preset value and actual output value, and make up correspondent control portfolio  $u(t)$  with deviations composed in different combination forms of PID.

AC servo control system of joint control of the robot often uses frequency converter as the driver of the motor, and it is composed of three closed loop control system, namely, position loop, speed loop and current loop [13]. The function of current loop is to adjust current to maintain constant output of torque, refrain current deviation and improve the performance of the system by enhancing the stability of current; the function of speed loop is to refrain load disturbance and velocity perturbation and effectively overcome dynamic errors and shorten adjustment time; the function of position loop is to ensure the system can position and follow in a correct way. The control principal and structure model of AC servo motor with three closed loops is shown in Fig. 2.

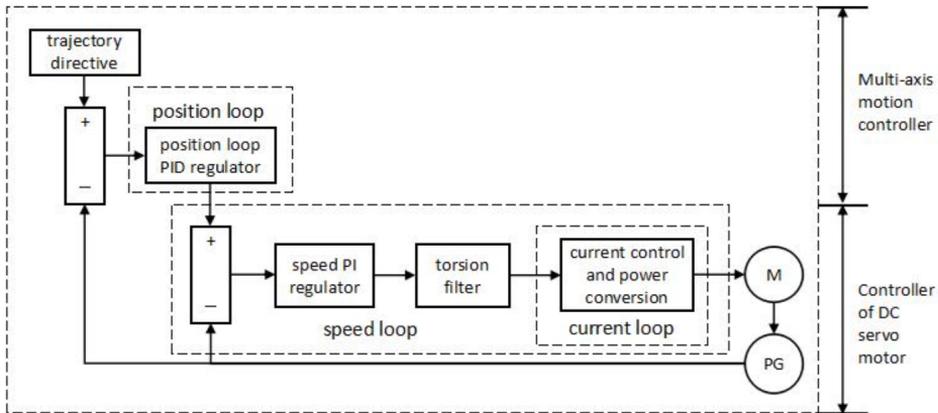


Fig. 2. Control model of AC servo motor with three closed loops

## 2.2. Mechanical-control co-simulation system of table tennis

As far as the mechanical-control co-simulation system of table tennis is concerned, the position and speed control co-simulation based on kinematics and control co-simulation with dynamics as the basic force are introduced in the study. The principle of the position and speed control co-simulation based on kinematics is to develop robot kinematics inverse solution and obtain joint position and information about speed, and to transform position information into information about joint driving torque through control system so as to drive the robot [14]. Information about force or torque between motor shaft, moving and rotating shaft is transferred by connecting ball screw, coupling and the robot. At this time, input and output of robot co-simulation system reflects robot kinematics relation. Its model is shown in Figs. 3 and 4.

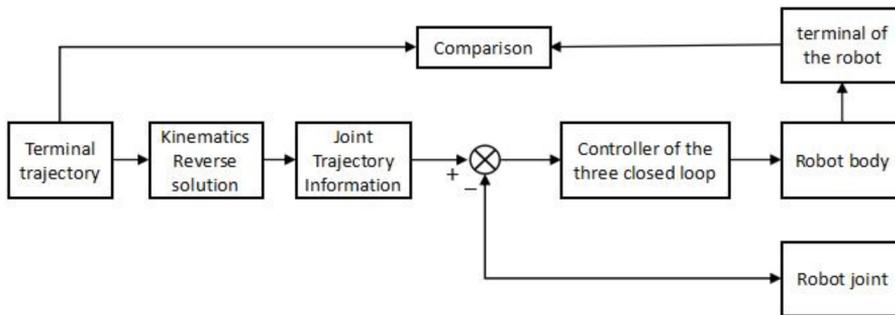


Fig. 3. Model of co-simulation system

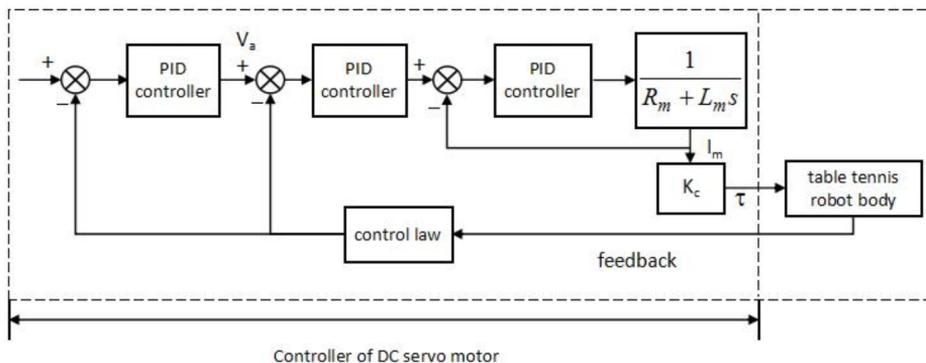


Fig. 4. Driving model of AC servo motor

The principle of control and co-simulation system based on kinematics. The principle of the control co-simulation with dynamics as the basic force is to develop inverse resolution through terminal position information by using robot dynamics so as to get information about joint force and torque, and then to connect each joint of

the robot through joint driving force and torque parameters after the output of the control system so as to drive the robot. At this time, co-simulation system of input and output robot reflects its dynamic relation. Its principle of the control system is shown in Fig. 5.

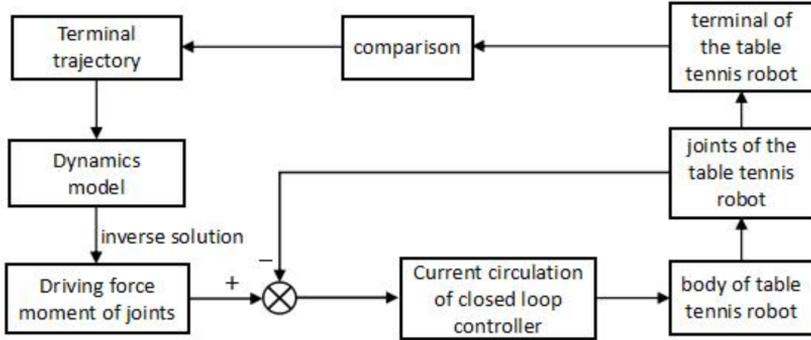


Fig. 5. Principle of dynamic control co-simulation system

What is shown in Fig.5 is the co-simulation model of robot with semi-closed loop. The features of table tennis robot are combined in the study. Terminal trajectory of the robot is received after developing inverse resolution to obtain joint information and controlling the robot through control principal of joint closed loop based on the tracking what has been programmed. Comparative analyses are made, in the following, between the terminal trajectory and programmed trajectory so as to commend on motion characteristics of the robot.

### 3. Experimental results and discussion

#### 3.1. *Virtual simulation environment when two table tennis robots are playing*

The virtual simulation environment mainly concludes robot body, serial 6R robot and hybrid robot 3-RPS+RP; the environment of two robots are playing with each other includes striking balls trajectory of robots, motion and collision model of table tennis; virtual control system of the robot, Paul three-loop servo control system, is set. Focus on the study is scenes simulation which requires bodies of the playing table tennis robot, sticking ball environment for the playing robots and its controller.

Virtual scene where two robots are playing can be established. The environment for double play is closely related with the motion model of the robot and table tennis. The stress state of the table tennis may affect the speed and accuracy of the robot's striking on balls. Terminal position of the mechanical arm is developed to joints by inverse resolution to simulate robots playing. Effects on the robot made by elements in striking ball environment are [15]: pat speed, pat position, direction the pat surface faces when it touch the ball and different trajectories of striking ball. In a word, the motion of the table tennis robot decides the motion of the ball,

and moving features of the ball affects the motion of table tennis robot. Mechanic-control co-simulation model of table tennis robot is shown in Fig.6, and results of simulation model is shown in Fig.6 and 7.

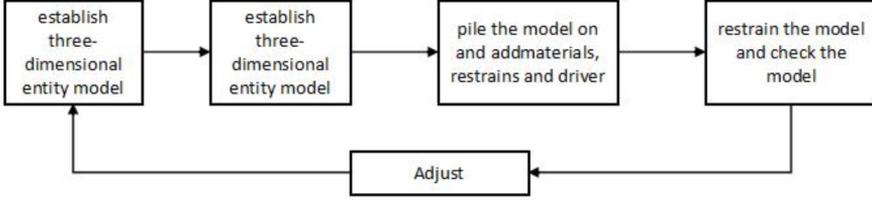


Fig. 6. Flow chart of the establishment of mechanical model

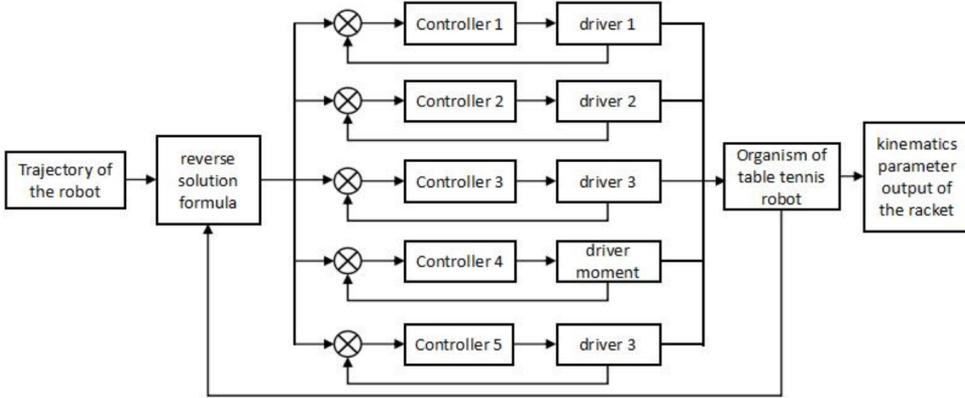


Fig. 7. Simulation system model

### 3.2. Evaluation and optimization of performance of the table tennis robot

Mechanical performance index is the important element to evaluate robot. Different indexes serve for evaluating robots with different tasks, which reflect the performance of robots. Table tennis robot with the task to striking on balls is evaluated from the perspective of spatial coincidence degree, speed response characteristics and speed stability. Spatial coincidence degree stands for the coincidence rate of working space the robot needs in practice and the design requires for, namely, volume percentage. The design demand for space when hybrid robot works and realizing parameters are shown in Table 1.

When there is no pose motion at the end of the robot, the working space is a cylinder, but the actual working space is bigger than a cylinder. It is concluded through calculation that space coincidence rate of table tennis robot is far over 1, meeting the working space requirement.

Velocity response characteristic stands for the speed of striking balls at the end of

the robot and acceleration characteristic when the robot is striking balls. The speed parameters the robot strikes the ball are shown in Table 2. The characteristic is that the results of accelerate speed of terminal of the robot show its overall non-smooth skipping and the skipping range of the second half is particularly large, thus it needs to be optimized.

Table 1. Test results of insulated resistance value ( $k\Omega$ )

Design area	Up and down working field (mm)	Horizontal working area (mm)	Longitudinal work area (mm)
Estimated parameters	$\pm 250$	1600	450
Realization method	Telescopic rod rotation	Telescopic rod circle rotation	Telescopic rod moving

Table 2. Speed parameters of striking balls

Environmental parameters	Full speed (m/s)	Rotation speed (r/s)	Maximum deformation (%)
Ball	0~27	0~172	27
Ball pat	5~12	/	seldom

ADA MS/Vibration vibration analysis module is adopted to analyze vibration in the study. When the response of the shock excitation is close to 2.5 Hz, the result of frequency response shows that the frequency where there is maximum acceleration response is near 2.7 Hz to 32.9 Hz and resonance can be avoided.

Works about the optimization of the table tennis robot are carried about from two aspects, namely, control and machinery. In terms of the control, parameters of the controller is regulated, control performance is improved and control evaluation index is stepped up, thus enhancing speed response characteristics of the robot; As for machinery, from the aspect of decreasing inertia, light materials meeting striking ball tasks are chosen to reduce inertia of mechanical system; linear motor driver is selected which means inertia of ball screw is ignored, and dynamics equation is simplified, thus making the robot more flexible. The diagram of optimization process is shown in Fig. 8.

Tests are carried out on the optimized robot. It can be found that its speed steps up, the acceleration stability is better than before and its stability improves too. And meanwhile, the impact force in the experiment produced by the ball and ball pat in the striking ball scene is 80.7493 N, which is about 102 N in practice coinciding with the actual impacting environment. Finally, some relatively ideal mechanical data about the robot is shown in Table 3.

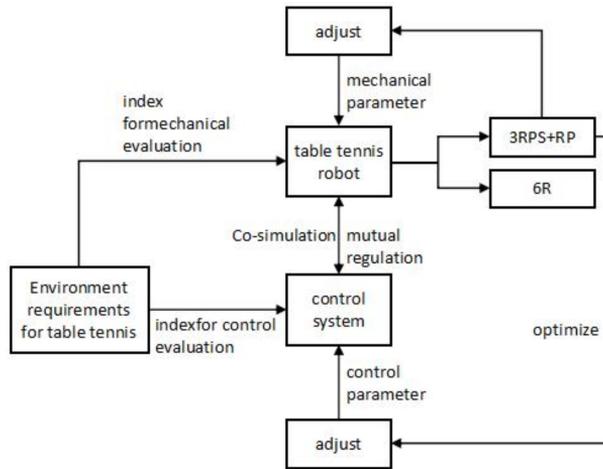


Fig. 8. Diagram of optimization process

Table 3. Parameters of mechanical system

Component name	Size (mm)	Component name	Size (mm)
Rectangular base	510×357×102 (L×D×H)	Elbow	Diameter: 62, Turning diameter: 306
Cylinder base support	Length:620, diameter:62	Double racket	Diameter: 153, thickness: 10
Fixed platform	Diameter: 410	Racket	Length: 367
Moving platform	Diameter: 305	Swigging are of the serial branch	46×46×470
Upper branch	Diameter: 34, length:494	Screw of the paralleling branch	Lead: Ph=15, Nominal diameter: d0 = 15, the distance is 400
Lower branch	Diameter: 56, length: 530	Swinging arm screw shaft of the serial branch	Lead: Ph = 12, Nominal diameter: d0 = 8, the distance is 400

## 4. Conclusion

The control system and mechanical-control co-simulation system of table tennis robot are studied. The performance of it when two robots are playing in the virtual simulation environment is evaluated and optimized. It is concluded, finally, that, three-loop control system of the robot is established according to the control demand of servo motor, and that co-simulation model of hybrid robot is created, which mean

mechanical-control simulation system is realized and it is feasible; simulation environment of double play between robots is created, methodologies for the evaluation and optimization for the whole system is put forward. In addition, hybrid robot with five degrees of freedom is optimized and relatively ideal parameters for the system are known, which lay a good theoretical foundation for the control system of the table tennis and its producing.

However, there are some limitations in the study due to limited experiment time and equipment. The further research will be carried out from the following two aspects: the mechanical-control co-simulation for the robot can be made in the way of adopting the driving combining linear motor and rotating servo motor. Furthermore, comparative analyses will be made between it and response from the robot driven by rotating servo motor. Through appropriately simplifying its dynamic equation, real-time control system with feedback closed loop is realized after the mechanical-control co-simulation.

## References

- [1] Y. HAYAKAWA, A. NAKASHIMA, S. ITOH, Y. NAKAI: *Ball trajectory planning in serving task for table tennis robot*. SICE Journal of Control, Measurement, and System Integration 9 (2016), No. 2, 50–59.
- [2] K. MÜLLING, J. KOBER, O. KROEMER, J. PETERS: *Learning to select and generalize striking movements in robot table tennis*. International Journal of Robotics Research 32 (2013), No. 3, 263–279.
- [3] G. MUNIVRANA, L. Z. PETRINOVIĆ, M. KONDRIĆ: *Structural analysis of technical-tactical elements in table tennis and their role in different playing zones*. Journal of Human Kinetics 47 (2015), No. 1, 197–214.
- [4] C. CHARBONNIER, S. CHAGUÉ, F. C. KOLO, A. LÄDERMANN: *Shoulder motion during tennis serve: Dynamic and radiological evaluation based on motion capture and magnetic resonance imaging*. International Journal of Computer Assisted Radiology and Surgery 10 (2015), No. 8, 1289–1297.
- [5] C. LIU, Y. HAYAKAWA, A. NAKASHIMA: *Racket control for a table tennis robot to return a ball*. SICE Journal of Control, Measurement, and System Integration 6 (2013), No. 4, 259–266.
- [6] M. LIU, Q. LI: *Analysis on the Application of Region-growing Algorithm in Table Tennis Trajectory Simulation*. Research Journal of Applied Sciences, Engineering and Technology 5 (2013), No. 9, 2779–2785.
- [7] K. MÜLLING, J. KOBER, O. KROEMER, J. PETERS: *Learning to select and generalize striking movements in robot table tennis*. The International Journal of Robotics Research 32 (2013), No. 03, 263–279.
- [8] D. SUN: *Neural network-based table tennis competition technique and tactics diagnosis and evaluation*. Open Cybernetics & Systemics Journal 9 (2015), No. 1, 1722–1727.
- [9] R. SINGH, P. N. PANDEY: *Molecular docking and molecular dynamics study on SmH-DAC1 to identify potential lead compounds against schistosomiasis*. Molecular Biology Reports 42 (2015), No. 3, 689–698.
- [10] R. KRUG, D. DIMITROV: *Model predictive motion control based on generalized dynamical movement primitives*. Journal of Intelligent & Robotic Systems 77, (2015), No. 1, 17–35.
- [11] R. M. HOU, R. Z. LIU, Q. GAO, L. WANG, T. B. DENG: *Application of adaptive fuzzy wavelet neural network in AC servo control system*. Acta Armamentarii 36 (2015), No. 05, 781–788.

- [12] F. CHÉNIER, P. BIGRAS, R. AISSAOUI: *A new dynamic model of the wheelchair propulsion on straight and curvilinear level-ground paths*. *Computer Methods in Biomechanics & Biomedical Engineering* 18 (2015), No. 10, 1031–1043.
- [13] A. MAHAPATRA, K. BISWAS, A. KUMAR, A. CHATTERJEE: *Motion control strategies based on PD control for a four degree-of-freedom serial robotic manipulator to mimic human index finger articulations*. *International Journal of Computer Applications* 109 (2015), No. 13, 35–42.
- [14] S. U. HU, X. U. DE, G. D. CHEN, Z. J. FANG: *Striking position selection based on two-step multi-purpose fuzzy decision method for robotic table tennis*. *Control Theory & Applications* 30 (2013), No. 05, 597–603.
- [15] K. WANG, W. WANG, H. ZHANG: *The mechanical properties of a wall-climbing caterpillar robot: Analysis and experiment*. *International Journal of Advanced Robotic Systems* 10 (2013), No. 1, 257–271.

Received June 29, 2017

# Security dynamic performance evaluation method of ad hoc network based on anti-attack model

XIAOLING XU<sup>1</sup>, WEICUI<sup>2,3</sup>, TIANYUN ZHANG<sup>1</sup>

**Abstract.** The purpose of this paper is to study the security of Ad hoc networks. The Ad hoc network is decentralized and self-organized. It has the unique characteristics of topological dynamic change and fully open wireless access. They are easily attacked by various network malicious nodes. Based on the anti - attack model, the security of Ad hoc network is analyzed. By using the model, the corresponding trust filtering mechanism is established. It can detect nodes stably. At the same time, the recommended nodes and nodes with collusion attack are found in the link. Through the theoretical analysis and experimental simulation, the results show that the proposed model can avoid the attack of malicious nodes, so as to realize the stable operation of the whole system. Based on the above findings, we conclude that this model can effectively solve the trust management problem in Ad hoc networks.

**Key words.** Anti-attack model, Ad hoc, network security, trust in peers.

## 1. Introduction

Ad hoc network is a kind of temporary autonomous distributed system in terms of its system, which shows a lot of natural disadvantages in the practical application. There are many problems that need to be solved urgently, among which the problem of network security is remarkable since without solution, it would be difficult to achieve the universal application of network, while the security of network node has always been one of the important issues conducted in research [1]. With the Ad hoc network model continuing to enter into thousands of households, how to better improve the security of Ad hoc system is one of the main problems that is in need to be solved. Some studies have indicated that the real-time monitoring of node control and the interception of malicious nodes in the network are gradually becoming the main methods to enhance the service capability of Ad hoc network [2].

---

<sup>1</sup>Information Network Center, Lanzhou City University, Lanzhou, 730070, China

<sup>2</sup>Educational Administration Office, Lanzhou Jiaotong University, Lanzhou, 730070, China

<sup>3</sup>Corresponding author

Meanwhile, the anti-attack model has gradually become the mainstream of research. This model, in the process of research, with direct access to the communication nodes in the whole system, ensures the smooth development of the entire system, which follows a specialized trust evaluation of nodes [3].

Usually, the wireless communication network mainly refers to the cellular mobile communication system or the wireless. The former usually adopts centralized control mode, and needs related infrastructure support, such as base station or switch. The latter is connected to the fixed network in a wireless access mode. What they all have in common is that mobile terminals require fixed network support. However, these are not suitable for certain occasions. Therefore, we urgently need a ubiquitous network, that is, mobile Ad hoc networks. Mobile Ad hoc network is a new type of wireless network. It does not depend on the control of the fixed infrastructure, but emphasizes the concept of multi hop, self-organization and no center. Nodes in a network have dual roles. It has both router function and host characteristics, which are connected as peer entities. The normal communication process between nodes must be relayed by other nodes in the network, so as to construct a multi hop wireless network. Although mobile Ad hoc network has many advantages, such as fast, flexible and strong invulnerability, it has many inherent security problems, such as no linearity, dynamics, self-organization and resource limitation. Therefore, its security has become a serious challenge for mobile Ad hoc network research. As a kind of computer network, the network also has to solve the basic network security problems, such as the availability of basic network services, the confidentiality of information and the integrity of network information. Authentication, communication, encryption and authorization are the main means to achieve the above goals. The specific methods include authentication protocol, digital signature, information encryption and so on. These methods still play an important role in ensuring the availability, integrity, authentication, and non-repudiation of network communications. However, these mechanisms require nodes with higher hardware requirements. As a result, its actual operation is relatively inefficient. In addition, security mechanisms in centralized networks are not suitable for mobile Ad hoc networks. It cannot prevent unsafe behavior, malicious behavior and selfish behavior in the network. In this network, the mutual trust between nodes is of great importance to the security and reliable operation of the network. The establishment of trust relation is the premise of normal communication. In order to establish security communication, a perfect trust model must be established among the nodes, and it is one of the main means to solve malicious attacks in the network by determining the trust relationship between nodes. Therefore, the research of mobile Ad hoc network trust model is a direction to solve the problem of network soft security. However, mobile Ad hoc network is a resource constrained autonomous system. Because nodes are controlled and used by many rational users, the nodes in the network often show their privacy. In the normal packet forwarding process, it is not willing to actively provide services for other nodes, which results in a dramatic decline in network performance. Therefore, in this environment, it is another direction to solve the problem of network soft security by motivating and mobilizing the cooperation enthusiasm of selfish nodes. At present, the research on secure routing

of mobile Ad hoc networks is based on table driven, on-demand and hybrid routing protocols. Security enhancement is carried out by using the idea of hard security. The introduction of trust and cooperation mechanisms provides a new approach for solving the soft routing problem in mobile Ad hoc networks.

The security protection of network is achieved in trust evaluation where the system attacks directly malicious nodes for the invasion into the system, thereby expelling these nodes out of the system. After the relevant research and evaluation of this problem, the preliminary analysis shows that in the network communication security, if the trust is increased or remain unchanged among these nodes with better behaviors, then the other nodes in the network will be more willing to carry out communication [4], on the contrary, the trust of these nodes with poor behaviors will continue to decline, and the other nodes will exclude or even refuse the communication with them. Where, the process seems to imitate the new relationship of new person in human society, with its operation model involving hypothesis, expectations and the corresponding behavior and environment, which are more appropriate to conform to the human social relations [5]. At the same time, the trust evaluation model is also an important driving force of the whole anti-attack model, which prevents the network from being attacked internally, and is suitable for distributed network communication [6]. In the actual use, it can strengthen the detection of malicious nodes to improve the security of the system [7], so the trust evaluation system has also gradually become the focus in researches on the security of Ad hoc network based on anti-attack model [8]. With the continuous development and progress of Ad hoc network's applications, how to establish the trust relationship between unrelated individuals is also the most important issue [9].

## 2. State of the art

In the simulation of Ad Hoc network, most people use mobile model (MM) based on single entity, such as RWPM (random waypoint mobility model), RWM (random walk mobility model), RDM (random direction mobility model) and GMMM (Gauss–Markov mobility model). RWPM is the basic model, and the most commonly used MM. In RWPM, the MN randomly takes the starting point S and the destination point D in the moving area, and randomly takes the velocity  $v$  and moves from S to D in a straight line. After MN arrives at D, randomly select a pause time to remain stationary, thus completing a step. Then, the next round step movement is performed with the present round destination point D as the start point S of the next movement, and the other MNs follow the above-described movement pattern and move independently of each other. Other models are mostly evolved from RWPM, when the pause time of MN in RWPM is zero, RWPM becomes RWM. RWM will produce unrealistic dynamic characteristics, such as sudden stop and sharp turn. RDM reduces the node density fluctuation on the basis of RWPM, but it cannot improve the network topology connectivity performance under RWPM. GMMM is based on the use of the RWPM Markov chain to establish the movement model, it attempts to describe a more realistic movement pattern, but due to MN in some direction cannot change the speed of movement, resulting in the dynamic

characteristics of the network is still flawed (such as node location distribution, etc.).

The EMWMSDP evaluation method proposed in this paper can quantitatively evaluate the temporal and spatial dynamic characteristics of interrelated MM. On this basis, the CCMM model is also proposed to describe the curve movement model with realistic scene. For simplicity, it is assumed that the MN moves within a circular region, but it can be extended to other moving regions (such as rectangles) and can quantitatively compare the temporal and spatial dynamics of different MM. Because the method is independent of any MM, the results of Ad Hoc network-related research (such as protocol optimization and performance evaluation) under different MMs are comparable and the conclusion is more reliable. In addition, the EMWMSDP method can evaluate the temporal and spatial dynamic characteristics of MM more comprehensively and make up for the deficiency of single dynamic characteristic research. In the Ad Hoc network, the network model describes and reflects the geometric characteristics of the mobile node (MN), such as the mode of motion, speed and direction, which is the main factor of the network topology change. In the existing research of Ad Hoc networks (such as MAC protocol and routing protocol optimization, network performance prediction, etc.) are often based on a specific model for the premise. The network model makes the Ad Hoc network show complex dynamic characteristics (such as the spatial distribution of node location, the duration of network link connectivity, etc.) both in time and space. Most of the existing models are based on simple, random (or blind) linear motion, it not only difficult to describe the reality of the curve of motion patterns, but also produce such as the border effect (border effect), speed attenuation and other defects, resulting in the simulation based on these models can not accurately reflect the real network behavior, it is necessary to establish a mobile model close to reality. However, how to analyze the dynamic characteristics and simulation results of different models of Ad Hoc network systematically and comprehensively, and theoretically require a general and quantifiable evaluation method [10]. At present, most of the researches on dynamic characteristics of mobile models are based on the moving model itself and its one side (such as temporal or spatial dynamic) to establish a specific computing model. It is difficult to be independent of the mobile model, because it is difficult to independent of the mobile model, so the lack of comparability and versatility [11].

In the network, trust and security are two different concepts. Security is mainly based on the security of the system. It can avoid or deal with illegal intrusions, malicious attacks and so on. However, in the current network, trust is to help the network entity to establish confidence or judge service cooperation, to reduce the risk and substantive cooperation [12]. At the same time, the two are closely related. Security can provide reliable communication and information protection for the creation of trust, and trust can enhance security. In order to correctly distinguish selfish nodes and malicious nodes on the network, many scholars have proposed the establishment of a trust model and the routing algorithm according to the history of interactions between entities, to achieve node priority trust, as well as the normal communication [13]. The weight of the score is determined by the confidence, credibility, timing, score, and distance of the existing scoring. This method overcomes the disadvantages of the simple trust model, which is too

rough to describe the trust value accurately, and the algorithm is simple, intuitive and easy to understand [14]. The method of probability theory is more suitable for the requirement of historical interaction information collection in trust model. Both first divides trust into direct trust and recommendation trust, calculates trust value by probability statistics, and proposes a method for synthesizing trust. The trust value is evaluated by calculating the binary probability or positive conditional probability value. However, it does not consider the nature of trust and decay over time. Bias network is based on the Bias theorem in probability theory, and is one of the common methods of uncertainty reasoning. It provides an accurate theoretical basis for calculating trust value [15]. In its calculation, the prior probability, the sufficiency measure, the necessity measure and the equivalence are mostly given by expert experience. Some papers call it the Subjective Bayes approach. However, this method relies too much on expert experience and is more subjective. It is difficult to implement in Ad hoc network [16]. The uncertainty of trust is expressed by entropy, and the model also calculates the direct trust value between entities based on Bayesian theory, and takes into account the factors that decay with time. However, the proposed model does not consider more than two recommended trust synthesis problems. It does not satisfy the combination rate of trust synthesis. It cannot deal with trust recommendation from multiple parties, and does not distinguish direct trust from indirect trust.

In the model proposed by George et al., the trust evaluation problem is considered as the shortest path problem of weighted directed graph trust graphs, using path based semi loop and distance based semi loop transfer and composite trust respectively [17]. The trust of the model is only based on the demand of intuitive subjective judgment, and lacks the proof of the theoretical basis of the evidence. Moreover, the trust is generated only on the basis of local observation, so the amount of information needed for the calculation of trust is not complete enough. Luo Jun-hai proposed a trust model based on strategy game. The trust relationship among nodes is modeled. It gives the corresponding node's policy selection and payment function, thus ensuring that the trust model can identify malicious nodes, ensure secure communication between trusted nodes, and improve the performance of the network. Combining the concept of trust, Zhang Tao proposed the modeling idea of routing algebra. Based on this, we define and quantify the routing metric metrics in current trust routing research, and validate the idea theoretically, and extend the model to other classic routing protocols. This method is the first to give a complete set of theoretical systems about trust routing, and it has also a very good scalability [18]. The dynamic, time-varying and lost characteristics of wireless links lead to poor quality and low stability of wireless links, which challenges the reliability of network throughput and transmission. In addition, the link instability caused by node mobility and the limited energy of nodes also make it difficult to design and optimize routing protocols. However, broadcast characteristics of wireless channels are inherent advantages. Opportunistic routing takes advantage of this characteristic of wireless channels to improve the transmission reliability and end-to-end throughput of wireless networks.

### 3. Methodology

#### 3.1. Synthetic update of direct trust

In the ubiquitous environment, the nodes in the network can form a network with relatively stable topology at extremely fast speed. In this process, the nodes, mainly through the sharing of services and the cooperation of equipment, complete the function between each other, which can also primarily realize the application of a variety of mobile business with the use of self-organizing network [19]. While, it is necessary to be based on self-organizing network nodes for the application so as to achieve a variety of service resources by cooperating with request nodes. However, in the actual working environment, owing to their own limited resources, nodes would become selfish nodes or nodes with malicious strategy [20].

Among these nodes, selfish nodes, which would lead to the emergence of various problems of the entire system, directly determine the stability and security of the entire system [21]. For a server, if the number of selfish nodes is excessive, it will have a great adverse impact on excellent nodes in the network for its combating the enthusiasm of these excellent nodes to provide services, which directly causes the difficulty of the entire network in continuing [22]. Furthermore, the selfish node—Free-Rider also has a huge adverse impact in the network layer where misbehavior nodes attain its selfish purpose for their avoiding to provide the appropriate services, or packet forwarding services for other nodes. Thus, if these selfish nodes exist longer in the system, the destructive degree to system will become greater, meanwhile, exerting serious impact on the normal operation of system communication [23].

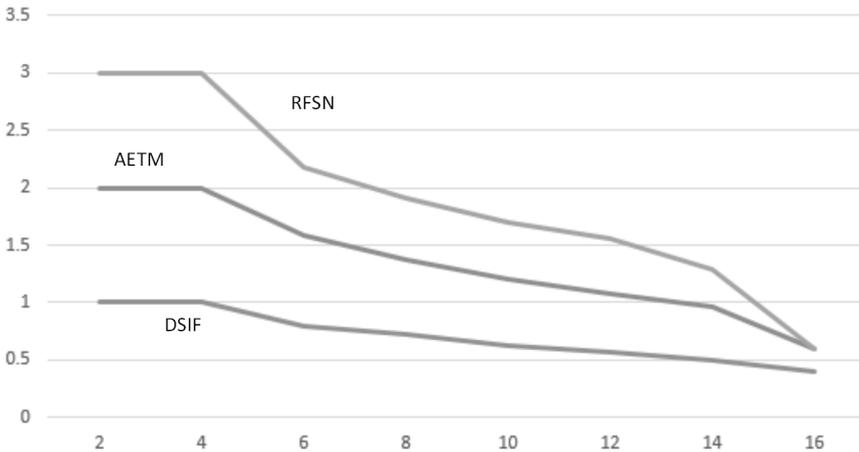


Fig. 1. Strategic attack of 1000 nodes under the change of direct trust

In the application, direct trust is based on the accumulation in interactive records of service request nodes and target nodes, and further provides strong evidence for the trust evaluation of the entire system, as shown in Fig.1. Here, RFSN is reputation-based framework for sensor networks, AETM is ad-hoc enabled trust

mechanism and, DSIF is dynamic sensor information fusion. The trust relationship established between two different nodes is the direct trust between these two nodes, which is mainly regarded as the degree of trust to the target nodes that have experienced in the services for request nodes according to its actual needs [24], with its significant importance for the safe operation of the entire system.

### *3.2. Design principles of incentive mechanism*

The incentive subject interacts with the incentive object, during which the process is reflected through a series of rationalized incentive mechanisms. In the practical application, the main rules of the incentive mechanism are as follows: First of all, one of the most important rules is open and intuitive principle, under which, purposes with the need to be incentive as well as methods require to be extremely clear, intuitive and open, so as to satisfy the open and intuitive principle of the entire system. Thus, the openness and intuitiveness of incentive mechanism is the primary link with significant focus when designing the network.

Secondly, it is the principle of objectives' combination that requires determination of goals of the whole mechanism with needs to be considered in the design process. However, in the construction of the network system, it is necessary to meet the individual requirements of participants for the realization of this link. In these processes, tangible and intangible incentives participating in activities together in the entire system will adopt different approaches to carry out classified incentives in different situations, thereby ensuring the balance and stability of the entire system for its long-term development and the maintenance of system's security.

Thirdly, the principle is praising virtue and punishing vice. Based on this principle, through incentive mechanism, participants who participate in the entire system would give up those behaviors that are harmful to the development and security of the system, and those who carry out reasonable operations and maintain the security would be further rewarded. In the actual incentive mechanism, the participants who are in accordance with target behaviors should be rewarded, and those who compromise the stability of system should be sanctioned from another point of view, as shown in Fig. 2. The symbols  $\alpha$  and  $\beta$  represent the adaptability of 2 kinds of behavior compromises of the system. The symbol  $\gamma$  represents the adaptability of the incentive behavior.

The fourth principle is to be incentive on demand, that is, the incentive mechanism must be based on its meeting the needs of nodes in the system. However, in different period of time, it is often different from the leading of stage nodes, so it is indispensable to meet the most urgent need in all needs of nodes in order to make these incentives can be carried out to the greatest extent.

The last one is the effectiveness principle of the system, which helps more nodes in incentive system to further bring about positive behaviors that are in line with the overall interests. At the same time, in order to effectively prevent those behaviors that do not meet the requirements and those violating the interests, the whole incentive mechanism can better promote the timely implementation of this part of punitive measures.

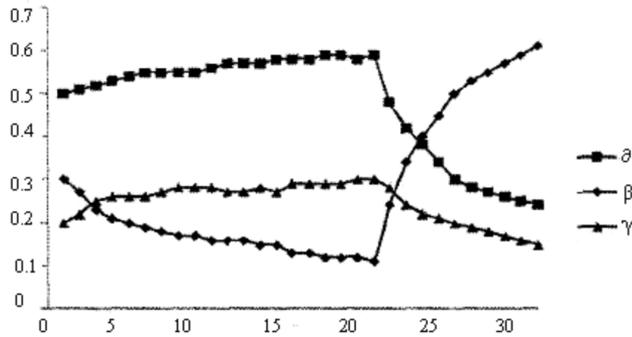


Fig. 2. Dynamic adaptability analysis

### 3.3. Analysis of simulation experiment

In the analysis of this problem, the source node, based on its own node, directly select the value of trust to send node's service request. The results are shown in Fig. 3. In the whole system, it is illustrated that all nodes have their own needs and rationality, desiring to choose a node with a higher trust value to provide more stable and reliable service. Thus, this part of excellent nodes with high value of trust, in the actual participation, often select those nodes that have the same good characteristic, so as to ensure that good nodes tend to combine with other good nodes for the stability of the entire system. Therefore, excellent nodes can be gradually gathered together to form a more solid network, so that all nodes within the region can obtain reliable network services, finally achieving the upgrade of priority. For those nodes with extremely low contribution rate and that are harmful to the stable development of the entire system, the anti-attack model is applied to eliminate them, and the malicious nodes are identified and cleared accordingly to ensure that all nodes are excellent, while to avoid the entering of many malicious nodes. Thus, the security and stable development of the network will gradually be achieved. Similarly, the request information sent by many nodes with malicious strategy, which is used by selfish nodes and malicious nodes can only be passed within the edge of the network, thus an incentive effect is available for the entire network where a successful interaction that is reliable and with relatively long-term adherence must be provided at the same time, if nodes in network want to get more superior web services.

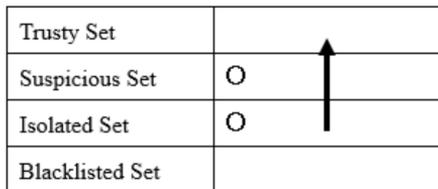


Fig. 3. Transition from an isolated set to a suspect set

In this study, the ability to resist malicious nodes of the entire network was mainly analyzed. The results are shown in Fig. 4. Among these nodes, one kind of node itself will attack maliciously in order to deceive other nodes in the network. Meanwhile, with it disguised itself as benign nodes of this system in the beginning to provide services, as soon as the trust is accumulated to a certain level, it will attack other nodes with the probability to damage the system. In the experiment, in order to further analyze the whole data model in depth, a service requesting node has been designed for the whole system according to various analysis models. In addition, a cheating node is designed for further analyzing how these two nodes participate in activities and how the cheating node obtain the trust in the system, thus it will be more convenient for malicious node to do its own malicious attacks. In these researches on model, the experiment is carried out under the same experimental conditions. The service request node sends 10 times of service request to the malicious node, among which the first three times allow the malicious node to provide the system with relatively stable reliable services, and provide unreliable services for the subsequent seven times.

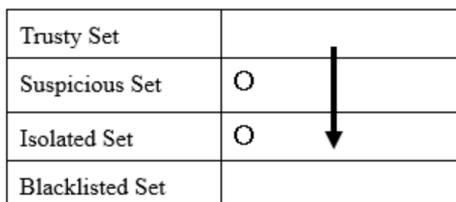


Fig. 4. Node from suspicious set to isolated set conversion

## 4. Result analysis and discussion

### 4.1. Design of discount factor

When designing and studying the whole system, the corresponding relationship analysis with D-S evidence theory illustrated that when the evidences are in conflict, if the consistency of some part of evidence is low compared with that of other evidences accounting for the majority, the impact of this part of evidence needs to be reduced to prevent the final fusion results being affected in varying degrees. On the contrary, if the evidence itself is highly consistent with other evidences in the entire system, thus this evidence can be caused corresponding attention in the network transmission, with direct impact on the final fusion results. In the analysis and research, through the comparison and analysis, it was found that in the process of integration, the weight that the evidence assigned to should be proportional to the assigned weight in the integration with other evidences. In this study, the relevant similarity matrix is applied to analyze these data, during which the data relationship is gradually clear, with effectively guaranteeing the accuracy of information every part obtained.

#### 4.2. Proportion of malicious nodes' finding

As for the proportion of malicious nodes, a more detailed data analysis and research is conducted. Through the comparison of data, it was illustrated that not only a kind of malicious node exists in the network, but many kinds of combination of malicious nodes with mutual influence and work together, if assuming that there are 500 malicious nodes. In the further analysis of this study, it was found that the data of different types of nodes among these malicious nodes is in a state of balance, and the proportion is relatively balanced, contributing to the further observation and analysis of these nodes in the experiment (see Fig. 5).

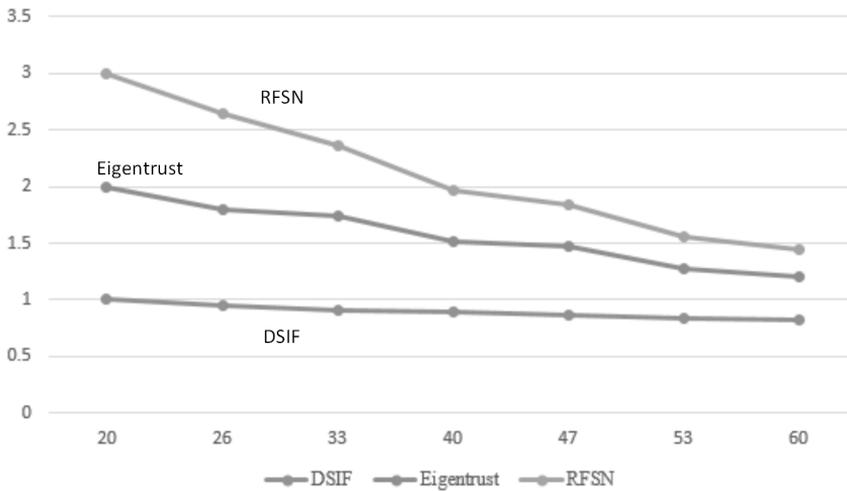


Fig. 5. Relationship between the success rate and the proportion of collusion nodes

In this experimental study, the probability of malicious nodes to be found has been greatly improved based on the research model, which will show a higher level after oscillating reactions. On the basis of this model, the errors caused by the prior distribution of RFSN can be well avoided, and the impact of subjective ambiguity can be more accurately identified, thus maximizing the accuracy of trust evaluation.

## 5. Conclusion

In this paper, the research on the main link in the trust management of Ad hoc network is conducted, and with it regarded as the background, the anti-attack model is analyzed. Because the network lacks the relative centralized server, and the trust of nodes must be completed by themselves, meanwhile, due to the changes in the network topology and its infinite conflict, some problems existing in the trust evaluation of the system are caused into fact. Then, in the process of establishing the trust evaluation model with the background of evidence theory, the randomness and subjective uncertainty when processing assessment do not require the analysis

of prior distribution. With this model, the relevant trust filtration mechanism can be carried out to detect nodes stably, while the recommended nodes and nodes with collusion attack are found in this link. In this paper, with theoretic analysis and results analysis in simulation, it was found that attacks of malicious nodes can be well avoided to achieve the stable operation of the entire system.

## References

- [1] V. PAHAL, S. KUMAR: *Degradation of ad-hoc network performance under wormhole attack*. International Journal of Computer Applications 49 (2012), No. 13, 25–29.
- [2] G. M. BORKAR, A. R. MAHAJAN: *A secure and trust based on-demand multipath routing scheme for self-organized mobile ad-hoc networks*. Wireless Networks (2016), 1–18.
- [3] H. C. CHEN: *TCABRP: A trust-based cooperation authentication bit-map routing protocol against insider security threats in wireless ad hoc networks*. IEEE Systems Journal 11 (2017), No. 2, 449–459.
- [4] M. J. FAGHIHNIYA, S. M. HOSSEINI, M. TAHMASEBI: *Security upgrade against RREQ flooding attack by using balance index on vehicular ad hoc network*. Wireless Networks (2016), 1–12.
- [5] A. HINDS, M. NGULUBE, S. ZHU, H. AL-AQRABI: *A review of routing protocols for mobile ad-hoc networks (MANET)*. International Journal of Information and Education Technology 3 (2013), No. 1, 1–5.
- [6] D. SARAVANAN, R. M. CHANDRASEKARAN, B. V. PRABHA, V. R. SARMA DHULIPALA: *Trust worthy architecture implementation for mobile ad hoc networks*. International Journal on Computer Science and Engineering 3 (2011), No. 7, 2601–2609.
- [7] W. LI, H. SONG: *ART: An attack-resistant trust management scheme for securing vehicular ad hoc networks*. IEEE Transactions on Intelligent Transportation Systems 17 (2016), No. 4, 960–969.
- [8] M. MOHANAPRIYA, I. KRISHNAMURTHI: *Trust based DSR routing protocol for mitigating cooperative black hole attacks in ad hoc networks*. Arabian Journal for Science and Engineering 39 (2014), No. 3, 1825–1833.
- [9] R. CHEN, J. GUO, F. BAO, J. H. CHO: *Trust management in mobile ad hoc networks for bias minimization and application performance maximization*. Ad Hoc Networks 19 (2014), 59–74.
- [10] H. XIA, Z. JIA, X. LI, J. LEI, E. H. M. SHA: *Trust prediction and trust-based source routing in mobile ad hoc networks*. Ad Hoc Networks 11 (2013), No. 7, 2096–2114.
- [11] R. A. SHAIKH, A. S. ALZHRANI: *Intrusion-aware trust model for vehicular ad hoc networks*. Security & Communication Networks 7 (2014), TOC No. 11, 1652–1669.
- [12] A. SHABBIR, F. KHALID, S. M. SHAHEED, J. ABBAS, M. ZIA-UL-HAQ: *Security: A core issue in mobile ad hoc networks*. Journal of Computer Science & Communications 3, (2015), No. 12, 41–66.
- [13] Z. MOVAHEDI, Z. HOSSEINI, F. BAYAN, G. PUJOLLE: *Trust-distortion resistant trust management frameworks on mobile ad hoc networks: A survey*. IEEE Communications Surveys & Tutorials 18 (2016), No. 2, 1287–1309.
- [14] E. DA SILVA, L. C. P. ALBINI: *SEMAN: A novel secure middleware for mobile ad hoc networks*. Journal of Computer Networks and Communications (2016), paper 3136853.
- [15] S. SICARI, A. RIZZARDI, L. A. GRIECO, A. COEN-PORISINI: *Security, privacy and trust in internet of things: The road ahead*. Computer Networks 76 (2015), 146–164.
- [16] S. TAN, X. LI, Q. DONG: *A trust management system for securing data plane of ad-hoc networks*. IEEE Transactions on Vehicular Technology 65 (2016), No. 9, 7579 to 7592.

- [17] D. LAVROVA, A. PECHENKIN: *Applying correlation and regression analysis to detect security incidents in the internet of things*. International Journal of Communication Networks and Information Security 7 (2015), No. 3, 131–131.
- [18] R. AASERI, P. CHOUDHARY, N. ROBERTS: *Trust value algorithm: A secure approach against packet drop attack in wireless ad-hoc networks*. International Journal of Network Security & Its Applications (IJNSA) 5 (2013), No. 3, 99–111.
- [19] B. BALAJI, M. H. KHAN, T. S. N. MURTHY: *A defense mechanism for EOLSR against DOS attacks in ad hoc networks using trust based system*. International Journal of Security and Its Applications [SCOPUS] 8 (2014), No. 5, 51–64.
- [20] F. ZHANG, Z. P. JIA, H. XIA, X. LI, E. H. M. SHA: *Node trust evaluation in mobile ad hoc networks based on multi-dimensional fuzzy and Markov SCGM(1, 1) model*. Computer Communications 35 (2012), No. 5, 589–596.
- [21] T. EISSA, S. A. RAZAK, M. A. NGADI: *A novel lightweight authentication scheme for mobile ad hoc networks*. Arabian Journal for Science and Engineering 37 (2012), No. 8, 179–2192.
- [22] R. VENKATARAMAN, M. PUSHPALATHA, T. R. RAO: *Regression-based trust model for mobile ad hoc networks*. IET Information Security 6 (2012), No. 3, 131–140.
- [23] Z. YAN, P. ZHANG, A. V. VASILAKOS: *A survey on trust management for internet of things*. Journal of Network and Computer Applications 42 (2014), 120–134.
- [24] M. KALIAPPAN, B. PARAMASIVAN: *Enhancing secure routing in mobile ad hoc networks using a dynamic bayesian signalling game model*. Computers & Electrical Engineering 41 (2015), 301–313.

Received June 29, 2017

# Seismic performance analysis of frame and frame-shear wall structures based on energy balance

HAIBO LIU<sup>1,2</sup>

**Abstract.** To explore the seismic performance of frames based on energy balance, based on the principle of energy analysis, the expression of energy relations of single degree-of-freedom system and the expression of energy response of multi-degree-of-freedom system are introduced. Based on the principle of energy analysis, the energy balance model is established, and a model for finite element analysis is set up. In addition, the traditional response spectrum is used to obtain the structure parameters. Entering the MIDAS finite element analysis software, the ability and performance of the frequently-seen earthquakes and rare earthquakes model are analyzed. Simulation experiments are carried out, and the results verified the validity of the above parameters. They are in line with the requirements of "High Standard", empirical formula and other requirements. The experimental results show that this method can effectively improve the seismic performance of reinforced concrete frame shear structures. At last, it is concluded that the structure can be well applied.

**Key words.** Energy balance, frame-shear structure, seismic performance.

## 1. Introduction

In the 1950s, Housner first proposed energy-based seismic design of structural thinking. He tried to use the energy analysis method to study the seismic response of the structure, that the structure of the earthquake response can be seen as a seismic energy input and dissipation process, as long as the damping and hysteretic energy dissipation capacity of the structure is greater than the earthquake input energy, structure can effectively resist the earthquake, and does not produce collapse. In the same time, Dr. George Housner explicitly proposed the concept of energy. The concept can better reflect the seismic intensity and spectral characteristics, and capture the inelastic deformation process of the structure under the action of strong earthquake from the input energy and dissipation of energy. The study of energy

---

<sup>1</sup>School of Management, China University of Mining and Technology, 221116, Xuzhou, China

<sup>2</sup>Jiangsu Collaborative Innovation Center for Building Energy Saving and Construction Technology, Jiangsu Vocational Institute of Architectural Technology, 221116, Xuzhou, China

method becomes an important development direction to improve the traditional seismic design method [1]. In the 1980s, Japan's Akiyama systematically researched the energy-based seismic design method. Based on the research results of the SDOF system and the MDOF system, the idea and method of energy-based seismic design were proposed, and some of them were applied in the Japanese seismic code. From the late 1980s to the 1990s, Fajfar studied the ground motion intensity index and the energy input and distribution relationship of the SDOF system, and proposed the N2 design method for the concrete structure [2]. After 1990s, Fajfar and other scholars carried out a large number of studies on the seismic design method of the comprehensive consideration of the cumulative hysteretic energy dissipation and deformation. After 2000, Chou and Shen established the energy-based seismic design method of steel frame structure, and put forward the practical design flow.

Since the 1980s, Chinese scholars have begun to study energy-based seismic design methods. China's scholars have done a lot of research work in the energy input, distribution and the law of cumulative hysteretic energy dissipation of the SDOF system and MDOF system and the test of energy dissipation capacity of the structure and components. After many efforts of researchers, the foundation work of the energy law has been tending to improve, the corresponding design framework is basically mature, but there is no systematic design method.

## 2. Materials and methods

### 2.1. Energy method theory

For an idealized single-layer structure affected by horizontal seismic motion, it can be regarded as an ideal single-degree-of-freedom system. The quality of the system focus on a point of the end, the single degree of freedom or the multi-degree of freedom system will have the damping, and dissipate the structural energy. If it is assumed that the axial deformation does not occur at the end, the system will have three degrees of freedom in the analysis of the static force, namely: the horizontal line displacement and the rotation angle of the two nodes. In the dynamic analysis under horizontal ground motion, the single-degree-of-freedom system requires only an independent horizontal displacement to determine the position of the particle under the action of the horizontal inertial force. Therefore, the system has only one lateral displacement degree of freedom. The displacement  $u'$  of the particle relative to the original rest position can be regarded as the superposition of two parts, which are the rigidity lateral displacement  $u_g$  of the whole structure and the horizontal displacement  $u$  of the particle relative to the structure base produced by the inertia force

$$u' = u_g + u \quad (1)$$

Thus, it is easy to get a motion equation of single degree of freedom system, which can be expressed as

$$m\ddot{u}(t) + C\dot{u}(t) + f_s(u, \dot{u}) = -m\ddot{u}_g(t) . \quad (2)$$

Here,  $C$  is the viscous damping coefficient and  $f_s(u, \dot{u})$  is the restoring force of the structure.

When the structure is still in the linear elastic deformation stage, the restoring force of the system can be expressed by the formula

$$f_s = Ku, \quad (3)$$

where  $K$  is the initial lateral stiffness of the system and  $u$  is the relative displacement of the system.

The initial stiffness of the structure is represented by the stiffness of the beam, column and wall which the structure belongs, and can be obtained by a certain algebraic calculation. Under the action of rare earthquakes, the deformation of the structure may be in the elastic-plastic stage. In this process, the relationship between the restoring force and the displacement of the structure becomes very complex. In general, the restoring force will not correspond to the displacement value, but will depend on the deformation path and the deformation state, and can be expressed as

$$f_s = f(u, \dot{u}). \quad (4)$$

In the theoretical study, the general mathematical model of restoring force which is simplified on the basis of the experimental data is adopted. The equation (2) can be understood as the system substrate is stationary, and a horizontal equivalent force  $P = -mu$  is affected in the particle. Because the force is proportional to the quality of the system, the system quality is greater, the stronger the earthquake. The both ends of the motion differential equations (2) relatively displace  $u$  points to the particle, and the energy equation relative to the displacement can be obtained [3]

$$\int_0^u m\ddot{u} du + \int_0^u C\dot{u} du + \int_0^u f_s du = - \int_0^u m\ddot{u}_g du. \quad (5)$$

Since the displacement  $u$  is a function of time  $t$ , a differential relation  $du = u dt$  can be obtained, so that the integration of the displacement can be converted to the integration of time  $t$ , and the above relationship can be substituted into (3) to obtain the expression of the seismic response energy calculation of the single degree of freedom system:

$$\int_0^t m\ddot{u}\dot{u}(t) dt + \int_0^t C\dot{u}^2(t) dt + \int_0^t f_s\dot{u}(t) dt = - \int_0^t m\ddot{u}_g\dot{u}(t) dt. \quad (6)$$

$$E_I = \int_0^u m\ddot{u}_g du = - \int_0^t m\ddot{u}_g\dot{u}(t) dt. \quad (7)$$

The equations on the left side of the above formula represent the different meanings, and the representing method is as follows:

Symbol  $E_K$  stands for the kinetic energy of the structure, that is given by the

expression

$$E_K = \int_0^u m \dot{u} du = \int_0^t m \ddot{u} \dot{u}(t) dt. \quad (8)$$

Quantity  $E_D$  is the damping dissipation energy of the structure, that is given as

$$E_D = \int_0^u C \dot{u} du = \int_0^t C \dot{u}^2(t) dt. \quad (9)$$

Symbols  $E_S$  and  $E_H$  denote the elastic deformation energy and inelastic hysteretic dissipation energy of the structure, respectively, that may be expressed as

$$E_S + E_H = \int_0^u f_s du = \int_0^t f_s \dot{u}(t) dt. \quad (10)$$

So, at any time  $t$ , the relationship of the energy balance in the structure is

$$E_K + E_D + E_E + E_H = E_I. \quad (11)$$

In mathematics, because the structure is a continuous medium, there should be infinite degrees of freedom, but in specific engineering practice, the structure is usually equivalent to the limited multi-degree-of-freedom system to analyze according to certain rules, and the specific expression is

$$[M] \{\ddot{u}(t)\} + [C] \{\dot{u}(t)\} + \{R(t)\} = -[M] \{r\} \ddot{u}_g(t), \quad (12)$$

where  $[M]$  is the diagonal matrix of the concentrated mass,  $[C]$  is the damping matrix of the structure,  $\{R(t)\}$  is the restoring force matrix of the structure and  $u(t)$ ,  $\dot{u}(t)$ ,  $\ddot{u}(t)$ , respectively, represent the displacement vector, the velocity vector and acceleration vector of the particle of the structure. Symbol  $\ddot{u}_g(t)$  is the acceleration of the ground motion; conversion column vector, and consider the structure of the damping matrix,  $\{r\}$  is the switching column vector, and the corresponding terms of the degree of freedom in the direction of action of the seismic inertial force is 1, and the rest is 0 [4].

When the earthquake continues to develop on the structure, the structure will enter the elastic-plastic stage, the stiffness matrix will change with time, and the specific performance is related to the position and the state of the each unit in their respective restoring force curve. Therefore, when solving differential equations, it is necessary to divide the whole seismic motion into a series of small steps of equal step length or different step length, and treat the structural parameters in each period as constants, and then use the stepwise integration method to solve.

In the structural analysis, the common step by step integration methods are linear acceleration method, the midpoint acceleration method, Marker  $\beta$  method and Wilson  $\theta$  method [5]. In general, considering the aspects of stability and calculation accuracy of the model, the midpoint acceleration method or the Wilson's  $\theta$  method are used to analyze the nonlinear dynamics of the multi-degree-of-freedom system.

Similar to the energy response equation of the single-degree-of-freedom system,

the general form of the energy-response equation of the multi-degree-of-freedom system can be expressed as a matrix:

$$\begin{aligned} \int_0^t \{\dot{u}(t)\}^T [M] \{\ddot{u}(t)\} dt + \int_0^t \{\dot{u}(t)\}^T [C] \{\dot{u}(t)\} dt + \int_0^t \{\dot{u}(t)\}^T \{R(u(t))\} dt = \\ = - \int_0^t \{\dot{u}(t)\}^T [M] \{r\} \ddot{u}_g dt. \end{aligned} \quad (13)$$

The above equation can be abbreviated, and the simplified equation is:

$$E_K(t) + E_D(t) + E_S(t) + E_H(t) = E_I(t). \quad (14)$$

Now the energy of the ground motion input structure is

$$E_I(t) = - \int_0^t \{\dot{u}(t)\}^T [M] \{r\} \ddot{u}_g dt. \quad (15)$$

The kinetic energy of the structure is

$$E_K(t) = \int_0^t \{\dot{u}(t)\}^T [M] \{\ddot{u}(t)\} dt. \quad (16)$$

The damping dissipation energy of the structure [6] is

$$E_{qrmD}(t) = \int_0^t \{\dot{u}(t)\}^T [C] \{\dot{u}(t)\} dt. \quad (17)$$

Elastic deformation energy and the hysteresis dissipation energy are

$$E_S(t) + E_H(t) = \int_0^t \{\dot{u}(t)\}^T \{R(u(t))\} dt. \quad (18)$$

As the nonlinear dynamic analysis process is very complex, the calculation is very large. So, in general, in the actual design and analysis, it is necessary to do corresponding simplify and assumptions to the structure to have better results in the application.

The energy-based seismic design method first establishes the seismic input energy spectrum. The empirical criteria based on the numerical analysis are the equal energy criterion, the maximum deformation criterion, the instantaneous energy criterion and the equivalent linearization method in geometry. The geometrical equivalent energy criterion considers that the elastic system and the elastic-plastic system are equal to the deformation energy obtained according to the geometrical calculation area in the uniaxial load-displacement curve. The maximum deformation criterion considers that the elasticity system is almost equal to the maximum deformation of the elastic-plastic system, is an empirical conclusion obtained based on seismic response time history analysis results [7]. The instantaneous energy criterion

is an effective method to predict the maximum response of the elastic-plastic system based on the energy concept. According to the input energy in the time interval close to the fundamental period of the building structure, the unidirectional maximum deformation is determined, that is, the instantaneous input energy is transformed into the energy dissipation of uniaxial load-displacement to predict the maximum deformation. The equivalent linearization is a method suitable for the elastic system. It needs to assume that the damage distribution of the system is the same as that of the elastic system.

## ***2.2. Establishment of analytical model of frame-shear structure***

The engineering model is a part of the actual two-phase engineering under construction. The structure type adopts the cast-in-place reinforced concrete frame-shear wall structure, and the structure safety grade is the second level. The foundation design level is B level; the class of the building aseismicity is C-class, and the seismic fortification is 7 degree (0.15g) [8]. The site soil is Class II, and the design seismic grouping is the first group. The seismic rating framework of the special-shaped column-frame shear-wall structure is the three-tier level.

The basic modal analysis is carried out first to analyze the structural model, and the most original results of the structure are obtained, as shown in Tables 1–3.

Table 1. Structural eigenvalue analysis

Modal number	Frequency (Hz)		Cycle (s)	Allowable error
	Rad/s	Cycles/s		
1	6.7	1.0663	0.9378	0
2	7.7762	1.2376	0.808	0
3	8.1754	1.3012	0.7685	0
4	22.4687	3.576	0.2796	0
5	27.3295	4.3496	0.2299	0
6	28.271	4.4995	0.2222	0
7	44.6431	7.1052	0.1407	2.29E-82
8	56.8744	9.0518	0.1105	1.76E-68
9	58.3491	9.2866	0.1077	1.13E-67
10	74.1298	11.7981	0.0848	2.54E-54
11	98.6677	15.7034	0.0637	4.18E-29
12	101.4837	16.1516	0.0619	1.9E-25

The first mode of vibration of the structure is the translational motion in the  $X$  direction, the second mode is the translation in the  $Y$  direction, the third mode is the torsion, and the ratio of the first mode to the third mode is 73%, which meets less than 90% of the provisions [9]. In the first mode and the second mode, the stress at the top of the structure is the largest, and in the third mode, the stress at the corner of the structure is the largest.

Table 2. Parameters and quality of the mode of vibration

Modal number	TRAN-X		TRAN-Y		TRAN-Z	
	Quality (%)	Total (%)	Quality (%)	Total (%)	Quality (%)	Total (%)
1	71.0445	71.0445	0.1332	0.1332	0.0001	0.0001
2	0.4741	71.5186	67.3234	67.4566	0	0.0001
3	1.7839	73.3025	3.652	71.1086	0.0001	0.0002
4	14.1655	87.4681	0.0082	71.1168	0.0001	0.0002
5	0.0255	87.4936	10.6149	81.7317	0.0025	0.0027
6	0.0097	87.5033	3.9686	85.7003	0.0006	0.0033
7	5.5002	93.0035	0.0007	85.7009	0.0001	0.0034
8	0.0019	93.0054	1.1037	86.8046	0.0011	0.0046
9	0.0029	93.0083	4.7371	91.5417	0.001	0.0055
10	2.6485	95.6568	0.0005	91.5423	0.0003	0.0058
11	0.0053	95.6621	0.0726	91.6148	0.0002	0.006
12	0.0002	95.6623	3.1349	94.7497	0.7585	0.7645

Table 1. Structural eigenvalue analysis

Modal number	TRAN-X	TRAN-Y	TRAN-Z
1	97.2254	0.1823	0.0001
2	0.6695	95.0804	0
3	2.4784	5.0737	0.0001
4	99.1613	0.0576	0.0006
5	0.1744	72.5288	0.0169
6	0.0686	28.0642	0.0043
7	99.9584	0.012	0.0017
8	0.0324	18.711	0.0194
9	0.0513	82.8047	0.01690
10	99.5879	0.0203	0.0115
11	0.1452	1.9918	0.0053
12	0.0049	73.7291	17.8393

### 3. Results

#### 3.1. Analysis of the displacement

Figure 1 shows the displacement of structure under the effects of Taft wave and E1 wave on frequent earthquakes, Fig. 2 then depicts the displacement of the structure under the effects of the Taft wave and E1 wave on rare earthquakes

The results show that the average displacement of the whole structure is 14.15 mm,

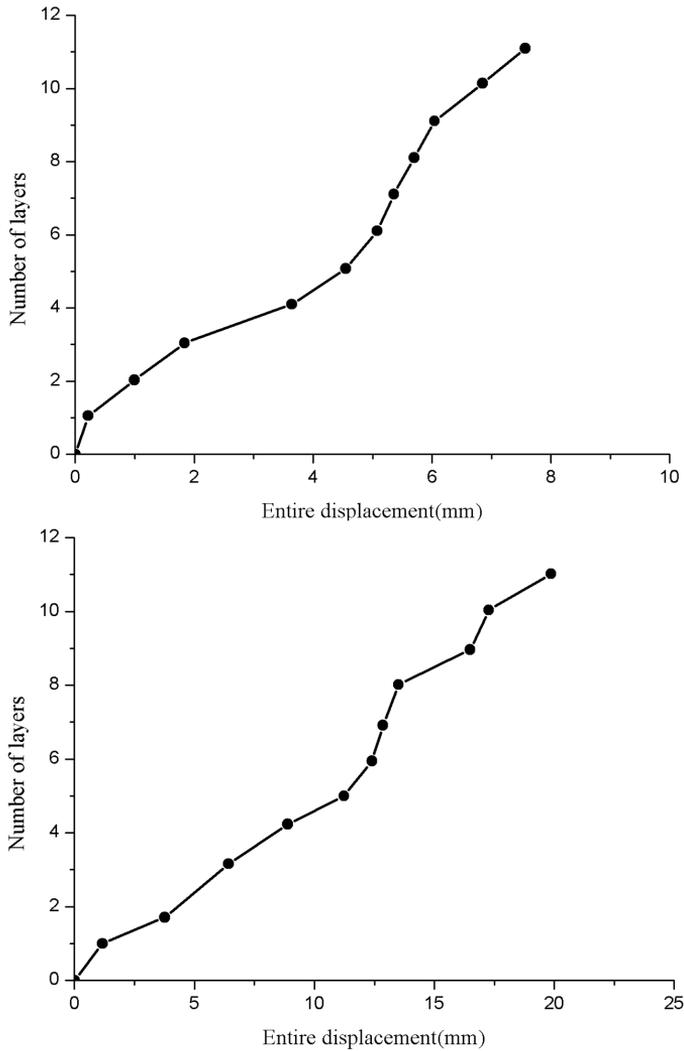


Fig. 1. Displacement of structure under the effects of Taft wave (up) and EL wave (bottom) on frequent earthquakes

which is in accordance with the specification. Comparing the floor displacement map, we can see that although the structure is subjected to different seismic waves, the response of the structure is similar. The structure is mainly composed of bending deformation in 1–4 layers, shear deformation in 5–8 layers and bending deformation in 9–11 layers [10]. Under different seismic waves, the displacements and deformations of the structures are different. The displacements of the EL waves are much larger, which is determined by the structural dynamic characteristics and the spectral characteristics of seismic waves. Overall, the overall deformation of the structure shows bending and shearing state. Therefore, it can be deduced that the displace-

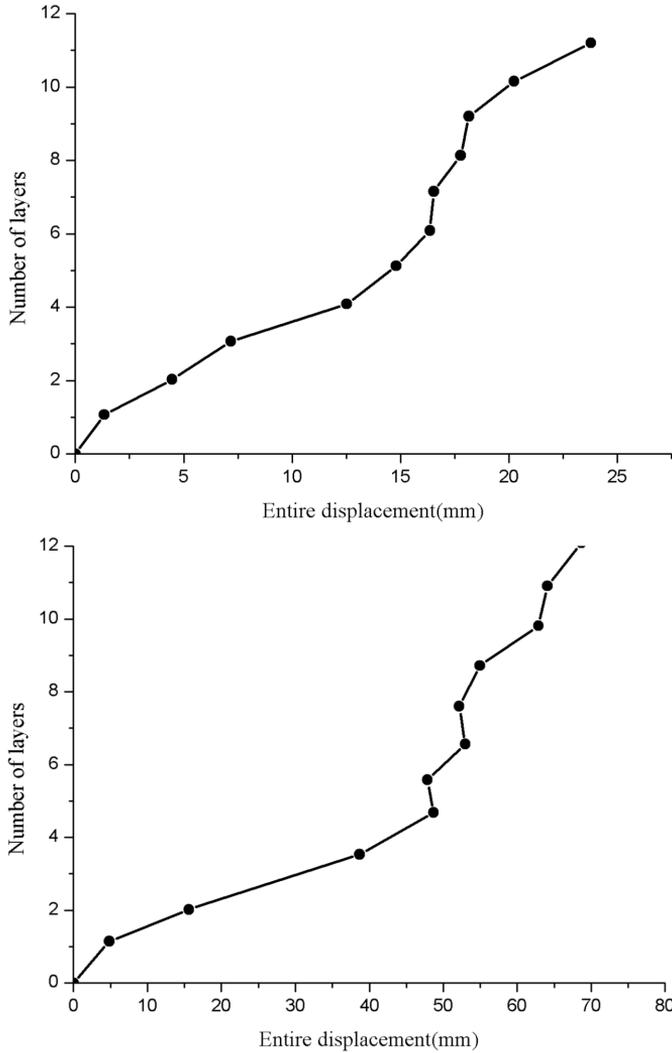


Fig. 2. Displacement of structure under the effects of Taft wave (up) and EL wave (bottom) on rare earthquakes

ment of frame-shear-wall structure under different seismic waves is different, but the deformation is uniform.

### 3.2. Analysis of the structure energy

The input structure model of the Tianjin, Taft and EL waves under the rare earthquake are taken to calculate the energy response of the above structure, at the same time, the peak value of seismic wave acceleration was taken as 2209, and the seismic response time was calculated as 20 s [11]. The analysis results of the total

energy time history is respectively shown in Figs. 3 and 4.

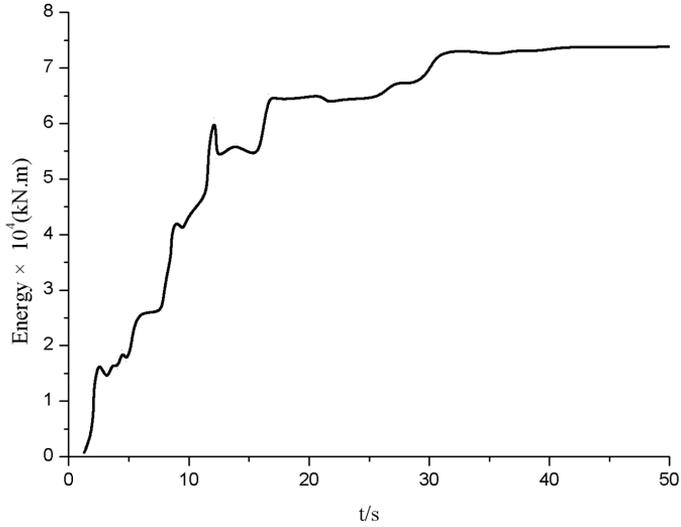


Fig. 3. Time history of Taft wave energy

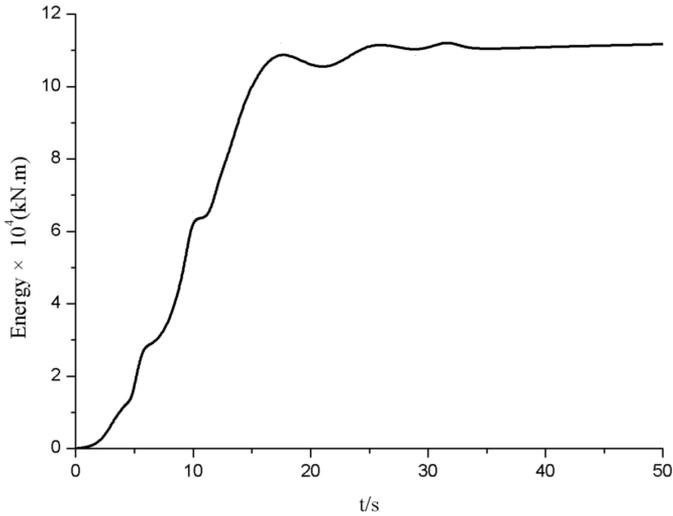


Fig. 4. Time history of EL wave energy

From the figure, we can see that although the seismic wave intensity and the duration are the same, the energy response of the structure is also quite different. The total energy basically increases and the shock increases. The time history curve of the total energy increases continuously in the beginning, which corresponds to the concentrated area of the seismic wave intensity.

From the comprehensive analysis of the effects of the frequent earthquake and

rare earthquake on the structure, it can be concluded that the deformation curve of the whole structure under the action of multiple earthquakes is curved scissors type. And the displacement of the top layer of the structure is relatively large, because the mass and the rigidity setting is smaller. So the layout of the structure cannot ignore the rational layout of the top structure [12]. Under the action of rare earthquakes, the interlayer displacement value of the bottom layer of the structure shows the maximum value, and it is most likely to yield first. Through comparison and analysis of the wave energy, the following conclusions can be drawn: the total energy input of the structure under the action of rare earthquakes increases with oscillation, and the energy has a rapid growth section, which is usually located in the concentration section of the seismic wave intensity. Because of the waveforms of the seismic waves and the structure, the results of this paper show that although the peak of acceleration and the duration of earthquake are the same, the energy response of the structure is very different. Under the action of earthquake, the energy of the input structure is finally balanced by damping energy dissipation and non-elastic hysteretic energy dissipation. On the contrast of energy dissipation, the damping energy dissipation of the structure is smaller than that of the non - elastic hysteretic energy. Under the action of Taft wave and EL wave, the input energy and energy dissipation of the earthquake will reach the maximum value quickly and will not increase again in the later period, which reflects that the response of the structure to earthquake is reduced, the strength and rigidity of the structure depredate, while hysteresis loop area decreases.

#### 4. Conclusion

Although this paper does not fully express the superiority of energy method and there are still some deficiencies, from the overall view, the energy analysis results are basically consistent with the actual situation of the project. And the results obtained are similar with those of existing methods, which suggests that it has a certain reference value. Energy based seismic design methods focus on the failure modes of structure and control of energy dissipation mechanisms, so the seismic performance of structures can be better understood as a whole. At the same time, in that the cumulative damage effect of structure is taken into account, so it can make necessary and reasonable supplement for the design of the structure bearing capacity and that based on displacement. And it is proved to be currently comparatively perfect seismic design method based on the performance. The reinforced concrete frame structure system is a widely used structural system in practical engineering. As a result, it is of great significance to study the seismic design of reinforced concrete frame shear structures based on energy.

#### References

- [1] F. REN, Y. ZHOU, G. CHEN, J. LIANG: *Experimental study on seismic performance of concrete-filled steel tubular frame-shear wall structure with buckling-resistant braces.*

- Structural Design of Tall & Special Buildings 24 (2015), No. 1, 73–95.
- [2] M. SURANA, Y. SINGH, D. H. LANG: *Seismic performance of concrete-shear-wall buildings in India*. Ice Proceedings Structures & Buildings 169 (2016), No. 11, 809–824.
  - [3] H. JIANG, X. LIU, L. HU: *Seismic fragility assessment of RC frame-shear wall structures designed according to the current Chinese seismic design code*. Journal of Asian Architecture & Building Engineering 14 (2015), No. 2, 459–466.
  - [4] J. SEO, J. HU, B. DAVAJAMTS: *Seismic performance evaluation of multistory reinforced concrete moment resisting frame structure with shear walls*. Sustainability 7 (2015), No. 10, 14287–14308.
  - [5] Z. DANG, X. LIANG, K. LI, H. ZHAO: *Seismic design of reinforced concrete frame-shear wall structure based on yield point spectra*. Tumu Gongcheng Xuebao/China Civil Engineering Journal 48 (2015), No. 6, 25–35.
  - [6] H. A. D. S. BUDDIKA, A. C. WIJEYEWICKREMA: *Seismic performance evaluation of posttensioned hybrid precast wall-frame buildings and comparison with shear wall-frame buildings*. Journal of Structural Engineering 142 (2016), No. 6, paper 04016021.
  - [7] M. SURANA, Y. SINGH, D. H. LANG: *Seismic performance of shear-wall and shear-wall core buildings designed for India codes*. Advances in Structural Engineering (2014), 1229–1241.
  - [8] M. SURANA, Y. SINGH, D. H. LANG: *Seismic characterization and vulnerability of building stick in hilly regions*. Natural Hazards Review (2017), No. 7.
  - [9] F. ZHANG, X. LI, Q. XU, C. GONG, X. CHEN, Q. LIU: *Experimental study on seismic behavior of frame-rocking wall structure*. Jianzhu Jiegou Xuebao/Journal of Building Structures 36 (2015), No. 8, 73–81.
  - [10] J. Y. LU, L. N. YAN, Y. TANG, H. H. WANG: *Study on seismic performance of a stiffened steel plate shear wall with slits*. Shock and Vibration (2015), paper 689373.
  - [11] M. ZEYNALIAN: *Numerical study on seismic performance of cold formed steel sheathed shear walls*. Advances in Structural Engineering 18 (2015), No. 11, 1819–1830.
  - [12] C. YUAN, J. HAO, C. FANG, C. FAN, N. HAN: *Experimental study on seismic behavior of semi-rigid steel frames with multi-ribbed grid composite steel plate shear wall structure*. Jianzhu Jiegou Xuebao/Journal of Building Structures 36 (2015), No. 4, 16–24.

Received June 29, 2017

# Fault diagnosis methods of rotating machinery based on mathematical morphology

JUMEI ZHANG<sup>1</sup>

**Abstract.** To make a fault diagnosis of rotating machinery, it is necessary to use the Local Characteristic-scale Decomposition (LCD) to remove the noise before the fractal method. The major reason is that the fractal method is sensitive to mechanical noise. LCD and mathematical morphology method are combined for the diagnosis of mechanical failure, and in this ways, more accurate results can be obtained than that under the box dimensions. In addition, the morphological fractal dimension is used to calculate the fractal dimension of the main component, and the degree of discrimination of each state can be clearly depicted by curve description. And the results showed that the fault state of rolling bearing can be effectively identified and fault diagnosis can be realized. At last, it is concluded that the method based on LCD decomposition and morphology fractal dimension can successfully do the fault diagnosis, which has great application value and good prospect.

**Key words.** Mathematical morphology, rotating machinery, LCD, fault diagnosis.

## 1. Introduction

With the development of economic globalization and science and technology, people had a higher request in the stable and efficient operation of the machinery and equipment in industrial. More and more attention had been paid to the equipment fault diagnosis technology. Rolling bearings are one of the most commonly components used in mechanical equipment, but they are also particularly vulnerable to damage, which is detrimental to the life of the entire system and normal production. So, it is necessary to detect and diagnose the failure state of machinery and equipment bearing.

In 1991, Koskinen extended the traditional mathematical morphology operator, and proposed soft mathematical morphology. Soft mathematical morphology means to replace the maximum and minimum operation in the traditional mathematical morphology as weighted ordering statistics. Its structure elements include two parts:

---

<sup>1</sup>Weinan Normal University, Weinan, Shaanxi, 710000, China

hardcore and soft boundary [1]. In 1990s, Sinha and other scholars introduced fuzzy theory into traditional mathematical morphology, and put forward fuzzy mathematical morphology. It used membership function to achieve the operation between structure operator and fuzzy images. As a result, compared with traditional mathematical morphology, fuzzy mathematical morphology has stronger noise suppression capability. At present, application of morphology based on single scale structure element is widely studied. Zhang Lijun used generator signal for the filtering processing, and then he applied morphology non-sampling wavelet for the rotor wave processing and Shen Lu made use of morphology wavelet and morphology non-sampling to make filtering processing of gearing and rolling, respectively [2]. All of these studies laid a foundation for the further study of mathematical morphology. Mechanical fault diagnosis is an actually signal processing. The fault signal is typically non-linearity and non-stationarity. The fractal dimension of the mathematical morphology can effectively analyze and characterize the nonlinear behavior of the fault signal. However, the measured bearing vibration signal often contains a lot of system background noise. Because the mathematical morphology fractal dimension is very sensitive to noise, so the measured signal must be denoised to get the accurate fractal dimension, and the traditional linear filter is usually not competent [3]. A new adaptive time-frequency analysis method—Local Characteristic-Scale Decomposition (LCD) is proposed, which can decompose the vibration signal into a single rotation component with physical meaning at instantaneous frequency. In this paper, the LCD method is combined with the fractal dimension of the morphology, and the vibration signal is decomposed by the LCD. The components of the main characteristic frequency are used as the fault signals to be analyzed. The morphological fractal dimension is used to calculate the fractal dimension of the main component, and the degree of discrimination of each state can be clearly depicted by curve description, which can effectively identify the fault state of rolling bearing and realize the fault diagnosis of rolling bearing.

## 2. Materials and methods

### 2.1. Mathematical morphology

The basic idea of using mathematical morphology to measure the complexity of the nonlinearity of vibration signals at different scales is that the results of covering the one-dimensional vibration signals with planar set  $B$  in the process of dealing with one-dimensional fault signals are regarded as one-dimensional structural elements  $g$  to detect the signal, which is an equivalent method, in which the structural element is the upper bound of the planar set  $B$ . Based on the above-mentioned idea, the concrete method steps are as follows [4]: the one-dimensional discrete vibration signal  $f(n)$  ( $n = 0, 1, \dots, N$ ),  $g(m)$  is the one-dimensional unit structure element defined on  $G = \{0, 1, 2, \dots, M - 1\}$ . One-dimensional morphology corrosion and expansion are carried out to measured signal, and the scale  $\varepsilon = 0, 1, 2, \dots, \varepsilon_{\max}$  represents the discrete scale range. The vibration signals are expanded and corroded by one-dimensional discrete function as the unit structure element at each scale.

### 2.2. Analysis of simulation signal

In order to verify the effectiveness of fractal dimension estimation method based on mathematical morphology, Weierstrass cosine function (WCF) is used as the fractal signal, which is defined as [5]:

$$W_H(t) = \sum_{k=0}^{\infty} \gamma^{-kH} \cos(2\pi\gamma^k t), \quad (0 < H < 1). \quad (1)$$

In the formula,  $\gamma > 1$ . WCF is a continuous but not differentiable signal, and the fractal dimension of  $W_H$  is  $D = 2 - H$  in theory. The sampling frequency of the simulation signal is 1024 Hz, the number of sampling points is 2048, and the parameters are set to  $\gamma = 5$  and  $k = 20$  [6–7]. Table 1 shows the WCF signals of three different dimensions ( $D = 1.4, 1.6, 1.8$ ).

At present, the Box-counting method is the most widely used signal estimation method, but because the box dimension divides the grid regularly, there is a problem that the fractal dimension estimation is inaccurate. However, the mathematical morphology is not affected by these factors, and the calculation results are more accurate [8]. Table 1 is the result of using box counting and fractal dimension estimation to deal with the signal. From Table 1, it can be seen that the fractal dimension of the box counting method is generally low, the error is larger than the mathematical morphology, and the morphology is more accurate, and the computational efficiency of mathematical morphology is higher than the box dimension.

Table 1. Fractal dimension estimation of WCF signal

Methods	Actual value	$D = 1.4$	$D = 1.6$	$D = 1.8$
Box-counting method	Estimated value	1.3497	1.4997	1.6553
	Relative error	3.59 %	6.26 %	8.04 %
Morphological method	Estimated value	1.4386	1.6200	1.8076
	Relative error	1.97 %	1.25 %	0.42 %

The WCF signal was morphologically covered with structural elements of sizes 32 and 64. Figure 1 shows the expansive corrosion results of  $W_{0.4}(t)$  at scales 32 and 64.

### 2.3. LCD decomposition method

LCD’s premise is that all the signals are regarded as to be composed of different single-component ISC, and any two ISC components are independent of each other. So that the signal  $x(t)$  can be decomposed into the sum of the independent components, and these feature-scale components meet the following two conditions:

In the whole interval, the maximum value is positive, the minimum is negative, and there is monotony between any two adjacent. In the whole interval, the extreme point of the component is set to  $(\tau_k, X_k), k = 1, 2, \dots, M, M$  being the number of

extreme points [9]. Therefore, the straight line

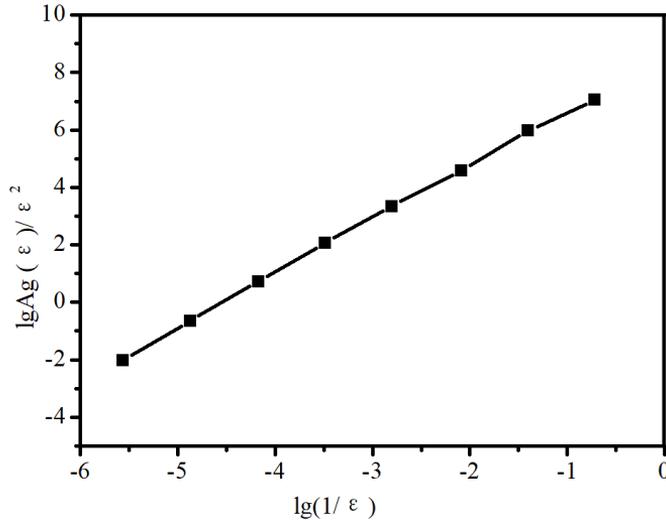


Fig. 1. Double logarithmic curve of  $W_{0.2}(t)$

$$l_k \left\{ y = X_k + \frac{X_{k+2} - X_k}{\tau_{k+2} - \tau_k} (t - \tau_k) \right\}$$

is determined by any two adjacent extremums (maximum or minimum)  $(\tau_k, X_k)$  and  $(\tau_{k+2}, X_{k+2})$  in the data segment, the ratio of the value  $A_{k+1}$  of this line in  $\tau_{k+1}$  and the function value at this point are unchanged, as shown in Fig. 2.

$$\frac{A_2}{X_2} = \cdots = \frac{A_6}{X_6} = \cdots \mu. \quad (2)$$

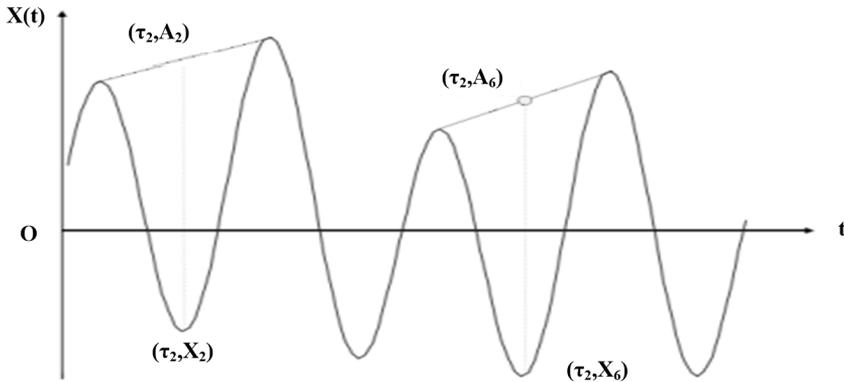


Fig. 2. Conditions for the intrinsic scale component

In the simulation signal,  $\mu$  may not change, but in the actual signal, its value must not be so stable, so it can be changed in a certain range of permission.

The general form is [10]:

$$aA_{k+1} + (1 - a) X_{k+1} = 0, \quad a \in (0, 1), \quad (3)$$

$$A_{k+1} = X_k + \left( \frac{\tau_{k+1} - \tau_k}{\tau_{k+2} - \tau_k} \right) (X_{k+2} - X_k). \quad (4)$$

When  $a = 0.5$ ,  $A_{k+1} = -X_{k+1}$ .

### 3. Results

The following signals are used to investigate the decomposing effect of the LCD:

$$x(t) = x_1(t) + x_2(t) = [1 + 0.5 \cos(20\pi t)] \sin(200\pi t + 200t^2) + \sin(40\pi t). \quad (5)$$

The simulation signal is composed of an amplitude modulation and frequency modulation signal and a sinusoidal signal.

From the decomposition results of the LCD, it can be seen that this method can decompose the frequency components in the simulation signal very well. CISC3 has a small fluctuation after processing the end effect by the extension method, and the processing result is ideal. The results show that LCD is an effective and feasible decomposition method.

In the case of adaptive time-frequency analysis, such as Empirical Mode Decomposition (EMD) and Local Mean Decomposition (LMD), these methods only give their components with the physical meaning [11]. The conditions that the intrinsic mode function (IMF) component EMD definite by EMD or the PF component in LMD need to meet are the sufficient conditions of instantaneous frequency with the physical meaning, but not a necessary condition, that is, as long as certain conditions are met, a single component signal can have physical significance. In addition, the cubic spline function is used to deal with the end-effect in the EMD method, which often has over-envelope, under-envelope phenomenon in the formation of the upper and lower envelopes, and the mode aliasing is serious. However, the LMD method has more end-effects than the EMD method, and the range of the LMD method is small, but the LMD algorithm has its own limitations [12]. In the decomposition of LMD algorithm, it uses the moving average algorithm to calculate the local mean function and local envelope function, while the moving average algorithm is a cycle of multiple iterations, so the calculation is very large. For these reasons, the components defined by EMD and LMD often result in unreliable methods. Therefore, this paper combines a new instantaneous frequency with physical meaning of the decomposition method LCD with the morphology. Experiments show that this method is effective.

The experimental data is from the bearing data of deep groove ball bearing

fault whose bearing model is 6205-2RS JEM SKF in United States West Reserve University [13]. The four failure types of the rolling bearing failure in normal test platform, rolling element failure, inner ring failure and outer ring failure are used.

It can be seen from Fig.3 that when the vibration signal is not denoised, the fractal dimension of the rolling element fault has two sudden ups and down at the position of sample 4 to 8, which cause the interference by overlapping with the curve of the normal state and the inner fault state. The curve of fractal dimension overlap twice in inner-circle fault and in the normal state, which makes the fault state unrecognizable. As the LCD method can decompose the original signal from high frequency to low frequency into several components, the high frequency component is often the best to reflect the fault feature information [14]. Therefore, usually the first component after the characteristic scale decomposition is calculated, and the vibration state of each failure can be effectively discriminate. Figure 3 is the first intrinsic component of the signal corresponding to Fig.4.

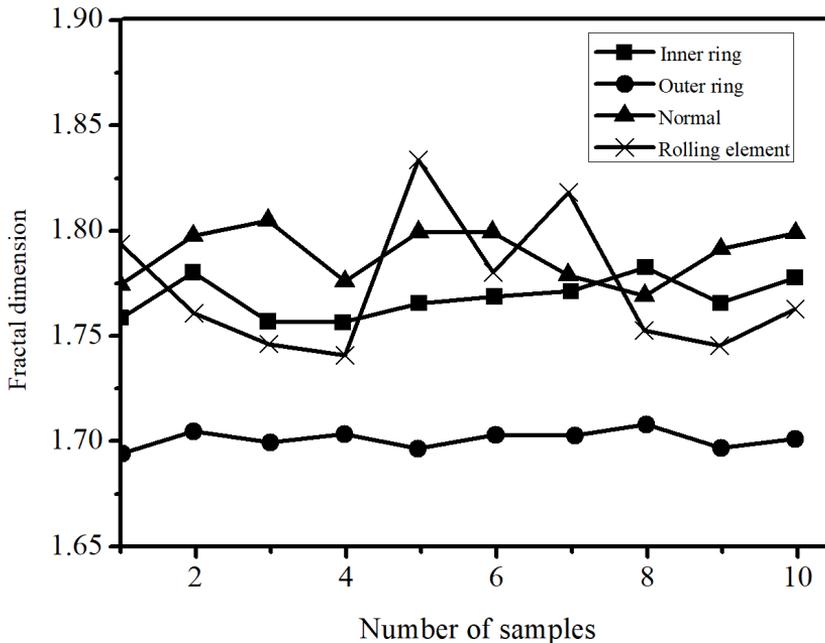


Fig. 3. Morphology fractal dimension of bearing faults signal

From the decomposition results of the LCD, it can be seen that this method can decompose the frequency components in the simulation signal very well, and the CISC3's fluctuation is small, and the processing result is ideal after dealing with the endpoint effect by the extension method. The results show that LCD is an effective and feasible decomposition method [15]. The first component is analyzed and processed by mathematical morphology, and a signal fractal is obtained as shown in Fig. 5. From the figure we can see that the fractal dimension is maximum in normal state of the bearing. Because there is no significant impact from the outside world

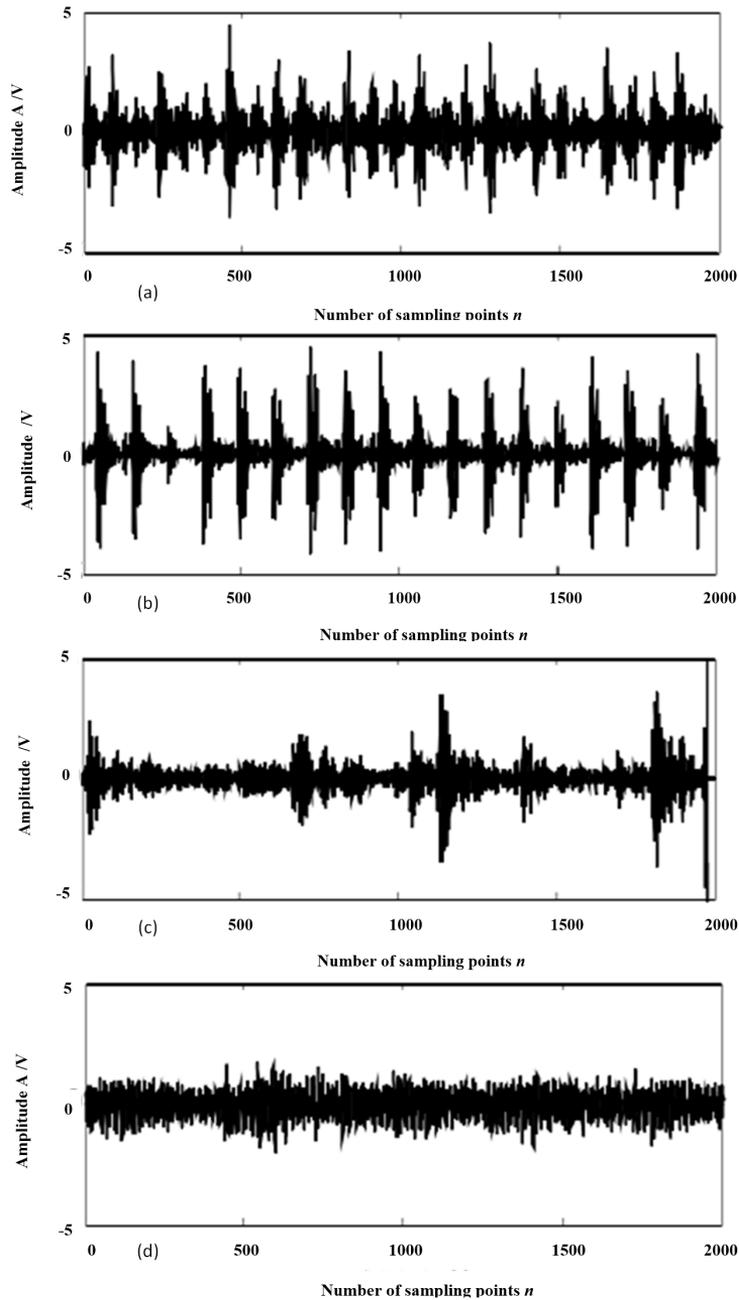


Fig. 4. The first component of the signal of four bearing fault conditions: (a)–first component of the rolling bearing inner ring, (b)–first component of the rolling bearing outer ring, (c)–first component of the rolling element of rolling bearing, (d)–first component of the normal state of rolling bearing

in normal state, and it is similar to the random signals in probability distribution, the randomness is equivalent to the instability in various states, which also coincides with the definition of the dimension in the mathematical morphology. While the inner ring fault and the rolling element failure are fault signals with obvious impact, so their dimensions are relatively small, but there is not much difference between the dimensions. While the impacts to the outer ring are relatively large, and the features are relatively obvious, so the dimension is smaller than the other fault types [16]. Although the fractal curves of the inner ring and the rolling element has once intersected, this does not affect the accurate discrimination of them. The fractal dimension of mathematical morphology based on LCD can distinguish the four states better. The experimental results show the effectiveness and feasibility of the method.

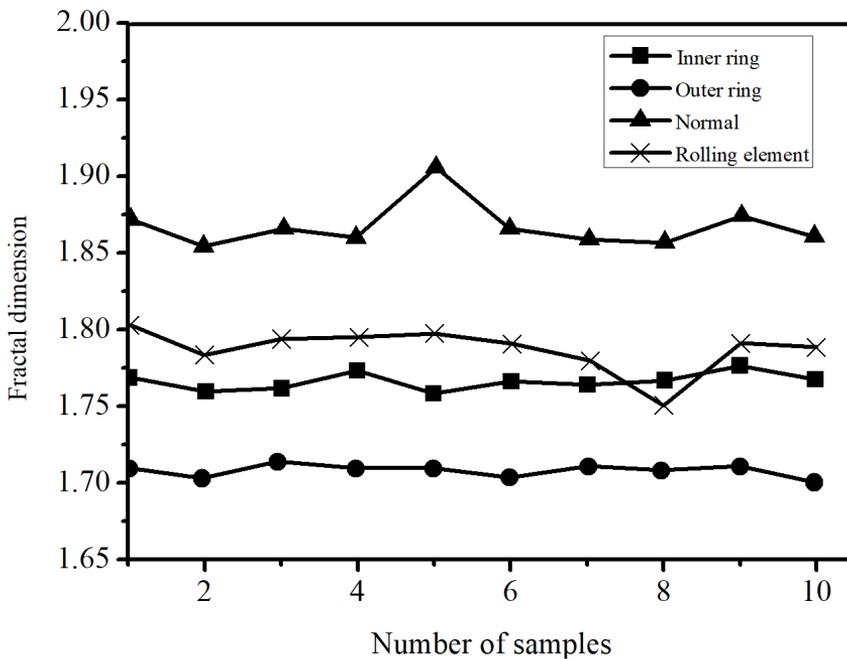


Fig. 5. The first CISC morphology fractal dimension of bearing fault signals

#### 4. Conclusion

In this paper, a fault diagnosis method of rolling bearing based local feature scale decomposition and fractal dimension is studied. The simulation results showed that the mathematical morphology has better accuracy and computational efficiency than the box dimension. The LCD method can separate the fault signal characteristic component from the background noise or other interference signal to improve the accuracy of the fault identification. After the fault signal is decomposed by the LCD

method, the ISC component of the characteristic fault signal is obtained. The fractal dimension of each ISC component is calculated and used as the characteristic parameter to judge the state. At last, the normal bearing, rolling element failure, inner ring fault and outer ring fault are analyzed. The results show that the method based on LCD decomposition and morphology fractal dimension can effectively realize the diagnosis of rolling bearing fault state.

## References

- [1] A. J. HU, G. J. TANG, L. S. AN: *Denoising technique for vibration signals of rotating machinery based on mathematical morphology filter*. Chinese Journal of Mechanical Engineering 42 (2006), No. 4, 127–130.
- [2] H. LI, R. WANG, S. CAO, Y. CHEN, W. HUANG: *A method for low-frequency noise suppression based on mathematical morphology in microseismic monitoring*. Geophysics 81 (2016), No. 3, V159–V167.
- [3] S. SUMIJAN, S. MADENDA, J. HARLAN, E. P. WIBOWO: *Hybrids Otsu method, feature region and mathematical morphology for calculating volume hemorrhage brain on CT-Scan image and 3D reconstruction*. TELKOMNIKA Telecommunication, Computing, Electronics and Control 15 (2017), No. 1, 283–291.
- [4] D. LEGLAND, I. ARGANDA-CARRERAS, P. ANDREY: *MorphoLibJ: Integrated library and plugins for mathematical morphology with Image*. Journal Bioinformatics 32 (2016), No. 22, 3532–3534.
- [5] B. BOURAOUI, C. RONSE, J. BARUTHIO, N. PASSAT, P. GERMAIN: *3D segmentation of coronary arteries based on advanced mathematical morphology techniques*. Computerized Medical Imaging and Graphics 34 (2010), No. 5, 377–387.
- [6] Y. LI, J. ZHAO, B. MO: *Small target segmentation algorithm based on wavelet transform and mathematics morphologic*. International Journal of Electronics and Electrical Engineering 1 (2013), No. 2, 76–79.
- [7] T. HAN, D. JIANG, X. ZHANG, Y. SUN: *Intelligent diagnosis method for rotating machinery using dictionary learning and singular value decomposition*. Sensors (Basel) 17 (2017), No. 4, paper 689.
- [8] J. ATIF, C. HUDELLOT, I. BLOCH: *Explanatory reasoning for image understanding using formal concept analysis and description logics*. IEEE Transactions on Systems, Man, and Cybernetics: Systems 44 (2014), No. 5, 552–570.
- [9] C. LU, Y. WANG, M. RAGULSKIS, Y. CHENG: *Fault diagnosis for rotating machinery: A method based on image processing*. PLoS ONE 11 (2016), No. 10, e0164111.
- [10] T. XU, G. WANG, H. WANG, T. YUAN, Z. ZHONG: *Gap measurement of point machine using adaptive wavelet threshold and mathematical morphology*. Sensors (Basel) 16, (2016), No. 12, e2006.
- [11] A. HU, L. XIANG, G. TANG, Y. DU: *Fault feature extracting method of rotating machinery based on mathematical morphology*. Chinese Journal of Mechanical Engineering 47 (2011), No. 23, 92–96.
- [12] L. SHEN, X. J. ZHOU, W. B. ZHANG, Z. G. ZHANG: *Denoising for vibration signals of a rotating machinery based on generalized mathematical morphological filter*. Journal of Vibration and Shock 28 (2009), No. 9, 70–73.
- [13] S. WANG, J. ZHANG, Y. LI, S. ZHANG: *Rotating machinery fault diagnosis based on mathematical morphology and fuzzy clustering*. Chinese Journal of Scientific Instrument 33 (2012), No. 5, 1055–1061.
- [14] Y. FENG, B. LU, D. ZHANG: *Multiscale singular value manifold for rotating machinery fault diagnosis*. Journal of Mechanical Science and Technology 31 (2017), No. 1, 99 to 109.

- [15] J. XIONG, Q. ZHANG, Z. PENG: *A diagnosis method for rotation machinery faults based on dimensionless indexes combined with K-nearest neighbor algorithm*. Mathematical Problems in Engineering (2015), No. 4, 1–9.
- [16] R. JIA, X. Y. WANG, Z. H. CAI, L. ZHANG, X. Q. LUO: *The Hilbet-Huang transform based on least squares support regression machine and its application in the fault diagnosis of hydroelectric generation unit*. Proceedings of the CSEE 26 (2006), No. 22, 128–133.

Received June 29, 2017

# Design and application of computer network system integration based on network topology

SHUNA ZHOU<sup>1</sup>

**Abstract.** To explore the computer network system that covers more and more areas, including the management of network, system, operating environment and other aspects, reasonable layout and unified planning to the network are carried out in relatively complex computer network system engineering. What's more, the network center, the open network room, the network management running system and the network security system based on the topology are also designed. In addition, the algorithm implementation is carried out to the topology construction process. In the design, the internal network topology is determined according to the actual internal distribution of the buildings in campus, and the basic engineering design is divided into work areas, buildings, trunk and management, and the NFCT algorithm operation and realization is carried out in the process of the topology construction. The results showed that the reliability and accuracy of the network system is realized through the final verification and analysis of the results. In a word, the computer network system integration based on network topology has good performances and it can be applied in computer related fields.

**Key words.** Network topology, computer network, system integration.

## 1. Introduction

Computer network is the foundation and core of promoting informatization, digitalization and globalization, because computer network system is an open and digital integrated information system. All kinds of application system based on computer network, through the comprehensive collection, storage, transmission, processing and utilization of the digital information, linked all kinds of social factors closely. As a result, the computer network has become the most important infrastructure in information society [1]. With the development of economy and the implementation of the strategy of invigorating the country through science and education, the construction of computer network has gradually become the basic construction project of many fields, and has become an important symbol to measure the informatization

---

<sup>1</sup>Manzhouli Branch of Inner Mongolia University, Inner Mongolia Manzhouli 021400, China

and modernization. The computer network construction project is the basic condition to realize the overall quality education and the cultivation of creative talents. Through the implementation of the computer network project, it not only can the network platform, but more importantly, it will be beneficial for the sharing and cooperation of educational resources, and it is conducive to realize modern distance education. The construction of computer network is a fundamental way to realize the educational informatization. Rich and colorful, healthy and clean computer network culture will become the new way to cultivate way of thinking, moral quality, and creative ability [2]. In addition, it becomes a new platform for all students, to cultivate all-round developed high-quality personnel.

Computer networking construction is the foundation and core of promoting informatization, digitization and globalization. The design of computer network integration based on network topology collects, stores, transmits, processes and utilizes various types of information through network topology to closely link with the various components and interconnect various devices through transmission media. Computer network system covers the management of operation environment, equipment and other aspects, is a synthesis of variety technologies such as information, management science, computer and network technology, and is a relatively complex systems engineering.

With the development of the economy and the implementation of the national strategy of rejuvenating the country through science and education, the university network construction has gradually become the basic construction project of the school, and become an important symbol of measuring the informatization and modernization of education [3]. In this paper, the design and application of computer network system integration architecture is mainly from the network topology, and the interconnection design based on the topology technology is taken the university computer network as the object. The campus network has the characteristics of large number of computers, rich network application, complex topology and high bandwidth requirement. The rational distribution of topology design is carried out to the university computer network will help solve the network bottleneck and congestion, further strengthen the reliability and accuracy of the integrated system.

## **2. Summary of network topology and system integration development**

The computer connective way is called the network topology. Network topology is the physical layout that uses transmission media to interconnect with various devices, especially the location of the computer distribution and how the cable through them. Classification of network topology: network topology can be divided into two categories according to the communication channel of the communication subnet, which are the topology of the broadcast communication channel subnet and the topology of the point-to-point communication subnet [4]. The basic topological structure of the broadcast communication channel subnet main has bus shape, tree shape, ring shape, wireless communication and satellite communication. The basic topological structure of point-to-point communication subnets are star shape, ring

shape, tree shape and mesh shape.

The topological structure of the network is divided into logical topological and physical topological structure. Bus topology is a topology based on multi-point connections, and all devices connect to a common transmission medium. Star topology puts a central computer in the center, each arm of the endpoint is placed one, all the data packages and messages through the central computer to communicate, and each has only one connection except the central machine. This structure requires a large number of the cable, and star topology can be seen as a layer of the tree structure that does not require multi-layer access contention. Star topology is more common in the network cabling [5]. Daisy-chain topology is similar to a ring topology, but with a pair of breakpoints in the middle. Several topologies can be mixed and star topology is more common in campus network planning.

With the rapidly development of the computer technology, the computer network system integration gradually developed. The development of computer network system integration technology is divided into three stages, namely, single integration technology, distributed integration technology and information integration technology. In order to ensure that the designed target of the network integration system is completed on time with high quality, the certain steps should be followed. The steps of the network system integration design mainly are network size, network topology structure, network protocols, equipment instruction, IP address planning, network security design and so on. In the actual operation, some procedures according to the actual situation, if necessary, can be omitted as appropriate. In real life, computer network integration technology has a lot of applications. Due to the rapidly development of computer network integration technology, many industries have the intersection with it, the computer network integration technology can effectively improve the security and reliability of the data.

### 3. Materials and methods

#### *3.1. Overall scheme design of universities network*

Because the campus network have a higher stability requirement on the network, it is necessary to use dual-core in the topology design to do two unit standby and double exports. This not only avoids the possibility of failure in a certain extent, but also makes users shunt, achieving load balancing. Campus network set up the core switchboard to interconnection with the help of the link, and the Gigabit fiber-optic link is adopted between all convergence switches and core switches [6]. When the two core switches can work properly, they share all the communication data of the inter-changer to achieve the load balancing of network access. When one of the core equipment is failure, other core equipment quickly assumes all the tasks.

The export route is planed according to the actual management mode, you can select an export device, according to the site that the user clicks to determine the user uses the education network or telecommunications or Netcom, etc., so that achieving the unified management to the users, as well as the control to the networked computer [7–8]. It can greatly improve the network security for management personnel.

Link redundancy is mainly for the communication lines between the core switches and aggregation interchanger; so that each aggregation interchanger can connect with two core switches with fiber-optic to ensure the stability of the network.

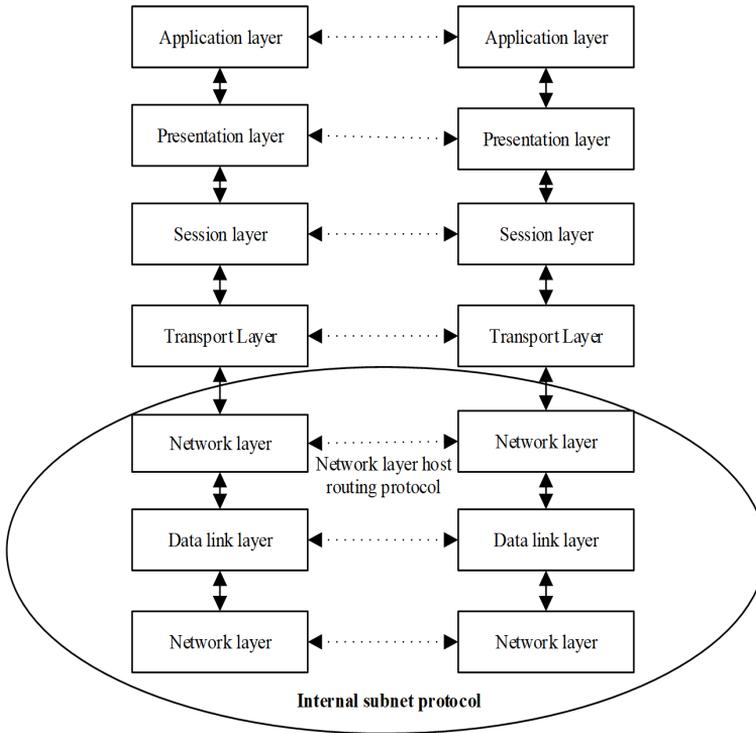


Fig. 1. System interconnection model

### 3.2. Development scheme design

The main functions of the university network include providing the basic network service function, providing the office automation of each management institution in the school, providing the interconnection with the other application system through the WAN interface. The design principle is set according to the specific environment. In the initial period of network construction, the access control is the key, and the system construction uses star topology to make the backbone of the network server to provide better maintenance and optimization of network service management with Gigabit Ethernet technology [9]. In the network system construction, the long-distance teaching and the multimedia application are supported, and the authentication charge management platform and the firewall are also supported. In the latter part of the network, the global security and the deep security are the focus, and the actual location of the building determine that the admissions room server and the client can not be directly connected with the core device. In doing isolation, it is enough to shield the others. The admissions room access is allowed to visit the

college internal resources. Export, trainees' general topology general firewall and core routing can ensure the reliability and stability of the network operation [10].

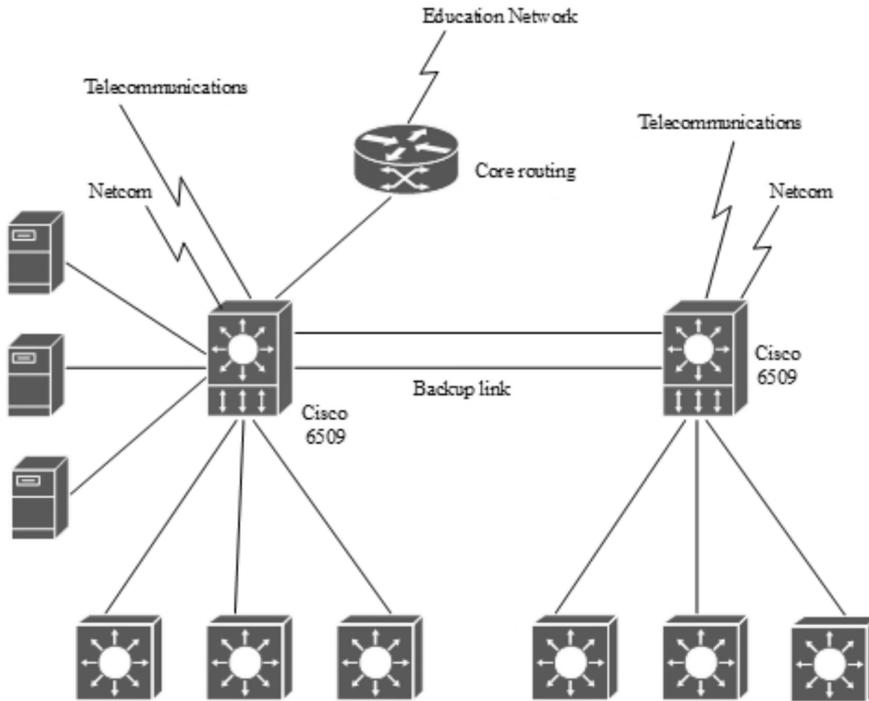


Fig. 2. General topology structure

### 3.3. Algorithm design

In the life cycle of the topology network, it can be divided into three phase according to the fault-tolerant clustering topology control algorithm (NFCT): cluster topology construction, fault detection and topology maintenance. In the phase of cluster topology construction, clustering and cluster head election mechanism are used to establish the cluster topology structure [11]. The topology fault detection phase is completed by the cluster construction in the network, and starts to execute the task and collects the information, also known as the running phase. At this phase, the fault detection mechanism is used to determine whether the cluster is faulty or not. If a fault occurs, the topology maintenance process is triggered. In the maintenance phase, it processes the faulty nodes or judges whether the cluster heads are still suitable as cluster heads or not. Once they are not suitable for cluster heads, the backup cluster heads are transformed into cluster heads to avoid affecting the operation of the cluster or even the stability of the entire network. In order to better describe the algorithm, the relevant definitions are introduced:

Cluster information: Chf. This is the association information that the node selects cluster head, which includes the cluster head ID, the cluster parameter cp of

the cluster head, and the residual energy  $\text{resE}$  of the cluster head [12].

The list of neighbors in the cluster:  $\text{CnList}$ . It is a list of neighbor nodes within the same cluster except the cluster head. Each entry in the list includes the neighbor's ID, clustering parameters, and residual energy.

Members of the cluster list:  $\text{CmList}$ . It is list of all members of the cluster head. Each entry in the list includes the member ID, clustering parameters, and residual energy.

The NFCT algorithm uses a local maintenance mechanism that combines the energy triggering with the fault triggering. The NFCT triggers the topology maintenance process when a network fault occurs. Then, the difference between the residual energy of the cluster head and the average energy in the cluster is calculated, when the energy difference is less than zero, the topology maintenance process is triggered.

The cluster head neighbor graph is composed of the cluster head  $u$  and its cluster members  $v_i$ ,  $i = 1, 2, \dots, k$ , expressed as

$$V_c \in \{v_i \in CM_u \cup u\}, v_i \in N_u, E_c = \{(v_i, u) | v_i \in CMList_u\}. \quad (1)$$

Here,  $CM_u$  is the head of cluster message,  $N_u$  is the collection of node's neighbors. Symbol  $E_c$  denotes the energy cost, and  $CMList_u$  is the head of cluster member list.

The sum of the residual energy, including the cluster-head  $u$  and all its member nodes  $v_i$ ,  $i = 1, 2, \dots, k$  in the neighborhood graph  $G_c(V_c, E_c)$  of the cluster head  $u$  is called the total residual energy of the cluster head  $u$  neighbor graph

$$\text{Sum}_c(u) = \sum_{i=1}^k E_{\text{res}}(v_i) + E_{\text{res}}(u). \quad (2)$$

Here,  $E_{\text{res}}(v_i)$  represents the residual energy of the cluster members nodes  $v_i$  and  $E_{\text{res}}(u)$  represents the residual energy of the cluster head  $u$ ,  $k$  being the number of member nodes in the cluster [13].

The ratio of the sum of the residual energy of the cluster head  $u$  and all its member nodes  $v_i$ ,  $i = 1, 2, \dots, k$  and the residual energy of the cluster head  $u$  in the neighbor graph  $G_c(V_c, E_c)$  of the cluster head  $u$  and the sum of these nodes is the average residual energy of the cluster- $u$  neighbor graph, expressed as

$$\text{Aver}_c(u) = \frac{1}{k+1} \left( \sum_{i=1}^k E_{\text{res}}(v_i) + E_{\text{res}}(u) \right). \quad (3)$$

The topology maintenance process is carried out in the neighborhood of any cluster head  $u$  in the network, only needs the local information of the cluster node, not the network's global information and location information. The local information can be obtained through the normal communication "incidentally", reducing the communication overhead [14]. In the maintenance phase of the algorithm, a trigger mechanism combining fault and energy is used to trigger the local topology maintenance in the neighbor graph of the cluster head when the network node fails or the

residual energy of the cluster head is lower than the average energy in the cluster, so the algorithm is applicable to topology control of energy-sensitive large-scale wireless sensor networks.

## 4. Results

In order to verify the validity of the NFCT algorithm, the NFCT algorithm is compared with the typical backup mechanism algorithm P-CDS, FTCB and k-connected algorithm CBCC ( $k = 2$ ). The simulation testing is carried out in the three methods from the cluster head size, cluster head node reliability, energy consumption, communication traffic, and network life cycle. In order to make the simulation results closer to the real wireless environment, and compare with other algorithms in the same environment, in addition to assumptions, the environmental simulation parameters items is set out which is shown in Table 1 to standardize the experimental environment, including network conditions, nodes attribute conditions, packet size and energy consumption and other related settings.

Table 1. Simulation parameters table

Parameter	Value
Network size	$500 \times 500 \text{ m}^2$
Number of nodes	50–400
Aggregation node position	(0, 0)
Primary energy	1 J
$E_{\text{elec}}$ (Electric energy consumption in info processing)	50 nJ/bit
$\xi_{\text{fs}}$ (Energy loss on forward)	10 pJ/bit/m <sup>2</sup>
$\xi_{\text{mp}}$ (Energy loss on cluster switch)	0.0013 pJ/bit/m <sup>2</sup>
$E_{\text{DA}}$ (Energy loss on delay)	5 nJ/bit
Transmission range $R$	100 m

The size of the backbone nodes not only affects the energy consumption of the network but also the communication efficiency of the network. Therefore, the backbone node is measured by the ratio of the backbone node to the total number of nodes in the network. In addition, the NFCT algorithm and P-CDS, FTCB algorithm through the backup mechanism to tolerate the network fault, so the nodes number of the two algorithms in network is also tested. The results are shown in Figs. 3 and 4.

It can be seen from Fig. 3 that the connected CBCC algorithm has more backbone nodes than the NFCT and P-CDS algorithm, because each backbone node is equipped with redundant nodes to construct the backbone network, which results in more nodes carrying backbone tasks. In addition, it can be seen from the figure, NFCT algorithm has smaller backbone nodes and backup nodes than FTCB and P-CDS, so it can improve the network's energy utilization efficiency.

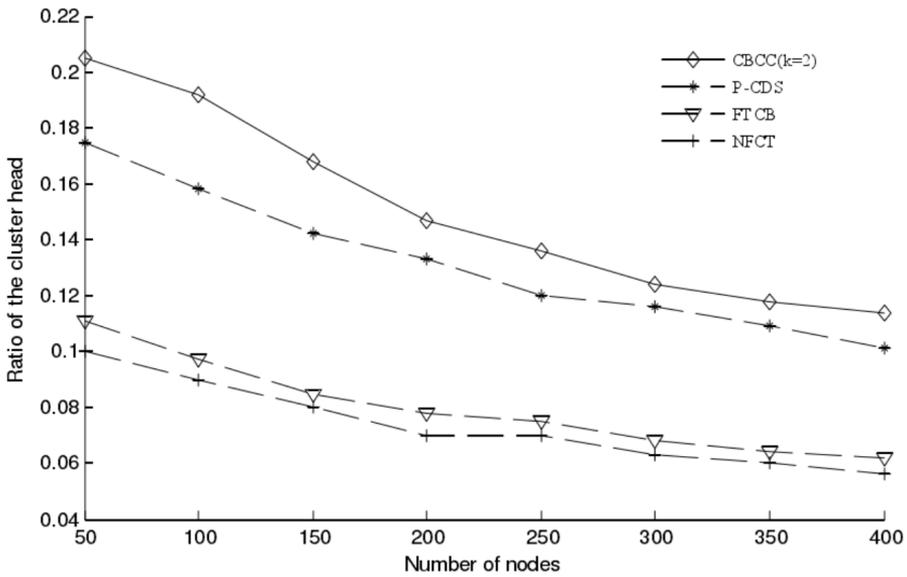


Fig. 3. Comparison of backbone nodes in network

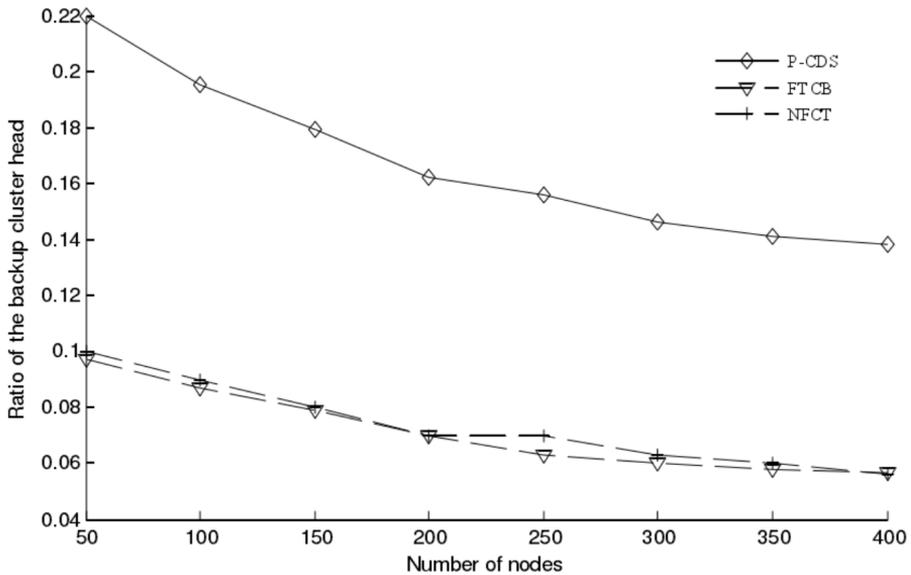


Fig. 4. Comparison of the backbone node and the backup node in the network

The simulation results of the typical algorithms  $k$ -connected FC-CBCC algorithm ( $k = 2$ ) and the P-CDS and FCTB algorithms based on the backup mechanism in the fault-tolerant topology control algorithm show that the NFCT algorithm has much smaller backbone nodes, higher communication traffic and higher backbone node

reliability than the other three algorithms, and it can be applied to the traditional, which can effectively extend the life cycle of the university computer network to ensure the stability and accuracy of the network topology structure and integration system [6].

In the design and application of system integration based on the network topology, the rational distribution and unified planning are carried out in the network, and the fault-tolerant design is considered at each stage of the NFCT algorithm. Firstly, in the cluster phase, the cluster head backup mechanism is used to enhance the fault-tolerant ability of clusters and improve the stability of cluster operation. Secondly, in the process of network operation, a lightweight fault detection mechanism is adopted to detect the failure of the cluster heads and members in the network in real time. Once the faults are detected, the failure nodes are processed to further improve the network fault tolerance. Finally, a topology maintenance mechanism is used to monitor the residual energy of the cluster head node in real time. When the residual energy is lower than the average residual energy value in the cluster, the backup cluster head becomes cluster head to maintain the stable operation of the network and improve the network fault tolerance.

## 5. Conclusion

The computer network system integration based on the network topology can carry out the rational distribution and unified planning to the network in relative complex computer network system engineering. Taking the university computer network as an example, we design and apply the university network integration system of the backbone network, the network center, the open network room, the network management running system and the network security system based on the topology. In the design, the internal network topology is determined according to the actual internal distribution of the buildings in campus, and the basic engineering design is divided into work areas, buildings, trunk and management, and the NFCT algorithm operation and realization is carried out in the process of the topology construction. The reliability and accuracy of the network system is realized through the final verification and analysis of the results.

## References

- [1] M. RAHAMAN, M. M. ISLAM: *A review on progress and problems of quantum computing as a service (Qcaas) in the perspective of cloud computing*. Global Journal of Computer Science and Technology: B Cloud and Distributed 15 (2015), No. 4, 15–18.
- [2] A. A. IBRAHIM, T. A. KHALAF, B. M. AHMED: *Design and implementation of iris pattern recognition using wireless network system*. Journal of Computer and Communications 4 (2016), No. 7, 15–21.
- [3] N. P. KUMAR, D. MISHRA, S. RAM: *Dynamic e-epidemic model for active infectious nodes in computer network*. Journal of Statistics and Management Systems 19 (2016), No. 2, 247–257.
- [4] R. T. GOSWAMI, B. K. MISHRA: *Information and the dynamics of SEIR eepidemic*

- model for the spreading behavior of malicious objects in Computer Network*. International Journal of Engineering Science and Technology (IJEST) 4 (2012), No. 10, 4275–4282.
- [5] N. YU: *The risk evaluation research of computer network security based on grey clustering*. International Journal of Security and Its Applications 10 (2016), No. 8, 191–200.
- [6] X. Y. QU, F. ZHANG, H. F. XUE: *Research on new algorithm of network topology design based on cloud integration*. Applied Mechanics and Materials 190–191 (2012), Chapter No. 1, 360–363.
- [7] D. Q. YAO, Y. TAO, H. R. YU: *Based on the integration of power grid topology model of distribution network fault analysis applications*. Advanced Materials Research 1006–1007 (2014), Chapter No. 13, 1194–1198.
- [8] S. WANG, J. XIE, W. YANG, B. SUN: *Preparative separation and purification of alky-lamides from zanthoxylum bungeanum maxim by high-speed counter-current chromatography*. Journal of Liquid Chromatography & Related Technologies 34 (2011), No. 20, 2640–2652.
- [9] J. W. WANG, H. F. WANG, J. L. DING, K. FURUTA, T. KANNO, W. H. IP, W. J. ZHANG: *On domain modelling of the service system with its application to enterprise information systems*. Enterprise Information Systems 10 (2016), No. 1, 1–16.
- [10] G. MISRA, V. KUMAR, A. AGARWAL, K. AGARWAL: *Internet of things (IoT) – A technological analysis and survey on vision, concepts, challenges, innovation directions, technologies, and applications (An upcoming or future generation computer communication system technology)*. American Journal of Electrical and Electronic Engineering 4, (2016), No. 1, 23–32.
- [11] J. L. HE, B. ZHANG, R. ZENG, G. YU: *Grounding system design of 1000 kV ultra-high voltage substation*. Proceedings of the CSEE 29 (2009), No. 07, 7–12.
- [12] J. L. ZHAO, L. C. CHEN, L. H. PAN, Y. J. ZHANG, B. H. XIE: *Quantitative evaluation method for component security based on FAHP*. Networks and Communications 37 (2011), No. 22, 67–69.
- [13] P. K. KUMAR, K. VIJAYALAKSHMI, R. BHARATHI: *Performance analysis of PCA and artificial intelligence techniques for the detection of application layer attacks*. International Journal of Applied Engineering Research 10 (2015), No. 2, 2531–2540.
- [14] M. KRARTI: *An overview of artificial intelligence-based methods for building energy systems*. Journal of Solar Energy Engineering 125 (2003), No. 3, 331–342.

Received June 29, 2017

# Evaluation of transformer insulation condition based on cloud matter element model

RONGSHENG LIU<sup>1</sup>, MINFANG PENG<sup>1,4</sup>, XUN WAN<sup>2</sup>,  
HAIYAN ZHANG<sup>1</sup>, WEI ZHOU<sup>3</sup>

**Abstract.** The purpose of this paper is to study the insulation state of transformer. Insulation system is an important part of the transformer, and it is the basic condition for the normal operation and operation of the transformer. The insulation condition assessment of transformer is of great significance to guide the condition based maintenance of the transformer, to enhance the life cycle management and to save the cost of operation. Based on the cloud matter element theory, a method for evaluating the insulation state of transformers is proposed. The results show that the transformer insulation state evaluation model can effectively integrate various state parameters, and accurately assess the insulation state of the transformer each insulation parts and the whole part. Based on the above findings, we conclude that the model is suitable for evaluating the insulation state of transformers.

**Key words.** Transformer, insulation state, cloud model, matter element theory.

## 1. Introduction

Power transformer is the core of energy-conversion and transmission in the grid network, it is one of the most important equipment in power system, and it is in a pivotal position in power system. Its running state directly affects the security and reliability of the whole power system. For a long time, the maintenance of transformer in China adopts the planned maintenance mode, but planning maintenance has serious limitations and defects, which may lead to "maintenance excesses" and "lack of maintenance", causing the equipment effective time utilization loss and waste of manpower, financial and material resources, and adding new risks. In re-

---

<sup>1</sup>College of Electrical and Information Engineering, Hunan University, Changsha, 410082, China

<sup>2</sup>State Grid Hunan Electric Power Company Electric Power Research Institute, Changsha, 410007, China

<sup>3</sup>Hunan Rural Credit Cooperatives Union, Hunan University, Changsha, 100101, China

<sup>4</sup>Corresponding author

cent years, with the gradual improvement of unit condition monitoring technology, the maintenance mode of transformer has changed from planned maintenance to predictive maintenance (condition maintenance) based on condition monitoring. It is to make use of life characteristics of some important parts of the transformer, through the advanced detection means for data acquisition, to compare and analyze the transformer operation history and the operation conditions of the same type of transformer, to evaluate the current operation state of transformers, and to predict its development trend.

A large number of data show that the main reason for the failure of equipment is the deterioration of its insulation performance, and most of the failure of electrical equipment are insulation failure. In consequence, the transformer insulation system is the basic condition for the normal operation and operation of the transformer. Evaluate the transformer insulation system condition, determine the insulation state of the insulation oil, solid insulation and insulation casing and other components. It not only provides theoretical support and reference data for the life prediction of the transformer, and provide scientific basis for effectively realizing the transformer condition maintenance, making reasonable operation, protection, and updating plan. As a result, it has important academic value and practical significance to study the insulation condition assessment of the transformer.

## 2. State of the art

The cloud model is, based on probability theory and fuzzy set theory these two theories mutually penetrating, by constructing a specific algorithm, to form the transformation model between qualitative concept and quantitative representation, and to reveal the internal relationship between the randomness and the fuzziness.

Let  $U$  be set to the quantitative domain composed of accurate numerical values, and  $A$  is a qualitative concept on  $U$ . For an arbitrary element  $x$  on  $U$ , there is a random number  $\mu(x) \in [0, 1]$  with a stable tendency, which is called the membership of  $X$  for  $A$ , and the distribution of membership in the domain is referred to as the cloud. Each  $x$  is called a cloud.

The numerical characteristics of the cloud are represented by using three values, respectively, the expectation ( $Ex$ ), the entropy ( $En$ ) and the hyper entropy ( $He$ ). The expectation indicates the expectation of the domain spatial distribution of droplet, referring to the value that can most represent the qualitative concept in the domain space, which reflects the gravity position of the cloud [1]. The cloud at this position 100 % belongs to the qualitative concept. Entropy is a measure of the uncertainty of the qualitative concept. In general, the larger the entropy, the vaguer the concept, and the more difficult it is to quantify the concept. The hyper entropy is the entropy of entropy, and it is the uncertainty measure of entropy [2]. It reflects the degree of coherence of all points in the domain space, namely the condensation degree of cloud droplets, the greater the hyper entropy, the greater the randomness of each cloud membership, and the greater the thickness of the cloud.

Cloud generation algorithm is called cloud generator, which establishes the mapping relationship of qualitative and quantitative inter-connection, inter-dependence,

and inter-transformation. Cloud generator, according to the calculation direction, can be divided into positive direction cloud generator and reverse cloud generator; according to the different dimensions of the cloud, it can be divided into one-dimension cloud generator and two-dimension cloud generator. The clouds involved in this paper are one-dimension clouds. Normal cloud generator is the most basic cloud generator, with universal applicability. The following will focus on the generation rules and related algorithms of normal cloud.

Positive normal cloud generator is a mapping from qualitative to quantitative, achieving the scope and distribution regular of quantitative data obtained from qualitative information expressed by the language value, which is a forward and direct process, as shown in Fig. 1.



Fig. 1. One-dimensional positive cloud generator

For the one-dimension normal cloud, three digital characteristics (Ex, En, He) of cloud are given, and the generated required cloud  $\text{drop}(x_i, \mu_i)$  algorithm is shown in Fig. 2.

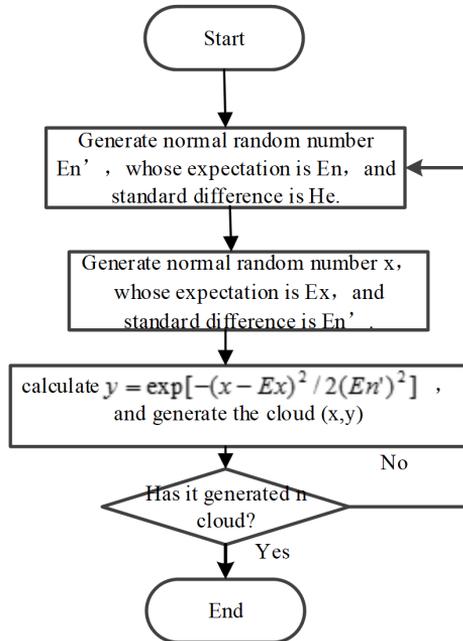


Fig. 2. Algorithm flow of one-dimensional normal positive cloud generating droplet

Reverse normal cloud generator effectively converts a certain number of accurate

numerical to the qualitative concept represented by digital features (Ex, En, He), which is a reverse and indirect process, as shown in Fig. 3.



Fig. 3. One-dimensional reverse cloud generator

Reverse normal cloud generator algorithm is based on the principle of statistics, including using the membership information and not using the membership information these two basic algorithms, and the reverse cloud generator by using the membership information is shown in Fig. 4.

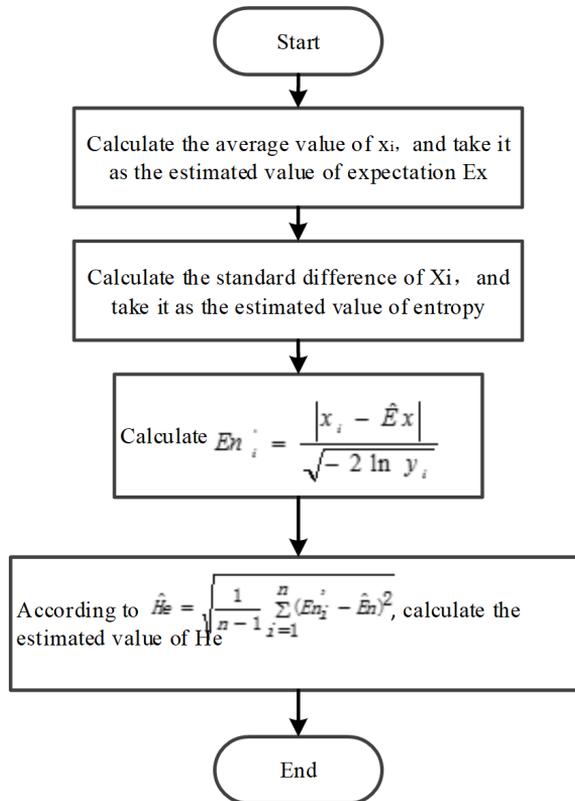


Fig. 4. Algorithm flow of reverse cloud generator by using membership information

For a given thing, represented by the name  $N$ ,  $N$  has the characteristics of  $c$ , and its value is  $v$  [3]. The names, characteristics and values of things are orderly combined to form  $R = (N, c, v)$  as a basic element to describe things, referred to as matter element.

If the matter  $N$  has  $n$  features  $c_1, c_2, \dots, c_n$ , then, using these characteristics and corresponding values of  $v_1, v_2, \dots, v_n$ , the description of thing  $N$  can be expressed as

$$R = \begin{bmatrix} N, & c_1, & v_1 \\ & c_2 & v_2 \\ & \dots & \dots \\ & c_n & v_n \end{bmatrix} = \begin{bmatrix} N, & c_1 & c_1(N) \\ & c_2 & c_2(N) \\ & \dots & \dots \\ & c_n & c_n(N) \end{bmatrix}. \quad (1)$$

Here,  $R$  is called a multidimensional matter element, in which  $R_i = (N, c_i, v_i)$ , ( $i = 1, 2, \dots, n$ ) is called the sub matter element of  $R$ .

### 3. Methodology

#### 3.1. Establishment of evaluation index system of transformer insulation

Transformer insulation system is a system composed of insulating oil and solid insulation. There is no standard index system used to evaluate the insulation condition of the transformer [4]. In this paper, according to the different classifications and compositions of insulation, the index hierarchy evaluation system is established, as shown in Table 1.

#### 3.2. Determination of index weight

In the evaluation process, the weight is an objective reflection of the index status and subjective measurement of relative importance of each index. According to the advantages of cloud model in expressing the fuzziness and randomness of natural language, this paper introduces the cloud model into the traditional analytic hierarchy process, and improves the analytic hierarchy process to make the evaluation result more objective.

In the improved analytic hierarchy process, the scale of the element importance is represented by the following 9 cloud models:  $C_1(Ex_1, En_1, He_1)$ ,  $C_2(Ex_2, En_2, He_2)$ ,  $C_3(Ex_3, En_3, He_3)$ ,  $C_4(Ex_4, En_4, He_4)$ ,  $C_5(Ex_5, En_5, He_5)$ ,  $C_6(Ex_6, En_6, He_6)$ ,  $C_7(Ex_7, En_7, He_7)$ ,  $C_8(Ex_8, En_8, He_8)$ ,  $C_9(Ex_9, En_9, He_9)$ . Among them, the expectations  $Ex_1$  to  $Ex_9$  were 1 to 9. For the cloud model of each scale, use the following assumptions: since that the meaning that 1, 3, 5, 7, and 9 these 5 grades express is relatively clear, expert's judgment about them is relatively clear [5]. While the judgment for other language value is relatively vague, so the cloud model used will not be the same. According to the principle of the golden section method, the entropy and the entropy of each cloud model are obtained in the form

$$En_1 = En_3 = En_5 = En_7 = En_9 = 0.382(\chi_{\max} - \chi_{\min})\alpha/6 = 0.437, \quad (2)$$

$$En_2 = En_4 = En_6 = En_8 = En_1/0.618 = 0.707, \quad (3)$$

Table 1. Index hierarchy evaluation system

	Item layer	Sub item layer	Index layer
Insulation state of transformer X	Oil insulation state X1	Oil chromatographic analysis X11	H <sub>2</sub> content X111
			C <sub>2</sub> H <sub>2</sub> content X112
			Total hydrocarbon content X113
		Oil test X12	Micro water X121
			Acid value X122
			Breakdown voltage X123
	Oil dielectric loss value X124		
	Solid insulation state X2	Aging index X21	CO+CO <sub>2</sub> content X211
			Furfural X212
			Polymerization degree X213
		Insulation test X22	Insulation resistance X221
			Absorption ratio X222
			Winding leakage current X223
			Dielectric loss factor X224
	Casing insulation state X3	Casing oil chromatographic analysis X31	H <sub>2</sub> content
			CH <sub>2</sub> content X312
			C <sub>2</sub> H total content X313
		Casing electrical test X32	Capacitance X321
Dielectric loss X322			
Main screen insulation resistance X323			
End shield to ground insulation resistance X324			

$$He_1 = He_3 = He_5 = He_7 = He_9 = 0.382(\chi_{\max} - \chi_{\min})\beta/3 = 0.073, \quad (4)$$

$$He_2 = He_4 = He_6 = He_8 = He_1/0.618 = 0.118. \quad (5)$$

According to the characteristic number of each cloud model, the cloud model of each scale grade can be obtained by using the generating algorithm of forward normal cloud.

The analytical hierarchy process method based on cloud model scale judgment matrix starts from its determination. The judgment matrix of the comparison of the elements importance in a certain layer after the aggregation of the experts group is as follows:

$$\begin{aligned}
 & \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} = \\
 & = \begin{bmatrix} C_{11}(Ex_{11}, En_{11}, He_{11}) & C_{12}(Ex_{12}, En_{12}, He_{12}) & \cdots & C_{1n} \\ C_{21}(Ex_{21}, En_{21}, He_{21}) & C_{22}(Ex_{22}, En_{22}, He_{22}) & \cdots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{n1}(Ex_{n1}, En_{n1}, He_{n1}) & C_{n2}(Ex_{n2}, En_{n2}, He_{n2}) & \cdots & C_{nn} \end{bmatrix}. \quad (6)
 \end{aligned}$$

In the above equation, the entropy and hyperentropy of diagonal elements of the cloud model is equal to 0, that is,  $C_{ii}(Ex_{ii}, En_{ii}, He_{ii}) = C(1, 0, 0)$ , element  $c_{ij}$  represents the importance degree of element  $i$  relative to element  $j$ , and for its reciprocal  $c_{ji}$ , it is obtained by cloud reciprocal computing method [7].

After obtaining the judgment matrix, according to the multiplication operation of the cloud model, the relative weights  $W_i(Ex_i, En_i, He_i)$  of the expectation, fuzziness and randomness of elements in the matrix are obtained.

After getting the weights of each index element, it is necessary to carry out consistency validation of the expectation. When it meets  $CR = \frac{CI}{RI} < 0.1$ , through consistency validation,  $CI = \frac{\lambda_{\max} - n}{n - 1}$ , RI corresponding random consistency index can be acquired from the table.

According to the different reflections of each item layer and index on the insulation state and performance of transformer, make use of the improved hierarchical analysis method based on cloud model to determine the weight of each factor. The specific method is: invite 5 industry or technical personnel, according to the comparison angle, compare the element of the target layer, that under the same item layer, and the index importance, and build the original  $A - U$  judgment matrix. This paper takes the oil test sub project level as an example, the  $A - U$  judgment matrix of 5 experts is:

$$W^{(1)} = \begin{bmatrix} 1 & 2 & 1/3 & 1/2 \\ 1/2 & 1 & 1/4 & 1/3 \\ 3 & 4 & 1 & 2 \\ 2 & 3 & 1/2 & 1 \end{bmatrix}, \quad W^{(2)} = \begin{bmatrix} 1 & 1 & 1/3 & 1/2 \\ 1 & 1 & 1/3 & 1/2 \\ 3 & 3 & 1 & 2 \\ 2 & 2 & 1/2 & 1 \end{bmatrix},$$

$$W^{(3)} = \begin{bmatrix} 1 & 1 & 1/4 & 1/2 \\ 1 & 1 & 1/3 & 1/2 \\ 4 & 3 & 1 & 3 \\ 2 & 2 & 1/3 & 1 \end{bmatrix}, \quad W^{(4)} = \begin{bmatrix} 1 & 1 & 1/3 & 1 \\ 1 & 1 & 1/3 & 1/3 \\ 3 & 3 & 1 & 2 \\ 1 & 3 & 1/2 & 1 \end{bmatrix},$$

$$W^{(5)} = \begin{bmatrix} 1 & 2 & 1/3 & 1/2 \\ 1/2 & 1 & 1/3 & 1/3 \\ 3 & 3 & 1 & 1 \\ 2 & 3 & 1 & 1 \end{bmatrix}.$$

For each judgment matrix, the element importance scale is represented by the cloud model, and the judgment matrix of the element importance by the experts is obtained, in which the judgment matrix of expert 1 is shown as follows [8]:

$$W^{(1)} = \begin{bmatrix} (1, 0, 0) & (2, 0.707, 0.118) & (1/3, 0.049, 0.008) & (1/2, 0.177, 0.030) \\ (1/2, 0.177, 0.030) & (1, 0, 0) & (1/4, 0.442, 0.007) & (1/3, 0.049, 0.008) \\ (3, 0.437, 0.073) & (4, 0.707, 0.118) & (1, 0, 0) & (2, 0.707, 0.118) \\ (2, 0.707, 0.118) & (3, 0.437, 0.073) & (1/2, 0.177, 0.030) & (1, 0, 0) \end{bmatrix}.$$

After getting the expert's cloud model judgment matrix, the clouds in the same position are assembled by using the group decision method, and the comprehensive judgment matrix of the 5 experts is obtained [9]:

$$W = \begin{bmatrix} (1, 0, 0) & (1.391, 0.563, 0.094) & (0.320, 0.048, 0.008) & (0.571, 0.209, 0.034) \\ (0.718, 0.291, 0.049) & (1, 0, 0) & (0.327, 0.484, 0.008) & (0.356, 0.062, 0.010) \\ (3.125, 0.470, 0.079) & (3.063, 0.454, 0.076) & (1, 0, 0) & (1.371, 0.547, 0.088) \\ (1.750, 0.640, 0.104) & (2.813, 0.488, 0.079) & (0.730, 0.291, 0.047) & (1, 0, 0) \end{bmatrix}$$

The index weight of the sub item of oil test is [10]

$$W^0 = [(0.157, 0.161, 0.162), (0.119, 0.116, 0.117), (0.420, 0.409, 0.409), (0.304, 0.313, 0.313)].$$

And then it is necessary to carry out the consistency test of the expectation, as shown in Table 2.

Table 2. Consistency validation results

Name	$\lambda_{\max}$	CI	RI	CR
Value	4.0167	0.0056	0.9	0.0062

From Table 2, we can see that the test result is  $CR = 0.0062 < 0.1$ , which meets the requirement of consistency. By using the method mentioned above, carry out weight calculation of other indexes in the index system and the item layer, and the index weight distribution is shown in Table 3.

Table 3. The weights distribution chart of each index

	Item layer	Weights	Sub item layer	Weights	Index layer	Weights
Transformer insulation condition	Transformer oil insulation state	0.301	Oil chromatographic analysis	0.364	H <sub>2</sub>	0.188
					C <sub>2</sub> H <sub>2</sub>	0.392
					Total hydrocarbon	0.420
		Oil test	0.636	Micro water	0.157	
				Acid value	0.119	
				Breakdown voltage	0.420	
				Oil dielectric loss	0.304	
	Solid insulation state	0.563	Aging index	0.485	Furfural	0.360
					Polymerization degree	0.538
					CO+CO <sub>2</sub> content	0.103
		Insulation test	0.515	Insulation resistance	0.139	
				Absorption ratio	0.170	
				Winding leakage current	0.254	
				Dielectric loss factor	0.436	
		Casing insulation state	0.136	Casing oil chromatographic analysis	0.381	H <sub>2</sub>
	CH <sub>4</sub>					0.222
	C <sub>2</sub> H total content					0.528
Casing electrical test	0.619		Capacitance	0.419		
			Dielectric loss	0.297		
			Main screen insulation resistance	0.154		
			End shield to ground insulation resistance	0.130		

## 4. Result analysis and discussion

### 4.1. Steps for evaluation of transformer insulation state based on the cloud matter element model

The insulation condition of power transformer is comprehensively reflected by all the indexes in the index system, and different indicators tend to have different magnitude and dimension, cannot adopt a uniform measurement standard for the analysis of all indexes. Therefore, before the evaluation, each characteristic state should be quantified [11]. This paper uses the relative deterioration degree to measure the insulation status degree of deterioration, the range of degree of deterioration of  $[0, 1]$ , with 0 suggesting insulation in good condition, 1 indicating the insulation in the fault state. The greater the value, the more serious the deterioration.

In this paper, with reference to "Oil immersed transformers (reactors) state assessment guide", considering the state development trend of the transformer, the transformer insulation condition is divided into excellent, good, general, attention, and serious these five state levels, and the deterioration degree of each index partition on the state level is shown in Table 4.

Table 4. Level division of transformer insulation state

Deterioration degree range	0–0.15	0.15–0.35	0.35–0.6	0.6–0.8	0.8–1
State levels	Excellent	Good	General	Attention	Serious

The specific steps for evaluating the insulation of the transformer are shown in Fig. 5.

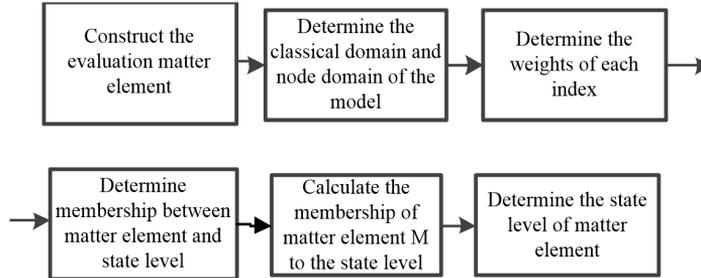


Fig. 5. Specific steps for transformer insulation state evaluation

### 4.2. Case analysis

Tested was a 240 MVA, 220 kV transformer, model for SFPSZ1-240000/220, in 2014 and 2015. Some parts of the number of preventive test values after its building are as shown in Table 5 [12]. The two spectrum tracking data in 2015 are shown in Table 6.

Using the above information, the concrete steps for evaluating the insulation condition of the transformer are as follows: According to the transformer insulation

condition assessment system established in the previous paper, the condition index of each sub item layer constitutes the element to be evaluated [13].

Table 5. Prediction test data

Test items	2014	2015
Absorption ratio	1.48	1.28
Winding leakage current change	23.1	122.7
Winding dielectric loss	0.275 %	0.59 %
Capacitive casing dielectric loss	0.57 %	0.72 %
Capacitance casing	-1.2 %	-1.4 %
Water content in oil	16 mg/l	21mg /l
Oil dielectric loss	1.57 %	2.41 %
Oil breakdown voltage	52 kV	48 kV
Furfural content in oil	0.27 mg/l	0.32 mg/l

Table 6. Oil chromatographic data

Test date	CO	CO <sub>2</sub>	H <sub>2</sub>	CH <sub>4</sub>	C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	Total hydro-carbon
2015.5.25	30.6	310	64	7.7	1.1	1.0	2.5	21.3
2015.8.02	197	2481	92.1	28	3.2	73.6	9.1	113.9

Using the above information, the concrete steps for evaluating the insulation condition of the transformer are as follows:

According to the transformer insulation condition assessment system established in the previous paper, the condition index of each sub item layer constitutes the element to be evaluated [13]:

Oil chromatographic analysis of matter element MX11:

$$R_{M_{X11}} = \begin{bmatrix} M_{X11} & c_{X111} & 0.518 \\ & c_{X112} & 0.64 \\ & c_{X113} & 0.722 \end{bmatrix}.$$

Oil chemical test matter element MX12:

$$R_{M_{X12}} = \begin{bmatrix} M_{X12} & c_{X121} & 0.6 \\ & c_{X122} & 0.675 \\ & c_{X124} & 0.47 \end{bmatrix}.$$

Aging index matter element MX21:

$$R_{M_{X21}} = \begin{bmatrix} M_{X21} & c_{X211} & 0.812 \\ & c_{X212} & 0.8 \end{bmatrix}.$$

Insulation test matter element MX22:

$$R_{M_{X22}} = \begin{bmatrix} M_{X22} & c_{X222} & 1 \\ & c_{X223} & 1 \\ & c_{X224} & 0.738 \end{bmatrix}.$$

Casing electrical test matter element MX32:

$$R_{M_{X32}} = \begin{bmatrix} M_{X32} & c_{X321} & 0.68 \\ & c_{X322} & 0.9 \end{bmatrix}.$$

The weights of each index are determined according to the index weights introduced in the previous paper, and the results are shown in Table 7.

Table 7. Weights of each index

Indexes	Objective indexes	Indexes	Objective indexes
X	0.301, 0.563, 0.136	X12	0.292, 0.383, 0.325
X1	0.364, 0.636	X21	0.376, 0.625
X2	0.485, 0.515	X22	0.163, 0.406, 0.430
X11	0.161, 0.345, 0.493	X31	0.529, 0.471

The evaluation results show that the insulation index of the transformer is in the "attention" state, the deterioration trend is obvious, especially the winding insulation absorption ratio and leakage current exceed the value of attention. In the face of various insulation parts of the transformer, the insulation oil is in the "attention" state, solid insulation and casing insulation are in the "serious" state [14]. On the whole, the transformer insulation is in the "attention" state, close to the "serious" state, the aging phenomenon is more serious, and the possibility of failure is high, supposed to strengthen the monitoring of the transformer, to pay close attention to the development of insulation condition, to arrange the repair as soon as possible, and to timely process insulating oil degassing and filtration.

In 2015, after 1 month of test running, we examined the transformer. We found that the casing cap did not seal the tank completely. Thus the insulation part of the transformer was damp. The insulation oil was dissolved with large amount of air. The above description fit the evaluation model built in the study.

## 5. Conclusion

Insulation system is an important part of the transformer, and it is the basic condition for the normal operation and operation of the transformer. The insulation condition assessment of transformer is of great significance to guide the condition based maintenance of the transformer, to enhance the life cycle management and to save the cost of operation. According to the composition and characteristics of transformer insulation system, a state evaluation model based on cloud matter element theory is established. The evaluation results show that the transformer

insulation index is in a dangerous state, and the trend of deterioration is obvious. In particular, the insulation absorption rate and leakage current of windings are more than the attention value. On the whole, the insulation aging of transformer is more serious, and the possibility of failure is very high. The government should strengthen the monitoring of transformers, pay close attention to the development of insulation status, arrange maintenance as soon as possible, and timely deal with the degassing and filtration of insulating oil. After a month's pilot work on transformer maintenance, we found that the lid was not sealed, causing insulation to damp. The insulation of the transformer is seriously degraded. It is consistent with the model evaluation conclusion. As a result, the model is proved to be correct and reliable according to the analysis of the case.

## References

- [1] T. BO CZAR, A. CICHON, S. BORUCKI: *Diagnostic expert system of transformer insulation systems using the acoustic emission method*. IEEE Transactions on Dielectrics and Electrical Insulation 21 (2014), No. 2, 854–865.
- [2] X. SHE, A. HUANG, R. BURGOS: *Review of solid-state transformer technologies and their application in power distribution systems*. IEEE Journal of Emerging and Selected Topics in Power Electronics 1 (2013), No. 3, 186–198.
- [3] M. GUTTEN, M. BARTLOMIEJCZYK, M. ŠEBÖK: *Mathematical analysis of transformer insulation state by means of composite indicator*. Przegląd Elektrotechniczny 89 (2013), No. 3a, 132–135.
- [4] F. SCHOBBER, A. KÜCHLER, C. KRAUSE: *Oil conductivity – an important quantity for the design and the condition assessment of HVDC insulation systems*. FHWS Science Journal 1 (2013), No. 2, 59–78.
- [5] Y. CUI, H. MA, T. SAH, C. EKANAYAKE: *Understanding moisture dynamics and its effect on the dielectric response of transformer insulation*. IEEE Transactions on Power Delivery 30 (2015), No. 5, 2195–2204.
- [6] W. SIMA, J. SHI, Q. YANG, S. HUANG, X. CAO: *Effects of conductivity and permittivity of nanoparticle on transformer oil insulation performance: Experiment and theory*. IEEE Transactions on Dielectrics and Electrical Insulation 22 (2015) 380–390.
- [7] R. VILLARROEL, D. F. GARCIA, B. GARCIA, J. C. BURGOS: *Moisture diffusion coefficients of transformer pressboard insulation impregnated with natural esters*. IEEE Transactions on Dielectrics and Electrical Insulation 22 (2015), No. 1, 581–859.
- [8] M. F. M. YOUSOF, C. EKANAYAKE, T. K. SAHA: *Examining the ageing of transformer insulation using FRA and FDS techniques*. IEEE Transactions on Dielectrics and Electrical Insulation 22 (2015), No. 2, 1258–1265.
- [9] D. MARTIN, Y. CUI, C. EKANAYAKE, H. MA, T. SAHA: *An updated model to determine the life remaining of transformer insulation*. IEEE Transactions on Power Delivery 30 (2015), No. 1, 395–402.
- [10] J. LIU, R. LIAO, Y. ZHANG, C. GONG, C. WANG, J. GAO: *Condition evaluation for aging state of transformer oil-paper insulation based on time-frequency domain dielectric characteristics*. Electric Power Components and Systems 43, (2015), No. 7, 759–769.
- [11] S. K. OJHA, P. PURKAIT, S. CHAKRAVORTI: *Modeling of relaxation phenomena in transformer oil-paper insulation for understanding dielectric response measurements*. IEEE Transactions on Dielectrics and Electrical Insulation 23 (2016), No. 5, 3190 to 3198.

- [12] Q. YUAN, J. HU, T. C. ZHOU, K. LUO, R. J. LIAO: *Temperature influence on frequency domain dielectric response to transformer oil-paper insulation aging*. High Voltage Apparatus 49 (2013), No. 2, 74–79.
- [13] Z. WANG, X. ZHANG, F. WANG, X. LAN: *Chemical characterization and research on the silicone rubber material used for outdoor current transformer insulation*. Phosphorus, Sulfur, and Silicon and the Related Elements 192 (2017), No. 1, 109–112.
- [14] Z. WANG, X. ZHANG, F. WANG, X. LAN, Y. ZHOU: *Effects of aging on the structural, mechanical, and thermal properties of the silicone rubber current transformer insulation bushing for a 500 kV substation*. Springerplus 5 (2016), No. 1, paper 790.

Received May 7, 2017

# Simulation study on automotive EMB system based on self-tuning fuzzy PID control

LIJUN XIE<sup>1</sup>, WENHUI YANG<sup>1,2</sup>

**Abstract.** To enhance stability of automobile brake, and shorten brake distance, study is made on anti-lock control and its control theory based on electronic mechanical brake system (EMB system). Models related to EMB system was built according to dynamics analysis for automobile braking system, and simulation was made to those models under Anti-Lock Braking System (ABS System). Then, self-tuning fuzzy PID controller was put forward to improve fuzzy control and PID control. Comparison was made between braking mode that without ABS control and self-tuning fuzzy PID. Finally, pavement simulation was made in order to verify the adaptiveness of self-tuning fuzzy PID controller. The results showed that models related to EMB system are effective and have met the national standard and code for braking system. Automobile braking performance was improved because fuzzy control and PID control was improved by self-tuning fuzzy PID control. And it is concluded that self-tuning fuzzy PID controller is good at identify pavement types which meets the control requirement and expectations.

**Key words.** Self-tuning fuzzy PID control, EMB, ABS, slip rate.

## 1. Introduction

The braking system is one of the most important parts of an automobile. Its main function is: first of all, to ensure stable parking at high speed, and when accidents occur, it is very important to its performance requirements. Secondly, its function is to ensure that the vehicles, in the event of poor road conditions and when necessary to reduce the speed, can slow down the parking; or in downhill road, when necessary to travel at a constant speed, to ensure that the brake has high reliability and does not fail due to friction sheet overheating [1–3]. The last one is the parking function. Parking is, with a constant braking force, to ensure that the vehicles stay in place after the end of driving, and do not move because of terrain and other reasons. These are the functions that the braking system should have.

---

<sup>1</sup>Changsha Vocational & Technical College, 410217, China

<sup>2</sup>Corresponding author

The braking system is mainly composed of a drive mechanism and a brake actuating mechanism. The braking transmission mechanism can be divided into mechanical transmission and hydraulic transmission in accordance with the transmission mode. The simple mechanical drive is seldom used at present, and hydraulic transmission is the most widely used transmission method [4–6]. However, with the development of intelligent systems, the number of electronic control systems is gradually increasing, which makes the hydraulic system with pipeline transmission become more complex and difficult to maintain. Therefore, it has become an important problem for people to study a kind of system with simple structure and reliable function.

As one of the most important performances for automobile, automobile braking performance, such as braking distant, side slipping and braking direction, is responsible for traffic accident. Under certain braking initial speed, braking distant is related to two factors, brake response time and wheels' utilization rate of pavement deceleration. Overall circuit transfer was used by EMB, thus response time for braking controller braking distant is shortened [7]. Braking distant is further shortened by adopting effective ABS control because pavement adhesion is used to the highest degree [8]. In terms of intelligence and safety, system structure is simplified by adopting CAN and EMB, because of which fault detection and maintenance is more convenient. Thus, fault detection and maintenance for braking system is improved [9]. In conclusion, security of automobile control can be ensured by adopting improved EMB system.

## 2. State of the art

Originally, EMB system is used for plane [10], and this system is in the improvement stage for automobile field. Significant achievements for study and application of fuzzy theory are mainly made in America or countries and regions in European. For research results on fuzzy ABS, slip rate predictor is introduced by Georg E. Mauer [11] and his partners based on ABS controller. Using this slip rate predictor, good control is achieved in simulation for single wheel model, and robustness is improved. Model-based control methods such as PID controller and fuzzy controller, was combined by R. Sun [12] and this controller combination verified that adaptability to different pavement is improved comparing with PID controller. Robust controller is built by Chin-Min Lin and his partners based on fuzzy controller. Certain control effects are made when fuzzy controller is adopted for braking control and robust controller is adopted for adjusting control errors of fuzzy control [13]. Regarding research status at home and abroad, there are disadvantages in fuzzy control though great progresses are achieved. Thus, further study and discussion is needed theoretically and pragmatically.

Compared with foreign countries, the application of fuzzy theory in our country started relatively late, but it has developed rapidly. In recent years, many universities and automobile research institutions in China have done a great deal of theoretical and experimental researches on ABS fuzzy control technology, which laid the foundation for the development of ABS fuzzy control in our country. The famous scholar Guo Konghui, based on the simplified model, designed the fuzzy controller

and the adaptive fuzzy controller. As a result, the vehicle can achieve better control in variable conditions, which overcomes the shortcomings of single control and so on. Li Jun, Yu Fan, Zhang Jianwu and other scholars proposed control strategy of road recognition in the process of steering braking. The strategy, according to the road adhesion conditions and the motion state, the vehicle pavement condition was estimated, wheel optimal slip ratio was real-time calculated, and the corresponding control strategies were made, so the braking and vehicle lateral stability was greatly improved. Mo Yousheng, Zhu Rong and Li Sien put forward adaptive fuzzy neural network control system with combination of fuzzy control and neural network control. In addition, it was compared with the fuzzy control, and the simulation was made to verify the validity of the control. Chen Jiong, Wang Huiyi and Song Jian designed a fuzzy controller based on slip ratio and speed reduction. The simulation was carried out on a vehicle model of freedom, and it was proved that the controller is more adaptive than the logic threshold method.

From the present research situation at home and abroad, although fuzzy control has been greatly developed, there is still a lack of fuzzy control. For instance, the creation and analysis methods of fuzzy control system were still in the primary stage, and the stability theory was not mature. In addition, the modeling of fuzzy systems, establishment of fuzzy rules and fuzzy inference methods were also not well resolved. As a result, both theory and application need to be further studied and discussed.

### 3. Methodology

#### 3.1. Dynamics modeling for EMB

Automobile EMB system is mainly made up of electronic pedal, electronic control unit (ECU), wheel braking system, and power [14, 15]. A set of braking system and wheel speed sensor (WSP) is installed on each wheel of the automobile. Each braking system contains a control unit (CU) to control performance of electric machine and the needed control signals is provided by ECU [16, 17]. For general structure, see Fig. 1.

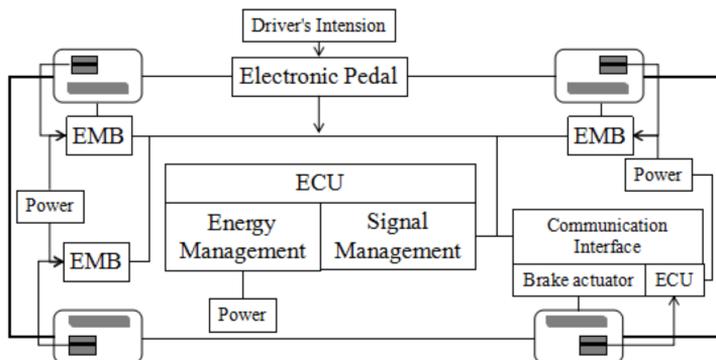


Fig. 1. General structure of EMB system

External force is needed to decelerate or stop automobile in a short time [18]. Among external forces, braking force is one of the most important one to decelerate automobile. Thus, braking performance is studied and this study is mainly focused on effectiveness of braking force to automobile movement.

When wheel braking on hard pavement, rolling friction couple, inertia force and inertia couple occurred during deceleration were ignored. Figure 2 shows stress state of wheel.

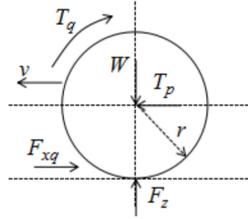


Fig. 2. Stress state of wheel during braking

In the above figure,  $T_q$  denotes braking torque (N·m) of brake,  $T_p$  denotes thrust of driving direction of wheel,  $F_q$  denotes braking force,  $F_{xq}$  denotes braking force of the pavement,  $W$  denotes loads of wheel  $F_z$  denotes normal reaction of pavement to wheel,  $r$  denotes action radius of wheel. Thus, the below equation is obtained.

$$F_{xq} = F_q = \frac{T_q}{r}. \quad (1)$$

During automobile braking, friction plate gradually touches brake disc with increasing braking force. At this state,  $T_q$  is not big enough to lock wheel. This is called state of friction and rolling. Under this state, pavement brake torque equals to  $T_q$ , and pavement brake torque is in direct proportion to  $T_q$ . When  $T_q$  keeps increasing, there is only friction state for wheel. At this state, pavement brake force is no longer in direct proportion to  $T_q$ , and its limit value is the adhesive force  $F_\phi$ , see the below equality

$$F_{xq} \leq F_\phi = F_{z\phi}. \quad (2)$$

During braking, state of wheel is changed from rolling to lock and slipping. In this state, wheel is rolling and slipping, and slipping is decided by slip rate. The below expression

$$S = \frac{v - v_r}{v} \times 100\% = \frac{v - r\omega}{v} \times 100\% \quad (3)$$

shows slip rate.

In the above expression,  $v$  denotes the automobile speed,  $v_r$  denotes the wheel speed and  $\omega$  denotes the angular speed of wheel. In the rolling state,  $v = v_r$ , thus  $S = 0$ . In state of rolling and slipping,  $0 < S < 100\%$ . In locked friction state,  $v_r = 0$  and  $S = 100\%$ . Steering capability of automobile is lost, which is very dangerous working state.

Modeling on automobile brake system based on above dynamics analysis. Mathematical model about automobile brake system is mainly made up of vehicle dynamics

model, tire model, and brake system model.

*3.1.1. Dynamics model of single wheel vehicle.* In order to verify controller performance and highlight its control law, a single wheel automobile is studied. Vehicle dynamic functions of two-degree-of-freedom is built based on wheel driving direction and direction of rotating around principle axis.

Vehicle movement function is

$$M\dot{v} = -F_{xq}. \quad (4)$$

Wheel movement function is

$$I\dot{\omega} = rF_{xq} - T_q \quad (5)$$

and the longitudinal friction of wheel is

$$F_{xq} = \phi F_z. \quad (6)$$

In the above expressions,  $M$  denotes 1/4 of the vehicle weight,  $F_{xq}$  denotes the cohesion force of tire to pavement and  $I$  denotes the rotational inertia of the wheel. Finally  $F_z$  denotes the normal reaction of pavement to wheel.

*3.1.2. Tire model.* Tire is the only part of vehicle that contact with pavement, and its cohesion to pavement, and its driving force, braking force, and trafficability to vehicle is of significant influence. Tire model can be used to approximate rapid analyze vehicle controllability and stability theoretically. Tire model reflects function relationships between pavement cohesion and other parameters. The below bilinear model is used to study tire model.

The bilinear model is a simplified tire model. Relationship of slip rate and adhesion coefficient is nonlinear. For the convenient of function solving, bilinear model is piecewise linearized, see Fig. 3. The function for bilinear tire model is deduced according to Fig. 3. **Figure 3 is missing.**

$$\begin{cases} \varphi = \frac{\varphi_p}{S_o} S & , \quad S < S_o \\ \varphi = \frac{\varphi_p - \varphi_s}{1 - S_o} S - \frac{\varphi_p - \varphi_s}{1 - S_o} S & , \quad S > S_o \end{cases} \quad (7)$$

In the above function,  $S$  denotes the wheel slip rate,  $S_o$  denotes the optimal slip rate,  $\varphi_p$  denotes the maximum adhesion coefficient;  $\varphi_s$  denotes slip adhesion coefficient.

The simplified bilinear tire model is adopted, and Table 1 shows the parameters of experimental made on typical pavement.

*3.1.3. Model for EMB braking system.* A brushless direct current motor is adopted as motive power of EMB system. Mainly, electronic machine in locked-rotor state is studied, which is focused on wheel braking state. The below function shows relationship between locked-rotor current and control signal of electronic machine

Table 1. Parameters of experiments on typical pavement

Types of pavement	$S_o$	$\phi_p$	$\phi_s$	Bilinear model
Concrete pavement	0.2	0.89	0.76	$\begin{cases} \phi = 4.5S & , S < 0.2 \\ \phi = 0.92 - 0.19S & , S > 0.2 \end{cases}$
Dry bitumen pavement	0.16	0.82	0.76	$\begin{cases} \phi = 4.7S & , S < 0.16 \\ \phi = 0.82 - 0.29S & , S > 0.16 \end{cases}$
Wet bitumen pavement	0.13	0.78	0.52	$\begin{cases} \phi = 6S & , S < 0.13 \\ \phi = 0.83 - 0.32S & , S > 0.13 \end{cases}$
Pavement covered by snow	0.06	0.22	0.15	$\begin{cases} \phi = 3.3S & , S < 0.06 \\ \phi = 0.23 - 0.06S & , S > 0.06 \end{cases}$

$$I_c = k_c \cdot \alpha, \quad (8)$$

where  $I_c$  denotes locked-rotor current,  $k_c$  denotes conversion relations between control signals and locked-rotor current,  $\alpha$  denotes inputted control signal. Below function shows relations between  $I_c$  and output torque of electronic machine.

$$T_m = 9.55k_G \cdot I_c, \quad k_G = \frac{E}{n_o}, \quad E = U_o - I_o r_o. \quad (9)$$

For the above function,  $T_m$  denotes the output torque of electronic machine,  $I_o$  denotes the non-load current,  $I_c$  denotes the current of locked-rotor,  $U_o$  denotes the non-load voltage of power,  $E$  denotes the counter electromotive force of the armature winding,  $r_o$  denotes the average resistance of armature winding,  $k_G$  denotes counter electromotive force coefficient, and  $n_o$  denotes the idle speed of electronic machine.

Planetary reducer is adopted by reducing mechanism model which is made up of sun wheel and planet carrier.  $T_m$  denotes input moment of sun wheel, and torque  $T_x$  denotes output planetary reducer. Below function shows the relations.

$$T_x = T_m \cdot i \cdot \eta_x, \quad (10)$$

where,  $i$  denotes transmission ratio of speed reducer,  $\eta_x$  and denotes transmission efficiency of planetary mechanism. Motion transfer device is made up of ball screw-nut pair. Finally,  $T_x$  denotes the input and  $P$  denotes outputted thrust of ball screw.

$$P = T_x \cdot \eta_g \cdot \frac{2\pi}{L_h}. \quad (11)$$

In the above formulas,  $L_h$  denotes lead of screw thread and  $\eta_g$  denotes transmission efficiency of ball screw. Ball screw and brake caliper were connected to each other and lining pad of brake caliper and brake disc are connected to each other through thrust  $P$ . Braking torque is produced due to friction of lining pad and brake disc. Below function shows relations between brake pressure and brake

moment when friction surface of lining pad and brake disc connects well.

$$T_q = 2P \cdot k_p \cdot R. \quad (12)$$

In above expression,  $k_p$  denotes brake friction coefficient,  $R$  denotes action radius, and  $P$  denotes the lead screw thrust. Table 2 shows the parameters of the used electric machine, and Table 3 shows the parameters of EMB brake actuator.

Table 2. Parameters of electric machine of permanent magnet DC motor

Parameter names	Non-load voltage (V)	Non-load current (A)	Armature resistance ( $\Omega$ )	Non-load speed (r/min)	Locked-rotor current (A)	Maximum locked-rotor current (A)
Symbols	$U_o$	$I_o$	$r_o$	$n_o$	$I_c$	$I_{c \max}$
Values	27	0.30	3.68	491	2.2	7

Table 3. Parameter of EMB brake actuator

Symbols of parameters	$i$	$\eta_x$	$L_h$	$\eta_g$	$k_p$	$R$
Values	20	0.95	0.016	0.95	2	0.2

### 3.2. Modeling on brake system of electric mechanical machine

Matlab/Simulink is used to model and simulate subsystems. Without control of ABS, effectiveness of the EMB brake system is tested.

3.2.1. *Subsystem of single wheeled vehicle.* After solving functions (4), (5) and (6), below expressions for  $v$ ,  $v_r$  and  $s$  can be obtained:

$$v = - \int \frac{F_{xq}}{M} dt, \quad v_r = \frac{r}{l} \int (rF_{xq} - T_q) dt, \quad s = - \int \left( \int \frac{F_{xq}}{M} dt \right) dt. \quad (13)$$

Simulation model of a single wheel can be built based on (13). With the input of  $T_q$  and  $F_{xq}$ , the curve of  $v$  and  $v_r$  changing with time can be obtained.

3.2.2. *Subsystem of slip rate.* Based on (3), subsystem of slip rate can be built. For this function, inputting  $v$ ,  $v_r$  and adopting  $F_{cn}$  function, the slip rate of vehicle is output.

3.2.3. *Subsystem of tire model.* Simulation model is built based on bilinear calculation function (7), which inputs slip rate and outputs longitudinal adhesion coefficient.

3.2.4. *Subsystem of brake model.* Brake actuator is made up of brake transmission mechanism and brake, whose simulation module is built based on relations

(8–12), inputting control signal  $v$ , and outputting braking torque  $T_q$ .

### 3.3. Design of self-tuning fuzzy PID controller

Self-tuning fuzzy PID controller is the combination of fuzzy control and traditional PID control, which tunes PID control parameters online using fuzzy theory related knowledge. Intelligence of fuzzy control and sensitivity of PID control is integrated by this tuning. Figure 3 shows the system structure.

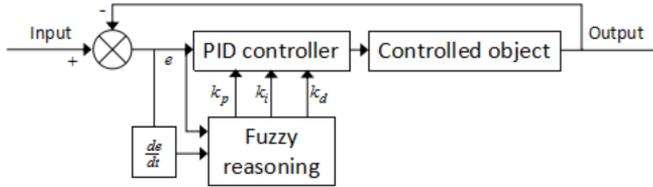


Fig. 3. Structure of self-tuning fuzzy system for automobile ABS

Self-tuning fuzzy PID controller is adjusting values of  $k_p$ ,  $k_i$ ,  $k_d$  in real time according to fuzzy control theory, thus objects is controlled. This controller inputs error  $e$  and its variation  $\Delta e$ , which is transmitted to  $E$  and  $\Delta E$  after a fuzzy process. Symbols  $k_p^*$ ,  $k_i^*$ ,  $k_d^*$  are fuzzy set of output quantities  $k_p$ ,  $k_i$ ,  $k_d$ . For input quantities  $e$  and  $\Delta e$ , the universe of fuzzy set is  $\{-3 - 2 - 1, 0, 1, 2, 3\}$ , whose fuzzy language variables are  $\{NB, NM, NS, ZO, PS, PM, PB\}$ . For output quantities  $k_p^*$ ,  $k_i^*$ ,  $k_d^*$ , the universe of fuzzy set is  $\{-3 - 2 - 1, 0, 1, 2, 3\}$ , whose fuzzy language variables are  $\{NB, NM, NS, ZO, PS, PM, PB\}$ , which represent big negative, medium negative, small negative, zero, small positive, medium positive, and big positive, respectively.

The fuzzy control Table 4 is built based on characteristics without PID control and different  $e$  and  $\Delta e$  input by the system.

## 4. Results analysis and discussion

### 5. Test on effectiveness of system model of electronic mechanical brake system

Concrete pavement with relatively high adhesive force is used to test brake performance of EMB. Relatively strong brake force can be produced on concrete pavement. Table 5 shows the vehicle parameters.

It can be seen from Fig.4 that in the initial braking phase, pavement braking torque is in direct proportion to  $T_q$  while pavement braking torque is in inverse proportion to  $T_q$  after reaching its value. Without ABS control, adhesive coefficient is decreasing with the rapid increasing slip rate of wheel, thus ground adhering moment is decreased. When  $T_q$  keeps increasing, ground adhering brake moment will remain the same torque value with that of in slipping state.

Table 4. Rule list of fuzzy control for  $k_p^*$ ,  $k_i^*$  and  $k_d^*$ 

$k_p^*$							
	NB	NM	NS	ZO	PS	PM	PB
NB	PB	PB	PM	PM	PS	ZO	ZO
NM	PB	PB	PM	PS	PS	ZO	NS
NS	PM	PM	PM	PS	ZO	NS	NS
ZO	PM	PM	PS	ZO	NS	NM	NM
PM	NS	ZO	NS	NM	NM	NM	NB
PB	PB	ZO	NM	NM	NM	NM	NB
$k_i^*$							
	NB	NM	NS	ZO	PS	PM	PB
NB	NB	NB	NM	NM	NS	ZO	ZO
NM	NB	NB	NM	NS	NS	ZO	ZO
NS	NB	NM	NS	NS	ZO	PS	PS
ZO	NM	NM	NS	ZO	PS	PM	PM
PS	NM	NS	ZO	PS	PS	PM	PB
PM	ZO	ZO	PS	PS	PM	PB	PB
PB	ZO	ZO	PS	PM	PM	PB	PB
$k_d^*$							
	NB	NM	NS	ZO	PS	PM	PB
NB	PS	NS	NB	NB	NB	NM	PS
NM	PS	NS	NB	NM	NM	NS	ZO
NS	ZO	NS	NM	NM	NS	NS	ZO
ZO	ZO	NS	NS	NS	NS	NS	ZO
PS	ZO						
PM	PB	NS	PS	PS	PS	PS	PB
PB	PB	PM	PM	PM	PS	PS	PB

Table 5. Parameters of single wheel vehicle

Name	1/4 Vehicle weight (kg)	Wheel radius (m)	Wheel rotary inertia (kg·m <sup>2</sup> )	Gravitational acceleration (m/s <sup>2</sup> )	Initial speed of braking (m/s)
Symbols	$M$	$r$	$I$	$g$	$v$
Values	1880	0.53	20	9.8	24

It is known from above simulation experiment that the wheel was locked in the 1.28 s after emergent brake under EMB brake system. Braking distance is 16.24 m, braking time is 2.34 s and the maximum vehicle deceleration is 8.85 m/s<sup>2</sup>. Besides, the average deceleration is 7.16 m/s<sup>2</sup>, and average braking coordination time is 0.14 s. This meets the national standard that requires coordination time should less than 0.35 s under such working state. Thus, EMB brake meets national requirement for braking performance well.

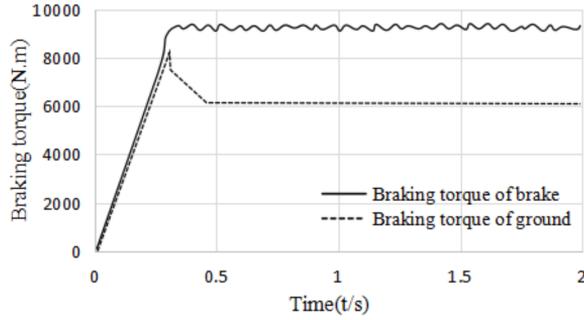


Fig. 4. Changes of braking torque of ground and brake without ABS control

**5.1. Automobile ABS simulation based on self-tuning fuzzy PID control**

Figure 5 shows automobile ABS simulation model built based on self-tuning fuzzy PID control.

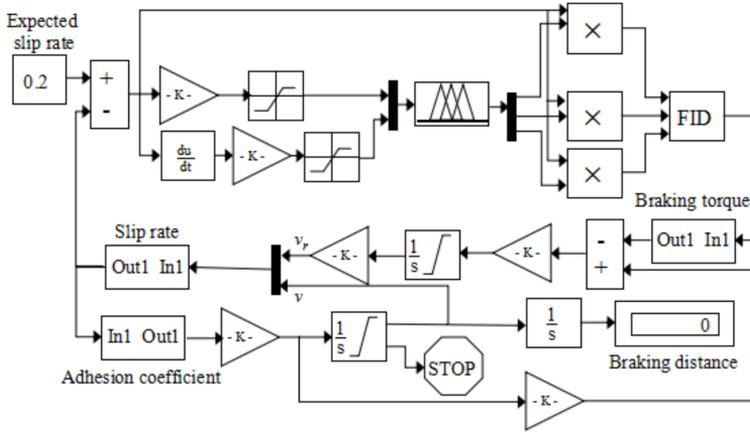


Fig. 5. Automobile ABS simulation model built based on self-tuning fuzzy PID control

In order to test the effectiveness of self-tuning fuzzy PID controller, comparison was made between braking systems with self-tuning fuzzy PID controller and system without ABS controller. Concrete pavement is chosen to be studied, with expected slip rate of 0.2, slip adhesion coefficient of 0.75, and initial braking speed of 24 m/s. Table 1 shows simulation parameters and Figs. 6 and 7 show the simulation results.

It can be seen from simulation results in Figs. 6 and 7 that the vehicle braking distance is 37.26 m and braking time is 3 s when the initial speed is 24 m/s. During braking, shown as Fig. 6, upper part, the slip rate remains the same as that of in 0.2 s and reaches its maximum in 0.3 s when without ABS control. It is shown on Fig. 6, bottom part, that without ABS control, vehicle speed is 23 m/s when wheel is locked. While under self-tuning fuzzy PID control, the wheel is locked when vehicle

speed is 0 m/s, which improves the braking stability. For Fig. 7, upper part, under ABS control, vehicle speed remains unchanged when it hits the highest deceleration  $8.85 \text{ m/s}^2$ . Comparing vehicle speed without ABS control, this takes full advantages of ground adhesion. Besides, under fuzzy control, the braking distance is shortened by 7.41 m, and braking time is shortened by 0.62 s. In the whole braking process, it is shown in Fig. 7, bottom part, that the output is relatively stable. Thus, self-tuning fuzzy PID controller meets the ABS control goal and requirements of safety compared with the state without ABS control.

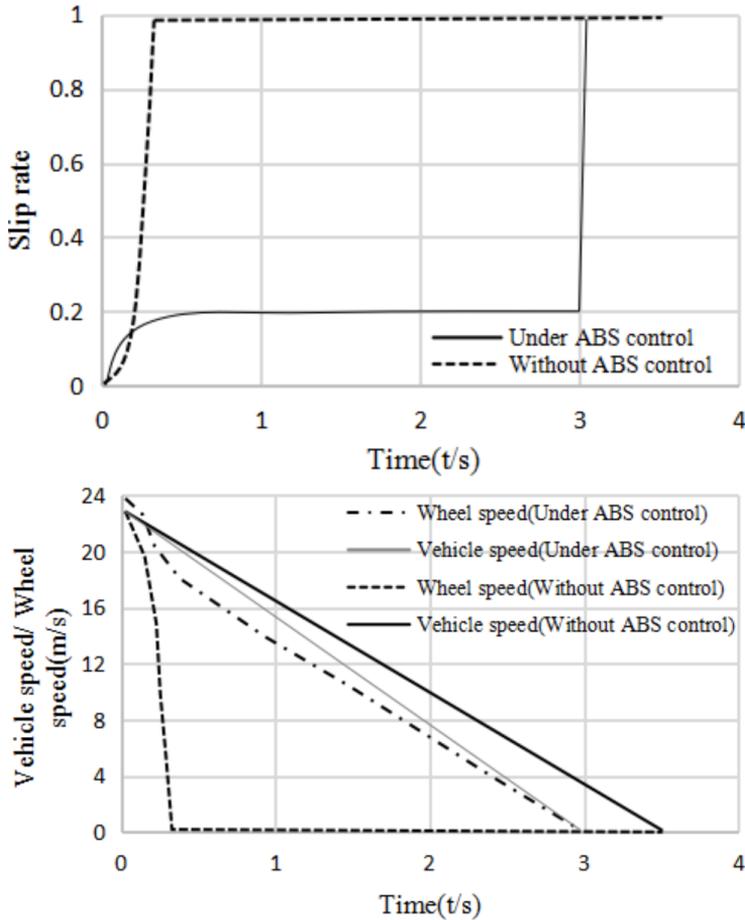


Fig. 6. Slip rate (up) and vehicle speed (bottom) for system with self-tuning fuzzy PID control and system without ABS control

The simulation and analysis was made based on bitumen pavement and pavement covered by snow. It is known that, in the 0.5 s, slip rate hits its optimal record of 0.06 and remains good stability when braking on the pavement covered by snow. In the 2 s after braking, when changing from pavement covered by snow to bitumen pavement, slip rate is unstable, which changed back to stable states 1 s later and

remains stable until vehicle stopped. For this process, braking distance is 79.37 m, and braking time is 5.11 s, which meets the requirements of changing from pavement covered by snow to bitumen pavement.

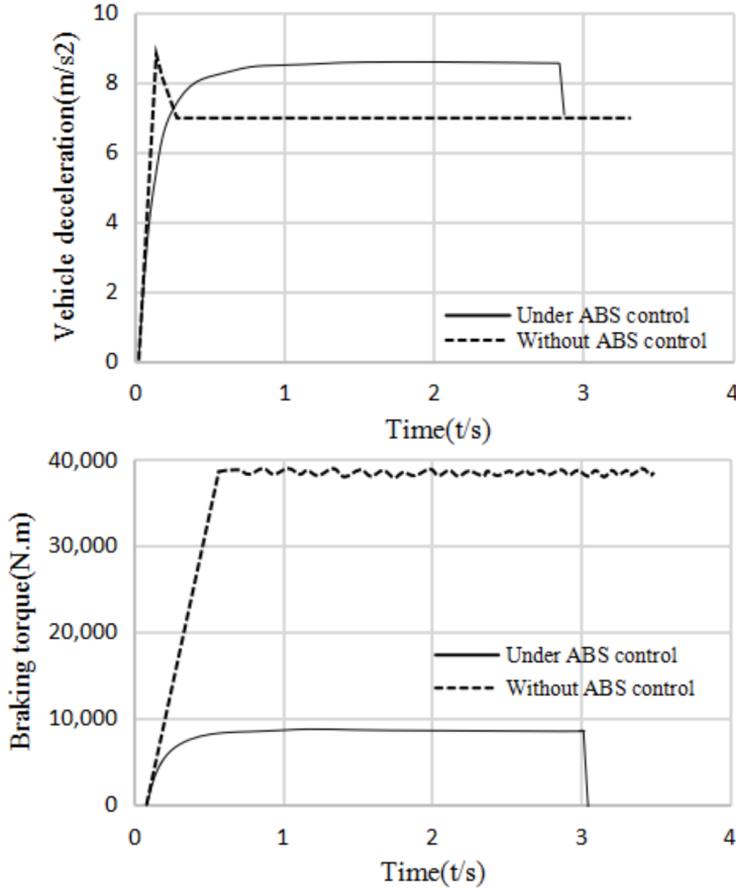


Fig. 7. Vehicle deceleration (up) and braking torque (bottom) for system with self-tuning fuzzy PID control and system without ABS control

## 6. Conclusion

This study is based on structure of braking system and braking methods, and below are the main research contents: EMB braking system model, tire model and single wheel vehicle model are built based on stress analysis of braking vehicle. Effectiveness of EMB braking system is verified that it met national standards for braking system. Improved self-tuning fuzzy PID controller is designed. And simulation results show that self-tuning fuzzy PID controller can adjust automatically according to external changes, which solves parameter setting problem of PID con-

troller because PID controller is nonlinearity and its parameters are changing with time. Adaptability of self-tuning fuzzy PID controller is verified by sudden changed pavement types. This simulation result shows that self-tuning fuzzy PID controller can control slip rate around target value, and identify optimal slip rate for different pavements, thus output of braking force is accurate controlled.

## References

- [1] W. SHABIR, S. A. EVANGELOU: *Real-time control strategy to maximize hybrid electric vehicle powertrain efficiency*. *Applied Energy* 135 (2014), 512–522.
- [2] C. MONTERO, A. OLIVA, D. MARCOS, E. GONZÁLEZ, C. BORDONS, M. A. RIDAO, E. F. CAMACHO, E. LÓPEZ: *Fuel cell and power control for a hybrid vehicle. Experimental results*. Proc. Annual Conference on IEEE Industrial Electronics Society (IECON), 25–28 October 2012, Montreal, Québec, Canada, IEEE Conference Publications (2012), 4121–4126.
- [3] F. S. AHMED, S. LAGHROUCHE, M. EL BAGDOURI: *Analysis, modeling, identification and control of pancake DC torque motors: Application to automobile air path actuators*. *Mechatronics* 22 (2012), 195–212.
- [4] X. ZHU, H. ZHANG, D. CAO, Z. FANG: *Robust control of integrated motor-transmission powertrain system over controller area network for automotive applications*. *Mechanical Systems and Signal Processing* 58–59 (2015), 15–28.
- [5] J. B. ZHAO, B. ZHOU, S. Y. BEI: *Bang-Bang-PID control of automotive EPS system under damping condition*. *Electric Machines and Control* 15 (2011), No. 11, 95–100.
- [6] J. B. ZHAO, B. ZHOU, X. L. LI, S. Y. BEI: *Switched control strategy and performance test of electric vehicle electric power steering system under different conditions*. *Electric Machines and Control* 17 (2013), No. 11, 105–109.
- [7] S. DI CAIRANO, J. DOERING, I. V. KOLMANOVSKY, D. HROVAT: *Model predictive control of engine speed during vehicle deceleration*. *IEEE Transactions on Control Systems Technology* 22 (2014), No. 6, 2205–2217.
- [8] X. Q. ZHANG, B. YANG, C. YANG, G. N. XU: *Research on ABS of multi-axle truck based on ADAMS/car and Matlab/Simulink*. *Procedia Engineering* 37 (2012), 120–124.
- [9] H. ALIPOUR, M. B. B. SHARIFIAN, M. SABAHI: *A modified integral sliding mode control to lateral stabilisation of 4-wheel independent drive electric vehicles*. *International Journal of Vehicle Mechanics and Mobility* 52 (2014), No. 12, 1584–1606.
- [10] V. R. APAROW, F. AHMAD, K. HUDHA, H. JAMALUDDIN: *Modelling and PID control of antilock braking system with wheel slip reduction to improve braking performance*. *International Journal of Vehicle Safety* 6, (2013), No. 3, 265–296.
- [11] M. DEMIRCI, M. GOKASAN: *Adaptive optimal control allocation using Lagrangian neural networks for stability control of a 4WS-4WD electric vehicle*. *Transactions of the Institute of Measurement and Control* 35 (2013), No. 8, 1139–1151.
- [12] C. LV, J. ZHANG, Y. LI, Y. YUAN: *Mode-switching-based active control of power train system with nonlinear backlash and flexibility for electric vehicle during regenerative deceleration*. *Proceedings of the Institution of Mechanical Engineers Part D: Journal of Automobile Engineering* 229 (2015), No. 11, 1429–1442.
- [13] B. LONG, S. T. LIM, H. R. JI, K. T. CHONG: *Energy-regenerative braking control of electric vehicles using three-phase brushless direct-current motors*. *Energies* 7 (2014), No. 1, 99–114.
- [14] S. YU, C. MAIER, H. CHEN, F. ALLGÖWER: *Tube MPC scheme based on robust control invariant set with application to Lipschitz nonlinear systems*. *Systems & Control Letters* 62 (2013), No. 2, 194–200.
- [15] Y. LIU, L. Q. JIN, X. L. LIANG, Z. A. ZHENG: *Research on BP based fuzzy-PID controller for anti-lock braking system*. *Applied Mechanics and Materials* 365–366 (2013), 401–406.

- [16] V. R. APAROW, K. HUDHA, F. AHMAD, H. JAMALUDDIN: *Development of antilock braking system using electronic wedge brake model*. Journal of Mechanical Engineering and Technology 6 (2014), No. 1, 37–63.
- [17] G. YIN, X. J. JIN: *Cooperative control of regenerative braking and antilock braking for a hybrid electric vehicle*. Mathematical Problems in Engineering (2013), No. 4, paper 890427.
- [18] K. BAYAR, G. RIZZONI, J. WANG: *Development of a vehicle stability control strategy for a hybrid electric vehicle equipped with axle motors*. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering 226 (2012), No. 6, 795–814.

Received May 7, 2017

# Application of structural bending differential equation algorithm in bracing composite structure

LI KE<sup>2,3</sup>, YIN KE<sup>2</sup>

**Abstract.** The purpose of this paper is to study the effect of bending differential equation algorithm on composite bracing structures. By analyzing the existing landslide composite supporting structure, a new type of diagonal bracing composite supporting structure is purposed. The diagonal bracing composite supporting structure changes the original cantilever bending structure into diagonal loads transfer structure. It reduces the load on the structure which under the huge thrust of landslide, and it has good mechanical properties. Based on the Fredholm–Volterra line load integral and displacement control equation, diagonal bracing composite supporting structure control condition is put forward, and the continuous structure algorithm based on the deflection equation is deduced. The analytical solution of diagonal bracing composite supporting structure the displacement and internal force are solved. Continuous structure algorithm based on deflection equation of brace composite supporting structure engineering design is optimized. The experimental results verify the flexural differential equation algorithm of continuous structure by the three-dimensional finite element analysis. Based on the above finding, it is concluded that the design of diagonal bracing composite supporting structure can be optimized by the flexural differential equation algorithm.

**Key words.** Flexural differential equation, continuous structure algorithm, diagonal bracing composite supporting structure, mechanics characteristic.

## 1. Introduction

Existing large-scale landslide treatment of timbering form is numerous at present, structure forms are also various, the main application structure forms are cantilever anti-slide pile, double row piles, etc. Cantilever anti-slide pile was applied at the earliest, related research is more thorough. But Cantilever anti-slide pile application depth is limited by its mechanical characteristics, and control deformation ability of

---

<sup>1</sup>This work was partially supported by the Research Fund for the Doctoral Program of Higher Education of China (RFDP), (Grant No.20110191110027).

<sup>2</sup>School of Civil Engineering, Chongqing University 400044, China

<sup>3</sup>Chongqing GaoXin Geotechnical Investigation Institute, Chongqing 400047, China

Cantilever anti-slide pile is poor. These characteristics of the cantilever anti-slide pile cannot satisfy the requirement of landslide with large thrust. Double row pile as supporting form more research in recent years, double row pile research has achieved some results in domestic and overseas [1]. Based on the theory of plastic deformation and plastic flow theory, the pile lateral soil pressure formula is derived. The design idea and design method of the single pile and Multi-pile are purposed [2]. On the basis of predecessors' research, some researchers test the pile groups model under the action of the displacement of soil in 1997, and test results with boundary element program computing results have good consistency [3]. The model of soil was used to study the relative displacement of piles and soil, pile spacing and pile arrangement of its influence on soil arching between piles [4]. The calculation and stress distribution of double-row anti-slide pile in three kinds of typical landslide is calculated [5]. The comparative analysis of prestressed concrete anchor-stabilizing double-row piles and rigid frame anti-slide pile is purposed [6]. Double row pile theory research has made some achievements, but still failed to in giant, large rock landslide to obtain the very good application. In southwest China, many types of geological disasters and landslide, multistage complex landslide along with the urban expansion and construction appear constantly. From the characteristics of the giant, large damage rock landslides, this paper present a new diagonal bracing composite structure base on the research of existing timbering structure. diagonal bracing composite structure mainly aimed at a large thrust (generally more than 2000 kN/m) rock landslides. Diagonal bracing composite structure calculation method is: Improved the deflection control equation under the condition of different boundary control, makes it can get the optimal transfer ratio of diagonal bracing composite structure (the ratio of landslide thrust which transmit by diagonal bracing composite structure into rock mass and the total thrust). The method also can solve the internal force and displacement of diagonal bracing composite structure.

Diagonal bracing composite structure is mainly composed of two parts, the vertical bearing structure and diagonal bearing structure. Vertical bearing structure differs from the anti-slide pile and double-row piles structure and so on, its main function is to transfer thrust from the rock free face to diagonal bearing structure; Diagonal bearing structure is transfer the load which from the vertical bearing structure through the column brace to support structure and finally pass it to the lower rock mass [7]. The characteristics of diagonal bracing composite structure model is: Change the normal original cantilevered pile mechanic (the normal original cantilevered pile depends on the shear strength of section and bending strength of pile to resist the load) into a transferring patterns which transfer landslide thrust is primary, pile resistance is accessory [8].

## 2. Materials and methods

Diagonal bracing composite structure can be divided into four sections from rigid joint and fixed point [9]. The internal force and displacement of each section can be obtained by the deflection differential equation under the control condition (bending moment and shear is zero at the node which connect the vertical brace and diagonal

bracing).

Using elastic foundation beam differential equation solve the internal force and displacement of vertical brace anchoring range, and using the displacement as the control parameter to unify the four-section deflection differential equation, end up with solving the simultaneous equations of the whole structure, finally be solving the simultaneous equations of the whole structure [10].

Assumption of continuous structure flexural differential equation algorithm:

Vertical brace of composite structure pile under the rectangular uniform load.

The cross-section width of vertical brace and diagonal bracing in composite structure is the same.

Vertical brace and diagonal bracing intersection node O is assumed a rigid node.

Rock mass of fixed section in vertical brace and diagonal bracing according to the elastic material consideration.

Each section of the block division in the table.

General solution of differential equation for each section can be solved, particular solution of differential equation should also be solving at the same time. Each section should be four initial conditions to completely solve the equation [11]. The displacement, rotation, bending moment and shear force of each section as the initial condition into the equation, solution of equations can be obtained.

Vertical brace cantilever section (OA) flexural differential control equation is

$$EI \frac{d^4 x}{dy^4} = q, \quad (1)$$

where  $q$  means distributed load,  $E$  is the elastic modulus and  $I$  is the moment of inertia. Finally,  $x$  is the displacement.

The displacement, rotation, bending moment and shear displacement equation

$$x(y) = x_0 + \phi_0 y + \frac{M_0 y^2}{2EI} + \frac{Q_0 y^3}{6EI} + \frac{qy^4}{24EI}. \quad (2)$$

Here,  $x_0$  means the initial displacement,  $\phi_0$  means the initial rotation,  $M_0$  means the initial bending moment and  $Q_0$  means initial shear force. Generally,  $\phi$  denotes the rotation function, given by the prescription

$$\phi = \phi_0 + \frac{M_0 y}{2EI} + \frac{Q_0 y^2}{6EI} + \frac{qy^3}{24EI}.$$

Finally,  $\beta$ , which is the vertical brace pile deformation coefficient, is given as

$$\beta = \left( \frac{KB_p}{4EI} \right)^{\frac{1}{4}} \quad (3)$$

### 3. Calculation and analysis for diagonal bracing composite structure

#### 3.1. Diagonal bracing composite structure in the original design and optimization design comparative analysis

Comparative analysis between the flexural differential equation algorithm of continuous structure and design calculation method of actual landslide control project can optimize the design of practical project engineering. A certain slope project of Wulong county in Chongqing City adopted in 3 different form diagonal bracing composite structure. Based on type A structure as an example. In this paper,  $h_1$  means the distance from intersection (Point A) of vertical bracing pile with diagonal bracing pile to the top of vertical bracing pile;  $h_2$  means the distance from intersection (Point A) of vertical bracing pile with diagonal bracing pile to the fixed point of vertical bracing pile;  $h_3$  means the distance of fixed section of vertical bracing pile. Symbol  $\theta$  means the angle of vertical bracing pile from the horizontal plane. The flexural differential equation for optimized structure internal force compared with the original design size and internal force (see Fig. 1).

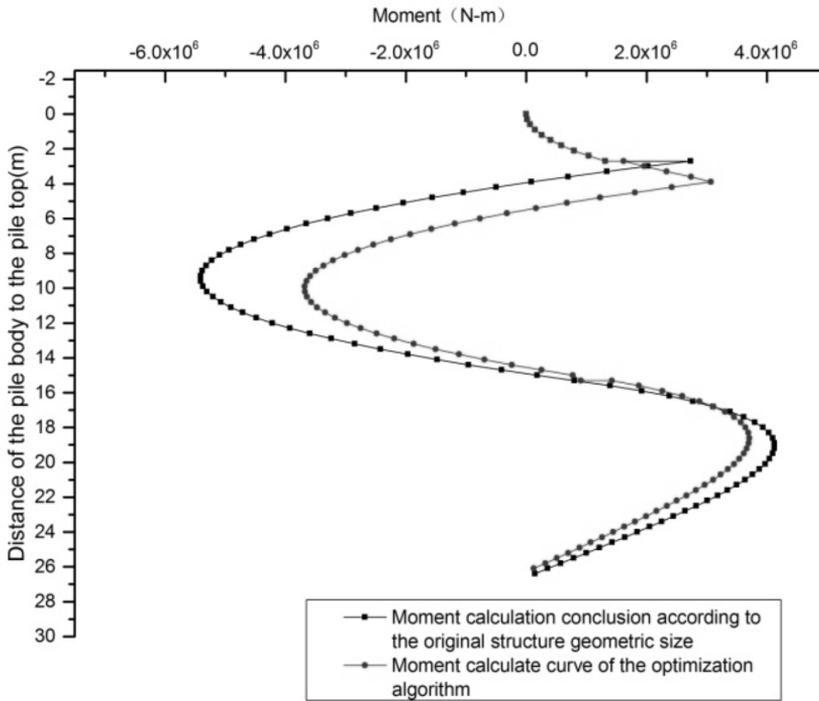


Fig. 1. Bending moment diagram of optimized structure and original design structure

From the figure, Point A to point B section of structure optimization design compared to the original design, the maximum bending moment increases 1727.3 kN.m

(the original design maximum bending moment is 5408.4 kN.m, optimization design maximum bending moment is 3681.1 kN.m), optimization of structure relative to the original structure to reduce bending moment value is about 31.9%. Point B to point C section of structure optimization design compared to the original design, the maximum bending moment increases 1019.8 kN.m (the original design maximum bending moment is 4124.2 kN.m, optimization design maximum bending moment is 3104.4 kN.m), optimization of structure relative to the original structure to reduce bending moment value is about 24.7%.

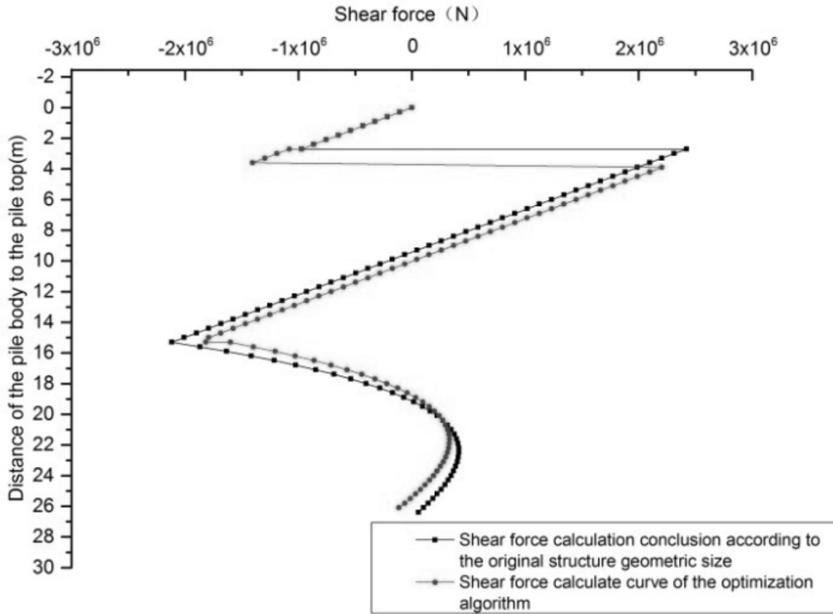


Fig. 2. Shear force diagram of optimized structure and original design structure

Figure 2 shows the shear force diagrams of optimized structure and original design structure. Point A to point B section of structure optimization design compared to the original design, the maximum shear increases 217 kN (the original design maximum shear is 2421 kN, optimization design maximum shear is 2204 kN), optimization of structure relative to the original structure to reduce shear value is about 8.9%. Point B to point C section of structure optimization design compared to the original design, the maximum shear increases 296.5 kN (the original design maximum shear is 2114.7 kN, optimization design maximum shear is 1818.2 kN), optimization of structure relative to the original structure to reduce shear value is about 14.1%.

Finally, Fig. 3 depicts the displacement diagrams of optimized structure and original design structure. In the original structural design of cantilever segments (point A to point O section), displacement variation is larger, the maximal displacement of the two methods difference in 1.2 mm; (increased from the two methods is 2.7 mm to 3.9 mm), in additional optimization design in the intersection (point A) place does not appear A turning point. Two methods of comparison that the displacement of

vertical bracing structure calculation difference of 1.2 mm, optimization of structure relative to the original structure to reduce bending moment value is about 44.4%.

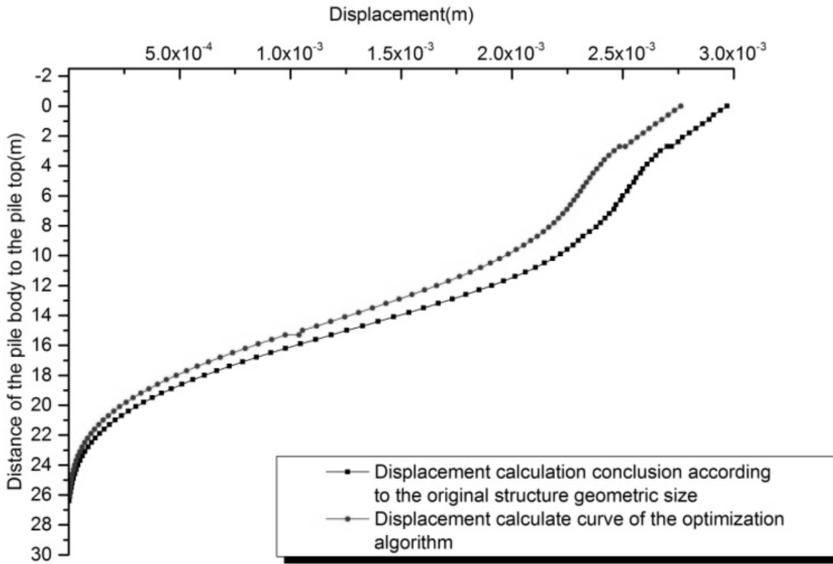


Fig. 3. Displacement diagram of optimized structure and original design structure

### 3.2. Diagonal bracing composite supporting structure and conventional supporting structure comparative analysis

For further verify the effect of diagonal bracing composite supporting structure. The internal force of anti-slide pile, reinforced anchorage pile, the original design and optimization design of diagonal bracing composite supporting structure were compared. Figure 4 shows the bending moment diagrams of anti-slide pile, reinforced anchorage pile and diagonal bracing composite supporting structure.

Through the comparative analysis, bending moment value of the diagonal bracing composite supporting structure both in the original design and optimization design is much smaller than the reinforced anchorage pile and anti-slide pile. The main cause of the different between above supporting structure is mechanism different. Maximum bending moment value from large to small is: anti-slide pile > anchor tensile pile > the original design of diagonal bracing composite supporting structure > optimization design of diagonal bracing composite supporting structure. Anti-slide pile structure because only rely on their own section strength (bending stiffness  $EI$ , tensile stiffness  $EA$  and shear stiffness  $GA$ ) resistance to slide thrust, and therefore in fixed point bending moment value is maximum ( $M_{\text{Max}} = 61.5 \times 10^3 \text{ kN.m}$ ).

Reinforced anchorage pile structure bending moment is smaller than anti-slide pile under the effect of pre-stressed anchor cable tension. Maximum bending moment is  $M_{\text{Max}} = 28.4 \times 10^3 \text{ kN.m}$ . Diagonal bracing composite supporting structure because of the difference of dynamic mechanism, a maximum bending moment are

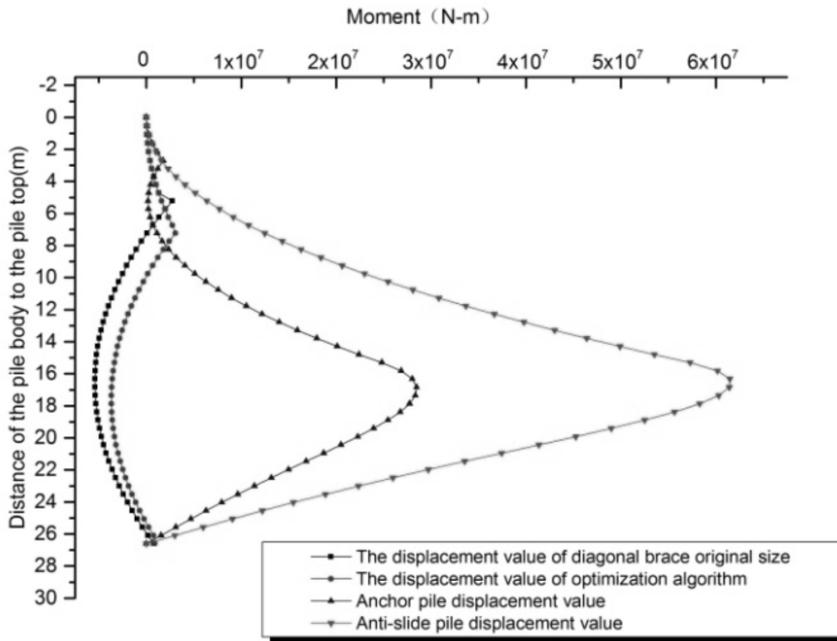


Fig. 4. Bending moment diagram of anti-slide pile, reinforced anchorage pile and diagonal bracing composite supporting structure

much smaller than the front two structures. Maximum bending moment of original design diagonal bracing composite supporting structure is  $M_{\text{Max}} = 5.41 \times 10^3$  kN.m. Maximum bending moment of optimization design diagonal bracing composite supporting structure is  $M_{\text{Max}} = 3.68 \times 10^3$  kN.m.

Fixed section of the diagonal bracing composite supporting structure (point B to point C) due to effect of the diagonal bracing, the shear graph is a linear change graph. This kind of phenomenon is different from the shear distribution graphics of reinforced anchorage pile. The anchor pile and anti-slide pile the maximum shear force was concentrated on the near of pile bottom. Diagonal bracing composite supporting structure shear maximum value appeared at intersection point of the vertical brace and diagonal bracing.

The shear diagrams of anti-slide pile, reinforced anchorage pile and diagonal bracing composite supporting structure are depicted in Fig. 5. Anti-slide pile maximum shear is  $Q_{\text{Max}} = 7.61 \times 10^3$  kN. Pre-stressed anchor cable under the effect of pre-stressed anchor cable tension share shear reduced overall structure, maximum shear of  $Q_{\text{Max}} = 4.41 \times 10^3$  kN. Original brace structure design calculation has maximum shear of  $Q_{\text{Max}} = 2.42 \times 10^3$  kN. Optimize the brace structure design calculation to get maximum shear of  $Q_{\text{Max}} = 2.20 \times 10^3$  kN.

Figure 6 shows the displacement diagram of anti-slide pile, reinforced anchorage pile and diagonal bracing composite supporting structure. A maximum displacement of anti-slide pile is  $X_{\text{Max}} = 75.4$  mm. Pre-stressed anchor cable under the

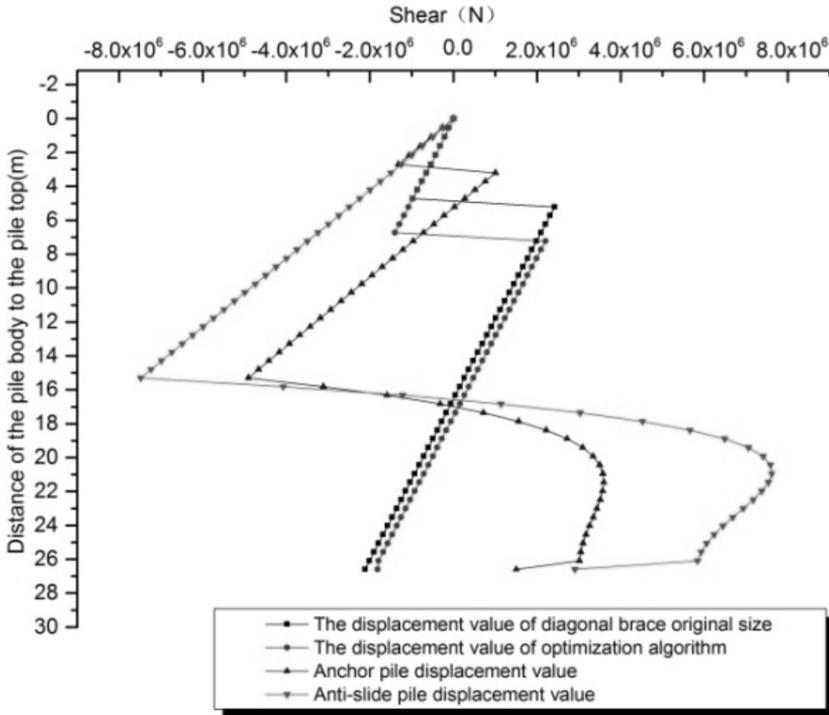


Fig. 5. Shear diagram of anti-slide pile, reinforced anchorage pile and diagonal bracing composite supporting structure

effect of pre-stressed anchor cable tension shares less overall structure displacement, the maximum displacement is  $X_{Max} = 33.9$  mm. The maximal displacement value of the original design diagonal bracing supporting structure is  $X_{Max} = 2.97$  mm. The maximal displacement value of the optimized design diagonal bracing supporting structure is  $X_{Max} = 2.76$  mm. The displacement value of diagonal bracing supporting structure relative to anti-slide pile is about 4%, the displacement value of diagonal bracing supporting structure relative to pre-stressed anchor cable anti-slide pile is about 8.7%.

#### 4. Results

Three-dimensional finite element analysis base on the practical project diagonal bracing composite supporting structure. The vertical support structure section size  $3.0\text{ m} \times 2.0\text{ m}$ , diagonal support structure section size  $2.0\text{ m} \times 2.0\text{ m}$ , angle between diagonal structure and horizontal is  $52^\circ$   $h_1 = 2700$  mm,  $h_2 = 12563$  mm,  $h_3 = 11298$  mm (details of three-dimensional finite element analysis of the entity body are depicted in Fig. 7).

According to the K.J.Bath theory of contact model between different materials,

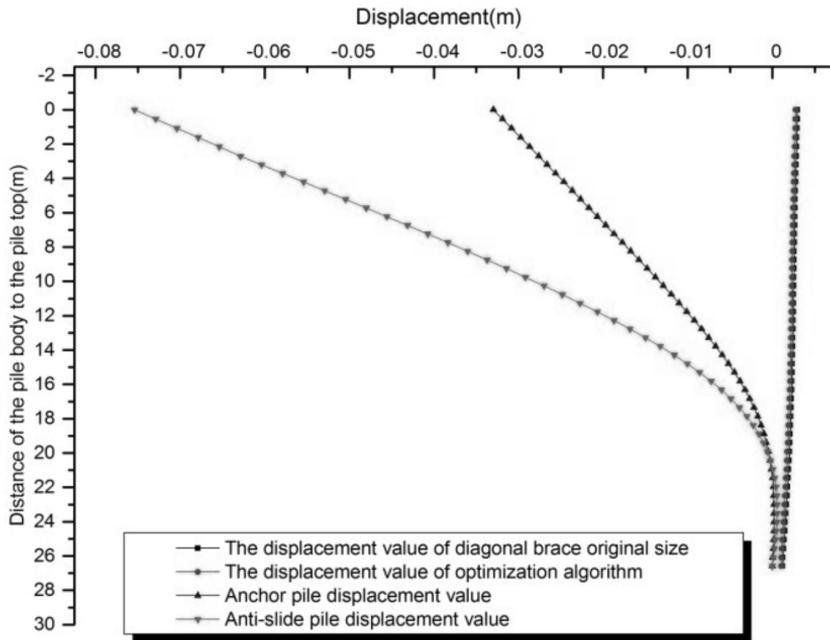


Fig. 6. Displacement diagram of anti-slide pile, reinforced anchorage pile and diagonal bracing composite supporting structure

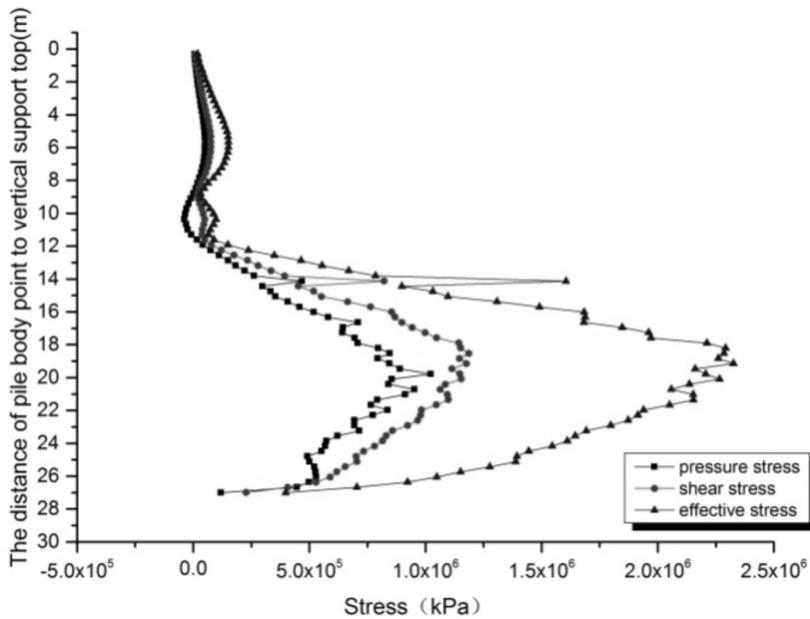


Fig. 7. Vertical brace back lateral principal stress/effective stress analysis diagram

vertical brace is obtained by numerical simulation to solve dorsal soil lateral principal stress analysis diagram (Fig. 8).

Known from the analysis diagram, although the cantilever segments with vertical compressive stress values are increase at intersecting point of vertical brace and diagonal bracing, but compared with the fixed section compressive stress values, cantilever section of the compressive stress value is very small; vertical supporting structure back lateral compressive stress since fixed point to the bottom of the pile increases gradually, near the middle of the fixed section grow to peak. Comparative analysis of the dorsal lateral compressive stress graphics and vertical brace back lateral stress diagram (Fig. 9), back lateral of diagonal bracing composite supporting structure is controlled by compressive stress, dorsal lateral is controlled by tensile stress, the compressive and tensile stresses appeared the corresponding relationship each other; both tensile stress and compressive stress sudden increase at intersecting point of vertical brace and diagonal bracing.

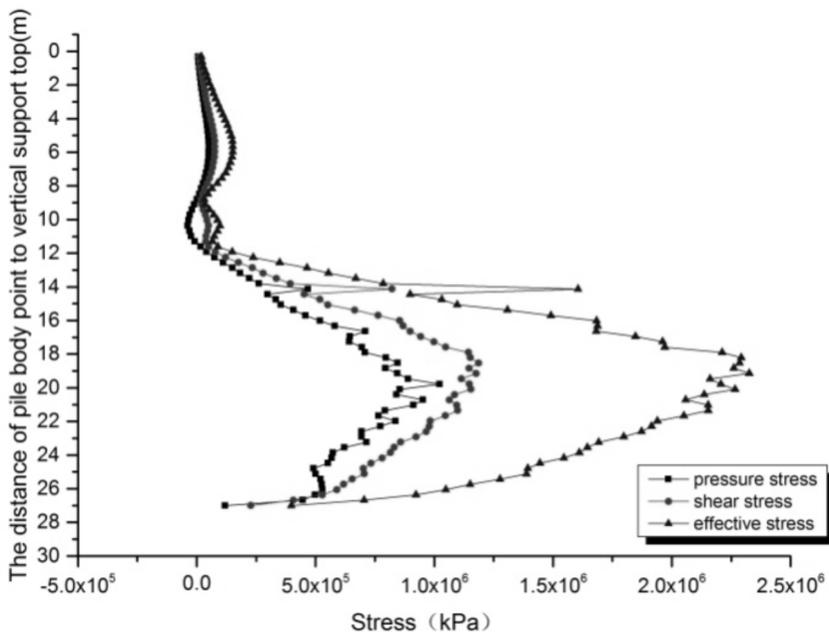


Fig. 8. Vertical brace back lateral compressive stress/shear stress analysis diagram

According to the vertical back lateral principal stress analysis diagram (Fig. 10), back lateral compressive stress larger value corresponds with shear stress of larger value.

Three-dimensional finite element analysis shows that the practical structure as influenced by the geometry size and diagonal bracing supporting foundation, theoretical calculation of the bending moment cannot possible appears zero value, it also does not exist the point in the structure has the zero bending moment, and only affected by the axial force. But through the flexural differential equation algorithm of continuous structure optimized structure can be found in bending moment minimum

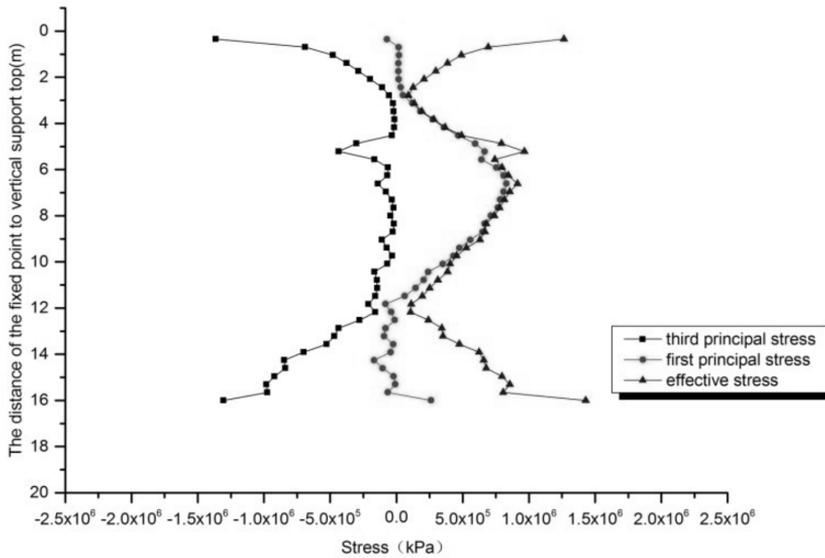


Fig. 9. Vertical brace dorsal lateral principal stress analysis diagram

and axial force of the largest point, so that the concrete compression characteristics of diagonal bracing could be full play, vertical brace of landslide thrust could be fully delivery.

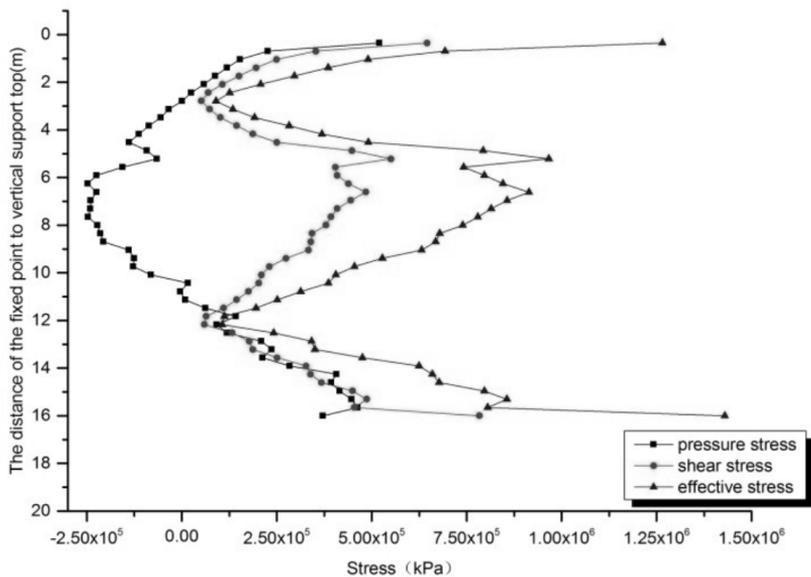


Fig. 10. Vertical brace dorsal lateral compressive stress/shear stress analysis diagram

## 5. Conclusion

The experimental part analyzes and calculates the diagonal bracing composite supporting structure by using flexural differential equation algorithm of continuous structure. After the experimental analysis, the mechanical characteristic of diagonal bracing composite supporting structure under the landslide thrust are obtained. According to the results for comparing the three-dimensional finite element model and the practical engineering of diagonal bracing composite supporting structure, the following conclusions can be summarized: First, diagonal bracing composite supporting structure with the common existing anti-slide pile supporting structure on the mechanical model has essential difference. Diagonal bracing composite supporting structure is a typical supporting structure is given priority to transfer landslide thrust. Second, through the analysis of the existing structure calculation method, the algorithm is proposed based on the deflection equation of continuous structure. Control conditions are put forward on the brace composite structure. A flexural differential equation algorithm of continuous structure is used to optimize for practical engineering calculation. Base on the above calculation results, optimization design with the original design of the calculated results are compared. At the same time, the original design and optimization design calculation results also will compare with the anti-slide pile and anchor pile calculation results. Finally, by comparison with the results, the internal force distribution characteristics of the diagonal bracing composite supporting structure and advantages are obtained.

Through the above points, it is theoretically verified that the algorithm of continuous structure based on the flexural differential equation is reliability. Meanwhile, optimization calculation for the practical engineering project also achieved good effect. Above all, diagonal bracing composite supporting structure is a kind of effective supporting structure of giant, large landslide control.

## References

- [1] P. E. SUDHAGAR, A. A. BABU, V. RAJAMOCHAN, P. JEYARAJ: *Structural optimization of rotating tapered laminated thick composite plates with ply drop-offs*. International Journal of Mechanics and Materials in Design 13 (2017), No. 1, 85–124.
- [2] R. FOULAD, M. MOFID, M. ZARRIN: *On the seismic performance of hat knee bracing system in low-rise multistory steel structures*. Advances in Structural Engineering 18 (2015), No. 3, 325–338.
- [3] J. JIANG, G. Q. LI, A. USMANI: *Effect of bracing systems on fire-induced progressive collapse of steel structures using opensees*. Fire Technology 51 (2015), No. 5, 1249–1273.
- [4] M. R. KAHYAOGU, G. IMANLI, O. ÖNAL, A. S. KAYALAR: *Numerical analyses of piles subjected to lateral soil movement*. KSCE Journal of Civil Engineering 16 (2012), No. 4, 562–570.
- [5] A. KARI, M. GHASSEMIEH, S. A. ABOLMAALI: *A new dual bracing system for improving the seismic behavior of steel structures*. Smart Materials and Structures 20 (2011), No. 12, paper 125020.
- [6] Y. J. SHEN, Q. LÜ, Y. Q. SHANG: *Effect of pile row distance on internal stress of double-row anti-slide piles*. Chinese Journal of Geotechnical Engineering 30 (2008), No. 07, 1033–1037.

- [7] N. MECHBAL, M. VERGÉ, G. COFFIGNAL, M. GANAPATHI: *Application of a combined active control and fault detection scheme to an active composite flexible structure*. *Mechatronics 16* (2006), Nos. 3–4, 193–208.
- [8] A. H. BHRAWY, M. A. ZAKY: *Numerical algorithm for the variable-order caputo fractional functional differential equation*. *Nonlinear Dynamics 85* (2016), No. 3, 1815–1823.
- [9] L. CUDDIHY, A. J. DANIELSSON, P. J. CAHILL, A. F. SAMDANI, H. GREWAL, J. M. RICHMOND, M. J. MULCAHEY, J. P. GAUGHAN, M. D. ANTONACCI, R. R. BETZL: *Vertebral body stapling versus bracing for patients with high-risk moderate idiopathic scoliosis*. *BioMed Research International* (2015), paper 438452.
- [10] X. R. LIU, M. X. OU, Y. R. ZHENG, Y. ZHANG, Y. ZHOU: *Model investigation of sliding resistance mechanism for h-type anti-slide pile*. *Journal of Civil, Architectural & Environmental Engineering 35*, (2013), No. 2, 33–37, 91.
- [11] M. P. NGUYEN, T. C. FUNG, K. H. TAN: *Structural behavior of CHS T-joints subjected to brace in-plane bending at elevated temperatures*. *Proc. International Symposium on Tubular Structures*, 12–14 Sept. 2012, London, United Kingdom, 607–614.

Received June 29, 2017



# Design and kinematic control simulation of wheeled mobile table tennis manipulator

CHEN YAN<sup>1</sup>

**Abstract.** The purpose of this paper is to design a wheeled mobile table tennis manipulator without the limited punching range caused by fixed arms of table tennis robots. By analyzing the structure, the technical characteristics of wheeled differential-drive mobile robot and table tennis robot are combined. In addition, kinematic control simulation is adopted to design wheeled mobile table tennis manipulator. First of all, the software and hardware exploitation platform are constructed by concerning with practical use of table tennis robot, and D-H module is also built. Moreover, the kinematics model of compartment is worked out based on the theory of building trolley module. In addition, the weighted least square method is taken as an example to solve joints limits obstacles-avoidance of mobile manipulator. The experimental results prove the validity of this method through simulation research. According to the requirements of control mission, the control method for redundant mobile manipulator is achieved. Based on the above finding, it is concluded that the relative algorithm can be used to realize the analysis on and research of sub-tasks performances characteristics.

**Key words.** Wheeled mobile table tennis manipulator, joint limits, D-H module, weighted least square method.

## 1. Introduction

### *1.1. Description of the problem*

The commonly used mobile robot mechanisms include wheeled mobile mechanism, legged mobile mechanism, tracked mobile mechanism and wheeled legged mobile mechanism. Among them, wheeled mobile mechanism has a long history, and relatively mature in mechanical design. Therefore, in practical applications, wheeled mobile robots, or mobile manipulators, are the most important compared to other types of mobile mechanisms. The shape structure of wheeled mobile manipulator mainly consists of two parts: wheeled mobile platform and mechanical arm.

---

<sup>1</sup>Department of P. E, North China University of Water Resources and Electric Power, 450011, China

Robotics involves many disciplines like kinematics, intelligent control and bionics, whose application scope expands from industrial production to people's daily life. Among which, table tennis robots not only can perceive and precast the surroundings, but make decisions [1–3], as well as make relative combined actions. This shows advanced intellectuality, realizing automatic countermeasures in sporting events [4]. There are researches on table tennis manipulators both at home and abroad, which basically realized man-machine playing and machine-machine playing [5–6], but its real development condition is quite inferior to the practical man-man playing. For example, man can move body to catch and serve ball in a wider scope, whereas robots' fixed body only allows a very narrow range [7–8].

At present, the motion modes of mobile robots involve wheeled mode and step-mode, the latter is more difficult to control, whereas the former has a relatively advanced technology and is easier to control [9]. We based on the playing condition of mobile table tennis manipulators, made the match more similar to the man to man playing condition. We added mobile parts for the manipulators, chose 4-wheel differential-drive mobile mechanism to make research, designed wheeled mobile manipulators and proved the validity of this method through stimulation.

### *1.2. State of the art*

The United States first proposed the concept of industrial robots, and invented the world's first industrial robot UNIMATE in 1961. Its robot technology has a long history of development and application. The Japanese robot, known as "Robot Kingdom", ranks first in the world no matter the number or density. The western European countries are also developing rapidly in the field of robotics by combining their own R & D and application. The research of mobile manipulator started earlier in foreign countries. The initial research mainly studies the architecture and information processing of outdoor robots from an academic perspective, and builds a buy test system to validate them. In the 90s, with the progress of science and technology, mobile manipulator began to develop more applications on the basis of "buy more now".

China's robot started in the early 1970s. After about 40 years of development, it has gone through three stages, from exploration, research and development to application. After several years of research, China has completed the anti nuclear reconnaissance vehicle, remote control mobile robot and wall climbing robot, and then developed anti explosion robots, eye guided vehicles and so on. Although China is a powerful country in table tennis, the research on table tennis robot started very late. The research on this aspect was first seen in the theoretical research of Shanghai Jiao Tong University in the 90s of the last century.

## 2. Construction of development environment

### 2.1. Hardware structure

We applied wheeled mobile mechanism to study the mobile manipulator in this research, whose hardware consists of compartment physical structure, dynamo, control box and CAN card [10–11]. The compartment structure of mobile manipulator applied in this research is showed in Fig. 1.

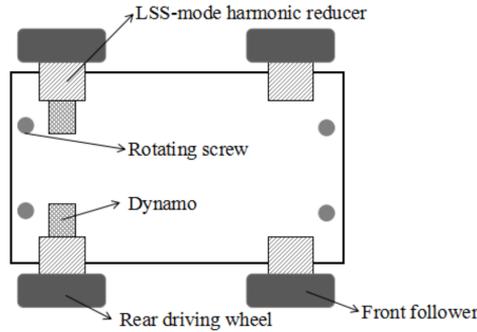


Fig. 1. Compartment structure of mobile manipulator

From Fig. 1, we can see that the compartment has 4 wheels, among which, the first two wheels are followers and the other two are driving wheels, and each one has an independent dynamo. Moreover, 4 wheels are all fixed, can only move back and forth to make the dynamo having more driving force. We utilized DC dynamo as the driving dynamo of compartment. When manipulator and dynamo are appropriately assembled, we can apply PID control to debug the dynamo( current, speed, location).

The framework diagram of debugging module is shown in Fig. 2. Input the current and peak current, after the limiting step, compare them with feedback current, and conduct PI control, then after pulse width modulation(PWM) set them as the input instruction of dynamo.

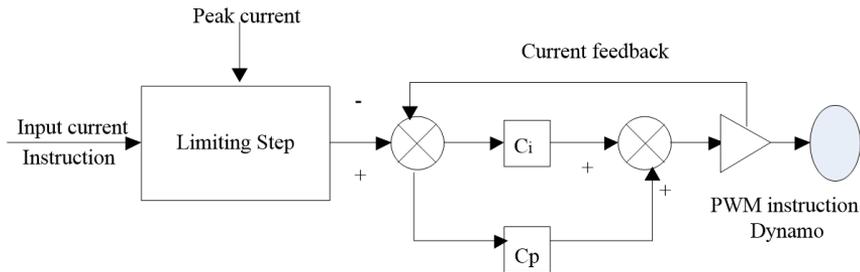


Fig. 2. Current loop debugging module of dynamo

The framework diagram of debugging module is shown in Fig. 3. After the limiting and filtering, work out the difference between the speed and feedback speed,

output it through filtering after the PI control. Two times of filtering process has greatly relieved the disturbance of noise.

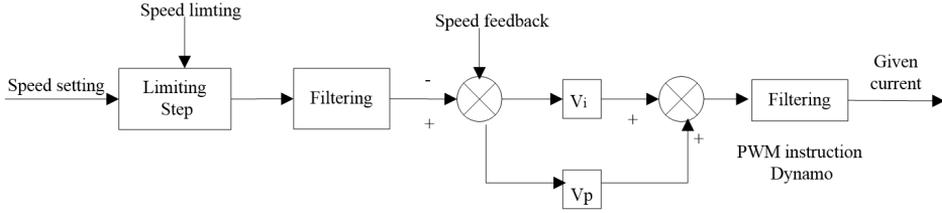


Fig. 3. Speed loop debugging module of dynamo

The framework diagram of debugging module is shown in Fig. 4. The forward circuit includes location, speed and acceleration, which can decrease error during location trailing. After debugging, the dynamo can reach a better tracing effect, laying a solid foundation for the following research.

**2.2. D-H parameter**

D-H (Denavit-Hartenberg) parameter includes 4 parameters: connecting rod length  $a_i$ , connecting rod corner  $\alpha_i$ , connecting rod offset distance  $d_i$ , and articulation angle  $\theta_i$ . The D-H coordinate system module is shown in Fig. 4.

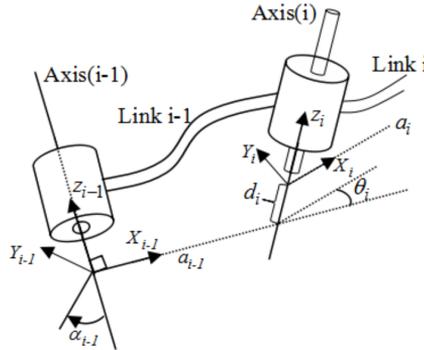


Fig. 4. D-H coordinate system module

Keep the number of  $a_i$  positive and not limiting the other three numerical values. Positive and negative values have different meanings. According to the established module, we can build various connecting rod fixed coordinate system [12–13]. The origin of fixed coordinate system can be set randomly. Choose the appropriate world coordinate system in the space based on different needs. According to the formula (1) describing the homogeneous coordinates transfer matrices on neighboring joints, we describe the space and location relationship among joints. After another transferring, work out the position of the target point in world coordination system. The formula

of homogeneous coordinates transfer matrices on neighboring joints is as follows

$${}^i T_{i-1} = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) & 0 & a_{i-1} \\ \sin(\theta_i) \cos(\alpha_{i-1}) & \cos(\theta_i) \cos(\alpha_{i-1}) & -\sin(\alpha_{i-1}) & -\sin(\alpha_{i-1})d_i \\ \sin(\theta_i) \sin(\alpha_{i-1}) & \cos(\theta_i) \sin(\alpha_{i-1}) & \cos(\alpha_{i-1}) & \cos(\alpha_{i-1})d_i \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

So the position of manipulator end to the world coordination system can be described as

$${}^0 T_n = {}^0 T_1 {}^1 T_2 {}^2 T_3 \cdots {}^{n-1} T_n. \quad (2)$$

In the formula,  $n$  refers to the number of joints.

### 2.3. Software development platform

We apply Visual Studio as the software development platform this time, Visual C++ as the programming language. There are two CAN cards, amounting to four channels, each has 2–3 dynamos, so we conduct programming with 4 threads.

## 3. Materials and methods

### 3.1. Kinematic modeling of the compartment

Most of mobile robots apply double-wheeled difference-drive mode, whose mobile module is shown in Fig. 5.

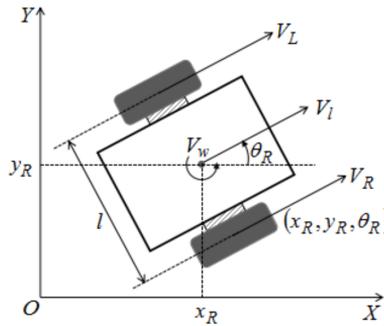


Fig. 5. Double-wheeled trolley kinematic module

In the figure,  $V_L$  and  $V_R$  refer to the left and right linear speed of two wheels respectively,  $l$  is distance of wheel shaft centers,  $V_I$  and  $V_w$  are the line speed and angular speed of the trolley [14]

$$V_I = \frac{V_L + V_R}{2}, \quad V_w = \frac{2V_R - V_L}{l}. \quad (3)$$

Moreover:

$$V_L = \omega_L R, \quad V_R = \omega_R R. \quad (4)$$

Here,  $\omega_L$  and  $\omega_R$  are the left and right angular speeds of the two wheels, respectively. And the centroid motion equation of the robot is

$$\dot{x}_R = V_l \cos(\theta_R), \dot{y}_R = V_l \sin(\theta_R), \dot{\theta}_R = V_\omega. \quad (5)$$

Given these, we can see that

$$\begin{bmatrix} \dot{x}_R \\ \dot{y}_R \\ \dot{\theta}_R \end{bmatrix} = \begin{bmatrix} \frac{R \cdot \cos(\theta_R)}{2} & \frac{R \cdot \cos(\theta_R)}{2} \\ \frac{R \cdot \sin(\theta_R)}{2} & \frac{R \cdot \sin(\theta_R)}{2} \\ \frac{R}{l} & \frac{R}{l} \end{bmatrix} \begin{bmatrix} \omega_R \\ \omega_L \end{bmatrix}. \quad (6)$$

To simplify formula (6), conduct decoupling process. And then

$$\begin{bmatrix} \dot{x}_R \\ \dot{y}_R \\ \dot{\theta}_R \end{bmatrix} = \begin{bmatrix} \cos(\theta_R) & 0 \\ \sin(\theta_R) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V_l \\ V_\omega \end{bmatrix}. \quad (7)$$

When the multi-wheeled trolley does rotational motion, there must be a center of rotation, that is instantaneous center of curvature (ICC) [15]. As to difference-drive trolley, ICC locates on the public axis of the two wheels, and can moves back and forth. The location is decided by the speed of the two wheels. The mobility model of trolley is shown in Fig. 6.

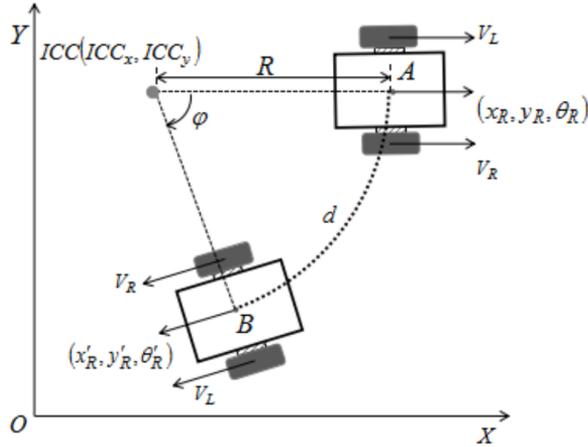


Fig. 6. Mobility model of trolley

In formula (7),  $R$  is the distance between ICC and the center of gravity A. So

$$\frac{v_L}{v_R} = \frac{R - \frac{l}{2}}{R + \frac{l}{2}} \rightarrow R = \frac{l}{2} \frac{v_R + v_L}{v_R - v_L}. \quad (8)$$

When the trolley makes direct linear movement,  $v_R = v_L$ , when  $v_R \neq v_L$ , the trolley makes rotational motion. When the time is  $t$ , trolley moves from A, after the time length of  $\varepsilon t$ , it moves to B. ICC can be calculated as follows:

$$ICC = [x_R - R \sin(\theta_R), y_R + R \cos(\theta_R)] . \tag{9}$$

According to  $(1 + \varepsilon)t$  and the angular speed  $v_\omega$ , we can work out the position of B:

$$\begin{bmatrix} x'_R \\ y'_R \\ \theta'_R \end{bmatrix} = \begin{bmatrix} \cos(\omega\varepsilon t) & \sin(\omega\varepsilon t) & 0 \\ \sin(\omega\varepsilon t) & \cos(\omega\varepsilon t) & 0 \\ 0 & 0 & 1 \end{bmatrix} . \tag{10}$$

In the same way, the motion distance  $d$  and rotating angle  $\phi$  can be worked out:

$$d = \int_t^{t+\varepsilon t} v_1 dt = \int_t^{t+\varepsilon t} \frac{v_L + v_R}{2} dt , \tag{11}$$

$$\phi = \frac{d}{R} = \frac{\int_t^{t+\varepsilon t} (v_L + v_R) dt}{l(v_L + v_R)} (v_R - v_L) . \tag{12}$$

Given these, we can work out the linear speed  $v_L$  and  $v_R$  of the two wheels, and the angular speed  $v_\omega$  of the trolley.

### 4. Trajectory planning of mobile manipulator

The trajectory planning of mobile manipulator refers to working out the reserved trajectory concerning with the relative kinematic knowledge and the requirements of controlling task. The controlling task may be the mobility of a certain joint of the manipulator, or the concrete mobility in space.

The trajectory planning process can be seen in Fig. 7. During the serving process, there are two different plannings, one is task space planning when serving the ball, the other one is joint space planning when restoring. The interpolation algorithm of these two plannings are not the same. See the following for more details.

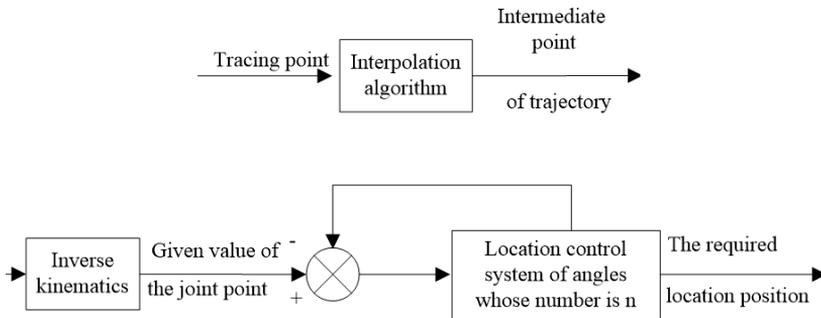


Fig. 7. Trajectory planning of mobile manipulator

We will introduce task space planning with the method of cubic polynomial interpolation, which can be classified into with and without intermediate point. If we want to work out the cubic polynomial interpolation, there must be 4 known conditions, or constraint conditions: the angle and speed limit at two positions.

Without passing intermediate point, if the angle constraint of two positions are within the range of joint limit, the speed constraint is 0. If the starting motion point of every joint is  $t_i$ , the joint angle is  $\theta_i$ . If the ending point is  $t_e$ , the joint angle is  $\theta_e$ . Then

$$\theta(t_0) = \theta_0, \theta(t_e) = \theta_e; \dot{\theta}(t_0) = \dot{\theta}(t_e) = 0 \quad (13)$$

and the cubic polynomial interpolation is

$$\theta(t) = A_0 + A_1t + A_2t^2 + A_3t^3. \quad (14)$$

Taking the derivative of  $\theta(t)$ , we can see the joint speed  $\dot{\theta}(t)$  and joint acceleration  $\ddot{\theta}(t)$ . Thus, four equations on factors can be drawn out, see the following formula

$$\theta_0 = A_0, \quad A_1 = 0, \quad \theta_e = A_0 + A_1t_e + A_2t_e^2 + A_3t_e^3, \quad A_1 + A_2t_e + 3A_3t_e^2 = 0. \quad (15)$$

The processing of joint space trajectory is a circulated process. Therefore, we only need to make planning following the steps above to accomplish the whole motion planning.

However, passing the intermediate point can be classified into two conditions: the speed is 0 or is not 0. When the speed is 0, we can process it according to the solutions above. When it is not, we need to process it based on the following steps. So the constraint condition that needs to be changed when speed isn't 0 is:

$$\dot{\theta}(0) = \dot{\theta}_0, \dot{\theta}(t_e) = \dot{\theta}_e. \quad (16)$$

The intermediate point value produced during joint space planning when using interpolation arithmetic is the joint angle value. The intermediate point value produced during task space planning is the three-dimensional coordinate values in Cartesian space. But the mobility of manipulators are realized by the coordinated motion of many joints, during which, a large number of arithmetical operations are needed, which leads to a longer control period.

When interpolating among task space, there are two frequently-used methods: linear interpolation method—work out the relative joint angle value through inverse operation, move to realize the required trajectory, and at the same time, the continuity of various points are approved; arc interpolation arithmetic, which includes planar arc interpolation arithmetic and space arc interpolation arithmetic.

## 5. Results

### 5.1. Kinematic control method

During the practical applying, the freedom degree of manipulator is about 8, to realize the change of end position, we need 6 task vectors at most. So, manipulator is of redundancy, and can be used to accomplish some extra submissions.

The kinematic equation of manipulator is

$$x = f(q), \quad \dot{x} = J(q)\dot{q}. \quad (17)$$

In this formula,  $x$  is the task space vector quantity of dimension  $m$  of the manipulator,  $q$  is the joint space vector quantity of dimension  $n$ ,  $\dot{x}$  and  $\dot{q}$  are the relative speed,  $f$  is the forward direction kinematic relation of them,  $J(q)$  is the Jacobian matrix in line  $m$  list  $n$ .

When  $m < n$ , the mobile manipulator is a redundant manipulator, resulting in non numerous inverse solution of formula (17). However, there are many methods to accomplish the control on mobile manipulator, whose frequent subtasks are a large number. We work out this equation by choosing weighted least square method and by which to solve joint limits and avoid problems.

The expected result of this research: when the joint of manipulator is close to joint limits, constraint the subtask of joint obstacles-avoidance to ensure the safety of its hardware. When the joint is far away from the joint limits, do not constraint the subtasks any more, thus to ensure the realize of main tasks, and gain a well accuracy in end task control.

The operation plan of weighted least square method is: introduce matrix and vector quantity to meet the equation

$$J_w = JW^{-\frac{1}{2}}, \quad \dot{q}_w = W^{\frac{1}{2}}\dot{q}. \quad (18)$$

Here,  $W \in R^{n \times n}$  is symmetric positive-definite weighted matrix, we can work out that

$$\dot{x} = J\dot{q} = J_w\dot{q}_w. \quad (19)$$

If the Jacobian matrix is nonsingular, then the weighted least square solution of formula (19) is:

$$\dot{q} = W^{-1}J^T (JW^{-1}J^T)^{-1} \dot{x}. \quad (20)$$

As to joint obstacles-avoidance, there is:

$$H(q) = \sum_{i=1}^n \frac{(q_u - q_d)^2}{4(q_u - q_i)(q_i - q_d)}. \quad (21)$$

In this formula,  $q_i$  is the number  $i$  joint angle,  $q_u$  and  $q_d$  are the upper and lower limits of the mobility range of joints respectively.

## 5.2. Simulation experiment

Let  $(x_0, y_0, z_0)$  refer to world coordinated system and  $(x_1, y_1, z_1)$  be the coordinate system of mobile joints. The left 7 are the coordinate system of 7 joints of the manipulator. The relative D-H parameter is shown in Table 1.

Now we will conduct simulation study on joint limits avoidance through weighted least square method. If the initial joint position is

$$q_i = [0.26 \ 200 \ 90 \ 80 \ 250 \ 105 \ 180 \ 135] ,$$

where the unit of 0.26 is “m”, the others are “°”. Then, the relative joint position is

$${}^0T_{8i} = \begin{bmatrix} 0.0502 & 0.1855 & 0.9811 & 0.3304 \\ 0.2005 & -0.9638 & 0.1722 & 0.8396 \\ 0.9779 & 0.1880 & -0.0853 & 0.2571 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Table 1. D-H parameter

Joint	J1	J2	J3	J4	J5	J6	J7	J8
Rotation angle $\alpha$ (°)	0	90	90	90	90	90	0	90
Connecting rod length $a$ (mm)	0	0	0	0	0	0	1.2	0
Connecting rod deflection distance $d$ (mm)	0.26*	68	5	156	325	-5	0	0
Joint angle $\theta$ (°)	90#	0	180	180	90	180	180	90

Note: \*is variation and # is fixed value.

Its position in task space is

$$x(0) = [0.4776 \ 0.8652 \ 0.2443 \ 0.2083 \ 1.3247] .$$

The whole motion cycle is 4 s, and the objective position of task space is

$$x(4) = [0.5501 \ 1.0647 \ 0.3998 \ 0.4699 \ 1.0528] .$$

The results are shown in Figs. 8–10. In Fig. 8 we can see that the whole motion process is steady and soft, realizing joint avoidance. Figure 9 shows that the speed changes smoothly during the whole process of motion, and without reaching the joint speed limits. From Fig. 10 we can see that the error precision of trailing the main task is  $10^{-3}$ , accomplishing both the main task and subtask of joint obstacles-avoidance.

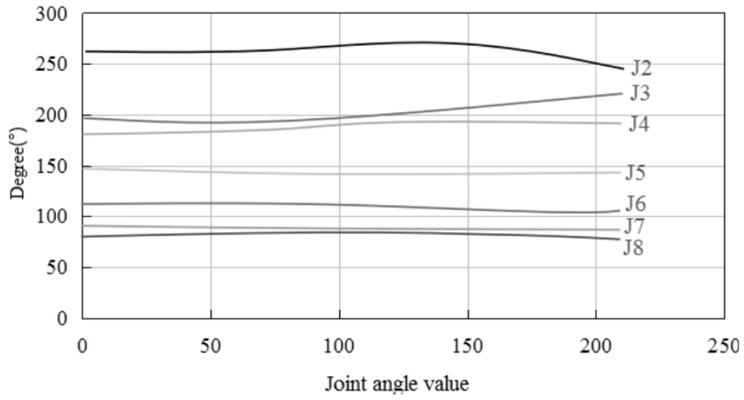


Fig. 8. Joint location (the horizontal axis is joint angle value, vertical axis is degree, i.e. “°”)

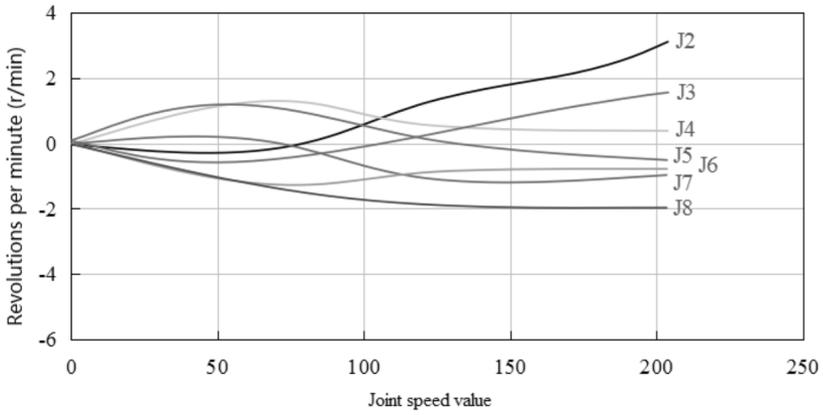


Fig. 9. Relationship of joint speed and revolutions per minute

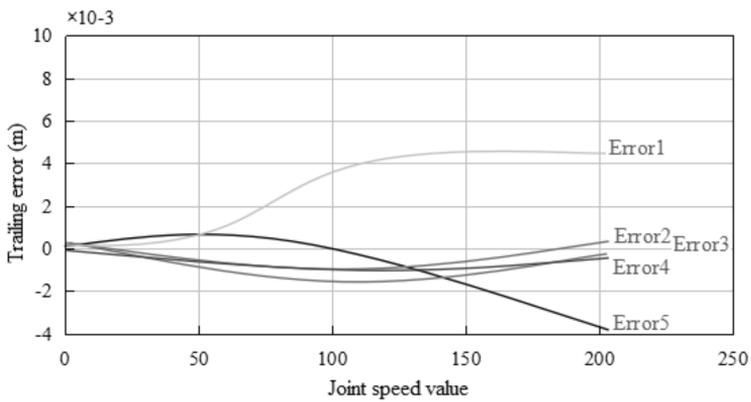


Fig. 10. Trailing error of the main task of joint limits avoidance

## 6. Conclusion

Based on the experimental results for the wheeled mobile table tennis manipulator, the mobile parts of the manipulators are designed to construct the whole mobile manipulator. From the perspective of freedom range, it is proved that mobile manipulator improves its redundancy and the flexibility when processing subtasks like joints obstacles-avoidance, space obstacles-avoidance and so on. In addition, the weighted least square method is selected to complete the research on subtask of the manipulator (joint limits obstacles-avoidance), which proves the validity of this method. Moreover, it is concluded that the heavy weight of the mobile parts can lead to a greater movement inertia. Therefore, it is necessary to avoid moving by a large margin to control accuracy during distributing the joint speed. Meanwhile, it is believed that this problem will also be refined specifically in the following researches.

### References

- [1] K. MÜLLING, J. KOBER, O. KROEMER, J. PETERS: *Learning to select and generalize striking movements in robot table tennis*. The International Journal of Robotics Research 32 (2013), No. 3, 263–279.
- [2] G. CAMPION, G. BASTIN, B. D’ANDREA-NOVEL: *Structural properties and classification of kinematic and dynamic models of wheeled mobile robots*. Journal Nelineinaya Dinamika 7 (2011), No. 4, 733–769.
- [3] M. J. HAYAWI, A. A. M. ISA, M. A. ANUAR, A. R. OMAR: *Research article kinematic and dexterity analysis of a 3-dof parallel manipulator*. Journal of Applied Sciences, Engineering and Technology 12 (2016), No. 2, 239–248.
- [4] Y. D. WANG, L. W. WANG, H. W. ZHANG, H. J. DONG: *Research and application for the regularity of distribution about curvature radius and curvature center of the linkage point track in a link mechanism*. International Journal of Plant Engineering and Management 17 (2012), No. 3, 160–168.
- [5] S. HAN, H. HA, Y. ZHAO, J. LEE: *Assumed model feedforward sliding mode control for a wheeled mobile robot with 3-DOF manipulator systems*. Journal of Mechanical Science and Technology 31 (2017), No. 3, 1463–1475.
- [6] Y. C. CHANG, Y. YAMAMOTO: *On-line path planning strategy integrated with collision and dead-lock avoidance schemes for wheeled mobile robot in indoor environments*. Industrial Robot 35 (2008), No. 5, 421–434.
- [7] H. KIM, B. K. KIM: *Online minimum-energy trajectory planning and control on a straight-line path for three-wheeled omnidirectional mobile robots*. IEEE Transactions on Industrial Electronics 61 (2014), No. 9, 4771–4779.
- [8] Y. D. PATEL, P. M. GEORGE: *Simulation of kinematic and workspace analysis of 3-PRS parallel manipulator*. International Journal of Modeling, Simulation, and Scientific Computing 6 (2015), No. 1, paper 1550007.
- [9] J. JOHNSON, R. PILLAI, S. M. VIGNESH, A. L. SRINIVAS: *Design and kinematic analysis of a 6-dof parallel manipulator for pharmaceutical application*. Journal of Chemical and Pharmaceutical Sciences 9 (2016), No. 4, 3283–3288.
- [10] D. LIANG, Y. SONG, T. SUN, G. DONG: *Optimum design of a novel redundantly actuated parallel manipulator with multiple actuation modes for high kinematic and dynamic performance*. Nonlinear Dynamics 83, (2016), Nos. 1–2, 631–658.
- [11] J. YAN, X. GUO, Y. LIU, J. ZHAO: *The design and kinematic analysis of a modular manipulator*. Journal of Harbin Institute of Technology 47 (2015), No. 1, 20–25.
- [12] D. ZHANG, B. WEI: *Design, kinematic and dynamic modeling of a novel tripod based manipulator*. Robotica 34 (2016), No. 10, 2186–2204.

- [13] T. ITO, S. TAKATA: *Existence theorem and error bound of numerical solution for optimal control problems of continuous system*. Transactions of the Society of Instrument and Control Engineers 18 (1982), No. 7, 647–652, Released (2009).
- [14] Z. SHAO, X. TANG, L. WANG, D. SUN: *Atlas based kinematic optimum design of the Stewart parallel manipulator*. Chinese Journal of Mechanical Engineering 28 (2015), No. 1, 20–28.
- [15] C. PARK, D. I. PARK, D. KIM: *The analysis of trajectory tracking error caused by the tolerance of the design parameters of a parallel kinematic manipulator*. Journal of Korea Robotics Society 11 (2016), No. 4, 248–255.

Received June 29, 2017



# Electric braking method in hydroturbine generating unit and transformation scheme

YANTIAN ZHANG<sup>1</sup>, YUHANG CHEN<sup>2</sup>, JINSONG TAO<sup>2,3</sup>

**Abstract.** The purpose of this paper is to study the electrical control and reconstruction of hydroelectric generating units. In general, there are two kinds of braking methods for hydraulic turbines: mechanical braking and electrical braking. The mechanical braking method has the disadvantage of low automation. Through the research and analysis of practical examples, a modified scheme of the combination of electric and mechanical braking water turbine generator sets is put forward. First of all, from the aspects of moment control and the design scheme, a detailed explanation on the electric braking method is introduced. The range and scope of application of various schemes concerning the advantages and disadvantages of different design schemes on the electric braking system is discussed. Secondly, based on the existing situation and transformation requisition of the generating unit electric braking system in Pingban hydropower station, this research is conducted. It is also the necessity and availability of the transformation. Finally, the concrete transformation scheme on the electric braking system in Pingban hydropower station is expressed in detail. The results show that the scheme improves the operating efficiency of hydroelectric generating units. Therefore, it is concluded that electrical control has important significance for hydroelectric generating units.

**Key words.** Hydroturbine generating unit, electric braking, mechanical braking, transformation scheme.

## 1. Introduction

During the working of electric system, the hydropower station generating unit is in the functions of peak load regulation of power grid and emergency standby [1–2]. Hydroturbine generating unit has two kinds of braking methods: electric braking and mechanical braking [3]. In general, the mechanical braking method creates the braking effect through the friction drag generated by the touching of brake valve and

---

<sup>1</sup>Wuhan Power Supply Design Institute, Wuhan, 430040, China

<sup>2</sup>School of Electrical Engineering, Wuhan University, 430072, China

<sup>3</sup>Corresponding author

brake ring, whereas the electric braking works through consuming the main power in a form of electricity and generating reverse electromagnetic force in the motor [4].

The generating unit employs traditional braking method: mechanical braking method, which is reliable, convenient and widely applicable, with low energy consumption [5]. Its braking effect will not be influenced by faults like power interruption or short-circuit of line owing to its high reliability and safety [6]. However, this mechanical braking method also has some shortcomings, for example, the powder generated during the braking process may enter the ventilation ducts, and powder accumulation during long periods of time will downsize the cross-section area of ventilation ducts and result in loss in cooling effect. During the process of braking, the surface temperature of the braking ring will increase rapidly, leading to thermal distortion and even crack. The airlock sometimes cannot fall itself, so our personnel will inspect the wind tunnel at every stop of it, which also lowers its level of automation [7].

To resolve the problems in mechanical braking, we conduct electric braking on hydroturbine generating unit here. Electric braking refers to a non-contact braking method based on synchronous motors electromagnetic induction principle, whose advantages lie in the high level of automation, high reliability and efficiency, as well as low abrasion and pollution [8]. On account of this, we analyze the electric braking moment control of generating unit and the influence on relay protection. We also design the transformation scheme for the electric braking system in Pingban hydropower station generating unit concerning the setbacks.

## 2. State of the art

During the practicing process of hydroturbine generating unit, the electric braking moment changes a lot, which is much more obvious especially when the generating unit speed decreases to be in a low statue [9]. In general, the electric braking moment is larger than the ultimate moment of the principal axis system [10]. So, there is some kind of danger. To ensure the safe operating of the generating unit, the electric braking system will be installed on the generating units that are frequently switched on and off, in order to better control the braking moment around rated torsion moment  $M_e$ . The ultimate moment of principal axis system  $T$  [11–12] can be expressed as:

$$T = KM_e. \quad (1)$$

In this formula,  $K$  refers to the safety factor (often being 2.5). And the torsion moment of the principal axis system in generating unit  $M$  can be identified as:

$$M = \frac{60 * P_M}{2\pi n}, \quad P_M = P_2 - P_{Cu}. \quad (2)$$

In this formula,  $P_M$  is the generator electromagnetic power,  $P_2$  is the electric load power. When the hydroturbine generating unit is under the rated speed ( $N_e$ ), the

rated power( $P_e$ ) is in the condition of ( $P_e = P_2$ ), then  $M_e$  can be identified as

$$M_e = \frac{60(P_e - P_{Cu})}{2\pi N_e} P_e = \sqrt{3} U_e I_e \cos \phi. \quad (3)$$

Here,  $\phi$  is generator power factor, and  $I_e$  is the excitation current. If the copper loss of the stator winding is omitted, then

$$M_e = \frac{60P_e}{2\pi N_e} = \frac{60(\sqrt{3}U_e I_e \cos \phi)}{2\pi N_e}. \quad (4)$$

If the electric braking moment  $M_E$  is not larger than  $M_e$ , then

$$\frac{60 \cdot 3I_k^2 R}{2\pi n} \leq \frac{60(\sqrt{3}U_e I_e \cos \phi)}{2\pi N_e}. \quad (5)$$

In this formula,  $I_k$  is the braking current. During the process of electric braking,  $M_E < M_e$  at first. If  $I_k = I_e$ , and remains constant, there is no need to modulate  $I_e$  to change  $I_k$ . This is the braking phase I. If the speed of generating unit keeps decreasing,  $M_E = M_e$ , until the generating unit stop completely. This is braking phase II.

If  $I_k = I_e$  and remains constant, and also  $M_E < M_e$  and  $n$  is the variable, then

$$\frac{60 \cdot 3I_k^2 R}{2\pi n} < \frac{60 \cdot 3(\sqrt{3}U_e I_e \cos \phi)}{2\pi N_e} \rightarrow \frac{\sqrt{3}I_e R}{U_e \cos \phi} < \frac{n}{N_e}. \quad (6)$$

The critical rotating speed value is

$$n_1^* = \frac{\sqrt{3}I_e R}{U_e \cos \phi}. \quad (7)$$

In this phase, the excitement current must be regulated, and keeping  $M_E$  around  $M_e$ . That is

$$\frac{60 \cdot 3I_k^2 R}{2\pi n} = \frac{60 \cdot (\sqrt{3}U_e I_e \cos \phi)}{2\pi N_e}. \quad (8)$$

Because  $I_k$  and  $n$  are all variables, then

$$I_k^2 = \frac{n}{N_e} \cdot \frac{U_e I_e \cos \phi}{\sqrt{3}R}. \quad (9)$$

From this, we can see that if keeping the value of  $M_E$  around  $M_e$ , we just need to regulate  $I_e$  and  $n$ , ensuring that they satisfy the relationships in formula (9).

### 3. Methodology

Hydroturbine generating unit is often equipped with independent electric braking device, as shown in Fig. 1.

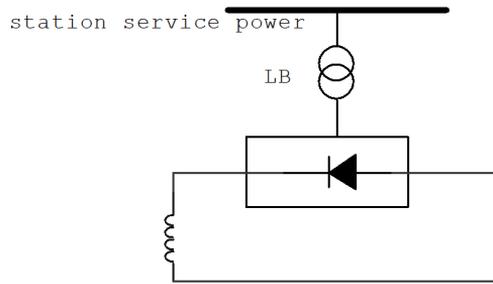


Fig. 1. Structure of independent electric braking device (note: LB—excitation transformer)

his device is applicable for all hydroturbine generating units. But it requires to set panel additionally, adding to the cost and difficulty in setting. However, to overcome the setbacks of independent electric braking method, some foreign researchers proposed a new mode of combining the electric braking of the generating unit with the excitation system [13–15], as shown in Figure 2.

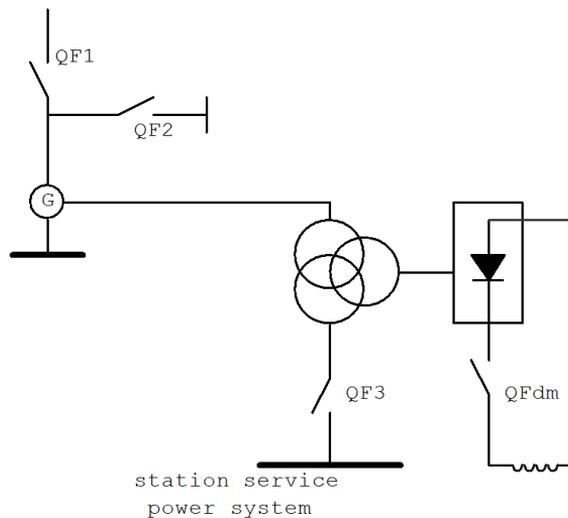


Fig. 2. System block diagram of combining electric braking with excitation system in generating unit (notes: QF1—circuit breakers, QF2—short circuiting switch, QF3—braking switch, QFdm—magnetic blow-out switch)

After analysis of Fig. 2, we can see that this mode can decrease costs to a large extent and also downsize the floor space. It belongs to an economic mode, deserving to be put into use.

In the practical process of operating, however, the mechanical braking of the generating unit is also an indispensable part, because electric braking only applies to the normal shutdown of generating unit or the emergency shutdown of non-

mechanical generating unit [16]. Therefore, to improve the operating situation of thrust-bearing liner at low rpm, most manufacturers employ composite braking of electricity and mechanics at low rpm. The braking effect is better when starting the mechanical braking system at 10 % of the rated rotation speed [17].

## 4. Result analysis and discussion

### 4.1. About the hydropower station

We conduct this research with pingban hydropower station as an example. There are 3 hydroturbine generating units with 135 MWh in this station. The area of reservoir-controlled basins is 56000 km<sup>2</sup>, the average annual flow rate is 616 m<sup>3</sup>/s, the average annual volume of runoff is  $1.94 \times 10^{10}$  m<sup>3</sup>. The normal storage level is 440 m, and the total reservoir storage is  $2.78 \times 10^8$  m<sup>3</sup>.

### 4.2. Analysis on the necessity and availability of transformation

The braking method that Pingban hydroturbine generating unit employs is mechanical braking. The operating time and braking time is a bit long. The braking method combining electricity and mechanics refers to decreasing the rotating speed of the generating unit by conducting mechanical braking fraction moment rather than electric braking moment on the generator during stop [2], thus to realize its braking effect. Furthermore, the hydroturbine generating unit in Pingban hydropower station is switched on or off frequently. So, to avoid intensifying the system abrasion and improving mechanical durability, it is necessary to transform the mechanical braking method to the method of combining electric braking and mechanical braking.

Based on the above analysis on operation process and other aspects of electric braking moment, we can work out the decelerated curve of the generating unit (Fig. 3). The following is about the calculation on the basic parameters of 4 hydroturbine generating units. The original computed parameters of generating unit are expressed in Table 1.

From Fig. 3, we can see that it needs 290 s for the generating unit to stop from the rated rotating speed, among which, the time length from the start of electric braking to stop is 230 s.

According to the development and construct situation of Pingban hrdropower station in recent years, we choose the electric braking point of short circuit at the generator outlet. From the generator outlet to the low voltage side of main transformer, the main circuits are PT cabinet (set at the middle level of the main power house), generator outlet CT, excitation transformer, and generator circuit-breaker respectively. In addition, the equipment needed additionally involves [18] an electric braking switch and braking transformer.

Table 1. Original computed parameter of generating unit

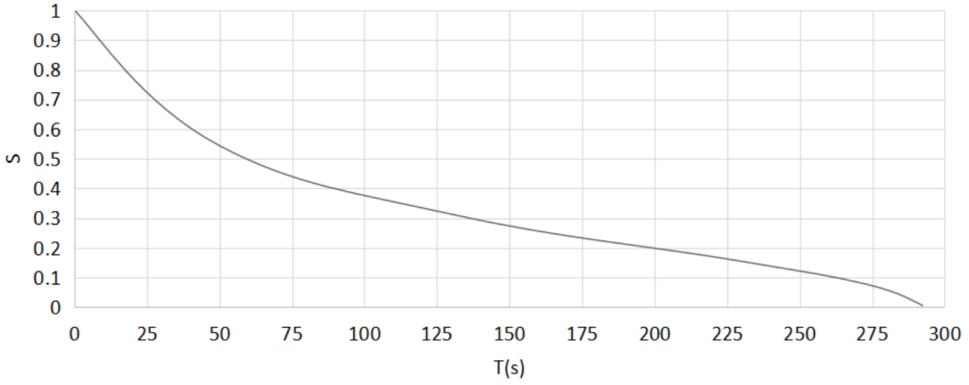


Fig. 3. Decelerated curve of the hydroturbine generating unit (notes:  $S = n/n_N$ , the rotating speed of electric braking is  $0.05n_N$ , which is 100 rpm and braking current  $I = I_N$ )

Parameter	Value	Parameter	Value
Rated power $P_N$	220 MW	Specific rotating speed $n_x$	150 rpm
Rated voltage $U_N$	18 kV	Stator core inside diameter $D_i$	8.06 m
Rated power factor $\cos \theta$	0.9	Stator core length $L_i$	2.70 m
Rated current $I_N$	7857 A	Stator direct axis relative synchronous reactance $X_d$	1.057
Rated rotating speed of the generating unit $n_e$	200 rpm	Stator winding phase resistance $R$	0.00156 $\Omega$
Flywheel moment $GD_2$	21250 t m <sup>2</sup>	Weight of rotating part of generating unit FR	690 t
Runner diameter $D$	4.5 m	Unit pressure of trust bearing	42.13 kg/cm <sup>2</sup>
Design head $H_d$	185 m	Mechanical time constant $T_{mec}$	9.487 s

### 4.3. Design on the transformation scheme

The core concept of this transformation scheme can be generalized as follows:

If setting the electric braking switch on the generator outlet, when the generating unit splits with the system, and the water distributor is closed, the rotating speed will decrease rapidly under the concerted effect of different moments [19]. When the speeding rate decreased to 50% of the rated rotating speed, make the electric

switch working. At the same time, decrease the rotating speed of the generating unit rapidly by adding excitation in the generator via electric braking transformer. When the rotating speed increase to 5 % or 6 % of the rated rotating speed, set the mechanical system to work, and the braking effect can be realized by decreasing the rotating speed to 0.

The relative parameters designed based on the designing principle of electric braking system are shown in Table 2.

Table 2. Relative parameters of electric braking system

Parameter	Value	Parameter	Value
Rated power	220 MW	Excitation transformer ratio	18 kV/670 V(676 V)
Rated terminal voltage	18 kV	Excitation transformer capacity	3×600 kVA
Rated excitation voltage	301 V	No-load excitation current	850 A
Rated excitation current	1565 A	No-load excitation voltage	122 V

The preliminary plan of this research is that: based on the general design mode for the electric braking system in the hydroturbine generating station, the electric braking system can adopt excitation system and share the same rectifying and controlling system with the electric braking system, and it is also possible to have two systems independent to each other. However, in sight of the practical operation of Pingban secondary power station, the primary excitation system has not concerned the practical situation of the electric braking function. The following text is about the detailed scheme of electric braking excitation and the controlling system and the allocation plan.

Without changing the preliminary excitation system, add a braking transformer, switch of low voltage side of excitation transformer, switch of direct current side of excitation and electric braking switch. What is more, add a set of independent rectifying device and electric braking controlling system.

The main circuit of the electric braking system employs independent electric braking device. To decrease the investment in this transformation scheme, the rectifying system uses the full bridge diode rectification. The schematic wiring diagram of this scheme is in Fig. 4.

Based on the above transformation scheme, considering the insulation of electric braking system and excitation device of the generating unit, an electric braking device completely independent from the former excitation system can be chose to ensure the operation of generating unit even when the electric braking device has faults, without influencing the normal on-load operation of the generating unit. Based on this, we designed the wiring diagram of primary electric braking equipment (seeing Fig. 5), which also increase the safety of the system.

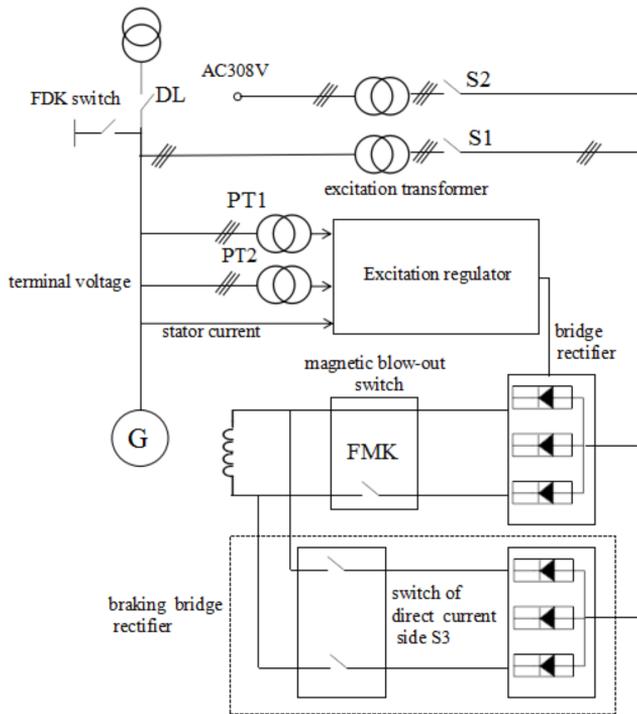


Fig. 4. Schematic wiring diagram

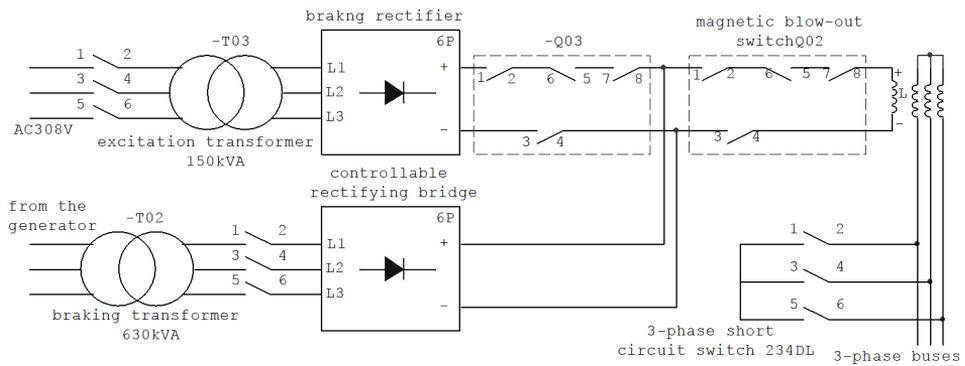


Fig. 5. Wiring diagram of electric braking primary equipment

### 5. Conclusion

Most large-scale hydropower generating units employ traditional mechanical braking method. To shortcut the operation time of the generating unit at low rpm, it must go into a penalty brake continuously. Using the mechanical braking system at high rotating speed may abrade the braking panel. Therefore, considering the safety

of the system, we should choose the electric braking method. We conduct a detailed study on electric braking and its transformation scheme from the following aspects: first, we analyze the principle of electric braking and divide the electric braking moment control of hydroturbine generating unit into 2 phases, and also construct the modules of these phases concerning their different features. The demarcation point of the two phases is critical speed of rotation. Second, in addition, we discussed those three schemes of the electric braking of generating unit respectively, and analyzed the applicable range and environment of different schemes, providing the theoretical support for the following transformation scheme. Third, concerning the transformation requirement of the braking system of generating unit in Pingban hydropower station, we analyzed and discussed the necessity and availability of the braking method transformation. Involving the practical engineering example, we designed the transformation plan for the electric braking system of Pingban hydropower station, and also express the possible safety issues of the system after transformation.

## References

- [1] M. P. CALASAN, D. S. PETROVIC, M. M. OSTOJIC: *Electrical braking of synchronous generators for combined generator and water turbine bearings as well as stray-load losses determination*. IET Electric Power Applications 7 (2013), No. 4, 313–320.
- [2] D. J. MORÁN-ZENTENO, G. TOLSON, R. G. MARTÍNEZ-SERRANO, B. MARTINY, P. SCHAAF, G. SILVA-ROMO, C. MACIAS-ROMO, L. ALBA-ALDAVE, M. S. HERNANDEZ-BERNAL, G. N. SOLÍS-PICHARDO: *Tertiary arc-magmatism of the Sierra Madre del Sur, Mexico, and its transition to the volcanic activity of the Trans-Mexican Volcanic Belt*. Journal of South American Earth Sciences 12 (1999), No. 2, 513–535.
- [3] M. ENTEZAMI, S. HILLMANSEN, P. WESTON, M. PH. PAPAELIAS: *Fault detection and diagnosis within a wind turbine mechanical braking system using condition monitoring*. Renewable Energy 47 (2012), 175–182.
- [4] D. M. YU, Y. H. HU, D. WANG, X. J. LI: *Thermal analysis based on finite element analysis for motorized spindle*. Applied Mechanics and Materials 556–562 (2014), 1170–1173.
- [5] S. KHASTGIR, P. WARULE: *Regenerative braking strategy for an unaltered mechanical braking system of a conventional vehicle converted into a hybrid vehicle*. Symposium on International Automotive Technology (SIAT), 16–19 January 2013, Pune, Maharashtra, India, SAE Technical Papers 2013-26-0155.
- [6] H. YU, S. TAHERI, J. DUAN, Z. QI: *An integrated cooperative antilock braking control of regenerative and mechanical system for a hybrid electric vehicle based on intelligent tire*. Asian Journal of Control 18 (2016), No. 1, 55–68.
- [7] G. BING, G. JIE: *Research of electro-mechanical actuator in electric braking system based on instantaneous torque control with SRM*. Open Automation and Control Systems Journal (2014), No. 6, 519–527.
- [8] D. H. HU, R. HE, J. B. YU, Q. ZHAO: *Research of regenerative braking control strategy of electric vehicle based on electro-hydraulic hybrid brake system*. Journal of Highway and Transportation Research and Development 31 (2014), No. 3, 148–152.
- [9] A. TAVASOLI, M. NARAGHI: *An optimized multi-stage scheme to coordinate steering and braking*. Iranian Journal of Science & Technology Transactions of Mechanical Engineering 37 (2013), No. 2, 161–174.
- [10] S. KIM, J. KIM, G. SUNG, J. LEE: *Evaluation and development of improved braking*

- model for a motor-assisted vehicle using MATLAB/simulink*. Journal of Mechanical Science and Technology 29 (2015), No. 7, 2747–2754.
- [11] J. DU, H. ZHANG, H. WANG: *Braking energy recuperation performance of input coupled power-split hydraulic hybrid powertrain*. Transactions of the Chinese Society of Agricultural Engineering 29 (2013), No. 13, 23–30.
  - [12] B. LONG, S. T. LIM, Z. F. BAI, J. H. RYU, K. T. CHONG: *Energy management and control of electric vehicles, using hybrid power source in regenerative braking operation*. Energies 7 (2014), No. 7, 4300–4315.
  - [13] D. PURCARU, I. M. GORDAN, A. PURCARU: *Implementation of hydrogenerator over-current protection with minimum voltage blockage using embedded systems*. Journal of Electrical and Electronics Engineering 8 (2015), No. 2, 29–34.
  - [14] J. C. AKIROR, A. MERKHOUF, C. HUDON, P. PILLAY: *Consideration of design and operation on rotational flux density distributions in hydrogenerator stators*. IEEE Transactions on Energy Conversion 30 (2015), No. 4, 1585–1594.
  - [15] Y. LIANG, Y. WANG, Y. LIU, M. FU: *Derivation of temperature distribution of stator winding with transposed conductors for a large air-cooled hydrogenerator*. IEEE Transactions on Industrial Electronics 63 (2016), No. 8, 4764–4772.
  - [16] R. HIRATANI, I. MAEDA, T. KANEKO, : *Characteristics of on-line and off-line partial discharge for hydrogenerator by acoustic and electric detection method*. Acta Geologica Sinica 89 (2015), No. s1, 27–28.
  - [17] B. LONG, S. T. LIM, J. H. RYU, K. T. CHONG: *Energy-regenerative braking control of electric vehicles using three-phase brushless direct-current motors*. Energies 7, (2014), No. 1, 99–114.
  - [18] X. GONG, S. CHANG, L. JIANG, X. LI: *Braking method of electric vehicle based on direct drive electro-hydraulic brake unit*. Open Mechanical Engineering Journal (2015), No. 9, 351–360.
  - [19] C. ZONG, T. ZHU, C. WANG, H. LIU: *Multi-objective stability control algorithm of heavy tractor semi-trailer based on differential braking*. Chinese Journal of Mechanical Engineering 25 (2012), No. 1, 88–97.

Received June 29, 2017

# Application of electronic information technology in modern logistics system

LIAN DAI<sup>1</sup>, YULIN LAI<sup>1</sup>, MENGLIANG SHAO<sup>1</sup>

**Abstract.** With the rapid development of the financial market, in order to further explore the application of electronic information technology in modern logistics system and the importance and necessity of electronic information technology for modern logistics enterprises, the characteristics of modern logistics were summarized through the method of empirical analysis in this paper, and the necessity and importance of electronic information technology for logistics system were put forward. And how to apply several modes of electronic information technology to logistics enterprises was also discussed. Finally, some innovative suggestions and views on the adjustment of social environment, resources and industrial structure and the implementation of effective and practical electronic information means were put forward.

**Key words.** Electronic information technology, modern logistics system, MRP.

## 1. Introduction

Since twenty-first Century, with the continuous development of electronic information technology, electronic information technology has played a very important role in many fields, and its influence has been constantly expanding. There is no doubt that this has affected the development of market economy and the improvement of people's living standard to some extent [1]. Electronic information technology has also been applied to the field of logistics systems. Electronic information technology has brought a lot of benefits and convenience to the logistics system, which has improved the work efficiency of the logistics system, solved the problem of many traditional problems in the logistics system, and reduced the error that can't be avoided in the logistics system, thus reducing the economic losses to the managers of the logistics system and providing a safe, practical and reliable guarantee.

In the current environment, it is the peak period of globalization, global development is an inevitable trend of the international, and the development is very fast [2]. It is this background that leads to the globalization of electronic information technology. In the western developed countries, electronic information technology

---

<sup>1</sup>Guangzhou Nanyang Polytechnic, Guangzhou, Guangdong, 510925, China

is not only popularized in the logistics system, but also has more and more obvious advantages in the purchasing, selling and production links of enterprises. Therefore, the electronic information technology has a very potential in the new era of development background, which will bring better development space for the modern logistics system [3].

## 2. State of the art

In twenty-first Century, the level of development of science and technology can reflect the comprehensive competitive power of a country, which can also reflect the status of a country in the world. With the development of network, electronic information technology has also developed rapidly [4]. By analyzing the content and structure of the electronic information technology, some problems in the course of development can be found out. More specific and objective analysis can be carried out only after understanding the background of the re-development of electronic information technology.

In the course of the development of electronic information technology, the training of talents is lacking. Although in the present situation, more and more talented people emerge, talent who is more suitable for electronic information technology is very scarce, and the specific research is also very simple. Therefore, it is necessary to meet the needs of some complex talents [5]. In the development environment of the domestic electronic information market, there are all kinds of counterfeit and false environment, which is not conducive to the development of China's electronic information technology and the updating of modern logistics system [6]. Under the current development background, it is difficult to develop the electronic information technology without strict laws. Electronic information technology is in a period of rapid updating and development, and it is a new technology with long influence and potential, which can not only improve people's quality of life, but also can play a certain role in the modern logistics system [7]. The application of electronic information to broaden the knowledge of data can make the logistics system more comprehensive and complete.

## 3. Methodology

Electronic information technology can ensure the efficiency of the logistics system, and information, funds and other materials can be ordered distribution. The modern logistics system mainly includes the computer software technology, the network technology, the bar code technology, the radio frequency identification technology and so on [8]. The general understanding of electronic information technology mainly includes computer hardware technology and computer software technology. For the training of modern logistics enterprises, if there is lack of the protection of electronic information technology, there is no corresponding legal support, and then electronic information technology will be difficult to continue to develop, and the fair competition in the market will be broken up [9]. Electronic information tech-

nology has also been applied to the field of logistics systems. Electronic information technology has brought a lot of benefits and convenience to the logistics system, which has improved the work efficiency of the logistics system, solved the problem of many traditional problems in the logistics system and reduced the error that can't be avoided in the logistics system. Electronic information technology has been also widely used in modern logistics system, and the main technical means used in modern logistics system include the following two branches [10]. The block diagram of electronic information technology is depicted in Fig. 1.

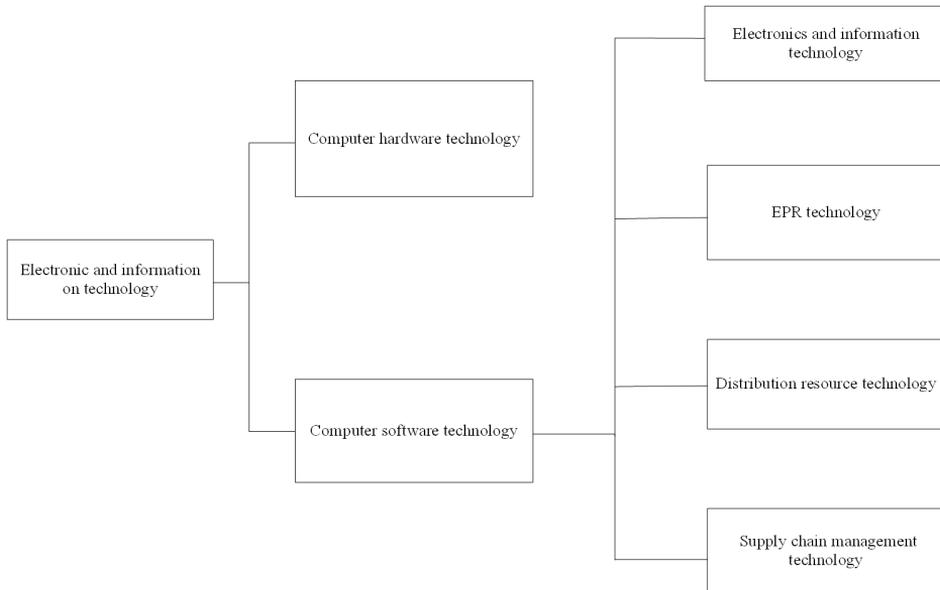


Fig. 1. Block diagram of electronic information technology

Computer hardware technology refers to a wide range of logistics systems in a basic technical means. It includes the traditional technology of computing hardware, such as operation technology and equipment technology, storage technology and data transmission technology and so on, which provides a certain guarantee through powerful and broad hardware facilities, and is widely used in modern logistics. Although computer hardware technology cannot determine the key to logistics management, it also affects the efficiency of its work. In computer software technology, a variety of software management and transportation management software and other technical means are needed [11]. The application of computer software technology mainly aims at providing technical support for the development of logistics activities in the logistics management system, and there are many branches of computer software technology classification.

The application of electronic information technology is very extensive, and different technologies play a far different role in the modern logistics system. Electronic information technology refers to a very important configuration technology in the

modern logistics system, which is mainly to plan and control the production and operation of enterprises. From the adjustment and control of the procurement of modern logistics enterprises, the product structure and procurement list can be understood, so as to achieve centralized logistics system effective management [12]. For the training of modern logistics enterprises, if there is lack of the protection of electronic information technology, there is no corresponding legal support, and then electronic information technology will be difficult to continue to develop, and the fair competition in the market will be broken up. Electronic information technology has been applied to the field of logistics systems. Electronic information technology has brought a lot of benefits and convenience to the logistics system, which has improved the work efficiency of the logistics system, solved the problem of many traditional problems in the logistics system and reduced the error that cannot be avoided in the logistics system. Enterprise resource planning technology, referred to as ERP technology, is a new technology implemented in the internal management of logistics enterprises, which uses modern electronic information technology to organize the concept, process and data analysis of logistics enterprise management, including manpower and material resources arrangement and so on. ERP technology is a new system which combines computer, electronic hardware and software. It not only greatly reduces the expenditure of logistics enterprises in management, but also improves the competitive ability of enterprises. The following table shows the comparison of several techniques.

Table 1. Comparison of several techniques

Variable	Value of ADF	Value of P	Inspection type
ln debt	-1.25	0.62	(c,0)
$D(\ln \text{ debt}(-1))$	-1.56	0.01	(c,0)
ln tax	-1.41	0.55	(c,0)
$D(\ln \text{ tax}(-1))$	-2.41	0.01	(c,0)
ln gov	-0.10	0.93	(c,0)
$D(\ln \text{ gov}(-1))$	0.53	0.02	(c,0)
ln gdp	-2.73	0.10	(c,0)
$D(\ln \text{ gov}(-1))$	1.90	0.10	(c,0)

Through comparison of several technologies, the application of modern logistics enterprises is not entirely suitable for modern logistics system, which still needs constant improvement to make greater influence on the logistics system. Through the investigation of various electronic information technologies, a new electronic information technology is found out, which is more perfect and more suitable for modern logistics enterprises, that is, supply chain management technology.

The emerging electronic information technology refers to the design planning and detailed control in logistics system, which can improve the competitiveness of logistics enterprises in the market economy, promote efficiency through the supply chain, and realize the purpose of increasing the efficiency of logistics enterprises.

In short, this technology is responsible for the management of sales orders in modern logistics enterprises. Through this foundation, the management ability between the enterprise and customer can increase continuously, and can make the customer leave a better impression to the enterprise [13]. The main targets are financial communication and customer interaction. At present, it has a strong competitive advantage, and it is one of the main new technologies of modern logistics system. And supply chain technology can ensure that each aspect can be carried out in an efficient and orderly way [14]. The above is the principle of supply chain technology, and with this principle, a small modern logistics system can greatly reduce labor costs, and save production costs, so that it is more suitable for most modern logistics enterprises [15]. The figure below shows how logistics enterprises manage warehousing goods through supply chain management techniques.



Fig. 2. How logistics enterprises manage warehousing goods through supply chain management techniques

## 4. Result analysis and discussion

### *4.1. Test of the influence of electronic information technology on logistics system*

Through the analysis, supply chain technology is more suitable for most modern logistics enterprises. Therefore, the empirical research on the enterprises which had already applied the supply chain technology was carried out. Taking supply chain technology as an example, the influence degree of electronic information technology on modern logistics was studied in this paper. The following table shows the global supply of supply chain technology in recent years, and the impact analysis of other data after using the supply chain of this electronic information technology.

In order to verify the feasibility and reliability of the above model, the relevant

data of logistics enterprises applying electronic information technology were statistically analyzed, and the empirical analysis was carried out. The details are shown in Table 2. The data was selected from years 1978–2014, so that there were in total 37 sets of data.

Table 2. Test on the impact of electronic information technology on logistics enterprises

Particular year	Electronics and information technology (Debt)	The increasing rate from electronics technology	Government expenditure (Gov)	GDP (Gdp)
1978	8058.1	2040.8	1303.6	9064.6
1979	8370.3	947.4	1107.7	7262.0
1980	8682.5	775.6	847.6	5998.5
1981	8994.7	700.0	713.6	5340.2
1982	9306.9	629.9	654.7	4896.0
1983	9619.1	571.7	637.4	4551.6
1984	9931.3	537.8	572.5	4067.7
1985	10243.5	519.3	480.0	3650.2
1986	10555.7	4255.3	5130.7	35524.3
1987	10867.9	3296.9	3933.9	27068.3
1988	11180.1	2990.2	3084.1	21895.5
1989	11492.3	2821.9	2576.1	18774.3
1990	11804.5	2727.4	2265.1	17090.3
1991	12116.7	2390.5	1901.7	15101.1
1992	12428.9	2140.4	1601.4	12102.2
1993	12741.1	2090.7	1446.0	10308.8
1994	13053.3	9262.8	12688.0	84883.7
1995	13365.5	8234.0	10930.4	79429.5
1996	13677.7	6909.8	9457.2	71572.3
1997	13989.9	6038.4	8152.8	61129.8
1998	14302.1	5126.9	6859.8	48459.6
1999	14614.3	10682.6	14767.0	88989.8
2000	17614.2	12581.5	16741.5	98562.2
2001	20614.0	15301.4	17908.4	108683.4
2002	23614.1	17636.5	19095.4	119765.0
2003	26614.5	20017.3	20297.7	135718.9
2004	29614.6	24165.7	22637.9	160289.7
2005	32614.2	28778.5	26371.8	184575.8
2006	35015.3	34804.4	30775.8	217246.6
2007	52074.7	45621.5	36645.4	268019.4
2008	53271.5	54223.8	42408.0	316751.7
2009	60237.7	59521.6	46432.1	345629.2
2010	67548.1	73210.8	53450.9	408903.0
2011	72044.5	89738.4	65047.2	484123.5
2012	77565.7	100614.3	73181.8	534123.0
2013	86746.9	110530.7	81245.9	588018.8
2014	95655.5	119175.3	86523.3	635910.0

As shown in the above Table 1, in the last five years, the electronic information technology had a great influence on the investment income and time cost of logistics enterprises. With the development of the market economy, the electronic information technology was positively related to the investment income of the logistics enterprises, and was negatively related to the time cost. After data collation, the data were checked by ADF. The result is shown below.

As can be seen from Fig. 3, after the first difference, all the roots were within the unit circle, which indicated that the time series was stable, and thus the stability of the model was determined. Finally, the pulse impact analysis was carried out, and the delivery error variable was used as the explanatory variable, and the remaining three variables were used as explanatory variables, so as to analyze the distribution error growth rate of the electronic information technology to the logistics enterprises and sample firms, and get the influence degree of the electronic information technology on the logistics enterprises. The result is shown in Fig. 4.

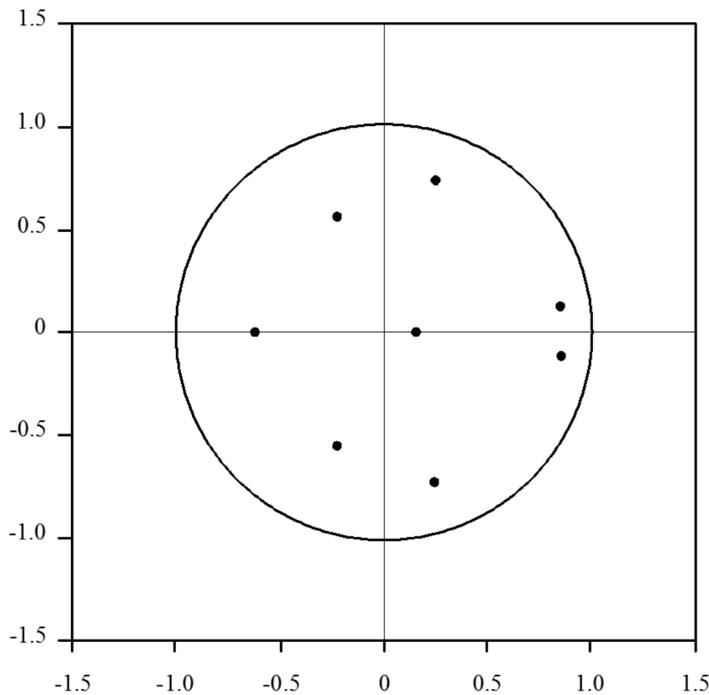


Fig. 3. Discrete rate of geographic information system

In China, there are many examples of the benefits of using electronic technology, such as Haier model - proprietary logistics system, and Haier logistics features can be summarized as the strength of the logistics company. In 1999, Haier began to implement business process reengineering with the "market chain" as a link, and took the order information flow as the center, so as to drive the logistics, business flow and capital flow operation, the logistics operation mode was increasingly at-

tracting people's attention. For Haier, the first is to achieve three "zero" goals: zero inventory, zero distance and zero working capital. The second is to win the core competitiveness in the market competition.

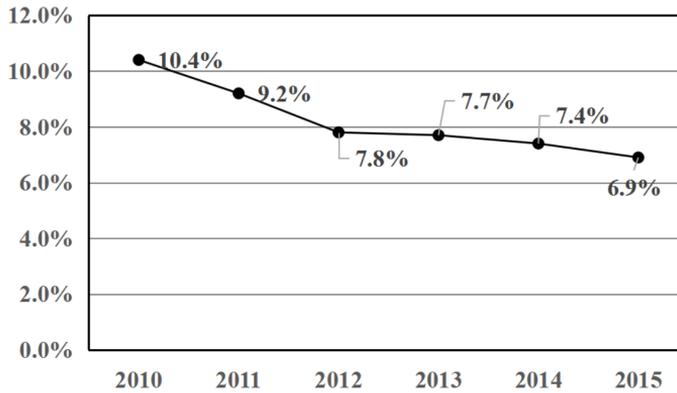


Fig. 4. Influence degree of electronic information technology on logistics enterprises

From the above detailed analysis, it is not difficult to draw a conclusion: with the development of time, the impact of new electronic technology on logistics enterprises has been playing a very important role, especially in the increase of investment income, the saving of enterprise time cost and the decrease of enterprise delivery error. However, in the application of specific electronic information technology, the actual situation of logistics enterprises should be analyzed, and we can't be anxious for success. For example, in the small scale of modern logistics system, the supply chain model of electronic information technology is more conducive to the development of logistics system. Electronic information technology is a necessary factor to improve the efficiency of modern logistics system, which can not only improve the comprehensive ability of the logistics system, but also can improve the competitiveness of logistics enterprises in the market economy, and bring more benefits for enterprises.

With the continuous progress and development of social economy, electronic information technology is bound to become a new technology means, and will also play an important role in modern logistics system. Electronic information technology has become one of the most popular technologies of the logistics industry. And its application is very extensive, including the cargo tracking, monitoring, system identification, and logistics management and warehousing and distribution links and so on. Especially in the modern logistics system, the container automatic identification function is the only one that can automate the tracking of cargo management in the current situation, which can identify all kinds of invoices, lists and labels of customers, reduce the utilization of labor, and greatly improve the efficiency of the management of enterprises. However, there are still many problems and some severe tests in the development of electronic information technology because of the immature development of electronic information technology, which should be solved in a planned way.

For the problems of the electronic information technology in the logistics enterprises, some suggestions can be through research: the social environment, resources and industrial structure should be adjusted, and the effective and practical electronic information means should be implemented. For example, the global positioning technology, RF technology and others can be used to improve the logistics system; the existing problems should be constantly summed up, so as to contribute to the development of electronic information technology and affect the efficiency of the modern logistics system.

## 5. Conclusion

In recent years, with the rapid development of the economic market, the electronic information technology has mushroomed. At present, the electronic information technology information has become an integral part of the modern logistics system, and the logistics system can't be separated from the operation of electronic information technology. Electronic information technology can ensure the efficiency of the logistics system, and information, funds and other materials can also be ordered distribution. The modern logistics system mainly includes the computer software technology, the network technology, the bar code technology, the radio frequency identification technology and so on. The characteristics of modern logistics were summed up through detailed data in this paper, and the necessity and importance of electronic information technology for logistics system were put forward. If the electronic information technology is integrated into the logistics system, and the logistics system and the electronic information technology are combined effectively, the modern logistics system can still have huge development space. In addition, how to further apply several modes of electronic information technology to logistics enterprises was also discussed. Through these patterns, the problems that should be paid attention to and the ways to solve them were expounded. Through the empirical research of the electronic information technology in the modern logistics system, it is concluded that the future development trend of the electronic information technology in the modern logistics system will certainly promote the global economy to be on the right track.

## References

- [1] X. Y. LI, H. WANG, X. N. WANG, W. ZHANG, Y. L. TANG: *The application analysis of cloud computation technology into electronic information system*. Applied Mechanics and Materials 556-562 (2014), Chapter No. 9, 5552-5555.
- [2] S. Y. TONG, J. WAN: *China's supply-side reform: The rationale and implications*. East Asian Policy 8 (2016), No. 4, 44-54.
- [3] S. LIMWATTANANON, S. NEELSEN, O. O'DONNELL: *Universal coverage with supply-side reform: The impact on medical expenditure risk and utilization in Thailand*. Journal of Public Economics 121 (2015), 79-94.
- [4] Y. C. LIU, G., X. YU, Y. Y. YAN: *Study on the development status and countermeasure of Xinjiang jujube industry*. Northern Horticulture 391 (2013), No. 18, 91-99.

- [5] D. B. BERLIN, M. J. DAVIDSON, F. J. SCHOEN: *The power of disruptive technological innovation: Transcatheter aortic valve implantation*. Journal of Biomedical Materials Research Part B: Applied Biomaterials 103 (2015), No. 8, 1709–1715.
- [6] B. J. ZHANG, M. B. YI, J. F. SONG, D. S. GAO, N. H. ZHU, R. H. WU, W. WANG: *High-speed and high-power 1.3  $\mu$ m InGaAsP/InP selective proton-bombarded buried crescent lasers with optical field attenuation regions*. Japanese Journal of Applied Physics Part 1: Regular Papers short notes and Review papers 38 (1999), No. 12A, 6729–6731.
- [7] Y. H. ZHAO, H. M. LI, L. F. QIN, H. H. WANG, G. Q. CHEN: *Disruption of the polyhydroxyalkanoate synthase gene in aeromonas hydrophila reduces its survival ability under stress conditions*. FEMS Microbiology Letters 276 (2007), No. 1, 34–41.
- [8] M. CROWLEY, D. TOPE, L. J. CHAMBERLAIN, R. HODSON: *Neo-taylorism at work: Occupational change in the post-fordist era*. Social Problems 57 (2010), No. 3, 421–447.
- [9] X. Y. LIU: *Design of logistics information system based on RFID technology*. Applied Mechanics and Materials 608–609 (2014), Chapter No. 3, 343–346.
- [10] M. EXON-TAYLOR: *Beyond MRP: The development of a modern scheduling system*. Logistics Information Management 8, (1995), No. 1, 17–23.
- [11] T. BHAVAN, C. XU, C. ZHONG: *Growth effect of aid and its volatility: An individual country study in South Asian economies*. Business and Economic Horizons 3 (2010), No. 3, 1–9.
- [12] J. HUDSON: *Consequences of aid volatility for macroeconomic management and aid effectiveness*. World Development 69 (2015), 62–74.
- [13] S. HAO, Y. NA: *More focus on the macro-economic management of the supply side: An active adjustment adapting to the change of international and domestic situations*. Contemporary Economy & Management 37 (2015), No. 4, 8997–9006.
- [14] E. SHNAIDER, N. HARUVY, A. YOSEF: *Do macro-economic factors influence financial management decision making? The Israeli economy in perspective*. Journal of Financial Management and Analysis 28 (2015), No. 2, 64.
- [15] G. SPOLANDER, L. ENGELBRECHT, A. P. SANSFAÇON: *Social work and macro-economic neoliberalism: Beyond the social justice rhetoric*. European Journal of Social Work 19 (2016), No. 5, 634–649.

Received June 29, 2017

# Field programming technology for instrument sensor's correction parameter

JIAN YANG<sup>1</sup>, LINBIN WU<sup>1,2</sup>

**Abstract.** In the process of industrialization, a large number of mechanical equipment has been gradually applied to the development of various industries. The field programming technology for instrument sensor's correction parameter has a very important impact on the maintenance of equipment. In order to better improve this technology, first of all, the related concepts of field programming technology for instrument sensor's correction parameter were clarified through the reading of relevant data in this paper. Then, the temperature sensor was used as an example to further verify that the field programming technique for instrument sensor's correction parameter is obviously superior to traditional correction technique in sensor repair. The purpose of this research is to provide a theoretical basis and scientific support for the development of China's machinery industry.

**Key words.** Instrument and meter, sensor, correction parameter, field programming technology.

## 1. Introduction

With the development of the times, the world's economic level has developed to a great extent in the current era. Especially with the new era theme of peace and development, all sectors of the world have made progress and promotion. In some machinery manufacturing industry, its development has gradually become the basis for the development of various fields of the times. The development of the machinery manufacturing industry has provided certain impetus and positive influence for each profession's development. In the development of machinery industry, the application of some equipment and the application of instrument sensor have been gradually generalized. Instrument sensors play an important role in many industries, which are very important to normalize the running state of some equipment and monitor the whole operation process of the equipment. However, in the traditional mechan-

---

<sup>1</sup>Key Laboratory of Earthquake Geodesy, Institute of Seismology, CEA, Wuhan 430071, China

<sup>2</sup>Corresponding author

ical manufacturing industry, because there is a big error in the making process of instrument sensor, which causes a certain negative impact on the normal operation and maintenance of related industries and equipment. With the rapid development of science and technology, more innovative science and techniques have been created and gradually applied to various industries, and they have brought some positive impetus for the development of these industries. As one of the important science and technology, computer technology plays an important role in the development and progress of various industries. And in this trend, the computer industry has gradually combined with various traditional industries, which have brought about certain results for the development of these industries. The purpose of this research is to analyze the influence of field programming technology on the instrument sensor's correction parameter, so as to provide a reference for the development and theoretical improvement of instrument sensing technology in our country.

## 2. State of the art

Since the industrial revolution, the present era has begun to develop in the process of industrialization. In this trend, more and more mechanical equipment has been gradually applied to people's production and life. The application of the mechanical equipment is very important to people's production and life [1]. However, as people begin to pursue more efficient production and the cost control, people are required to use more sophisticated and complicated equipment in the process of production. With the continuous complication of the mechanical equipment, the relevant technicians are required to monitor the equipment more accurately and grasp the operation state of the whole machine in real time. Only in this way can we better carry out timely maintenance of mechanical equipment, so as to improve the efficiency of the entire production line and reduce the operating cost directly or indirectly [2]. On this basis, the application of instrument sensors is very important for the related technicians to monitor their mechanical equipment. Many researchers have begun to study the precision of instrument sensors for further improvement, and combined computer technology with the traditional instrument sensors. This has brought about a very important impact on the normal operation of the related machinery, and further reduced the cost of industrial operation, thus making the comprehensive economic level of the world greatly improved [3].

## 3. Methodology

The economic development of the current era has begun to show a gradual upward trend. Under the background of this great era and the trend of economic development, the economic level of our country has also been affected to some extent. China's various industries have had a certain degree of progress and development, and the overall strength of our country has begun to be further improved. Especially since the reform and opening to the outside world, China has gradually entered the industrialization era. More enterprises have begun to use a large number of mechani-

cal equipment for the production of enterprises. The mechanical equipment not only has liberated the hands of the people, but also greatly improved China's economic level because of higher production efficiency and lower production cost [4]. However, in the use process of mechanical equipment, further maintenance is needed, which can better ensure the normal operation of the machine and more accurately grasp the relevant data for information of the operation process. Therefore, the instrument sensors are gradually used in various mechanical operations, which provide a certain technical support for the efficient operation of the entire production line [5].

Nowadays, the manufacturing industry in our country has begun to increase gradually, and the use of different instrument sensors has gradually increased, and instrument sensors have begun to diversify (see Fig. 1). Because all kinds of data in the development of the industry have been stored, a large number of correction parameters have also been introduced into the instrument sensors in China to make the data more accurate. These correction parameters include the temperature parameters and correlation parameters of mechanical sensitivity. And some sensors have many nonlinear correction parameters in operation, which is very important for the further improvement and normal operation of instrument sensors [6]. However, there are many shortcomings and defects in our instruments and meters. In our country, the traditional way is usually used for the change of the instrument sensor's correction parameters. But in the repair of traditional instrument and meter equipment, the outer shell of the instrument sensor is removed and the CPU and some related components are repaired. Furthermore, it needs to rewrite the program for the storage device that stores the data, so as to restore its true long running performance [1]. Therefore, when the instrument sensors are damaged to some extent, the traditional repair methods may achieve the ultimate goal of restoration. However, the steps involved in this repair process are more complicated. Because only targeted programs are rewritten for the wrong program, the program may not be rewritten to match the previous program, and the operation process of the whole instrument sensor is restricted to a certain extent [7]. Therefore, the traditional repair technology for instrument sensor may make the reliable performance of the whole mechanical equipment gradually decreased, and further affect the development of related industries in China. However, with the rapid development of computer technology, the technology has more information sharing. Through the preparation of related procedures, it can effectively realize the accurate operation of related instruments and programs. Different from the traditional human perception, this technique can be used to evaluate the performance of related machines more objectively [8]. Our country has gradually introduced this kind of technology and related theories into the development of related industries in our country, and has made certain achievements. In this paper, based on the present situation, the shortcomings and defects of the correction factor in the development of traditional sensor technology in our country were pointed out. In addition, through the analysis and application of the field programming technology for instrument sensor's correction parameters, the research aimed at providing a theoretical basis and scientific support for the development of related industries in China.

First of all, through the reading and summary of relevant information, the relative



Fig. 1. Application of instrument sensor

concepts and advantages of field programming techniques for instrument sensor's correction parameters were analyzed to further make the relevant concepts clearer [9]. On this basis, the related properties of the main control element EEPROM chip of field operation technology were analyzed [10]. The relevant properties of the chip are shown in Table 1.

The main research object of this research was the correction parameter of temperature sensors in common instruments. Firstly, the relevant attributes of the main control element EEPROM chip were assigned. Then, the nonlinear compensation technology model was introduced to compare and analyze the related characteristics before and after the application of field programming technology [11]. The related model for nonlinear compensation techniques is shown below:

$$P'(U) = a_0 + a_1U + a_2U^2 + \cdots + a_nU^n, \quad (1)$$

or

$$P'(U_i) = \sum_{j=1}^n a_j U_i^j \quad (2)$$

$$\sigma = \sum_{i=0}^m [P'(U_i) - P_i]^2 = \sum_{i=0}^m \left( \sum_{j=0}^m a_j U_i^j - P_i \right)^2. \quad (3)$$

Finally, through the investigation of related technicians, the importance of this correction parameters technique to sensor repair in our country was confirmed.

#### 4. Result analysis and discussion

This study first read and summarized the related researches, and then summarized the related concepts of field programming techniques for instrument sensor's correction parameters. Some researchers believe that in field programming techniques, the interface of the sensor is connected with the interface of computer technology. Then, when the sensor operation program is found to be insufficient, the

sensor can be repaired by simply knocking the computer keyboard. The main features of field programming technology are reflected in that when the correlation parameters of the sensor is repaired, it breaks the single way that relies on manual detection to repair in traditional technology. In addition, the system realizes the monitoring of related error data relying on the transmission between computers.

Table 1. Summary of the related attributes of the main control element EEPROM chip

Attribute	Interpretative statement	No.
CommPort	Serial communication, that is, string slogan;	X1
Setting	Serial number settings, mainly consists of baud rate, parity, data bits and stop bits, in the middle with a comma interval;	X2
InBufferSize	Gets or set the current receive buffer size;	X3
InBufferCount	Get the number of bytes in the current receive buffer;	X4
InputLen	Determine the number of bytes read by the nInput property, which reads the full buffer when 0;	X5
InputMode	0 indicates receipt of text type, and 1 indicates binary reception;	X6
OutBufferSize	Set or get the current send buffer size;	X7
OutBufferCount	Get the number of bytes in the current send buffer;	X8
Rthreshold	Set the number of characters in the receive buffer to trigger the OnCom event, and 0 indicates the non-triggering event;	X9
Sthreshold	Set the number of characters in the send buffer to trigger the OnCom event, and 0 indicates the non-triggering event;	X10
OutBuffersize	Set or get the current send buffer size.	X11

In order to better ensure the correct operation of the entire repair process and the work of the sensors, the computer operating circuit and the overall monitoring system of the field programming technology of the instrument sensor's correction parameters are further improved [12]. In the related circuit setting, in order to obtain better performance of instrument sensor and store the monitoring data, the position of the sensor correction factor must be better controlled, such as putting it in the EEPROM chip. Only in this way can the relevant data information stored by these sensors be not lost. In the current instrument, the related chip devices are mainly composed of serial chips and parallel chips. And the two chip technologies have their own advantages and disadvantages in sensor repair. The differences between them are mainly in terms of the transmission speed between sensors and computers, the occupation of related resources and the formation of related circuits. Therefore, in order to operate and correct the correction factor of the instrument sensor more accurately, it is necessary to generalize and clarify the related theory of the circuit. On this basis, through the actual repair needs, the line is further selected accurately,

and then the programming technique of the download path of the sensor correction factor is analyzed [13]. In the process of programming, we mainly analyze the ways of obtaining data, the data transformation related systems and data transmission ways, and determine the relevant programming commands. After the commands are entered, the computer passes the command through the interface to the sensor. The instrument sensors that receive the relevant instructions will automatically detect their temperature, sensitivity and other related indicators, and further analyze the detected data comprehensively. The analyzed data is transmitted to the computer terminal via the interface to the computer. And through the display of the computer model, people can determine the current state of the relevant mechanical equipment.

Therefore, through the connection of computer system and sensor and the use of related programming technology, the real-time monitoring of mechanical equipment is realized. Compared with the traditional sensor correction parameters technology, the field programming technology of the instrument sensor's correction parameters can be free from human subjective factors. And because it can detect real-time automatic completion of the relevant indicators, more information can be applied to the later analysis of machine running status, which makes the analysis accuracy higher and analysis results more credibility. Under these advantages, the field programming technology is gradually combined with the sensor correction factor, and applied to the actual operation and monitoring of mechanical equipment. It has a positive impact on the maintenance of mechanical equipment, and provides technical support for enterprise operation and reduces efficiency and operation cost further, so as to provide a reliable basis for the comprehensive promotion of the world economic level (see Fig. 2). In this study, the field programming technique of the instrument sensor's correction parameters was compared with the traditional sensor correction coefficient technique [14]. The results are shown in Table 2.

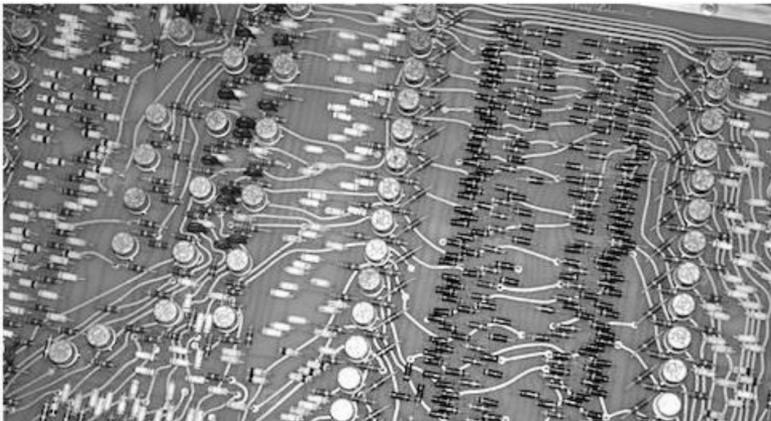


Fig. 2. Development of field programming techniques for instrument sensor's correction parameters

Table 2. Advantages of field programming techniques for instrument sensor's correction parameters

	Field programming techniques for instrument sensor's correction parameters	Correction factor technique of traditional sensors
Accuracy	High precision, multiple functions automatically	Low precision, relatively simple function
Reliability	It can automatically compensate for the drift of the system characteristics due to changes in operating conditions and environmental parameters, and the reliability is higher.	Relatively low stability
Stability	Compensate for the drift of the temperature and the drift of sensitivity, automatically change the range, and can carry out the self-test of the system, and the stability is higher.	Relatively low stability
SNR	The technology has the functions of data storage, memory and information processing, and can eliminate the noise in the input data by digital filtering, so it has a relatively high SNR.	Relatively high SNR
Resolution	This technique can eliminate the influence of complex sensitivity in multi-parameter state by data fusion technology, so as to ensure high resolution for measuring specific parameters.	Relatively low resolution
Adaptability	The technique has the function of judgment, analysis and processing in the process of information data transmission, which makes the system work in the optimized low power state and the optimized transfer rate.	Relatively poor impact

Taking the correction parameters of the common instrument temperature sensor as the example of this research, the various indexes of its central space were designed as shown in Table 3. Before and after the application of this technique, the temperatures of the sensors were measured. The result is shown in Fig. 3. The results show that this technique can effectively improve the performance of sensors.

Table 3. Design of the relevant attributes of the main control element EEPROM chip

Attribute	X1	X2	X3	X4	X5	X6
Set-point	2	9600,n,8,1	1024	1	1	1
Attribute	X7	X8	X9	X10	X11	
Set-point	1024	1	1	0	1024	

Finally, through the questionnaire, the correlation between field programming technology and traditional technology for temperature sensor was compared. The results are shown in Fig. 4.

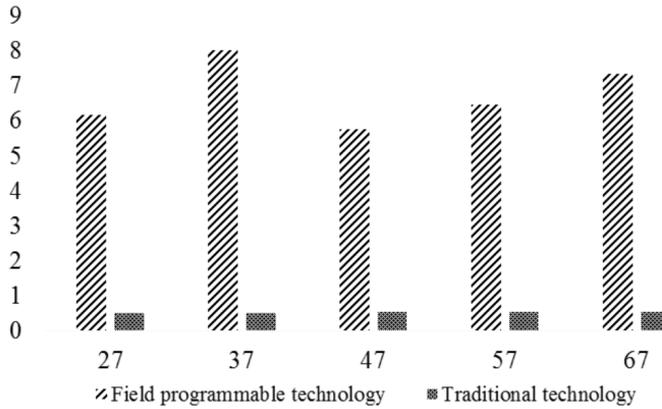


Fig. 3. Comparison of field programming technology and traditional technology for temperature sensor

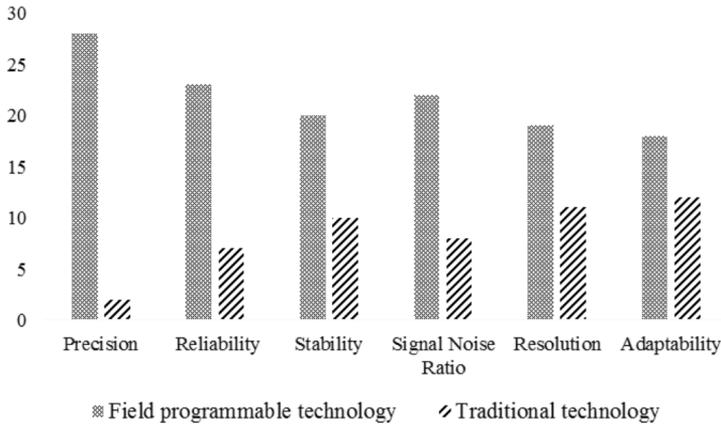


Fig. 4. Comparative analysis of field programming technology and traditional technology for temperature sensor

### 5. Conclusion

With the development of the times, nowadays, our society is developing towards industrialization. The extensive use of machinery and equipment is one of the most important forms of industrialization. In the background of this era, China has introduced a large number of industrial equipment in various industries. In the process of production, the relevant running equipment is further controlled by its operation and working state. Only in this way can equipment maintenance be realized. In addition, the industrial production cost is reduced, so as to further enhance the industry's economic development level. It is very important for instrument sensor to monitor the mechanical equipment. With the development of the times, the field

programming technology of the instrument sensor's correction parameters is very important for the application of instrument sensors. In this paper, the comprehensive analysis and the discussion of the related theory concept were firstly carried out. On this basis, the strength of the temperature sensor was explained. The purpose of this research is to provide theoretical basis for the development and theoretical improvement of our instrument sensor, and provide a reference for the continuous improvement of the comprehensive economic level in all sectors of our country.

## References

- [1] S. M. SAMBORSKI, N. TREMBLAY, E. FALLON: *Strategies to make use of plant sensors-based diagnostic information for nitrogen recommendations*. American Society of Agronomy Journal *101* (2008), No. 4, 800–816.
- [2] M. H. SCHULZE, H. HEUER, M. KÜTTNER, N. MEYENDORF: *High-resolution eddy current sensor system for quality assessment of carbon fiber materials*. Microsystem Technologies *16* (2010), No. 5, 791–797.
- [3] P. S. WAGGONER, H. G. CRAIGHEAD: *The relationship between material properties, device design, and the sensitivity of resonant mechanical sensors*. Journal of Applied Physics *105* (2009), No. 5, 054306–054314.
- [4] A. A. ZVYAGIN, D. V. NENAKHOV, S. N. KORCHAGINA, A. V. SHAPOSHNIK, V. V. KOTOV, V. A. YUKISH: *Determination of ammonia in the air using piezoelectric resonance sensors coated with humic acids*. Journal of Analytical Chemistry *65* (2010), No. 4, 414–417.
- [5] P. SREESHYLAM, K. RAVISANKAR, S. PARIVALLAL, K. AND S. SRIDHAR: *Condition monitoring of prestressed concrete structures using vibrating wire sensors*. International Journal of COMADEM *11* (2008), No. 3, 46–54.
- [6] F. STEFANI, D. MACII, A. MOSCHITTA, P. CARBONE, D. PETRI: *A simple and time-effective procedure for ADC INL estimation*. IEEE Instrumentation and Measurement Technology Conference Proceedings, 16–19 May 2005, Ottawa, Ont., Canada, IEEE Conference Publications *2* (2005) 1027–1032.
- [7] F. C. GÓMEZ DE LEÓN, P. A. MEROÑO-PÉREZ: *Discrete time interval measurement system: fundamentals, resolution and errors in the measurement of angular vibrations*. Measurement Science and Technology *21* (2010), No. 7, paper 075101.
- [8] Y. SONG, D. ZENG, Y. TIAN: *Accurate time interval measurement method based on vernier caliper principle*. International Conference on Image Analysis and Signal Processing, 9–11 April 2010, Zhejiang, China, IEEE Conference Publications (2010), 553–555.
- [9] M. ZIELIŃSKI, M. KOWALSKI: *Review of single-stage time-interval measurement modules implemented in FPGA devices*. Metrology and Measurement Systems *16*, (2009), No. 4, 641–647.
- [10] R. SALOMON, R. JOOST: *BOUNCE: A new high-resolution time-interval measurement architecture*. IEEE Embedded Systems Letters *1* (2009), No. 2, 56–59.
- [11] T. NAKAGAWA, R. FUJIWARA, G. ONO, M. MIYAZAKI: *UWB-IR receiver with accurate time-interval-measurement circuit for communication/location system*. IEEE International Symposium on Circuits and Systems, 24–27 May 2009, Taipei, Taiwan, IEEE Conference Publications (2009), 397–400.
- [12] P. PANEK: *Time-interval measurement based on SAW filter excitation*. IEEE Transactions on Instrumentation and Measurement *57* (2008), No. 11, 2582–2588.
- [13] S. SUGIYAMA, M. A. FARIZUL, H. TANOUÉ, Y. SUDA, H. TAKIKAWA, S. OKE, K. KAWASHIMA: *PV-array pyranometer with 4 sensors in greenhouse and study on temperature correction [in Japanese]*. Journal of Japan Solar Energy Society *37* (2011), No. 3, 55–61.

- [14] P. PANEK, I. PROCHAZKA: *Time interval measurement device based on surface acoustic wave filter excitation, providing 1 ps precision and stability*. Review of Scientific Instruments 78 (2007), No. 9, paper 094701.

Received June 29, 2017

# Integration management of strategic supply chain based on genetic algorithm

ZIHUI ZHENG<sup>1,2,3</sup>, XIAOHU ZHOU<sup>1</sup>

**Abstract.** Strategic supply chain research needs to take into account the three major stages of supply chain, namely, procurement, production, distribution, and their interactions. In order to study the limitations of the existing supply chain design model and establish a strategic supply chain integrated management system which is suitable for different customer needs, facilities matching relationship and supplier priority, a hybrid integer nonlinear programming (MINLP) model was established. Two-step method was used to solve the constraint problem in the model, and the adaptive genetic algorithm (AGA) was used to solve the model. Finally, the experimental results show that the proposed mixed integer nonlinear programming model can effectively solve the problem of supply chain coordination in strategic supply chain design, and can get a better supply chain design.

**Key words.** Strategic supply chain design model, mixed integer nonlinear programming, adaptive genetic algorithm.

## 1. Introduction

As China's bicycle export market scale increases every year, China has become the world's largest bicycle manufacturer. According to statistics, China produced 80.26 million bicycles and 32.57 million automatic bicycles in 2015. Large bike manufacturers and parts suppliers form a strategic logistics partnership while producing the bicycles. The expansion of the bicycle manufacturer examines its ability to generate energy, but also puts pressure on the supplier's ability to supply. As the environment changes, if bicycle manufacturers continue their relatively inflexible supply mechanism, this will greatly reduce customer confidence and loyalty. Nowadays, the key for the success or failure of an enterprise is the ability to coordinate the

---

<sup>1</sup>School of Economics & Management, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, China

<sup>2</sup>Haier Group, Qingdao, Shandong, 266101, China

<sup>3</sup>Corresponding author

complex network relationships between supply chain facilities and the integration of integrated management. Therefore, it is very important for the bicycle company to establish a reliable and effective supply chain cooperation alliance and continuously optimize it. The theoretical significance of this paper lies in optimizing the comprehensive configuration of the strategic and operational levels of the supply chain, namely, considering the product parts procurement, product design, production and transportation and other aspects of the products, so as to provide a new research perspective for bicycle manufacturers and improve the competitiveness of the supply chain.

## 2. State of the art

Supply chain design model has important significance, and it has been more concerned by domestic and foreign scholars. For example, in the design of low-carbon supply chain network, the least sum model is built among logistics costs, distribution centers, fixed costs, and carbon emissions costs through concave functions, the Lagrangian relaxation method is used to decompose the model into vest pack problems and a single supplier location problem and analyze supply chain network design and operating costs, so that the number of distribution centers based on the cost of carbon emissions can be obtained, and the distribution center utilization can be improved [1]. Closed-loop supply chain network design uses non-linear cost and continuous variable. Functional activity arcs and variational inequalities are used to design cost functions and build models to analyze the impact of recycling rates on sales, operations and recovery costs, design costs, profits, and demand for market demand, recovery and value [2]. Based on the general closed-loop supply chain, a multi-objective mixed integer linear programming model can be proposed to optimize the supply chain network. The model not only determines the number of products and nodes in the closed-loop supply chain network, but also selects the best suppliers and manufacturers [3]. The utility software system (GAMS) is used to solve and obtain the corresponding optimization scheme. But the GAMS can only solve small-scale data problems. In order to ensure the stable performance of the supply chain when the design parameters are perturbed, a robust optimization supply chain design model is established from the upstream selection supplier to the downstream location and distribution requirements, and the method of determining the value of the regret value is proposed, and the tabu search algorithm of supply chain node configuration is designed. Robust optimization can effectively avoid investment risk [4]. Supply chain involves a number of equipment such as raw materials and parts suppliers, the final product producers, distribution centers. According to the requirements of the customers, the supply chain members are designed and integrated, the strategic supply chain design integration model is proposed to enable collaborative scheduling of production and product offerings to ultimately achieve optimal supply chain performance [5]. Aiming at the existence of multiple nodes in the recycling and remanufacturing supply chain network, the equilibrium variational inequality model with both positive and reverse logistics is established. The advantage of this model is that it qualitatively and quantitatively analyzes the

competition behaviors among several manufacturers and among retailers, the supply chain competitiveness is determined by the overall efficiency of the members in the chain [6].

### 3. Methodology

A set of products is designed and manufactured via a supplier, a manufacturer, a distribution center (DC), and ultimately a customer base, as shown in Fig. 1, which form a supply chain. The manufacturer is responsible for the overall design and production of the product and the supplier provides the required parts or the design and production of the intermediate part according to the needs of the manufacturer [7]. Each production manufacturer completes product design, production and sales according to customer requirements. However, due to the manufacturer's own production conditions, it is impossible for a single manufacturer to produce enough products to meet market demand, so different manufacturers are required to produce the same product together [8]. The current global manufacturing practice is that a number of manufacturers of a product belong to the same enterprise, while suppliers and distribution centers (DC) do not belong to the same enterprise. In addition, the decisions made by the enterprise on suppliers, manufacturing manufacturers and distribution centers are within the scope of the supply chain. It can be seen from the above that the supply chain is the best choice when dealing with international companies [9].

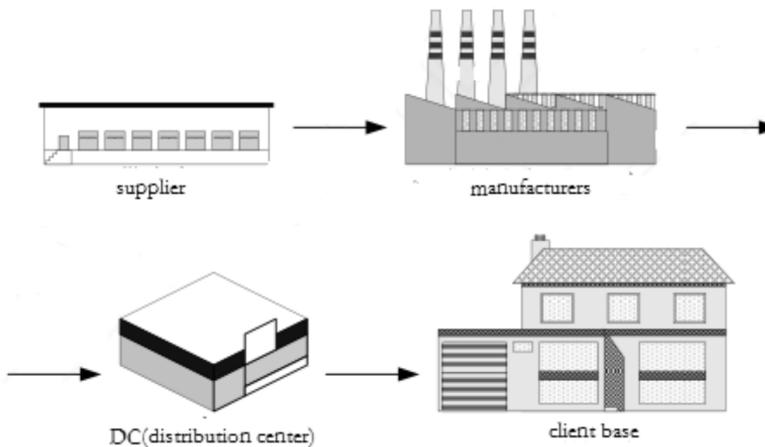


Fig. 1. Schematic diagram of product path

The model is based on the following assumptions: each manufacturer can produce any product; the product within the manufacturer is divided into a limited set of sub-tasks, that is, each product is supplied by one or more suppliers; if a supplier assumes the task of designing a component, it is responsible for the production of the part; each manufacturer procures parts from multiple suppliers, each of which serves

a number of manufacturers; each distribution center is open to any product; the distribution center transports all the products received to the customer; according to the location, the customer is divided into different regions to form a customer base; the design and production capacity of the supplier and the manufacturer is known; the capacity of the distribution center is known; and the transportation costs between the facilities are known [10].

The objective function  $Q$  is the minimum value of the total cost of the whole supply chain period (yuan/period), as shown in equations (1)–(4):

$$\text{Min}C = f_1 + f_2 + f_3, \quad (1)$$

$$f_1 = \sum_{rj} U_{rj} \frac{\eta_{rj}}{\delta_{rj}} + \sum_{rj} U_{rj} f_{rj}^S + \sum_{rj} U_{rj} p_{rj}^S Z_{rj} + \sum_{rjk} U_{rj} c_{rjk}^{\text{SP}} R_{rjk}, \quad (2)$$

$$f_2 = \sum_{ik} V_{ik} \frac{\eta_{ik}}{\delta_{ik}} + \sum_{ik} V_{ik} f_{ik}^P + \sum_{ik} U_{ik} p_{ik}^P X_{ik} + \sum_{ikl} V_{ik} c_{ikl}^{\text{PD}} Q_{ikl}, \quad (3)$$

$$f_3 = \sum_{il} W_{il} f_{il}^D + \sum_{il} W_{il} p_{il}^D Y_{il} + \sum_{ilm} W_{il} c_{ilm}^{\text{DZ}} P_{ilm}. \quad (4)$$

In the above equations,  $i$  is the set of products provided to the customer,  $j$  is the candidate supplier set,  $k$  is the set of potential manufacturers,  $l$  is the distribution center set,  $m$  is the customer group set and  $r$  is the set of parts provided to the supplier. Symbol  $f_{rj}^S$  is the production preparation cost of supplier  $j$  on production parts  $r$ ,  $f_{ik}^P$  is the production preparation cost of supplier  $k$  on production parts  $i$ ,  $f_{il}^D$  is the setup cost at which the distribution center  $l$  delivers the product  $i$ . Symbol  $\delta_{ik}$  is the capacity of the manufacturer  $k$  to design the product  $i$ ,  $\delta_{rj}$  is the capacity of the supplier  $j$  to design the component  $r$ ,  $c_{rjk}^{\text{SP}}$  is the unit transportation cost (yuan/unit) from supplier  $j$  to manufacturer  $k$  to transport part  $r$ ,  $c_{ikl}^{\text{PD}}$  is the unit transportation cost (yuan/unit) from the manufacturer  $k$  to the distribution center  $l$  to transport product  $i$ ,  $c_{ilm}^{\text{DZ}}$  is the unit transportation cost (yuan/unit) from the distribution center  $l$  to the customer base  $m$  to transport products  $i$ ,  $p_{rj}^S$  is the unit purchase cost (yuan/unit) of part  $r$  from the supplier  $j$ ,  $p_{ik}^P$  is the unit production cost (yuan/unit) of manufacturer  $k$  to produce product  $i$ ,  $p_{il}^D$  is the unit cost of the capacity of the product  $i$  at the distribution center  $l$  (i.e. the cost of processing and inventory) (yuan/unit). Symbol  $\eta_{rj}$  is the cost coefficient of the supplier  $j$  to design component  $r$ ,  $\eta_{ik}$  is the cost coefficient of the supplier  $k$  to design component  $i$ ,  $D_{im}$  is the average demand (unit/period) of product  $i$  in customer group  $m$ ,  $X_{ik}$  is the quantity (unit/period) of product  $i$  manufactured by manufacturer  $k$ ,  $Y_{il}$  is the quantity (unit/period) of the product  $i$  received by the distribution center  $l$ ,  $Z_{rj}$  is the quantity (unit/period) of the component  $r$  provided by the supplier  $j$ ,  $Q_{ikl}$  is the quantity (unit/period) of the product  $i$  being transported from the manufacturer  $k$  to the distribution center  $l$ ,  $R_{rjk}$  is the quantity (unit/period) of the part  $r$  being transported from the supplier  $j$  to the manufacturer  $k$ . Symbol  $P_{ilm}$  is the quantity (unit/period) of the product  $i$  being transported from the distribution center  $l$  to

the customer base  $m$ ,  $U_{rj}$ : if the supplier  $j$  provides components with the  $r$  value, it is equal to 1, otherwise it is 0,  $V_{ik}$ : if the manufacturer  $k$  produces the product  $i$ , its value is 1, otherwise 0,  $W_{il}$ : if the distribution center  $l$  accepts the product  $i$ , its value is 1, otherwise it is 0.

The total cost in the objective function includes fixed and variable. It contains the three stages costs of the supply chain: procurement, production and transportation phase. In the procurement phase, the cost includes parts design, production preparation, parts procurement and transportation costs [11]. Among them, the design cost comes from the ability of the supplier to design a specific component. At the production stage, the cost includes product design, production preparation, production up costs, and transportation costs from the manufacturer to the distribution center. During the transportation phase, the cost includes the cost of the equipment, the inventory costs of the distribution center, and the transportation costs from the distribution center to the customer base [12].

Constraint processing is a very important problem in model solving, and a two-step approach is proposed to deal with constraints. In the first step, constraints are divided into two types: ability-related constraints and priority-related constraints. Competence-related constraints are the constraints of the entity in terms of production capacity, set-up costs, and distribution center capacity. Priority related constraints are thinking about solutions from a holistic perspective, so there are bill of materials constraints in the model, supplier preference and facility pairing. In other words, the ability-related constraints have an effect on the entity itself, and the priority-related constraints affect the structure of the supply chain [13]. In the second step, the constraint processing module has a screening function in the process of solving the solution. Initially, the constraints are stored in a collection. Whenever a new solution is generated, the constraint processing module checks the constraints in the collection from the aspects of valid a priori constraints. In the process of solving, it is necessary to include viable and unfeasible solutions in the interactive search process, so as to promote diversified development [14]. In this study, different penalty factors are assigned to varying degrees of competency-related constraints and prior correlation constraints. As shown in Fig.2, there is a pair relationship between supply chain facilities.

The solution corresponding to a supply chain can be divided into several segments. Each fragment represents a supply chain stage, such as supply, production and distribution, which is mainly composed of two attributes: location and value. The location represents the corresponding index in the different fragments, which can describe the state or priority. As shown in Fig. 3, among the suppliers, suppliers 2 and 4 are selected; among the manufacturers, 1, 2 are selected; and in the transit center, 2, 3 are selected. There are three segments in the second paragraph, each of them represents a stage. The three segments correspond to the product from the supplier to the manufacturer, the manufacturer to the transit center, and the transit center to the customer base.

The combination of roulette and "elite" strategies is used to select the operator, the top 20% of the individual is retained, so that they do not cross and mutate.

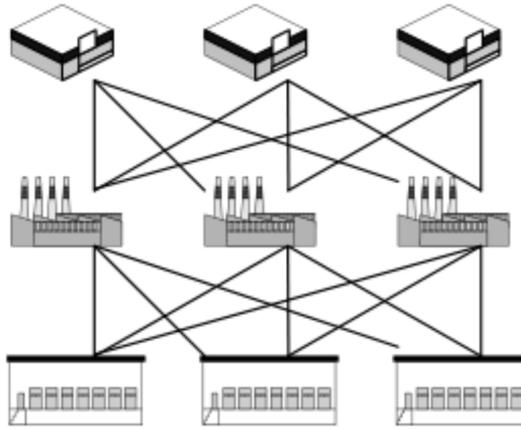


Fig. 2. Pairwise relationships between supply chain facilities

Supplier				Plant			DC		Stage 1			Stage 2			Stage 3				
1	0	1	0	1	1	0	0	1	1	2	3	1	3	2	1	1	3	4	2

Fig. 3. Chromosome coding

The fitness function of this paper is shown in the equation

$$\text{Fitness} = \begin{cases} \frac{1}{1+\exp((f-f_{\text{avg}})/c)}, & g \geq 20\%n, \\ \frac{1}{1+\exp(f-f_{\text{avg}})}, & g < 20\%n. \end{cases} \quad (5)$$

Here,  $f$  is the original fitness: this fitness function is for solving the minimum value of the objective function. If the problem itself is solved as the minimization problem,  $f$  directly represents the objective function.

The crossover probability  $p_c$  and the mutation probability  $p_m$  are the key parts of the genetic algorithm. Compared with the traditional genetic algorithm, the adaptive genetic algorithm (AGA) can adaptively change the crossover and mutation probability in the optimization process according to the number of iterations and its own fitness function, which avoids the problem that the standard genetic algorithm (GA) falls into the local solution or premature convergence in the solution (Wang et al. 2015) [15].

The crossover probability  $p_c$  and the mutation probability  $p_m$  are given as

$$p_c = \begin{cases} 0.6 - \frac{0.6-0.8}{1+\exp(10(2-\frac{3(f_{\text{max}}-f)}{f_{\text{max}}-f_{\text{avg}}+1}))}, & f' \geq f_{\text{avg}} \\ 0.8 - \frac{0.8-0.9}{1+\exp(10(1-\frac{3(f_{\text{avg}}-f)}{f_{\text{avg}}-f_{\text{min}}+1}))}, & f' < f_{\text{avg}} \end{cases} \quad (6)$$

$$p_m = \begin{cases} 0.005 - \frac{0.005-0.01}{1+\exp(10(\frac{3(f_{\max}-f)}{f_{\max}-f_{\text{avg}}+1}-2))}, & f \geq f_{\text{avg}}, \\ 0.001 - \frac{0.001-0.005}{1+\exp(10(\frac{3(f_{\text{avg}}-f)}{f_{\text{avg}}-f_{\min}+1}-1))}, & f < f_{\text{avg}}. \end{cases} \quad (7)$$

### 4. Results analysis and discussion

In this paper, the feasibility and potential of mixed integer nonlinear programming supply chain model and adaptive genetic algorithm were explained by bicycle. In general, the bicycle has three direct subassemblies, including the front and rear wheels (W1), the main frame (Mf) and the foot (also including the chain) (Pa), among them, the main frame (Mf) further comprises a seat (Sa) and a frame (Fr). (Note: here the wheels (W1) represented the front wheels and rear wheels; each bicycle needed a unit of wheels, and the number of each unit component was 1.) Table 1 shows the four program characteristics, Table 2 shows the manufacturer’s design/production capacity under the program 1, as well as the manufacturer and distribution center setup costs and customer requirements. Table 3 shows the transportation costs between the two facilities in scenario 1.

According to the AGA parameters and the solving steps set up above, the model was solved by AGA. The convergence is shown in Figs. 4-7. When the program was different, the workload of suppliers, manufacturers and distribution centers was obtained as shown in Tables 4-6.

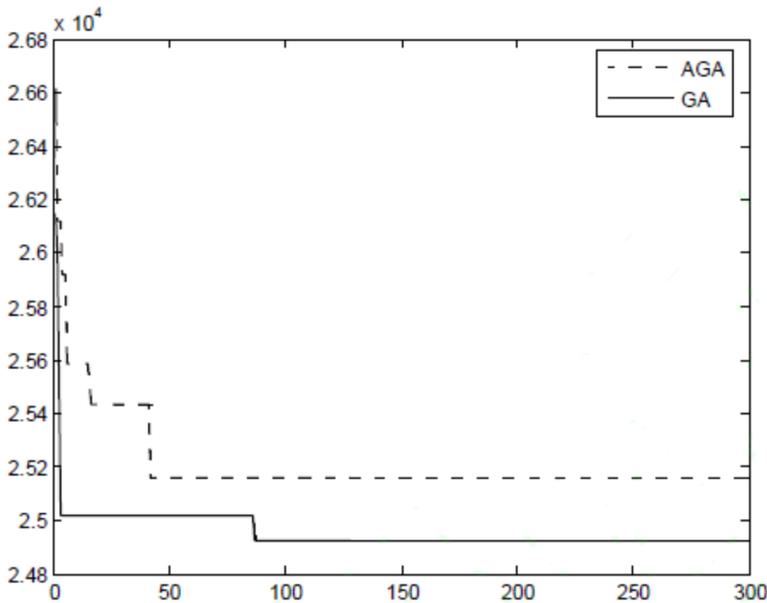


Fig. 4. Convergence graph of adaptive genetic algorithm for scheme 1



Table 4. Quantity of components shipped by supplier to manufacturer

Supplier/Plant	1	2	3
1	-	805	1015
2	-	805	0
3	-	0	1015
4	-	805	1015

Table 5. The number of parts shipped by the manufacturer to the distribution center

Plant/DC	1	2	3
1	-	-	-
2	434	371	-
3	435	580	-

Table 6. The number of components shipped by customer base to distribution center

DC/CZ	1	2	3	4
1	580	234	50	5
2	0	196	410	345
3	-	-	-	-

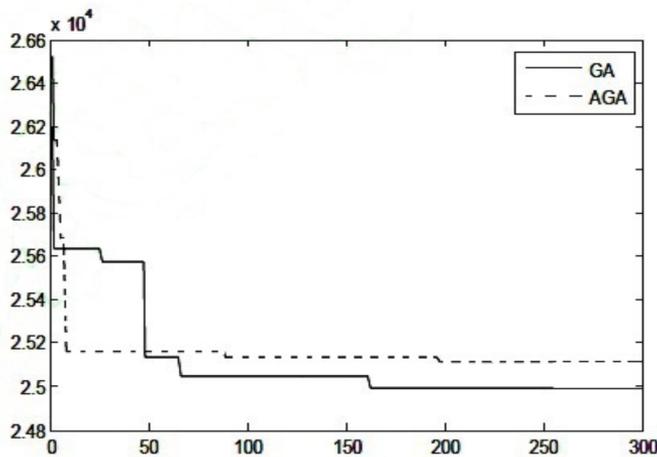


Fig. 5. Convergence graph of adaptive genetic algorithm for scheme 2

It can be seen from the above graph that a mixed integer nonlinear programming (MINLP) model was proposed based on considering the interaction between the strategic and operational layers and the associated constraints. The lowest overall cost of the supply chain was used as the optimization objective. The AGA which can deal with the constraint was used to optimize the model, so that the corresponding

optimization scheme was obtained. The experimental results show that the proposed mixed integer nonlinear programming model can effectively solve the problem of supply chain collaborative optimization in strategic supply chain design, and can get a better supply chain design.

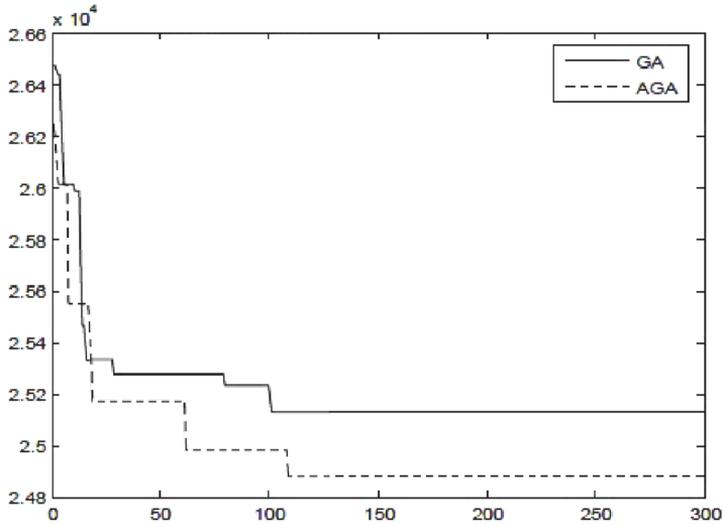


Fig. 6. Convergence graph of adaptive genetic algorithm for scheme 3

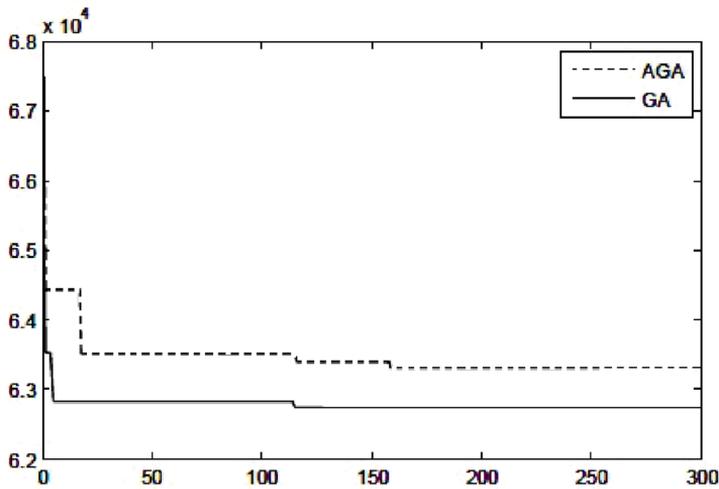


Fig. 7. Convergence graph of adaptive genetic algorithm for scheme 4

## 5. Conclusion

The purpose of this paper is to optimize the strategic and operational layers of the supply chain. In order to achieve the goal, in this paper, the mixed integer nonlinear programming model and the supply chain designed by the adaptive genetic algorithm (GA) were established on the basis of the consideration from the aspects of the supply chain, and the lowest cost of supply chain was optimized. Based on the adaptive genetic algorithm, the model was solved, and the optimal supply chain design scheme was obtained. Finally, the following conclusions were drawn: the model and the solution method provide the decision support in the supply chain design. The mixed integer nonlinear programming model takes into account the three main stages of the procurement, production and distribution of the supply chain and the interaction among them, and considers the optimal allocation of the supply chain from a holistic perspective, and puts the implementation of paired relations and supplier preferences in the constraints, which helps the actual design of the supply chain. In this paper, the optimal design of the supply chain was implemented, and the optimal supply chain design was obtained through the use of AGA. However, this paper still has some limitations, for example, when the enterprise is pursuing multiple performance indicators at the same time, such as cost, income and so on, this design cannot meet it, it is necessary to comprehensively configure the supply chain in a multi-objective form.

## References

- [1] A. HASANI, A. KHOSROJERDI: *Robust global supply chain network design under disruption and uncertainty considering resilience strategies: A parallel memetic algorithm for a real-life case study*. Transportation Research Part E: Logistics and Transportation Review 87 (2016), 20–52.
- [2] M. BRANDENBURG, K. GOVINDAN, J. SARKIS, S. SEURING: *Quantitative models for sustainable supply chain management: Developments and directions*. European Journal of Operational Research 233 (2014), No. 2, 299–312.
- [3] S. H. AMIN, G. ZHANG: *An integrated model for closed-loop supply chain configuration and supplier selection: Multi-objective approach*. Expert Systems with Applications 39 (2012), No. 8, 6782–6791.
- [4] S. H. AMIN, G. ZHANG: *A three-stage model for closed-loop supply chain configuration under uncertainty*. International Journal of Production Research 51 (2013), No. 5, 1405–1525.
- [5] M. ESMAELIKIA, B. FAHIMNIA, J. SARKIS, K. GOVINDAN, A. KUMAR, J. MO: *A tactical supply chain planning model with multiple flexibility options: an empirical evaluation*. Annals of Operations Research 244 (2016), No. 2, 429–454.
- [6] A. S. SAFAEI, S. M. M. HUSSEINI, R. Z. FARAHANI, F. JOLAI, S. H. GHODSYPOUR: *Integrated multi-site production-distribution planning in supply chain by hybrid modeling*. International Journal of Production Research 48 (2010), No. 14, 4043–4069.
- [7] M. EL-SAYED, N. AFIA, A. EL-KHARBOTLY: *A stochastic model for forward–reverse logistics network design under risk*. Computers & Industrial Engineering 58 (2010), No. 3, 423–431.
- [8] A. HASANI, S. H. ZEGORDI, E. NIKBAKHS: *Robust closed-loop global supply chain network design under uncertainty: The case of the medical device industry*. International Journal of Production Research 53 (2015), No. 5, 1596–1624.

- [9] A. JINDAL, K. S. SANGWAN: *Closed loop supply chain network design and optimisation using fuzzy mixed integer linear programming model*. International Journal of Production Research 52 (2014), No. 14, 4156–4173.
- [10] F. ARIKAN: *A fuzzy solution approach for multi objective supplier selection*. Expert Systems with Applications 40, (2013), No. 3, 947–952.
- [11] A. T. E. PÉREZ, M. CAMARGO, P. C. N. RINCÓN, M. A. MARCHANT: *Key challenges and requirements for sustainable and industrialized biorefinery supply chain design and management: A bibliographic analysis*. Renewable and Sustainable Energy Reviews 69 (2017), 350–359.
- [12] H. BADRI, M. BASHIRI, T. H. HEJAZI: *Integrated strategic and tactical planning in a supply chain network design with a heuristic solution method*. Computers & Operations Research 40 (2013), No. 4, 1143–1154.
- [13] A. SIFALERAS, L. KONSTANTARAS: *Variable neighborhood descent heuristic for solving reverse logistics multi-item dynamic lot-sizing problems*. Computers & Operations Research 78 (2017), 385–392.
- [14] B. FAHIMNIA, J. SARKIS, A. ESHRAGH: *A tradeoff model for green supply chain planning: A leanness-versus-greenness analysis*. Omega 54 (2015), 173–190.
- [15] K. GOVINDAN, A. JAFARIAN, R. KHODAVERDI, K. DEVIKA: *Two-echelon multiple-vehicle location–routing problem with time windows for optimization of sustainable supply chain network of perishable food*. International Journal of Production Economics 152 (2014), 9–28.

Received May 7, 2017

# A computer behavioral analysis algorithm based on image color classification statistics<sup>1</sup>

JIA WANG<sup>2</sup>

**Abstract.** At present, the analysis of computer operation behavior is mainly through the artificial desktop monitoring, recording button content, file operation records and other ways to monitor and analyze the behavior of the computer. In order to study how to use image color classification statistics to analyze computer behavior, a method based on color image clustering analysis and K-means algorithm for statistical analysis of graphics to achieve behavior analysis was proposed in this paper. And according to the type, range and depth of the color, the computer operation images were classified and counted. In this way, the entertainment, work and other operations performed by the computer at a given point of time were determined. The final experimental results show that the algorithm is more practical, and the effect of the deep colored game and movie analysis is remarkable.

**Key words.** Color image clustering algorithm, K-means algorithm, behavior analysis.

## 1. Introduction

With the continuous development of science and technology and the continuous expansion of computer applications, the image processing method came into being. The purpose of which is to classify the image and analyze the image information intelligently by using computer equipment. Nowadays, the image processing and recognition have been applied more and more widely. However, as far as the current level is concerned, the computer's perception of the external is still relatively weak, and a lot of manpower and material resources are needed to study the theory and application of digital image processing and recognition. As a result, all walks of life have greater demand for the accuracy and intelligence of digital image processing technology, especially in aerospace, biomedical engineering, industrial testing, robot

---

<sup>1</sup>This work is supported by Project from The Applied Basic Research Youth Programs of Science and Technology Department of Yunnan Province, Analysis on Screen Soft Proofing based on ICC Profile (No .2015FD039)

<sup>2</sup>Yunnan Open University, Kunming City, Yunnan Province, 650500 China

vision, public security, justice, culture, art and other fields. In this paper, according to the characteristics of digital image processing technology, it was applied to the analysis of universal computer operation behavior, and an innovative function of intelligent analysis of its operation behavior based on the image color appearing in the process of computer operation was realized.

## 2. State of the art

$K$ -means algorithm is the most common classical algorithm in data clustering analysis, which has been widely used because it has the advantages of short time and simple algorithm when clustering data [1]. However, there are some shortcomings in the algorithm, for example, high-dimensional and non-spherical data are difficult to be clustered, and greatly affected by the selection of initial cluster centers during clustering, which will fall into the local optimal solution during the solution process, so that the clustering result is not good, and the number  $K$  of cluster centers needs to be determined before clustering analysis [2]. Therefore, many experts and scholars have carried out a more in-depth study of the  $K$ -means algorithm, so as to find out the new algorithm to make up for the shortcomings of the  $K$ -means algorithm. When some commonly used clustering algorithms are used to deal with some high-dimensional complex data and nonlinear data, there are some problems such as long computing time, low efficiency and low accuracy of clustering. Therefore, a new distance cost function is proposed, which can calculate all kinds of data, including intra class and inter class data distance, and the computing efficiency is higher [3]. The division of data is done by the principle of maximum and minimum distance. After the data is partitioned, the required  $K$  value can be automatically determined without the user's prior experience. There is a clustering algorithm, which can determine the optimal number of clusters  $K_{\text{opt}}$  for the selected data. The specific process is as follows: firstly, the upper and lower boundaries of the selected object are determined, thus effectively reducing the scope of cluster search. Then in the algorithm, the  $k$  parameter should be set up, and  $K_{\text{min}}$  is usually set to 2. The setting of  $K_{\text{max}}$  needs to be determined according to the type of data selected and the data AP, and then the optimal clustering number  $K_{\text{opt}}$  is determined by the Silhouette clustering validity index [4]. An improved  $k$ -means algorithm is proposed for isolated data, the main idea of which is to determine the initial clustering center based on the average distance method. Firstly, the isolated data is removed from the general data, and then the distance between the remaining data is calculated, and a set of data with the smallest distance in these data is selected. The center of the set of data is used as the initial clustering center. Then the same method is used to obtain the next smallest set of data between the data sets, and the data in the set is also chosen. The distance between the two cluster centers is calculated, and finally whether the distance is less than the average distance between all the data is determined. If it is less than the clustering among all the data, the center is taken as the second clustering center, and this is continued to find  $K$  clustering centers [5].

### 3. Methodology

Color image clustering analysis and  $K$ -means algorithm are the keys of this topic. In this paper, through color image clustering analysis and  $K$ -means algorithm, the image was transformed into 3 kinds of modules, which were normal image, rendering image and gray image. These 3 types of modules were clustered simultaneously, and its influence color block interval and influence depth interval that can affect behavioral analysis were worked out. Finally, based on the combination of two kinds of intervals, the original image was judged according to the threshold value, and the analysis conclusion of operation behavior was obtained [6].

$K$ -means algorithm is a traditional and classic data clustering algorithm, the principle of which is to analyze the similarity between different spatial data by Euclidean distance [7]. The algorithm's data clustering process is: firstly, the data to be clustered is divided, and the  $K$  value of the algorithm is set. The data is divided into  $k$  families by means of  $K$  values, and the principle of similarity is used to make the data in the dataset divided into  $K$ -class families, and the partitioned data is similar in the same class. The similarity of data in different classes is small, but the algorithm is used to cluster data [8].

If  $Q$  is a finite set  $X = \{x_1, x_2, \dots, x_n\}$  of spatial  $S^Q$ , the initialization is randomly divided into  $K$ , remarked as  $C_1, C_2, \dots, C_K$ . If there are  $n$  objects in a class, the clustering center of class  $i$  is defined as  $Z_1, Z_2, \dots, Z_K$ , as is given by the expression

$$Z_i = \frac{1}{n} \sum_{j=1}^n x_j, j \in [1, K]. \quad (1)$$

The defined objective function is given by the expression

$$J = \sum_{i=1}^K \sum_{j=1}^{n_j} D_{x_j, z_i}^2. \quad (2)$$

$D_{x_j, z_i}^2$  represents the distance from the first  $j$  text to the cluster center of class  $i$ , that is, the Euclidean distance.

There are three main steps in  $K$ -means algorithms running.

Step 1:  $K$  objects are selected from the finite set  $X$  as the clustering center.

Step 2: the distance between objects are calculated according to the Euclidean distance, and the objects of the same distance are divided into the corresponding clusters.

Step 3: according to formula (2), the cluster center is continued to calculate, and step 2 is repeated until the algorithm converges [9]. The flow chart is shown in Fig. 1.

$K$ -means algorithm is a classical clustering algorithm for data processing, which usually requires two stages of processing when dealing with data. These two processes are: firstly, the data algorithm is set by class, and is equally allocated to these classes. Then, the spatial distances of the data are computed. It can be found that if the distance is shorter, the similarity will be high, the distance between races will

be large, and the similarity will be low. If it is found that the initial clustering center is not good, then the cluster center needs modified [10].

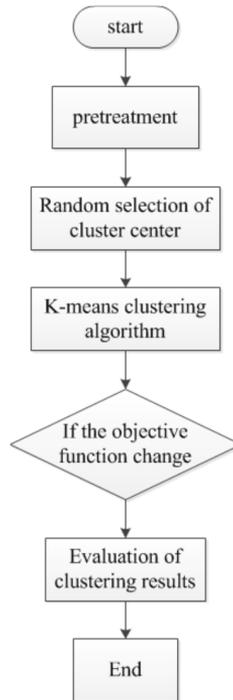


Fig. 1. *K*-means algorithm flow chart

In order to implement the algorithm in this paper, firstly, the computer image and the size of the current form were obtained. Then an image with the current form as a template and a bitmap Bitmap drawing surface were created. The handle of the form and the handle to the image were obtained. The clods of raster operation code was copied, the API function was called, so as to achieve a form of capture, release the handle, and save the image [11].

In this example, the image results are shown in Fig. 2.

According to the behavior analysis image - custom color of the system, the image was unified into RGB format, the pixels of the image was traversed, the type and proportion of colors in the image and the position of the color system were obtained, and the weight of the color system in which the main colors were located was counted (Sun et al. 2008) [12]. The image content was analyzed according to the custom color model. Based on the RGB format of the color image, the three-dimensional coordinate map was established, among them, R was the *X* axis, G was the *Y* axis, B was the *Z* axis, and the coordinate axis length was 255. According to the slight difference of the approximate color, the three channel pigment was divided into 14 colors - 14 color blocks. After that, according to the temperature and the brightness of the color blocks, the existing operation behavior and image results were matched, and the specific threshold was obtained (Zhang et al. 2011) [13].



Fig. 2. Original drawing obtained by the computer

Figure 3 shows a partial chromatic graph of a partition. According to the color image, all pixels were planned into 14 color blocks. After that, all pixel points were traversed and incorporated into class groups. The pixels of the same class were rendered to the same color.

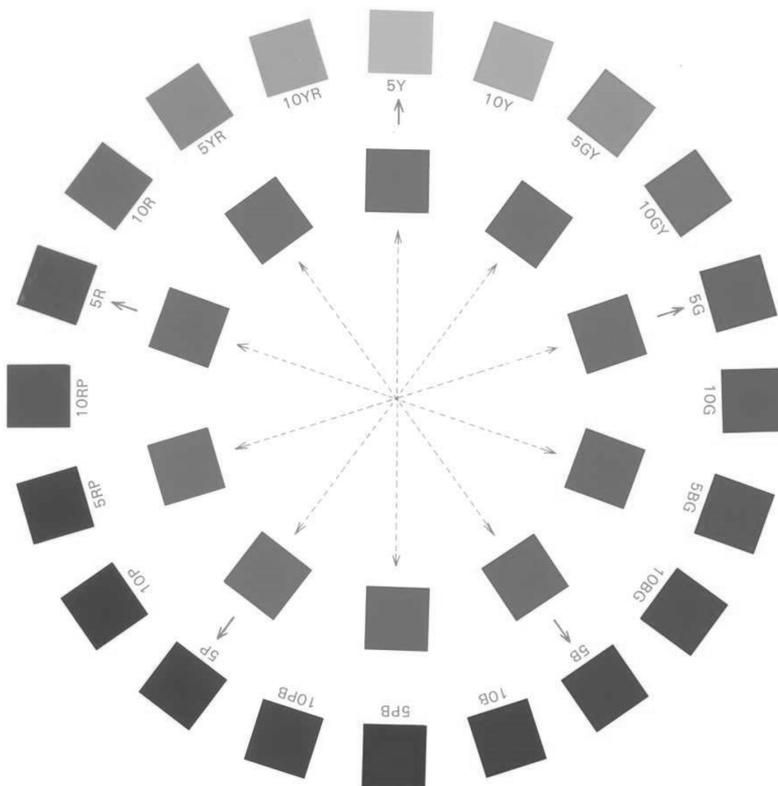


Fig. 3. Part divided color map

Figure 4 shows the rendering. Through repeated test experience accumulation, the six color areas affected by the color proportion in the color block were analyzed according to the lightness, temperature, hue and brightness of the color. An influence color interval was constructed, so as to store the pixels in the six color zones that can ultimately affect the results of the behavioral analysis. And then these six color areas were counted. When the proportion of pixels in a color area exceeded the specified threshold (this threshold was a range of behavior color distinction boundary summed up after repeated trials, pictures, analyses, calculations), the proportion was saved in the influence color interval (Zhang et al. 2011) [14]. Figure 5 shows the proportion of dominant colors.



Fig. 4. Clustering renderings



Fig. 5. Scale diagram of main color system

In this research, the clustering analysis of gray image is mainly based on  $K$ -means algorithm. The advantages of the adopted  $K$ -means clustering algorithm are mainly concentrated in: the algorithm is fast and simple; for large data sets, it is more efficient and scalable; the time complexity is near linear, and it is suitable

for mining large-scale datasets [15]. The time complexity of the  $K$ -means clustering algorithm is  $O(nKt)$ , among them,  $n$  represents the number of objects in the dataset,  $t$  represents the number of iterations of the algorithm, and  $K$  represents the number of clusters.

The specific process of  $K$ -means algorithm consists of the following steps:

The first step is to randomly select  $K$  clusters:  $\mu_1, \mu_2, \dots, \mu_K \in R^n$ .

The second step is to repeat the following process until it converges.

For each sample  $i$ , the class that should belong to is calculated, as shown in the following expression

$$c^{(i)} = \arg \min_j \|x^{(i)} - \mu_j\|^2. \quad (3)$$

For each class  $j$ , the centroid of the class is recalculated, as shown in the formula

$$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\}x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}}. \quad (4)$$

In the  $K$ -means algorithm, the selection of  $K$  is manually specified, and the selection of the  $K$  value is very difficult to estimate. The summary was made on the basis of a large number of raw data experimental results in this paper, and it was concluded that when the  $K$  value was selected to be 5, the operation performance analysis had the best precision.

In the  $K$ -means algorithm, firstly, an initial partition is determined according to the initial clustering center, and then the initial partition is optimized. The choice of the initial cluster center has a great impact on the clustering results. Once the initial value is not selected properly, the clustering results may not be valid. In this study, the initial clustering center was set according to the center of the position of the dominant color system in which the color image was the largest.

## 4. Result analysis and discussion

The  $K$ -means algorithm was used in this paper, and through Table 1 and Table 2, further analysis was carried out.

Table 1. The result analysis table of League of Legends

	Entertainment	Office learning	Other
The proportion	79 %	8 %	13 %

Table 2. The result analysis table of word

	Entertainment	Office learning	Other
The proportion	7 %	69 %	24 %

The entertainment games, movies, office software and learning software which were frequently occurred in the market were tested in this study, the  $K$ -means algorithm was more practical. The analysis of deep colored movies and games was very effective, such as the League of Legends shown in Fig. 1.

For light colored games, the analysis was disturbed in a particular scene, such as JX Online Version Three, referred to as the JX Three. The image color range of the game was fluctuated after a halo treatment of the picture. In view of the phenomenon, a lot of picture testing was carried out, and the game image was summarized, thereby providing a very good data correction for the future algorithm modification.

In addition, in traditional  $K$ -means algorithms, Euclidean distance is used as similarity measure. From the characteristics of Euclidean distance, the Euclidean distance can effectively measure the similarity between the spherical data with uniform distribution, while the Euclidean distance cannot effectively measure the similarity relation between inhomogeneous and non-spherical data. In other words, for each attribute of each data, Euclidean distance treats it the same, and the weights are same. However, in practical problems, the different attributes of data have different effects on the results of data. Therefore, one of the main drawbacks of using Euclidean distance as a similarity measure is that the traditional  $K$ -means algorithm is not suitable for non-spherical data sets with uneven distribution.

At the same time, through the analysis of the  $K$ -means algorithm and its privacy leakage problem, it can be seen that the key point of privacy leakage is the cluster center point. The clustering center is obtained by dividing the sum of the data points in the cluster by the data points, and detailed data point information is not needed when clustering data sets into a data set. As a result, only by publishing the approximate values of each cluster center point can the data privacy be protected, and meanwhile, the accuracy of the clustering results will not be affected.

In the massive data processing algorithms of this paper and the term, most of them adopt the tree index structure. The biggest disadvantage of this algorithm is the curse of dimensionality. As the amount of data increases, the tree structure expands exponentially, and the memory space becomes larger and larger. Especially for high dimensional data in the image, it is virtually impossible to put it into memory completely. In order to solve this problem, the external storage of hard disk storage is introduced. Although it can solve the problem of storage space, the search efficiency is reduced. There is no doubt that this is a way to exchange space for time, which can greatly reduce the user experience.

## 5. Conclusion

The purpose of this paper is to test the current entertainment games, movies, office software and learning software. In this paper, the color image clustering algorithm was used, and the color of the image was classified and counted. The influence color interval was constructed and matched, and the interval threshold was demarcated. At the same time,  $K$ -means algorithm was introduced to process the image. Through experimental data, as well as empirical selection,  $K$  values were selected to

suit behavioral analysis calculations. In addition, the depth influence interval was constructed, and the color depth of the image was classified. Finally, combining the influence of color interval, the boundaries of behavior analysis were delineated, and the innovation of intelligent analysis of computer operation behavior was realized. Research showed that the accuracy rate of the combination of the image abstract semantic features extracted by  $K$ -means algorithm and the image low-level visual features for image classification and retrieval was higher than traditional method. The search results were further checked and the results of erroneous retrieval were eliminated, so that higher precision can be achieved. Then, the abstract semantic features were fused with the image color features to compensate for the insensitivity of the CNN algorithm to the color information. However, there are still some problems in this study, for example, how to make the analysis of light color images more effective needs further study.

## References

- [1] R. ACHANTA, A. SHAJI, K. SMITH, A. LUCCHI, P. FUA, S. SUSSTRUNK: *SLIC superpixels compared to state-of-the-art superpixel methods*. IEEE Transactions on Pattern Analysis and Machine Intelligence *34* (2012), No. 11, 2274–2282.
- [2] T. KANUNGO, D. M. MOUNT, N. S. NETANYAHU, C. D. PIATKO, R. SILVERMAN, A. Y. WU: *An efficient  $k$ -means clustering algorithm: Analysis and implementation*. IEEE Transactions on Pattern Analysis and Machine Intelligence *24* (2002), No. 7, 881–892.
- [3] L. H. JUANG, M. V. WU: *MRI brain lesion image detection based on color-converted  $K$ -means clustering segmentation*. Measurement *43* (2010), No. 7, 941–949.
- [4] K. CHEN, Y. ZHU: *A summary of machine learning and related algorithms*. Statistics & Information Forum (2007), No. 05.
- [5] A. K. JAIN: *Data clustering: 50 years beyond  $K$ -means*. Pattern Recognition Letters *31* (2010), No. 8, 651–666.
- [6] S. L. YANG, Y. S. LI, X. X. HU, Y. R. PAN: *Optimization study on  $k$  value of  $K$ -means algorithm*. Systems Engineering-theory & Practice *26* (2006), No. 2, 97–101.
- [7] L. GRADY: *Random walks for image segmentation*. IEEE Transactions on Pattern Analysis and Machine Intelligence *28* (2006), No. 11, 1768–1783.
- [8] H. D. ZHU, Y. ZHONG, X. H. ZHAO: *An optimization initial center  $K$ -means algorithm for text clustering*. Journal of Zhengzhou University(Natural Science Edition) *41* (2009), No. 02.
- [9] K. R. ŽALIK: *An efficient  $k$ '-means clustering algorithm*. Journal Pattern Recognition Letters *29* (2008), No. 9, 1385–1391.
- [10] M. MIGNOTTE: *A de-texturing and spatially constrained  $K$ -means approach for image segmentation*. Pattern Recognition Letters *32* (2011), No. 2, 359–367.
- [11] M. MIGNOTTE: *Segmentation by fusion of histogram-based  $K$ -means clusters in different color spaces*. IEEE Transactions on Image Processing *17* (2008), No. 5, 780–787.
- [12] J. FAN, M. HAN, J. WANG: *Single point iterative weighted fuzzy  $C$ -means clustering algorithm for remote sensing image segmentation*. Pattern Recognition *42* (2009), No. 11, 2527–2540.
- [13] A. K. JAIN, F. FARROKHNI: *Unsupervised texture segmentation using Gabor filters*. Pattern Recognition *24* (1991), No. 12, 1167–1186.
- [14] C. FOWLKES, S. BELONGIE, F. CHUNG, J. MALIK: *Spectral grouping using the nystrom method*. IEEE Transactions on Pattern Analysis and Machine Intelligence *26* (2004), No. 2, 214–225.

- [15] D. L. PHAM, C. XU, J. L. PRINCE: *Current methods in medical image segmentation*. Annual review of biomedical engineering 2 (2000), 315–337.

Received May 7, 2017

# Auxiliary method of transformer intelligent partial discharge detection based on UWB technology

CHENGBO HU<sup>1,4</sup>, YONGLING LU<sup>1</sup>, FENGBO TAO<sup>1</sup>,  
WEIHUA CHENG<sup>2</sup>, WENNAN WU<sup>3</sup>

**Abstract.** At present, the method of determining the sensor placement position in partial discharge (PD) detection technology is carried out by manual measurement. The method has many drawbacks such as low efficiency, thus an auxiliary method of PD intelligent detection based on ultra-wideband (UWB) technology is proposed. In this method, UWB modules are deployed around the transformer, and localized PD detection sensors on the transformer surface are analyzed by UWB technology. The establishment of the auxiliary method model is analyzed in detail, and the methods of determining the coordinates of base station and tags are discussed. Through the realization of discovery, ultra-wideband in the metal environment can maintain a high ranging accuracy. Through the CAD simulation, it is concluded that the proposed auxiliary method is feasible in theory and can promote the intelligent development of PD detection.

**Key words.** UWB, transformer, auxiliary method, partial discharge.

## 1. Introduction

Partial discharge refers to a discharge phenomenon in which the field strength generated in the electric device when applied voltage is sufficient to cause the discharge of the insulating portion area but does not form a fixed discharge channel in the discharge region. Transformer partial discharges detection [1–4] can effectively detect the status of the insulation for the state of the transformer to provide accurate guidance. Partial discharge detection is also known as PD detection, it has been studied by many scholars at home and abroad [5–9]. Currently used PD detection methods are ultrasonic positioning method [10], electrical positioning method [11], infrared detection method [12] and so on. TDOA (Time Difference of Arrival)

---

<sup>1</sup>Jiangsu Electric Power Company Research Institute, Nanjing, 211167, China

<sup>2</sup>Jiangsu Power Info-tech Co. Ltd., Nanjing, 211167, China

<sup>3</sup>Jiangsu BDS Application Industry Institute, Nanjing, 211167, China

<sup>4</sup>Corresponding author

method is usually used to determine the three-dimensional coordinates of the points [13]. TDOA method is mature and stable, and has been widely used in many fields [14–15].

Take ultrasonic positioning method as example, the coordinates of the ultrasonic positioning sensor in the transformer surface are needed to achieve TDOA positioning. Usually we take the point at the bottom of transformer as the origin of the coordinates, the coordinates of the sensor placement point are measured by manual tape. Manual tape measurement can lead to the following three drawbacks:

- i. It will increase labor costs.
- ii. Because the transformer surface is not absolute plane, human measurement will inevitably bring errors or gross margin.
- iii. Man-made processes are cumbersome and inefficient. Therefore, it is considered to optimize it, in order to realize intelligent PD detection of substation.

UWB(Ultra Wide Band) is a low-power, low-cost but high-speed wireless communication technology. Recently, UWB positioning technology has been extensively researched by researchers. Normal operating frequency of UWB is between 3.0 GHz to 10.6 GHz [16]. UWB is characterized by not using carrier communication, but with very short time interval (nanosecond or less than nanosecond time interval) of the baseband narrow pulse communication, with penetrating power, anti-multipath effect of anti-interference ability outstanding. Especially in the metal or liquid environment that have great impact on signal attenuation, UWB plays a stronger performance than other wireless positioning technology [17].

In this paper, an auxiliary method based on UWB technology is proposed to determine the coordinates of the ultrasonic sensor placement point. This method use UWB sensors as little as possible to locate ultrasonic positioning points on the four sides of the transformer. Finally, through CAD simulation test environment, the correctness and feasibility of the auxiliary scheme are verified, which lays a theoretical foundation for practical operation.

## 2. Intelligent partial discharge detection assistant model

### 2.1. Principle of UWB ranging

UWB uses two-way time of flight (TW-TOF, two way-time of flight) for ranging, ranging information and other information can be transmitted between the modules. Distance measurement principle is shown in Fig. 1.



Fig. 1. Principle of UWB ranging

Each module generates a separate timestamp from the start. The transmitter of

module A transmits a pulse signal of the requested nature at the time  $T_{a1}$  on its time stamp, and the receiver of module B receives  $T_{b1}$  on its time stamp. After a certain processing of the signal means, the module B in  $T_{b2}$  time launches a response to the nature of the signal, and the module A receives it in its own time stamp in time  $T_{a2}$ . The distance  $d$  between the module A and the module B is calculated by the formula (1), where  $c$  is the propagation velocity of the light.

$$d = c \cdot \Delta t = c \cdot \frac{(T_{a2} - T_{a1}) - (T_{b2} - T_{b1})}{2}. \quad (1)$$

### 2.2. Model establishment

The auxiliary model of intelligent PD detection is shown in Fig. 2. The rectangular box in the figure represents a high-voltage transformer. The model consists of 7 UWB modules and 4 mounting posts, of which 6 base station modules (A, B, C, D, E, F) and a tag module. In the model, the base station modules A and B are in the vertical state, and the base station modules E and F are in the vertical state, that is, the straight line AB and the straight line EF are respectively parallel to the high voltage transformer. AB lever increases a G placement position, which can be used to place the module. It should be noted that the G point is located in the middle space between ABC plane and the transformer front side. In addition, the placement of F is closer to the ground than C point.

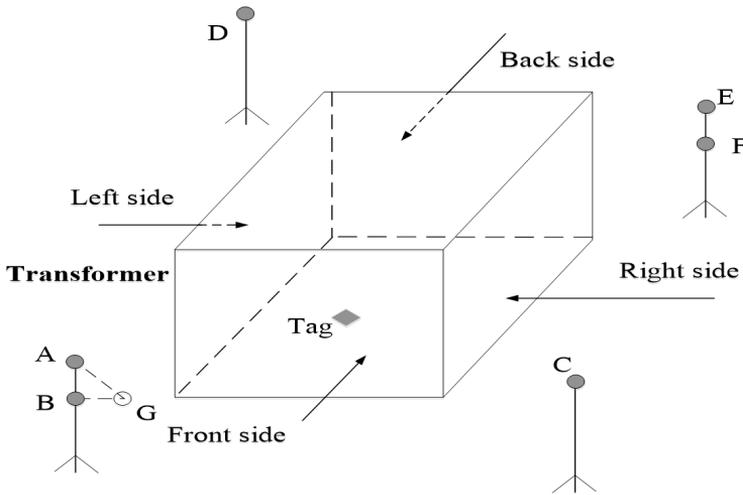


Fig. 2. The model

The coordinate system in the auxiliary model is established as described below. B is the coordinate origin, BA is the positive direction of the  $x$ -axis, ABC is  $x0y$  plane, the  $y$ -axis is perpendicular to BA and B points to C, the  $z$ -axis is perpendicular to ABC ( $x0y$  plane) and points to the side of the transformer.

### 3. Determination of base station and tag coordinates

#### 3.1. Determination of base station coordinates

In order to facilitate observation, the transformer in Fig. 2 is faded away, and several space auxiliary line are added, thus we get Fig. 3. Figure 3 is the base station coordinates of the spatial distribution. The projection of the base station in the  $y0z$  plane cannot be absolutely guaranteed as a matrix due to the actual placement of the benchmarks, which is magnified in the figure. The base station D is placed closer to the transformer side where the base station D is located on the right side of the  $x0z$  plane. When placed, G is located to the left of the  $x0z$  axis.

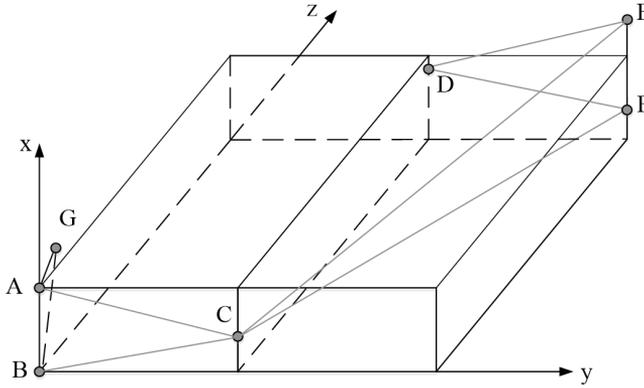


Fig. 3. Spatial distribution of the base station

From Fig. 3, we can see that the spatial coordinates of the base station B are  $(0,0,0)$ , and the spatial coordinates of the base station A are  $(d_{AB},0,0)$ , and  $d_{AB}$  can be obtained by the UWB module ranging. The  $z$  coordinate of the spatial coordinates of the base station C is 0, that is, A, B and C are located in the  $x0y$  plane.  $\angle ABC$  can be obtained from the formula

$$\angle ABC = \arccos \frac{d_{AB}^2 + d_{BC}^2 - d_{AC}^2}{2d_{AB}d_{BC}}. \quad (2)$$

The  $x$  and  $y$  coordinates of base station C are

$$\begin{cases} x_C = x_B + \Delta x_{BC} = x_B + d_{BC} \cos \angle ABC = d_{BC} \cos \angle ABC, \\ y_C = y_B + \Delta y_{BC} = y_B + d_{BC} \sin \angle ABC = d_{BC} \sin \angle ABC. \end{cases} \quad (3)$$

The spatial coordinates of the base station C are  $(d_{BC} \cos \angle ABC, d_{BC} \sin \angle ABC, 0)$ . Putting the module as a label at G, we can list three observation equations, as

shown in formula

$$\begin{cases} (x_A - X_G)^2 + (y_A - y_G)^2 + (z_A - z_G)^2 = d_{AG}^2, \\ (x_B - X_G)^2 + (y_B - y_G)^2 + (z_B - z_G)^2 = d_{BG}^2, \\ (x_C - X_G)^2 + (y_C - y_G)^2 + (z_C - z_G)^2 = d_{CG}^2. \end{cases} \quad (4)$$

Thus we can get the coordinates of point G, but because only three observation equations are used to get the solution of G coordinates, it will be two solutions. The  $x$  and  $y$  values of the two solutions are the same, and the values of  $z$  are opposite to each other. As can be seen from Fig. 3, the solution having the positive  $z$  value is reserved as the coordinate of the base station G.

Similarly, from the base station D and A, B, G three distances can be listed from three distance observation equations. For the two solutions of D, the following rule is used: in the two solutions of D, the values of  $x$  and  $z$  are the same, and the values of  $y$  are opposite to each other. The solution with a positive value of  $y$  is reserved as the coordinate of the base station D.

At this time, the coordinates of the base station A, the base station B, the base station C, and the base station D are determined, and the coordinates of the base station E base station F remain to be determined.

The base station is projected onto the  $y0z$  plane, as shown in Fig. 4. Line DE and line DF are the same in the  $y0z$  plane, which is  $d_1$ . Projection length of line CE and line CF is the same in the plane, and its length is  $d_2$ .

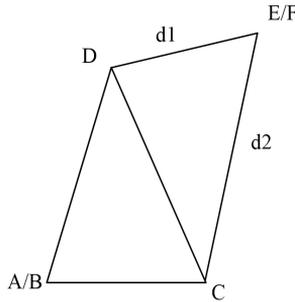


Fig. 4. Projection of base stations in  $y0z$  plane

The projection length of the CD in the  $y0z$  plane can be obtained using the  $y$  coordinate and the  $z$  coordinate of the base station C and the base station D. We can get distance CF, distance CE, distance EF through base station C, base station D and base station E. Symbol  $d_2$  denotes the height of the EF side in  $\triangle CEF$ , where the area of the triangle can be obtained from the formula (5), Helen formula [18], and then  $d_2$  is obtained from (6). The value of  $d_1$  can be obtained using the same method.

$$\begin{cases} p = \frac{d_{CE} + d_{EF} + d_{CF}}{2}, \\ S^2 = p(p - d_{CE})(p - d_{EF})(p - d_{CF}), \end{cases} \quad (5)$$

$$d_2 = \frac{2 \times S}{d_{EF}}. \quad (6)$$

At this time,  $\angle EDC$  and  $\angle ECD$  are obtained from  $y$  and  $z$  coordinates of base station C and base station D, and  $d_1, d_2$  using the cosine theorem. We take  $\angle EDC$  as  $\alpha$ , and we take  $\angle ECD$  as  $\beta$ . The  $y$  coordinate and the  $z$  coordinate of the base station E and the base station F are obtained from the formula

$$\begin{cases} y = \frac{y_D \cot \beta + y_C \cot \alpha + (z_C - z_D)}{\cot \alpha + \cot \beta}, \\ z = \frac{z_D \cot \beta + z_C \cot \alpha - (y_C - y_D)}{\cot \alpha + \cot \beta}. \end{cases} \quad (7)$$

The three-dimensional distance equation can be listed by the base station C and the base station F as shown in the formula

$$(x_C - x_F)^2 + (y_C - y_F)^2 + (z_C - z_F)^2 = d_{CF}^2. \quad (8)$$

There are two solutions to the  $z$  coordinate of the base station F, leaving the solution of the coordinates less than the coordinates of the base station C as the  $z$  coordinate of the base station F. On this basis, the coordinates of the base station E can be obtained by adding the  $z$  coordinate of the base station F to the distance between the base station E and base station F.

### 3.2. Determination of tag coordinates

After determining the base stations' coordinates, we ensure that the transformer around the four sides with three base stations can pass. At this point, the label module placed in the desired location of ultrasonic sensor can be get in real time.

When the label is placed with the front side of the transformer, the  $z$ -coordinate is kept positive. When the label is placed on the left side of the transformer, the  $y$  coordinate is retained as a larger solution. When the label is placed on the back side of the transformer, the smaller solution of the two  $z$  coordinates is retained. When the label is placed on the right side of the transformer, the smaller of the two  $y$ -coordinates is preserved.

### 3.3. Experimental verification

The prerequisite of the auxiliary method model is that ultra-wide band technology can accurately locate in the metal environment state. Therefore, the first choice of experimental scenarios is for ultra-wideband ranging technology to verify the accuracy. Two UWB modules are placed in a metal environment, as shown in Fig. 5, left and right parts.

The distance between two modules is 3.228 m determined by laser range finder. The results of 200 ranging measurements using UWB equipment are shown in Fig. 6. It is found that there are two obvious errors in the distance measurement. After removing the two data, the results is shown in Fig. 7.



The results of ranging data are analyzed and shown in Table 1.

Table 1. Analysis of distance results

Ranging times	Success rate of ranging	average value (m)	RMSE (mm)
200	99%	3.233	15

It can be seen that the UWB equipment has strong anti-multipath effect in the metal environment, and can maintain high accuracy of ranging. This provides a guarantee for the auxiliary method proposed in this paper.

Auto CAD software was used for simulation test. The ranging information is shown in Table 2.

Table 2. Ranging results

Name of distance	Distance value (m)
AB	2.000
AC	10.002
BC	10.161
AG	1.375
BG	1.972
CG	11.033
AD	5.886
BD	6.216
DG	5.408
DE	8.652
DF	8.992
CE	5.025
CF	5.518

The coordinates of each point are solved as shown in Table 3.

Table 3. Coordinates of each point

Point number	$x$ coordinate (m)	$y$ coordinate (m)	$z$ coordinate (m)
A	2.000	0.000	0.000
B	0.000	0.000	0.000
C	1.801	10.000	0.000
D	1.998	0.999	5.801
E	1.500	9.598	4.975
F	-0.500	9.598	4.975
G	1.500	-1.000	0.800

After comparing with the coordinates of the simulation point in CAD, it is found that the coordinates of the coordinates solved is similar to real value in CAD, which indicates that the improved method is theoretically feasible.

## 4. Conclusion

The traditional method of determining the position of the sensor in the PD detection of power equipment has low efficiency and is easy to be influenced by human factors. In view of this, this paper proposes an intelligent auxiliary PD detection method based on UWB technology. Ultra-wideband technology has a strong characteristics of anti-multipath ability in the metal environment, which maintains a high range of precision. The spatial location algorithm in this method is simple and practical, and it can improve the accuracy of real-time positioning effectively. The three-dimensional positioning is realized by only three base stations, which saves the hardware cost. The auxiliary method completely avoid the influence of human factors, which greatly improves the work efficiency. In the later stage, PD detection sensor is combined with UWB label to integrate the sensor location and PD detection, and to realize the positioning of the PD location in a more intelligent way.

## References

- [1] J. G. YANG, J. C. LIU, Y. Y. JIA, S. GAO, F. TAO, G. ZHANG: *Design and comparison of UHF antenna for partial discharge detection*. Electrical Measurement & Instrumentation 53 (2016), No. 18, 21–27.
- [2] S. QIAN, H. CHEN, Y. XU, L. ZHONG, L. SU: *Acoustic fiber optic sensors for partial discharge monitoring*. IEEE Electrical Insulation Conference (EIC), 19–22 June 2016, Montreal, QC, Canada, IEEE Conference Publications (2016), 109–112.
- [3] Y. TIAN, B. QI, R. ZHUO, M. FU, C. LI: *Locating partial discharge source occurring on transformer bushing by using the improved TDOA method*. Proc. IEEE International Conference on Condition Monitoring and Diagnosis (CMD), 25–28 Sept. 2016, Xi'an, China, IEEE Conference Publications, (2016), 144–147.
- [4] W. SHI, J. G. YANG, L. GUO, L. CHEN, M. J. JIANG: *Study of UHF frequency response characteristics of the semicircle dipole inner sensor for partial discharge detection in GIS*. Proc. International Conference on Mechatronics, Manufacturing and Materials Engineering (MMME), 15–16 October 2016, Wuhan, China, MATEC Web of Conferences 63 (2016), Article No. 01016.
- [5] W. HE, H. LI, D. LIANG, H. SUN, Z. SUN, B. LIU: *An improved directional sensor for on-line partial discharge measurement in overhead covered conductors*. Electric Power Components and Systems 44 (2016), No. 14, 1543–1550.
- [6] X. ZHENG, L. H. LU, W. A. FU: *On-line monitoring technique for partial discharge of high voltage explosion-proof motor*. High Voltage Engineering 42 (2016), No. 05, 1651–1658.
- [7] F. CH. SONG, G. LU, D. M. LIU, C. LIN, X. SUN, A. JIN: *Example analysis on partial discharge in the GIS bushing based on UHF technology and positioning method in three dimensional space*. High Voltage Apparatus 52 (2165), No. 09, 55–60.
- [8] F. Y. YANG, H. SONG, X. CHENG, Z. GAO, S. TAO, D. DUAN, G. SHENG, X. JIANG: *Partial discharge feature extraction based on multi-resolution analysis of higher-order singular spectrum entropy*. Power System Technology 40 (2016), No. 10, 3265–3271.
- [9] S. D. MITCHELL, M. SIEGEL, M. BELTLE, S. TENBOHLEN: *Discrimination of partial discharge sources in the UHF domain*. IEEE Transactions on Dielectrics and Electrical Insulation 23 (2016), No. 2, 1068–1075.
- [10] H. Y. LI, H. CHEN, D. G. WU, Y. ZHOU: *Diagnosis and analysis of two partial discharge in CIL equipment by ultrasonic detection method*. High Voltage Apparatus 52, (2016), No. 2, 68–73.

- [11] S. Y. LI, Y. N. WANG, X. B. SONG, L. ZANG: *Application of combined acoustic-electric-chemical detection to GIS partial discharge defect caused by free metal particles*. High Voltage Apparatus 52 (2016), No. 09, 78–82.
- [12] J. L. LIU, M. DONG, S. H. AN, L. ZANG, S. KUANG, W. ZHANG: *Review of partial discharge live detection and location technology for power transformer*. Insulating Materials 48 (2015), No. 8, 1–7.
- [13] J. TANG, L. HUANG, F. ZENG, X. ZHANG: *A positioning approach based on successive approximation of multi-samples for partial discharge source*. Transactions of China Electrotechnical Society 31 (2016), No. 10, 119–126.
- [14] C. ZHAI, Z. DENG, J. JIAO, N. LI, Y. ZHOU, C. LI: *Dynamic weighted data fusion algorithm based on TDOA/RSSI for indoor location*. Proc. China Satellite Navigation Conference (CSNC), 18–20 May 2016, Changsha, China, Lecture Notes in Electrical Engineering, Springer, Singapore 2 (2016), No. 24, 365–374.
- [15] A. POLLARA, A. SUTIN, H. SALLOUM: *Improvement of the detection of envelope modulation on noise (DEMON) and its application to small boats*. IEEE OCEANS 2016 (MTS), 19–23 September 2016, Monterey, CA, USA, IEEE Conference Publications (2016), 1–10.
- [16] R. MAALEK, F. SADEGHPOUR: *A low-profile printed antenna for UWB applications*. Automation in Construction 63 (2016), 12–26.
- [17] R. A. SANTOS, A. C. SODRÉ, S. E. BARBIN: *Study the thermal gradient effect on frequencies of a trapezoidal plate of linearly varying thickness*. Proc. IEEE International Conference on Electromagnetics in Advanced Applications (ICEAA), 19–23 September 2016, Cairns, QLD, Australia, IEEE Conference Publications (2016), 905–908.
- [18] B. CHAKRABORTY: *Pythagorean theorem from Heron's formula: Another proof*. Resonance 21 (2016), No. 7, 653–655.

Received May 7, 2017

# Design of medical track logistics transmission and simulation system based on internet of things<sup>1</sup>

WEI LIU<sup>2,3</sup>, SHUANG LIU<sup>2</sup>

**Abstract.** With the quacking pace of hospital modernization, the introduction of efficient automated logistics system has become imminent. In order to realize the efficient operation of the whole system, the scheduling planning and coordination strategy of the object vehicle in TVS system were studied in this paper. First of all, the single TV path planning was studied. Through the research and analysis of the TVS system, the environment map model was established. The Dijkstra algorithm based on the shortest path principle was adopted to realize the optimal path planning of the single TV. On this basis, the dynamic path planning method with time window principle and Dijkstra algorithm was applied to path planning of TVS system, and the path planning of each TV in the system was realized. The final experimental results show that the TVS system can run at the highest working efficiency during the delivery of the logistics tasks, analyze the conflict and its type in the process of driving, and put forward the strategy of conflict coordination.

**Key words.** Rail logistics transmission system, scheduling planning, coordination strategy.

## 1. Introduction

Logistics transmission system refers to the system of transporting goods in the set area by means of a series of technologies and facilities, such as information technology, photoelectric technology and mechanical transmission device, which is mainly used in airports, shopping malls, banks, factories, libraries and other fields. With the development of electronic information and control technology, the degree of automation of logistics transmission system is higher and higher. In recent years, the logistics transmission system has been popular because of its efficient transporta-

---

<sup>1</sup>The Education Department of JiLin Province Science and Technology Research Project of "13th Five-Year (Subject Source)"; Design of Hospital Logistics Trolley Based on Internet of Things (Topic Name), Project Number: JiLin Branch of The Department of Science and Technology [2016] 107th

<sup>2</sup>JiLin Technology College of Electronic Information, JiLin, 132021, China

<sup>3</sup>Corresponding author

tion and manpower saving, and the application field has gradually expanded to the medical field. A hospital is a place where personnel flow and goods flow are very concentrated. Once there is the high incidence of the epidemic, the drawbacks of this logistics mode will be more exposed. However, the automated logistics transmission system can solve this problem very well. It can not only improve the efficiency of hospital work and save the valuable time of patients, but also save the operation cost of hospital and reduce the flow of staff inside the hospital. The logistics system is an important component of the hospital logistics support system. The automated logistics transmission system with the integration of digital control and photoelectric control technology used in the hospital can not only improve the overall operation and management level of the hospital and improve the overall operation efficiency of the hospital, but also provide more efficient service for the patients. The application of logistics transmission system has become an important symbol of modernization of hospital construction and modernization of hospital management, and it has great application value. There are many kinds of hospital logistics transmission systems, among them, the track transmission system is more and more popular because of its small space occupation, large load space, stable operation and high fault tolerance.

## 2. State of the art

In the developed countries and regions, the introduction of logistics transmission system was relatively early. For example, most of the hospitals in Britain, Germany and France were equipped with automated logistics transmission systems [1]. In 1990s, there were more than ten thousand sets of logistics systems in the use only in Europe. In 1997, a special research project of multi robot system, MARTHA, was set up, namely "a multi autonomous robotic system for handling" [2]. In the past ten years, the application of TVS system has gradually increased. There are hundreds of users in the world only for Swisslog brand track logistics transmission system. A number of research institutes and universities in the United States and Japan have also done a lot of researches on it, and have made fruitful research in theory and practice [3]. Japanese hospitals began using logistics handling equipment in the 1960s to solve the problem of lack of hospital nurses and high labor costs. After decades of development, Japan has developed various large and medium-sized logistics transmission systems, which have achieved rapid promotion and popularization. So far, more than three thousand hospitals in Japan have been equipped with automated logistics systems [4].

Compared with foreign countries, there is a big gap between the abroad and domestic hospital logistics systems both in practice and theoretical research. As a professional field, the research of domestic hospital logistics is still in the initial stage of development. At present, some hospitals with better conditions have begun to equip logistic transmission systems. In 2002, the Affiliated Cancer Hospital of Zhongshan University introduced the first track logistics transmission system in China. In addition, the Third People's Hospital of Yancheng, Suzhou-Xiangcheng People's Hospital, and Ningbo-Beilun District People's Hospital also have introduced the medical TVS system [5]. However, on the whole, the domestic hospital logistics

system is still in the functional management stage. The design of the logistics network is not good, and the framework of the hospital logistics system based on process has not yet been established.

### 3. Methodology

Track logistics transmission system (TVS) is an advanced automated logistics transmission system that is widely used in hospitals. It refers to a system for carrying articles on a predetermined track by means of an intelligent rail vehicle under the control of the computer [6]. Hundreds of sites such as each ward's emergency room, nursing room, operation room, laboratory, blood bank, central pharmacy, central supply room, inpatient department and office area are connected by logistics transmission track in this system. Medical TVS system usually consists of central control system, transceiver workstation, intelligent rail carrying vehicle, logistics track, track converter, automatic isolation door, empty storage area and other equipment [7].

A time window is the time period in which a car enters and leaves. In this period of time, this section can only be used for the car. Other cars are not allowed to pass the section during that time. Based on the traffic condition of the system, the time window method can realize the search for the best path without collision in the bidirectional directed graph. Dynamic path planning method combining Dijkstra algorithm with time window method is used to route planning for TVS system [8].

Suppose that the TVS system has  $n$  platform vehicles, and the car set is  $R = \{r_1, r_2, \dots, r_n\}$ . There is currently  $m$  TVs performing tasks, with the task set of  $M = \{m_1, m_2, \dots, m_m\}$ . The starting execution time of each task  $m_i$  is  $t_s(i)$  with a priority of  $P_i$ . For each task  $m_i$ , there exists a corresponding path, which is a collection of column edges. From the starting point  $O_i$  to the destination point  $D_i$ , the path can be represented by  $\sigma_i = \{e_j, e_k, \dots, e_q\}$ . And  $e_j, e_k, \dots, e_q \in E$  are the paths in a set of edges of the mathematical model of environmental map. Each task can be represented as shown in formula

$$m_i(t) = (O_i, D_i, t_s(i), \sigma_i(t), r_i, P_i). \quad (1)$$

For each task, the starting point, the target point and the assigned TVs do not change over time. The quantity  $\sigma_i(t)$  will not change if the paths of the TV do not conflict during the run. If a collision occurs when the vehicle is executing the task, which only the changing path can be resolved, then the traffic path will be changed dynamically. Each TV has a priority when it is initially assigned tasks. In the course of task execution, as the time goes on, the priority of the task is kept unchanged or set gradually higher according to the actual demand [9].

In the course of running, the carrier vehicle continuously enters and drives out the section in the path. During that time, the section is occupied by the TV, and other TVs are not allowed to enter. For task  $m_i$ , the time window  $S_{ij}$  for the vehicle  $r_i$  to enter the section  $e_j$  is defined by the formula

$$S_{ij} = \{m_i, r_i, l, t_{ij}^{\text{in}}, t_{ij}^{\text{out}}\}. \quad (2)$$

The upper formula defines the time window of the section  $e_j$ , where  $l$  represents that the section  $e_j$  is the  $l$  edge in the path  $\sigma_i(t)$  among the paths found by the car  $r_i$  within the task  $m_i$ . Time  $t_{ij}^{\text{in}}$  is the time to get into the section  $e_j$ , and  $t_{ij}^{\text{out}}$  is the time to leave the section  $e_j$ . Now, the formula

$$t_{ij}^{\text{out}} = t_{ij}^{\text{in}} + \omega_j \quad (3)$$

is satisfied. Here,  $\omega_j$  represents the running time of the car on the section  $e_j$ , as shown in the formula

$$\omega_j = L_j/v_{ij} + t_j, \quad (4)$$

where,  $L_j$  represents the actual length of the section  $e_j$ , and  $v_{ij}$  represents the speed of the car  $r_i$  on the section  $e_j$ . Time  $t_j$  is the buffer time on the section  $e_j$ , and it is usually taken as  $0.05L_j/v_{ij}$ .

Generally, in order to improve the adaptability of the system, it is acceptable for the car to reach the end of the section  $e_j$  within the range of  $[t_{ij}^{\text{out}} - t_j, t_{ij}^{\text{out}} + t_j]$ .

If the section  $e_j$  is the starting edge of the path  $\sigma_i(t)$ , the time for the car to enter the section  $e_j$  is the start time  $t_s(i)$  of the task. If  $e_j$  is not the starting edge, the time to drive into  $e_j$  is the time when the car leaves the  $l - 1$  edge of  $\sigma_i(t)$ , as shown in the formula

$$t_{ij}^{\text{in}} = \begin{cases} t_s(i), l = 0, \\ t_{i(j-1)}^{\text{out}}, l \geq 1. \end{cases} \quad (5)$$

Through the constant iteration of formula (3) and formula (5), the time window of the car  $r_i$  into and out of all the sections and nodes in the path  $\sigma_i(t)$  can be obtained.

The time window of all cars passing through the section  $e_j$  is represented by the time window vector  $e_j = \{S_{1j}, S_{2j}, \dots, S_{mj}\}$ . The dimension of the vector is equal to the number of cars, and it varies with time. If the task  $m_i$  does not enter the section, the driving time and departure time are set to be 0.

By introducing the time window method, the overlap of each task time window is checked. It can detect the conflicts between each task and the type, and then adjust the path with relevant policies, so that the time windows of each task are not overlapped. The path planning of TVS system is realized by checking the node time window [10].

The task scheduling is made according to the generation time of system tasks. If two or more tasks are generated at the same time, they are sorted based on the priority. Then the scheduling strategy based on the priority is used to schedule the tasks. Firstly, the task with the highest priority is selected to search for an idle TV. By using the Dijkstra algorithm, the shortest running path of this TV is designed to perform the task. Then the time of arrival and departure of all occupied nodes during the execution of the task is calculated, and the time windows of each node are initialized [11]. Secondly, the secondary task is the planning path, whether there is the free TV searched. If not, it will enter the waiting state. If so, the Dijkstra algorithm will be used to compute the time windows of all nodes passing in and out of the task, and update the time window vector tables of each node. Thirdly,

the time window vector table has conflicts is checked, and if not, path planning is complete. If it exists, the conflicting nodes are labeled, and the conflict type is judged. The relevant coordination strategy is adopted to eliminate the conflict, and the algorithm is re-programmed by solving the invalid problem. The process is repeated until there is no conflict [12]. Fourthly, the new generation task is assigned to the car and the path is planned. The time window vector table of the system is updated, and the conflict is checked. The above process is repeated. Through such continuous calculation, multi-task scheduling can be completed and a collision free path is planned for the car [13].

Generally, the simulation steps of track logistics transmission system include system investigation, system analysis, system modeling and simulation, strategy scheme optimization, simulation result analysis and so on. The concrete steps are shown in Fig.1.

TVS system simulation can be divided into single vehicle simulation and multi-vehicle simulation. Single simulation can be divided into single point simulation, multi-point simulation and the shortest path simulation. Single point simulation is the simulation of TV running from one transceiver station to any other transceiver station. Multi-point simulation is a simulation of TV running between a transceiver station and a plurality of transceiver stations ( $n_2$ ). The shortest path simulation is the simulation of TV running according to the shortest path between two or more points [1]. Multi-vehicle simulation refers to the simulation of multiple object vehicle ( $r_2$ ) running between designated transceiver stations. It is divided into the shortest path simulation and anti-collision simulation. The shortest path simulation is the same as the single vehicle process. Anti-collision simulation is the simulation to take corresponding actions according to the coordination strategy when multiple vehicles receive the task at the same time, in order to prevent collisions during the running.

Simulation target is the foundation and premise of establishing simulation model. The main target of TVS system simulation is to help researchers understand the behavior and performance of the carrier vehicle in the idle state, the movement with the object and the coordination of the conflict, and assist the effective operation management and scheduling planning of TVS system in the practical application, so as to make the system achieve higher work efficiency during operation [2]. The overall goal of TVS system simulation is to make the whole logistics transmission system run efficiently and orderly, so as to give full play to the efficiency of the carrier car. It can verify the TVS system scheduling planning algorithm, so as to achieve the shortest path between stations for a single TV. During the operation of TVS system, the logistic task is realized and the automatic allocation of the carrier is carried out. In the multi-vehicle system with multiple cars, the shortest path planning of each TV is completed, and the collision between each other can be avoided and the normal operation of the system can be realized.

#### 4. Results analysis and discussion

As shown in Table 1, the TVS system studied in this paper consists of 13 transceiver workstations and 7 carrier TVs. The performance parameters of the

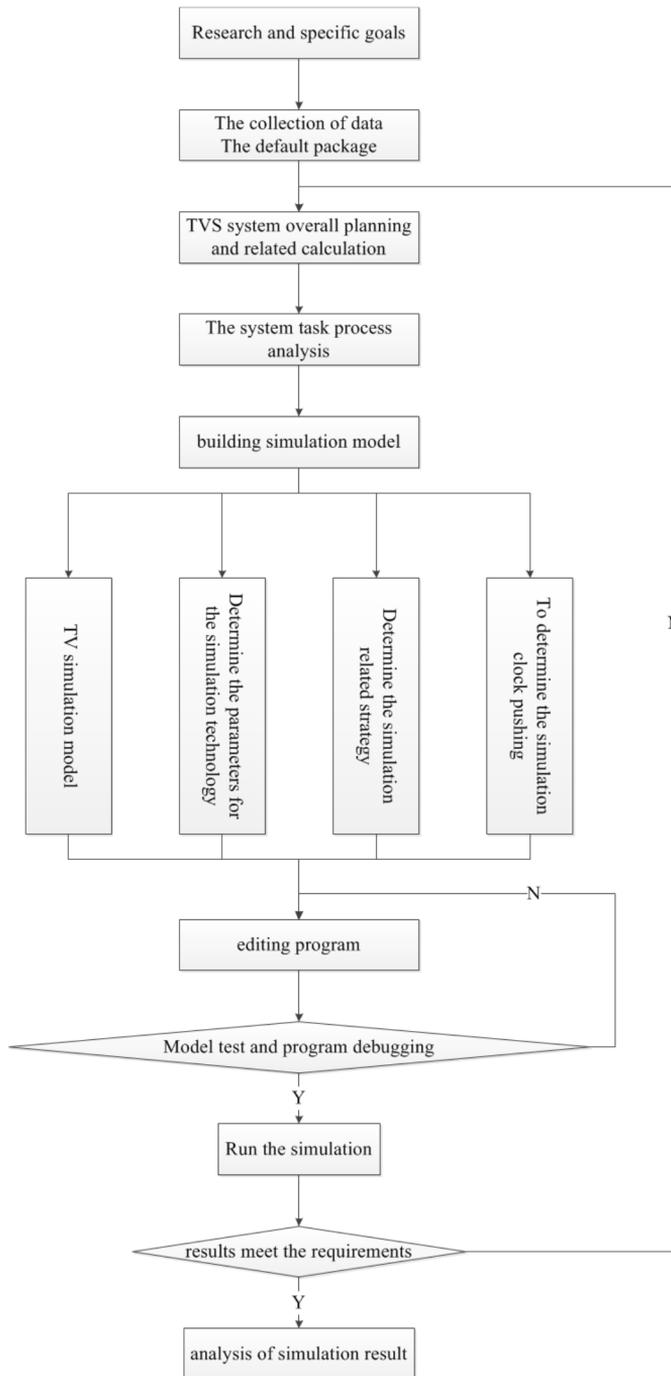


Fig. 1. Flow chart of TVS simulation system design

TVS system are shown in Table 2.

Table 1. Composition of TVS system

The number of transceiver stations	13
The number of empty storage stations	1
The number of junction	21
The number of total nodes	45
The number of TV $n$	7

Table 2. Performance parameters of TVS simulation system

Number of storage paths between two points $k$	5
Safe distance $d$	2 m
Deceleration distance $s$	5 m
TV normal running speed	0.5 m/s
TV speed regulation range	0–1 m/s

According to the TVS system model set up above, it is assumed that the logistics tasks of the hospital in a given period of time are shown in Table 3. Through simulation, the task allocation and operation of the entire system was achieved.

Table 3. Logistics task assignment table for TVS system

Vehicle No.		Logistic task (€)				
		Speed (m/s)	Start station	Passing station	End station	Priority
1	TV#1	0.5	ID05	ID01	ID06	I
2	TV#2	0.5	ID04	ID11	ID09	III
3	TV#3	0.5	ID03	ID10	ID08	II
4	TV#4	0	\	\	\	III
5	TV#5	0.5	ID12	\	ID05	II
6	TV#6	0.8	ID02	\	ID07	I
7	TV#7	0.5	ID03	\	ID07	I

Notes: The \ indicates no input information, and indicates that the carrier vehicle TV travels back and forth between the two sites. The lower the priority value is, the higher the priority is.

After entering the TVS simulation system, the initialization parameters were set up, then the simulation was run. The whole process was displayed in animation. The various states of collisions during the operation are shown in Figs. 2–8.

Figures 2 and 3 show the process of simulating the sending and receiving of an item in a TVS system from the sending station and the receiving station. Figure 4 is a simulation effect diagram for simulating the shortest path of a single vehicle.

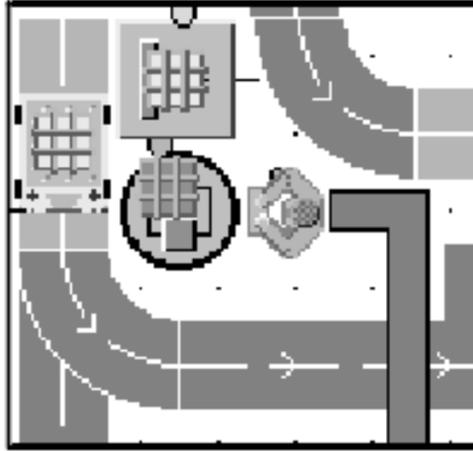


Fig. 2. Delivery at the transmitter station

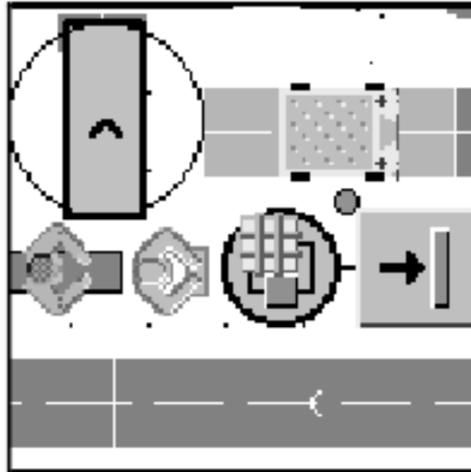


Fig. 3. Unloading at receiving station

Figures 5–8 simulates that in the running process of TVs, the cars can reasonably avoid a variety of possible conflicts. Figure 5 shows the changed path and waiting policy to avoid the first type of conflict. Figure 6 shows the change path policy taken to avoid the second type of conflict. Figure 7 shows the waiting and changed path policy to avoid the third type of conflict. And Fig. 8 shows the waiting strategy used to avoid the fourth type of conflict.

The simulation results show that the proposed scheduling planning algorithm and coordination strategy are feasible and correct. Using eM-Plant to simulate the actual situation of the system can provide good decision-making basis and technical support for the scheduling and planning of the TVS system and the reasonable selection of the driving route. Moreover, the 3D visual animation can provide more

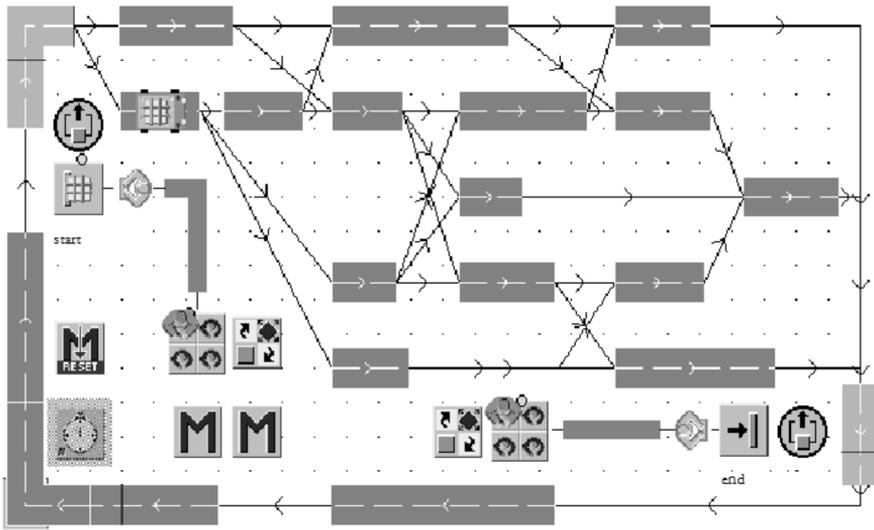


Fig. 4. Simulation effect diagram of the shortest path of the single vehicle

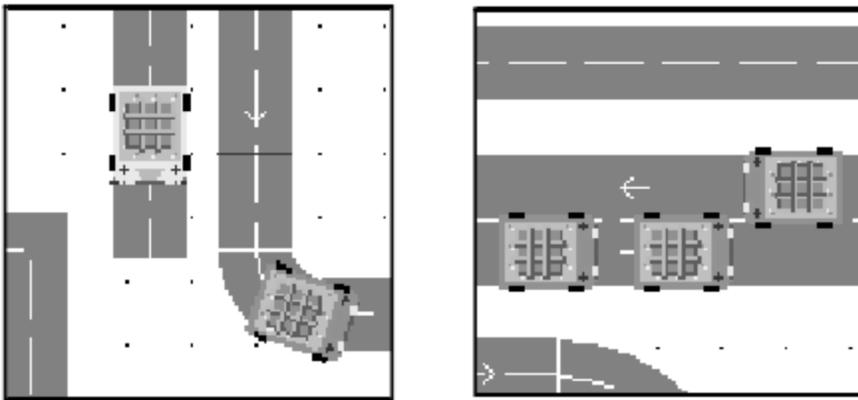


Fig. 5. The first kind of conflict

intuitive running effect for researchers, as shown in Fig. 9.

In addition, for the same logistics task, the general online scheduling strategy and the two-stage scheduling strategy were adopted to simulate the running process of two cars respectively. The resulting run time profile is shown in Fig.10.

In the above figure, the generally online scheduling strategy is used in the left figure, and the two-stage scheduling strategy is used in the right picture. In the same picture, the left histogram shows the first car, and the second car is shown on the right side of the histogram. The meaning of the four time distributions is as follows: red means the car is in the running state. Green means that the car is in idle state, which mainly refers to the time of stay in the work site. Blue indicates the dwell time of the car in the empty storage area. Yellow means the car is in a wait state,

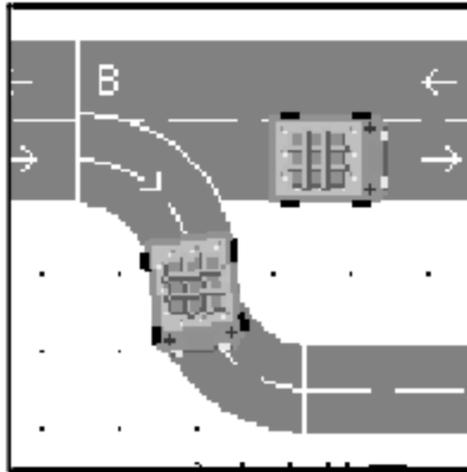


Fig. 6. The second kind of conflict

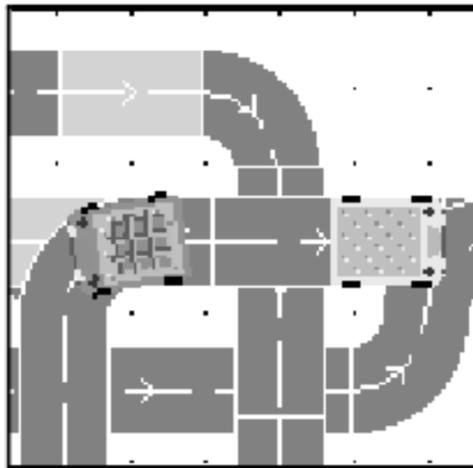


Fig. 7. The third kind of conflict

including the time of command waiting, collision waiting, and troubleshooting. As can be seen from Fig. 10, the two-stage scheduling strategy reduces the waiting time of the logistics car to a certain extent. This is because the collision problem of the car in the off-line phase has been considered by the two-phase control scheduling strategy, which can reduce the on-line computation time and the waiting time, so that the efficiency of the whole system is improved.

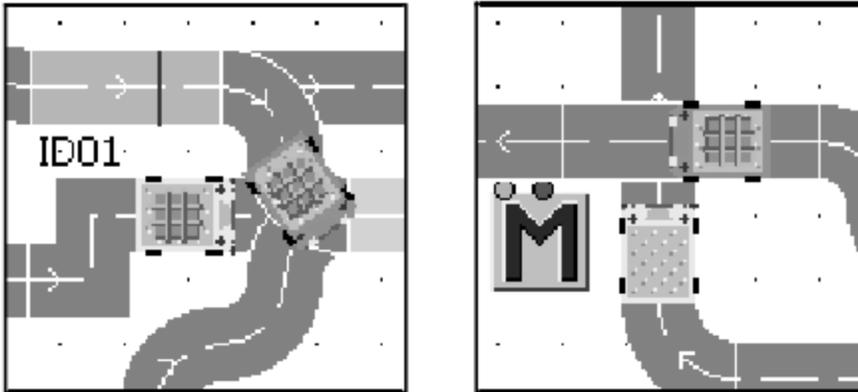


Fig. 8. The fourth kind of conflict

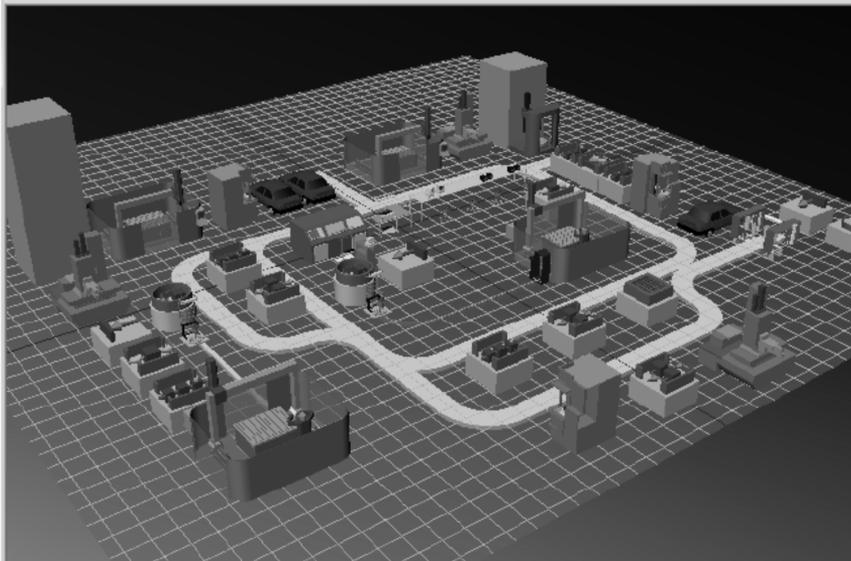


Fig. 9. 3D visual simulation model of TVS system

## 5. Conclusion

At present, the track logistics transmission system is used more and more widely, and the planning of the car is the key problem in the whole system, which is also the purpose of this paper. In this paper, the structure of TVS control system, the function of each part and requirements were studied. Based on the modeling of system path and stations, examples of practical application were given. The environment map modeling was completed, and the TVS scheduling system model was formed. The results of this research show that on the basis of single TV path planning, the dynamic path planning method combining Dijkstra algorithm with time window method can realize the path planning of TVS system, ensure the collision

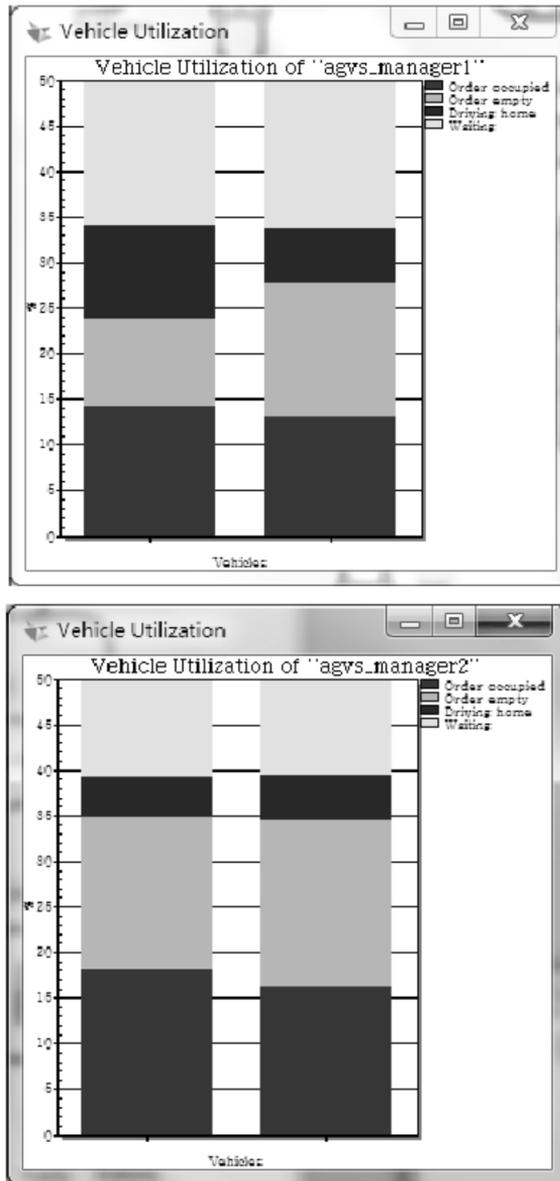


Fig. 10. Comparison of TVS running time under different strategies

free running between the cars, and solve the problem of online traffic control as well as realizing the collision free running TV with each other. In addition, a conflict coordination strategy based on speed regulation and path regulation was proposed, and the automatic collision between TV in TVS system was realized. It can provide a lot of convenience for the adjustment process of the system, and improve the

overall operation efficiency of the system. However, the research of this paper is still in the early stage. When the number of cars is large, the existing system will have some limitations, which needs further study.

## References

- [1] A. BEINAROVIČA, M. GOROBECs, A. LEVČENKOVs: *Algorithm of energy efficiency improvement for intelligent devices in railway transport*. Electrical, Control and Communication Engineering 10 (2016), No. 1, 29–34.
- [2] V. DIGANI, L. SABATTINI, C. SECCHI: *A probabilistic eulerian traffic model for the coordination of multiple AGVs in automatic warehouses*. IEEE Robotics and Automation Letters 1 (2016), No. 1, 26–32.
- [3] A. BEINAROVIČA: *Algorithm of efficiency improvement for railway transport using intelligent devices*. The Scientific Journal of Riga Technical University - Electrical, Control and Communication Engineering 10 (2016), 29–34.
- [4] A. GOUJARD: *Improving transport and energy infrastructure investment in Poland*. OECD Economics Department Working Papers (2016), No. 1302.
- [5] X. YANG, X. LI, B. NING, T. TANG: *A survey on energy-efficient train operation for urban rail transit*. IEEE Transactions on Intelligent Transportation Systems 17 (2016), No. 1, 2–13.
- [6] N. SMOLIC-ROČAK, S. BOGDAN, Z. KOVACIC, T. PETROVIC: *Time windows based dynamic routing in Multi-AGV systems*. IEEE Transactions on Automation Science and Engineering 7 (2010), No. 1, 151–155.
- [7] M. ESMAELIKIA, B. FAHIMNIA, J. SARKIS, K. GOVINDAN, A. KUMAR, J. MO: *A tactical supply chain planning model with multiple flexibility options: An empirical evaluation*. Annals of Operations Research 244 (2016), No. 2, 429–454.
- [8] C. TURNER, A. TIWARI, A. STARR, K. BLACKTOP: *A review of key planning and scheduling in the rail industry in Europe and UK*. Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 230 (2016), No. 3, 984–998.
- [9] Y. K. NA, S. Y. OH: *Hybrid control for autonomous mobile robot navigation using neural network based behavior modules and environment classification*. Autonomous Robots 15 (2003), No. 2, 193–206.
- [10] R. LAL, Y. KUMAR: *The analysis of transit assignment with variable transit price based on time schedule*. Advanced Materials Research 1030–1032, (2014), Chapter No. 8, 2260–2263.
- [11] R. SHI, S. NI: *Theory and application technology of railway intelligent transportation systems*. International Conference on Transportation Engineering (ICTE), 26–27 September 2015, Dalian, China, ASCE Library, (2015), 626–632.
- [12] M. G. PARK, J. H. JEON, M. C. LEE: *Obstacle avoidance for mobile robots using artificial potential field approach with simulated annealing*. Proc. IEEE International Symposium on Industrial Electronics (ISIE) Proceedings (Cat. No.01TH8570), 12–16 June 2001, Pusan, South Korea, IEEE Conference Publications, 3 (2001), 1530–1535.
- [13] N. WU, M. C. ZHOU: *Deadlock resolution in automated manufacturing systems with robots*. IEEE Transactions on Automation Science and Engineering 4 (2007), No. 3, 474–480.

Received June 29, 2017



# Image segmentation algorithm based on improved fuzzy clustering<sup>1</sup>

CHANGXING GENG<sup>2</sup>, PENG WANG<sup>2</sup>, PENGBO WANG<sup>2</sup>

**Abstract.** Image segmentation is a key part of image processing and it plays a very important role in image analysis and information application. The principle of image segmentation based on fuzzy theory was used as the basis in this context, and the application steps of fuzzy C mean clustering in image segmentation were studied. Aiming at the problems existing in the algorithm, the objective function was improved and a new algorithm model was established. The improved fuzzy clustering image segmentation algorithm was compared with the traditional algorithm through the experiment. The improved algorithm had better image segmentation effect. The feasibility of the improved algorithm was proved, which contributed to the research of image segmentation algorithm based on improved fuzzy clustering.

**Key words.** Image segmentation, fuzzy theory, fuzzy C means clustering, improved algorithm

## 1. Introduction

With the rapid development of science and technology, the storage and dissemination of information become cheaper and cheaper, which leads to explosive growth in the amount of information that people are exposed to in their daily lives. Images become the main carrier of information dissemination because of their intuitive, image, easy to understand, and rich in information [1]. About 70% come from images of all the information people acquire. However, it is very important to extract valuable information from the images obtained by various devices to meet the needs of different applications [2]. Useful information in an image is often concentrated in a particular area after researching people develop. The task of image segmentation is how to separate these regions from the image background effectively. Image segmentation can be divided into several regions according to the predefined features,

---

<sup>1</sup>This work is supported by Jiangsu Province Natural Science Foundation for Young Scholars, No. BK20140325.

<sup>2</sup>Robotics and Microsystems Centre, Soochow University, Suzhou 215021, China

which can be gray, color and texture [3]. With the development of image segmentation technology, it has become one of the most basic and important research contents in computer vision. It also occupies a very important position in the process of image analysis, processing and description. Image segmentation is based on the prior knowledge of the target and the background. Firstly, the target and background are marked and located in the image. And then, the targets that need to be identified are separated from the pseudo target according to these tags. Only in this way can the follow-up work of a series of information processing such as target identification and accurate positioning be realized. This result is also important because it can have a major impact on subsequent processing [4]. Image segmentation is very important for image recognition, tracking and understanding during the process of image analysis. The accuracy of image segmentation is very demanding in the process of image processing. Therefore, how to segment the required target from the background quickly and effectively has become the key to the research process [5].

## 2. State of the art

Zadeh proposed fuzzy set theory for the first time in 1965. Thus, fuzzy mathematics was created as a new discipline. Fuzzy set theory is extended on the basis of traditional set theory. For traditional set theory, an element belongs to a set, or does not belong to a collection. For a fuzzy set, each element belongs to a set to a certain extent, and can also belong to multiple sets at the same time [6]. Fuzzy sets theory can describe the fuzziness and randomness in human vision exactly. Fuzzy set theory has also been applied to various levels of pattern recognition. For example, the input pattern can be represented as a membership matrix for the feature layer (representing the extent of ownership of a given property). Membership values of fuzzy patterns can be used to provide information for representation and loss estimation at the classification layer [7]. Scholars have reformed some image segmentation algorithms in the course of research according to the characteristics of fuzzy sets. These algorithms can be divided into fuzzy threshold segmentation and fuzzy clustering segmentation. Among them, the fuzzy C means algorithm (FCM) is the most classical and widely used image segmentation algorithm [8]. The application of fuzzy theory to image segmentation is a classic example of the computer having some of the visual functions of human beings [9].

### 2.1. Methodology

Clustering is the separation and classification of things of similar nature. Clustering analysis is the classification of a given object on the basis of mathematical methods. A classical classification usually begins with a single factor or a few limited factors and uses experience or knowledge to classify things [10]. This category has a very clear line of categories and can cluster together the same thing effectively. However, with the cognition of people deeply, it is found that not everything can find a clear line of its classification, and many things are fuzzy boundaries in the image. For example, many regions is not clear enough [11]. Therefore, in order to better

classify these fuzzy boundaries, fuzzy mathematics arises at the historic moment. Its generation provides a mathematical basis for this kind of soft classification, which is also called fuzzy cluster analysis. Fig. 1 shows the location of image segmentation in image engineering.

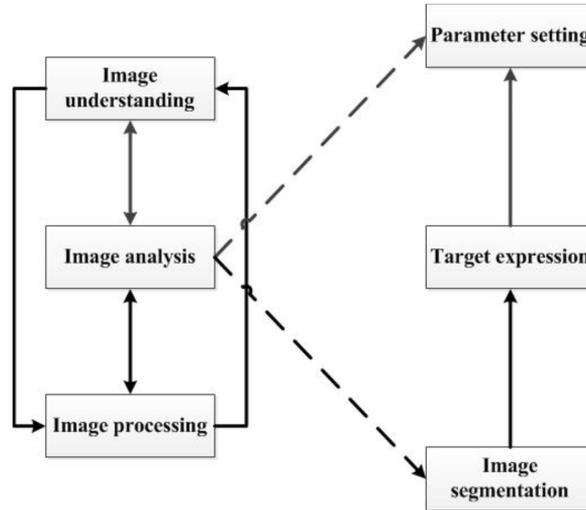


Fig. 1. Location of image segmentation in image engineering

Cluster analysis is a mathematical method of grouping similar data points into a class according to a specific criterion [12]. For the current theoretical development, the fuzzy C means clustering algorithm (Fuzzy C-Means, FCM) is one of the most famous fuzzy clustering methods. This algorithm was first proposed by Dunn. And later Bezdek improved it and gave Fuzzy C-Means Clustering an iterative optimization algorithm based on the least square method. By proving its convergence, it is proved that the algorithm can converge to an extremum [13]. Fig. 2 is the application of image processing technology.



Fig. 2. Application of image processing

The fuzzy C means clustering algorithm uses the double layer iterative method to obtain the extremum of the objective function. The inner iteration is used to correct the membership matrix and the clustering center. Inner iteration is a gradient

descent method essentially. The next step of the optimization direction is determined by calculating the gradient of the current state. The outer iteration is used to test the convergence condition to determine whether the iteration meets the requirement of convergence [14]. The membership degree of each element in each class can be obtained from the membership matrix after the completion of the iteration. The division of elements can be determined according to the degree of membership.

The fuzzy C means clustering algorithm defines the following form of objective function:

$$J = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m \|x_i - v_j\|^2 \quad (1)$$

The objective function satisfies the following constraints:

$$\sum_{j=1}^c \mu_{ij} = 1, \quad \forall i \in \{1, 2, \dots, n\} \quad (2)$$

Here,  $\mu_{ij}$  is the membership degree of the element  $I$  about class  $j$ . Symbol  $x_i$  denotes the characteristic vector of element  $i$ ,  $v_j$  is the clustering center of class  $j$ , and  $m$  is the weight coefficient. According to the validity criterion, the range of  $m$  is 1.5–2.5. In general,  $m = 2$ .

Lagrange conditional extremum method can be applied to deduce the corresponding iterative formula [15].

$$\begin{aligned} F &= \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m \|x_i - v_j\|^2 + \lambda \left( 1 - \sum_{j=1}^c \mu_{ij} \right) = \\ &= \lambda \left( 1 - \sum_{j=1}^c \mu_{ij} \right) + \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m d_{ij}^2. \end{aligned} \quad (3)$$

In order to maximize the function  $J$ , the first order necessary condition is as follows:

$$\frac{\partial F}{\partial \lambda} = 1 - \sum_{j=1}^c \mu_{ij} = 0, \quad (4)$$

$$\frac{\partial F}{\partial u_{ij}} = m (u_{ij})^{m-1} (d_{ij})^2 - \lambda = 0, \quad (5)$$

From (5), the following formula can be obtained

$$u_{ij} = \left[ \frac{\lambda}{m (d_{ij})^2} \right]^{\frac{1}{m-1}}. \quad (6)$$

Formula (6) is put in formula (4) to obtain that:

$$\sum_{j=1}^c \mu_{ij} = \sum_{j=1}^c \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} \left[\frac{1}{d_{ij}^2}\right]^{\frac{1}{m-1}} = \left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} \left\{ \sum_{j=1}^c \left[\frac{1}{d_{ij}^2}\right]^{\frac{1}{m-1}} \right\} = 1. \quad (7)$$

Thus, there is:

$$\left(\frac{\lambda}{m}\right)^{\frac{1}{m-1}} = \frac{1}{\sum_{j=1}^c \left[\frac{1}{d_{ij}^2}\right]^{\frac{1}{m-1}}}. \quad (8)$$

The formula (8) is fused together with (6) to get that

$$u_{ij} = \frac{1}{\sum_{i=1}^c \left[\frac{d_{ij}^2}{d_{il}^2}\right]^{\frac{1}{m-1}}}. \quad (9)$$

In view of the possibility that  $d_{ij}$  has taken as 0, the discussion is separated. For  $\forall i \in \{1, 2, 3, \dots, n\}$ , the definition of  $T_i$  and its complement set  $T_i^c$  are defined as

$$T_i = \{j \mid 1 \leq j \leq c, d_{ij} = 0\}, \quad (10)$$

$$T_i^c = \{1, 2, 3, \dots, c\} - T_i, \quad (11)$$

In order to obtain the minimum of objective function  $J$ , the value of  $u_{ij}$  should be

$$\begin{cases} \mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}}, T_i = \phi, \\ \mu_{ij} = 0, \forall j \in T_i^c, \sum_{j \in T_i} \mu_{ij} = 1, T_i \neq \phi. \end{cases} \quad (12)$$

The iterative formula of the cluster center  $v_j$  can be obtained by using a similar iterative method:

$$v_j = \frac{\sum_{i=1}^n \mu_{ij}^m x_j}{\sum_{i=1}^n \mu_{ij}^m}. \quad (13)$$

The fuzzy C means clustering algorithm is to approximate the extreme value progressively from a random initial value by the constant iteration of formula (12) and formula (13). The steps of the algorithm are described below briefly:

Step 1: The cluster number  $c$  and parameter  $m$  are set.

Step 2: The cluster center  $v_j$  is initialized.

Step 3: Repeat the following operations until the value of the target function obtained by the current two rounds of iterations satisfy  $|J_t - J_{t+1}| < \varepsilon$ . It can be considered that better clustering results have been obtained, thus stopping updating. Among them,  $\varepsilon$  is a small positive number, which is called tolerance error.

When the algorithm converges, all kinds of clustering centers and the membership values of each object can be obtained, and then the fuzzy clustering is completed. If it is necessary, we can also blur the fuzzy clustering results. That is to say,

fuzzy clustering is transformed into deterministic classification with certain criteria. The most common criterion is the maximum membership criterion. The object is partitioned into classes corresponding to the maximum of its membership.

Image segmentation using clustering analysis is a widely used and very important method in the field of image segmentation. Both color and gray scale images can be segmented by using them. Clustering is the aggregation of things of the same nature, and the differentiation of things that have significant differences in order to achieve the classification of things. Image segmentation is the process of classifying the pixels in an image according to their similarity. Naturally, people think of clustering analysis into image segmentation. The human eye has a certain degree of subjectivity and is often fuzzy to the division of the region. Therefore, it is more suitable to use fuzzy means for image segmentation. On the other hand, image segmentation often has the shortage of training samples and poor representation, and unsupervised learning methods can deal with such problems. They do not require the support of training samples during the process of processing. We can know that fuzzy clustering is a good choice from the needs of these two aspects. Fuzzy clustering is a powerful tool in the field of image segmentation.

The traditional FCM based image segmentation algorithm can get good segmentation results in more complete areas when the image has high signal-to-noise ratio and contrast ratio. Traditional FCM uses the gray value of pixels as the only feature to cluster without considering the abundant spatial location information, so there are some defects which are sensitive to noise. The objective function of FCM was improved, and the spatial location information was introduced into the calculation of the objective function so as to achieve better segmentation results in this paper.

The space constraint term is added on the basis of the traditional FCM's objective function, and the new objective function is defined as follows:

$$J = \sum_{i=1}^n \sum_{j=1}^c \left[ u_{ij}^m \left\| x_i - v_j \right\|^2 + \sum_{\substack{k \in N_i \\ k \neq i}} \frac{\left\| x_k - v_j \right\|^2}{\left\| x_k - v_j \right\|^2 + 1} u_{ij} (1 - u_{ij}) \right], \quad (14)$$

where  $N_i$  is the collection of field points of the pixel  $I$  in the upper form, which uses a rectangular, diamond or cross range of fields generally. The space constraint term introduced is to consider the domain of the pixel in the calculation of the objective function. To some extent, this suppresses the number of pixels in a domain belonging to multiple classes. The iterative formula of the new objective function is as follows:

$$v_j = \frac{\sum_{i=1}^n \mu_{ij}^m x_i}{\sum_{i=1}^n \mu_{ij}^m}, \quad (15)$$

and

$$u_{ij} = \sum_{l=1}^c \left( \frac{\left\| \frac{\|x_k - v_j\|^2 + \sum_{\substack{k \in N_i \\ k \neq i}}{\|x_k - v_j\|^2 + 1}}{u_{ij}} (1 - u_{ij})}{\left\| \frac{\|x_k - v_l\|^2 + \sum_{\substack{k \in N_i \\ k \neq i}}{\|x_k - v_l\|^2 + 1}}{u_{il}} (1 - u_{il})} \right\|} \right). \quad (16)$$

For the improved FCM algorithm, the space function is defined as follows:

$$h_{ij} = \sum_{k \in N_i} u_{jk}, \quad (17)$$

where  $N_i$  is the domain collection of pixel  $i$  in the formula. This function reflects the degree of membership of the domain pixels of the pixel  $i$ . The new membership is defined as follows:

$$u_{ij}^* = \frac{(u_{ij})^p (h_{ij})^q}{\sum_{k=1}^c (u_{ik})^p (h_{ik})^q}. \quad (18)$$

In formula (18),  $p$  and  $q$  are adjustment parameters to control the proportion of the original membership and space functions. If the pixel  $i$  and its domain pixels belong to the same class, the spatial function only increases the original membership degree. If the pixel  $i$  and its domain pixels do not belong to the same class, the spatial function will suppress the original membership degree. Thus, the pixels gathered together belong to the same class as much as possible, and the noise can be suppressed. The new clustering center functions are

$$v_j = \frac{\sum_{i=1}^n (u_{ij}^*)^m x_k}{\sum_{i=1}^n (u_{ij}^*)^m}. \quad (19)$$

The new target function is

$$J = \sum_{j=1}^c \sum_{i=1}^n (u_{ij}^*)^m \|x_i - v_j\|^2 \quad (20)$$

The specific algorithm steps are as follows:

Step 1: Initial parameters are set, including cluster number  $c$ , initial clustering center  $v_j$ , tolerance error  $e$ , and domain size and corresponding  $t$  values.

Step 2: The FCM is used to segment the image under the maximum membership constraint.

Step 3: The class of pixels is modified.

Step 4: The original image is chosen to smooth the noise combining the generic information of the domain pixels.

Step 5: FCM is used to segment the smoothed image for two times under the constraint of the maximum membership degree.

### 3. Result analysis and discussion

The following is the experimental image segmentation results.

The rice grain diagram was used as the experimental image. When the clustering was carried on, the clustering number was  $c = 2$  and the tolerance error was  $e = 0.01$ . The initial clustering center was 0, 255. The weighting function took  $m = 2$ . Figure 3 shows the rice image from left to right. A Gauss noise figure with a mean value of 0, a variance of 400, and a grain of rice with 5% salt and pepper noise were added. Figure 4 shows the first segmentation result of the Gauss noise map, the selective equilibrium results and the two segmentation results. Figure 5 shows a segmentation result of a salt pepper noise map, a selective smoothing result and a second segmentation result. The following table is a comparison of the indexes in the segmentation process.

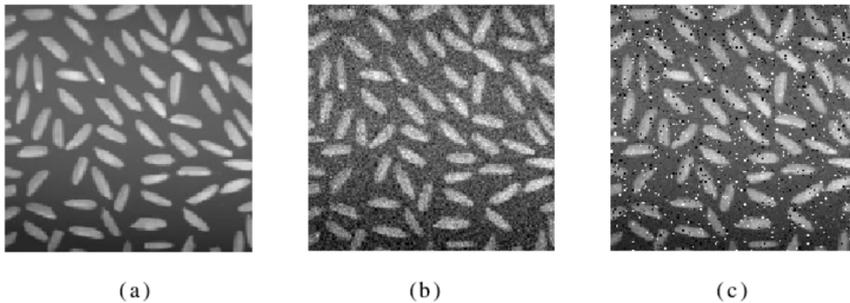


Fig. 3. Artwork, Gauss noise map, and salt and pepper noise map

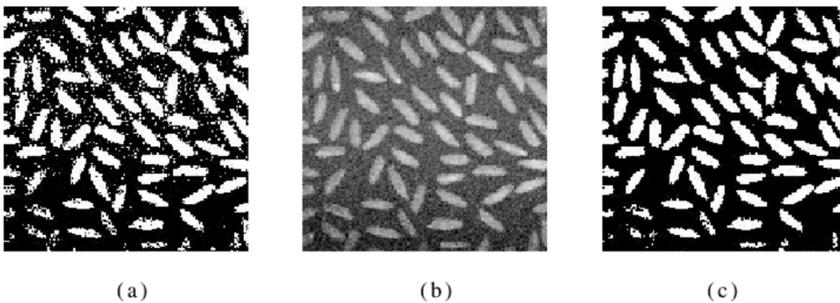


Fig. 4. Intermediate results and final results of the Gauss noise map

It can be seen from the previous table that selective smoothing does have an inhibitory effect on noise, especially for salt and pepper noise. Selective smoothing not only improves the signal to noise ratio of the image, but also makes the clustering center of the smoothed image closer to the cluster center of the original image, thus reducing the influence of noise effectively.

The experimental image used in this paper was a synthetic map with a resolution of  $300 \times 300$ . The hardware environment of the simulation experiment was AMD

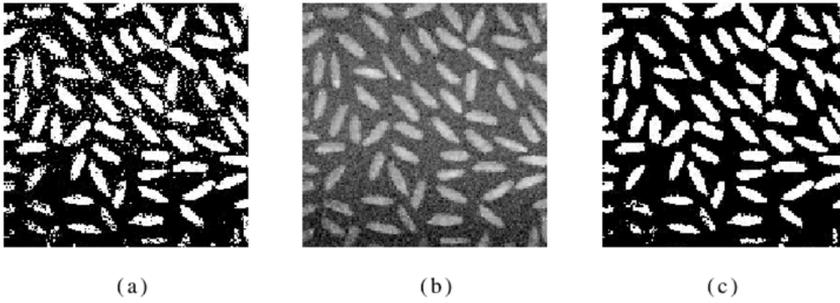


Fig. 5. Intermediate results and final results for segmentation of salt and pepper noise maps

Athlon (tm) 64 X2 Dual Core, Processor4400++. The CPU main frequency was 2.30 GHz, and the memory was 2 GB. The programming environment was Matlab7.0. Among them, the Gauss noise map was added to the original image with a mean value of 0. The variance was 400 of the Gauss noise. Salt and pepper noise map was based on the original added 5 % salt and pepper noise. The time consumed by segmentation and the correct segmentation rate are shown in Table 2. Comparison of correct segmentation rates is in Table 3.

Table 1. Comparison of the indexes in the segmentation process

	Gauss diagram	Salt pepper diagram
Artwork clustering center	{84.964,167.66}	{84.964,167.66}
Noise map clustering center	{80.281,164.8}	{81.26,175.08}
Smoothed cluster centers	{80.799,167.96}	{83.212,169.6}
Signal to noise ratio before smoothing	33.7763	15.2494
Signal-to-noise ratio after smoothing	46.0306	59.4954

Table 2. Comparison of running time

	FCM	KFCM	Paper algorithm
Gauss noise map	83.80 s	83.00 s	169.56 s
Salt and pepper noise map	85.19 s	84.25 s	179.25 s

Table 3. Comparison of correct segmentation rates

	FCM	KFCM	Paper algorithm
Gauss noise map	97.2 %	98.4 %	100 %
Salt and pepper noise map	96.3 %	98.1 %	99.8 %

It can be seen from the two objective indexes of segmentation accuracy and

time consuming that FCM and KFCM were the least time-consuming and sensitive to noise. The segmentation algorithm took about 2 times as much as FCM, but the noise suppression was obvious, which improved the robustness of the algorithm greatly. This type belongs to the FCM image segmentation method of correction and spatial constraint, which can modify the class of pixels according to class dispersion. And then, the image is selectively smoothed according to the class of neighborhood pixels, and finally the two segment of the modified image is segmented. Experimental results show that the proposed algorithm has good anti-noise capability. In addition, the complexity of distance and membership calculation formula is not increased when spatial information is introduced. Therefore, the operation is the same as the traditional FCM, but the segmentation effect is better than the traditional FCM algorithm.

## 4. Conclusion

One of the basic problems in computer vision is image segmentation. The research of image segmentation has promoted the development of computer vision greatly, and has also promoted the development of intelligent information processing technology. There is a lot of uncertainty about the image itself. As a theory that can define fuzzy boundaries, fuzzy theory can describe the uncertainty of images effectively. Many researches have focused on the application of fuzzy clustering algorithm in image segmentation in recent years. The research status of image segmentation technology was explained and the development process of fuzzy theory was introduced in this paper. The fuzzy C means clustering algorithm was analyzed in detail. Aiming at the problem that the fuzzy C means clustering algorithm has poor noise immunity function, an improved fuzzy clustering algorithm was proposed. The algorithm presented in this paper performed FCM image segmentation by changing the objective function and based on domain pixel class constraint, which solved the problem of introducing spatial information effectively. The traditional algorithm was compared with the improved algorithm through the contrast experiment. It was proved that the algorithm can suppress the noise and improve the robustness of the algorithm without changing the time complexity by defining the objective index, including the running time and the segmentation accuracy. To some extent, the research in this paper can promote the theoretical research of image segmentation algorithm based on improved fuzzy clustering.

## References

- [1] K. S. CHUANG, H. L. TZENG, S. CHEN, J. WU, T. J. CHEN: *Fuzzy c-means clustering with spatial information for image segmentation*. Computerized Medical Imaging and Graphics 30 (2006), No. 1, 9–15.
- [2] S. CHEN, D. ZHANG: *Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure*. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34 (2004), No. 4, 1907–1916.

- [3] Y. HAN, P. SHI: *An improved ant colony algorithm for fuzzy clustering in image segmentation*. *Neurocomputing* 70 (2007), Nos. 4–6, 665–671.
- [4] D. L. PHAM, J. L. PRINCE: *An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities*. *Pattern Recognition Letters* 20 (1999), No. 1, 57–68.
- [5] L. O. HALL, A. M. BENSaid, L. P. CLARKE, R. P. VELTHUIZEN, M. S. SILBIGER, J. C. BEZDEK: *A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain*. *IEEE Transactions on Neural Networks* 3 (1992), No. 5, 672–682.
- [6] S. SHEN, W. SANDHAM, M. GRANAT, S. STERR: *MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization*. *IEEE Transactions on Information Technology in Biomedicine* 9 (2005), No. 3, 459–467.
- [7] M. GONG, Y. LIANG, J. SHI, W. MA, J. MA: *Fuzzy c-means clustering with local information and kernel metric for image segmentation*. *IEEE Transactions on Image Processing* 22 (2013), No. 2, 573–584.
- [8] Z. YANG, F. L. CHUNG, W. SHITONG: *Robust fuzzy clustering-based image segmentation*. *Applied Soft Computing* 9 (2009), No. 1, 80–84.
- [9] X. GAO, J. LI, H. JI: *A multi-threshold image segmentation algorithm based on weighting fuzzy C-means clustering and statistical test*. *Acta Electronica Sinica* 32 (2004), No. 4, 661–664.
- [10] M. FOROUZANFAR, N. FORGHANI, M. TESHNEHLAB: *Parameter optimization of improved fuzzy c-means clustering algorithm for brain MR image segmentation*. *Engineering Applications of Artificial Intelligence* 23, (2010), No. 2, 160–168.
- [11] Y. CHEN, J. Z. WANG: *A region-based fuzzy feature matching approach to content-based image retrieval*. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002), No. 9, 1252–1267.
- [12] P. VASUDA, S. SATHEESH: *Improved fuzzy C-means algorithm for MR brain image segmentation*. *International Journal on Computer Science & Engineering* 2 (2010), No. 5, paper 1713.
- [13] W. CAI, S. CHEN, D. ZHANG: *Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation*. *Pattern Recognition* 40 (2007), No. 3, 825–838.
- [14] B. N. LI, C. K. CHUI, S. CHANG, S. H. ONG: *Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation*. *Computers in Biology and Medicine* 41 (2011), No. 1, 1–10.
- [15] S. DAS, S. SIL: *Kernel-induced fuzzy clustering of image pixels with an improved differential evolution algorithm*. *Information Sciences* 180 (2010), No. 8, 1237–1256.

Received June 29, 2017



# Research on an improved UWB space location method

CHENGBO HU<sup>1,4</sup>, XIAOBO WANG<sup>1</sup>, YONGLING LU<sup>1</sup>,  
WEIHUA CHENG<sup>2</sup>, WEI DAI<sup>3</sup>

**Abstract.** Because of its strong anti-multipath capability and other characteristics, UWB technology is widely used in positioning. While using UWB, unstable phenomenon of the positioning results still remained. In this paper, we proposed an improved UWB space location method. We first proposed a distance optimization method and then introduced the distance intersection of space method which uses three base stations to solve spatial location. The experimental results show that the proposed method is simple and easy to implement. Besides, it is with small computation cost and high accuracy of positioning. At the same time, because the method only needs three base stations to measure the three-dimensional coordinate of the tested point, it can save the hardware cost to a certain extent. Finally, this method is suitable for warehouse location and other scenes.

**Key words.** UWB, ranging optimization, distance intersection of space method.

## 1. Introduction

At present, location information has become an important basis for the development of all walks of life. GNSS (GPS, GLONASS, BDS, Galileo) ensures the efficient and convenient realization of outdoor positioning. With the development of wireless communication technology, emerging technologies, such as Bluetooth, WiFi, ZigBee and UWB, have played an important role in indoor positioning, and they make up for the shortcomings that GNSS cannot work in the indoor and other complex environments. UWB [1] is a low-power, low-cost but high-speed wireless communication technology. Recently, UWB positioning technology has been extensively researched by researchers. Normal operating frequency of UWB is between 3.0 GHz to 10.6 GHz [2]. UWB is characterized by not using carrier communication, but with very short time interval (nanosecond or less than nanosecond time interval) of the baseband

---

<sup>1</sup>Jiangsu Electric Power Company Research Institute, Nanjing, 211167, China

<sup>2</sup>Jiangsu Power Info-tech Co. Ltd., Nanjing, 211167, China

<sup>3</sup>Jiangsu BDS Application Industry Institute, Nanjing, 211167, China

<sup>4</sup>Corresponding author

narrow pulse communication, with penetrating power, anti-multipath effect of anti-interference ability outstanding. Especially, in the metal or liquid environment that have great impact on signal attenuation, UWB plays a stronger performance than other wireless positioning technology [3]. In view of the outstanding characteristics of UWB technology, many scholars use UWB technology for positioning research. The key to achieve high-precision positioning is to obtain high-precision ranging data through UWB. According to the characteristic parameters used in UWB ranging, UWB ranging methods include RSS (Received Signal Strength), AOA (Angle of Arrival), TOA (Time of Arrival) TDOA (Time Difference of Arrival) measurement [4]. In [5], AOA, TOA, and TDOA are considered to be the focus of attention. Typical algorithms are based on the arrival time of UWB direct path components. Works [6–7] studied the direct path, and analyzed the multi-path resolution. The range data were divided into LOS (Line of Sight) and NLOS (Not Line of Sight), two types. In order to get high accuracy positioning results, many scholars have studied the algorithm of coordinate solution. Currently, Chan algorithm [8], Fang algorithm [9] and least square algorithm [10] are usually used. These algorithms are based on the improvement of the algorithm itself, and they do not take the optimal data type into account. In addition, these algorithms do not make full use of data for the location of quasi-real-time applications. In this paper, a method of 2-second data preprocessing of continuous time data is proposed. The method makes full use of the high-frequency characteristics of UWB technology. The method is simple and feasible and easy to implement. At the same time, the method can determine the three-dimensional coordinates of the tags with only three base stations, thus saving the hardware cost of the UWB module. Besides, it is applicable to the application scenario where the real-time requirement is not very high.

## 2. Modeling of underwater robot kinematics

### 2.1. Positioning system

United States Time Domain P440 module are used as both base station and tag in positioning platform. This product is with superior performance, and it has a good anti-multipath effect, and can achieve high-precision ranging. Between the module and the module can achieve two-way communication ranging. At the same time, the operating frequency of the module can also be artificially set. Three modules are used as the base station of the UWB positioning platform, one module is used as the label of the UWB positioning platform. Thus three-dimensional coordinates of the label are obtained by the three base stations.

### 2.2. Algorithmic flow

Because the module can realize the two-way communication distance measurement (TW-TOF), the range type of the positioning platform can be obtained by the formula

$$\text{Amount} = C_n^m, \quad (1)$$

where  $m$  is the total number of modules in the positioning platform, ( $m = 4$  in this platform), and  $n = 2$  because of two-way ranging. Then the normal range of the platform type is 6.

The flow of the algorithm is shown in Figure 1.

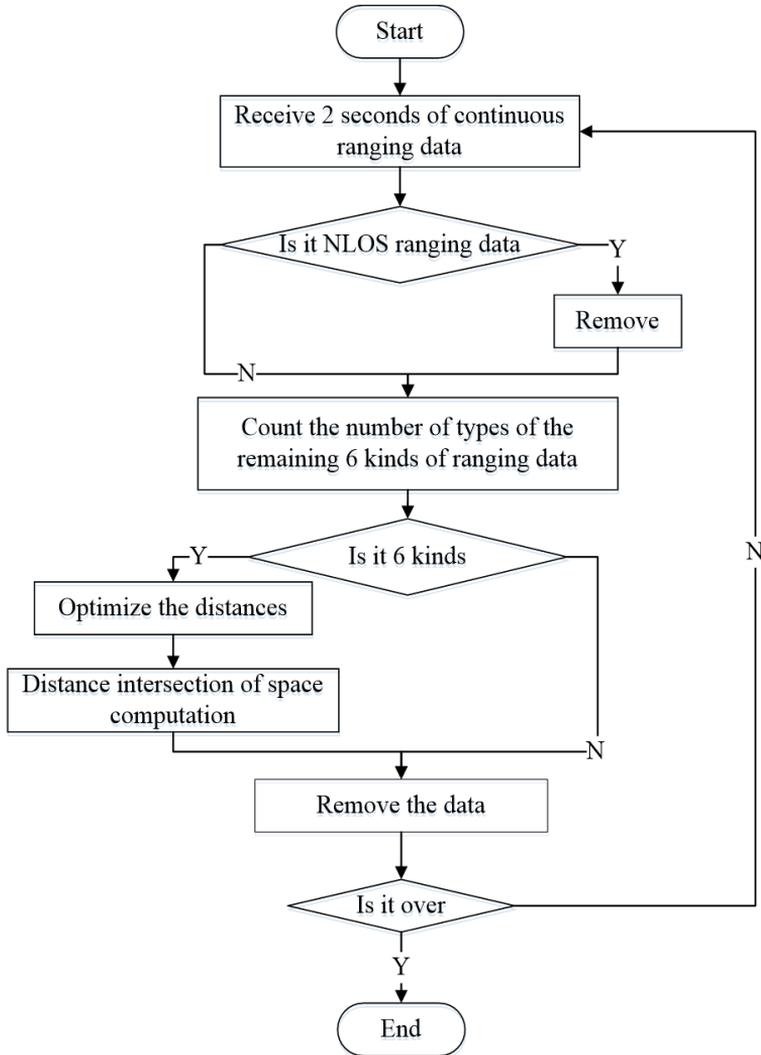


Fig. 1. Algorithm flowchart

First, the positioning platform receives the ranging data for two consecutive seconds, the interval time of the module is set to 10 ms. Thus the period of the six kinds of ranging polling is 60 ms, that is, 16.67 Hz. Therefore, about 33 ranging can be obtained theoretically for 2 seconds.

Then, it is judged whether or not NLOS type data is received for the received

ranging data. As the UWB actual communication can be effected by the external environment and its own stability. There are two types of data, LOS and NLOS. In the NLOS state, the TOA and TDOA data will produce the error of excessive time delay, so the direct use of the NLOS ranging data for spatial location will bring large error of positioning error [11]. If NLOS data exists, the NLOS data should be discarded.

After all processing is complete, the remaining data are all LOS type data. If there are 6 kinds of remaining data, there are 6 ranging information in the positioning platform, which can be used to optimize the edge distance, and then used in the follow-up space measurement; if the type is less than 6, it indicates that at least one kind of ranging type data in the ranging data is completely removed during the process of removing the NLOS type data, in which case the distance intersection of space can not be performed. Thus the data of the continuous 2 seconds is deleted. Then determine whether it is the end of the data. If not, continue to measure the range of data processing, otherwise go the end of the process.

### 3. Algorithm description

#### 3.1. NLOS type range determination

As described above, NLOS-type ranging data need to be eliminated because the error of the NLOS-type ranging value is large and it can leads a low positioning accuracy.

Based on the premise that the standard deviation of the NLOS data is greater than the standard deviation of the LOS data, the standard deviation of the actual range data is compared with the standard deviation of the LOS environment to determine whether the measurement period contains NLOS error [7].

At time  $t_i$ , the distance from the base station A to the tag is as shown in the formula

$$r_m(t_i) = r_m^0(t_i) + n_m(t_i) + \text{NLOS}_m(t_i), \quad (2)$$

where  $r_m^0(t_i)$  is the measurement distance determined by the LOS signal,  $n_m(t_i)$  is the system noise caused by the ranging error and  $\text{NLOS}_m(t_i)$  is the NLOS error distance.

Since  $\text{NLOS}_m(t_i)$  is always a nonnegative random variable, let  $\text{NLOS}_m(t_i)$  be a range of  $0 \leq \text{NLOS}_m(t_i) \leq \beta_m$ . And because  $n_m(t_i)$  obeys the normal distribution whose average value is 0, its value range is  $-\alpha_m \leq n_m(t_i) \leq \alpha_m$ , and then error scope of NLOS type range survey data is shown in the formula.

$$-\alpha_m \leq \text{NLOS}_m(t_i) + n_m(t_i) \leq \alpha_m + \beta_m. \quad (3)$$

Compared to the LOS type, the error becomes significantly larger, that is,  $\alpha_{\text{mNLOS}} < \alpha_{\text{mLOS}}$ . Then NLOS data can be judged.

As P440 module outputs range measurement data, it also outputs data type of the range data. Based on this feature, this positioning algorithm can extract the

data to solve.

### ***3.2. Optimization of measured distance***

The purpose of optimization of measured distance is to further eliminate NLOS data and improve the accuracy of ranging. The data collected for 2 consecutive seconds are processed to obtain the optimal distance in this period. The idea is as follows:

a) Obtain the data median  $N_1$  and average  $N_2$ . As the P440 module has a high ranging accuracy, the nominal ranging accuracy can reach 2 cm, and the data subject to normal distribution. So when the sample is sufficient,  $N_1$  is theoretically the same as  $N_2$ . Due to there are system errors during the actual processing process, there is some differences between  $N_1$  and  $N_2$ .

b) Judging by experience, the difference between  $N_1$  and  $N_2$  is set 0.01 m, that is 1 cm. If  $|N_1 - N_2| < 0.01$ , then the data is considered good, there is no large error value, the data obeys the normal distribution. We take  $N_1$  as the optimal distance value. If it is not satisfied, it is assumed that there is a ranging value containing the NLOS type value in the set of data.

Pauta Criterion is usually used in statistics to remove outliers [12]. It takes three times as error correction threshold of discrimination. Therefore, this paper, based on Pauta Criterion, proposed a method using  $N_1$  as the initial true value. In this way,  $N_1 \pm 3\sigma$  (standard deviation) is determined as decision interval of Pauta Criterion. The method uses the data in the interval to get its average value and takes it as the optimized ranging value.

### ***3.3. Optimization of measured distance***

At present, many scholars focus on the positioning of UWB in two-dimensional plane positioning [13], they lack the three-dimensional positioning of the label research.

The usual coordinates are solved by least squares [14]. The optimization of measured distance uses distance intersection of space method [15], which can be used to derive the three-dimensional coordinate of the label with three base stations.

The diagram of location is shown in Fig. 2.

In Fig. 2, A, B and C are the three base stations, P is the pending label. The relationship between A, B, C and P fits the law of the right hand, that is, when the right thumb is pointing to P, the other four fingers of the natural bending direction is A, B, C. It is to be noted that the three base stations cannot be located on the same straight line.

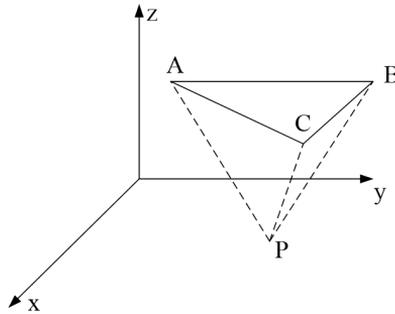


Fig. 2. Schematic diagram of location

Based on the spatial geometry, the volume formula of tetrahedron P-ABC is

$$V = \frac{1}{6} \begin{vmatrix} x_P - x_A & y_P - y_A & z_P - z_A \\ x_P - x_B & y_P - y_B & z_P - z_B \\ x_P - x_C & y_P - y_C & z_P - z_C \end{vmatrix} = \frac{1}{6} S_1 S_2 S_3 \begin{vmatrix} \cos \alpha_A & \cos \beta_A & \cos \gamma_A \\ \cos \alpha_B & \cos \beta_B & \cos \gamma_B \\ \cos \alpha_C & \cos \beta_C & \cos \gamma_C \end{vmatrix}, \quad (4)$$

where  $\cos \alpha_A$ ,  $\cos \beta_A$  and  $\cos \gamma_A$  represent the direction cosines of the vector AP,  $\cos \alpha_B$ ,  $\cos \beta_B$ ,  $\cos \gamma_B$  represent the direction cosines of the vector BP, and  $\cos \alpha_C$ ,  $\cos \beta_C$ ,  $\cos \gamma_C$  represent the direction cosines of the vector CP. Using  $K$  to represent the direction cosine matrix in formula (4), we obtain

$$V = \frac{1}{6} S_1 S_2 S_3 |K|. \quad (5)$$

We can obtain  $N$  as

$$N = |KK^T| = \sin^2 \varphi_{AB} + \sin^2 \varphi_{AC} + \sin^2 \varphi_{BC} + 2 \cos \varphi_{AB} \cos \varphi_{AC} \cos \varphi_{BC} - 2, \quad (6)$$

where:

$$|K| = \sqrt{\sin^2 \varphi_{AB} + \sin^2 \varphi_{AC} + \sin^2 \varphi_{BC} + 2 \cos \varphi_{AB} \cos \varphi_{AC} \cos \varphi_{BC} - 2} \quad (7)$$

Among them,  $\varphi_{AB}$ ,  $\varphi_{AC}$  and  $\varphi_{BC}$  are the angles between vectors PA and PB, vectors PA and PC, vectors PB and PC. The values of trigonometric function can be solved from triangles PAB, PAC and PBC.

Also define the vectors

$$\begin{aligned} \vec{AB} &= (x_B - x_A, y_B - y_A, z_B - z_A) = (X_B, Y_B, H_B), \\ \vec{AC} &= (x_C - x_A, y_C - y_A, z_C - z_A) = (X_C, Y_C, H_C), \\ \vec{AP} &= (x_P - x_A, y_P - y_A, z_P - z_A) = (X_P, Y_P, H_P). \end{aligned} \tag{8}$$

From the product of the inner product and the mixed product in formula (8), we can get

$$\left\{ \begin{aligned} &\left| \begin{array}{cc} Y_B & H_B \\ Y_C & H_C \end{array} \right| X_P + \left| \begin{array}{cc} H_B & X_B \\ H_C & X_C \end{array} \right| Y_P + \left| \begin{array}{cc} X_B & Y_B \\ X_C & Y_C \end{array} \right| H_P = M_1, \\ &X_B X_P + Y_B Y_P + H_B H_P = M_2, \\ &X_C X_P + Y_C Y_P + H_C H_P = M_3. \end{aligned} \right. \tag{9}$$

where  $M_1 = S_1 S_2 S_3 |K|$ ,  $M_2 = (D_{AB}^2 + S_A^2 - S_B^2)/2$ ,  $M_3 = (D_{AC}^2 + S_A^2 - S_C^2)/2$ , so the formula (9) can be solved, and we get

$$\left\{ \begin{aligned} X_P &= \frac{1}{\left| \begin{array}{cc} X_B & Y_B \\ X_C & Y_C \end{array} \right|} \left[ \left| \begin{array}{cc} Y_B & H_B \\ Y_C & H_C \end{array} \right| H_P - \left| \begin{array}{cc} Y_B & M_2 \\ Y_C & M_3 \end{array} \right| \right], \\ Y_P &= \frac{1}{\left| \begin{array}{cc} X_B & Y_B \\ X_C & Y_C \end{array} \right|} \left[ \left| \begin{array}{cc} H_B & X_B \\ H_C & X_C \end{array} \right| H_P - \left| \begin{array}{cc} M_2 & X_B \\ M_3 & X_C \end{array} \right| \right], \\ H_P &= \frac{M_1 \left| \begin{array}{cc} X_B & Y_B \\ X_C & Y_C \end{array} \right| + \left| \begin{array}{cc} Y_B & H_B \\ Y_C & H_C \end{array} \right| \left| \begin{array}{cc} Y_B & M_2 \\ Y_C & M_3 \end{array} \right| + \left| \begin{array}{cc} H_B & X_B \\ H_C & X_C \end{array} \right| \left| \begin{array}{cc} M_2 & X_B \\ M_3 & X_C \end{array} \right|}{\left| \begin{array}{cc} X_B & Y_B \\ X_C & Y_C \end{array} \right|^2 + \left| \begin{array}{cc} Y_B & H_B \\ Y_C & H_C \end{array} \right|^2 + \left| \begin{array}{cc} H_B & X_B \\ H_C & X_C \end{array} \right|^2}. \end{aligned} \right. \tag{10}$$

Then the three-dimensional coordinates of P point are  $x_P = X_P + x_A$ ,  $y_P = Y_P + y_A$ ,  $z_P = H_P + z_A$ .

### 4. Example analysis

Using the positioning method shown in Fig. 2. In the test environment, the label is placed on the metal surface, and there are fewer other obstacles in the environment. The improved spatial localization algorithm is validated and analyzed by MATLAB software.

#### 4.1. Analysis of measured distance optimization

We analyzed UWB modules' ranging data for two seconds. The results are shown in Fig. 3. In order to facilitate the distinction, the line which means not optimized is bold in Fig. 3.

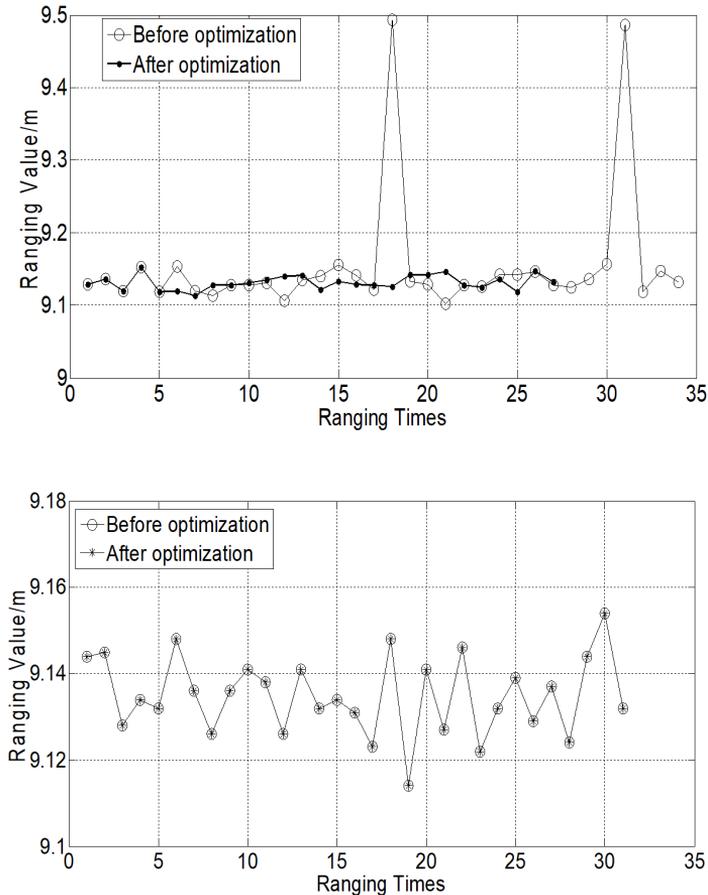


Fig. 3. Optimization results of edge detection

From Fig. 3, upper part, it can be seen that measured distance optimization method can eliminate the ranging value with obvious errors. All the ranging values after the elimination are located between 9.1 m and 9.15 m. Excluded values usually contain NLOS type data and other error effects. In addition, we can see that measured distance optimization method eliminates a total of seven ranging data, including two larger data error which can be visually seen from the figure. After the removal, overall data is very concentrated.

It can be seen from Fig. 3, bottom part, that the data obtained by the test itself is more concentrated, located between 9.11 m and 9.16 m. Standard deviation in Fig. 3,

upper part, is 0.0864 m, while it the bottom part it is 0.0097 m. So the measured distance optimization method can effectively improve the distance accuracy.

In addition, from both parts of Fig. 3 it can be found, because the P440 module has been processed ranging data itself, that if UWB measured turned failure then it will return 0, and the 0 data will not be considered. That is why there is a difference in the number of distance data from the two tests.

#### *4.2. Analysis of 3D positioning results*

Four known points are selected for experiment, each known point is continuously observed for a period of time. The distance intersection of space algorithm of the optimal distance is used to solve the three-dimensional coordinates for two consecutive seconds. The errors of these coordinates and the known precise coordinates are solved, as shown in Table 1.

Table 1. Test results of insulated resistance value (k $\Omega$ )

Point number	Axis	Mean value (m)	True value (m)	Standard deviation (cm)	Average point error (cm)
1	X	3.7262	3.749	0.0039	6.11
	Y	1.6907	1.635	0.0072	
	Z	1.6593	1.670	0.0061	
2	X	5.6325	5.685	0.0042	7.67
	Y	1.6819	1.637	0.0101	
	Z	1.7074	1.674	0.0082	
3	X	3.7777	3.751	0.0049	5.77
	Y	3.4606	3.439	0.0110	
	Z	1.6216	1.668	0.0183	
4	X	5.7034	5.682	0.0032	3.91
	Y	3.4267	3.444	0.0033	
	Z	1.6432	1.671	0.0048	

It can be seen from Table 1 that the average of several measurements is close to the true value, and the standard deviation of each solved value is small, indicating that each solved results has a low degree of dispersion, and results can be well gathered near the true value.

Limited to space, take point 1 as an example, the three-axis error and point error shown in Fig. 4. It can be seen from Fig. 4 that the three-axis error and point error are not randomly distributed along the zero, there is a certain systematic error that may be caused in the optimization algorithm. At the same time, it is found that the

point error is in the range of 8 cm, and the precision is high.

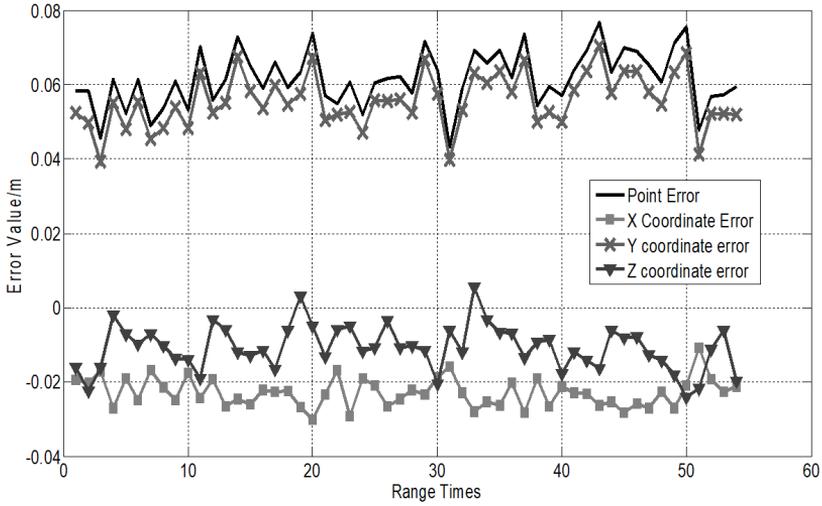


Fig. 4. Three-axis error and point error

Point 1 was continuously tested 41-times, and the results are shown in Fig. 5. It is considered that the positioning accuracy is high,  $x$ -axis value is between 3.71 m and 3.74 m,  $y$ -axis is between 1.67 m and 1.71 m, and  $z$ -axis is between 1.64 m and 1.68 m.

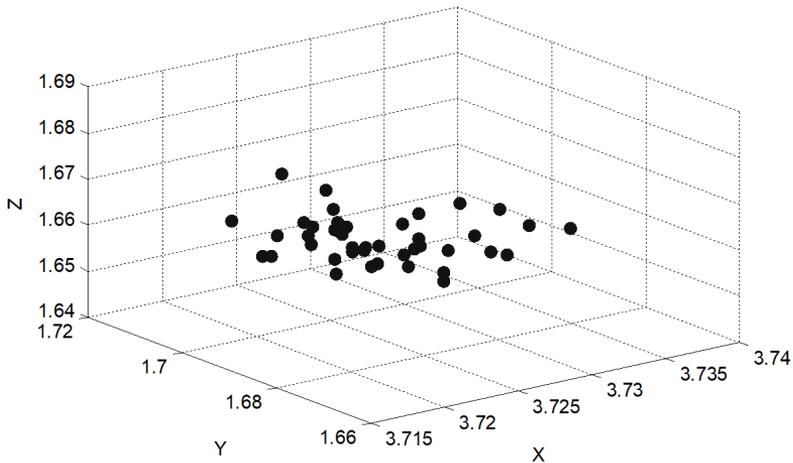


Fig. 5. Test results of point 1

## 5. Conclusion

In order to improve the localization accuracy of UWB, an improved UWB space location method is proposed in this paper. The method optimizes the ranging, and then use the distance intersection of space method to solve 3D coordinates.

Based on continuous 2 seconds of data processing, the method optimizes the ranging and effectively improves the accuracy of ranging. The space location algorithm in the method is simple and practical, and its computation cost is small. Tag's 3D coordinates can be solved by only three base stations, which saves the hardware cost. It can be applied to large-scale warehouse equipment management and other application scenarios, it may not suitable for real-time demanding applications.

In addition, a future study on ranging optimization method, such as study the empirical model, to eliminate system errors, which can improve the ranging accuracy for static positioning to provide a higher positioning accuracy, remains as a future work.

## References

- [1] R. MAALEK, F. SADEGHPOUR: *Accuracy assessment of ultra-wide band technology in locating dynamic resources in indoor scenarios*. *Automation in Construction* 63 (2016), 12–26.
- [2] R. A. SANTOS, A. C. SODRÉ, S. E. BARBIN: *A low-profile printed antenna for UWB applications*. Proc. International Conference on Electromagnetics in Advanced Applications (ICEAA), 19–23 September 2016, Cairns, QLD, Australia, IEEE Conference Publications (2016), 905–908.
- [3] G. YANJING, A. LO, I. NIEMEGEERS: *A survey of indoor positioning systems for wireless personal networks*. *IEEE Communications Surveys & Tutorials* 11 (2009), No. 1, 13–32.
- [4] X. SUI, G. YANG, Y. HAO, C. WANG, A. XU: *Dynamic positioning method based on TDOA in underground mines using UWB ranging*. *Journal of Navigation and Positioning* 4 (2016), No. 03, 10–14,34.
- [5] B. ALAVI, K. PAHLAVAN: *Modeling of the TOA-based distance measurement error using UWB indoor radio measurements*. *IEEE Communications Letters* 10 (2006), No. 4, 275–277.
- [6] B. SILVA, G. P. HANCKE: *IR-UWB-Based non-line-of-sight identification in harsh environments: Principles and challenges*. *IEEE Transactions on Industrial Informatics* 12 (2016), No. 3, 1188–1195.
- [7] X. LIANG, H. ZHANG, T. LV, X. CUI, T. A. GULLIVER: *NLOS identification approach based on energy block for 60GHz systems*. *International Journal of Database Theory and Application* 9 (2016), No. 9, 59–74.
- [8] X. LIU, Y. HUANG, Y. QIAN, J. WANG: *A TDOA localization algorithm based on sequential quadratic programming*. *SERSC Advanced Science and Technology Letters* 138 (2016), 134–139.
- [9] S. A. A. SHAHIDIAN, H. SOLTANIZADEH: *Path planning for two unmanned aerial vehicles in passive localization of radio sources*. *Aerospace Science and Technology* 58 (2016), 189–196.
- [10] E. OKAMOTO, M. HORIBA, K. NAKASHIMA, T. SHINOHARA, K. MATSUMURA: *Low-complexity indoor UWB localization scheme using particle swarm optimization*. *IEICE Nonlinear Theory and Its Applications* 6, (2015), No. 2, 194–206.

- [11] J. KHODJAEV, S. TEDESCO, B. O'FLYNN: *Improved NLOS error mitigation based on LTS algorithm*. PIER Progress In Electromagnetics Research Letters 58 (2016), 133–139.
- [12] L. LI, Z. WEN, Z. WANG: *Outlier detection and correction during the process of groundwater level monitoring base on pauta criterion with self-learning and smooth processing*. Proc. Asia Simulation Conference and SCS Autumn Simulation Multi-Conference, 8–11 October 2016, Beijing, China, Springer Singapore (2016), 497–503.
- [13] B. DAI, X. LÜ, X. LIU, Z. LI, : *A UWB-based four reference vectors compensation method applied on hazardous chemicals warehouse stacking positioning*. Journal of Chemical Industry and Engineering(China) 67 (2016), No. 03, 871–877.
- [14] M. NDOYE, J. M. M. ANDERSON, D. J. GREENE: *An MM-based algorithm for-regularized least-squares estimation with an application to ground penetrating radar image reconstruction*. IEEE Transactions on Image Processing 25 (2016), No. 5, 2206 to 2221.
- [15] A. LIU, Y. DU, H. SUN: *Research on positioning solutions of distance intersection applied with laser range-finder*. Modern Electronics Technique 38 (2015), No. 19, 24–32.

Received June 29, 2017

# Analysis of RFID technology based on technology principles and construction of development model

TIAN BOJIN<sup>1</sup>, CHEN KE<sup>1</sup>, AI QIANKE<sup>1</sup>, QIU XIAOPING<sup>2,3</sup>

**Abstract.** Radio frequency identification (RFID) system uses radio frequency communication technology to realize the information transmission between reader and tag, and realize the identification, location, monitoring and tracking of target. In order to obtain a better performance of the passive UHF radio frequency identification system, a high-performance open test platform for RFID systems was developed in this paper; influences of factors such as reader, tag antenna polarization mismatch, mutual coupling effect of tag antenna and multipath effect on system performances were analyzed; with the development of test platform, influencing factors in the study were measured. The final experimental results show that: combined with the RFID system simulation application environment based on PLC and OPC technology, the testing platform can provide complete RFID system testing solutions including system performance and conformance testing, third party monitoring, and application testing, and provide better performances.

**Key words.** Video recognition, performance testing, research development.

## 1. Introduction

As one of the core technologies in the development of Internet of things, radio frequency identification technology has attracted more and more attention from governments, scientific research institutions and enterprises. Based on the RFID system, combined with the existing network technology, the database technology and middleware technology, it is composed of a large number of networked readers and numerous mobile labels. The Internet of things has become a trend in the development of RFID technology, which is more massive than the current Internet. In the architecture of Internet of things, RFID tags contain items of information that conform to uniform standards with the interoperability. Through wireless and

---

<sup>1</sup>State Grid Chongqing Information & Telecommunication Company, Chongqing, 400000, China

<sup>2</sup>Chongqing University of Technology, Chongqing, 400054, China

<sup>3</sup>Corresponding author

wired network, it can collect relevant information to the central information system to realize the identification of objects. Through the open computer network, it can realize the information exchange and share, so as to realize the effective management of the goods. As the carrier of identification information, RFID tags have the advantages of data visibility, the character of real-time, security, environmental adaptability, and the potential of extending the lifecycle management of products. The research of RFID technology mainly includes two aspects: the industrialization key technology and the application key technology. The key technologies of RFID industrialization include tag chip design and manufacturing, antenna design and manufacturing, RFID tag packaging technology and equipment, RFID tag integration and reader design. The key technologies of RFID application include: RFID application architecture, RFID system integration and database management, RFID public service system and RFID testing technology and specifications.

## 2. State of the art

Foreign RFID related organizations, enterprises and research institutes have established their own RFID testing platforms to promote the development of RFID technology, as well as the industry and application. The Infineon Technologies Company founded the RFID solutions exhibition and evaluation center and system laboratory in Austria in 2004. The lab can provide solution information for the Infineon Technologies Company's RFID systems, including software and systems integration platforms, infrastructure, readers and tags, and other related equipment [1]. The RFID testing center, established by Sun Microsystems in Dallas, USA, focuses on the tag optimization and backend data integration issues. The terminal users of Sun Microsystems products can use the test center to ensure that the products they use can meet the application requirements and can test the RFID system before the actual deployment of the device [2]. At present, some institutions and research institutes in China have also made some progress in the construction of RFID test platform. With the support of the national high technology researches and development programs, in October 2004, Beijing Zhongjiaoguo Logistics Technology Development Co. Ltd. and Chinese Academy of Sciences Institute of Automation Research Center RFID established the first national RFID test center in China. Through a number of reliability indexes of the key technology of the RFID test, the center put forward the reliability evaluation system to provide the basic data support and direction for further researches [3]. In 2004, the Auto-ID China Laboratory, affiliated to the State Key Laboratory of ASIC and systems, Fudan University, Shanghai was established. The platform can analyze and test the problems in practical applications, provide references for the relevant production enterprises, and provide the theoretical basis and technical supports for the establishment of independent RFID standards in China [4].

### 3. Methodology

Typically, a radio frequency identification system consists of an electronic tag, a reader, and a computer communication network, as shown in Fig. 1. For each part, the functions are as follows. Electronic tag: as a memory, it contains information about objects and is usually placed in objects. Each tag has a unique ID number UID; reader: as a recognition device, it exchanges information mainly through the RF technology and electronic tags, and it can be designed as hand held or fixed type. The signal power of the reader is much larger than the backscattered signal from the tag, and this has the same frequency as the received signal; computer network communication: it can complete communications, and it is used for data management. Some systems can also connect to the upper computer via a reader's RS232 or RS485 interface [5]. The working principle of the system is that when the reader is in working mode, the antenna will emit radio waves of sufficient powers, and when an object with an electronic tag is near the reader and in the radio frequency range of the reader, the RFID tag will be activated and sent information to the reader. The reader receives and demodulates the RF signals from the tag and sends them to the computer's main systems for processing. The main system makes corresponding processing and control according to the logical operation, and sends out the instruction signal [6].

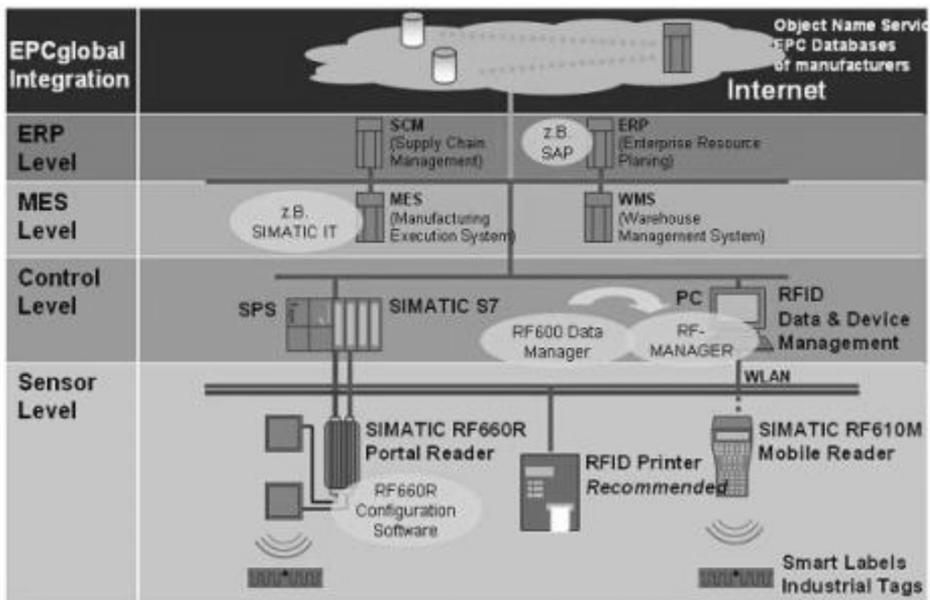


Fig. 1. Radio frequency identification system

RFID system test means testing related equipment through scientific testing methods, testing instruments and testing platforms, the readers, tags, antennas and middleware, and so on in the R & D design phase of RFID system or under the actual deployment environment set by terminal users [7]. Based on the related

test contents, test results should be scientifically analyzed; according to different frequency bands and application field of RFID products, corresponding production standards and test specifications shall be established; the system of testing standards of the perfect RFID system should be established to improve the product performance and optimize the equipment deployment, and to promote the progress of RFID technology and the wide range of product promotions [8].

The application test of RFID system was aimed at the specific application environment. By changing the location of the tag, the location of the antenna, the tag, the material attached, and other environmental factors, system identification area, recognition rate, recognition rate, reliability and interoperability performance of RFID system under specific conditions were tested, and specific test contents are shown in Table 1.

Table 1. Application test of RFID system

Label location	Direction
	Angle
	Spacing
Antenna deployment	Direction
	Angle
	Position
Dielectric material	Frequency drift
	Dielectric constant
	Absorption reflex
Environmental factor	Multipath effect
	Electromagnetic interference
	Environmental temperature and humidity

The application test of RFID system faces specific application environment, so it has more practical reference values to terminal users [9]. The environmental factors involved in RFID application testing include: the change of parameters such as the direction, angle and relative position of labels caused by the tag position; the change of parameters such as the square, the angle and the height of the antenna caused by the deployment of the antenna; the change of frequency drift, dielectric constant and absorption reflection coefficient caused by dielectric materials; other changes in the application environment, such as multipath effects, electromagnetic interference, and environmental temperature and humidity parameters. Through the test, among the above parameters, in one or several hours, the identification area, recognition area, recognition rate, recognition rate, reliability and interoperability of RFID systems were changed; the internal relations between various environmental factors and system performance in application environment were analyzed; the influence degree of environmental factors on the system performances was graded; a visual system evaluation model was proposed; references were provided for the optimized and rapid deployment of equipment for terminal users [10].

According to the application requirements of the development of RFID testing

technology, combined with the hardware and software design of test platform and four working modes, main functions of the RFID system test platform are shown in Fig. 2.

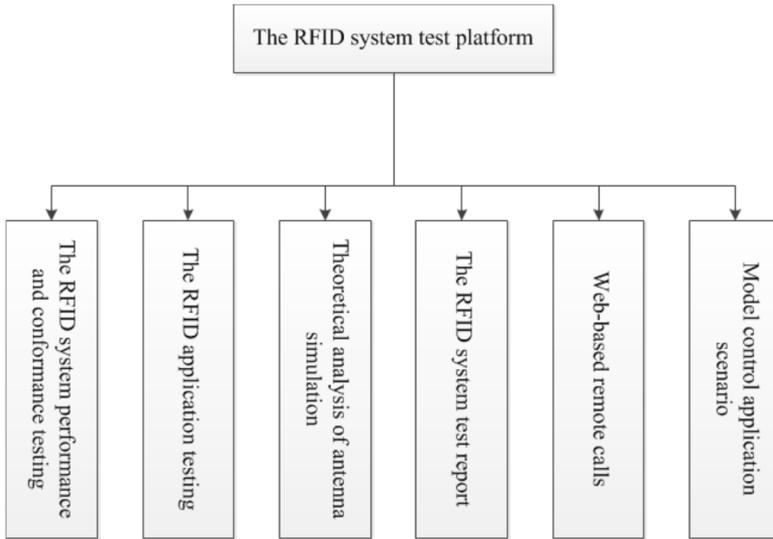


Fig. 2. Main functions of the RFID system test platform

The system performance and conformance testing of RFID mainly completes the conformance testing of air interface physical layer, protocol layer performance standards and other provisions of the standard RFID system under the environment of the system (such as anechoic chambers and open spaces, etc.), as well as the recognition, recognition rate, sensitivity, label RCS and system performance test [11]. RFID system conformance testing includes physical layer and protocol layer test. Physical layer tests include the time domain parameter tests, such as the power on the reader, the time, the pulse width, the duty cycle, the polling time and the state and storage time of the reader; frequency parameter tests such as frequency range, frequency offset, carrier stability, occupied bandwidth, intra band power, and adjacent channel power leakage ratio [1]; modulation domain parameter tests such as the modulation method, modulation index, modulation depth, envelope level value, envelope rising value, falling edge level and envelope value, envelope rising time, fall time, envelop overshoot and undershoot; and the joint time-frequency analysis of signal in time domain, frequency domain and modulation domain. The test protocol layer includes a data storage test; testing of state machines such as ready, arbitrate, answer, confirm, open, protect and inactivate; the test of instruction execution such as save, query, read and write; testing of data frames such as preamble, checksum, data rate, and delimiter length; and the collision test [12].

According to the relevant standards and testing standards, the consistency test of the air interface communication protocol used the developed test platform to complete the test of the reader, label performance and system communication process, as shown in Fig. 3.

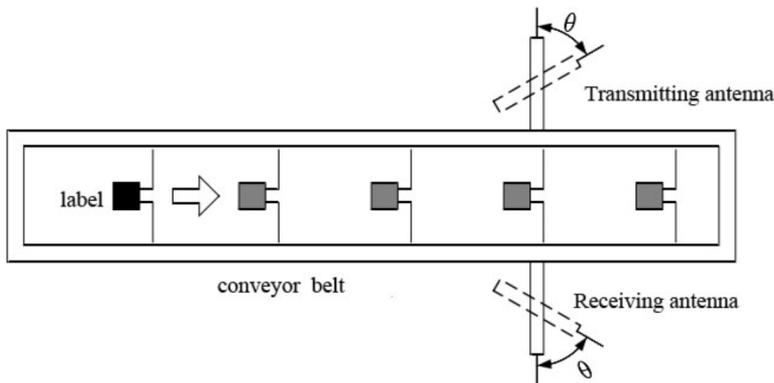


Fig. 3. Conformance testing of communication parameters of the tag's air interface

The test process is as follows: the first step: after the device deployment and initialization were completed, the PCI-5640R, PCI-5610 and PCI-5600 card aliases were selected, and the carrier mode was set to Burst. Second step: the carrier center frequency was set to 915 MHz; the transmit power was set to 27 dBm; the reference power was set to  $-10$  dBm; the capture signal length was set to 10 ms. The third step: the firing command was selected as Inventory Sequence. According to the ISO/IEC 18000-6C standard, the relevant parameters of the instruction are set and the instructions were emitted. In the fourth step, the time domain and frequency domain characteristic parameters of instruction and response signal were analyzed and tested according to the captured instruction and the response signal waveform. By calculating, the differential radar scattering interface of the tag was obtained. The fifth step: according to the test result, the test report was given.

The real-time simulation system RFID reader test can identify the scope by using the developed test platform, system identification and test the reader and the tag antenna angle by rotating the reader antenna with the change of the angle between the reader and the tag antenna. The test example of the ALN-9640 tag of Alien Company was given. A symmetrical half wave antenna was used to test the antenna.

Test process: Step 1: the ALN-9640 tag was attached to the conveyor belt, and the labels and readers were set parallel to the antenna. At this point, the azimuth of the transmit and receive antennas was assumed to be:  $\theta = 0^\circ$ . The second step: the conveyor speed was set to 0.1 m/s; the transmitting power of the reader antenna was set to 27 dBm. The carousel was activated to record where the label was successfully identified. The third step: the angle of the transmitting and receiving antenna was set to 15 degrees, from 0 to 165 degrees. At different angles, the second step was repeated. The fourth step: the identification location of the tag was counted with the change of  $\theta$ .

On the basis of the performance and conformance testing of the RFID system and the testing function of the RFID system application, the platform can also realize third party monitoring and data stream disk function. Based on the remote

invocation function based on Web of the platform, the platform can provide users with data storage and open testing services, and further study and analyze the stored data.

### 3.1. Result analysis and discussion

The results of the system identification test are shown in Fig. 4.

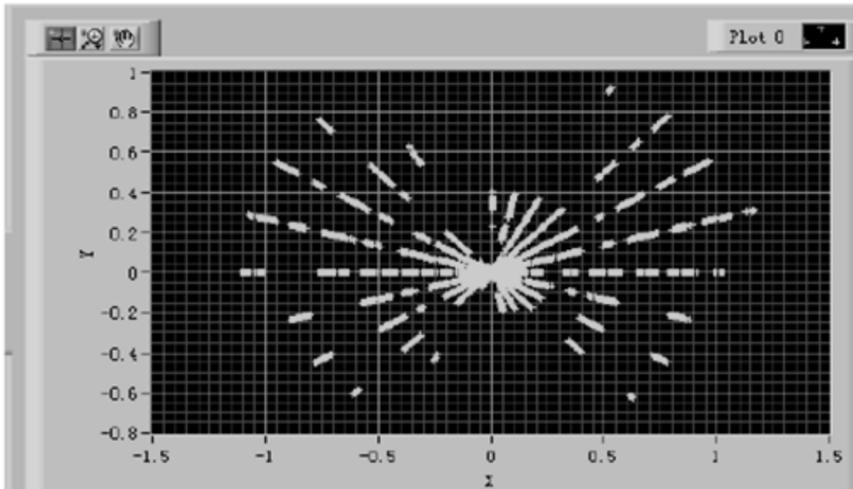


Fig. 4. The results of the system identification test

It can be seen that due to the effect of antenna, reader and tag reader and tag antenna main lobe direction angle, reader and tag antenna polarization mismatch labels and multipath effects and other factors, the recognition range was scrappy and uneven. In the maximum recognition range of the system, the tag had the location blind area which can't be identified, and it affected the recognition range and recognition rate and other system performance parameters.

The third party monitoring test can monitor the air interface of RFID system, and test the standard compliance of the system reader and tag equipment. The test platform supports the DSB-ASK, PR-ASK modulation, Miller encoding, NRZ-L encoding, PIE encoding, Manchester encoding and other encoding methods, and it can test the time domain waveform analysis, frequency domain analysis, and the baseband signal parameters of the whole communication process. It has the testing function of protocol layer parameters such as protocol state machine, data rate, and frame format, mandatory and user-defined command, so the sensitivity of monitoring test is relatively high, accompanied by simple testing methods. In the test, it is necessary to note that the tag response signal is weak, so in the case of limited antenna sensitivity, the monitoring antenna should be placed near the tag to obtain better listening effect. In order to further analyze, handle, and store the performance of the device to be tested, the test platform has realized the function of the signal flow disk. The acquired air interface signal can be entered into a

microcomputer to carry out the advanced protocol analysis. The waveform data is in the standard format and it is compatible with software such as Matlab.

The test results of minimum transmit power of the reader antenna are shown in Table 2.

Table 2. Test results of minimum transmit power of the reader antenna

Tag number	1	2	3	4	5	6	7	8	9
$\bar{P}_{\text{seni}}$ (dBm)	23	33	34	17	16	17	21	20	21
$P_{\text{ni-min}}$ (%)	-8.7	-9.1	-8.7	-5.9	-6.3	-5.9	-9.5	-10	-9.5
$P_{\text{ni-max}}$ (%)	4.4	4.8	4.4	11.8	12.5	11.8	4.8	10	4.8
$\Delta P_{\text{ni}}$ (%)	13.1	13.9	13.1	17.7	18.8	17.7	14.3	20	14.3

When the tag spacing is less than 1.5 times the system operating frequency wavelength, the mutual coupling effect has great influences on the system performance; the effect of mutual coupling on the system performance is nonmonotonic, so it can be enhanced or reduced; for double label, the range of the minimum transmit power of the reader antenna is (-7%, 11.6%); increasing the transmit power of the reader antenna can reduce the influence of mutual coupling effect on the system performance; for the label double plane case, the range of the minimum transmit power of the reader antenna of target plane label is (-10%, 12.5%); the influence of the interference planar tag on the recognition rate of the target plane label system is similar to that of the metal plane.

## 4. Conclusion

In this paper, combined with the requirements of the project and the actual application requirements of RFID, the RFID system test platform based on software defined radio and virtual instrument technology was developed. By using the test platform and commercial RFID equipment, influences of polarization mismatch between reader and tag antenna, antenna inter-coupling effect and multipath effect on the performance of RFID system were analyzed and tested. Theoretical guidance and references were provided to the RFID technology research and the rapid deployment of equipment of terminal users. Through the research of this paper, some conclusions were obtained as follows.

Compared to existing RFID test equipment, the test platform developed in this paper has the characteristics of low cost, multi support protocol, user-defined test function and typical deployment scenarios, and it can provide complete solutions for the RFID system testing, such as the RFID system performance, the protocol conformance and application testing. When the tag spacing is less than 1.5 times the system operating frequency wavelength, the mutual coupling effect has great influences on the system performance; the effect of mutual coupling on the system performance is nonmonotonic, so it can be enhanced or reduced; increasing the transmit power of the reader antenna can reduce the influence of mutual coupling

effect on the system performance; for double label and label biplane situations, the ranges of the minimum transmit power of the reader antenna of target plane label are (-7%, 11.6%) and (-10%, 12.5%). Compared to free space, the degradation rate of system path loss in indoor multipath environment is faster; the propagation of the antenna from the reader to the tag electromagnetic wave is the first Fresnel region, and the additional loss is caused by the barrier so as to increase the loss rate of the system path loss; when Fresnel clearance is 1.5 times larger than the radius of the first Fresnel District, edge obstacles have less influence on the system path loss; compared with the traditional logarithm distance path loss model, the standard deviation of the two-slope model proposed in this paper can be reduced by more than 10%. The reader antenna should be oriented to the geometric center of the target region; commercial reader antennas are mostly elliptical polarized; polarization mismatch is still an important factor that should not be neglected; compared to the single antenna case of a single reader, the multi reader antenna can effectively improve the target area recognition rate; multi antenna coherent multipath interference can produce new blind spots, but the improper use of it will reduce the identification rate of target areas; the system performance optimization method based on label set and phase switch can improve the target area recognition rate by 10% and 7.6%.

Because of the limitations of research level and condition, this research still has some problems: the expression of mutual impedance between tag antennas and the influence of mutual coupling effects on the performance of RFID system are only applicable to the condition where the antenna is in the far field of the antenna radiation, and the tags in practical applications may be in the near-field region of antenna induction; evaluation methods cannot reflect the density distribution of target region recognition rate; tag set optimization method increased system costs; when the target object is small, if the distance is less than 1.5 times the frequency of the label system wavelength, the mutual coupling effect between antennas may lower system performances; the phase switch optimization method increased the number of read times of readers, and reduced the recognition rate of the system.

## References

- [1] E. W. T. NGAI, K. K. L. MOON, F. J. RIGGINS, C. Y. YI: *RFID research: An academic literature review (1995–2005) and future research directions*. International Journal of Production Economics 112 (2008), No. 2, 510–520.
- [2] Q. B. SUN, J. LIU, S. LI, C. X. FAN, J., J. SUN: *Internet of things: Summarize on concepts, architecture and key technology problem*. Journal of Beijing University of Posts and Telecommunications 3 (2010), No. 3, 1–9.
- [3] Y. JIAO: *Construction and application of logistics information tracking system based on RFID technology*. Journal of Convergence Information Technology 8 (2013), No. 10, 837–845.
- [4] K. DZIADAK, B. KUMAR, J. SOMMERVILLE: *Model for the 3D location of buried assets based on RFID technology*. Journal of Computing in Civil Engineering 23 (2009), No. 3, 148–159.
- [5] Y. XIAO, S. YU, K. WU, Q. NI, C. JANECEK, J. NORDSTAD: *Radio frequency identification: Technologies, applications, and research issues*. Wireless Communications and Mobile Computing 7 (2007), TOC No. 4, 457–472.

- [6] K. PENTTILA, M. KESKILAMMI, L. SYDANHEIMO, M. KIVIKOSKI: *Radar cross-section analysis for passive RFID systems*. IEEE Proceedings - Microwaves, Antennas and Propagation 153 (2006), No. 1, 103–109.
- [7] J. S. PARK, J. W. JUNG, S. Y. AHN, H. H. ROH, H. R. OH, Y. R. SEONG, Y. D. LEE, K. CHOI: *Extending the interrogation range of a passive UHF RFID system with an external continuous wave transmitter*. IEEE Transactions on Instrumentation and Measurement 59 (2010), No. 8, 2191–2197.
- [8] S. J. THOMAS, E. WHEELER, J. TEIZER, M. S. REYNOLDS: *Quadrature amplitude modulated backscatter in passive and semipassive UHF RFID systems*. IEEE Transactions on Microwave Theory and Techniques 60 (2012), No. 4, 1175–1182.
- [9] D. G. KUESTER, D. R. NOVOTNY, J. R. GUERRIERI, A. IBRAHIM, Z. B. POPOVIC: *Simple test and modeling of RFID tag backscatter*. IEEE Transactions on Microwave Theory and Techniques 60 (2012), No. 7, 2248–2258.
- [10] B. S. WANG: *Internet of things technology research*. Journal of Electronic Measurement and Instrument 23, (2009), No. 12, 1–6.
- [11] V. DERBEK, C. STEGER, R. WEISS, J. PREISHUBER-PFLÜGL, M. PISTAUER: *A UHF RFID measurement and evaluation test system*. e & i Elektrotechnik und Informationstechnik 124 (2007), No. 11, 384–390.
- [12] X. ZHU, S. K. MUKHOPADHYAY, H. KURATA: *A review of RFID technology and its managerial applications in different industries*. Journal of Engineering and Technology Management 29 (2012), No. 1, 152–167.

Received June 6, 2017

# Application of ZigBee wireless sensor network in gas monitoring system<sup>1</sup>

JUN HU<sup>2</sup>

**Abstract.** The coal mines in our country are in great demand, and mining technology level is relatively backward compared with other developed countries. The monitoring system has the disadvantages of complicated wiring and high cost. In order to overcome the shortcomings of China's current coal and gas system, the application of ZigBee wireless sensor network technology in coal mine gas monitoring system was put forward. The hardware and software of wireless sensor network node and the overall scheme were designed, and the performance of the system was tested in this paper. The results show that the monitoring system can carry out a comprehensive real-time intelligent monitoring to gas accidents, which reduces the probability of gas accidents effectively and has certain practical significance.

**Key words.** ZigBee, wireless sensor, gas monitoring, application.

## 1. Introduction

Coal industry has always been the focus of China's industrial development. It is not only related to China's economic development and social progress, but also the key industry for the revitalization of the nation. Base on this, the status of coal industry in the development of China's industry can be seen. However, the coal industry in our country now has such disadvantages as low overall productivity, serious natural disasters, complicated mining conditions, low safety and frequent accidents. In addition, there are many small coal mines which are privately exploited in our country, and the security awareness is poor. Therefore, the coal mine industry is the key industry in our country's safety accidents. According to the survey of China's Security Supervision Bureau, China's coal mine accidents caused by death toll reached 5670 people in 2001 and the number of deaths caused by coal mines in China reached 6995 people in 2002. 7023 people died due to coal mine accidents in

---

<sup>1</sup>This work is supported by a fund from Scientific Research Project of Education Department of Hunan province. Project name: Research on key technology of gas concentration monitoring in coal mine based on Wireless Sensor Networks, No. 15C0612.

<sup>2</sup>College of Software, Hunan Vocational College of Science and Technology, Changsha, Hunan, 410118, China

2004. Thus, the safety awareness of China's coal industry has not increased year by year, but accidents happen frequently year by year. The study shows that gas explosion has the highest proportion and is the most dangerous among all accidents in these coal mine accidents. Therefore, gas monitoring is a very important work, and it is the guarantee of safety production in coal mining industry.

Based on the demand of coal mine industry safety guarantee, the hardware and software of the overall scheme of gas monitoring and wireless sensor network node were designed aiming at the coal mine gas monitoring system by using the Zigbee wireless sensor network technology based on the basic theory in this paper. In the second section, the concept of infinite sensor network was introduced, and the current development of the world wide sensor network was summarized briefly. In the third section, Zigbee technology was used to design the hardware and software of the wireless sensor network node. CC 2430 was used as the sensor node, and a whole gas monitoring system was designed. The data of the preliminary application of the system were summarized and analyzed in the fourth section. Finally, the research process and results were summarized in the fifth section.

## 2. State of the art

Through consulting relevant research documents, it can be seen that China began to apply gas monitoring technology in coal mine industry from the beginning of eightieth century [1]. The developed countries, such as the United States and Britain, have established the gas monitoring system with long time. Gas monitoring system is to monitor all kinds of harmful gas and working environment mainly, such as the concentration of oxygen and carbon dioxide gas, the temperature and humidity under the mine, and the highly dangerous methane gas and so on [2]. Once there is insecurity, people can find it in time. At first, our country imported related equipment and instruments from abroad, and later developed the K 14 monitoring system through the integration of the actual situation and characteristics of our country [3]. With the development of science and technology and the arrival of electronic era, the traditional gas monitoring system was replaced by K 195 and other high-tech, computer or electronic technology systems gradually [4]. The national government has also set up relevant laws and regulations to regulate the development of coal mine industry. It is stipulated that the gas monitoring system must be equipped regardless of the size of the mineral enterprise [5]. With the continuous development of China's coal mining industry, more and more small enterprises appear in the line of sight. Competition among enterprises is fierce, and it also promotes the development speed and quality level of China's coal industry. Gas monitoring system has also changed from traditional single microcomputer monitoring system to networked monitoring system or networked monitoring system [6]. However, the monitoring of safety is still in the international backward level.

### 3. Methodology

At present, the gas monitoring system in our country has the following problems compared with the foreign countries, as shown in Table 1. And ZigBee technology characteristics and gas monitoring system problems can be fused together. Specific features of ZigBee are shown in Table 2. Therefore, targeted research is conducted on existing issues in this article. Sensor nodes, management nodes and sink nodes are three major components of a sensor [7]. The nodes of the general sensor are located in the monitored area randomly, and all nodes are connected to form a huge network system. Sensor nodes can communicate the monitored signals to each other, and aggregate in the pooled nodes, and then transmit to the management node via satellite and Internet processing. People can manage and set up the entire sensor network according to the signal of the management node. At the same time, it can also monitor the release of the monitoring task and obtain the statistical results [8]. The general structure of the sensor network is shown in Fig. 1. Sensor nodes usually exist in the form of embedded systems [9]. Each node has a path to the node terminal, and it can also propagate the signal to other nodes. Each node has the function of processing data and storage, so the core of the sensor network is even the terminal monitoring node [10].

Table 1. Existing problems of gas monitoring system in China compared with those in other countries

Serial number	Problems existing in gas monitoring system in China
1	The wired network has complex wiring, high labor intensity and high maintenance cost of communication lines. Therefore, its scalability and flexibility are inadequate for industrial buses. The communication line is easy to destroy, and it is easy to waste the resources because of the increase of cost.
2	The network structure is relatively fixed, and it is not suitable for the dynamic change of tunneling.
3	The monitoring point is relatively fixed, and the detection blind area is prone to occur.
4	The processing level of primary instrument (sensitive component) in safety instrument is much lower than that of foreign advanced level, which makes the accuracy and reliability of detecting gas data insufficient.

It is well known that wireless sensor nodes are very important for the whole monitoring network. A general wireless sensor node consists of the following sections: SCM and wireless communication module, energy supply module, sensor module and processor module [11]. The detailed structure is shown in Fig. 2. The most important component of the infinite sensor nodes is the microcontroller and wireless communication module. The biggest difference between the chip and CC2420 is the integration of Zigbee, 8KB and RF front-end SARM and large capacity storage space on the basis of CC2420, and the advanced performance of 8051MCU enhanced

industry standard is joined. The chip has an analog digital converter, CO processor, time and sleep mode timer, and has the core MJC4/3.0L original carrier catalytic gas concentration monitoring circuit and a temperature and humidity sensor.

Table 2. Technical features of ZigBee

Serial number	Features of Zigbee Technology
1	Excellent wireless reception sensitivity and powerful immunity
2	Only 0.9 microns of flow is consumed in sleep mode, and external interrupts or RTC can wake up the system. In standby mode, the flow loss is less than 0.6 a, and external interrupts can wake up the system.
3	Wider voltage range (2.0~3.6 V)
4	Hardware support for CSMA/CA functionality
5	Digital RSSI/LQI support and powerful DMA capabilities
6	ADC integrated with 14 bit analog to digital converter
7	There are 2 powerful USART protocols that support several groups of protocols, 1 MAC timers that conform to the IEEE 802.15.4 specification, 1 regular 16 bit timers and two 8 bit timers.
8	Powerful and flexible tools for development.

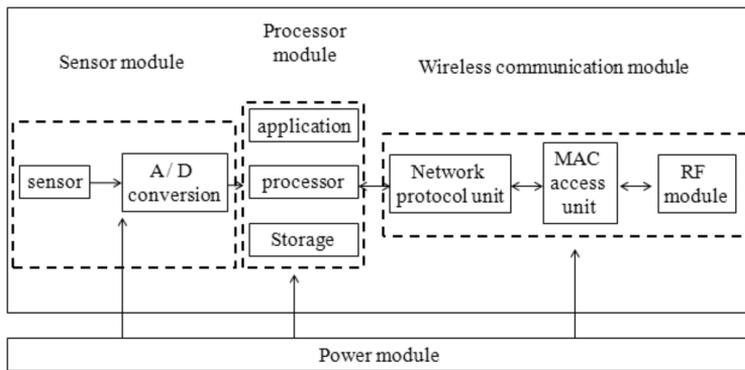


Fig. 1. Structure of wireless sensor

The CC2420 chip has a working band of 2.400–2.4835 GHz. The current consumption is low and the receiving sensitivity is  $-99$  dBm. The CC2420 chip adopts the IEEE 802.15.4 standard direct sequence spread spectrum (DSSS) mode. The speed is 2 MChip/s, and the output power is controlled by the programming program. The internal has VCO, LNA, PA, and current rectifier. The power supply voltage is 2.1–3.6 V, and the anti-interference channel capability is stronger. And the MAC layer application of IEEE802.15.4 can support the generation of automatic

frame format, simultaneous insertion and detection, 16bit, CRC collation, inspection, monitoring for power supplies and full MAC layer security protection. With 4 bus SPI interface, development tools are fairly complete, including the development of documents and presentations, the size of  $7 \times 7$  mm, and QLP-48 packaging. Although all aspects of CC2420 performance have been able to meet the general monitoring system, CC2430 chip has many more powerful features. Various aspects of performance of CC2430 have been improved greatly on the basis of CC2420. Details are shown in Table 3. The power supply of this system is a low voltage RE113-3 regulator chip, which can provide stable voltage of 3V continuously. The principle of an infinite sensor node is shown in Fig. 3.

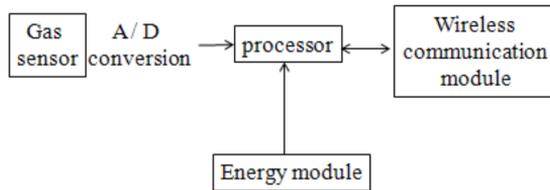


Fig. 2. Structure of wireless sensor node

Table 3. Features of CC 2430 chips

Characteristic	performance parameter
Data rate	868 MHz, 20 kbit/s, 915 MHz, 40 kbit/s, 4 GHz, 250 kbit/s
Communication range	up to about 150 m in the open environment, and within 100 meters generally,
Communication delay	About 15 ms
Channel number	868/915 MHz 11, 24 GHz 16
Band	868/915 MHz and 2.4 GHz
Addressing mode	64 bit IEEE address, 8 bitnet work address
Channel access	CSMA-CA and Time slot CSMA-CA
Temperature	-40~85 °C

The key information of this system comes from the user's special keyboard mainly. The input program in section 1 consists of 8 bytes. When a key is pressed, the first byte represents the corresponding value of the keyboard. When two keys are pressed, the first byte and second byte represent the corresponding values of the keyboard. Based on the same principle, when 8 keys are pressed, all bytes represent the corresponding values of the keyboard. However, when the key pressed is less than 8, the bytes that are not represented will be displayed in the form of "00H". The status of the keyboard is scanned regularly and the scanning information is sent to the fast detection module. When CH375 completes the task of receiving the information, the application is interrupted to the microcontroller. The single chip

microcomputer takes the corresponding value as the judgment and the operation processing after the judgment mark.

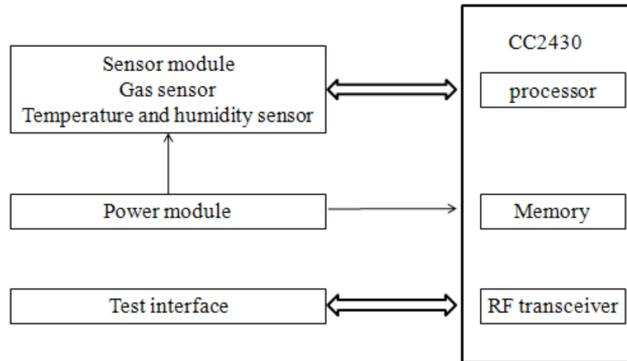


Fig. 3. Principle of infinite sensor nodes

#### 4. Results analysis and discussion

The infinite sensor node structure adopted in this paper not only saves the cost, but also solves the problem of the connection between single chip computer and wireless communication module. The software used is the C51RF-3-CC 2430-PK wireless single-chip microcomputer development system of Chengdu Antenna Dragon Company. The process of designing a wireless sensor network node by using ZigBee technology is as follows: firstly, ZigBee C51RF-3 was used as the development platform to build the platform for wireless sensor nodes. The platform consisted of two ZigBee based host computers and two wireless extended performance boards. The wireless expansion board was used to connect the antennas to complete the function of sending and receiving information. The host and wireless extensions board were connected via the RS-232 interface. Secondly, the driver software for each function was developed. The development environment for IAR7.2C51 was installed after the establishment of the wireless sensor network platform. And then through the software design, the application development was completed according to the need, and the function of each module was realized. Thirdly, the simulation test of the system was carried out. The online emulator system was connected to the host via the USB interface. At the same time, the 10 wire simulation cable is used to connect the system to the wireless MCU target board of CC2430 chip. The wireless network node system was simulated and tested online. The overall system flow is shown in Fig. 4.

Generally, when the gas concentration is more than 1% in the mine, the gas monitoring system will give an alarm prompt. The prescribed gas monitoring system shall not exceed 0.05% of the prescribed error range. Different concentrations of methane gas were prepared in the laboratory and stored in closed cylinders. Then, the probe of wireless sensor was extended into the bottle, and the on-line simulation

test of the system was carried out. The results show that the gas concentration of the gas in the cylinder is determined by the alarm concentration of the gas under the mine.

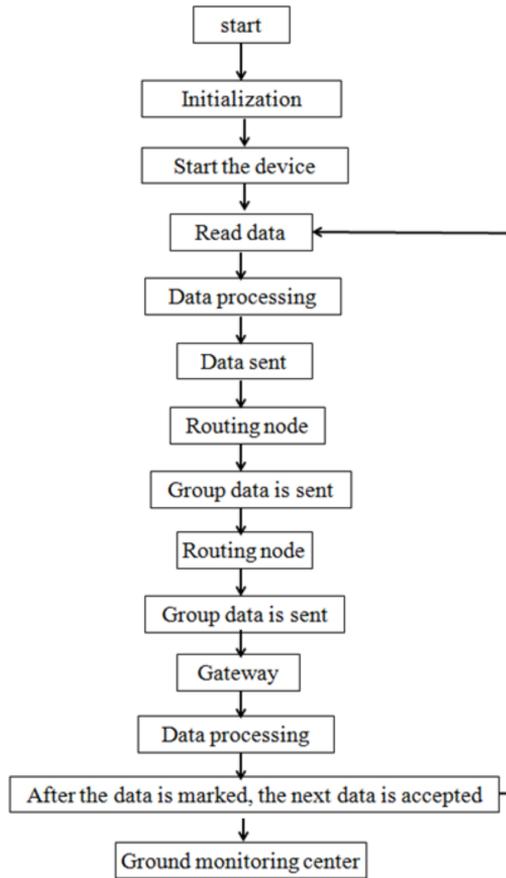


Fig. 4. System construction diagram

Some source files and software functions in the coordinator used in the system are shown in Table 4.

The protocol stack is responsible for assisting the coordinator to build the network mainly in routing nodes. The functions of sending information and data are realized by connecting the RFD nodes. Some source files and software functions used in the system are shown in Table 5.

Mote View visual monitoring software was used in the design of real-time monitoring system in this paper. The software got all the data passed by all the nodes from the Postre SQL database successfully, and achieved real-time monitoring. As long as the manager can observe all the data of the Mote View in the display screen or the situation diagram and the topological graph of the node, the environment information of all the nodes can be mastered for the first time. The software can

also be set to mobile mode, which means that when there is an alarm message, it will be sent to the manager's mobile phone in the form of SMS or e-mail. The security hidden danger can be discovered in the first time and the prompt measures are made.

Table 4. Coordinator source and software functions

Application layer program	Master coordinated control application (coord.c) , SPI master control interface (MSPI.c) , RS-232 terminal program (Console.c) , Zigbee application layer program (zAPL.c) , Zigbee application support sublayer (zAPS.c) , Zigbee device object (ZDO.c) , Proximity tables and binding tables are set up (NeighborTable.c) , Dynamic storage manager for indirect transport buffers (SRALLOC.c) etc.
Network layer program	Zigbee network layer program (zNWK.c)
MAC layer program	IEEE802.15.4MAC layer program (zMAC.c)
PHY layer program	CC2430 specific PHY program (zPHYCC2430.C)

Table 5. Source code and software function of routing node

Application layer program	Routing application (ROUTER.c), SPI master interface (MSPI.c), RS-232 terminal program (Console.c), Zigbee (zAPL.c), application layer Zigbee application support sublayer (zAPS.c), Zigbee (ZDO.c), the device object near the table and set up the binding table (NeighborTable.c), dynamic indirect transmission buffer storage manager (SRALLOC.c) etc.
Network layer program	Zigbee network layer program (zNWK.c)
MAC layer program	IEEE802.15.4MAC layer program (zMAC.c)
PHY layer program	The specific PHY program for CC 2430 (zPHYCC2430.C)

The monitoring results of the gas monitoring system established in this paper are shown in Table 4. Gas content of 0 cylinder monitoring results were in full accordance with reality in the three minute test, and the monitoring result was 0. That is to say, the error was 0. For the gas content of 0.79 cylinder test results, the monitoring result was 0.77, and the actual error size was 0.01%, which accorded with the corresponding standard of gas monitoring concentration. For the gas content of 1.19 cylinder test results, the monitoring result was 1.16 and the actual error size was 0.02%, which also conformed to the gas monitoring concentration corresponding standard. For the gas content of 1.69 cylinder test results, the monitoring result was 1.65, and the actual error size was 0.03%, which also conformed to the gas monitoring concentration corresponding standard. For the gas content of 1.49 cylinder test results, the monitoring result was 1.55, and the actual error size was 0.03%, which accorded with the corresponding standard of gas monitoring concentration. For the gas content of 1.09 cylinder test results, the monitoring result was

1.08, and the actual error size was 0.01%, which was in line with gas monitoring concentration of the corresponding standards. The error of gas concentration and actual gas concentration detected by wireless sensor based on ZigBee was less than 0.05%. The average error of all experiments was only 0.019%. According to gas monitoring standard and error range under coal mine, the wireless sensor network system based on ZigBee conforms to the standard of corresponding gas monitoring index fully.

Table 6. Monitoring results of gas monitoring system established in this paper

Gas concentration (%)	Test time (min)	Determination result (%)	Error (%)
0	3	0	0
0.79	3	0.77	0.01
1.19	3	1.16	0.02
1.69	3	1.65	0.03
1.49	3	1.55	0.03
1.09	3	1.08	0.01

## 5. Conclusion

In order to establish a more perfect gas monitoring system and reduce the occurrence rate of mine accidents, the node hardware of the gas monitoring system based on the ZigBee wireless sensor network of single chip microcomputer MSP430F149 and radio frequency chip CC2430 was designed. The sensor module based on the ZigBee protocol was designed to complete the information acquisition and collection in the monitoring area and control the whole system. A 2.4 GHz IEEE 802.15.4/ZigBee standard was used to establish a wireless sensor gas monitoring system for monitoring the gas concentration in the target environment. The system was used for simulation and test in this paper. The conclusions were drawn as follows: the wireless sensor network node ZigBee designed by MCU MSP430F149 and RF chip CC2430 designed in this paper can accomplish the functions of information collection, aggregation and sending. The application of Mote View software also can realize the real-time monitoring of the monitoring environment and the prompt of real-time alarm successfully. However, the environment under the mine is very complex, and the simulation of the cylinder is different from each other. So it is necessary to further optimize and adjust the system for the underground mine. There is no doubt that ZigBee based wireless sensor network monitoring system will become more and more mature, and its application will be expanded in the near future.

## References

- [1] J. HAN, C. S. CHOI, W. K. PARK, I. LEE, S. H. KIM: *Smart home energy management system including renewable energy based on ZigBee and PLC*. IEEE Transactions on Consumer Electronics 60 (2014), No. 2, 198–202.

- [2] N. C. BATISTA, R. MELÍCIO, J. C. O. MATIAS, J. P. S. CATALÃO: *Photovoltaic and wind energy systems monitoring and building/home energy management using ZigBee devices within a smart grid*. *Energy* 49 (2013), 306–315.
- [3] F. LECCESE, M. CAGNETTI, D. TRINCA: *A smart city application: A fully controlled street lighting isle based on raspberry-pi card, a ZigBee sensor network and WiMAX*. *Sensors* 14 (2014), No. 12, 24408–24424.
- [4] Y. LIU, Y. HE, M. LI, J. WANG, K. LIU, X. LI: *Does wireless sensor network scale? A measurement study on GreenOrbs*. *IEEE Transactions on Parallel and Distributed Systems* 24 (2013), No. 10, 1983–1993.
- [5] D. ZHANG, G. LI, K. ZHENG, X. MING, Z. H. PAN: *An energy-balanced routing method based on forward-aware factor for wireless sensor networks*. *IEEE Transactions on Industrial Informatics* 10 (2014), No. 1, 766–773.
- [6] S. TYAGI, N. KUMAR: *A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks*. *Journal of Network and Computer Applications* 36 (2013), No. 2, 623–645.
- [7] F. LEPOIVRE, A. GRIMAUD, D. LARCHER, J. M. TARASCON: *Long-time and reliable gas monitoring in Li-O<sub>2</sub> batteries via a swagelok-derived electrochemical cell*. *ECS-Journal of The Electrochemical Society* 163 (2016), No. 6, A923–A929.
- [8] A. E. ANDREWS, J. D. KOFLER, M. E. TRUDEAU, J. C. WILLIAMS, D. H. NEFF, K. A. MASARIE, D. Y. CHAO, D. R. KITZIS, P. C. NOVELLI, C. L. ZHAO, E. J. DLUGOKENCKY, P. M. LANG, M. J. CROTWELL, M. L. FISCHER, M. J. PARKER, J. T. LEE, D. D. BAUMANN, A. R. DESAI, C. O. STANIER, S. F. J. DE WEKKER, D. E. WOLFE, J. W. MUNGER, P. P. TANS: *CO<sub>2</sub>, CO, and CH<sub>4</sub> measurements from tall towers in the NOAA earth system research laboratory's global greenhouse gas reference network: Instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts*. *Atmospheric Measurement Techniques* 7 (2014), No. 2, 647–687.
- [9] B. GRUBER, S. KELLER, T. GROEGER, G. MATUSCHEK, W. SZYMCZAK, R. ZIMMERMANN: *Breath gas monitoring during a glucose challenge by a combined PTR-QMS/GC\*GC-TOFMS approach for the verification of potential volatile biomarkers*. *Journal of Breath Research* 10 (2016), No. 3, paper 036003.
- [10] S. B. HAMMERSCHMIDT, T. WIERSBERG, V. B. HEUER, J. WENDT, J. ERZINGER, A. KOPF: *Real-time drilling mud gas monitoring for qualitative evaluation of hydrocarbon gas composition during deep sea drilling in the Nankai Trough Kumano basin*. *Geochemical Transactions* 15, (2014), paper 15.
- [11] T. WIERSBERG, A. M. SCHLEICHER, K. HORIGUCHI, M. L. DOAN, N. EGUCHI, J. ERZINGER: *Origin and in situ concentrations of hydrocarbons in the Kumano forearc basin from drilling mud gas monitoring during IODP NanTroSEIZE Exp. 319*. *Applied Geochemistry* 61 (2015), 206–216.

Received June 29, 2017

# Image registration algorithm based on SIFT feature descriptor<sup>1</sup>

CHANGXING GENG<sup>2</sup>, PENG WANG<sup>2</sup>, PENGBO WANG<sup>2</sup>

**Abstract.** With the development of society, the single image has been unable to meet the needs of mankind. The new sensor allows the ability to capture images quickly. In order to meet this demand, the image mosaic technology is born. In this paper, a registration algorithm based on feature descriptors was proposed. The feature descriptor could fully reflect the shape and texture features of the image because the feature descriptor could quantify the local structural features of the image. The SIFT algorithm was used to extract image feature points and analyze the influence of different distance selection on the feature points matching in the similarity criterion. RANSAC algorithm was used to eliminate some misunderstanding matching, which improved the accuracy. The speed of pairing was analyzed by the parameter theory of Pyramid in SIFT algorithm.

**Key words.** SIFT, feature descriptor, image registration, algorithm.

## 1. Introduction

Through the human visual sensory system, the information described in an image can be analyzed by looking at an image. Human information is transported mainly through images because they can objectively display the real form of things, so that people can observe and understand [1]. The information expressed by voice and text is relatively simple, but the image is different, which contains many large amounts of information. The abundance of information makes it difficult for people to extract useful information from them. As the most important science in image information processing, digital image processing technology also has its branches [2], including image acquisition, registration, object detection, recognition and classification and so on. Image registration is the most basic task [3], and we only need to be responsible for that the same object or the same pixel image can match with others. No matter what the conditions are, we must obtain the same space in the final image [4].

---

<sup>1</sup>This work is supported by Jiangsu Province Natural Science Foundation for Young Scholars, No. BK20140325.

<sup>2</sup>Robotics and Microsystems Centre, Soochow University, Suzhou, 215021, China

Whether they are single sensors or multiple sensors, they have a certain relationship within the same data at different angles [5]. If the redundant information and complementary information can be used, a reliable basis can be obtained to improve the signal-to-noise ratio. Therefore, the integration of multiple resource information is the best way to get the information that we need [6].

## 2. State of the art

Image feature technology can be used in a variety of occasions requiring image processing. Foreign scholars have invested a lot of effort to study image processing techniques [7]. Beginning in the late 1970s, images began to be studied. Corner features were proposed by Moravec in 1977, but corner feature detection had many limitations [8]. For example, rotation invariance and noise sensitivity were not included. Harris and Stephens improved the Moravec's detector in 1988, which significantly improved detection rates and repetition rates with invariance to rotation and grayscale changes compared with the previous features [9]. In 1998, Lindeberg systematically proposed the scale space theory of signals, which realized scale invariant feature extraction. The method of region detection for maximum stable extremum was proposed by Matas in 2002, and affine invariant was obtained in strict sense [10]. In 2004, in order to solve many matching problems such as translation, rotation, affine transformation, perspective transformation, image features and so on, Lowe proposed scale invariant feature transformation (SIFT) algorithm, which was applied to a variety of situations.

In order to improve the speed of feature matching and the performance of matching, many scholars at home and abroad have done a great deal of improvement and research on the features according to the SIFT algorithm. In 2007, Tinne Tuytelaars improved the SIFT algorithm on the basis of gray information. Scholars from various countries have done a great deal of work and have reviewed the local feature detection operators [11]. After the proposal of n SIFT feature descriptor based on gradient image [12], SURF, GLOH and PCA-SIFT were improved based on SIFT feature descriptor [13], and was widely used in the follow-up.

### 2.1. Methodology

Image feature technology has been applied in many fields, such as image recognition, graph retrieval, image registration, image stitching, texture recognition and other fields. Its wide range of applications makes its research more thorough. Image registration means that no matter what conditions we are shooting, as long as we can be responsible for the same image or the same pixel, the same space can be obtained at last. Some local picture information obtained in the practical application is only by single sensor or multiple sensors in a same image or items, but the data is far more than the object or scene itself. In this paper, a registration algorithm based on feature descriptors is proposed. The feature descriptor can fully reflect the shape and texture features of the image, the reason is that the feature descriptor can quantify the local structural features of the image. In order to recognize objects, the

first thing is to represent the image in a reasonable way. This is to make it easier for us to match, so that only the same target can be matched in the absence of conditions [14]. In addition, it is necessary to take into account factors such as time, resolution, light, posture, etc. Why the same object appears differently in different images because the various unstable factors affect the state of the target itself and the environment in which the scene is situated. But even with such uncertainty, people can be distinguished from different nationalities. When the characteristics of the information are judged, people can identify some objects through some local characteristics of the same object [15]. Local commonality allows us to use less resource to get the information that we need, so that the time and effort can be saved without tedious data analysis. Image local feature descriptor is one of the best matching and most widely studied algorithms. It not only has the characteristics of translation, scaling and rotation invariance, but also has good robustness to changes in illumination, affine and projection.

Local feature descriptor has the following characteristics: (1) local feature of image has stability and invariance to some extent; (2) The matching is fast and accurate with unique characteristics, and the information is rich; (3) it has a large quantity, and a few objects can produce a large number of eigenvectors; (4) it has the high speed of the optimized matching algorithm; (5) the scalability with other forms of feature vectors is very convenient. There are two steps of image registration by using SIFT algorithm: one is the extraction of SIFT feature points; another is the feature point matching.

Just like a light in a dark area and a black spot in a bright region, it still remains the same even when the light conditions change. In addition to the very stable feature points extracted by the SIFT registration algorithm and the corner and edge points in the image, there are some local extreme points. Usually, the transform relation of the image can be calculated by matching points. Typically, the matching points have the following characteristics. If the two images are registered in the same target area, the corresponding feature points and the corresponding relationship based on the SIFT feature points of the two images can be obtained. In the SIFT algorithm, the most important thing is to extract feature points. However, the premise of extraction is to build a multi-scale space with stable feature points to extract the invariant feature points of these scales. But in order to further accurately determine the location of the feature point, the interference of the unstable points should be eliminated, such as noise elimination and so on. These feature points are extracted for registration and the generation of SIFT feature descriptor at last. Through the main direction and auxiliary direction of feature points, the feature points can be rotationally invariant.

Usually it is arranged in the shape of Pyramid, and the resolution of each layer of the image in Pyramid is raised from top to bottom. In order to generate a space of Pyramid image, low resolution filtering and sampling are used for input brightness images, and multi-resolution processing mechanism is generally used. The bottom is high resolution image, and the top is low resolution image. The structure of the image gradually decreases from bottom to top and becomes smoother gradually. According to the spatial structure of Pyramid, the effect of noise on images can

be reduced. The proposal of Pyramid provides strong evidence for our subsequent analysis. The scale of feature points is invariant because Gauss convolution kernel is used to establish scale space.

The two-dimensional Gauss function is

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2},$$

where  $(x, y)$  is the pixel coordinate position and  $\sigma$  is the scale factor. The larger the  $\sigma$  value, the more blurred the image. The image is represented with  $I(x, y, \sigma)$ . The image  $I(x, y, \sigma)$  of different scales is obtained by calculating the product of image and two-dimensional Gauss function, and the formula is

$$L(x, y, \sigma) = G(x, y, \sigma) \times I(x, y, \sigma).$$

The Gauss difference scale space is  $D(x, y, \sigma)$ , which is obtained by different Gauss differential kernel product images:

$$D(x, y, \sigma) = G(x, y, k\sigma) - G(x, y, \sigma) \times I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma).$$

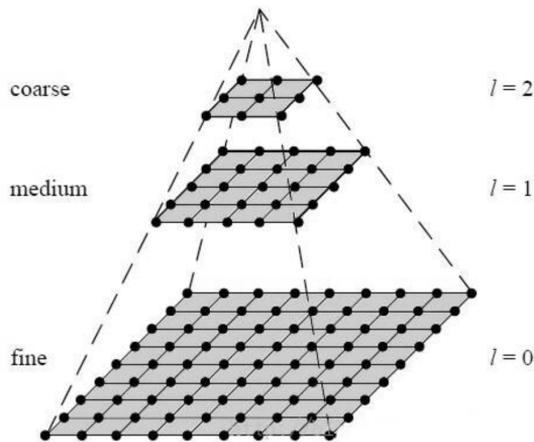


Fig. 1. Image pyramid

In order to detect the maximum and minimum points of  $D(x, y, \sigma)$ , the key points are selected according to whether all of the sampling points can be greater than or less than the 26 adjacent points. So in order to judge, these 26 points are used to compare with each sampling point. Since DOG values are sensitive to noise and edge, the candidate feature points can be detected when the above method for detection is used: there are some low contrast points or some edge response points. These unstable candidate feature points need to be eliminated. The extremum point  $x_0$  and the corresponding extreme value  $D(x_0)$  are obtained by the derivation from the second-order Taylor expansion of the DOG function. If the value of  $D(x_0)$  is not

greater than the set threshold, the point is deleted.

In addition, it is necessary to take the edge response point into account and eliminate it. The method of trace and determinant ratio of Hessian matrix is used.

The Hessian matrix is defined as

$$H = \begin{bmatrix} D_{XX} & D_{XY} \\ D_{XY} & D_{YY} \end{bmatrix},$$

the trace of the matrix is

$$\text{Tr}(H) = (D_{XX} + D_{YY})$$

and the determinant of the matrix is

$$\text{Det}(H) = D_{XX}D_{YY} - (D_{XY})^2.$$

If the value of  $\text{Tr}(H)^2/\text{Det}(H)$  at the key point is not greater than the set threshold, the point is removed. The final threshold is set at 10, which is calculated by Lowe continuous experiments.

A feature descriptor with rotation invariant properties has been proposed by Schmid. But it is limited by its poor uniqueness and poor matching accuracy, and the descriptor does not take into account the direction. Therefore, the neighborhood pixel gradient direction is added into the SIFT algorithm. In order to improve the matching accuracy, it is guaranteed that the feature descriptor has rotation invariance. The following formulas represent the gradient direction and amplitude calculation of the feature points  $(x, y)$ :

$$\theta(x, y) = \arctan \left[ \frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)} \right], \quad (1)$$

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2}, \quad (2)$$

where,  $L(x, y + 1)$ ,  $L(x, y - 1)$ ,  $L(x - 1, y)$ ,  $L(x + 1, y)$  represent the gray values of the upper, lower, left, and right pixel points of the point  $M$ .

When the key point is the center, the histogram is used to calculate the gradient direction. According to the histogram of the gradient direction, the angle range is 0–360 degrees, and 36 columns with 10 degrees as a column can be obtained. The next thing is to find our peak representation, which represents the direction of our key points. But there are likely to be several main directions, at this time, it is necessary to judge. According to the robustness of the matching, our auxiliary direction needs to be judged whether it has 80% energy of the main peak.

With the premise of not rotating, take the key point as the center and go to a  $16 \times 16$  window. The main direction of the key is rotated around the coordinate axis. Then, the gradient direction and amplitude of each pixel are calculated with the  $16 \times 16$  pixels as the center. After these pixels are computed, the Gauss weighting is

also necessary. Through the pixels above, the window can be divided into  $4 \times 4$  windows. Histograms of gradient directions in 8 directions are computed on each small window. The cumulative values for each gradient direction are plotted. From the above division, the feature vector of 128 dimensions can be obtained, which consists of  $4 \times 4$  16 seed points, each of which has vector information in 8 directions. Through such a joint domain method, we can effectively enhance our anti-noise performance and have a good fault tolerance for the positioning error.

The matching relation between images is set up by matching, so the image mosaic is finished. However, matching pairs are often found after extracting SIFT features, in which there are always incorrect matches. In order to reduce this situation, the matching must be focused on the extracted feature points. For the initial matching between two pairs of images, it is necessary to take the following two steps to match them.

1. Through the RANSAC algorithm, the error matching of initial matching points need to be eliminated to improve the accuracy. The initial matching points require that the nearest feature points are divided by the distance near them, and the ratio is the initial matching point that we need.

2. By referring to two original images, a feature point can be extracted from it. An efficient search of the nearest and sub adjacent points in a floating image is performed by using the BBF algorithm. But the initial matching must be judged by the Euclidean distance between the nearest feature points and the Euclidean distance between the sub adjacent feature points. The ratio is determined by the size of a specified proportional threshold.

The incorrect matches directly affect the parameters of the registration affine transformation, and then affect the image after stitching. The serious error matching results in the inaccurate transformation relation between the images. This situation occurs in the initial matching pair, which eventually leads to the poor accuracy of the stitching. The correct number of matches is obtained by using the RANSAC algorithm. After continuous screening, in order to be able to match more precise and let the final mosaic can have good accuracy, RANSAC algorithm is used to further screen in the correct matches. The original RANSAC algorithm is improved, and its improved algorithm steps are as follows.

1. Four pairs of matched pairs are extracted at random from the initial pair of points. The four pairs are set as initial interior points, and the transformation matrix  $H$  is calculated.

2. The distance between outside point of the set and the matching points after the transformation matrix is calculated. By setting a distance threshold  $T$ , the distance obtained can be judged. If it is greater than  $T$ , the remaining points are continued to be judged. If it is smaller than  $T$ , all previous points are added to the inner set.

3. The number of interior points under the transformation matrix  $H$  is counted.

4. After repeating the above three steps, one of the largest number of inner points is selected to compare with the threshold. If the quantity is greater than the threshold, the interior point is used as the initial value of the RANSAC. The change matrix  $H$  is re-calculated and the RANSAC is estimated. So set of interior points

under the new transform matrix  $H$  is the union of the original set of interior points and the new set of interior points.

5. The accurate matching point pairs are the all the feature points included in the best set with the largest number of interior points.

### 3. Result analysis and discussion

PC (Intel (R), Pentium (R), Dual, T2330@1.60GHZ, 8 G, memory, Windows, XP) were used as the adopted setting environment. The following points were needed to draw conclusions: the first step was to extract features from the SIFT and match the features. For the overlapping regions of the two images, feature extraction and matching were performed. The second step was screening, the initial screening was carried out. After screening, images were filtered by the improved RANSAC method for the second time to obtain accurate matching. The effects are compared in Figs. 2–7. Then images were fused by wavelet transform to obtain the final stitching image. The results show that in order to obtain more accurate stitching image, it is necessary to use the feature registration method described in this paper under the premise of improving the image registration parameters. In the process of matching the original image with the SIFT algorithm to register, it is necessary to adjust the value of the scale parameter. Table 1 shows the results for various values of threshold.

Table 1. Data results for various threshold values

Ratio threshold	Time (s)	Feature points extracted from reference images	Feature points extracted from floating images	Extraction duratio (s)	Matchpairs (unit)	Match time (s)	Total match time (s)
0.6	3.28	1783	1875	2.04	336	4.66	9.98
0.5	2.79	1783	1875	2.39	295	4.68	9.86
0.4	2.30	1783	1875	2.39	247	4.79	9.48

Fig. 4 shows the matching result of the two images with the proportional threshold of 0.6. Fig. 5 shows the matching result of the two images with the proportional threshold of 0.5. Fig. 6 shows the matching result of the two images with the proportional threshold of 0.4. The above experimental data shows that the total use time of the above three proportional thresholds was not much different. So the total match time did not have much influence on the proportional threshold. However, when the threshold was not used, the match pairs had obvious difference. Therefore, the appropriate proportion of the threshold was mainly reflected in the number of matching. When the threshold was set to 0.6, the number of incorrect matches was too high, but the total number of matches was higher. But when the threshold was proportional to 0.4, the number of matches was relatively small. Regardless of the



Fig. 2. Reference image



Fig. 3. Floating images



Fig. 4. Matching results with a threshold of 0.6

number of incorrect matches, it directly affected the final stitching effect. When the



Fig. 5. Matching results with a threshold of 0.5



Fig. 6. Matching results with a threshold of 0.4



Fig. 7. Stitched image using the method proposed in the paper

scale threshold was set to 0.5, the number of matches was big, and more accurate matches could be obtained. Therefore, the threshold in SIFT was selected as 0.5.

#### 4. Conclusion

With the increasing demand for image acquisition in society, an accurate image registration method based on SIFT features was proposed in this paper by analyzing SIFT arithmetic extraction feature descriptor, so as to improve the registration

accuracy, and the SIFT descriptor had a strong matching rate. However, in order to achieve more accurate of the matching number, the proportion of the threshold was adjusted, and the influence of matching threshold was analyzed according to the change of proportional threshold. In order to obtain seamless stitching images, the initial matching was obtained by using the Euclidean distance in the similarity criterion. The improved RANSAC method could further purify the matching pairs and obtain the desired mosaic images. The reason why the precisions of matching information obtained were different was that different thresholds were used with the RANSAC method to filter initial matches, although the different proportional thresholds had less impact on the total match time, they could directly affect the number of pairings, so as to affect the accuracy. So in the future, the SIFT feature points of a region can be calculated for corresponding registration, so as to improve our further registration effect through more direct and specific goals.

## References

- [1] W. HU, L. ZHANG, S. LIU, C. SHI: *An algorithm on registration of multi-view range images based on SIFT feature matching*. Journal of Computer-Aided Design & Computer Graphics 22 (2010), No. 04, 654–661.
- [2] Y. NA, D. WEN: *An Effective Video Text Tracking Algorithm Based on SIFT Feature and Geometric Constraint*. Pacific-Rim Conference on Multimedia, 21–24 September 2010, Shanghai, China, Conference Paper LNCS, Springer, Berlin, Heidelberg 6297 (2010), 392–403.
- [3] T. Y. BAI, X. B. HOU: *An improved image matching algorithm based on SIFT*. Transactions of Beijing Institute of Technology (2013), No. 06.
- [4] J. Z. J. ZHAO, L. J. XUE, G. Z. MEN: *Optimization matching algorithm based on improved harris and SIFT*. International Conference on Machine Learning and Cybernetics, 11–14 July 2010, Qingdao, China, IEEE Conference Publications 1 (2010), 258–261.
- [5] X. T. WANG, Y. XU, F. GAO, J. Y. BAI: *An image matching algorithm based on SIFT and invariability of feature points set*. Applied Mechanics and Materials 121 to 126 (2011), Chapter No. 2, 701–704.
- [6] D. Z. CHENG, L. I. YAN-JUN, Y. U. RUI-XING: *Image matching method based on improved SIFT algorithm*. Computer Simulation (2011), No. 07.
- [7] M. Y. YIN, F. GUAN, P. DING, Z. F. LIU: *Implementation of image matching algorithm based on SIFT features*. Applied Mechanics and Materials 602–605 (2014), Chapter No. 5, 3181–3184.
- [8] H. Q. ZHANG, L. G. CAO: *An image matching algorithm based on SUSAN-SIFT algorithm*. Applied Mechanics and Materials 325–326 (2013), Chapter No. 14, 1588–1592.
- [9] B. ZITOVA, J. FLUSSER: *Image registration methods: A survey*. Image and Vision Computing 21 (2003), No. 11, 977–1000.
- [10] F. TIAN, Y. B. YAN: *A SIFT feature matching algorithm based on semi-variance function*. Advanced Materials Research 647, (2013), Chapter No. 4, 896–900.
- [11] A. WANG, D. LU, Z. WANG, Z. FANG: *Research on non-rigid medical image registration algorithm based on SIFT feature extraction*. Journal of biomedical engineering 27 (2010), No. 4, 763–768,784.
- [12] W. T. WANG: *Multi-sensor image registration algorithm based on SIFT points and canny edge features matching*. Computer Science 38 (2011), No. 07, 287–289.
- [13] X. DAI, S. KHORRAM: *A feature-based image registration algorithm using improved chain-code representation combined with invariant moments*. IEEE Transactions on Geoscience and Remote Sensing 37 (1999), No. 5, 2351–2362.

- [14] X. ZHANG, Y. S. ZHANG, H. YAO: *Image feature matching based on SIFT algorithm*. Applied Mechanics and Materials 644–650 (2014), Chapter No. 6, 4157–4161.
- [15] Y. W. WANG, H. L. YU: *Medical image feature matching based on wavelet transform and SIFT algorithm*. Applied Mechanics and Materials 65 (2011), 497–502.

Received June 29, 2017



# Data acquisition and fusion system based on wireless sensor

DAN QIU<sup>1</sup>, SHULI GONG<sup>2</sup>

**Abstract.** In order to further promote the research on the technology of data acquisition and fusion in the process of using the wireless sensor technology and to improve the optimization of the installation, use, maintenance, performance and many other indicators of wireless sensor devices, the requirements of wireless communication field were analyzed, and the fusion algorithm was innovatively adopted with the research background of motor experiment platform and the data acquisition, and the optimal weighting factor of a single wireless sensor which can effectively reduce invalid data and improve the transmission efficiency was obtained in this paper; then, it was applied to the data fusion system, so that a new data acquisition and fusion system based on the wireless sensing was successfully obtained. The experimental results prove that this system model has good effects in aspects of data acquisition, transmission, fusion, low-power consumption, communication test and so on.

**Key words.** Data acquisition system, wireless sensor, fusion algorithm.

## 1. Introduction

With the development of microelectronics, machinery, information transmission, wireless sensor networks have been widely used in many fields, such as data acquisition, medical and health care, environmental detection, smart home, and national defense, military affairs and so on. At the same time, data acquisition has also undergone a series of changes with the in-depth research of computer technology, so it has put forward higher requirements in terms of transmission quality, efficiency, accuracy and so on.

In 1990, the United States used wireless sensor networks to carry out military research works to monitor battlefield conditions, locate targets accurately and monitor enemy equipment and so on.) [1]. Zhang Guojun of Beijing Forestry University introduced the wireless sensor network technology with low-power consumption into forest fire monitoring and created a forest fire monitoring system based on wireless

---

<sup>1</sup>Zhongshan Vocational College, Nanjing, Jiangsu, 210000, China

<sup>2</sup>College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 210000, China

sensor technology to monitor the temperature, humidity and other related indicators of designated areas in the forest and to send them to the computer for analysis, at the same time, this system achieved the advantages of energy saving [2]. In 2005, the U.S. Military used wireless sensor technology to successfully test the location system based on the gunfire and effectively construct the electronic defense system, so as to provide the technical basis for obtaining the enemy's military intelligence. Melbourne University in Australia and James Cook University constructed a wireless sensor system for detecting the coastal climate, wind direction, water temperature, water pressure, and so on, and realized the promotion of environmental monitoring [3].

The article is arranged in five chapters. In the first chapter, the current international researches on wireless sensor technology to collect data were mainly introduced; in the second chapter, the general design idea of wireless sensor system for data acquisition and fusion was analyzed; in the third chapter, combined with the analysis of some examples, the data fusion algorithm for wireless sensor was studied through the creation of application model; in the fourth chapter, in terms of the data acquisition, communication, data fusion, performance and power consumption, the data acquisition and fusion system of wireless sensor was run and tested; in the fifth chapter, researches of data acquisition and fusion system for wireless sensor were summarized with the prospect of future researches.

## 2. State of the art

After entering the 21 century, China has made progress in the field of the research of wireless sensor network. At the same time, university teaching has opened courses about the wireless and sensing technology. In addition, our country has provided the financial assistance for some representative wireless sensor projects. Among them, in 2009, China's Ministry of Industry and Information Technology launched a new broadband mobile communications network, and set up the research and development and industrialization of the short distance wireless Internet and wireless sensor network.

In the process of in-service, wireless sensor equipment is often difficult to be maintained due to the effects of the environment and force majeure and other factors, so the security problem becomes the biggest problem. Therefore, with its limited memory spaces and computing powers, it is a breakthrough point to improve the equipment adaptive environment and data acquisition, calculation, transmission and other capabilities [4]. In terms of the data fusion, the number of ports needs to be reduced as much as possible to improve the data inefficiency and to achieve more optimized data fusion. Based on the technology development of wireless sensor network, system detection performances and construction costs have broad prospects for development.

### 3. Methodology

#### 3.1. Routing algorithm

The network involved in this study can support the mesh routing algorithm and tree routing algorithm of the two algorithms. In which, the tree routing algorithm is mainly applied to the tree structure network in the communication protocol, its principle is the allocation mechanism based on the network address, and each router has a certain address space, which is used to allocate ports. Therefore, the tree routing algorithm was adopted according to the requirements of this study.

Communication protocols usually need to allocate addresses and follow them throughout the device's use cycle to identify devices or send data over the network. In the early stage of network construction, the number of logical port needs to be correlated and constructed, and the sequence relation of solidification needs to be determined. When data is transmitted, this sequence of relationships needs to be strictly followed, and the address should be assigned completely after the completion of the network construction. Assuming that the maximum number that a higher level device may have the next level of equipment is  $C_m$ , and the maximum number of routers is  $R_m$ , while the maximum depth of the network is  $L_m$ . If these parameters are determined by the communication protocol, then the network address offset can be calculated as follows:

$$Cskip(d) = \begin{cases} 1 + C_m * (L_m - d - 1), & R_m = 1, \\ \frac{1 + C_m - R_m - C_m * R_m^{L_m - d - 1}}{1 - R_m}, & \end{cases} \quad (1)$$

When a device has an offset of 0, it does not have the ability to redistribute the port address of a lower device. That is, it is unable to access the network through other ports. If the offset is greater than 0, then it can accept other devices as its subordinate and assign the corresponding network address.

Table 1. Network depth and offset

Network depth	Offset
0	21
1	5
2	1
3	0

#### 3.2. 3.2 The self-processing mechanism of routing failures

In the network of meshed structures, routers usually have two operating states: the network operating state and the backup standby state. For the network, the most important problem is the switching of the two states. If the normally operating router has not consumed all of its energy or has not yet completed the task, the backup port will replace it, but which will waste energy and reduce the network

lifespan. However, when the router uses up energy, the backup route is started, which will also cause some data loss [5].

In general, the tree routing algorithm in communication protocols takes a passive maintenance mechanism. When it can communicate but cannot transmit data, it attempts to resend data; when it is completely unable to transmit data, it will try to re-establish the connection. As a result, the quality of data transmission can be seriously affected. In order to solve this problem, a self-processing mechanism for feedback information needs to be created [6].

During the process of network operation, whether the routing port is normal can be determined according to the detected link quality indexes, so that whether to carry out the conversion between different states can be determined.

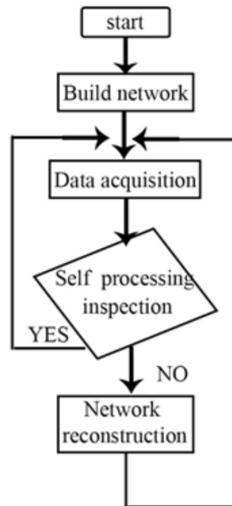


Fig. 1. Self-processing mechanism model

When an abnormal network is determined, the communication protocol will clear the fault port from the network through the coordinated processing, and then it will inform the standby port in time to switch to the working state. Prior to this, the lower port of the faulty port will also switch to the working state. On this basis, the network structure table will be updated and the fault information will be cleared [7].

### 3.3. Data fusion algorithm

Due to the different measuring data accuracies and measuring environments of the wireless sensor, there will be errors in the measurement. Therefore, each sensor data needs to be computed, and arithmetic averages should be taken to reduce the accuracy of the results. The adaptive fusion algorithm can adjust the original data measured by each sensor and fuse the minimum variance data fusion value. The mean square deviation after fusion is less than the mean square deviation of one or more sensors [8]. Assuming that there are  $n$  sensors to collect data, and then

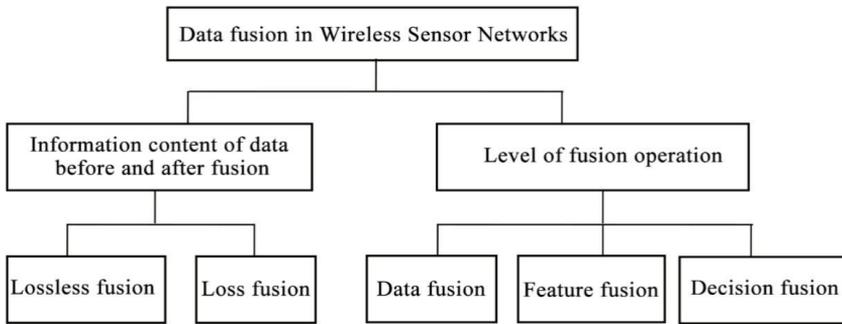


Fig. 2. Classification of data fusion in Wireless Sensor Networks

the adaptive weighted fusion algorithm is to find the corresponding factors of each sensor as  $W_1$ ,  $W_2$  and  $W_3$  under the condition that the total mean square error is minimum, so as to make the fused  $\hat{x}$  reach the optimum.

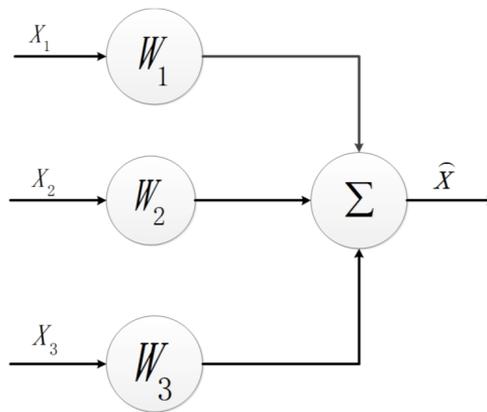


Fig. 3. Adaptive weighted fusion algorithm model

In the above model, the wireless sensor will produce a large number of related data on the mining network, and its terminal system will collect the data in time. Taking the collected humidity as an example, the terminal node will collect data every once in a while and store the collected data into memorizer according to the changes of humidity. When multiple data is collected, the fusion calculation can be carried out, and the mean and variance can be obtained, finally these mean and variance can be sent to the router. When the router receives the mean and variance of multiple sensor nodes, the optimal values and fusion values of each sensor can be calculated and sent to the coordinator.

With the humidity as an example, the humidity data of 4 sensor nodes is collected. In order to verify the effectiveness of the adaptive weighted fusion algorithm in eliminating error data and improving the accuracy, one of the nodes is artificially intervened to simulate the occurrence of the fault.

Table 1 shows the five humidity data collected by the five sensor nodes during a collection cycle in summer (unit: %rh).

Table 2. Five sets of humidity data

	The first time	The second time	The third time	The fourth time	The fifth time
node 1	40	50	45	49.9	50
node 2	45	45.7	46	50	51
node 3	60	59	59.5	60	60
node 4	75	74.1	74.9	70	70
node 5	50	52.3	53	55	50

### *3.4. Data fusion algorithm in the system*

In the adaptive weighted fusion algorithm, the adjustment of weighting factors can effectively eliminate the impact on the final results, and calculate the average value of the data collected by the sensor. In order to make the results more accurate, the more accurate results of wireless sensor nodes must be obtained finally to achieve the most accurate final estimate. In this research, the fractional operation of the sensing measured data was carried out, and the corresponding values were calculated. On this basis, the overall measurement variance was estimated through the weighted calculation [9].

In the wireless sensor data acquisition system designed in this paper, the data measured by each sensor was fused. The estimated value and variance were calculated by the batch estimation fusion algorithm, and then the adaptive weighted fusion algorithm was used to obtain the measured estimated value and variances of the individual sensors, so that the optimal weighting factor of each sensor was calculated, and the final results were calculated. Thus, even in the presence of error data, the two steps of using batch estimation and adaptive weighting fusion can also gradually weaken the influence of error data on the final result and obtain more accurate data.

## **4. Result analysis and discussion**

### *4.1. Hardware design of low-power network nodes*

Wireless sensor network data acquisition system is composed of low-power network nodes, and it mainly includes coordinator and router and many other different terminal nodes with different functions. In general, a system composed of wireless communication, sensors, power supplies, processors and other auxiliary modules has the characteristics of low-power consumption. The corresponding module can be selected according to the different types of low-power network nodes. For example,

the coordinator node sends the data sent by the terminal node to the host computer and select different modules according to different requirements [10].

Wireless sensor network nodes in the wireless sensor data acquisition and fusion system use the latest chip CC2430 launched in the United States, the chip has the characteristics of low energy consumption, so it can realize the signal transmission with low-power. Semiconductors fabricated with metal oxide materials can meet the needs of this application, thus achieving low-power requirements and reducing costs [11]. Chip CC2430 integrates a low-power video front-end, microcontroller, and memory on a single chip; in addition, it also has powerful peripheral resources, so the current consumption is low in the pick-up mode. Chip CC2430 has the function of low-power consumption and high performances. The current consumption is small in the resting state, and a large number of circuits are integrated in the interior. There is serial port circuit on the low-power coordinator node of data collection. In the design, the serial port is used to realize the data transmission between PC and chip CC2430.

In the design process, the problem of level switch needs to be solved to achieve the communication between the chip CC2430 and the PC serial port, and the driver is responsible for the level switch.

In the research of the system, the sensor circuit design mainly includes the sensors in the voltage, the current, and the temperature of the three aspects. For the voltage sensors used in voltage collection, the circuit board of a welded voltage sensor in the process of using needs to pay special attention to the problem of zero drift of voltage sensor, and which needs the timely compensation. Therefore, in the process of using, zero point circuit also needs to be designed to ensure the accuracy of measurement data. The use and design of the current sensor are the same as the voltage sensor, because it also requires the actual zero point circuit to ensure that the system output is 0 without output. The temperature sensor sends or receives information via a single wire interface, so it only needs one interface.

#### *4.2. System software design*

In the software design of this system, the cross compiler and debugger are the most perfect development tools, including the optimizing compiler, assembler, connection locator, and editor, and so on. The integrated compiler has the characteristics of efficient code, fast interrupt processing and optimization of memory mode. The C/C+ compiler of EW can generate efficient and reliable executable codes, and its applications are large with obvious results. Compared with other development tools, the system can use both global and specific chip optimization techniques [4].

In the whole low-power wireless communication technology network, the coordinator can create a new network. Works that the coordinator node needs to complete include: firstly, the coordinator node needs to establish a new network to accept other nodes and allocate the specific network address for them; secondly, the coordinator node needs to accept the data sent by each sensor node and send it to the PC machine through the serial port, and uses serial port to debug the data, so as to show the data. The establishment of a short-time and low-power wireless

communication network is mainly initiated by a network coordinator. A fully functional node device with the coordinator capability can effectively monitor scanning and the possible interferences, thus establishing a short-time and low-power wireless communication network. When a suitable channel is selected, the network address, coordinator, network address, network identifier and topological parameters of the network can be determined. After each parameter is determined and the network is established successfully, the coordinator node can accept other nodes to join the network [12].

After the router node in the system is initialized, the relevant parameters can be set up, so as to search the network, join the network and get its own network address as well as determine its working states. It is important to note that the backup nodes do not perform data forwarding. Routers can receive new nodes and add them into the network to find the corresponding network address. As required, the router node which works normally can complete the data forwarding, at the same time, in order to obtain more accurate data and reduce the amount of data sent to the coordinator, the router node also needs to complete data fusion. Data acquisition nodes can join the network via routers and send the data collected by sensors [13].

### *4.3. Low-power design*

In the wireless sensor network data acquisition system, once the energy of the data acquisition node is exhausted, the node will be ineffective and the system life cycle will be terminated. Therefore, under the premise of ensuring the normal operation of the system, minimizing the power consumption and prolonging the life cycle are the key aspects of the system design. The coordinator node which is responsible for data collection is powered by mountain power, so there is no need to consider the energy consumption, but the nodes which are responsible for data acquisition need to be taken into account. Among them, the energy supply module is responsible for providing energy to other three modules to ensure their normal operation. Energy consumption is mainly divided into three parts: the energy consumption associated with communications, the energy consumption of sensor module data acquisition and the energy consumption of the processor module for data processing [14].

### *4.4. Running and testing*

In view of the model designed and researched in this paper, the network construction, operation and test works on the aspects of data acquisition, communication and integration and so on were carried out.

Firstly, 9 nodes were constructed under laboratory conditions, including 2 serial modules which can be used for coordinating data acquisition and transmission, 3 sensor modules which can be used to collect data and the routers and terminals that can be used to test content.

Secondly, a tree based network which consists of 1 coordinated port, 2 router ports, and 5 terminal ports was constructed, as shown in Fig. 4.

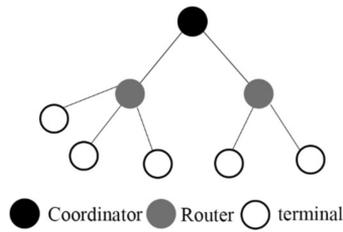


Fig. 4. Network structure diagram

After the completion of the network structure, the terminal port can collect relevant data periodically according to the pre-set command. In the process of data transmission, the data loss often occurs due to the far distance, obstacles, interference and other reasons. In order to solve this problem, the data transmission cycle needs to be limited, and the reasonable acquisition cycle should be set.

After the test, it can be seen that through the formation of the overall system, each port can operate stably and achieve efficient and reliable data transmission; loss phenomenon can be greatly reduced; the power consumption can be effectively reduced; and the working time can get a greater degree of protection.

The whole process of data acquisition and fusion of the wireless sensor is shown in Fig. 5.



Fig. 5. Data acquisition and fusion of wireless sensor

## 5. Conclusion

In order to better study the popularization work of wireless sensor, the wireless communication technology with low-cost and low-power was selected in this paper. By understanding the current development situation of wireless sensor in data acquisition and fusion, the basic situation of data acquisition and fusion system generated by wireless sensor was analyzed. Based on the use of these two technology systems, the core products of the hardware platform were determined. In view of the deficiency of this protocol, appropriate improvements were carried out, and an active detection mechanism that can effectively reduce the failure rate was designed. At the same time, weighted algorithms for data fusion were added, and related performance tests were carried out. The result reached the basic expectation. Finally, the overall test of the system was carried out on the basis of the completion of the functional modules, and the expected effect of this research was achieved.

The above functions were implemented in this paper. However, in order to promote the technology to more areas, there are few places that can be improved: the first is how to effectively expand the sensor power supply ways to extend the service life; the second is to consider the address optimization and try to reduce the waste phenomenon in the network design; the third is to improve the human intelligence development of the system, so as to maximize the role of remote control and role management functions.

## References

- [1] L. Q. YANG, J. C. ZHAO, Y. ZHOU, S. MA: *Design of data acquisition system based on wireless sensor network*. *Microelectronics & Computer* 24 (2007), No. 9, 68–71.
- [2] S. XIE, Y. WANG: *Construction of tree network with limited delivery latency in homogeneous wireless sensor networks*. *Wireless Personal Communications* 78 (2014), No. 1, 231–246.
- [3] J. H. SHORE, M. ALDAG, F. L. MCVEIGH, R. L. HOOVER, R. CIULLA, A. FISHER: *Review of mobile health technology for military mental health*. *Military medicine* 179 (2014), No. 8, 865–878.
- [4] J. N. WANG, W. WANG: *The electric quantity data acquisition system based on wireless sensor network*. *Applied Mechanics and Materials* 239–240 (2013), 879–883.
- [5] W. QIU, Y. ZHANG: *Physiological signal acquisition system based on wireless sensor networks*. *Journal of Southeast University* 26 (2010), No. 1, 73–77.
- [6] X. WANG, H. QIN: *Implementation of fall detection system based on data fusion technology*. *International Journal of u- and e- Service, Science and Technology* 9 (2016), No. 4, 1–8.
- [7] J. WANG, X. LI: *Progresses on data fusion technology of crop growth model and multi-source observation information*. *Remote Sensing Technology and Application* 30 (2015), No. 2, 209–219.
- [8] H. FOURATI: *Heterogeneous data fusion algorithm for pedestrian navigation via foot-mounted inertial measurement unit and complementary filter*. *IEEE Transactions on Instrumentation and Measurement* 64 (2015), No. 1, 221–229.
- [9] S. KIM, J. Y. CHOI, S. HAN, M. R. YONG: *Adaptive weighted fusion with new spatial and temporal fingerprints for improved video copy detection*. *Signal Processing: Image Communication* 29 (2014), No. 7, 788–806.

- [10] F. BOCCARDI, R. W. HEATH, A. LOZANO, T. L. MARZETTA, P. POPOVSKI: *Five disruptive technology directions for 5G*. IEEE Communications Magazine 52, (2014), No. 2, 74–80.
- [11] B. M. HUNTER, J. D. BLAKEMORE, M. DEIMUND, H. B. GRAY, J. R. WINKLER, A. M. MÜLLER: *Highly active mixed-metal nanosheet water oxidation catalysts made by pulsed-laser ablation in liquids*. Journal of the American Chemical Society 136 (2014), No. 38, 13118–13121.
- [12] C. X. WANG, F. HAIDER, X. GAO, X. H. YOU, Y. YANG, D. YUAN, H. AGGOUNE, H. HAAS, S. FLETCHER, E. HEPSAYDIR: *Cellular architecture and key technologies for 5G wireless communication networks*. IEEE Communications Magazine 52 (2014), No. 2, 122–130.
- [13] A. OSSEIRAN, F. BOCCARDI, V. BRAUN, K. KUSUME, P. MARSCH, M. MATERNIA, O. QUESETH, M. SCHELLMANN, H. SCHOTTEN, H. TAOKA, H. TULLBERG, M. A. UUSITALO, B. TIMUS, M. FALLGREN: *Scenarios for 5G mobile and wireless communications: the vision of the METIS project*. IEEE Communications Magazine 52 (2014), No. 5, 26–35.
- [14] A. KINALIS, S. NIKOLETSEAS, D. PATROUMPA, J. ROLIM: *Biased sink mobility with adaptive stop times for low latency data collection in sensor networks*. Information Fusion 15 (2014), 56–63.

Received June 29, 2017



# Mining urban active point circle based on spatio - temporal constraint data<sup>1</sup>

RONG TAO<sup>2,3</sup>, MENGLUO JI<sup>2</sup>

**Abstract.** In order to detect the active points of cities and analyze the active circles in time and accurately, this paper proposes a method to detect urban active points and active circles based on the data of social network. A data preprocessing model based on discrete point is proposed to solve the problem of large data volume, discrete space-time constraint data storage and cluster analysis efficiency. Spatio-temporal constraint data were tested by spatial autocorrelation, which indicated that it had significant spatial clustering characteristics. This paper proposes an active clustering method based on spatio-temporal constraint data and explores the geography distribution of business factors to obtain active circle information. A city, for example, on the positioning network (www.dingwei.com) until September 30th, time and space constraints data test the active city point detection and active circle excavation test. The results show that there is a strong correlation between the active circle distribution and urban planning active circle based on spatio-temporal constraint data mining, which can be used to forecast the urban social economic development and regional economic planning.

**Key words.** Public-source geographic data, space-time constraint data, data mining, active-point detection, active-cycle distribution.

## 1. Introduction

As one of the driving forces of urban economic development, urban active circle is an important part of urban comprehensive competitiveness. The active circle dynamic measurement is an important basis for guiding the economic layout of the city. It plays a very important role in bringing into full play the social benefits and overall functions of the active circle, promoting the process of urbanization and promoting the sustained development of the national economy. At home and abroad active circle research mainly has macro angle and micro angle two aspects

---

<sup>1</sup>This work was support by the Henan province foundation and advanced technology research project in 2015 (No.152300410115).

<sup>2</sup>College of Computer and Information Engineering, Luoyang Institute of Science and Technology, 471023, Luoyang, China

<sup>3</sup>Corresponding author

[1]. The former extension of the measurement range is too wide, will make active lap determination doping many other factors, which focuses on the micro level of the enterprise active lap analysis, are not from the city perspective to the city active lap determination. In addition, commonly used active circle measurement method is generally used questionnaire survey method, takes more time and effort, limited scope of the investigation, affecting the active circle determination of accuracy and comprehensiveness [2].

The crowd sourcing geographic data is the open geographic data collected by the public and provided to the public [3–5]. The representative geographical data includes GPS trajectory data, map data compiled by users' collaborative annotation, and points of interest (POIs) for various social networking sites such as Twitter, Facebook, and the street (www. Dingwei. Com), etc. [6]. Compared with the traditional geographic information data, the public geographic data from the non-professional public has the characteristics of large data volume, good potentiality, abundant information and low cost, and has become the active research area of international geographic information science in recent years [7–10]. Space-time constraint data is a kind of data which has spatial, temporal and social attribute information produced by the intelligent terminal with GPS. It records the life trajectory and reflects people's daily life behavior, an important source of geographical data. Spatiotemporal constraints data concentrated in the city, and to the public to sign the points of interest as the main form. To positioning network, for example, since May 13, 2010 positioning network officially launched since the location network registered users to 20% per week to maintain the rapid growth rate, as of September 2011, the number of positioning network users has more than 1.2 million. As more than 70% of registered users will sign-up information and social platform binding, so the average sign-in will have 400 audiences. Location network every 5s (based on 24h basis) to update a user check-in information, including a wealth of location information, semantic information and behavioral information. Therefore, the space-time constraint data obtained from the positioning network not only rich in data, but also good potential, from the side to reflect the city's economic and cultural distribution situation. Based on spatial and temporal constraint data of positioning network, this paper presents a method to detect urban active point and active circle based on all-source spatial and temporal constraints data. Through data preprocessing, exploratory spatial analysis and spatial clustering analysis, constraint data high-value clustering active point.

## 2. Exploratory clustering and analysis of spatio-temporal constraint data

Spatio-temporal constraint data is a discrete GIS point object with spatial coordinates and user attributes. First, the discrete log-in data is lattice processed, and the large data volume and discrete sign-in point are transformed into spatial continuity and adjacency. Grid data that reflects the density of check-in events. Secondly, spatial correlation of measured data is measured by exploratory spatial data analysis (ESDA), and its spatial structure and global distribution pattern are

measured to determine the best mode of active point detection and active circle clustering. Thirdly, clustering analysis was used to identify the locations of active, cold and spatial anomalies with statistical significance. Finally, by measuring the spatial distribution of the clustering results, the spatial characteristics of clustering geographic elements are obtained, including the range of the active circle, the central trend and the direction of development. The concrete algorithm flow is shown in Fig. 1.

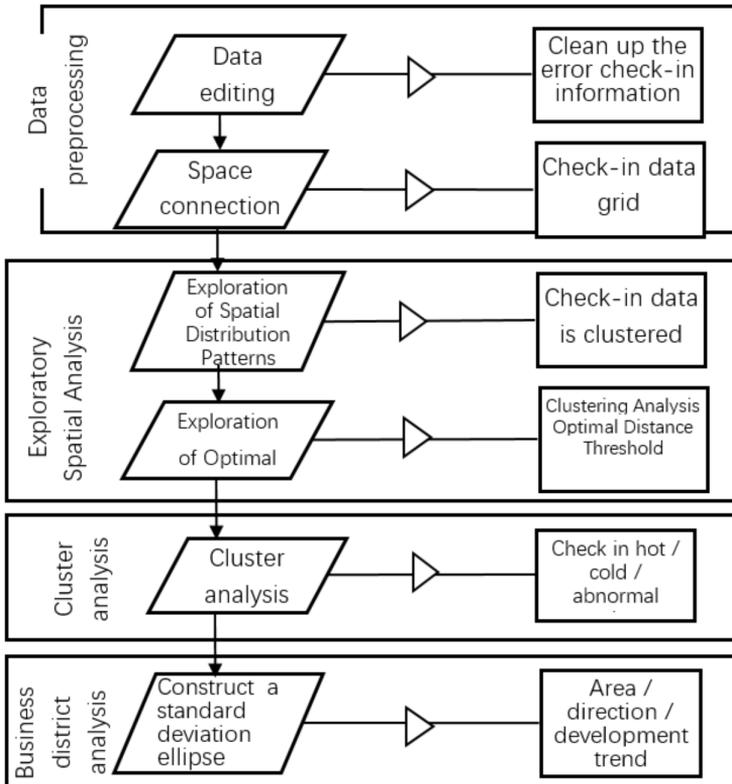


Fig. 1. Flowchart of active point detection and active circle mining algorithm based on spatio-temporal constraint data

### 2.1. Spatio-temporal constraint data grid processing

Spatiotemporal constraint data is a large number of discrete feature points, which does not have obvious spatial continuity and adjacency in space, and is not conducive to the exploration of spatial data analysis method to measure its spatial distribution pattern. In order to make space-time constraint data reflect both spatial continuity and proximity, and to preserve the characteristics of sign-in times and the statistical properties of key objects, this paper chooses the area covered by sign-in data as the research area, where  $G(W)$  is the grid  $G$ , and  $G$  is the space between the point-to-

sign and data-key attributes of the grid-containing data, and the corresponding grid attributes are mapped to the corresponding grid attributes [11, 12] of the sign

$$\left. \begin{aligned} G(W) &= \sum_{i=1}^n Np_i \times \sigma p_i, \quad \text{among them } P_i \subset (\{p\} \cap G), \\ G(T) &= T_k, \quad \text{among them } \sum \sigma p, T_k = \\ \max \{ &\sum \sigma p, T_1, \dots, \sum \sigma p, T_i, \dots, \sum \sigma p, T_s \}. \end{aligned} \right\} \quad (1)$$

Here,  $G(T)$  represents the area type of the grid  $G$ ,  $n$  represents the number of check-in points in the grid  $G$ ,  $Np_i$  represents the total number of check-in points of the  $i$ th grid in the grid  $G$ ,  $\sigma p_i$  represents the weighting level of the check-in point,  $\sum \sigma p, T_i$  represents the sum of the weights of all check-in points belonging to the  $K$ th class in the grid  $G$ . Figure 2 shows the algorithmic flow of gridding of discrete sign-on data grids.

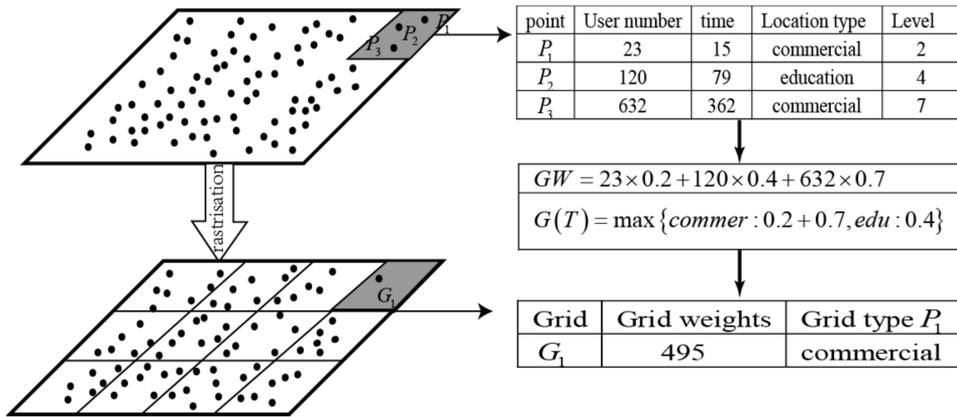


Fig. 2. Schematic diagram of grid-based processing of temporal-spatial constraint data

From Fig. 2, the discrete sign-in data is transformed into the grid-signed data set with the sign-in frequency as gray scale, which not only simplifies the discrete point data, but also preserves the time-space characteristic and thematic attribute characteristic of the signed data, satisfies the exploratory spatial analysis and data Mining requirements.

### 2.2. Spatio-temporal constraint data space autocorrelation test and analysis

Spatial autocorrelation is an important form of spatial dependency and a prerequisite for exploratory spatial data analysis (ESDA). In this paper, global spatial autocorrelation of space-time constraint data is studied by global Moran's statistical method, and Ripley's  $K$ -statistic method is used to explore the spatial distribution pattern with the strongest sign-in feature to provide a basis for sign-in data mining.

Given a set of spatio-temporal constraint data and their check-in frequency,

Global Moran's  $I$ -statistics is evaluated as a clustering pattern according to (2), where  $n$  is the number of check-in points,  $z_i$  being the sign-in position. Here

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{i,j} z_i z_j}{S_0 \sum_{i=1}^n z_i^2}, \quad (2)$$

where  $z_i$  is the check-in frequency of the deviation  $(x_i - \bar{X})$  from the mean,  $w_{i,j}$  is the spatial join matrix of the check-in position,  $S_0$  is the sum of all spatial weights, given by the formula  $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{i,j}$  denoting molecular normalization by variance, the index value being in the range  $-1.0$  to  $+1.0$ . Its positive value indicates that the sign-in frequency of the sign-in position has a positive correlation, and a negative value indicates a negative correlation, indicating that the spatial distribution of the spatial object is not correlated. This can be calculated in accordance with equation (3) to check that the value of the sign frequency is statistically significant—EOT.

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}}. \quad (3)$$

Here,

$$E[I] = \frac{-1}{n-1}, \quad V[I] = E[I^2] - E^2[I].$$

Figure 3 shows the spatial distribution pattern of spatial-temporal constraint data based on Global Moran's  $I$  statistical computation. The score value for the test is 8.003898 times the standard deviation, much larger than 2.58, indicating that the null hypothesis probability value is 0, consistent with the requirement of 99% confidence value (probability likelihood value). The global spatial autocorrelation of the spatial pattern of constraint data is in accordance with the statistical characteristics of the typical clustering model, which can be used for cluster analysis of active points and active circles.

### ***2.3. Geographical distribution measures of check pointing data clustering active points***

The high-value active point detected by local autocorrelation clustering can be regarded as the center of active circle, and its range, center change and direction should be further defined. In this paper, the active circle is studied by measuring the geographical distribution of commercial clustering active points, Specific steps are as follows:

*2.3.1. Standard deviation elliptical structure.* The standard deviation ellipses are constructed with the clustering active points as the center and the sign-in positions and their associated attribute values (frequency of sign-in). The ellipse center is the weighted average center of the feature in the clustering area, and the ellipse length and short axis of the feature distribution are defined by the standard distance



Fig. 3. Distribution of urban active points

in the  $x$  and  $y$  directions.

$$SDE_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}},$$

$$SDE_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}}, \quad (4)$$

Here  $SDE_x$  and  $SDE_y$  are the long and short axes of the standard deviation ellipse.

*2.3.2. Calculation of active circle range based on standard deviation ellipse.* Calculation of active circle range based on standard deviation ellipse. With the standard deviation ellipse length and the short axis as the spatial distribution of the active circle, the major axis is the main trend of the central trend of the active cycle.

*2.3.3. Determination of active circle direction.* The proportion of the long and short axes represents the flatness of the active circle distribution, and the rotational azimuth of the standard deviation ellipse is the development trend of the active circle. The rotation angle of standard deviation ellipse is calculated according to the formula

$$\tan \theta = \frac{A + B}{C}. \quad (5)$$

Here,

$$\Lambda = \left( \sum_{i=1}^n \tilde{x}_i^2 - \sum_{i=1}^n \tilde{y}_i^2 \right), \quad B = \sqrt{\left( \sum_{i=1}^n \tilde{x}_i^2 - \sum_{i=1}^n \tilde{y}_i^2 \right)^2 + 4 \left( \sum_{i=1}^n \tilde{x}_i \tilde{y}_i \right)^2},$$

$$C = 2 \sum_{i=1}^m \tilde{x}_i \tilde{y}_i.$$

Here,  $\tilde{x}_i$  and  $\tilde{y}_i$  are the standard deviations of the clustered feature points with respect to the center point of the ellipse.

The local standard deviation ellipse is constructed according to the geographical distribution measurement method and the range, the central trend and the trend of each active circle are calculated. The result shows the spatial-temporal constraint data of a certain city.

### 3. Experiment and discussion

#### 3.1. Analysis of detection results of active city based on spatio-temporal constraint data

According to the attribute information of the high-value sign-in point of each active area, this paper divides the active area of the clustering into commercial active points (such as Jiangnan Road), educational active points (such as a city university), tourism (Such as the East Lake area), traffic class active points (such as light rail station), living class active point (such as South Lake district) and other types of active points (such as restaurants, libraries) six categories, the spatial distribution shown in Fig. 3.

As shown in Fig. 3, the commercial activity points are clustered in the spatial distribution, which reflects the active circle distribution of a city. The number of commercial activity points aggregated by each active circle reflects the popularity of the active circle. Figures 3 and 4 in the mouth and the mouth of the street, the mouth of the street area of commercial high active point number, indicating a relatively large population of street population, the economy has maintained a relatively active state in the Simingkou area. On the contrary, the region’s commercial activity is only 3 points, indicating that the door area population flow is not, the economy is not active and there are signs of recession, and the two regional economic development in line. Figure 4 shows the statistical distribution of different types of active points. Of the 172 high-value sign-in elements, there are 90 commercial activity points, 43 educational activity points, 12 tourism activity points, 10 traffic activity points, class activity points 3, 14 other types of active points. It can be seen that the proportion of business activity points in space-time constraint data is the most, which validates the rationality of using spatio-temporal constraint data to analyze active circle.

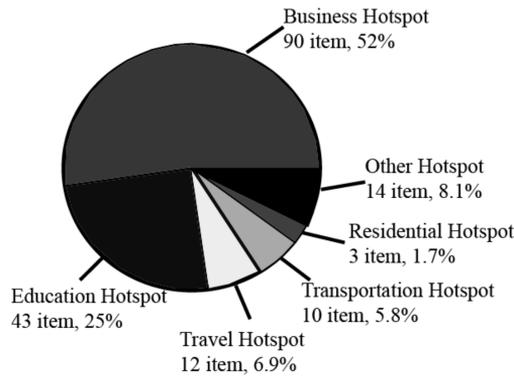


Fig. 4. Classification map of urban active points

### *3.2. Analysis of urban active circle of a city based on spatio-temporal constraint data*

The results show that a city has formed commercial and business sections with Wuchang, Hankou and Hanyang as the main pattern, and dozens of large and medium - sized active laps with a certain scale. Both the traditional sense of the old active circle, such as Jiangnan Road, Zhongshan Road, Wangjiawan, Zhongjiacun, Xudong, the door of the door, fruit lake, active circle in the South, there are emerging in recent years active circle, World, street mouth, Lu Xiangguang Valley active circle, but also includes some active in the construction and development of the circle, such as water chestnut Lake, Dunkou zone active circle. For each of the 90 commercial economic activity points, the standard deviation ellipses are used to compute the range, trend and direction of each active circle, and the spatial distribution of each active circle in a city based on sign-in data is obtained (Fig. 5). Through the spatial superposition analysis, the statistical results of the number of sign-in points, the number of sign-in times, the number of registered users, and the average number of sign-in times of users in each active circle are obtained.

It can be seen from Table 2, based on spatial and temporal constraints of data analysis of a city active circle with a city three towns active circle actual distribution. According to the statistics of active circles in three cities, the statistical distribution characteristics of Wuchang, Hankou and Hanyang in terms of business activity points, sign-in number and number of users coincide with the regional functional characteristics and population distribution characteristics. The number of active users in Wuchang has exceeded that of Hankou, and the average number of its users is the highest among the three towns. It can be seen that Wuchang, as the center of traditional science, culture and politics, has a tendency to overtake Hankou as a traditional economic center is relying on the East Lake High-tech Development Zone, the formation of policy-oriented Lvxiang Optics Valley active circle and the integration of Asian Trade Shopping Center, the IT port of the IT market in Guangzhou port active lap effect. In contrast, in a city of Hanyang Economic and Technological Development Zone, commercial scale in the initial stage of de-

velopment, although the average number of attendants and Wuchang quite, but the active circle of commercial outlets rather small. Compared with the Hankou business model, the Wuchang active circle is generally discrete in geographical distribution, showing a clear regional distribution of active laps. Hankou area, the traditional Jiangnan Road active circle, Wuhan-Guangzhou World Trade active circle is still playing a pillar of the role of mid-stream, a new city active circle, water chestnut Lake (Wanda) active circle also plays an increasingly important role.

Table 2. Active lap information statistics table

Active circle	Business active points	Attendance	User number	The average number of attendance
Central and South active circles	126	2206	1615	1.3569
Division door active circle	104	3019	2374	1.2717
Street mouth active circle	658	15412	9179	1.6791
Fruit Lake active circle	67	1038	672	1.5446
Lu Xiangguang active circle	407	14771	9264	1.5948
Xudong active circle	95	3463	2069	1.6738
Marshland active circle	71	2851	1704	1.6731
A city active circle	131	4955	3592	1.3795
Jiangnan Road active circle	421	4790	3293	1.4546
Northwest Lake active circle	258	4133	2882	1.4341
Wuhan-Guangzhou World Trade active circle	300	7781	5050	1.5408
Lake water chestnut active circle	48	3900	2621	1.4880
Zhongjiacun active circle	170	2184	1445	1.5114
Wangjiawan active circle	66	1377	957	1.4389
Dunkou active circle	34	869	515	1.6874
Wuchang	1457	39912	25173	1.5855
Hankou	1229	28410	19142	1.4842
Hanyang	270	4430	2917	1.5187

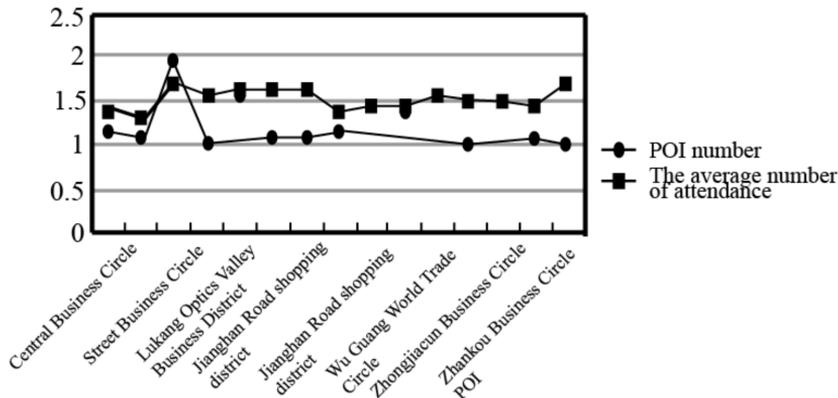


Fig. 5. Active circle POI and the average number of attendance

It can be seen from Fig.5 that in all active circles of a city, the number of commercial outlets and the average number of attendance in the street lap are all at the highest level in the whole city, and it is the largest IT computer in the city of Guangbaihuo, New World, Market, but also with a city traffic is currently the most congested point of the same, we can see that the source of space-time constraint data for active lap active point detection is very effective.

## 4. Conclusion

The emergence of all-source geographic data has provided a new data source for spatial data updating, and provided a new research direction for spatial data mining. In this paper, we propose a method to detect and analyze the active points of the active points of interest accumulated in the city for a long time. The clustering analysis based on the number of the endorsements of the points of interest and the construction of the standard deviation ellipse can accurately determine the city Active range and distribution of the active circle, from the meso-level use of spatial analysis and mining and other means of urban active and active circle detection and analysis. Compared with the traditional method of active circle measurement and analysis, this method has the characteristics of objectivity, real-time and high accuracy. It is shown that the spatial and temporal constraint data has obvious clustering characteristics by the test results overlaid with a city administrative map. This paper is based on the analysis of the data from the spatial and temporal constraints to find the distribution of the active circle in a city is consistent with the objective facts and more detailed. This result reflects the high correlation between daily life behavior and the distribution of commercial economy in a certain city area. It provides a new method for monitoring the distribution and development trend of urban commercial circle. It also provides a new method for city planning and administration Decision-making provides a more intuitive reference. It is necessary to further study the automatic active point detection method based on spatio-temporal constraint data. At the same time, the active circle analysis can be used to get the active circle dynamic changes, such as the change of the active circle and the change

of the active circle range, real-time monitoring the active circle's rise, and decline.

## References

- [1] Z. X. LIAO, W. C. PENG: *Clustering spatial data with a geographic constraint: Exploring local search*. Knowledge and Information Systems 31 (2012), No. 1, 153–170.
- [2] L. J. YANG, H. L. ZHU, J. P. WU: *Market area analysis in retailing based on GIS*. Remote Sensing technology and application 18 (2003), No. 3, 144–148.
- [3] J. GILES: *Wikipedia Rival Calls in the Experts*. Nature 443 (2006), No. 7111, paper 493.
- [4] J. HOWE: *The rise of crowdsourcing*. Wired magazine 14 (2006), No. 6, 1–4.
- [5] C. HEIPKE: *Crowdsourcing geospatial data*. ISPRS Journal of Photogrammetry and Remote Sensing 65 (2010), No. 6, 550–557.
- [6] B. T. HAWORTH: *Implications of volunteered geographic information for disaster management and GIScience: A more complex world of volunteered geography*. Annals of the American Association of Geographers (2017) 1–15.
- [7] B. J. HECHT, M. STEPHENS: *A tale of cities: Urban biases in volunteered geographic information*. International Conference on Weblogs and Social Media (ICWSM), 1–4 June 2014, Ann Arbor, Michigan, USA, Publisher AAAI press (2014), 197–205.
- [8] P. NEIS, D. ZIELSTRA: *Recent developments and future trends in volunteered geographic information research: The case of opensreetmap*. Future Internet 6 (2014), No. 1, 76–106.
- [9] M. F. GOODCHILD, J. A. GLENNON: *Crowdsourcing geographic information for disaster response: A research frontier*. International Journal of Digital Earth 3 (2010), No. 3, 231–241.
- [10] S. FRITZ, I. MCCALLUM, C. SCHILL, C. PERGER, R. GRILLMAYER, F. ACHARD, F. KRAXNER, M. OBERSTEINER: *Geo-Wiki.org: The use of crowdsourcing to improve global land cover*. Remote Sensing 1, (2009), No. 13, 345–354.
- [11] A. M. LU, C. M. LI, Z. J. LIN, W. Z. SHI: *Spatial continuous surface model of population density*. Acta Geodaetica Et Cartographica Sinica 32 (2003), No. 4, 344–348.
- [12] Z. HALIM, M. M. GUL, N. UL HASSAN, R. BAIG, S. UR REHMAN, F. NAZ: *Malicious users' circle detection in social network based on spatio-temporal co-occurrence*. International Conference on Computer Networks and Information Technology, 11–13 July 2011, Abbottabad, Pakistan, IEEE Conference Publications (2011), 35–39.

Received June 29 7, 2017



# Fuzzy adaptive control for boiler based on nonlinear depth recursion<sup>1</sup>

NI NA ZHOU<sup>2,3</sup>, LI WANG<sup>2</sup>

**Abstract.** Fuzzy adaptive control for boiler (FACB) is a kind of boiler data drive control method, which has advantages of simple calculation, strong robustness and no modeling. At present, the fuzzy adaptive control methods for boiler generally do not take the non-linear control problem of the boiler that may occur into consideration. In view of this paper, an improved algorithm is designed in this paper for the situation that the upper limit of the actuator execution capability. The algorithm uses the nonlinear depth recursive method to solve the problem by introducing the constraint condition to the depth recursion, with the advantages of simple programming and small computation amount. On the basis, the closed-loop stability is analyzed and proved. Finally, the effectiveness of the algorithm is verified by comparing the simulation experiment with the boiler as the control object.

**Key words.** Fuzzy adaptive control for boiler, boiler nonlinear control, robustness, boiler data drive.

## 1. Introduction

Since the 1950s, the model-based control theory has been developed and perfected rapidly. The process of establishing the control system has gone through three stages, followed by the establishment of the model, the analysis of the model and the design of the control law by the model [1]. However, with the controlled object becoming more and more complex, how to model the controlled object effectively has become increasingly difficult. The reason lies in that for a system, if the constructed mathematical model is too complex, it will be difficult to design the control law or hard to implement the control law obtained in the engineering [2–3]. While if it is too simple, it will be difficult to reflect the dynamic characteristics of the actual system, thus the control law designed based on this model will be hard to achieve the desired control effect in the practical application. Secondly, the main methods

---

<sup>1</sup>This work was support by the Shaanxi Provincial Department of Education Project 14JK1049, Baoji University of Arts and Sciences project YK1510.

<sup>2</sup>Baoji University of Arts and Sciences, Baoji, Shaanxi, 721013 China

<sup>3</sup>Corresponding author

for the establishment of the system model include the mechanism modeling and system identification. No matter which method is used, the model established is the approximation to the real system, while the actual system will inevitably have the impact on the unmodeled dynamic of the control system robustness and other uncertain factors [4–5].

At present, there are two main methods to solve the above problem. The first category of methods is based on the boiler data drive control method. Currently there are three typical methods including the virtual reference feedback tuning (VRFT), synchronous disturbance random (SPSA) and fuzzy adaptive control for boiler (FACB). VRFT was proposed by Guardabassi et al. in 2000, which is characterized by the design of control law based on off-line data. The method is based on the model reference adaptive, and the controller parameters are obtained directly by the parameter identification, so that the dynamic characteristics of the closed-loop system are approximated to the reference model [6]. As the dynamic information of the actual system cannot be fully obtained by one excitation, the controller obtained by using the method design is difficult to ensure the stability of the closed-loop system. SPSA was proposed by Spall in 1993, which is characterized by iterative identification of system parameters, the control effect is susceptible to changes in system structure or parameter changes. It is difficult to guarantee the stability of closed-loop system [7]. FACB method was put forward by Hou Chongsheng in 1994 as a fuzzy adaptive control method for boiler [8], which is characterized by the controller design does not require the system model information, but is based on input and output data to directly calculate and obtain the control value. Literature [5, 9–10] proved the stability of closed-loop system with compact format and partial format dynamic linearization method. The second category of methods is the feature modeling method proposed by Wu Hongxin, et al. The feature modeling method is widely used in the spacecraft and industrial control, which is a kind of modeling method which takes the dynamic characteristics of the object and the performance requirements of the control into account [2]. This method is similar to the full-format dynamic linearization model proposed in literature [5]. The difference lies in that the feature modeling method emphasizes that the feature model is equivalent to the input and output of the original model when the sampling period is sufficiently small [2], while the dynamic linearization method based on the full format considers the situation where time-varying linear system and the original system is equivalent [5, 9–10].

Hou Zhongsheng proposed the fuzzy adaptive control for boiler. Due to its rigorous theory and small amount of calculation, it has broad application prospect in the engineering. After years of development, Hou Zhongsheng designed the fuzzy adaptive control for boiler based on the compact format, the partial format and the full format respectively for the single-input single-output system and multi-input multi-output system.

Based on the PJM (Pseudo Jacobian matrix) identification technique in literature [5], a fuzzy adaptive control method based on the compactness scheme for the boiler is put forward in this paper. And the closed-loop stability of the control method is proved. The simulation experiment of boiler Wood/Berry further shows that the

method has stronger tracking ability and is insensitive to the initial value. In this paper, the rate saturation and position saturation are optimized for the actuator at the same time.

## 2. Problem description

This section briefly describes the related concepts and calculation ideas of the fuzzy adaptive control method for boiler proposed in literature [5]. On this basis, the defects of the existing methods are analyzed, and the problems to be solved in this paper are put forward.

The following multiple input multiple output discrete system is considered:

$$y(k+1) = f(y(k), \dots, y(k-n_y), u(k), \dots, u(k-n_u)), \quad (1)$$

in which,  $u(k) \in R^m$ ,  $y(k) \in R^n$ , which are the input and output vector at the system time  $k$ . Symbols  $n_y$  and  $n_u$  are unknown integers and  $f(\cdot)$  is an unknown non-linear function. Assuming that  $f(\cdot)$  is the partial derivative continuity related to  $u(k)$ , and the system (1) satisfies the generalized Lipschitz hypothesis, theorem 1 [5] can be obtained.

Theorem 1. For the nonlinear system (1) that satisfies the generalized Lipschitz hypothesis as the partial derivative continuity related to  $u(k)$ , when  $\|\Delta u(k)\| \neq 0$ , there must be a time-varying parameter  $\Phi_c(k) \in R^{n \times m}$  known as the PJM (Pseudo Jacobian matrix), so that the system (1) can be transformed into the compact form dynamic linearization (CFDL) as follows

$$\Delta y(k+1) = \Phi_c(k) \Delta u(k), \quad (2)$$

where

$$\Phi_c(k) = \begin{bmatrix} \phi_{11}(k) & \phi_{12}(k) & \cdots & \phi_{1m}(k) \\ \phi_{21}(k) & \phi_{22}(k) & \cdots & \phi_{2m}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{n1}(k) & \phi_{n2}(k) & \cdots & \phi_{nm}(k) \end{bmatrix} \in R^{n \times m}. \quad (3)$$

and for any time  $k$ ,  $\|\Phi_c(k)\|$  is bounded.

The fuzzy adaptive control of boiler based on the compact format dynamic linearization is to use the method of parameter identification to dynamically calculate the value of PJM time-varying parameter and control it on this basis. The detailed calculation steps can be found in literature [9–10]. In the existing method, only literature [10] considered the problem of the incomplete non-linear runaway of boiler. For the actual physical system, the actuator execution capability is limited, which is reflected in the control of the magnitude and rate of change. The execution capability of the actuator can be completely expressed as the following

$$\begin{cases} \Delta u_{\min}(k) \leq \Delta u(k) \leq \Delta u_{\max}(k), \\ u_L \leq u(k) \leq u_v, \end{cases} \quad (4)$$

in which,  $\Delta u_{\min}$  and  $\Delta u_{\max}$  stand for the minimum and maximum values of the control variable rate, respectively.  $u_L$  and  $u_v$  stands for the minimum and maximum values of the control amplitude, respectively.

### 3. Nonlinear runaway optimization of the fuzzy adaptive control for boiler

In this section, an optimized fuzzy adaptive control for boiler based on the compact format boiler is put forward, and a control algorithm by comprehensively analyzing the execution ability of the actuator is provided.

In order to strictly analyze the closed loop stability of the improved algorithm, the following assumptions are made:

*Hypothesis 1:* There is a sufficiently large  $\lambda$  that makes  $\Phi E^{-1} (F + M^T x)$  positive definite.

If the assumption 1 is not satisfied, it shows that the improved algorithm cannot guarantee the closed-loop stability of the system in the event of the non-linear boiler runaway, and it requires more complicated control algorithm so as to carry out effective control.

Theorem 2 can be proved based on Hypothesis 1.

*Theorem 2:* For the nonlinear system (1), when the amplitude  $\Delta u(k)$  is bounded, the iterative algorithm proposed by an identification scheme below has the following properties when the hypothesis 1 is met: When  $y^*(k+1) = y^* = \text{const}$ , there is a positive number  $\lambda_{\min} > 0$ , which enables the following when  $\lambda \geq \lambda_{\min}$ :

1) The system tracking error sequence is bounded, that is,  $\|y(k+1) - y^*\|$  is bounded.

2) The closed-loop system is BIBO (Bounded-input bounded-output) stable, that is, the output sequence  $\{y(k)\}$  and the input sequence  $\{u(k)\}$  are bounded.

*Proof.*

Prove that  $\|\tilde{y}(k)\| = \|y(k) - \hat{y}(k)\|$  is bounded. According to the theorem in literature [5],  $\|\Phi_c(k) - \hat{\Phi}_c(k)\|$  is bounded, and  $\Delta u(k)$  is bounded, assuming that

$$\|\Phi_c(k) - \hat{\Phi}_c(k)\| \|\Delta u(k)\| \leq b. \tag{5}$$

Then the following can be obtained

$$\begin{aligned} \|\tilde{y}(k+1)\| &= \|y(k+1) - \hat{y}(k+1)\| = (1-K) \left\| y(k) + \tilde{\Phi}_c(k) \Delta u(k) \right\| \leq \\ &(1+K) \left( \|\tilde{y}(k)\| + \left\| \tilde{\Phi}_c(k) \Delta u(k) \right\| \right) \leq (1+K) \|\tilde{y}(k)\| + (1+K)b \leq \dots \leq \\ &(1+K)^k \|\tilde{y}(1)\| + \frac{(1+K)b(1-(1+K)^k)}{K} \end{aligned} \tag{6}$$

Hence  $\|\tilde{y}(k)\|$  is bounded, assuming it to be  $c$ , that is,  $\|\tilde{y}(k)\| \leq c$ .

Prove that there is  $\lambda$  that enables

$$I - \Phi_c(k) S(k) \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k), \tag{7}$$

the absolute values of the eigenvalue are all less than 1. In this case,  $S(k)$  is a diagonal matrix,  $S_i(k)$  stands for the  $k$ th diagonal element in the matrix  $S(k)$ . In addition, denote  $\delta(k)$  as an  $m$ -dimensional vector, and  $\varepsilon_0$  is a positive number greater than 0. And  $S(k)$  and  $\delta(k)$  satisfy the following.

for  $i = 1 : m$   
if  $(E^{-1}F)_i \leq \varepsilon_0$  then

$$S_i(k) \leftarrow \frac{(E^{-1}(F + M^T x))_i}{\varepsilon_0}$$

$$\delta_i(k) \leftarrow \frac{(x)_i}{S_i} - (E^{-1}F)_i$$

else

$$S_i(k) \leftarrow \frac{(E^{-1}(F + M^T x))_i}{(E^{-1}F)_i}$$

$$\delta_i(k) \leftarrow 0$$

It is easy to see that

$$x = S(k) E^{-1}F + S(k) \delta(k) \quad (8)$$

Use  $\|\cdot\|_2$  to represent the spectral norm of the matrix,  $\rho(\cdot)$  to represent the spectrum of the matrix. As  $\Phi_c(k)$ ,  $S(k)$ ,  $\hat{\Phi}_c^T(k)$  and  $\delta(k)$  are all bounded, assuming that

$$\begin{aligned} \|\Phi_c(k)\|_2 \|S(k)\|_2 \left\| \hat{\Phi}_c^T(k) \right\|_2 &\leq e_1 \|S(k)\|_2 \left\| \hat{\Phi}_c^T(k) \right\|_2 \leq \\ &\leq e_2 \|\Phi_c(k)\|_2 \|S(k) \delta(k)\|_2 \leq e_3. \end{aligned} \quad (9)$$

Take  $e_0 = \max\{e_1, e_2\}$ ,  $\lambda_{\min} = e_0 + \rho(\hat{\Phi}_c(k) \hat{\Phi}_c^T(k))$ , as  $(\hat{\Phi}_c \hat{\Phi}_c^T(k) + \lambda I)^{-1}$  is a symmetric matrix, and when  $\lambda > \lambda_{\min}$ , the following is established

$$\left\| (\hat{\Phi}_c \hat{\Phi}_c^T(k) + \lambda I)^{-1} \right\|_2 = \rho \left( (\hat{\Phi}_c \hat{\Phi}_c^T(k) + \lambda I)^{-1} \right) < \frac{1}{e_0}. \quad (10)$$

Then it is further established that

$$\begin{aligned} \rho \left( \Phi_c(k) S(k) (\hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I)^{-1} \hat{\Phi}_c^T(k) \right) &\leq \\ \|\Phi_c(k)\|_2 \|S(k)\|_2 \left\| (\hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I)^{-1} \right\|_2 \times \left\| \hat{\Phi}_c^T(k) \right\|_2 &< 1. \end{aligned} \quad (11)$$

It can be seen from the Hypothesis 1 that, when  $\lambda > \lambda_{\min}$ , all the absolute values of the eigenvalue of Equation (11) are less than 1.

Prove that the tracking error is bounded. As can be seen from Step 2

$$\rho \left( I - \Phi_c(k) S(k) \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k) \right) < 1. \quad (12)$$

Then there is sufficiently small  $\epsilon$  and norm  $\|\cdot\|$  that enables the following

$$\begin{aligned} & \left\| I - \Phi_c(k) S(k) \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k) \right\| \leq \epsilon + \\ & \rho \left( I - \Phi_c(k) S(k) \times \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k) \right) < 1 \end{aligned} \quad (13)$$

For any  $k$ , take the following

$$\begin{aligned} & \left\| I - \Phi_c(k) S(k) \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k) \right\| < d_1 < 1 \\ & \left\| \Phi_c(k) S(k) \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k) \right\| < d_2 \end{aligned} \quad (14)$$

Then

$$\begin{aligned} \|e(k+1)\| &= \|y^*(k+1) - y(k+1)\| = \|y^*(k+1) - y(k) - \Phi_c(k) S(k) \times \\ & \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k) (y^*(k) - \hat{y}(k)) - \Phi_c(k) S(k) \delta(k)\| \leq \\ & \left\| I - \Phi_c(k) S(k) \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k) \right\| \times \|y^*(k) - y(k)\| + \\ & \left\| \Phi_c(k) S(k) \times \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k) \right\| \|\tilde{y}(k)\| + \|\Phi_c(k) S(k) \delta(k)\| \\ & \leq d_1 \|e(k)\| + d_2 c + e_3 \leq \dots \leq d_1^k \|e(1)\| + \frac{(d_2 c + e_3)(1 - d_1^k)}{1 - d_1}. \end{aligned}$$

Then  $\|e(k)\|$  is bounded, assuming its boundary is  $f_0$ , that is,  $\|e(k)\| \leq f_0$ .

As  $y^*(k)$  is bounded, it can be known that  $y(k)$  is bounded. Also because in the process of solving  $x$ , the control value position saturation limit is added,  $\|u\|$  can meet the constraints, hence it is bounded. Then it can be further known that  $u(k)$  is bounded.

*Inference 1.* If the filter is not used, that is,  $\hat{y}(k) = y(k)$ . And when  $\delta(k) = 0$  is established, the conclusion in Theorem 2 can be strengthened as the following:

- 1) system tracking error sequence convergence;
- 2) The closed-loop system is BIBO stable, that is, the output sequence and the input sequence are bounded.

*Proof.*

$$\begin{aligned} \|e(k+1)\| &\leq \|y^*(k+1) - y(k) - \Phi_c(k) S(k) \times \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \\ &\hat{\Phi}_c^T(k) (y^*(k) - \hat{y}(k))\| \leq \left\| I - \Phi_c(k) S(k) \left( \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \lambda I \right)^{-1} \hat{\Phi}_c^T(k) \right\| \\ &\times \|y^*(k) - y(k)\| \leq d_1 \|e(k)\| \dots \leq d_1^k \|e(1)\|. \end{aligned} \quad (15)$$

Therefore,  $\lim_{k \rightarrow \infty} \|y(k+1) - y^*\| = 0$ . Hence the conclusion 1) is proved. The proving of the conclusion 2) is similar to the proving of the conclusion 2 in Theorem 2.

The physical meaning of  $\delta(k) \equiv 0$  in Inference 1 is that under the premise of the absence of the constraints of equation (7), the result obtained by the calculation only according to Equation (7) is not zero. In the actual system, due to the existence of noise and numerical calculation error, the resulting value from solving equation (7) will rarely be zero, therefore, the hypothesis in the actual system is of certain significance.

#### 4. Simulation experiment

Boilers are widely used in the chemical industry. However, the delay occurred inside the boiler has caused great difficulties in the design of the control law. The fuzzy adaptive control for boiler can be used to conduct effective tracking control. In this simulation, Wood/ Berry boiler is used, as shown in Fig.1, in which,  $u_1$  stands for the reflux rate (IB/min),  $u_2$  stands for the steam flow (IB/min),  $y_1$  stands for the upper part component (mol% methanol), and  $y_2$  stands for the bottom part component (mol% methanol). The following discrete system is used as the Wood/Berry boiler.

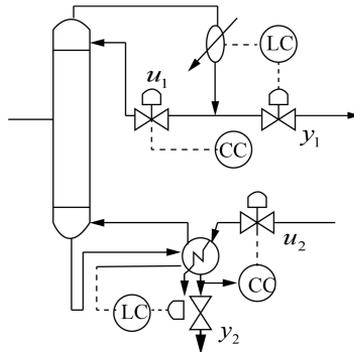


Fig. 1. Boiler control system

$$\begin{cases} y_1(z) = \frac{0.7665}{z-0.9419}u_1(z) + \frac{0.9z^{-2}}{z-0.9535}u_2(z), \\ y_2(z) = \frac{0.6055z^{-6}}{z-0.9124}u_1(z) + \frac{1.3472z^{-2}}{z-0.90311}u_2(z). \end{cases}$$

The expected output signal is as the following

$$\begin{aligned} y_1^*(k) &= \begin{cases} 40, k \leq 1000 \\ 90, k > 1000 \end{cases} \\ y_2^*(k) &= \begin{cases} 40, k \leq 1000 \\ 85, k > 1000 \end{cases} \end{aligned} \quad (16)$$

In order to compare the control effect of the original method and the improved method upon the occurrence of non-linear runaway of the boiler, the calculation method for  $\hat{\Phi}_c(k)$  before the improvement is described as the following:

$$\hat{y}_i(k+1) = \hat{y}_i(k) + \Delta u^T(k) \hat{\phi}_i^T(k) + k_i \tilde{y}_i(k), \quad (17)$$

$$\hat{\phi}_i^T(k+1) = \hat{\phi}_i^T(k) + 2\Delta u(k) \left( \|\Delta u(k)\|^2 + \mu_i \right)^{-1} \times (\tilde{y}_i(k+1) - F_i \tilde{y}_i(k)), \quad (18)$$

in which  $\hat{y}_i(k)$  is the estimate value of the  $i$ th output component and  $\tilde{y}_i(k) = y_i(k) - \hat{y}_i(k)$  is the corresponding estimate error. Quantity  $F_i = 1 - k_i$ , and  $k_i$  is the corresponding element on the diagonal of the matrix.  $\hat{\phi}_i^T(k)$  is the  $i$ th row vector of the matrix  $\hat{\Phi}_c(k)$ . The calculation method for  $u(k)$  before the improvement is as the following:

$$u(k) = u(k-1) + \hat{\Phi}_c^T(k) \left[ \hat{\Phi}_c(k) \hat{\Phi}_c^T(k) + \alpha \right]^{-1} \times \quad (19)$$

$$[y^*(k+1) - \hat{y}(k) - K\tilde{y}(k)], \|\Delta u(k)\| \leq \delta,$$

$$u(k) = u(k-1) + \delta \cdot \text{sgn}(\Delta u(k)), \|\Delta u(k)\| > \delta. \quad (20)$$

In the simulation process of this paper,  $\alpha = \text{diag}\{0.003, 0.0015\}$  and  $\delta = 0.02$ . The following parameters are shared with the improved algorithm before and after the improvement: the sampling cycle  $T_s = 1s$ ,  $K = \text{diag}\{0.9, 0.9\}$ ,  $\mu_1 = \mu_2 = 9$ , and PJM is the initial value of the PJM parameter.

$$\hat{\Phi}_c(0) = \begin{bmatrix} 910 & 750 \\ 450 & 520 \end{bmatrix}.$$

At the same time limiting the execution capacity of the controlled system actuator

as the following

$$\begin{cases} 0 \leq \mu_1 \leq 1, & -0.02 \leq \Delta\mu_1 \leq 0.02 \\ 0 \leq \mu_2 \leq 4, & -0.02 \leq \Delta\mu_2 \leq 0.02 \end{cases} \quad (21)$$

On this basis, the algorithm is simulated before and after the improvement and Fig. 2 can be obtained.

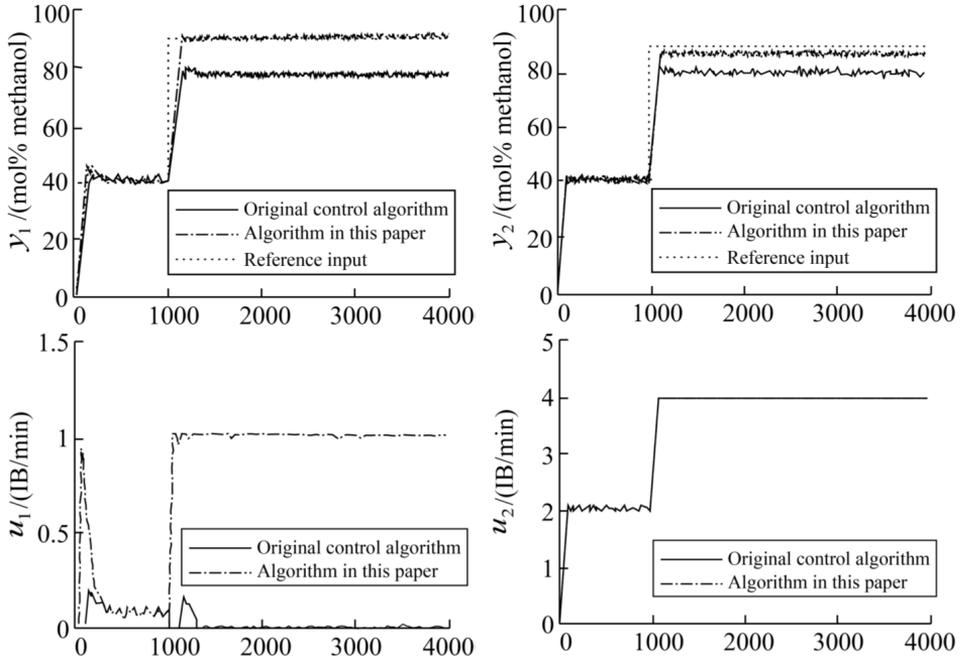


Fig. 2. Control performance comparison chart

In order to compare the sensitivity of the control algorithm to the initial parameters before and after the optimization of the initial parameters, assume the original parameter

$$\hat{\Phi}_c(0) = \begin{bmatrix} 1000 & 1000 \\ 1000 & 1000 \end{bmatrix}.$$

Then, Fig. 3 can be obtained after performing simulation. In the absence of limited execution capacity, the boiler fuzzy adaptive control can better track the signal. But it can be found after comparing with Fig. 3 that, the fuzzy adaptive control of the boiler before the improvement shows significant static difference upon the occurrence of the non-linear runaway of the boiler. This is because the controller in the calculation of control output does not take into account the implementation of the actuator capacity, indirectly causing the situation that the system cannot correct the errors of the PJM parameters, and finally leading to the violent jitter in the control algorithm, and basically losing the capability to track the reference input. Under the improved boiler fuzzy adaptive control algorithm, it fully considers the execution capability of the actuator, which can effectively track the reference input,

showing the effectiveness of the algorithm.

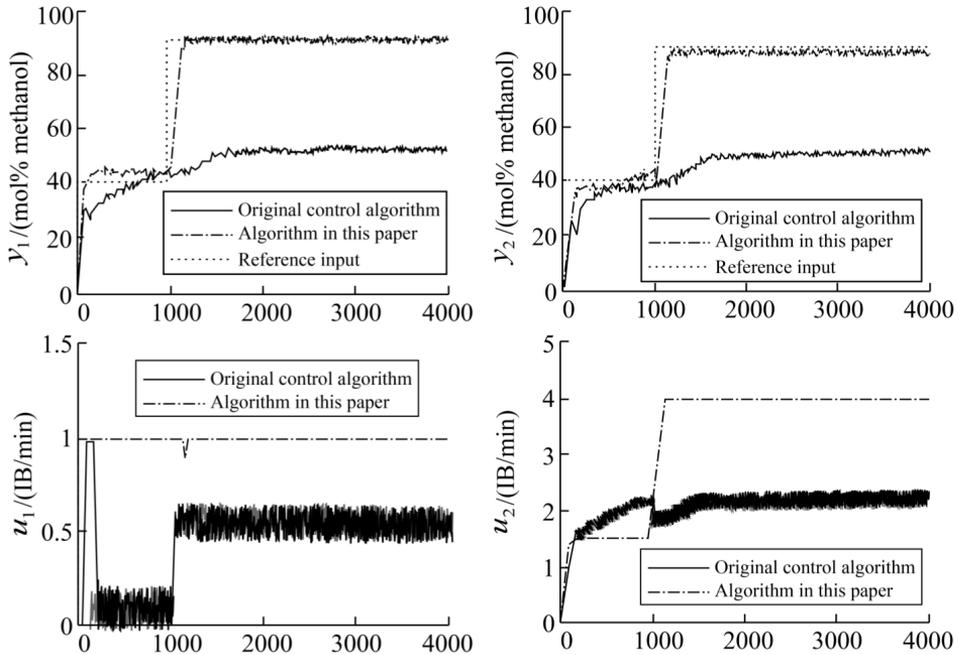


Fig. 3. Control performance comparison chart after the change of the initial parameters

## 5. Conclusion

In this paper, a fuzzy adaptive control algorithm for boiler is proposed to solve the failure to deal with the non-linear runaway problem of the boiler. And the closed loop stability of the algorithm is proved rigorously. The algorithm has the advantages of simple to implement and small calculation amount. The control effects of the algorithm combined with the Wood/Berry model before and after the improvement are compared. The simulation results show that, the improved algorithm has the advantages of strong tracking capability and weak dependence on the initial parameters, which can effectively deal with the non-linear runaway problem of the boiler.

## References

- [1] B. Z. GUO, J. Q. HAN, F. B. XI: *Linear tracking-differentiator and application to on-line estimation of the frequency of a sinusoidal signal with random noise perturbation*. Journal of Systems Science 33 (2002), No. 5, 351–358.
- [2] H. WU, J. HU, Y. XIE: *Characteristic model-based all-coefficient adaptive control*

- method and its applications*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 37 (2007), No. 2, 213–221.
- [3] M. LEWIN, P. SÉVERINE: *A numerical perspective on Hartree-Fock-Bogoliubov theory*. Mathematical Modelling and Numerical Analysis (ESAIM: M2AN) 48 (2014), No. 1, 53–86.
  - [4] Z. HOU: *On model-free adaptive control: The state of the art and perspective*. Control Theory and Applications 23 (2006), No. 4, 586–5922.
  - [5] Z. HOU, S. JIN: *Model free adaptive control. Theory and applications*. CRC Press, Taylor and Francis Group, Boca Raton (2016).
  - [6] G. O. GUARDABASSI, S. M. SAVARESI: *Virtual reference direct design method: An off-line approach to data-based control system design*. IEEE Transactions on Automatic Control 45 (2000), No. 5, 954–959.
  - [7] J. C. SPALL: *Multivariate stochastic approximation using a simultaneous perturbation gradient approximation*. IEEE Transactions on Automatic Control 37 (1992), No. 3, 322–341.
  - [8] P. SADEGH, J. C. SPALL: *Optimal random perturbations for stochastic approximation using a simultaneous perturbation gradient approximation*. IEEE Transactions on Automatic Control 43 (1998), No. 10, 1480–1484.
  - [9] Z. HOU, S. JIN: *A novel data-driven control approach for a class of discrete-time nonlinear systems*. IEEE Transactions on Control Systems Technology 19 (2011), No. 6, 1549–1558.
  - [10] R. CHI, Z. HOU: *Optimal higher order learning adaptive control approach for a class of SISO nonlinear systems*. Journal of Control Theory and Applications 3, (2005), No. 3, 247–251.

Received June 29, 2017



# Software test data generation algorithm based on multi-dimensional space-time granularity

WU DAFEI<sup>1</sup>

**Abstract.** Based on the multi-dimensional space-time granularity, a kind of software test data generation algorithm was proposed. The algorithm first uses the space-time function multi-dimensional approach and granularity method, finds all feasible paths in the program. And it generates proper initial software test data set for the partial feasible paths automatically; when the correct software testing data can not be obtained by using the space-time function multi-dimensional approach and granularity method, it can depend on the principal of making software testing data set smallest and multi-dimensional space-time granularity thought. The software testing data can be supplemented according to the predicate and sub path that never covered by initial test data set. The new algorithm has a combination with predicate slice and DUC expression, so it is able to judge the feasibility of the sub path from the source. Then it can effectively to reduce the influence of unfeasible path on the algorithm performance. The algorithm analysis and experimental results show that the algorithm can effectively reduce the software test data bulk, and improve the test performance.

**Key words.** Software test, software test data automatic generation, multi-dimensional space-time granularity, multi-dimensional approach; space-time function granularity.

## 1. Introduction

At present, dynamic software test has become a research focus of software testing. In the dynamic software testing process, the generation of test cases is the key and difficulty of the task. According to statistics, about 40% of the test costs was spent on the design test cases [1]

Relying on a variety of test models and standards, the generation methods of software test data like dependent on the syntax, predicate slices, program specifications, symbol execution and program execution were proposed [2–6]. Path coverage is a typical test standard. The goal is to require that all paths of the program be tested at least once at the end of the test. In literature [7, 8], a software test data

---

<sup>1</sup>Hunan University of Science and Engineering, Yongzhou, Hunan, 425199, China

generation algorithm based on path coverage test is given by symbolic execution, as the implementation of the symbol on the form of the array and the pointer is difficult to deal with; In literature [9, 10], program execution and space-time function granularity are used to solve the drawbacks of the array execution and pointers. However, since each test only considers one predicate and one input variable, a large number of iterations are required to automatically generate a new input that satisfies the condition even if the branch condition in the path is a linear function of the input. A new algorithm for a given path-dependent iteration is proposed in Literature [11]. The algorithm takes into account multiple predicates and multiple input variables at same time. A set of initial inputs satisfying the given path is automatically generated by iteration using a randomly selected set of initial inputs. The algorithm only uses the linear function to iterate, so it is only effective for the linear function, and less effective for the nonlinear function.

In this paper, a software test data generation algorithm based on the idea of multi-dimensional space-time granularity was proposed. The new algorithm first uses the multidimensional approximation and granularity method of the space-time function to find out the feasible path of the program and automatically generate the appropriate initial software test data set for some feasible paths. When the space-time function multi-dimensional approach and granularity methods can not get the software test data correctly, it can depend on the principal of making software testing data set smallest and multi-dimensional space-time granularity thought for the predicate and sub path that never covered by initial test data set. The software testing data can be supplemented according to the thought. Since the new algorithm uses the DUC expression [6], it is possible to determine whether the sub path is feasible from the source, so as to effectively reduce the influence of unfeasible paths on the algorithm. The new algorithm combines the advantages of space-time function multi-dimensional approach and granularity methods. And it relies on multi-dimensional space-time granularity to supplement the software test data, thus reducing the number of software test data and improving the test performance.

*Definition 1.* Path. A program module  $M$  can be regarded as a directed graph  $G = (V, E, s, e)$ , where  $V$  is the set of nodes in  $M$ ,  $E$  is the set of edges in  $M$ ,  $s$  is the unique source node of  $M$ , indicating the beginning of the program module  $M$ ;  $e$  represents the only sink nodes for  $M$ , meaning the end of program module  $M$ . A node  $n$  means a piece of declaration statements or a conditional expression. From node  $n_i$  to node  $n_j$ , a possible control transfer is regard as an edge  $(n_i, n_j) \in G$ . A subpath  $P = \{n_1, n_2, \dots, n_{k+1}\} \in G$  is a sequence of node  $n_i$ , and  $(n_i, n_{i+1}) \in E$ ,  $\forall i \in [1, k]$ ; if  $n_1 = s, n_{k+1} = e$ , then  $P$  is called a piece of path in program module  $M$ .

*Definition 2.* Input variable. If a variable  $i_k$  appears in an input statement program of program module  $M$  or an input parameter of  $M$ , then  $i_k$  is called the input variable of  $i_k$ . The definition domain  $D_k$  of input variable  $i_k$  is a set for all the possible values of  $i_k$ . An input vector  $I = \{i_1, i_2, \dots, i_m\} \in (D_1 \times D_2 \times \dots \times D_m)$  is called a program input of program module  $M$ , which is called an input of  $M$  for short. Herein,  $m$  is the number of the input variables in  $M$ .

*Definition 3.* Predicate. Simple predicate refers to the predicate only containing

single relational operator  $op$ ; obviously, the simple predicate can be expressed as  $E_1 op E_2$ . Herein,  $op \in \{<, \leq, >, \geq, =, \neq\}$ ,  $E_1, E_2$  is the algebraic expression. Compound predicate means the predicate of two or more than two simple predicates connected by Boolean connectives AND or OR together. If there is a boolean variable in the predicate, then TRUE can be expressed by 0 or a positive number, and FALSE can be expressed by a negative number.

*Definition 4.* An unfeasible path. All predicate explanations [8] on a subpath compose its path condition, which defines a path definition field. It ensures the program to consist of all inputs that execute along the subpath. If the path definition field of a sub-path is empty ( $\phi$ ), the sub-path is called an unfeasible sub-path.

The following *Lemma 1* is clearly defined by the *Definition 4*: from *Definition 4*, it is clear that the following *Lemma 1* comes into existence.

*Lemma 1.* If a sub-path is an unfeasible sub-path, then all paths containing the sub-path are unfeasible paths.

*Definition 5.* Space-time function. A simple predicate  $pr : E_1 op E_2$  can be converted into the following form:  $F rel 0$ ; herein:  $rel \in \{<, \leq, =\}$ ;  $F$  is a direct or indirect function of the input variable, and called the corresponding space-time function of the predicate  $pr$ .

*Definition 6.* Multi-dimensional approach. For arbitrary space-time function  $F_1$  and a set of inputs  $I_0$ , it is called the multi-dimensional approach of linear function  $f_1 = \sum_{i=1}^m a_i^1 x_i + b_1$  for  $F_1$ . Herein,  $x_i$  is the input variable;  $a_i^1$  is the constant coefficient;  $b_1$  is the constant;  $a_i^1$  and  $b_1$  can be obtained from the following equations:

$$\left\{ \begin{array}{l} f_1(I_0) = F_1(I_0), \\ f_1(I_0 + (\Delta_{x_1}, 0, \dots, 0)) = \\ F_1(I_0 + (\Delta_{x_1}, 0, \dots, 0)), \\ \vdots \\ f_1(I_0 + (0, \dots, 0, \Delta_{x_{m-1}}, 0)) = \\ F_1(I_0 + (0, \dots, 0, \Delta_{x_{m-1}}, 0)), \\ f_1(I_0 + (0, \dots, 0, \Delta_{x_m})) = \\ F_1(I_0 + (0, \dots, 0, \Delta_{x_m})). \end{array} \right.$$

*Theorem 1.* For a given input  $I_0$ , supposing the corresponding multi-dimensional approach  $f_1$  of the space-time function  $F_1$  for the predicate  $p_1$  is  $f_1$ , and the operator is of  $F_1$  is  $rel_1$ , then: 1) if  $f_1(I_0) rel_1 0 \Rightarrow I_0$  ensures the program to execute along the sub-path  $p_1$ ; 2) if  $f_1(I_0) rel_1 0 \Rightarrow I_0$  ensures the program to execute along the sub path  $p_1$ .

*Proof.* It can be seen from *Definition 6*:  $f_1(I_0) = F_1(I_0)$ , then  $f_1(I_0) rel_1 0$  is set

up  $\Rightarrow F_1(I_0)rel_10$  is set up  $\Rightarrow I_0$  ensures the program to execute along the sub path  $p_1$ . Proving by the same method,  $f_1(I_0)\neg rel_10$  is set up  $\Rightarrow I_0$  ensures the program to execute along the sub-path  $p_1$ , Q.E.D.

## 2. Software test data generation algorithm based on multi-dimensional space-time granularity

### 2.1. Generation algorithm of initial software test data set

Based on the definitions, theorems and lemmas in section 1, the initial software test data set generation algorithm is described as follows:

Step 1. By using the predicate slicing algorithm [6], the DUC expression  $r_j.d_j : U_j : c_j$  and operator  $rel_j$  corresponding in  $M$  of all the  $n$  predicates  $p_j$  ( $j \in [1, n]$ ) were obtained; and all the  $n$  discrete input variables  $x_1, x_2, \dots, x_m$  in  $M$  are obtained.

Step 2. Randomly generate a program input  $I_0 = (x_1^0, x_2^0, \dots, x_m^0)$ , order  $I^0 = I_0$ .

Step 3. For  $p^k$  ( $k \in [1, n]$ ) and input  $I^{t_1 t_2 \dots t_{k-1}}$  (herein,  $t_i \in \{0, 1\}$ ,  $i \in [1, k-1]$ ), and when  $k = 1$ , order  $I^{t_1 t_2 \dots t_{k-1}} = I^0$ .

Step 3.1. If  $k \geq 2$  and  $\sim p_1^{t_1}, \sim p_2^{t_1 t_2}, \dots, \sim p_{k-1}^{t_1 t_2 \dots t_{k-1}} \notin c_k$ , then skip to Step 3.2, otherwise skip to Step 3.5.

Step 3.2. By using *Definition 6*, according to the input  $I^{t_1 t_2 \dots t_{k-1}}$ , the multi-dimensional approach  $f_k$  of  $p^k$  corresponded space-time function  $F_k$  can be solved. Substitute  $I^{t_1 t_2 \dots t_{k-1}}$  into  $f_1^{t_1}, f_2^{t_1 t_2}, \dots, f_{k-1}^{t_1 t_2 \dots t_{k-1}}, f_k$  (herein  $f_i^{t_1 t_2 \dots t_{i-1}} = f_i, i \in [1, k-1]$ ) and obtain a group of values  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ .

Step 3.3. If  $\theta_k rel_k 0$ , then  $I^{t_1 t_2 \dots t_{k-1}}$  ensures the program execute along the sub-path  $p_1^{t_1}, p_2^{t_1 t_2}, \dots, p_{k-1}^{t_1 t_2 \dots t_{k-1}}, p_k$ . Order  $I^{t_1 t_2 \dots t_{k-1}} = I^{t_1 t_2 \dots t_{k-1}}, f_k^{t_1 t_2 \dots t_{k-1}} = f_k, p_k^{t_1 t_2 \dots t_{k-1}} = p_k$ , and record the path and its input as well as the multi-dimensional approach set of space-time function; then, solve the vector  $I$  relying on *Theorem 2*, order  $I_1 = I^{t_1 t_2 \dots t_{k-1}} + I$ .

Step 3.3.1. If  $I_1$  ensures the program execute along the sub-path  $p_1^{t_1}, p_2^{t_1 t_2}, \dots, p_{k-1}^{t_1 t_2 \dots t_{k-1}}, \sim p_k$ , then order  $I^{t_1 t_2 \dots t_{k-1}} = I_1, f_k^{t_1 t_2 \dots t_{k-1}} = f_k, p_k^{t_1 t_2 \dots t_{k-1}} = \sim p_k$ , and record the path and its input as well as the multi-dimensional approach set of space-time function; and then skip to Step 3.5.

Step 3.3.2. Otherwise, the solution of the program executing along the above subpath can be found according to the space-time function granularity thought. If it is able to find the input  $I'_1$  meeting the conditions in a given maximum number (if cannot, then order  $I'_1 = \phi$ ), then use the *Definition 6* and obtain the new multi-dimensional approach  $f'_1, f'_2, \dots, f'_{k-1}, f'_k$  of  $F_1, F_2, \dots, F_k$ , order  $I^{t_1 t_2 \dots t_{k-1}} = I'_1, f_1^{t_1} = f'_1, f_2^{t_1 t_2} = f'_2, f_{k-1}^{t_1 t_2 \dots t_{k-1}} = f'_{k-1}, f_k^{t_1 t_2 \dots t_{k-1}} = f'_k, p_k^{t_1 t_2 \dots t_{k-1}} = \sim p_k$ , and record the path and its input as well as the multi-dimensional approach set of space-

time function; and then skip to Step 3.5.

Step 3.4. If  $\theta_k \text{ rel}_k 0$  is not set up, then  $I^{t_1 t_2 \dots t_{k-1}}$  ensures the program execute along the sub-path  $p_1^{t_1}, p_2^{t_1 t_2}, \dots, p_{k-1}^{t_1 t_2 \dots t_{k-1}}, \sim p_k$  order  $I^{t_1 t_2 \dots t_{k-1}} = I^{t_1 t_2 \dots t_{k-1}}, f_k^{t_1 t_2 \dots t_{k-1}} = f'_{k-1}, f_k^{t_1 t_2 \dots t_{k-1}} = f'_k, p_k^{t_1 t_2 \dots t_{k-1}} = \sim p_k$ , and record the path and its input as well as the multi-dimensional approach set of space-time function; then, solve the solution vector  $I$  according to Theorem 2, order the input  $I_1 = I^{t_1 t_2 \dots t_{k-1}} + I$ . Finally, record the path and its input as well as the multi-dimensional approach set of space-time function; and then skip to Step 3.5.

Step 3.5. For another set of  $t_1 t_2 \dots t_{k-1}$ , repeat the Step 3.1 to 3.4, until all the groups in  $t_1 t_2 \dots t_{k-1}$  are tested, and then skip to Step 4.

Step 4. Repeat the Step 3 for  $k = k + 1$ , till  $k = n$ , and then skip to Step 5.

Step 5. End of the algorithm: for record of the obtained path and its input as well as the multi-dimensional approach set of space-time function, if the path and input are not empty, then the path is a feasible path, and the corresponding input is a software test data meeting the path.

## 2.2. Supplement algorithm of software test data

Considering the test overhead, it is impractical to construct the appropriate software test data for all paths in the program module  $M$ . Therefore, in this section, we propose a new software test data addition algorithm, which is based on the idea of multi-dimension. Before the description of a specific algorithm is given, the relevant definitions are given as follows:

*Definition 7.* Sub-path pair. For software test data  $t_i$ , assume that the corresponding test path is  $P_i = b_{i1} b_{i2} \dots b_{it}$ , herein,

$$b_{i1} b_{i2} \dots b_{it} \in \{p_1, \sim p_1, p_2, \sim p_2, \dots, p_k, \sim p_k\}, p_1, p_2, \dots, p_k$$

is all  $k$  predicates in program module  $M$ , then call  $(b_{i1}, b_{i2})$  is the subpath pair covered by software test data  $t_i$ , herein  $j \in [1, t - 1]$ . The subpath pair set covered by software test data  $t_i$  is recorded as  $\Psi_i = b_{i1}, b_{i2} \{(b_{i1}, b_{i2}), (b_{i2}, b_{i3}), \dots, (b_{it-1}, b_{it})\}$ . According to the software test data generation algorithm in Section 2.1, an initial software test data set can be obtained. Multi-dimensional space-time granularity refers to that based on the initial software test data set, first supplement the software test data for uncovered predicate. And then selectively supplement the redundancy software test data relying on the idea of subpath pair covering, in order to achieve better test coverage.

According to the above definition, the supplementary algorithm can be described as follows:

1) Assume that after using the initial software test data set generation algorithm, the obtained initial software test data set is  $\{t_1, t_2, \dots, t_n\}$ , and the covered paths are  $\{P_1, P_2, \dots, P_N\}, p_1, p_2, \dots, p_k$  are all the  $k$  predicates in program module  $M$ . Assume that the set of  $P_j$  covered predicates is  $\Omega_j$ . If  $i$  exists and makes  $p_i \notin \bigcup_{j=1}^N \Omega_j$

or  $\sim p_i \notin \bigcup_{j=1}^N \Omega_j$ , then supplement new software test data  $t'$  for  $p_i$  or  $\sim p_i$ .

2) Assume the obtained software test data set is  $\{t_1, t_2, \dots, t_m\}$  after the supplement of software test data in Step 1. For any software test data  $t_i \in \{t_1, t_2, \dots, t_m\}$ , assume the covered sub-path pair set is  $\Psi_i$ . Then the sub-path pair set covered by software test data set  $\{t_1, t_2, \dots, t_m\}$  is  $\Psi = \bigcup_{j=1}^M \Psi_j$ . If sub-path pair set  $\{(b_1, b_2), (b_2, b_3), \dots, (b_{x-1}, b_x)\} \in \Psi$  exists, but  $\forall \in [1, M]$ , there is

$$\{(b_1, b_2), (b_2, b_3), \dots, (b_{x-1}, b_x)\} \notin \Psi,$$

then supplement new software test data  $t''$  for sub-path  $b_1, b_2, b_3 \dots, b_x$ .

### 3. Algorithm analysis and experiment

The complexity of the algorithm is mainly concentrated in the initial software test data set generation algorithm Step 3. In the worst case, for a program module containing  $n$  predicates, the worst-case algorithm has a time complexity of  $O(2^n)$ , since there may be at most  $2^n$  different paths in the program. However, in practice, the actual complexity of the algorithm is much lower because the condition  $\sim p_1^{t_1}, \sim p_2^{t_1 t_2}, \dots, \sim p_{k-1}^{t_1 t_2 \dots t_{k-1}} \notin c^k$  in Step 3.1 of the initial software test dataset generation algorithm will cause a large number of unfeasible paths to be removed in time. In addition, the method of automatically generating software test data by using the spatio-temporal function granularity method, even for the linear function, needs to be tested several times in order to find the appropriate software test data; The new algorithm proposed in this paper combines the advantages of both the multidimensional approximation and the spatio-temporal function granularity. Therefore, the new algorithm requires only one iteration of the linear function to get the appropriate software test data. The use of non-linear function, multidimensional approximation and spatio-temporal function granularity method makes the new algorithm converge faster from the initial input to the satisfying software test data. Moreover, the new algorithm relies on the idea of multidimensional spatio-temporal granularity to update the software test data for predicates and sub-paths that are not covered by the initial software test dataset, So there is better test coverage performance. The program module  $M_1$  is illustrated as an example, and the program module  $M_1$  is shown in Fig. 1.

For  $M_1$ , according to the algorithm described in Section 2, the specific implementation steps are as follows.

First solve the DUC expression of program module  $M_1$ . Assume the automatically generated initial input is  $I_0 = (1, 2)$ .

For  $p_1$ , it can be seen from algorithm Step 3.1 that  $I_0$  ensures the execution of  $\sim p_1$ . According to algorithm Step 3.4, another group of solutions  $I_1 = (2, 0)$  can be obtained. And  $I_1$  ensures the execution of  $p_1$ . Then get  $\Gamma \leftarrow \leftarrow -x + y, p_1, (3, 2) \right\rangle$ .

For  $p_2$ : for the record 1 in  $\Gamma$   $1 : \leftarrow -x + y, \sim p_1, (1, 2) \right\rangle$ , it can be seen from algorithm Step 3.1 that  $(1, 2)$  ensures the execution of  $\sim p_2$ . Record  $\Gamma \leftarrow \leftarrow$

$\{-x + y, x + 3y\}, \{\sim p_1, \sim p_2\}, (1, 2) >$ , then another group of solutions  $I_1 = (-1, 0)$  can be obtained from algorithm Step 3.2 and 3.3; obviously,  $(-1, 0)$  ensures the execution of path  $\sim p_1, p_2$ . So

$$\Gamma \leftarrow \langle \{-x + y, x + 3y\}, \{\sim p_1, \sim p_2\}, (-1, 0) \rangle .$$

<i>r</i>	<i>d</i>	<i>U</i>	<i>c</i>
<i>read</i> ( <i>x, y</i> )	<i>x, y</i>		
if ( <i>x</i> > <i>y</i> ) then	<i>p</i> <sub>1</sub>	<i>x, y</i>	
<i>w</i> = <i>x</i> + 2 * <i>y</i>	<i>w</i>	<i>x, y</i>	<i>p</i> <sub>1</sub>
else			
<i>w</i> = <i>y</i>	<i>w</i>	<i>y</i>	: <i>p</i> <sub>1</sub>
endif			
if ( <i>w</i> + <i>y</i> < 0) then	<i>p</i> <sub>2</sub>	<i>w, y</i>	
<i>x</i> = <i>x</i> - 2	<i>x</i>	<i>x</i>	<i>p</i> <sub>2</sub>
<i>y</i> = <i>y</i> + <i>w</i>	<i>y</i>	<i>w, y</i>	<i>p</i> <sub>2</sub>
write( <i>y</i> )	<i>o</i> <sub>1</sub>	<i>y</i>	<i>p</i> <sub>2</sub>
else			
if ( <i>x</i> + <i>y</i> - <i>y</i> <sup>3</sup> < 0) then	<i>p</i> <sub>3</sub>	<i>w, y</i>	: <i>p</i> <sub>2</sub>
write( <i>y</i> )	<i>o</i> <sub>2</sub>	<i>y</i>	<i>p</i> <sub>3</sub>
else			
write( <i>w</i> )	<i>o</i> <sub>3</sub>	<i>w</i>	: <i>p</i> <sub>3</sub>
endif			
endif			

Fig. 1. Program module  $M_1$

For the record 2 in  $2 : \langle -x + y, p_1, (3, 2) \rangle$ , we also can get

$$\Gamma \leftarrow \langle \{-x + y, x + 3y\}, \{p_1, p_2\}, (0.5, -0.5) \rangle$$

and  $\Gamma \leftarrow \langle \{-x + y, x + 3y\}, \{p_1, \sim p_2\}, (3, 2) \rangle$ .

For  $p_3$ , since  $c_3 = \{\sim p_2\}$ , so it can be known from algorithm Step 3.4 that only two pieces of records need to be considered in  $\Gamma, \langle \{-x + y, x + 3y\}, \{\sim p_1, \sim p_2\}, (1, 2) \rangle$ .

For  $\langle \{-x + y, x + 3y\}, \{\sim p_1, \sim p_2\}, (1, 2) \rangle$ , we can only obtain

$$\Gamma \leftarrow \langle \{-x + y, x + 3y, x - 4y + 6\}, \{\sim p_1, \sim p_2, p_3\}, (1, 2) \rangle ;$$

while for path  $\{\sim p_1, \sim p_2, \sim p_3\}$ , it cannot solve the software test data meeting the condition.

For  $\langle \{-x + y, x + 3y\}, \{p_1, \sim p_2\}, (3, 2) \rangle$ , we can only obtain

$$\Gamma \leftarrow \langle \{-x + y, x + 3y, x - 4y + 6\}, \{p_1, \sim p_2, p_3\}, (3, 2) \rangle .$$

However, for path  $\{\sim p_1, \sim p_2, \sim p_3\}$ , it cannot solve the software test data meeting the condition.

The obtained paths  $\Gamma \leftarrow \langle \{-x + y, x + 3y, x - 4y + 6\}, \{\sim p_1, \sim p_2, p_3\}, (1, 2) \rangle$  are feasible paths, and the corresponding software test data are  $t_1=(0.5, -0.5)$ ,  $t_2=(-1, 0)$ ,  $t_3=(3, 2)$  and  $t_4=(1, 2)$ , respectively. The covered sub-path pairs and predicates are shown in Tables 1 and 2.

Table 1. Subpaths covered by initial software test data set

Test data	$(p_1, p_2)$	$(p_1, \sim p_2)$	$(\sim p_1, p_2)$	$(\sim p_1, \sim p_2)$	$(p_2, p_3)$	$(\sim p_2, p_3)$
$t_1$	*					
$t_2$			*			*
$t_3$		*				*
$t_4$				*		*

Table 2. Predicates covered by initial software test data set

Test data	$p_1$	$\sim p_1$	$p_2$	$\sim p_2$	$p_3$	$\sim p_3$
$t_1$	*		*			
$t_2$		*	*			
$t_3$	*			*	*	
$t_4$						

Table 1 and Table 2 show that the sub-path  $\sim p_3$  is not covered by  $t_1, t_2, t_3, t_4$ . According to Step 1 in software test data supplementary algorithm, the obtained software test data set is the subpath pairs and predicate covered by

$$\{t_1 (0.5, -0.5), t_2 (-1, 0), t_3 (3, 2), t_4 (1, 2), t_5 (2, 1), t_6 (1, 1)\} ,$$

which is shown in Tables 3 and 4.

Table 3. Sub-paths covered by software test data set after the first supplement

Test data	$(p_1, p_2)$	$(p_1, \sim p_2)$	$(\sim p_1, p_2)$	$(\sim p_1, \sim p_2)$	$(p_2, p_3)$	$(\sim p_2, p_3)$	$(\sim p_2, \sim p_3)$
$t_1$	*						
$t_2$			*			*	
$t_3$		*				*	
$t_4$				*		*	
$t_5$		*					*

Table 4. Predicates covered by software test data set after the first supplement

Test data	$p_1$	$\sim p_1$	$p_2$	$\sim p_2$	$p_3$	$\sim p_3$
$t_1$	*		*			
$t_2$		*	*			
$t_3$	*			*	*	
$t_4$						
$t_5$						

Table 3 and Table 4 show that the sub-paths  $(\sim p_1, \sim p_2), (\sim p_2, \sim p_3)$  meet the conditions of Step 2 in software test data supplementary algorithm. Therefore, software test data should be added. Assume that the new added software test data is  $t_6(1, 1)$ , then the obtained software test data set is the subpath pairs and predicate covered by  $\{t_1(0.5, -0.5), t_2(-1, 0), t_3(3, 2), t_4(1, 2), t_5(2, 1), t_6(1, 1)\}$ , which is shown in Tables 5 and 6.

Table 5. Sub-paths Subpaths covered by software test data set after the second supplement

Test data	$(p_1, p_2)$	$(p_1, \sim p_2)$	$(\sim p_1, p_2)$	$(\sim p_1, \sim p_2)$	$(p_2, p_3)$	$(\sim p_2, p_3)$	$(\sim p_2, \sim p_3)$
$t_1$	*						
$t_2$			*			*	
$t_3$		*				*	
$t_4$				*		*	
$t_5$		*					*
$t_6$				*			*

Table 6. Predicates covered by software test data set after the second supplement

Test data	$p_1$	$\sim p_1$	$p_2$	$\sim p_2$	$p_3$	$\sim p_3$
$t_1$	*		*			
$t_2$		*	*			
$t_3$	*			*	*	
$t_4$						
$t_5$	*				*	*
$t_6$		*		*		*

Obviously, Table 5 and Table 6 no longer meet the conditions of the software test data supplement algorithm. Thus, the finally obtained software test data set is  $\{t_1(0.5, -0.5), t_2(-1, 0), t_3(3, 2), t_4(1, 2), t_5(2, 1), t_6(1, 1)\}$ .

In particular, the software test data

$$\{t_1(0.5, -0.5), t_2(-1, 0), t_3(3, 2), t_4(1, 2), t_5(2, 1), t_6(1, 1)\}$$

obtained in this case covers all the 6 pieces of feasible paths  $p_1p_2, \sim p_1p_2, p_1 \sim p_2p_3, \sim p_1 \sim p_2p_3, p_1 \sim p_2 \sim p_3$  and  $\sim p_1 \sim p_2 \sim p_3$  in the program. Finally, in order to further verify the performance of the algorithm, experiments were conducted depending on the test pool data in literature [12–14]. The results are shown in Fig. 2. The test results are the average value of each group of program test.

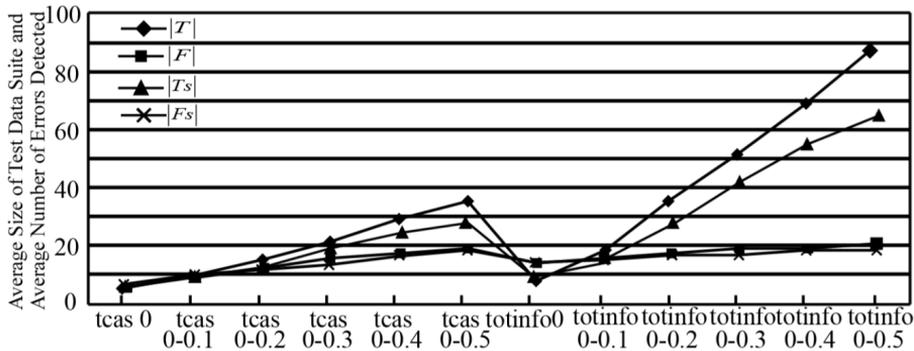


Fig. 2. Average size of the software test data set and the average number of detected errors

In Fig. 2,  $|T|$  represents the average size of the original software testing data set.  $|F|$  represents the average number of detected errors with the original software test data set. Quantities  $|T_s|$  and  $|F_s|$  mean the average number of average size and error detection software test data obtained by using the proposed algorithm, respectively. The experimental results show that the novel algorithm proposed in this paper can effectively reduce the number of software test data on the basis of guarantee the testing performance.

### 4. Conclusion

Software test is a very cumbersome and complex but extremely important stage in software development, especially for large-scale system software and application software. If the potential errors and defects in the software are not detected in time, they will cause serious consequences. In this paper, a software test data generation algorithm based on the idea of multi-dimensional space-time granularity was proposed. The algorithm analysis and experimental results show that the new algorithm can effectively reduce the influence of the unfeasible path to the algorithm, reduce the number of software test data, and improve the test efficiency. The next step is mainly to approach and solve approximation appropriate software test data for complex calculus functions, and to test software of structured program containing the string and numerical calculation and to generate test data automatically.

## References

- [1] H. Y. CHEN, T. H. TSE: *Equality to equals and unequals: A revisit of the equivalence and nonequivalence criteria in class-level testing of object-oriented software*. IEEE Transactions on Software Engineering *39* (2013), No. 11, 1549–1563.
- [2] J. W. CANGUSSU, R. A. DECARLO, A. P. MATHUR: *A formal model of the software test process*. IEEE Transactions on Software Engineering *28* (2002), No. 8, 782–796.
- [3] T. J. OSTRAND, E. J. WEYUKER, R. M. BELL: *Predicting the location and number of faults in large software systems*. IEEE Transactions on Software Engineering *31* (2005), No. 4, 340–355.
- [4] H. LI, S. DING: *Research of individual neural network generation and ensemble algorithm based on quotient space granularity clustering*. Applied Mathematics & Information Sciences *7* (2013), No. 2, 701–708.
- [5] X. YAO, D. GONG, W. WANG: *Test data generation for multiple paths based on local evolution*. Chinese Journal of Electronics *24* (2015), No. 1, 46–51.
- [6] R. L. ZHAO, Y. H. MIN: *Automatic test data generation of character string based on predicate slice*. Journal of Computer Research and Development *39* (2002), No. 4, 473–481.
- [7] C. CHIU, P. L. HSU: *A constraint-based genetic algorithm approach for mining classification rules*. IEEE Transactions on Systems, Man, and Cybernetics, Part C: (Applications and Reviews) *35* (2005), No. 2, 205–220.
- [8] A. RATAJ, M. NOWAK, P. PECKA: *Modelling CTMC with a standard programming language and using conventions from computer networking*. Theoretical and Applied Informatics *23* (2011), Nos. 3–4, 229–243.
- [9] W. MILLER, D. L. SPOONER: *Automatic generation of floating-point test data*. IEEE Transactions on Software Engineering *SE-2* (1976), No. 3, 223–226.
- [10] M. J. GALLAGHER, V. L. NARASIMHAN: *ADTEST: A test data generation suite for Ada software systems*. IEEE Transactions on Software Engineering *23*, (1997), No. 8, 473–484.
- [11] N. GUPTA, A. P. MATHUR, L. M. SOFFA: *Automated test data generation using an iterative relaxation method*. International Symposium on Foundations of Software Engineering (SIGSOFT), 1–5 November 1998, Lake Buena Vista, FL, USA, ACM SIGSOFT Software Engineering Notes *23* (1998), No. 6, 231–244.
- [12] M. J. BALCER, W. M. HASLING, T. J. OSTRAND: *Automatic generation of test scripts from formal test specifications*. Symposium on Software Testing, Analysis, and Verification (TAV3), 13–15 December 1989, Key West, FL, USA, ACM SIGSOFT Software Engineering Notes *14* (1989), No. 8, 210–218.
- [13] D. SHIN, E. JEE, D. H. BAE: *Comprehensive analysis of FBD test coverage criteria using mutants*. Software & Systems Modelings *15* (2016), No. 3, 631–645.
- [14] D. BIDEAU, R. HÉKINIAN: *A dynamic model for generating small-scale heterogeneities in ocean floor basalts*. Journal of Geophysical Research *100* (1995), No. B6, 10141 to 10162.

Received June 29, 2017



# Chaotic time series prediction based on the fusion of multi-source collaborative data feature constraints

BAOYAN ZHANG<sup>1</sup>

**Abstract.** With the continuous development of multi-source collaborative data network technology, multi-source collaborative data has been widely applied and become an important infrastructure in the field of information technology. However, the intelligent prediction based on the fusion of multi-source collaborative feature constraints often fails to obtain the accurate predictive relevant real-time information as the chaotic time series perceptual information demand of intelligent applications normally cannot be translated into simple query requests and multi-source sensor underlying query interfaces. In view of this problem, a multi-source collaborative data network information resource description, reasoning and application model based on the fusion of multi-source collaborative data feature constraints is proposed in this paper, which combines, and on the basis of the chaotic time series prediction support as the application foundation, the chaotic time series prediction technology based on the fusion of the multi-source collaborative data feature constraints is studied. Experiments prove that, the improved model can realize the accurate positioning of the specific multi-source sensor in the multi-source collaborative data network and obtain the real-time sequence prediction mechanism of the corresponding perceptual information.

**Key words.** Multi-source collaborative data, feature constraints; prediction, multi-source collaborative data network; chaotic time series.

## 1. Introduction

With the rapid development of artificial intelligence, the chaotic time series has been widely applied in the fields of intelligent transportation [1], military application [2], emergency handling [3], disaster relief [4] and other aspects. Chaotic time series often needs to obtain a lot of real-time information about the physical world in the course of the completion of the prediction tasks. For example, in the disaster rescue, when rescuing the people in the disaster area, it is required knowing the location of the wounded person, the surrounding environment, injury situation and so on, so that rapid and targeted rescue can be carried out. The emerging multi-

---

<sup>1</sup>School of Information Technology and Engineering , Jinzhong University, Jinzhong, Shangxi, 030619, China

source collaborative data network technology has provided a brand new perceptual information service mechanism with wide coverage, strong real-time performance and other features, and this information service mechanism is on the basis of the multi-source sensor network system with mutual communication and interoperability in the multi-source collaborative data network, to provide chaotic time series with extensive and comprehensive real-time information, so as to significantly improve the efficiency of the chaotic time series to accomplish the prediction task.

The disunity of multi-source data collaboration and heterogeneity of sensor information is a major bottleneck in the application of the autonomous planning and prediction support of chaotic time series in the multi-source collaborative data network [5–7]. It is a hot research topic in recent years on how to describe the multi-source sensor effectively and make use of the description information of the multi-source sensor in the acquisition and application of the perceptual information. Although the aforementioned research methods can achieve effective semantic representation of the multi-source sensors, they are not suitable for the direct application to the multi-source collaborative data network.

Firstly, the multi-source collaborative data network is a service network based on the fusion of the sensing resources [8–10]. Therefore, the semantic modeling for the multi-source collaborative data network must be service-oriented semantic modeling, and not just the semantic modeling for the multi-source sensor;

Further, in consideration of the availability of the sensor resources and the close correlation with the temporal and spatial features, in the semantic modeling, it is necessary to conduct unified modeling for the temporal and spatial features and the effectiveness of services [11–12].

On this basis, in this paper, a semantic modeling approach for the multi-source collaborative data network service resources and their temporal and spatial features. In the modeling approach: We first encapsulate the perceptual resources in the multi-source collaborative data network through the restful service to generate the corresponding multi-source collaborative data network resource service;

Furthermore, through constructing a semantic model for the restful services the semantic model, the semantic description of the multi-source collaborative data network service resources is realized. Finally, we introduce the temporal and spatial description method in the semantic model, and achieve the effective semantic modeling for the temporal and spatial features of the multi-source collaborative data network resource service.

In this paper, different from the former fusion based multi-source data collaborative feature constraint work, the multi-source collaborative data feature constraint prediction task planning method in view of the multi-source collaborative data feature constraints is put forward. The method can make full use of the semantic description meta-information in view of the multi-source collaborative data network resources, and realize the dynamic mapping from high-level semantic description information requirement to the underlying multi-source sensor resources by integrating the planning reasoning of the dynamic description logic. Based on this method, the chaotic time series prediction model can effectively utilize various existing multi-source collaborative data networks to realize the effective prediction task assignment,

the real-time prediction task execution monitoring for various emergency prediction tasks in the physical world, and dynamic prediction task adjustment and prediction task re-planning, etc.

The complex prediction task here refers to the prediction task which needs to be completed by several agents collaboratively and collaboratively. Based on the team-oriented plan (hereinafter referred to as multi-source data collaborative feature constraints) [6], this paper constructs a new method for the prediction of the features of complex tasks. The reason of the fusion based multi-source data collaborative feature constraints to construct the multi-source data collaborative feature constraints is that: the multi-source data collaborative feature constraints can provide a framework to construct the prediction task feature constraint plan according to the behavioral capacity of each individual in the team. With the integration of this framework, effective breakdown of complex prediction tasks can be realized.

## **2. Chaotic time series prediction model based on the fusion of multi-source collaborative data feature constraints**

### ***2.1. Chaotic time series prediction mechanism***

When an abstract complex prediction task is accomplished in the chaotic time series, it requires a lot of real-time information in the physical world as the prediction basis. However, the multi-source collaborative data network can be regarded as a database with the capability to provide massive information about the physical world. Therefore, we propose a chaotic time series prediction support system based on the fusion of the multi-source collaborative feature constraints (hereinafter referred to as CTSPSS for short). CTSPSS obtains the prediction task from the prediction task interface and passes it to the chaotic time series; the chaotic time series constrains the abstract complex task features to be completed as a series of sub-prediction tasks with lower coupling degree according to the prediction task knowledge in the knowledge base; and then, each sub prediction task is assigned to the specific agent.

The agent needs to perceive the real-time state of the physical world when it accomplishes the task, and it can make reasonable prediction on the basis. However, as the agent system usually describes the environment states that need to be perceived in the high-level semantic form (such as description logic), while for the heterogeneous multi-source collaborative data network entities (such as gateways and sensing nodes) in the multi-source collaborative data network, the perceptual information can only be exported through the specific query oriented method, therefore, CTSPSS can construct a multi-source collaborative data feature constraints based on the semantic model of the multi-source collaborative data network resources. The multi-source collaborative data feature constraints realizes unified description on the individual attributes and access interfaces of the multi-source collaborative data network resources through the resource description method of the fusion ontology, so that the chaotic time series can obtain the corresponding multi-source collaborative data network resources by analyzing the semantic description. The chaotic

time series can constrain the complex prediction task features into a series of simple prediction tasks, and obtain the chaotic time series perceptual information needed to accomplish the simple prediction task in the multi-source collaborative data network.

To sum up, whether it is possible to obtain the required information resources to predict the execution of the task from the multi-source collaborative data network determines whether the prediction task can be successfully executed. Therefore, when the prediction task feature constraints are conducted, the chaotic time series needs to determine whether the resources in the multi-source collaborative data network can meet the information requirements of the prediction task execution. In the following section, we will discuss in details how the chaotic time series can perform the prediction task feature constraints according to the available information resources in the multi-source collaborative data network. First of all, it is required making formal specification on the problem of chaotic time prediction task feature constraint prediction for the multi-source collaborative data network.

*Definition 1.* A prediction task  $T$  is a two-tuples with the form of  $T = \langle Init, Goal \rangle$ .

$Init$  and  $Goal$  are both state description sets,  $Init$  and  $Goal$  are composed of a set of description logic formula. In which,  $Init$  represents the system state before the execution of the task; and  $Goal$  is the expected system state after the task is executed, that is, the target state.

*Definition 2.* The breakdown scheme of a prediction task  $T$  is a four-tuples with the form of  $Schema(T) = \langle TaskSet, State, Action, QuerySet \rangle$ .

Quantity  $TaskSet = \{T_1, T_1, \dots, T_n\}$  is the set of the sub-prediction task broken down by the chaotic time series according to the knowledge base,  $T_k$  is the  $k$ th sub prediction task,  $State = \{state_1, state_2, \dots, state_n\}$  is the environmental information needed to accomplish the sub prediction task, where  $state_k$  is the environment information needed to accomplish the abstract complex prediction task  $T_k$ ,  $Action = \{A_1, A_2, \dots, A_n\}$  is the behavioral prediction needed for the agent to accomplish the prediction task, in which,  $A_k = Decision(T_k, state_k)$  is the action prediction taken by the agent to accomplish the sub prediction task  $T_k$ ,  $QSet_k = \{Q_1, Q_2, \dots, Q_M\}$ ,  $QSet_k \in QuerySet$  is the set of the environment information required to accomplish the sub prediction task  $T_k$ . Obviously,  $QSet_k$  is determined by  $state_k$ .

For example, assuming that the chaotic time series receives an abstract complex prediction task "Fire fighting for building B", through breakdown, a number of sub-prediction tasks can be obtained, in which, the sub-prediction task  $T_i$  is "The fire engine arrives at the scene of the fire". In order to accomplish the sub-prediction task, the required environment state perception is state= "Traffic state from the fire site to the fire center". Although the multi-source collaborative data feature constraint is constructed, and the semantic information fusion query can be realized, but as the query granularity is relatively large, the multi-source collaborative data network cannot directly understand and complete this type of environmental state query, therefore, it is necessary to further break down the abstract state query, and ultimately generate the semantic query operation that can be directly understood and executed by the machine,  $QuerySet = \{Q_1 = \text{"Beijing No. 1 Road traffic pressure multi-source sensor state"}, Q_2 = \text{"Beijing No. 2 Road traffic pressure multi-}$

source sensor state",  $Q_3 = \text{"Beijing No. 1 Road traffic lights state"}$ ,  $Q_4 = \text{"Beijing No.2 Road traffic lights state"}$ , ... }. Through these specific query statements, the chaotic time series can obtain the necessary environmental information from the multi-source collaborative data network.

## *2.2. Chaotic time series prediction based on the fusion of multi-source data collaborative feature constraints*

As mentioned earlier, action prediction relies on the real-time sensing of the environment information state. However, state is usually a relatively abstract state of the environment, making it difficult for the machine to directly map the state queries to the directly executable multi-source collaborative data network atomic query statements. Therefore, CTSPSS requires the query breakdown process that maps the state abstract query to a specific atomic query set QuerySet. According to QuerySet, CTSPSS can determine whether the multi-source collaborative data network can provide the necessary information support for the action execution.

The process of breaking down the state query is a semantic breakdown process with decreasing abstraction degree by layer, and the semantic breakdown by layer can be achieved through the prediction task feature constraint knowledge in the multi-source data collaborative feature constraint knowledge base. However, the multi-source data collaborative feature constraint knowledge base is usually oriented to a specific prediction task, while state usually represents a more general physical world state. Therefore, it is very difficult to identify a multi-source data collaborative feature constraint breakdown Map to correspond to the state. Therefore, we try to construct a state query breakdown method based on the fusion ontology.

It can be known from Algorithm 1 that the information that needs to be queried from the multi-source collaborative data network is a subset of the action execution preconditions set, and the action execution prerequisite represented by the dynamic description logic consists of two forms: the conceptual assertion formula with the form of  $C(x)$  and the relation assertion formula with the shape of  $R$ . Their respective implication is that: If individual  $x$  is an instance of  $C$ , the precondition holds; if there is  $R$  relationship between individual  $x$  and individual  $y$ , this prerequisite holds. The prerequisites of both types can be regarded as the state verification of the individual in the physical world, for example, HighTemperature represents the verification whether the temperature of room  $r$  is too high, and InSameRoom represents the verification whether  $x$  and  $y$  are in the same room, etc.

To simplify the discussion, we assume that the multi-source collaborative data network only provides the perceptual information on the attributes of various physical objects in the physical world, such as the physical object location, running speed, running direction and so on. At the same time, such attribute information can be obtained from the multi-source sensor associated with the physical object. Therefore, the problem of the state verification for a physical object can be converted to the acquisition of the perceptual information about the multi-source sensor associated with the physical object. Furthermore, since the physical object-related chaotic time series implies that the information acquisition of multi-source sensors requires temporal-spatial constraints, the state verification of the physical object can

be mapped to the acquisition behavior of certain types of perceptual information in certain time period and spatial domain. Based on the aforementioned analysis, we can transform the verification of the physical object state of the action execution prerequisite to the breakdown query of the perceptual information with the temporal and spatial constraints, and finally form a series of query requests related to the multi-source sensor in the multi-source collaborative data network.

Through the ontology, we can represent a variety of perceptual information required for the validation of a certain physical object state, such as the state validation on whether a room is on fire can be expressed as the concept definition of and regarding to RoomOnFire:

$$\text{RoomOnFire} \equiv \text{Room} \sqcap \text{hasSensingDevice} . (\exists \text{hasLocation} . (\text{InRoom} \sqcup \text{NearRoom}) \sqcap \text{observes} . (\text{Pmperty} \sqcap \text{Smoke}) \sqcup \text{observes} . (\text{Pmperty} \sqcap \text{HighTemperature})))$$

The specific meaning of this concept definition is that: If the multi-source sensor that is located in a room near the room detects smoke or temperature increase, the room has an outbreak of fire.

According to this definition, it is possible to determine the state of RoomOnFire (No.116) by obtaining the chaos time series perceptual information about the smoke and temperature of the multi-source sensor in the room No. 116 or in the vicinity of the room.

It can be noted that the concept definition that describes the physical entity state only describes the logical relationship of the perceptual information related to the state concept itself, while lacks the temporal and spatial constraints. Therefore, in the actual query breakdown, it is necessary to combine the concept definition with the temporal and spatial concept to form the query with the temporal and spatial constraints. For example, the temporal and spatial constraints as the following:

$$\text{RoomFireState} \equiv \text{RoomOnFire} \sqcap \text{hasTime} . ("2012-1-8") \sqcap \text{hasLocation} . ("168Express Hotel")$$

The concept of the temporal and spatial constraints corresponds to a query on whether there was the information on an outbreak of fire in a room at 168 Express Hotel on July 8, 2012. After the temporal and spatial constraint is added, the status validation  $\text{RoomFireState}(\text{No.116})$  on the physical object "No.116" is conducted. When "No.116" is located at the hotel "168 Express Hotel", the time is "2012-7-8", and the state of RoomFireState (No.116) is true,  $\text{RoomFireState}(\text{No.116})$  returns the true value.

It can be seen from the aforementioned analysis that, the query breakdown for the conceptual assertions is the basis of the whole physical object state verification. Therefore, we need to design a query breakdown method for the conceptual assertions. The basic idea of this query breakdown method is as the following:

For a conceptual assertion  $C(x)$ , first of all, search for the concept definition of  $C$  in the ontology, if it cannot be found, it will return that system cannot verify  $C(x)$ ; If the corresponding concept definition is found, the system will breakdown the concept definition according to the nested form of the concept into a tree structure (as shown in Fig. 1), and try to gradually match with the multi-source collaborative data feature constraints from the leaf nodes, to explore whether there is instance object in the multi-source collaborative data network information database that can

satisfy the semantics of the node.

Then, the sibling nodes of the breakdown tree are merged to generate a new concept node, and the new concept node is matched with the multi-source collaborative data feature constraint and so on until the root node of the tree is matched. If the root node of the tree is matched, and there is object in the multi-source collaborative data network that belongs to the concept of the root node, the object is further filtered by the temporal and spatial constraint.

Finally, if there is an instance of the concept corresponding to the root node that satisfies the temporal and spatial constraints, query breakdown can be implemented on the system returns.

In order to improve the breakdown efficiency of the conceptual assertions, we define certain specification on the form of concept definition. The specification states that: The root node of the breakdown tree corresponding to the concept definition must contain a child node with the form of  $\exists hasSensingDevice.()$ , and the child node has intersection relations with other sub nodes. That is, the concept definition must be the intersection of the concept and the sets with certain multi-source sensors.

Furthermore, through the analysis on the sub-tree with  $\exists hasSensingDevice.()$  as the root node, it is possible to obtain the corresponding multi-source sensor set which can verify the state of the object, and then through the temporal and spatial constraint relationship, the corresponding perceptual query set can be determined.

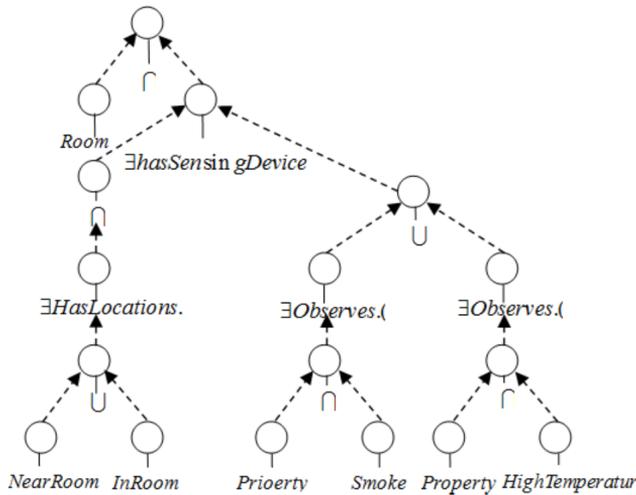


Fig. 1. Concept hierarchical tree example

### 3. Experimental analysis

In order to validate the proposed semantic prediction support technical scheme of the multi-source collaborative data network based on the fusion of the chaotic time series proposed in this paper, we analyze the availability and advancement

of this design by the application of an implemented insurance accident handling command support time series prediction system as an example. The implemented overall framework of the system is shown in Fig. 2.

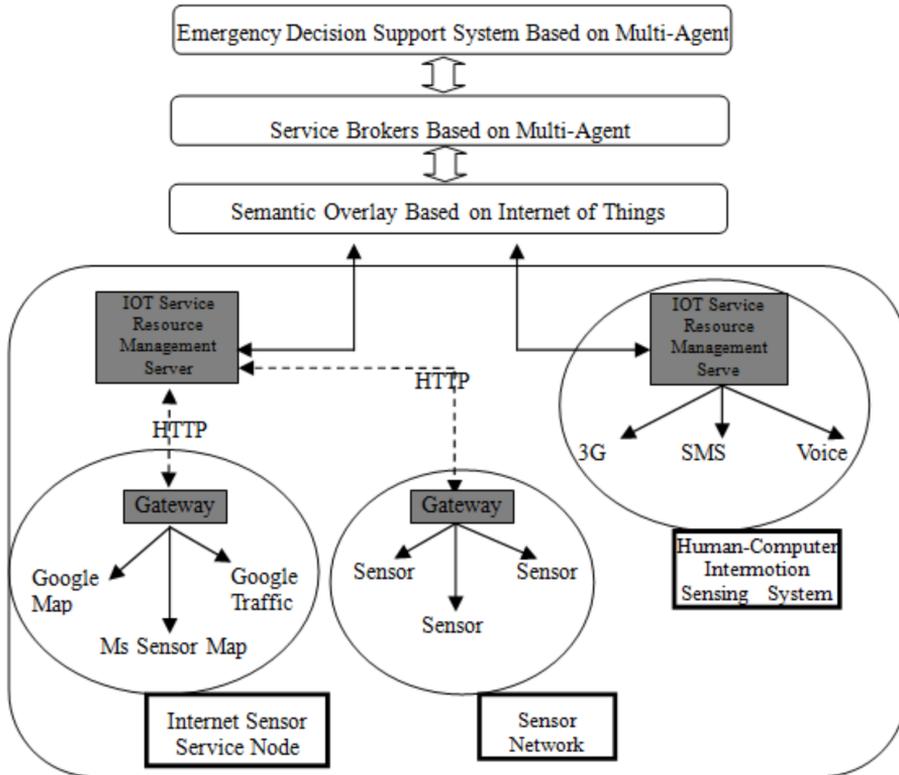


Fig. 2. Frame diagram of time series prediction system for command support in insurance accident handling

In this system, we define the multi-source collaborative data network sensor node can be connected through the 6LowPAN and other protocols to the gateway. The gateway registers the service information of the multi-source sensor node with which it is connected to the IOT service resource management server. This process can be achieved through manual construction or automatic discovery and other methods. Due to the lack of the corresponding physical multi-source sensor support, we adopt Google Map and Microsoft Sensor Map and other virtual multi-source sensor services for the case verification. At the same time, the program takes the smart phone as a special type of human-computer interaction multi-source sensor, and the smart phone can obtain user information and status through 3G, voice, text messages and other methods, and can also give commands to the designated executor. In respect of the physical multi-source collaborative data, we have defined the resource management server to generate its semantic meta-information about the IOT registration service. There are multiple IOT service resource management servers

in the system; therefore, the service resources and its corresponding semantic meta-information are distributed. Distributed semantic meta-information stored in each resource management server has formed the multi-source collaborative data feature constraints. The service agent system composed of the chaotic time can achieve the multi-source collaborative data network resource discovery and service agent, etc. through the distributed semantic reasoning. In this system, all the breakdown, matching, resource positioning on the query is completed by the service agent system. And the chaotic time emergency accident handling system is above the service agent system, which obtains the IOT real-time information through the service proxy system, and implements the corresponding prediction task feature constraints and execution for the insurance accident emergency accidents.

For the specific traffic accidents, the system has formed the specific implementation scheme through the prediction task feature constraints as shown in Figure 3. By the query breakdown of the fusion based multi-source collaborative data feature constraints, the variety of information requirements in the prediction task is finally transferred into a series of sensor information query and management information query. According to the finally obtained prediction task query system and the underlying multi-source collaborative data, information is acquired and returned to the user for the predictive support.

The implementation process of the insurance accident handling command support time series prediction system is shown in Figure 8, and its basic process is as the following:

When a traffic accident occurs, the system makes the accident notification, chaotic time series breaks down the prediction task, first conducts inspection on the policy, and sends the detailed information of the accepted policy to the forecaster.

Call out all underwriting vehicles information in the vicinity where the accident occurs. The system will automatically generate the location where the accident occurs on the map on the basis of the accident report, and at the same time, according to the vehicle status, mark the current location of the underwritten vehicle and its free status on the map, and automatically prioritize all the underwriting vehicles executing the prediction task for the forecaster to select the underwriting vehicle to perform the prediction task;

According to the selected underwriting vehicles, the multi-agent system provides the current road condition and multiple routes planning through the GoogleMap road condition. The forecaster selects the route of the prediction task execution according to the road condition.

When the route selection is successful, the system automatically generates the prediction task execution information, and sends text messages to the staff on board the vehicle through the system, to inform them the execution of the prediction task;

The chaotic time series obtains the information of the underwriting vehicle that executes the prediction task through the underlying multi-source collaborative data in real time to perform real-time surveillance on the underwriting vehicle. If the underwriting vehicle deviates from its prediction task execution route, the abnormal alarm feedback is provided to the forecaster, to help the forecaster track the latest state in real time and ensure that the prediction task is completed in a reasonable,

fast and error-free manner, so as to avoid insurance fraud and other commercial crime incidents.

In the system implementation effect, we compare the efficiency of the accident treatment method of this system with the traditional accident telephone treatment method, including two aspects: the customer service of the insurance company head-quarter reports the accident handling and the scheduling of the underwriting vehicle to the customer by phone. In the data analysis, we have intercepted the holiday peak travel period, that is, the period of time when the traffic accidents and congestion are most prone to occur, and the insurance company makes analysis on the ordinary accident handling records.

Table 1 shows the comparison data of the customer service handling time after the customer accident claim, as can be seen from the statistical results that: the support time series prediction system based on the fusion of the chaotic time can intelligently call out the customer insurance records, underwriting vehicle information and other materials quickly, to help customer service greatly shorten the accident service handling and service waiting time, and largely enhance the user experience effect.

Table 1. Test results of insulated resistance value ( $k\Omega$ )

	Average customer service handling time per small accident (minutes)	Average customer service quantity per hours (number)	Average customer waiting time (minutes)	Average waiting queue (persons)
Traditional telephone accident treatment method	15	5	10	4
Time series prediction system for command support in insurance accident handling	5	10	Almost do not wait	0

On the other hand, in Table 2 we have provided the statistical comparison of the efficiency of the underwriting vehicles on the handling of the ordinary accident. It can be seen from the statistical results that: the system can provide intelligent prompts to the fastest way to arrive at the scene in the terminal of the underwriting vehicle driver through the multi-source sensor analysis on the traffic state, which has greatly reduced the time for the underwriting vehicle to reach the scene of the accident, lowered the consumption of fuel caused by the traffic congestion in the road, and enhanced the utilization efficiency of the underwriting vehicle. For example, the accident handling efficiency of the underwriting vehicle is increased by 28.8%, while at the same time, the vehicle fuel consumption is reduced by 15.7%.

Table 2. Efficiency comparison of the selection of underwriting vehicles on the accident handling route

	Average time of arrival of an accident (minute)	Efficiency enhancement	Reduction in monthly fuel consumption of underwriting vehicle
Driving based on the experience of driver	45	-	-
Driving based on the recommended route by the chaotic time prediction support system	32	28.8 %	15.7 %

#### 4. Conclusion

The dynamic information service support of multi-source collaborative data network is the basis of the realization of the complex prediction task automation of the chaotic time series based on the fusion of the multi-source collaborative feature constraints. In this paper, the multi-source collaborative data feature constraint model of the multi-source collaborative data network is studied, and the breakdown method for the abstract complex prediction task based on the multi-source collaborative data feature constraints fusion and multi-source data collaborative feature constraints. Furthermore, the breakdown of the abstract environment state queries for the sub-prediction tasks is realized by the application of the multi-source data collaborative feature constraints and ontology knowledge base in this paper, and finally the information queries can be obtained directly from the multi-source sensors in the multi-source collaborative data network. Through the breakdown of the abstract prediction task twice, the system model proposed in this paper has successfully identified the corresponding multi-source sensor in the multi-source collaborative data network to provide the necessary sequence prediction for the chaotic time series.

#### References

- [1] W. Y. MU, R. LI, Z. Z. YIN, Q. WANG, B. Y. ZHANG: *Fusion prediction of mine multi-sensor chaotic time series data*. Journal of Computer Applications 32 (2012), No. 6, 1769–1773.
- [2] X. CUI, M. JIANG: *Chaotic time series prediction based on binary particle swarm optimization*. AASRI Procedia 1 (2012), 377–383.
- [3] Y. DANG, X. LV: *A chaotic time series crop forecasting model based on Bayesian semi-supervised SVR algorithm*. Journal of Computational Information Systems (2014), No. 10, 4179–4186.

- [4] M. ASSAAD, R. BONÉ, H. CARDOT: *A new boosting algorithm for improved time-series forecasting with recurrent neural networks*. *Information Fusion* 9 (2008), No. 1, 41–55.
- [5] C. BORMANN, A. P. CASTELLANI, Z. SHELBY: *CoAP: An application protocol for billions of tiny internet nodes*. *IEEE Internet Computing* 16 (2012), No. 2, 62–67.
- [6] F. ZAMBONELLI, A. OMICINI: *Challenges and research directions in agent-oriented software engineering*. *Autonomous Agents and Multi-Agent Systems* 9 (2004), No. 3, 253–283.
- [7] M. COMPTON, P. BARNAGHI, L. BERMUDEZ, R. GARCÍA-CASTRO, O. CORCHO, S. COX, J. GRAYBEAL, M. HAUSWIRTH, C. HENSON, A. HERZOG, V. HUANG: *The SSN ontology of the W3C semantic sensor network incubator group*. *Web Semantics: Science, Services and Agents on the World Wide Web* 17 (2012), 25–32.
- [8] L. CHANG, Z. SHI, T. GU, L. ZHAO: *A family of dynamic description logics for representing and reasoning about actions*. *Journal of Automated Reasoning* 49 (2012), No. 1, 1–52.
- [9] W. WANG, S. DE, G. L. CASSAR, K. MOESSNER: *Knowledge representation in the internet of things: Semantic modelling and its applications*. *Automatika—Journal for Control, Measurement, Electronics, Computing and Communications* 54 (2013), No. 4, 388–400.
- [10] I. F. SU, Y. C. CHUNG, C. LEE, Y. Y. LIN: *Efficient skyline query processing in wireless sensor networks*. *Journal of Parallel and Distributed Computing* 70, (2010), No. 6, 680–698.
- [11] K. P. FERENTINOS, T. A. TSILIGIRIDIS, K. G. ARVANITIS: *Energy optimization of wireless sensor networks for environmental measurements*. *Proc. IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA)*, 20–22 July 2005, Messian, Italy, IEEE Conference Publications (2005), 250–255.
- [12] H. GHOLIZADE-NARM, S. CHAFI, M. REZA: *Using repetitive fuzzy method for chaotic time series prediction*. *Journal of Intelligent & Fuzzy Systems* 28 (2015), No. 4, 1937 to 1946.

Received June 29, 2017

# Practical online learning model based on big data balance<sup>1</sup>

QU JIE<sup>2</sup>

**Abstract.** In view of the problem that the number of parameters is huge in Markov online learning, the convergence rate is slow and the online learning can't be implemented, a practical Markov online learning method based on the big data balance was proposed. Firstly, the learning parameters were represented in a practical manner to reduce the number of the learning parameters; and then, according to the a priori knowledge and observation data, the Markov method was used to learn and optimize the exploration and utilization of the balance relationship between the two; finally, Markov online learning method based on the big data balance was used to realize the rapid convergence of the learning process, so as to achieve the purpose of the online learning. The simulation results show that the algorithm can meet the requirements of real-time system performance.

**Key words.** Markov decision process, Markov online learning, big data balance.

## 1. Introduction

In the real dynamic system, the state transition function is usually unknown and dynamic. If the steady state model is used to describe the dynamic system, it will cause the distortion of the dynamic system modeling, so that the obtaining of the real approximation optimal value function and the optimal strategy cannot be guaranteed in theory. In view of this problem, it is necessary for the agent to learn in the interaction with the dynamic uncertain environment. Online learning is an effective optimal control learning method, which can be used under the conditions of complex model or uncertainty, etc. to realize the system multi stage optimization of the learning control based on the drive of the data [1–2].

The classical enhanced learning algorithms can be divided into two categories

---

<sup>1</sup>This work was support by the 13th Five-Year Research Project "Study on Cultivating Normal Students: Ability of the Information Technology Application Under the Background of Internet Plus Era", Supported by Education Science in Shaanxi Province (No.SGH16H169). The Key research Subject "Research on University Students' MOOC Study Status and Related Factors" Supported by Baoji University of Art and Sciences (No.ZK16073).

<sup>2</sup>College of Education of Baoji University of Art and Sciences, Baoji, Shaanxi, 721016, China

according to whether it is based on the big data, namely, the algorithms based on the big data balance and the algorithms based on the model-free data: the algorithms based on the big data balance and the ones based on the model-free. The algorithms based on the big data balance include TD learning, Q learning, SARSA [3] and other algorithms. The algorithms based on the model-free include DYNA-Q, priority sweep and other algorithms. The above classical enhanced learning algorithms have proved the convergence of the algorithms in theory. In the practical application areas, the number of learning parameters is huge. This is a typical NP difficult problem, which makes it hard to optimize the exploration and utilization of the balance of the two [4—5].

Markov Online Learning (referred to as MOL for short) conducts the modeling for the unknown model parameters by using the model priori knowledge, then makes the update to the posterior distribution of the unknown model parameters based on the observed data, and finally, conducts the planning according to the posterior distribution, thus obtaining the maximized expected reward value. Essentially, MOL transforms the learning problem into the planning problem. Since MOL can conduct the modeling for the unknown parameters and unknown models by using the priori distribution of the states, it has provided a perfect solution to optimize the balance of the optimized exploration and utilization, at the same time, it has gained extensive attention from the scholars at home and broad, and has become the hot spot in the research of the online learning field at present [6—7]. But there are two difficult problems existing in MOL: Firstly, the number of learning parameters is huge, and the scale increases exponentially [8]; secondly, the solution of the planning problem in the space of all the posterior belief states will encounter the “Curse of dimensionality” [9]. The above two difficult problems lead to the result that the existing MOL algorithms can only solve small scale problems, but cannot realize the online learning for large scale problems.

In this paper, a practical Markov online learning method based on the big data balance was put forward to solve the above problems, and the practical representation method was used to reduce the scale of the learning parameters; in the condition that the DBNs (Dynamic Markov Networks) structure (independent relationship between the variables) was unknown, Markov method was adopted to learn the unknown structure and parameter; finally, the Pointed-Based MOL (referred to as PBMOL for short) based on the big data balance was used to make the action choice in the posterior belief space, thus realizing the online planning and learning. The experimental and simulation results show that the proposed algorithm can effectively reduce the number of parameters and realize the online learning on the dynamic system.

## 2. Markov online learning modeling

The Markov Decision Processes (referred to as MDPs for short) can be described by a quaternion  $\langle S, A, T, R \rangle$ . The state set  $S = \{s_1, s_2, \dots, s_n\}$  contains all the possible states of the agent; the action set  $A = \{a_1, a_2, \dots, a_n\}$  contains all the possible actions of the agent; state transition function; the state transition function

$T(s, a, s') = P(s' | s, a)$ , taking the probability of the transfer from the action  $a$  to the state  $s'$  when the agent is in the states; the reward function  $R(s, a, s')$ , taking the reward value obtained from the transfer of the action  $a$  to the state  $s'$  in the state  $s$ .

In the online learning, the state transition function  $T(s, a, s')$  is unknown learning parameter  $\theta^{s,a,s'}$ . According to the literature [10], the Markov online learning based on the big data balance is defined as the Partially Observable Markov Decision Processes (referred to as POMDPs for short), which is described using the sextet-set  $\langle S_P, A_P, Z_P, T_P, O_P, R_P \rangle$ . Here,  $S_P$  stands for the cross product of the discrete state  $S$  and the continuous unknown parameter  $\theta^{s,a,s'}$ ; the action set  $A_P$  is the same as the action set  $A$ ;  $Z_P = S$ . The state transition function  $T_P(s, \theta, a, s', \theta') = P(s', \theta' | s, \theta, a)$  can be decomposed into the product of the two conditional distributions as the following:

$$\begin{aligned} T_P(s, \theta, a, s', \theta') &= P(s', \theta' | s, \theta, a) \\ &= P(s' | s, \theta, a, \theta') P(\theta' | s, \theta, a) \\ &= \theta^{s a s'} \delta^{\theta \theta'}. \end{aligned} \quad (1)$$

Here,  $\delta^{\theta \theta'}$  is the Kroneike function, which meets the following

$$\delta^{\theta \theta'} = \begin{cases} 1, & \theta' = \theta \\ 0, & \text{otherwise} \end{cases}. \quad (2)$$

The observation function  $O_P(s', \theta', a, z) = P(z | s', \theta', a)$  stands for the probability that when the agent executes the action  $a$  and the state and parameters are transferred to  $s'$  and  $\theta'$ , the observation is  $z$ . Since the reward function does not depend on  $\theta$  or  $\theta'$ , the reward function  $R_P(s, \theta, a, s', \theta') = R(s, a, s')$  is the same as the reward function of MDPs.

According to the definition of Markov online learning, the MDPs problem is transformed into the POMDPs problem. In POMDPs, as the state is unknown, the probability distribution  $b(s)$  of the state  $S$  is introduced, which is known as the belief. Through the introduction of the concept of belief,  $\theta$  learning can be performed by using the belief monitoring method [11]. Making use of the Markov update rule, the belief is updated as the following:

$$b^{s,a,s'}(\theta) = \eta b(\theta) P(s' | \theta, s, a) = \eta b(\theta) \theta^{s,a,s'}, \quad (3)$$

where  $\eta$  is the normalization factor.

The belief monitoring method is effective only when the priori and posterior distribution of beliefs are in the same distribution family, and Markov online learning adopts the Dirichlet distribution to represent the priori and posterior belief. As-

suming that the priori belief is  $b(\theta) = \prod_{s,a} D(\theta^{sa}; n^{sa})$ ,  $n^{sa}$  is the vector of the super-parameter  $n^{s,a,s'}$ , then its posterior distribution is given as

$$\begin{aligned} b^{s,a,s'}(\theta) &= \eta \theta^{sas'} \prod_{s,a} D(\theta^{sa}, n^{sas'}) \\ &= \prod_{s,a} D\left(\theta^{sa}, n^{sa} + \delta_{\bar{s}, \bar{a}, \bar{s}} \left(s, a, s'\right)\right). \end{aligned} \quad (4)$$

Here,  $\delta$  is still the Kroneike function, when  $s = \bar{s}, a = \bar{a}$  and  $s = \bar{s}$ , it is 1, and in other conditions, it is 0.

The goal of the online learning is to find the optimal strategy to optimize the balance exploration and utilization so as to obtain the maximum long-term reward value according to the posterior state of the current model under the condition that the state transition function is unknown. In the POMDPs, the strategy  $\pi$  is the mapping from the belief  $b$  to  $a$ , that is,  $\pi(b) \rightarrow a$ . The optimal strategy  $\pi^*$  is the strategy corresponding to the optimal value function  $V^*$ .

### 3. Practical Markov online learning based on the big data balance

In order to ensure that a good model is still available in the case of uncertainty, it is necessary to collect large amount of data during the learning process, which results in the exponential explosive growth of the learning parameters, thereby causing the failure of MOL to achieve rapid convergence. The practical representation method is an effective method to solve the ‘‘Curse of dimensionality’’ problem of the learning parameters [12]. In the practical representation methods, if the independent relationship between the variables in the DBNs is known, the size of the learning parameters can be easily compressed. However, in the practical application field, the structure of the DBNs is unknown. Therefore, the structure and parameters of the DBNs need to be learned at the same time.

#### 3.1. Practical learning representation

In most real-world models, through the analysis of the internal structure of the state variables that, it can be found that the state variables can be represented by a set of random variable sets, which are called practical properties. This internal characteristic is known as the practical characteristic. The practical state is represented by a random finite variable set  $X = \{X_1, X_2, \dots, X_n\}$ , in which each  $X_i$  stands for a characteristic of the state variable,  $X_i$  stands for the value set of each variable in the set  $X$ . One state can be represented as  $s = \{X_1 = x_1, \dots, X_n = x_n\}$ , where  $x_i \in X_i$ , which can also be represented by  $s = \{x_i\}_{i=1}^n$  for short. The state variable space is  $|S| = \prod_{i=1}^n |X_i|$ . After making the state variables practical, all the state transition function, the observation function, and the reward function can be represented using DBNs in compression [12].

Quantity  $G(a)$  is defined as a two-layer directed acyclic graph, in which,  $a \in A$ ,

the node is  $X = \{X_1, \dots, X_n, X'_1, \dots, X'_n\}$ . Symbol  $\theta_{G(a)}$  is defined as a conditional probability table, then the state transition function  $T$  can be expressed by the  $G(a)$  and  $\theta_{G(a)}$ . Quantity  $X_i$  is defined as the  $i$ th characteristic variable under the current state,  $X'_i$  is defined as the  $i$ th characteristic variable in the next moment. Finally  $X_{\rightarrow i}^{G(a)}$  stands for the value taken for the parent node when characteristic variable  $X'_i = x_i$ . Then the state transition function is calculated as follows

$$T(s, a, s') = T(X, a, X') = \prod_{i=1}^n P(X'_i | X_{\rightarrow i}^{G(a)}), \quad (5)$$

where  $T(s, a, s')$  is the state transition function at the time when the decomposition representation is not performed, and  $T(X, a, X')$  is the state transition function at the time of the practical representation.

In the FMDPs (Factored MDPs), since the state transition function, the observation function and the reward function can all be expressed by the conditional probability table of the DBNs, the above unknown models can be learned at the same time by repeating the calculation of the belief.

### 3.2. Belief posterior update

The practical Markov online learning can obtain the observation data, learning of unknown parameters and unknown structures through the interaction of the agents with the environment, so as to establish the state transition model and the reward model. In a deterministic environment, the initial belief  $b(s)$  is given, and  $b_{a,z'}(s')$  of its belief is calculated as

$$b_{a,z'}(s') = \eta \delta\left(\left[s'\right]_{Z'} = z'\right) \sum_s b(s) P(s' | s, a), \quad (6)$$

where,  $\eta$  is the normalized constant;  $\left[s'\right]_{Z'}$  is the subset of the state variable values corresponding to the observed set of variables  $Z'$ ;  $\delta$  is the Kronecker function, which returns the value 1 when  $\left[s'\right]_{Z'} = z'$  is true and returns value 0 when it is false. Since the model and the structure are unknown, according to the knowledge of the previous section, it can be known that the update process of the belief state is as the following:

$$b(X', \theta_{G(a)}) = \eta \delta \sum_X P(X' | X, a, \theta_{G(a)}) b(X, \theta_{G(a)}), \quad (7)$$

Here,  $X$  and  $X'$  are the practical representation of the variable characteristics,  $a$  is the action,  $Z$  is the observation data set,  $z$  is the subset in  $Z$ ,  $\theta_{G(a)}$  is the unknown parameter, and  $\delta$  is the Kronecker function.

However, since the belief state update requires the historical information, it is necessary to traverse all the historical observations and actions, resulting in the failure of convergence of Equation (7). According to literature [9], the belief state update is close-looped on the Dirichlet mixture product. Therefore, Dirichlet mixture product can be used to represent the state of the belief. And the representation form of the Dirichlet mixture product of the belief priori probability is as

$$b(X, \theta_{G(a)}) = \sum_i c_{i,X} \prod D_{i,X} \left( \theta_{G(a)}^{X_{G(a)}^i} \right), \quad (8)$$

where  $c_{i,X}$  is the Dirichlet coefficient,  $D$  is the Dirichlet distribution function,  $\theta_{G(a)}^{X_{G(a)}^i} = P(X' \mid \text{parents}(X'))$ . Hence, the posterior belief after the update of the state of the belief is

$$b_{a,z'}(X, \theta_{G(a)}) = \sum_j c_{j,X'} \prod D_{j,X'} \left( \theta_{G(a)}^{X_{G(a)}^j} \right). \quad (9)$$

### 3.3. Value function parameterization

According to the above knowledge, it can be known that the Markov online learning in the FMDPs field can be modeled by DBNs with the model variable  $\theta_{G(a)}$ . If the unknown model variable  $\theta_{G(a)}$  is used as the hidden variable of FMDPs, the FMDPs with unknown parameters can be transformed into the FPOMDPs (Factored POMDPs, POMDPs). According to the above conclusions, the existing FPOMDPs planning algorithms can be used to solve the MOL problem.

Porta and others put forward a BEETLE algorithm for iterative online learning for big data balance [9], which is only applicable to the MDPs field. On this basis, Porta et al. [10] proposed an improved BEETLE to process the continuous BEETLE algorithm and its improved algorithms take full advantage of the fact that the optimal value functions are parametric forms of the interface on the function set  $\alpha-$ , and the function  $\alpha-$  is a multivariate polynomial. However, they are on the basis of MDPs or POMDPs, and cannot be generalized to the field of FPOMDPs [4]. This section draws on BEETLE's ideas, and extends the value function parameters to the FPOMDPs field.

The optimal value function of the discrete POMDPs has the characteristic of the piecewise linear convexity, that is, the optimal value function can be expressed by the upper interface of the linear segment set  $\Gamma$  (known as the vector  $\alpha-$ ). The formula description is as the following:

$$V^*(b) = \max \alpha(b). \quad (10)$$

Each  $\alpha$  is the linear combination of the probability value of each characteristic variable, that is  $\alpha(b) = \sum_X c_X b(s)$ . For a discrete state space, the number of states is bounded, and  $\alpha$  can be expressed as the Dirichlet coefficient vector, that is  $\alpha(X) = c_X$ . For the continuous state POMDPs, the optimal value function is the upper envelope of the linear function (function  $\alpha-$ ) set, and the equation description

is

$$\alpha(b) = \int_X c_X b(X) dX. \tag{11}$$

In the practical online learning, assuming that the optimal value function at time  $k$  is  $V^k(b)$ , the set of function  $\alpha-$  is  $\Gamma^k$ , then the following can be obtained

$$V^k(b) = \max_{\alpha \in \Gamma^k} \alpha(b). \tag{12}$$

According to Bellman’s update equation, the optimal value function at the time  $k + 1$  is  $V^{k+1}(b)$ , the set of function  $\alpha-$  is  $\Gamma^{k+1}$ . Due to the introduction of the function  $\alpha-$ , the Bellman update equation can be rewritten as

$$\begin{aligned} V^{k+1}(b) = & \max_{\alpha \in A} \sum_X b(X) R(X, a, \theta_{G(a)}) + \\ & + \gamma \sum_{z'} P(z' | b, a, \theta_{G(a)}) \max_{\alpha \in \Gamma^k} \alpha(b_{a,z'}), \end{aligned} \tag{13}$$

According to the proof in literature [7], the function  $\alpha-$  is the linear combination of the Dirichlet product. In each Bellman backup, the number of Dirichlet products in the linear combination is equal to the size of the state space. Therefore, the linear combination size of the function will increase with the decision time exponentially in the scale.

### 4. Experimental results

For the Chain problem, the BEETLE algorithm in literature [9] and the MC-MOL algorithm in literature [6] put forward in recent years can represent the level of the current Markov online learning algorithm. In view of the Mountain climbing problem, comparison with the EFSL algorithm in literature [8] is conducted.

#### 4.1. Experimental results

In the Chain problem, there are two actions  $\{a, b\}$ , five states  $\{1, 2, 3, 4, 5\}$ , and the transfer probability of the action  $P = 0.2$ . Once the state 5 is achieved, the reward value is 10. Chain has Chain\_Tied, Chain\_Semi, Chain\_Full and other three versions. Chain\_Full refers that both the state transition function  $T(s, a, s')$  and state transition structure  $G$  are unknown. Chain\_Semi refers that the state transition structure is known, the state transition function is unknown and there is a dependency between actions. Chain\_Tied refers that the dynamic system is known, the action transition probability is unknown, and the action and state are independent. Therefore, the Chain problem has a variety of uncertainties, which is the ideal platform for the evaluation of the online learning algorithm.

In this paper, the three versions of the Chain problem were tested in 500 experiments, and each experiment executed 1000 steps (the number of iterations); then

for the reward value, the experimental results of the average and standard deviation was taken. The greater the value of reward, the more superior the algorithm shall be. Table 1 shows the comparison of the reward values of different algorithms, in which,  $n$ . Symbol  $v$  stands for not available, Optimal stands for the optimal value under the ideal condition; BEETLE is the value iterative algorithm based on the big data balance. The algorithm uses DDNs (Dynamic Decision Networks) to decompose the state. MC-MOL is the Markov online learning algorithm based on Monte Carlo, Q-Learning is a kind of  $\varepsilon$ - greedy strategy, and the value of  $\varepsilon$  ranges from 0 to 0.5, PB-MOL is the algorithm proposed in this paper. The experimental data is the sampling results of  $K = 1000$ .

Table 1. Comparison of the reward values for different algorithms

Problem	BEETLE	MC-MOL	Q-Learning	PB-MOL
Chain_Tied	3650 $\pm$ 41	3618 $\pm$ 29	1616 $\pm$ 24	3659 $\pm$ 20
Chain_Semi	3648 $\pm$ 41	<i>n.v.</i>	1616 $\pm$ 24	3661 $\pm$ 21
Chain_Full	1754 $\pm$ 42	1646 $\pm$ 32	1616 $\pm$ 24	2565 $\pm$ 23

As can be known from the experimental data in Table 1, the average value of PB-MOL and BEETLE and MC-MOL is not the same in the Chain\_Tied and Chain\_Semi problems with less uncertain factors, but the PB-MOL algorithm is closer to the true optimal value In the large-scale Chain\_Full problem, the state transition function and state transition structure are unknown, the uncertain factors are more, PB-MOL average reward value is 2565, which is significantly higher than BEETLE and MC-MOL algorithm. Therefore, BEETLE has more Good performance. As the use of Monte Carlo sampling method can effectively reduce the scale of the problem, the exploration and use can be balanced better. As Q-Learning is a free online learning method of this model, it is related to state transition function and other models are independent, so its reward value in the three versions of Chain remains unchanged. BEETLE, MC-MOL and PB-MOL are three different types of Markov online learning methods. As can be seen from Table 1, Kofu online learning method makes full use of the priori knowledge, which can effectively enhance the learning effect and improve the reward value.

Figure 1 shows the situation of the change of the accumulated reward value of four algorithms BEETLE, MC-MOL, Q-Learning and PB-MOL with the number of iterations. The experiment is the iterative result of the first 1000 steps. As an be known from Figure 1, the accumulated reward value of the PB-MOL proposed in this paper is maximum, while the accumulated reward value of the Q-Learning algorithm is minimum. The results of the BEETLE algorithm and MC-MOL algorithm are close to those of the PB-MOL algorithm. As can be known from the comparative experiment of the accumulated reward value, Markov online learning has more superior performance than the Q learning. In the iteration of the first 1000 steps, the learning rate of the algorithm is constant.

Table 2. Comparison of algorithm calculation time (in ms)

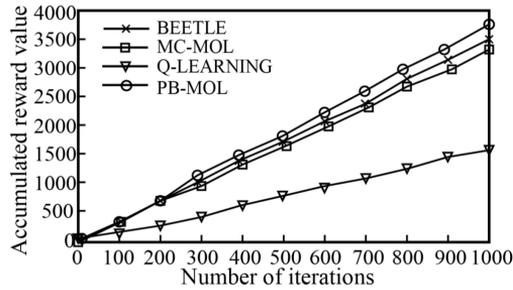


Fig. 1. Comparison of accumulated reward values

Problem	BEETLE		MC-MOL		PB-MOL	
	Offline	Online	Offline	Online	Offline	Online
Chain_Tied	400	1500	1.8e+6	32	400	18
Chain_Semi	1300	1300	<i>n.v.</i>	<i>n.v.</i>	1300	22
Chain_Full	14800	18000	<i>n.v.</i>	<i>n.v.</i>	14800	37

Table 2 shows the comparison of the calculation time for different Markov online learning algorithms, where *n.v.* stands for not available. From the data in the table, it can be seen that it is PB-MOL, where *n.v.* stands for not available. The data in the table shows that PB-MOL and MC-MOL on-line computations are less time consuming, which have higher real-time. However, from the value of the reward shown in Table 1, MC-MOL shows relatively big error in the large scale problem solving. PB-MOL off-line calculation method and BEETLE are the same, and there is the time-consuming problem. The offline pre-calculation will not affect the online real-time performance of the algorithm; at the same time, offline training can help obtain better priori knowledge, so as to get as large accumulated reward value as possible, which has properly solved the difficult problem of exploration and utilization in the online learning.

#### 4.2. Simulation experiment of car climbing problem

In the literature on the related online learning, the car climbing learning control is usually used as a typical continuous state space online learning problem to verify the learning efficiency and generalization performance of the algorithm. The goal of the car is to climb to the top of the mountain. Before climbing the top of the mountain will not get positive feedback, hence the car has no knowledge on the environment that it is in.

Due to the lack of power, the car cannot climb directly to the top of the mountain. Therefore, it must first climb to the left to get sufficient kinetic energy, so as to reach the top of the mountain. The car's kinetic equation is as the following:

$$\begin{cases} x_{t+1} = x_t + v_{t+1}, \\ v_{t+1} = v_t + 0.001a_t - 0.0025 \cos(3x_t), \end{cases} \quad (14)$$

where  $x$  stands for the location of the car,  $x \in [-1.2, 0.5]$ ,  $v$  stands for the speed of the car,  $v \in [-0.07, 0.07]$ ,  $a_t$  stands for the space of the action and  $A(s) \in \{-1, 0, 1\}$ . When  $x_t = -1.2$ , the speed of the car is 0; when  $x = 0.5$ , the goal of the car is achieved. And the beginning point of the car is  $x = -0.5$ ,  $v = 0.0$ .

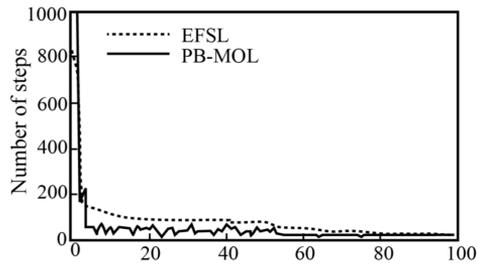


Fig. 2. Car climbing problem learning curve

The main evaluation index of the car climbing learning system is the number of steps of the car from the starting point position to the target position and the number of episodes required to achieve the steady state. The less the number of studies, the less the algorithm. In order to test the effectiveness of the proposed PBMOL learning algorithm, the algorithm is combined with EFSL (Enhanced Fuzzy Sarsa Learning). In this paper, the EFSL (Enhanced Fuzzy Sarsa Learning) and algorithm were compared to test the effectiveness of the PB-MOL learning algorithm proposed in this paper, and the experimental results were obtained, as shown in Fig. 2. In this experiment, the maximum number of steps was the 1000, the learning rate was 0.1, the discount factor was 0.9, the maximum and minimum of the temperature parameter were 0.1 and 0.001, respectively; and the sampling period was 0.02s. As can be seen from Fig. 2, the PB-MOL algorithm achieved the goal of climbing the car in 15 to 20 time steps after about 8 times of learning, while the EFSL needed about 100 time steps after about 10 times of learning, thus achieving the goal of the car climbing the mountain. If the EFSL needed to achieve the goal of the car within 20 time steps, it needed to learn at least 75 times. The experimental results show that, PB-MOL has better convergence and real-timeness than EFSL.

## 5. Conclusion

In view of the “curse of dimensionality” problem of the learning parameters in the Markov online learning based on the big data balance, the practical method was used to reduce the dimension of the unknown learning parameters in this paper. For the dynamic system of the model under the uncertain environment, the simultaneous learning of DBNs structure and the unknown parameters were used to effectively realize the real modeling of the dynamic system in the uncertain environment and solve the problem of the difficulty in the application modeling. The PRACTICAL

ONLINE LEARNING MODEL of the 11 unknown parameters were regarded as the hidden variables of MDPs, and the MDPs learning problem was transformed into the planning problem of POMDPs. All the existing POMDPs planning methods were applied to the MDPs learning through the transformation from MDPs to POMDPs, so as to solve the difficult problem of the online learning generalization. Finally, an online value iterative algorithm based on the big data balance was proposed to realize the online planning and learning.

## References

- [1] R. COSTANZA: *Science and ecological economics: Integrating of the study of humans and the rest of nature*. Bulletin of Science, Technology and Society 29 (2009), No. 5, 358–373.
- [2] H. T. LIU, B. R. HONG, S. H. PIAO, X. M. WANG: *Evolutionary algorithm based reinforcement learning in the uncertain environments*. Acta Electronica Sinica 34 (2006), No. 07, 1356–1360.
- [3] Q. LIU, J. LI, Q. M. FU, Y. C. FU, Z. M. CUI: *A multiple-goal Sarsa algorithm based on lost reward of greatest mass*. Acta Electronica Sinica 41 (2013), No. 8, 1469–1473.
- [4] S. ROSS, J. PINEAU, B. CHAIB-DRAA, P. KREITMANN: *A bayesian approach for learning and planning in partially observable Markov decision processes*. Journal of Machine Learning Research 12 (2011), No. 4, 1729–1770.
- [5] Y. GAO, J. K. HU, B. N. WANG, D. L. WANG: *Elevator group control using reinforcement learning with CMAC*. Acta Electronica Sinica 35 (2007), No. 2, 362–365.
- [6] F. DOSHI-VELEZ, J. PINEAU, N. ROY: *Reinforcement learning with limited reinforcement: Using Bayes risk for active learning in POMDPs*. Artificial Intelligence 187–188 (2012), 115–132.
- [7] H. ITOH, H. FUKUMOTO, H. WAKUYA, T. FUKUKAWA: *Bottom-up learning of hierarchical models in a class of deterministic POMDP environments*. International Journal of Applied Mathematics and Computer Science 25 (2015), No. 3, 597–615.
- [8] V. NA, H. NGO, S. LEE, T. C. CHUNG: *Approximate planning for bayesian hierarchical reinforcement learning*. Applied Intelligence 41 (2014), No. 3, 808–819.
- [9] J. ASMUTH, L. LI, M. L. LITTMAN, A. NOURI, D. WINTAGE: *A bayesian sampling approach to exploration in reinforcement learning*. Conference on Uncertainty in Artificial Intelligence (UAI), 18–21 June 2009, Montreal, Quebec, Canada, AUAI Press Arlington, Virginia, USA (2009), 19–26.
- [10] S. M. SMITH, M. JENKINSON, M. W. WOOLRICH, C. F. BECKMANN, T. E. BEHRENS, H. JOHANSEN-BERG, P. R. BANNISTER, M. DE LUCA, I. DROBNJAK, D. E. FLITNEY, R. K. NIAZY, J. SAUNDERS, J. VICKERS, Y. ZHANG, N. DE STEFANO, J. M. BRADY, P. M. MATTHEWS: *Advances in functional and structural MR image analysis and implementation as FSL*. Neuroimage 23, (2004), Suppl. No. 1, S208–219.
- [11] M. KEARNS, S. SINGH: *Near-optimal reinforcement learning in polynomial time*. Machine Learning 49 (2002), Nos. 2–3, 209–232.
- [12] H. L. S. YOUNES, M. L. LITTMAN, D. WEISSMAN, J. ASMUTH J: *The first probabilistic track of the international planning competition*. Journal of Artificial Intelligence Research 24 (2005), No. 1, 851–887.

Received June 29, 2017



# Analysis algorithm for internet of things big data based on multi-granularity functional

SUN JING<sup>2</sup>

**Abstract.** The Internet of things big data is a kind of important object of data mining. In Internet of things big data analysis process, if the time difference of the data is not considered, the misjudgment of the correlation will occur. First of all, from the perspective of the time warping, the reason and characteristics of two types of correlation errors were analyzed; Then, according to the asymptotic distribution of correlation coefficients, the boundary of correlation coefficients in a certain degree of significance level was obtained. The correlation method based on time shift sequence correlation coefficient characteristic was obtained by integrating both of them; finally, the multi-granularity functional model based on the maximum correlation coefficient was put forward, which has a wider application scope than AISE principle. The model is based on Multi-granularity maximize functional (MGMF) algorithm for solving the time warping function. The numerical experiment results in constructing data and real data show that, in the spurious regression identification, the correlation method is more effective than 3 kinds of conventional correlation coefficients and Granger causality test; in most cases the proposed MGMF algorithm is better than continuous monotone registration method (CMRM), self-modeling registration (SMR) and maximum likelihood registration (MLR).

**Key words.** Spurious regression, time warping; correlation, curve registration, Internet of things big data.

## 1. Introduction

The Internet of things big data is one of the most common data type of in data mining, which has been applied in many fields, such as the monthly runoff of certain river, the average monthly temperature and precipitation of local area, China's consumer price index (CPI) and gross domestic product (GDP), the seismic sequences of a number of observation points when the earthquake happened and so

---

<sup>1</sup>This work was support by the Xi'an Innovation Special Fund of science and technology project, CXY1352WL13.

<sup>2</sup>School of Mechanical and Materials Engineering, Xi'an University, Xi'an, Shaanxi, 710065, China

on. The analysis of these Internet of things big data can get some useful conclusions. For example: through the study of river history flow, temperature and precipitation characteristics, the level of flood forecast can be effectively improved; the utilization of CPI and GDP can analyze the extent of inflation and economic development momentum in countries or regions; according to the seismic wave sequences, the source and magnitude of earthquake can be accurately located [1]. Certain non-Internet of things big data can also be transformed by Internet of things big data for analysis, for example: the distance from the edge to the centroid of leaf can be used to describe its characteristics, and a series of data from different angles can be obtained, and then the type of leaf can be identified [2].

In the Internet of things big data analysis process, if the time difference of the data was not considered, it is easy to be affected by intuition or prejudice, and introduce erroneous judgement of the correlation; but the time difference does not consider the correlation sequence is of no significance. That is to say, determining the sequence correlation needs to consider the time difference. It is necessary to consider the time difference and correlation data, so the correlation and the time sequence among data are restricted with each other. At present, the correlation analysis of Internet of things big data is faced with some problems, such as the data relation is complex, data contains noise, missing data or abnormal data [3]. Homogeneous data (data with the same source or attribute, such as the seismic wave data with same earthquake obtained from multi areas) have natural similarity, which do not need to determine the correlation and does not exist correlation and time difference constraint problem. It is generally used for classification or clustering. While for the heterogeneous data (data with the different sources or attributes, such as precipitation and river runoff, CPI and GDP), it is necessary to determine their relevance. If there does exist relevance, regression analysis can be carried out, etc. Therefore, the main object of the Internet of things big data analysis is heterogeneous data.

The main reason of correlation error happened is the two groups of Internet of things big data have time warping. The coordination of both can be realized as long as time conversion is conducted. In reality, it is generally nonlinear time warping or dynamic time warping (DTW) [4, 5]. This requires the help of functional data analysis (FDA) method to transfer the Internet of things big data into function data for time correction, which is usually called curve registration or curve alignment. The main reason of correlation error happened is the two groups of Internet of things big data have time warping. The coordination of both can be realized as long as time conversion is conducted. In reality, it is generally nonlinear time warping or dynamic time warping (DTW) [4, 5]. This requires the help of functional data analysis (FDA) method to transfer the Internet of things big data into function data for time correction, which is usually called curve registration or curve alignment [6]. Kneip and Gasser regarded the extremum as landmark registration. But it is not suitable for the curves with inconspicuous feature points, and the selection of the feature points has great influence on the results. A more general method is: to determine an objective function or a curve, and to align the local features of other curves or minimize some metrics (such as the mean square distance between

each curve and the target curve) [7, 8]. Ramsay et al. proposed the continuous monotone registration method (CMRM), to ensure the continuity and consistency of the time warping function [9]. Wang and Gasser put forward a curve registration method based on dynamic time warping model [10, 11]. Kneip et al. used the local nonlinear regulating method about the time warping function to align adjust curves.

The homogeneous data has natural similarity, which can be curve registration function directly. To solve the restriction problem of heterogeneous Internet of things big data correlation and the time difference, the time difference is fixed to judge the correlation of each time shift sequence. Based on sequence correlation, The time difference function is then refined by the curve registration function. When the heterogeneous data are doing correlation judgment, on the one hand, since the sample correlation coefficient has deviation with the overall correlation coefficient, so upper and lower bounds on the overall correlation coefficient are studied; On the other hand, in order to prevent the emergence of two types of correlation errors, starting from the main reason of its occurrence, the characteristics of the two kinds of correlation errors are studied and a method to determine the corresponding correlation judgment method is put forward. The curve registration method suitable for the heterogeneous data is also applicable to homogeneous data, while the criterion suitable for homogeneous data (such as AISE) is not applicable to heterogeneous data (non-uniform dimension and negative correlation etc.). Therefore, mainly according to the characteristics of heterogeneous data, this paper proposes curve registration criterion based on the maximization of correlation coefficient (absolute value), and uses MGMF algorithm to solve the problem.

## 2. Correlation analysis method of curve registration

Because we can only get the sample data in solving practical problems, and it will produce deviation when using samples to estimate the population, this paper uses the sample correlation coefficient to infer the overall correlation coefficient in a certain degree of the boundary; At the same time, in order to prevent the occurrence of two related errors, this paper studies the characteristics of two kinds of shift sequence correlation coefficients, and then excludes two types of correlation errors. These two aspects can be determined to obtain two groups of correlation decision method of the Internet of things big data.

### 2.1. *Correlation decision of correlation sequences with time warping*

In order to determine the correlation of the sequence, it is necessary to infer the upper and lower bounds of the overall correlation coefficient. The overall correlation coefficient has upper and lower bounds in a certain degree. It was obtained from two asymptotic distributions of sample correlation coefficient. And then with the combination of the first type of correlation error, the correlation decision method of correlation sequences with time warping was obtained.

2.1.1. *Boundary of the correlation coefficient.* Pearson correlation coefficient is the most common metric form for measuring the sequence correlation. If there are two groups of corresponding data  $\{(x_i, y_i), i = 1, 2, \dots, n\}$  ( $n$  being the sample capacity), which are the samples come from binary normal population  $(x, y) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ , then the sample correlation coefficient is

$$\begin{aligned} \hat{\rho}(X, Y) &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \\ &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \cdot \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}}. \end{aligned} \tag{1}$$

Here,  $\bar{x}$  and  $\bar{y}$  are the sample averages of  $X$  and  $Y$ , respectively.

The sample correlation coefficient  $\hat{\rho}(X, Y)$  can be used as unbiased and consistent estimator of the correlation coefficient  $\rho$  of two normal populations  $(X, Y)$ . But the correlation coefficient has an obvious disadvantage that its degree closing to 1 is related to the number of data group  $n$ . It is easy to give an illusion. When  $n$  is small, the fluctuation of correlation coefficient is larger. The absolute value of some samples' correlation coefficient is close to 1, especially when  $n=2$ , the absolute value of the correlation coefficient is 1; When  $n$  is larger, the absolute value of the correlation coefficient is small. The distribution of results of the sample correlation coefficient, the sample size and the bivariate normal population correlation coefficient were given by many scholars.

Under the assumption that  $(X, Y)$  is the bivariate normal population and  $\rho=0$ , the distribution is as follows:

$$T = \frac{\sqrt{n-2}\hat{\rho}}{\sqrt{1-\hat{\rho}^2}} \sim t(n-2). \tag{2}$$

When  $\rho = \rho_0$ , Fisher proposed a more complex probability density function of  $\hat{\rho}$ . The following asymptotic distribution is obtained after proper transformation:

$$z = \frac{\phi(\hat{\rho}) - \phi(\rho)^{n \rightarrow \infty}}{2\sqrt{n-3}} \sim N(0, 1). \tag{3}$$

Here,  $\phi(x) = \ln \frac{1+x}{1-x}$ . When the sample size is large, the correlation coefficient can be used to estimate the population correlation coefficient.

Then  $n$  samples are extracted in the two normal states, and the following asymptotic distributions are obtained:

$$\sqrt{n}(\hat{\rho} - \rho)^{n \rightarrow \infty} \sim N\left(0, (1 - \rho^2)^2\right), \text{ namely } \frac{\sqrt{n}(\hat{\rho} - \rho)^{n \rightarrow \infty}}{(1 - \rho^2)} \sim N(0, 1). \tag{4}$$

In this paper, the population correlation coefficients are estimated based on the two asymptotic distributions.

Since  $\phi(x)$  is a monotonically increasing function in Formula (2), then

- When  $\rho \geq \hat{\rho}$ , then

$$P \left\{ \rho \leq \phi^{-1} \left[ \phi(\hat{\rho}) + 2z_{1-\frac{a}{2}} \cdot \sqrt{n-3} \right] \right\} = 1 - a. \quad (5)$$

- When  $\rho \leq \hat{\rho}$ , then

$$P \left\{ \rho \geq \phi^{-1} \left[ \phi(\hat{\rho}) - 2z_{1-\frac{a}{2}} \cdot \sqrt{n-3} \right] \right\} = 1 - a. \quad (6)$$

Herein,  $\phi^{-1}(x) = \frac{e^x - 1}{e^x + 1}$ ,  $Z_a$  is the  $a$ th quantile of standard normal distribution, namely  $P(x \leq Z_a)$  and random variable  $x \sim N(0, 1)$ .

On the basis of Formula (3), the bounds of the total correlation coefficient are deduced, namely:

- When  $\rho \geq \hat{\rho}$ , then

$$P \left\{ \frac{-\sqrt{n} - Q}{2z_{1-\frac{a}{2}}} \leq \rho \leq \frac{-\sqrt{n} + Q}{2z_{1-\frac{a}{2}}} \right\} = 1 - a. \quad (7)$$

- When  $\rho \leq \hat{\rho}$ , then

$$P \left\{ \frac{\sqrt{n} - R}{2z_{1-\frac{a}{2}}} \leq \rho \leq \frac{\sqrt{n} + R}{2z_{1-\frac{a}{2}}} \right\} = 1 - a. \quad (8)$$

In (7) and (8)

$$Q = \sqrt{n + 4\sqrt{n}z_{1-\frac{a}{2}} \cdot \hat{\rho} + 4z_{1-\frac{a}{2}}^2}$$

and

$$R = \sqrt{n - 4\sqrt{n}z_{1-\frac{a}{2}} \cdot \hat{\rho} + 4z_{1-\frac{a}{2}}^2}.$$

Synthesizing Formulas (4)–(7), when  $a = 0.05$ , it has the following approximation:

- When  $\rho \geq \hat{\rho}$ , then

$$\inf_{a=0.05} \rho = \max \left\{ \frac{-\sqrt{n} - \sqrt{n + 8\sqrt{n} \cdot \hat{\rho} + 16}}{4}, \hat{\rho}, -1 \right\} \quad (9)$$

and

$$\sup_{a=0.05} \rho = \min \left\{ \frac{-\sqrt{n} - \sqrt{n + 8\sqrt{n} \cdot \hat{\rho} + 16}}{4}, \phi^{-1} \left[ \phi(\hat{\rho}) + 4\sqrt{n-3} \right], 1 \right\}. \quad (10)$$

- When  $\rho \leq \hat{\rho}$ , then

$$\inf_{\alpha=0.05} \rho = \max \left\{ \frac{\sqrt{n} - \sqrt{n - 8\sqrt{n} \cdot \hat{\rho} + 16}}{4}, \phi^{-1} [\phi(\hat{\rho}) - 4\sqrt{n-3}], -1 \right\} \tag{11}$$

and

$$\sup_{\alpha=0.05} \rho = \min \left\{ \frac{\sqrt{n} + \sqrt{n - 8\sqrt{n} \cdot \hat{\rho} + 16}}{4}, \hat{\rho}, 1 \right\}. \tag{12}$$

Figure 1 shows the upper and lower bounds of the total correlation coefficient in different sample sizes  $n$  and sample correlation coefficients. As can be seen from the graph, the upper and lower bound curves presented in this paper have the following characteristics:

- The larger the sample size, the more compact the upper and lower bounds.
- When the sample size is the same, the upper and lower bound curves are central symmetry.
- The greater the absolute value of the correlation coefficient, the more compact of the upper and lower bounds.

The above characteristics can easily be proved by Formulas (8)–(11).

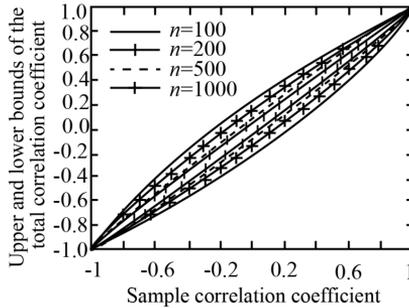


Fig. 1. Bounds of the total correlation coefficient (significance level  $\alpha = 0.05$ )

*2.1.2. Correlation determination method.* In order to describe the characteristics of the sequence, the definition of time-lag series is given.

Assuming there are two sequences  $(X, Y) = \{(x_i, y_i), i = 1, 2, \dots, n\}$ , the following sequence is defined as the time-lag series

$$(X_t, Y_{t+m}) = \{(x_i, y_{i+m}), i = 1, 2, \dots, n - m\}, 1 \leq m < n, m \in N^+, \tag{13}$$

$$(X_t, Y_{t-m}) = \{(x_i, y_{i-m}), i = m + 1, 2, \dots, n\}, 1 \leq m < n, m \in N^+.$$

For the first type of regression error, that is, there is no correlation between the related sequences, and just considering the initial sequence, then the inevitable correlation is small; if considering the correlation of time-lag series, then there does exist  $m_0 (1 \leq |m_0| \ll n, |m_0| \in N^+)$ , making the correlation coefficient of  $(X_t, Y_{t+m_0})$

larger.

Then the correlation decision of correlation sequences with time warping is obtained: if the correlation coefficient of time-lag series changes with  $m$ , and reached the maximum at  $m_0$ , namely  $\rho = \hat{\rho}(m)$  in the curve shows obvious convex phenomenon, then the range of correlation coefficient can be estimated according to Formulas (8)–(11). If  $|\rho(m_0)| > \rho_0$  (that is more than a given threshold such as 0.6), then the time-lag series  $(X_t, Y_{t+m_0})$  has correlation and the curve registration and regression analysis can be carried out.

## 2.2. A curve registration model based on functional correlation coefficient

The correlation coefficient of time-lag series can determine whether the sequences are correlated or not. If there is correlation in the two sequences, but with time deviation, it needs to align them with curve registration method to eliminate the differences in phase (time axis). For heterogeneous data, when using the AISE criterion, the results will change with the change of dimension. Therefore, we need to put forward a dimensionless criterion to align the heterogeneous data curves.

The Pearson correlation coefficient is a dimensionless measure describing two sequences correlation and similarity, but applies only to describe the correlation of discrete data. The correlation of continuous function can be represented by the inner product. At the same time, in order to make the inner value standardized, the norm of the two functions is divided. The curve registration criterion composed by heterogeneous data can be constructed through the functional correlation coefficient

$$\max_{h(t)} |\rho(x_1^*, x_2)| = \max_{h(t)} \left| \frac{\int_T x_1^*(s) x_2(s) ds}{\sqrt{\int_T [x_1^*(s)]^2 ds} \cdot \sqrt{\int_T [x_2(s)]^2 ds}} \right|. \quad (14)$$

Here,  $x_1^*(t) = x_1[h(t)]$  means the aligned function. In view of the complexity of the continuous function and the high dimensional feature of the optimization criterion, the corresponding discretization is given in this paper.

Assume the sample sequence of two functional data  $x_1(t)$  and  $x_2(t)$  at the sampling time points  $T = (t_1, t_2, \dots, t_n)$  are  $x_1(T) = [x_1(t_1), x_1(t_2), \dots, x_1(t_n)]$  and  $x_2(T) = [x_2(t_1), x_2(t_2), \dots, x_2(t_n)]$ , and now for function  $x_1(t)$  it is necessary to do curve registration as function  $x_2(t)$ . Order  $\Delta = (\delta_1, \delta_2, \dots, \delta_n)$  to be the offset of  $x_1(t)$  relative to  $x_2(t)$  at the time point  $T$ , namely the time warping function satisfies  $h(T) = T + \Delta$ , then the aligned time-lag series samples changes into  $x_1(T + \Delta) = [x_1(t_1 + \delta_1), \dots, x_1(t_n + \delta_n)]$ . The two groups of functional data sample sequences after aligned should have high correlation. The curve registration problem can be solved by transforming into

$$\max_{\Delta} |\rho[x_1(T + \Delta), x_2(T)]|. \quad (15)$$

Generally, the time warping function has uniform monotonicity, that is, satisfying  $t_{i-1} + \delta_{i-1} < t_i + \delta_i < t_{i+1} + \delta_{i+1}$ . However, the offset vector will cause the time

warping function does not satisfy uniform monotonicity, so  $\delta_i^{k+1} \in (bndl_i, bndr_i)$  is limited. Herein,  $bndl_i = t_{i-1} + \delta_{i-1}^k - t_i$ ,  $bndr_i = t_{i+1} + \delta_{i+1}^k - t_i$ ,  $\delta_i^k$ , means the value of  $\delta_i$  in the  $k$  iteration. In realization, the search range of  $\delta_i^{k+1}$  is reduced to closed interval  $[bndl_i + p \cdot (bndr - bndl_i), bndr_i - p \cdot (bndr - bndl_i)]$ . Herein,  $p$  is the constant in  $(0, 0.5)$ .

In the end, the problem of curve registration is changed to solve the following constrained optimization problem:

$$\begin{cases} \Delta^* = \arg \max_{\Delta} |\rho[x_1(T + \Delta), x_2(T)]|, \\ s.t. \delta_i \in [bndl_i + p \cdot (bndr - bndl_i), bndr_i - p \cdot (bndr - bndl_i)]. \end{cases} \quad (16)$$

Finally, the time offset vector  $\Delta^*$  is transformed into function form, and the time offset function  $d(t)$  is obtained. The corresponding time warping function is  $h(t) = d(t) + t$ .

### 3. Experimental results and discussion

Based on the simulated data (7 kinds of man-made Internet of things big data and two sets of Sinc function), the proposed correlation decision method and curve registration method were verified. The relevant data for the time warping did the curve registration using this method. On the one hand, the sensitivity of the method to parameters was analyzed; and on the other hand, the existing methods are compared and analyzed.

#### 3.1. Correlation decision

*3.1.1. Spurious regression decision.* For the spurious regression decision problem, 7 kinds of main data generating process (DGP) was selected. The correlation coefficient of the time-lag series is used to identify the spurious regression. First of all, the correlations in the 7 groups were analyzed by routine correlation analysis. The results are shown in Table 1.

Model 3 shows that: the correlation coefficients of model 3 and model 5 are relatively small, and have no spurious regression phenomenon. The 3 kinds of correlation coefficients of other groups autocorrelation sequences are relatively high, showing that the conventional correlation coefficient does not have the identify ability to the spurious regression phenomenon of autocorrelation sequence. Model 5 shows that: the Granger causality test can identify the correlation of most of the sequence, but there are still 3 identification errors. The model 7 is two-order autocorrelation sequence, and the bidirectional Granger causality tests were wrong. So for a simple model, Granger causality test can identify spurious regression; when the model is complicated, it can not identify the spurious regression. As the 3 correlation coefficients and Granger causality test can not decide the correlation of autocorrelation sequence, these methods are no longer used in the subsequent experiment for

correlation decision.

Figure 2 shows the time-lag series correlation coefficient varying diagram of 7 models. Figure 2 shows that: the absolute values of time-lag series correlation coefficient of model 3 and model 5 are smaller, which will not cause spurious regression. High correlation coefficients are observed in the other 5 models, but the correlation coefficient of time-lag series has little change with  $m$ . This method can quickly and accurately identify spurious regression. Although this first-order autocorrelation sequence spurious regression discriminant method is given in this paper, it can be seen from the results of model 7: for the two order autocorrelation sequence, it can also obtain reliable correlation results.

Table 1. Comparison of algorithm calculation time (in ms)

Model	Single integral order		Granger causality test (Lag=0)			
	X	Y	Cause→Effect	F statistics	Significance	Granger or not Reason
1	1	1	Y→X	0.006	0.993	No
			X→Y	0.287	0.750	No
2	0	1	Y→X	1.185	0.307	No
			X→Y	0.019	0.980	No
3	1	1	Y→X	0.412	0.662	No
			X→Y	0.129	0.878	No
4	1	1	Y→X	0.412	0.662	No
			X→Y	0.129	0.878	No
5	1	1	Y→X	0.513	0.599	No
			X→Y	0.164	0.848	No
6	4	1	Y→X	6.463	0.002	Yes
			X→Y	1.799	0.168	No
7	4	1	Y→X	4.035	0.019	Yes
			X→Y	16.653	0.000	Yes

*3.1.2. Correlation decision of time warping series.* Select the Sinc function with volatility ( $\text{Sinc}(x) = \sin \pi x / \pi x$ ,  $x \in [-6, 6]$ ) as the correlation research object of the simulated data, and do the following two functions:  $d_1(t) = 0.01t^2 - 0.36$  and  $d_2(t) = 0.005t(t-6)(t+6)$ . In the two cases, the correlation coefficient varying trends with the standard Sinc function and time-lag series are shown in Fig. 3.

It can be seen from Fig. 3, right part: two time-lag series correlation coefficient curves have obvious convex phenomenon, and bounds of two correlation coefficients are  $[0.991, 0.996]$  and  $[0.914, 0.962]$ . Then it can be decided that the two groups of

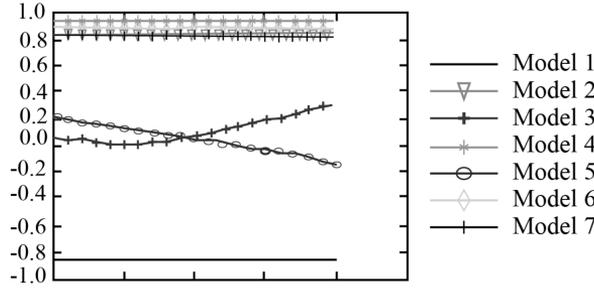


Fig. 2. Diagram of time-lag series correlation coefficient changing

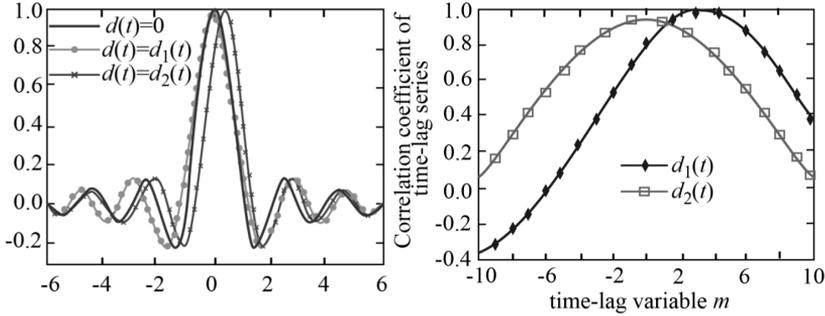


Fig. 3. Sinc function and the correlation coefficient variation of time-lag series: left—Sinc function and the function in two time differences, right—correlation coefficient variation of time-lag series under two kinds of time differences

series have correlation. And the average hysteresis amount with sequence of time-lag function of  $d_1(t)$  and  $d_2(t)$  and the standard Sinc sequence are 0 and 3, respectively.

### 3.2. Curve registration

This section is mainly testing the performance of the MGMF algorithm and compare with the classical CMRM algorithm [8], maximum likelihood registration (MLR) and self-modeling registration (SMR). For the sake of fairness, CMRM algorithm and MLR results are the average results of 5 experiments run. Machine configuration is: Intel Quad CPU (2.83 GHz dominant frequency), 3G memory.

The time difference functions are  $d_1(t) = 0.01t^2 - 0.36$  and  $d_2(t) = 0.005t \cdot (t - 6)(t + 6)$ , respectively. Put them with the noise-containing Sinc function and the standard Sinc function to do curve registration. In the two kinds of time difference functions, the align effects of 4 kinds of alignment methods after parameter adjustment are shown in Fig. 4.

Figure 4 shows that: when the time difference function is  $d_1(t)$ , the MLR alignment effect is poor, and the 3 others are closer to the actual value; when the time difference function is  $d_2(t)$ , the MLR and CMRM alignment effect is poor, and the SMR and MGMF are very close to the time difference function.

The above experiments show that for the simple time difference function, the

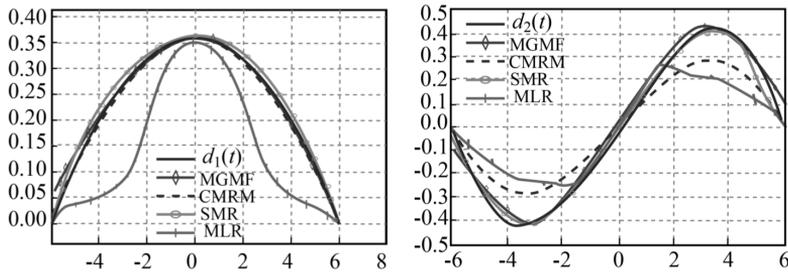


Fig. 4. Align effects of 4 kinds of methods: left-time difference function  $d_1(t)$ , right-time difference function  $d_2(t)$

precisions of CMRM, SMR and MGMF are high, but the CMRM efficiency is poor. When the time difference function is complicated, the SMR and MGMF results are in better alignment, but the efficiency and stability of SMR is better than MGMF.

## 4. Conclusion

In this paper, the upper and lower bounds of the population correlation coefficient are given at a certain significance level and used to discriminate the correlation. Spurious regression problems arise for more reasons, and there is not yet find a rigorous and accurate identification method. In this paper, starting from the main reasons of the spurious regression, we get the characteristics of the correlation coefficient of the time shift sequence, which can eliminate most of the common spurious regression phenomenon; For another kind of correlation error, the correlation coefficient can be determined from the characteristics of the time series correlation coefficient. Based on the correlation sequence with time warping, a model based on correlation coefficient maximization and an improved MGMF algorithm are established. The applicability of the model is more extensive than the AISE criteria. The experimental results show that the correlation method in this paper is more effective than Pearson linear correlation coefficient, Spearman rank correlation coefficient, Kendall rank correlation coefficient and Granger causality test in spurious regression recognition. The proposed MGMF algorithm is significantly superior to CMRM, SMR and MLR in most cases.

## References

- [1] F. SHANG, Y. JIANG, A. XIONG, W. SU, L. HE: *A node localization algorithm based on multi-granularity regional division and the lagrange multiplier method in wireless sensor networks*. *Sensors (Basel)* 16 (2016), No. 11, e1934.
- [2] J. DAI, X. LIU, F. HU: *Research and application for grey relational analysis in multi-granularity based on normality grey number*. *Scientific World Journal* (2014), No. 2, ID 312645.
- [3] A. ARRIBAS-GIL, H. G. MÜLLER: *Pairwise dynamic time warping for event data*. *Computational Statistics & Data Analysis* 69 (2014), 255–268.

- [4] P. M. THOMPSON, J. MOUSSAI, S. ZOHOORI, A. GOLDKORN, A. A. KHAN, M. S. MEGA, G. W. SMALL, J. L. CUMMINGS, A. W. TOGA: *Cortical variability and asymmetry in normal aging and Alzheimer's disease*. *Cerebral Cortex* 8 (1998), No. 6, 492–509.
- [5] X. HUANG, N. PARAGIOS, D. N. METAXAS: *Shape registration in implicit spaces using information theory and free form deformations*. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 28 (2006), No. 8, 1303–1318.
- [6] D. TELESKA: *Bayesian analysis of curves shape variation through registration and regression*. Book: *Frontiers in Probability and the Statistical Sciences, Nonparametric Bayesian Inference in Biostatistics*, Springer (2015), 287–310.
- [7] A. KNEIP, J. O. RAMSAY: *Combining registration and fitting for functional models*. *Journal of the American Statistical Association* 103 (2008), No. 483, 1155–1165.
- [8] T. CHAU, S. YOUNG, S. REDEKOP: *Managing variability in the summary and comparison of gait data*. *Journal of Neuroengineering & Rehabilitation* 2 (2005), No. 1, paper 22.
- [9] A. KNEIP, X. LI, K. B. MACGIBBON, J. O. RAMSAY: *Curve registration by local regression*. *Canadian Journal of Statistics* 28 (2000), No. 1, 19–29.
- [10] S. A. AHMADI, T. SIELHORST, R. STAUDER, M. HORN, H. FEUSSNER, N. NAVAB: *Recovery of surgical workflow without explicit models*. Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention, 1–6 October 2006, Copenhagen, Denmark, Springer Nature, LNCS 4190 (2006), Part No. 1, 420–428.
- [11] K. WANG, T. GASSER: *Asymptotic and bootstrap confidence bounds for the structural average of curves*. *Annals of Statistics* 26, (1998), No. 3, 972–991.

Received June 29, 2017

# Image analysis of workpiece joints with uncertainty for cooperative feature constraints<sup>1</sup>

XIAOFENG LV<sup>2</sup>

**Abstract.** In order to improve the denoising effect of the solder joint image of the uncertainty workpiece, a new method of noise removal for the solder joint image based on the cooperative feature constraint and the radial singular function. The image is decomposed by the cooperative feature constraint through the analysis of the image characteristic constraint coefficient, and skews high-frequency part and the low-frequency part of the feature constraint coefficients. In order to improve the denoising performance, an improved self-adaptive eigenvalue algorithm is proposed by calculating the median absolute variance estimation corresponding to the characteristic coefficient, and the image is subjected to the second denoising. Finally, the median filter is used to smooth the image and get the final denoising image. The results show that compared with the traditional method, the proposed algorithm can not only improve the denoising performance and denoising effect, but also keep the uncertainty of the workpiece solder joint image edge information.

**Key words.** Uncertainty, workpiece solder joint, image denoising, cooperative feature constraint, radial singular function.

## 1. Introduction

In the case of uncertainty in workpiece image detection and recognition [1], due to the presence of noise of different properties in the image, the noise reduces the image quality, makes the image blurred, and sometimes submerges some of the effective features of the image, and brings difficulty on the analysis and research. Because of the existence of mixed noise in the solder joint, the traditional method is used to remove the uncertainty of the workpiece image denoising, the denoising effect is not ideal, and it is easy to destroy the edge information of the image. In many cases, salt and pepper noise and Gaussian noise or multiplicative noise at the same time there is uncertainty in the workpiece solder joint image [2]. The feature is a generalized

---

<sup>1</sup>This work was supported by the Baoji science and technology plan project. Project name: Workpiece solder joint quality detection system. Project number: 16RKX1-7.

<sup>2</sup>Baoji University of Arts and Sciences, Baoji, Shaanxi, 721016, China

concept of wavelet [3–4]. Feature analysis can provide a more sophisticated analysis of the signal because it not only divides the low frequency band into multiple levels but also for the multi-resolution analysis does not subdivision of the high-frequency part of the further decomposition. So the feature technology has a wider application value, has been applied to pattern recognition, image processing and many other areas [5–6]. The skewness filter is a classical linear smoothing filter and is a local adaptive linear filter based on the principle of minimum mean square error. It can adjust the output according to the local variance. The skew filter has a good denoising feature for the noise of the known noise distribution [7]. The Median filter is a nonlinear filtering technique which can effectively suppress the image noise and improve the image signal to noise ratio. Value filter can effectively filter out random noise and salt and pepper noise, while the image can be fine to retain the edge of the details. In order to improve the denoising performance, Wang [8] proposed a feature combined with median filtering algorithm (WPM) to improve the denoising performance to achieve a certain effect.

In this paper, the noise removal method of the workpiece is analyzed. The residual image of the workpiece is decomposed by the uncertainty of the workpiece, and the noise of the original image is estimated by the median absolute variance estimation method. By combining with the skewness filter and improving radial singular and median filter, it denoises the image at the same time protect the edge less subject to fuzzy.

## **2. Uncertainty workpiece contact point image removal algorithm based on cooperative feature constraint and radial singular function**

### ***2.1. Feature constraint system data skew filter***

The image is decomposed into high frequency and low frequency parts. The image noise is mainly concentrated in the high frequency part, and the useful information of the image is mainly concentrated in the low frequency part. The high frequency part and partial low frequency part of the feature decomposition are filtered by using the skew filter to effectively remove the mixed noise in the image while preserving the edge and high frequency detail information of the image. For different levels of feature decomposition, the use of different window size of the skew filter, the characteristic coefficient skew filter process shown in Fig. 1.

### ***2.2. Radial singular function denoising algorithm***

The traditional feature threshold method is to keep the image coefficients and reduce most of the noise figure to zero. In practice, there are three common thresholds, namely hard threshold, soft threshold and half soft threshold. Whether the threshold selection is appropriate or not directly affects the effectiveness of the denoising algorithm. The threshold selection is too large, so that too many feature decomposition coefficients are set to zero, and too much image detail is destroyed;

the threshold is too small to achieve the expected denoising effect [9–10]. As a result of the traditional threshold processing method cannot achieve the requirements of denoising, the paper proposes an improved adaptive threshold denoising method, and the basic idea is as follows.

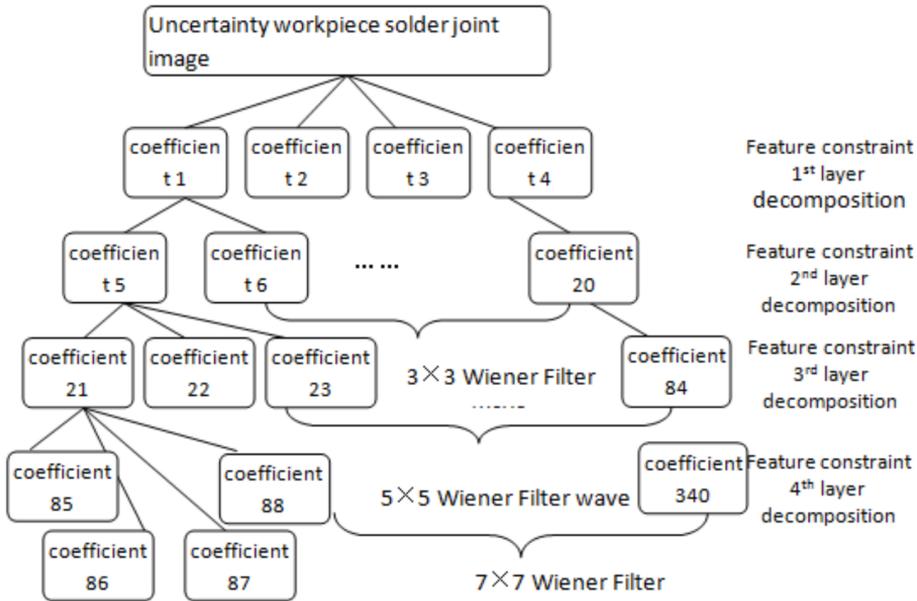


Fig. 1. Feature constraint coefficient skew filter

(1) Calculate the variance of each characteristic coefficient of the high frequency part of the feature layer decomposition

$$\sigma = \frac{\text{Median}(|w_{J,k}(s)|)}{0.6745}. \tag{1}$$

Here,  $J$  is the number of layers to be decomposed. Symbol  $k$  denotes a characteristic high frequency coefficient sequence and Median is the median function. Finally,  $w_{J,k}(s)$  is the characteristic coefficient and  $s$  is the noise image.

(2) The absolute value of the variance corresponding to the characteristic coefficient calculated according to equation (1) is arranged in descending order.

(3) According to the result of the arrangement obtained in step (2), the ordered intermediate value is  $\sigma_{J,M}$  and corresponding characteristic coefficient is  $w_{J,s}(m)$ . Symbol  $m$  is the number of characteristic coefficients.

(4) According to the result obtained in step (1), the optimal threshold value corresponding to the high frequency coefficient  $w_{J,s}(s)$  in the characteristic constraint in  $w_{J,d}(s)$ ' location  $P_{J,s}$  is calculated according to the formula (2). The obtained

optimal threshold  $\lambda_s$  is:

$$P_{J,s} = N(s)^2 - \frac{J^J N(s)}{\sigma_{J,s}}, \quad (2)$$

$$\lambda_s = |w_{J,s}(s)|_{P_{J,s}}. \quad (3)$$

In the above formula,  $N(s)$  is the width of the image after the decomposition of the  $J$ th layer.  $J^J$  is the function after the decomposition of the  $J$ th layer.

(5) Threshold processing of the characteristic coefficient gives

$$w_{J,s} = \begin{cases} w_{J,s}, & |w_{J,s}| \geq \lambda_s, \\ 0, & |w_{J,s}| \leq \lambda_s, \end{cases} \quad (4)$$

$$w_{J,s} = \begin{cases} w_{J,s} - \lambda_s, & w_{J,s} \geq \lambda_s, \\ 0, & |w_{J,s}| < \lambda_s, \\ w_{J,s} + \lambda_s, & w_{J,s} \leq -\lambda_s. \end{cases} \quad (5)$$

When  $s \leq m$ , the formula (4) is executed, and when  $s \geq m$ , the formula (5) is executed. After the characteristic high frequency coefficients are subjected to the optimal threshold processing, a new feature coefficient is obtained and the image is reconstructed. The image is subjected to secondary denoising, and most of the noise in the workpiece's solder joint image has been removed, but there may be a relatively large noise that cannot be removed. Smoothing noise is used to filter the edge of the image, and finally the denoised image is obtained.

### 3. Based on the synergistic feature constraint and the radial singular function of the uncertainty of the workpiece solder joint image denoising algorithm flow

Based on the above principle of image noise removal based on synergistic feature constraint and radial odd function, the noise removal algorithm of workpiece free solder joint image is designed. The process is as follows.

(1) Select the highest level  $N$  of feature decomposition, the characteristics of the workpiece solder joint image decomposition; get the characteristic coefficient of each layer.

(2) Using the skewness filter method, the image feature constraint coefficient is filtered.

(3) Calculate the absolute value of the variance corresponding to the characteristic coefficient, and sort it from descending order to ascending order, and obtain the intermediate value and the corresponding characteristic coefficient.

(4) According to the above-mentioned improved adaptive feature threshold calculation method, the optimal threshold is calculated.

(5) Reconstruct the feature coefficient.

(6) Smoothing the image using the median filter, the third denoising of the image,

and finally get the unpaired workpiece solder joint image. The algorithm flow is shown in Fig. 2

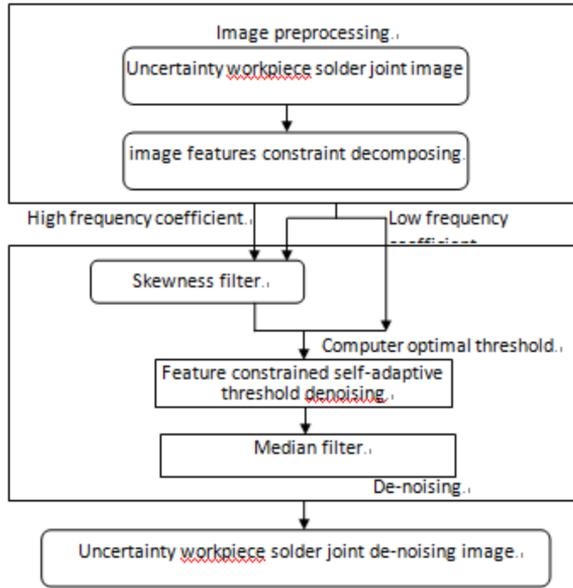


Fig. 2. Based on the synergistic feature constraints and radial singular function of the uncertainty of the workpiece solder joint image denoising algorithm flow

### 4. Denoising performance comparison

In order to compare the other methods and the denoising performance of the proposed method, the minimum mean square error (MSE) is used as the comparison parameter. The "Lena" images with the probability of salt and pepper noise of 0.02 and different degrees of Gaussian noise are denoised by different denoising methods to obtain the corresponding minimum mean square error, and to compare the different methods of denoising performance. The Median filter and the skew filter are 3×3, and 5×5, respectively. The synergistic feature constraints are respectively using 2, 3 and 4 layer decompositios. Only the best denoising results are listed in this paper, and the MSE results are calculated as shown in Table 1.

Table 1. Comparison of several methods of denoising performance

Method	(MSE) $\varepsilon$					
	$\sigma=15$	$\sigma=2$	$\sigma=25$	$\sigma=30$	$\sigma=35$	$\sigma=40$
Median	66.59	99.43	140.73	190.64	249.10	315.88
Skewness	177.08	184.30	194.37	207.69	224.83	246.11
WPM	96.64	103.30	111.33	120.62	131.17	142.75

It can be seen from Table 1 that the median filter can achieve better denoising effect when the mixed noise intensity is weak, but the performance of the characteristic combined median filter (WPM) is better than when the mixed noise is stronger. Other methods.

Using SRM method, respectively, using the characteristics of two layers of decomposition, skew the window pixel size  $3 \times 3$ , median filter window pixel size  $3 \times 3$  (method 1); Feature 2 layer decomposition, skewed window pixel  $5 \times 5$ , median filter window pixel  $3 \times 3$  (method 2), feature 3 layer decomposition, skewed window pixel  $3 \times 3$ , median filter window pixel  $5 \times 5$  (method 3). And for feature 3 layers decomposition, skewed window pixels  $5 \times 5$ , median filter window pixels  $5 \times 5$  (method 4), the results are shown in Table 2

Table 2. Mining SRM method image mixed noise removal

Method	(MSE) $\epsilon$					
	$\sigma=15$	$\sigma=2$	$\sigma=25$	$\sigma=30$	$\sigma=35$	$\sigma=40$
1	66.55	78.07	91.93	107.94	126.08	146.13
2	68.45	79.59	92.90	108.31	125.75	145.24
3	102.1	105.22	126.05	135.57	141.54	147.28
4	103.0	104.66	120.21	131.55	133.98	139.05

## 5. Uncertainty of workpiece solder joint image denoising results and comparison

Multiplicative noise is also one of the common noises in the solder joint image of uncertain parts. We choose the gray-scale image of the airfoil pin in IPC-A-610D as an example, as shown in Fig. 3, left part. After adding salt and pepper noise and multiplication, the noise is shown in Fig. 3, right part.

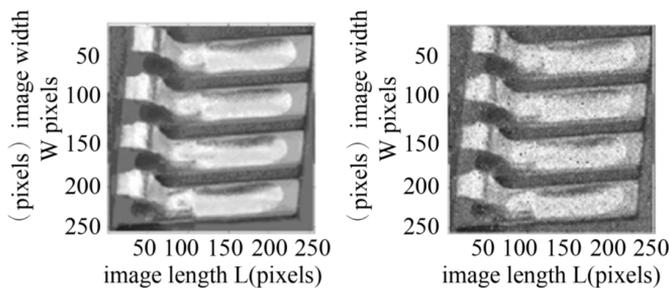


Fig. 3. Uncertainty workpiece solder joint image and noise image: left part—raw image of SMT solder joint, right part—added mixed noise

For the noise image, the median filter (window pixel size  $3 \times 3$ ), skew filter (window pixel size  $3 \times 3$ ), the characteristic soft threshold method for denoising, and denoising results are shown in Fig. 4.

It can be concluded from Fig. 4 that the median filter and the skew filter are not ideal for the denoising effect of the fixed part solder joint image. WPM method to

remove the smaller noise has a certain effect, but for most of the noise, the removal effect is not very good, still remain in the image, using the proposed SRM method, the median filter window pixel size were  $3 \times 3$ ,  $5 \times 5$ , the synergistic feature constraint is decomposed by 2, 3 layers. The results are Indicated in Fig. 5. It can be concluded from Fig. 5 and Fig. 3, right part, that the noise of the solder joint in Fig. 5 has been substantially eliminated, the denoising is achieved and the edge information is well preserved. The denoising effect is better than that shown in Fig. 4, and the effect of Fig. 5, right part, is slightly better than the effect of Fig. 5, left part.

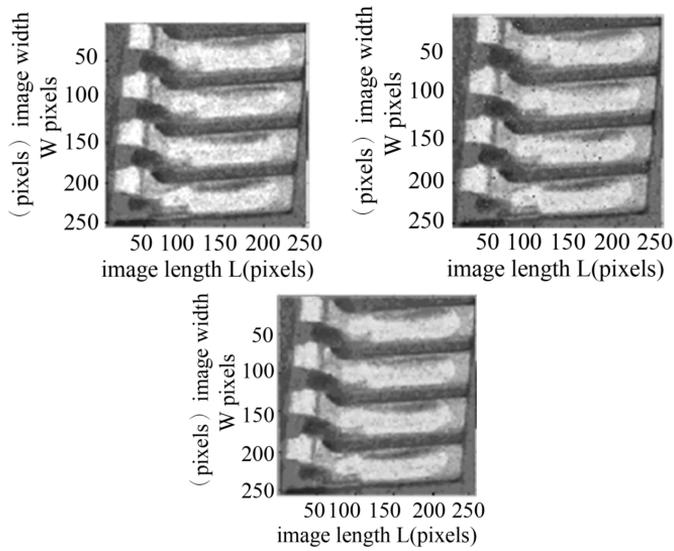


Fig. 4. Uncertainty workpiece disassembly results: left part up–median filter, right part up–skewness filter, bottom part–WPM method

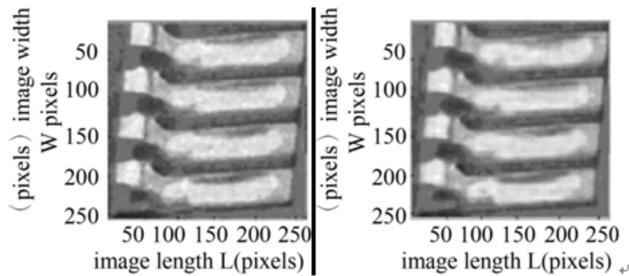


Fig. 5. Denoising results of uncertainty workpiece solder joint image: left part–AWPWM (2 layers decomposition,  $3 \times 3$ ), right part–AWPWM (2 layers decomposition,  $5 \times 5$ )

## 6. Conclusion

(1) Using the skew filter, the characteristics of the decomposition of the high-frequency part and part of the low-frequency part of the appropriate filtering, is conducive to improving the uncertainty of workpiece solder joint image denoising effect.

(2) The improved adaptive feature threshold method is used to deal with the threshold of the characteristic coefficient, which is more reasonable than the traditional feature threshold denoising method, which can effectively remove the mixed noise in the workpiece joint image.

(3) Using the median filter on the uncertainty of the workpiece solder joint image smooth processing, while protecting the edge of the image information.

## References

- [1] D. J. ZHOU, C. Y. HUANG, Z. H. WU, C. Q. LI: *SMT solder joints quality assurance based on solder joints virtual evolving technology*. Computer Integrated Manufacturing Systems 12 (2006), No. 08, 1267–1272.
- [2] N. K. S. LEE, G. YU, A. JONEJA, D. CEGLAREK: *The modeling and analysis of a butting assembly in the presence of workpiece surface roughness and part dimensional error*. International Journal of Advanced Manufacturing Technology 31 (2006), Nos. 5–6, 528–538.
- [3] L. BRECHET, M. F. LUCAS, C. DONCARLI, D. FARINA: *Compression of biomedical signals with mother wavelet optimization and best-basis wavelet packet selection*. IEEE Transactions on Biomedical Engineering 54 (2007), No. 12, 2186–2192.
- [4] W. G. ZHANG, F. LIU, X. B. GAO, L. C. JIAO: *An image denoising algorithm using adaptive multiscale products thresholding*. Journal of Electronics & Information Technology 31 (2009), No. 8, 1779–1785.
- [5] S. KOC, E. ERCELEBI: *Image enhancement by lifting-based wavelet domain E-median filter*. IEEE Signal Processing and Communications Applications, 11–13 June 2007, Eskisehir, Turkey, IEEE Conference Publications (2007), 1–4.
- [6] J., SILVA, S. S. NARAYANAN: *Discriminative wavelet packet filter bank selection for pattern recognition*. IEEE Transactions on Signal Processing 57 (2009), No. 5, 1796–1810.
- [7] X. ZHA, R. FU, Z. DAI, B. LIU: *Noise reduction in interferograms using the wavelet packet transform and wiener filtering*. IEEE Geoscience and Remote Sensing Letters 5 (2008), No. 3, 404–408.
- [8] X. J. ZHA, R. S. FU, Z. Y. DAI, B. LIU, Z. G. SHAO, T. X. XUE: *The influence on SAR interferograms noise reduction due to the selection of wavelet base function*. Remote Sensing Information 21 (2008), No. 2, 17–21.
- [9] Y. KOPSINIS, S. MCLAUGHLIN: *Development of EMD-based denoising methods inspired by wavelet thresholding*. IEEE Transactions on Signal Processing 57 (2009), No. 4, 1351–1362.
- [10] J. HAN, M. VAN DER BAAN: *Microseismic and seismic denoising via ensemble empirical mode decomposition and adaptive thresholding*. Geophysics 80, (2015), No. 6, KS69–KS80.

# Expression of moving object group movement in large scale traffic network<sup>1</sup>

YAN ZHENGSHU<sup>2</sup>

**Abstract.** In order to express the traffic scenes composed of crisscross road networks, large scale moving objects (such as the moving object) and their movement, a kind of expression technology is proposed for the movement of moving object group in large scale traffic network. First, a kind of hierarchical road network logic model was designed. According to the input vector data of road network lane line, the intricate traffic road network was described from geometry, topology and other hierarchies by logic modeling. Then, the lower computational efficiency of moving object group movement based on the microscopic method of individuals in the existing moving object group movement expression method, and the complex traffic phenomena cannot be expressed through the method based on small scale flow. Thus, a group of novel small scale flow equation was proposed to describe the movement of the moving objects in road network, which can offer an organic combination with the small scale flow model and the model of lane changing behavior of moving objects, as well as describe various complex traffic phenomena realistically. The experimental results show that, on the premise of achieving simulation in detail about moving object movement in large scale traffic network, the computational efficiency of the method in this paper has the same order of magnitudes with the general small scale flow model. In practical application, the virtual reality model of the real traffic road network was built, and the movement of moving object group was expressed on it. So the validity of this method was further verified.

**Key words.** Moving object group, large scale traffic network, hierarchical logic model, lane changing behavior of moving objects, small scale flow model, virtual reality.

## 1. Introduction

Transportation is one of the four aspects involved in human daily life of 'clothes, food, shelter and travel'. It is a significant component of human society. In order to exhibit the reality world in the spanning space of the computer, the traffic road network was constructed in virtual environment and the moving object and its

---

<sup>1</sup>This work was support by the teaching model reform of virtual reality technology "VR +flipped classroom" of Zhejiang Province Department of education research project (No. Y201636001).

<sup>2</sup>Information School, Ningbo City College of Vocational Technology, Zhejiang, 315100, China

movement in road network was expressed. Both of them are the essential steps in constructing virtual world [1–3].

The moving object movement in reality life has the following outstanding features: first, the structures of traffic road networks are complicated. All kinds of level crossings and various overpass routes are crisscross and intricate. Then the scale of the moving objects is huge. Therefore, how to realize the authentic simulation of moving object group movement in large scale traffic network is the significant problem in virtual reality. It is an effective technological method in the design, analysis and assessment of emergency evacuation plans and urban traffic projects as well. Its meaning and application prospects can be concluded in the following aspects:

- 1) Realize the rehearsing traffic simulation, and reinforce the ability to deal with special circumstances and emergencies (such as the large numbers of people evacuation in the Fukushima nuclear crisis).

- 2) Guide the design, assessment, programming of the traffic and so on.

- 3) Users experience in the enhanced virtual scene.

With the rapid development of virtual reality technology, the conceptions of virtual city, virtual tourism and others are widely proposed, integrating the distinct traffic simulation into them, which can further deepen the users' immersive feeling. The small scale simulation method is based on the model of fluid. The changing of traffic movement in the road is simulated from the small scale aspect like density and velocity [4–6]. The computation time of simulation has no relationship with the number of moving objects, which is mainly used in the simulation expression of moving object group movement. The TRANSYT system exploited by TRL in Britain and KRONOS software system exploited by University of Minnesota in America are using this kind of method. As a tool of traffic evaluation, programming and design, these system mainly study the roughly movement trend of the traffic flow. In order to enhance the computational efficiency, the model simplifies or even overleaps the simulation of individual moving objects movement behavior. It is unable to realistically describe the movement behavior of moving objects in various intricacy traffic scenes (such as various road changing behaviors), thereby reduces the veracity and authenticity of the simulation results.

In addition, the road network is the carrier of traffic flow, as well as the considerable basis data in traffic expression. The road network data in traffic expression is the mathematical abstract and standardized description of reality road network. Thus the road network which can be identified by computer is constructed. The road network data are usually required to describe the geometry information and logic information of the road network. In the expression computation of moving object group movement, the mentioned method also need to take road network data for basis. However, the present method has higher requirement for the input road network. It can first be reflected from larger amounts of input data quantity, for instance, in the plane intersections, the traffic phase is required to import or manually set to realize the detachment of conflicts routes. It has large amounts of workload and easy to generate mistakes. Then, when the model is applied to the expression of moving object group movement, the solution process of the position coordinate and

direction angle of moving objects needs to seek the road network data and process approximation calculation of numerical values. It costs lots of time and influences the efficiency of large scale expression computational.

To solve the mentioned problem, we proposed a kind of method to express the moving object group movement in large scale traffic network. First, aiming at the road network data and its modeling, we proposed a kind of road network hierarchy logic model. We conducted topology generation and hierarchy logic organization for the imported lane line geometry data, and described intricate traffic road network from different hierarchies. Then, we proposed a group of novel small scale flow equation to describe the movement of moving objects in the road network, in accordance with the movement expression computation of moving object group. This can offer an organic combination with the lane changing behavior of moving objects model, as well as obtain the movement condition of moving objects by optimized solution.

## 2. Related work

According to the degrees of the described details, the modeling of moving objects movement can be mainly divided into microscopic model and small scale model. The microscopic model describes the movement behavior and reaction of moving objects from individual moving objects. Car-following model and cellular automaton model are the two kinds of common used microscopic models. In the car-following model, according to the condition of the front car in current lane, its acceleration limit as well as the road curvature, the driver can ensure the accelerated and deceleration speed of the moving objects [7]. The basic thought of cellular automaton model is to disperse all the continuous variables [8]. The microscopic model can describe the individual behavior of moving objects in detail and combine with the model of lane changing behavior of moving objects. In this model, we can get a better traffic expression simulation result to show various complex traffic phenomena, traffic accidents, such as the road congestion, the intersection of lots of vehicle flows and so on. However, the computational efficiency of microscopic model has relationship with the scale of moving objects. Facing with the moving object group expression in large scale traffic network, its computational efficiency is relatively low.

The small scale model are usually called continuous flow model. This kind of model is mainly used in the traffic simulation and analysis of moving object group. Now, there have been a lot of classical small scale flow models. The anisotropy model adopts a mass conservation equation and accelerating equation to describe the traffic flow dynamics [9, 4]. The dispersed LWR equation is used to describe the traffic flow in the cell transmission model [10]. The other classical traffic flow model is grid flow model. Being is by optimal velocity model, Nagatani first proposed the grid dynamics model based on traffic flow [11]. The driver can react according to the local density ahead of the current lane. The movement of moving objects was calculated as the entire kinetic fluid. If these models are applied to the movement expression of moving object group, we are unable to know the detail information like whether the moving objects can change road or not and how to change road. Their precision is limited and they cannot express various traffic accidents.

Though the microscopic model can describe the movement behaviors of moving objects in detail, its computational efficiency is relatively low [6]. The small scale model can be used in the simulation of moving object group movement. However, the simulation accuracy is too rough, and the veracity and the visualization effect of the simulation results are limited, especially the incapable of simulating the various complex traffic phenomena and traffic accidents. Therefore, the modeling and expression of the moving objects movement in large scale traffic network are facing huge challenge.

### **3. The logic model of the moving object group movement expressions**

Now most of the traffic simulation software used the microscopic model based on individuals to express the movement behavior of moving objects. This kind of model has low computational efficiency, which is hard to be applied in the movement expression of moving object group. The small scale method based on sequence flow is used to simulate the entire movement trend of the traffic flow from the aspects of entire vehicle flow. This method is very suit for the expression of moving object group movement in large scale road network. However, the present small scale flow model is limited in the rough scale to simulate the movement behavior of moving objects. It excessively simplifies the movement details of moving objects, especially ignores the road change behavior and process of the moving objects, which cannot realistically exhibit all kinds of complex traffic phenomena. In this letter, we modified the classical small scale flow model - co-pilot grid flow model [5]. A group of novel small scale flow equation was proposed to conveniently couple with and solve the lane changing behavior of moving objects.

#### ***3.1. The road network logic model***

As mentioned above, considering the concrete problems in expression such as path programming, coordinate setting, movement smoothness and conflict avoidance, the expression of moving object group movement has higher requirement to the road network information. However, these methods have inadequate information, or the lower inquiring and computational efficiency. Therefore, we proposed the road network hierarchies logic model shown as the figure. From bottom to top, the model can be divided into lane layer, road segment layer and the path layer. The lane layer, only the axle wire Lane in lane is included, is the basis for the entire road network logic model. The road segment layer and the path layer describe the logic relationship of road network element Lane, including the section Link based on the definition of Lane, the Segment based on the definition of Link sets, the Connector represents the connection in sections, the Intersection represents the topology connection in sections and the Road represents the location dependent relationship.

The lane layer means the generalized lane line definition in this paper-Lane. In order to describe the moving orbit of moving objects, the Lane not only includes the lane axle wire with reality meaning, but also includes the curve describing the travel

orbit of moving objects in the crossing, which is called virtual Lane (as shown in Fig. 1). The vector data information of Lane can be described with  $P$ . Here,  $P$  is the directed polyline constituted by point range, and meets the requirement if certain 'smoothness'. The nature of broken line has no smoothness. The smoothness means that the point ranges of this broken line should approach to the smooth curve fitted by point range as much as possible. Then, except for the crossing area, the average distance has the consistent Lane trend in threshold values and the roughly equal length. Finally, the out-degree and in-degree of all the inner nodes, respectively, are equal to 1.

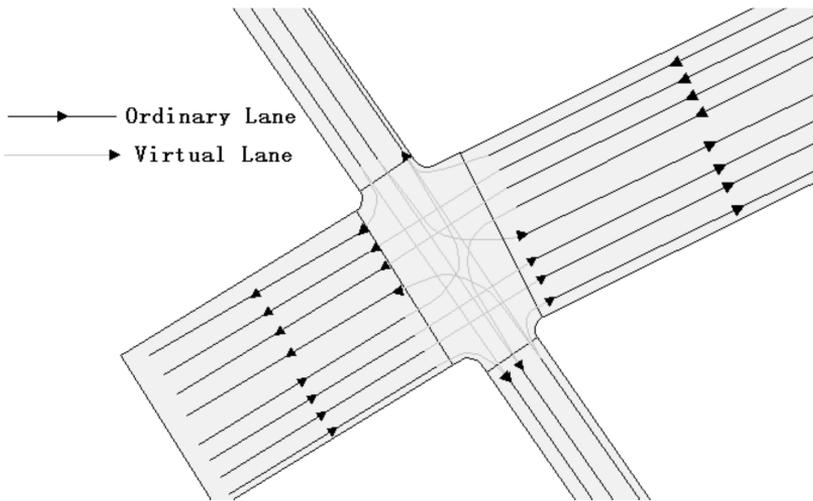


Fig. 1. Schematic diagram of Lane

Based on the mentioned logic description, the road network model can not only describe any complicated road network structures, including plane intersection and interchange, but also realize the phase distribution at the intersection by the same Intersection relates to the intersection of Lane information automatic. Thus, it can realize the automatic conflict separation of lines with different direction. All the logical information in the model can be automatically generated by the topological analysis of the Lane data, and the specific generating process is shown in Fig. 2.

### 3.2. *The small scale flow model*

In the movement process of the moving object, the driver can not only get the density and velocity information of the current lane, but also obtain the density velocity information of the left and right side of the lane. Based on this, this paper

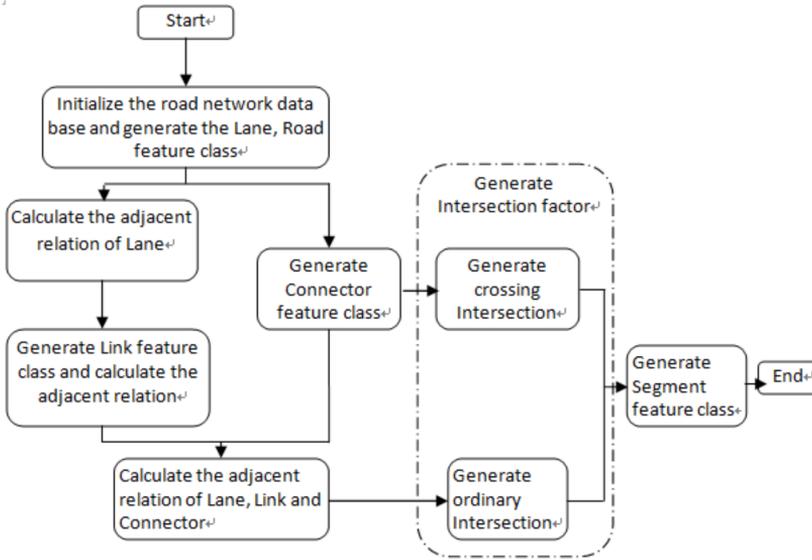


Fig. 2. Road network logic data generated flow chart

builds the following interactive cooperative driving grid flow model:

$$\begin{cases} \partial_t \rho + \nabla(\rho v) = 0, \\ \rho_{i,j}(t + \Delta t) v_{i,j}(t + \Delta t) = \\ \rho_0 \sum_{m=0}^2 [f_{2-m}(H_{i,j}(t)) V(\rho_{i+1,j-1+m}(t), \dots, \rho_{i+N,j-1+m}(t), \rho_j^*)] \end{cases}$$

The difference between the above equations and the classical co-pilot grid flow model equation [5] is that the interaction term is added.

$$\sum_{m=0}^2 [f_{2-m}(H_{i,j}(t)) V(\rho_{i+1,j-1+m}(t), \dots, \rho_{i+N,j-1+m}(t), \rho_j^*)]$$

This equation is used to describe the effects of the moving object in current lane, or in front of the adjacent lanes on the moving object in current position. Here,  $(i, j)$  represents the  $(i, j)$  grid directed along and perpendicular to the lane.

Here,

$$f_0(*) = \left\lfloor \frac{*+1}{2} \right\rfloor, \quad f_1(*) = ||*| - 1|, \quad f_2(*) = \left\lfloor \frac{|*-1|}{2} \right\rfloor,$$

$\rho_j^*$  means the car density in the  $(i, j)$  position in the current lane and the balanced place.  $H_{i,j}(t)$  is the direction of the car flow. When the car flow is directed to the left lane,  $H_{i,j}(t) = -1$ . When the car flow is directed to the right lane,  $H_{i,j}(t) = 1$ .

In other circumstance,  $H_{i,j}(t) = 0$ .

By using the finite difference method, the mentioned differential difference equations are fully discretized and the numerical stability analysis is performed. The equation has the only stable solution as long as determines  $H_{i,j}(t)$ . With the application of the behavior model, it is possible to determine whether the lane changing is feasible, so as to determine  $H_{i,j}(t)$ , and solve the movement state of the moving object.

### 4. Experimental results

Based on the mentioned method, we perform a moving object group movement expression system for large scale traffic networks, and carry out the practical application and validity test for the method in the urban road network. The efficiency experiment in this article is implemented on a workstation configured as follows: Core 8 Xeon (R) E31240 3.4 GHz CPU, 4.0 GBRAM.

#### 4.1. Results and analysis of hierarchical road network model

First, the road network model proposed in this paper is compared with Wilkie’s road network model [2] from the view of logic generation. Table 1 exhibits the comparison of the automation degree generated by logical data. It shows that the Wilkie’s model requires additional manual interaction input for the intersection data, and needs to further install the traffic phase distribution for the conflict traffic routes as well, which possesses heavy workload and is prone to generate errors. The required input data in this model is just lane line data. Other data, including intersection data and traffic phase data describing the conflict relationship, can be generated automatically, reducing the input of two kinds of data. This not only reduces the workload of manual input data, but also avoids the introduction of artificial errors. In addition, the definition of intersections in this paper is not directed to specific intersections, instead, an Intersection is automatically generated according to the connecting and intersecting topologies. Therefore, any complex intersection models generated by the model can compare with the existing model, which has remarkable advantages in the data generated automation degree.

Table 1. Comparison of automation degree generated by logical data

	Intersection	Adjacency relationship	Connection relationship	Conflict relationship
Wilkie method	Manually interactively generated	Automatically generated	Automatically generated	Manually interactively generated
Suggested method	Automatically generated	Automatically generated	Automatically generated	Automatically generated

Second, the results of road network logical data automatic modeling and efficiency test are represented. The radius of 5 km of a city urban road, with a total of about 8,000 lanes was processed with the road network logic modeling. The logical modeling is based on the input Lane data. The results show that all the actual adjacent and connected lanes in this area are completely correct in this logical data. All the intersections, including all kinds of intersections, ring road junctions, various three-dimensional intersections and other conflict lines are also automatically generated by the intersection of the traffic phase to achieve the separation of the conflict, which proves the validity of the logical model of road network. Fig. 3 (left) and Fig. 3 (right) separately show the vector data of road network local roadway and road network local hierarchical relationship. The hierarchical relationship data in the schematic diagram uses different colors distinguish the different Links. The blue polygons mean the terminal information of Lane involved in the Intersections.

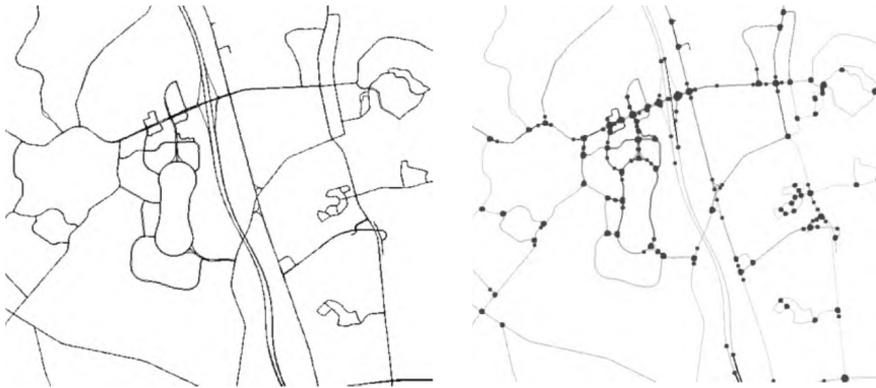


Fig. 3. Logical modeling results for a real regional road network

#### ***4.2. Result and analysis of movement expression of moving object group***

The calculation efficiency and effects of moving objects in real-time expression were tested with the mentioned method. In the given road network scale, the computational efficiency of the system is almost irrelevant to the number of moving objects (Fig. 4), and has the same order of magnitudes with the classical small-scale flow model [5] (one-dimensional co-driver grid flow model). It shows that the proposed method can be used for moving object group movement representation. On the machine with the mentioned configurations, this method can express the movement of moving objects in the road network scale with total mileage of 600 km. The scale of the road network in this experiment is as follows: the total length of the lane line is about 600 km, and the uneven road segment accounts for about 20% of the total road network. Red line refers to the classical co-driver grid flow model [5] simulation results.

To further verify the effectiveness of the method, we selected traffic lane changing behavior of instance data with this method of moving objects occur frequently

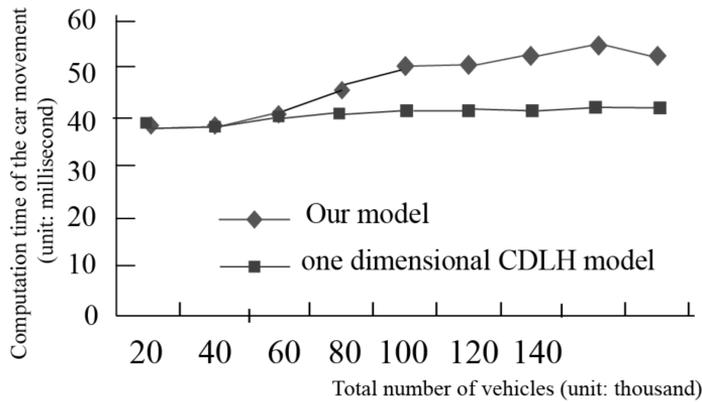


Fig. 4. Relationship between the number of moving objects and the expressed computation time

compared. We compared it with the traffic instance data which were published on the Next Generation Simulation (NGSIM) website. NGSIM was a kind of website initiated by the Federal Highway Administration of Transportation (FHWA). We used an up-ramp segment on the Interstate 80 road of the city of San Francisco on the site (as shown by the two thick straight-line markings in Fig. 5). We track all lane change objects within 1 minute of the road segment. The section of the highway is on the ramp section, moving objects from the ramp into the main road after the frequent lane change behavior.

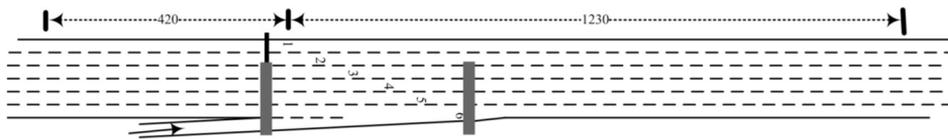


Fig. 5. Road conditions of instance data acquisition area

The measured data in the section of the period of time occurred in a total of 12 lane behavior. A car from a lane through the lateral movement (perpendicular to the lane direction) into the adjacent lane is called once lane change behavior. The start time of the lane in the measured data is the time when the moving object starts to shift sideways to the target lane. The end time is the time when the lane of the lane reaches the target lane and the moving object is oriented in parallel to the lane direction. The duration of the lane is the difference between the lane end time and the lane change start time.

Next, we compare the duration of the lane simulation and the length of lane change in the measured data. Fig. 6 represents the comparison diagram of the actual time length and the analog time length for 12 times of lane change behaviors in the comparison period. It shows that in the vast majority of lane changes, the model simulation results and the measured data have roughly equivalent lane length. Further analysis shows that the average error of the lane length and the measured results is about 1.1 s.

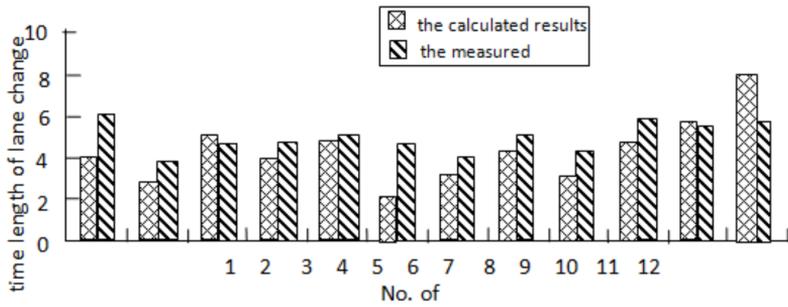


Fig. 6. Comparison of the duration for lane change

## 5. Conclusion

With the development of virtual reality technology, the research of virtual city has been greatly developed. The high-detail, high-fidelity mobile object motion simulation technology has integrated into the city simulation. It plays an important role in improving the reliability of simulation and enhancing the visual experience, as well as traffic design and traffic planning. The existing moving object motion simulation method can not only describe the general trend of traffic flow and cannot simulate the interaction behavior of moving objects between lanes, but also simulate the movement of urban moving objects, which reduces the realism and credibility of the simulation results. In this paper, a method is proposed for motion representation of moving objects within a large scale transportation network. And the technical details of the method are expounded from the aspects of road network logic modeling, moving object group motion expression calculation and so on. The experimental results and application examples show that this method has the following significant advantages: First of all, the method of road network data input requirements are low, only need to enter the lane line vector data to meet the needs of animation expression; secondly, the efficiency of the real - time simulation method is almost independent of the number of moving objects, making the method applied to the simulation of moving object group motion; finally, the method can be described in detail when moving objects and how to change lanes and other acts, which improves the simulation of fidelity and accuracy.

## References

- [1] J. SHEN, X. JIN: *Detailed traffic animation for urban road networks*. Graphical Models 74 (2012), No. 5, 265–282.
- [2] Q. CHAO, J. SHEN, X. JIN: *Video-based personalized traffic learning*. Graphical Models 75 (2013), No. 6, 305–317.
- [3] M. TREIBER, A. HENNECKE, D. HELBING: *Congested traffic states in empirical observations and microscopic simulations*. Physical Review E 62 (2000), No. 2, 1805–1824.
- [4] H. M. ZHANG: *A non-equilibrium traffic model devoid of gas-like behavior*. Transportation Research Part B: Methodological 36 (2002), No. 3, 275–290.

- [5] S. SHARMA: *Effect of driver's anticipation in a new two-lane lattice model with the consideration of optimal current difference*. *Nonlinear Dynamics* 81 (2015), Nos. 1–2, 991–1003.
- [6] M. M. PEDERSEN, P. T. RUHOFF: *Entry ramps in the Nagel-Schreckenberg model*. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 65 (2002), No. 5, paper 056705.
- [7] L. Y. DONG, Q. X. MENG: *Effect of relative velocity on the optimal velocity model*. *Journal of Shanghai University (English Edition)* 9 (2005), No. 4, 283–285.
- [8] V. I. SHVETSOV: *Mathematical modeling of traffic flows*. *Automation and Remote Control* 64 (2003), No. 11, 1651–1689.
- [9] A. B. KISELEV, A. V. KOKOREVA, V. F. NIKITIN, N. N. SMIRNOV: *Mathematical modelling of traffic flows on controlled roads*. *Journal of Applied Mathematics and Mechanics* 68 (2004), No. 6, 933–939.
- [10] C. F. DAGANZO: *The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory*. *Transportation Research Part B: Methodological* 28, (1994), No. 4, 269–287.
- [11] T. NAGATANI: *TDGL and MKdV equations for jamming transition in the lattice models of traffic*. *Physica A: Statistical Mechanics and its Applications* 264 (1999), Nos. 1 to 3, 581–592.

Received June 29, 2017



# High-dimensional data express model based on tensor<sup>1</sup>

ZHANG JING<sup>2,4</sup>, GUO XINCHANG<sup>3</sup>

**Abstract.** Taking into account the requirements of four-dimensional field data with high-dimension analysis and expression, the model of high-dimensional data structure analysis and dynamic expression based on tensor was constructed, the tensor definition, basic operators and tensor decomposition method were briefly introduced, and the process of multi-dimensional characteristic analysis based on tensor decomposition was given in this paper. Then, according to the multi-dimensional fusion feature, the high-dimensional data was organized and expressed, and the unified organization and storage method of high-dimensional data was designed. At the same time, by using the tensor decomposition method, the analysis and dynamic reconfiguration of high-dimensional data structure characteristics in different dimensions were realized, and the multi-dimensional data with high-dimensional analysis model and feature-driven high-dimensional linkage data based on tensor were established to exhibit the strategy. In addition, the experiments were verified by the grid data of the India ocean satellite RUEB, so that the multi-dimensional perspective, subset extraction, contour surface rendering and spatial data representation function based on tensor were achieved. Finally, the analysis and extraction of FARP event time type and space type were succeeded by using tensor decomposition, and the data representation driven with time, longitude and latitude coefficient was realized. The examples verified that this method can exhibit the spatial and temporal patterns and the dynamic evolution characteristics of FARP event better, and the multi-dimensional perspective in the evolution process of FARP spatial can be realized.

**Key words.** Tensor decomposition, the four dimensional field, feature-driven, data representation.

## 1. Introduction

The analysis and expression of high-dimensional data features are the main contents of high-dimensional data expression. Due to the continuous development of

---

<sup>1</sup>This work was supported by the Sponsored by a project of National Social Funds named “a Research on China-Japan Maritime Crisis Monitoring and Warning” (No. 16BGJ028).

<sup>2</sup>Basic courses Department, Qingdao Ocean Shipping Mariners College, Qingdao, Shandong, 266071, China

<sup>3</sup>Ocean University of China, Qingdao, Shandong, 266071, China

<sup>4</sup>Corresponding author

the network observation, a large number of remote sensing images, satellite and radar images have been obtained. All the images have the characteristics of large scale and huge amount [1–2]. Due to the large scale and high-dimensional data, the significant structure of the data can be extracted, the direct data observation is the identification approach of the solutions and rules, and the high-dimensional statistical method is generally used. This method is based on matrix analysis and statistics as the theoretical basis, which can be widely used in the analysis of space and time. The disunity of the multi-dimensional operation may lead to the asymmetry in time-space dimension. General geoscience shorthand software uses the mentioned method to express the data in three-dimensional space, instead of utilizing the underlying mathematical basis for multidimensional data expression. The general high-dimensional dynamic and analytic expression method is different, which can generate an impact on the current analysis of the geographical space and time. According to the different requirements of the data analysis, new support data can be found to analyze the space and time better, which is a useful solution for high-dimensional data analysis. The data analysis is used to extend the traditional vector and data. The analytic method of tensor decomposition is used to analysis the data center, which is conducive to ensure the form of the structure and reveal the relationship among different data dimensions, so that more sophisticated data analysis method can be proposed. This method has been widely used in plenty of fields. The general analysis methods are analyzed on the basis of tensor analysis, however, there is no the comprehensive analysis of tensor analysis in the aspects of data organization and characteristic analysis. In this paper, the tensor structure was taken as the basis to discuss the data organization and storage of the tensor [3–5]. By using this decomposition method, the field data was analyzed and the multi-dimensional space-time form was represented. The results show that this study is favorable for the validation of satellite data in the India Ocean.

## 2. The multi-dimensional solution of high-dimensional data features based on tensor

### 2.1. *Tensor and tensor operation*

Tensor is an expansion of the vector quantity matrix, and is a kind of expression model for multi-dimensional data to express the physical properties of multiple forms, so it can be used to express and deal with the data of multi-dimensional points. An  $M$ -step tensor can be written as  $A \in R^{L_1 \times L_2 \times \dots \times L_M}$ . Here,  $L_i$  means the dimension of the  $i$ th order. A lot of operational symbols can be used in the tensor product and Kronecker product, which can be applied in different algebraic tensors and the earthquakes. The symbolic significance in different conditions are identical. Generally, a bilinear calculation is required to conduct and expressed with  $\otimes$ . The main difference is the identical restructuring method and structure reset of the tensors [12]. Assuming that  $U$  and  $V$  are two tensors, the tensor product can be expressed as

$$U \otimes V = \begin{bmatrix} u_{11}V & u_{12}V & L \\ u_{21}V & u_{22}V & \\ M & \ddots & O \end{bmatrix} = \begin{bmatrix} u_{11}u_{11} & u_{11}u_{12} & L & u_{12}u_{11} & u_{12}u_{12} \\ u_{11}u_{21} & u_{11}u_{22} & & u_{12}u_{21} & u_{12}u_{22} \\ M & & & & \\ u_{21}u_{11} & u_{21}u_{12} & O & & \\ u_{21}u_{21} & u_{21}u_{22} & & & \\ M & & & & \end{bmatrix}. \quad (1)$$

## 2.2. *QRB tensor decomposition*

Tensor decomposition is mainly extended by the method of component analysis, which can make data analysis of high order tensor through the low order tensor. The singular value decomposition method is used in this method, and the tensors are calculated by least square method. The tensors can be separated through the high dimensional space, and then the reliability test can be carried out.

$$\begin{aligned} \sigma_i &= \max_{\substack{\|\psi\|_s=1 \\ \|\varphi\|_v=1 \\ \|\phi\|_t=1}} Y..(\psi \otimes \varphi \otimes \phi) \\ &= Y..(\psi_i \otimes \varphi_i \otimes \phi_i). \end{aligned} \quad (2)$$

Here,  $\sigma_i$  is the  $i$  characteristic value of the tensor as well as the peacekeeping vector. The operator can be shown as indentation operation, which can be called the accumulation of tensor or the decomposition of tensor coefficient and tensor. Then, dimensionality based on the tensor decomposition can be constructed.

$$\begin{aligned} Y..(\psi \otimes \varphi \otimes \phi) &= (Y.. \psi) .. (\varphi \otimes \phi) \\ &= (Y.. \varphi) .. (\psi \otimes \phi) \\ &= (Y.. \phi) .. (\psi \otimes \phi). \end{aligned} \quad (3)$$

For  $k$ -dimensional tensor, the tensor decomposition model can be constructed with the dimensionality  $k-1, k-2, \dots, k-n, \dots, 1$ .

## 2.3. *Multi-dimensional feature reconstruction based on tensor decomposition*

Generally, tensor decomposition is mainly to carry out the structure characteristics analysis of high-dimensional data, which is conducive to the data reconstruction for multi-dimensions. Taking the four-dimensional field data as an example, the sum of tensors can be used in this four-dimensional field data, so that the tensors can obtain the coefficient information of direction axis and the 3 directions. Then, the process matrix can be constructed, and different dimension coefficient can represent the information in practical works, so that the characteristics of spatial and temporal process in different dimensions can be obtained. Finally, the process of dimension perspective can be conducted on the basis of specific dimensions. For example, by constructing kinds of spaces, the specific combination of time and space can be dis-

tributed through different visual angles. The common dimensions are the tensor products in two-order tensors of a tensor and two dimensions, which can be used to represent the space-time evolution process. A large number of decomposition results guarantee the correlation of different dimension structures, which is beneficial to the performance of dynamic data in different dimensions, and it is also in favor of the analysis and expression of the unit structure. However, the dimensionality of data may be at the expense of partial data and information, which is good for high-dimensional data to be mapped into a low-dimensional space. Because there are some losses of the original information, the essential premise of the original data needs to be kept, so as to express the high-dimensional data in low-dimensional better.

### **3. Characteristic data representation of high-dimensional data**

#### ***3.1. High-dimensional data analysis and data expression process based on tensor***

Figure 1 shows the decomposition conditions of tensor data between the two-order tensors. This process can be used in the management and analysis of the original data. The data have better decomposition, so that the data coefficients can be obtained, which is not only beneficial for the data reconstruction to obtain the data with physical meaning, but also good for the analysis of space-time and import parameters of external time to find the data meaning in different dimensions, so that the dimension fusion of the tensor data organization can be conducted, and the different fusion results can be reconstructed. Finally, the observation and design pattern can be gained to express the data for simple structures. Each of the flow chart is compatible with a variety of multi-dimensional data, which can not only prompt a variety of data organization and the unified expression of the process, but also has a variety of dimension and analysis functions.

#### ***3.2. High-dimensional data organization based on tensor***

Based on the increment, the current four-dimensional data has been constructed by the multi-dimensional array to make the object retrieval and object reorganization and analysis process. This is conducive to the expression of four-dimensional data object, analysis of the history value, and has significant meanings to the construction of geographic models. It can carry on the reorganization of the market according to the storage mode. It can analyze and express the data by using tensors, and the multidimensional spatial data can be analyzed. It is possible to gain a better effect, but there may be more of the value of irregular boundary and missing, so the estimation of all kinds of parameter is required. Generally, based on the stereoscopic space-time model, a variety of model data can be stored through the positive structure, including all of the core components and the tensor storage structure. Therefore, the thought data and the attribute of tensors can be analyzed,

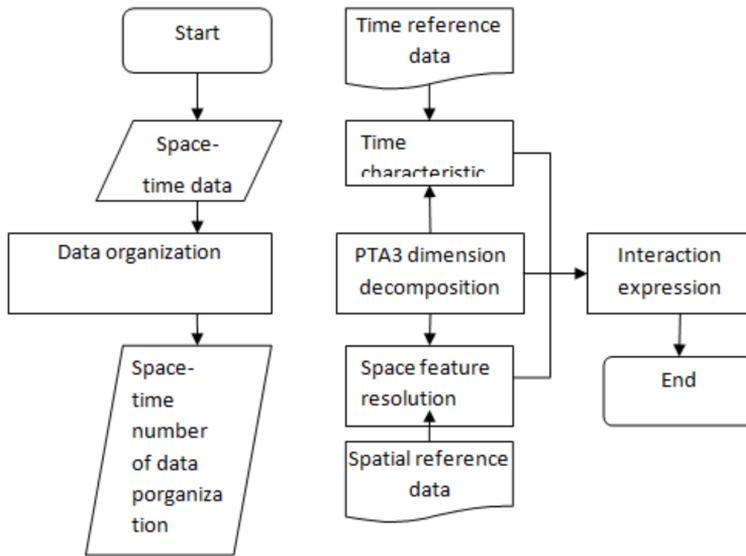


Fig. 1. High-dimensional data analysis and dynamic expression process

and the market can be described. According to the methods of the last operation tensor and a variety of data operation plans, the decomposition of tensor and tensor data analysis, multi-dimensional tensors perspective in the same group of tensors of these components, the goal of object modeling can be achieved, which is favor of the retrieval of capital market data. Based on this, kinds of strategies can be constructed, so that the structural characteristics and perspective target can be obtained easily, which is conducive to the management and operation of the relevant data, reorganization and calculation of all the data, and the analysis and function expansion of the follow-up data.

### 3.3. Data representation of feature driven

The goal of geographic data is to reveal all kinds of news in organizations, which is conducive to analyze the indicators, and is helpful to achieve the current data expression. The system is good for the interpretation and extraction of the data through the method of dimension reduction. This data obtained should be in the method of matrix operation, which has a great effect on the high-dimensional data. The data analysis should be linkage and the tensor data should have various fusion characteristics, so as to promote the combination and dispatch among different data. According to the integrality and reconstruction of the tensor decomposition, there are internal linkage and perspective. Since the characteristics of this idea can be obtained, as shown in the scheme, the tensor form organization data can perform the dimension reorganization and penetration, which is conducive to the integration of dimensions. By adding the geography index, the signal can be obtained from high level data. According to the different analytical results, different expression

strategies can be developed. According to the characteristics of data and function, all the result data can be used to carry out the time and space data election. Then the observer pattern is set to express the dynamic state and characteristics in different dimensions, so as to provide a better environment for data analysis, which is conducive to the analysis to the solution of the geoscience problem.

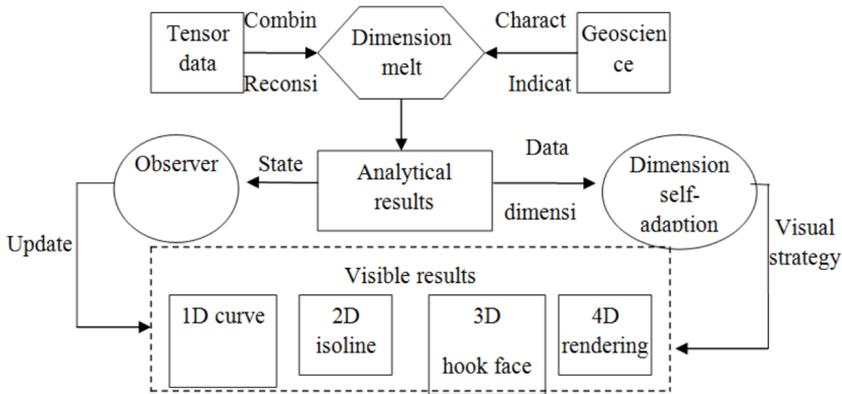


Fig. 2. Scheme of data representation for feature driven

## 4. Experiment and verification

### 4.1. Original data

The BQHTP not only can release the mixture of T/P and Jason4 in India ocean grid data, but also can perform all kinds of experimental verifications. The original data is the NetCDF format with the spatial resolution of  $1.4 \times 1.4$ , the time scale being from 1995 to 2003, and the spatial range being from 15 degrees north latitude to 15 degrees south latitude, 150 degrees east longitude to 150 degrees west longitude. In addition, there are 92 time slices, 21 time dimensions and 421 longitude intervals, which are compared with the standard indicators. This system is conducive to the decomposition of the tensor and representation of the data, and the import and convert of data are good for the unified investigation of various tensors, so as to draw all kinds of VTK pictures. Figure 3 shows the perspective four-dimensional field data along the time direction.

### 4.2. Multi-dimensional analysis of data characteristics

Since the sea surface is treated average monthly processing, abnormal high-dimensional field data may occur in the sea surface. Each of the tensor has its own contribution rate to the data, so that the fusion refactoring of the solved results can be performed, and the time change graphs can be obtained. The MEI index can be introduced to express the events' degree, which is conducive to the final analysis



Fig. 3. Data representation based on tensor data

of the characteristics. As shown in Fig. 4, the main tensor can decompose the space and time, so as to obtain the time factor, which has better correlation. Based on this effect, the general FARP evolution has differences and shows the obvious difference of latitude in space. Through a comprehensive comparison of reconstruction at different time, it can be found that the sea space changes with time; and the two images indicate the sea location, the amplitude and distribution, respectively.

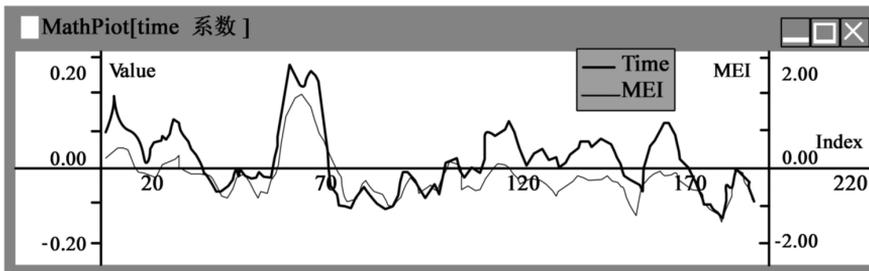


Fig. 4. Space and time decomposition of the main tensor

#### 4.3. Dynamic data representation with feature driven

The three coefficients can be obtained by using tensor decomposition and the data can be expressed. This method is favorable of the perspective different vectors and different dimensions. Through the decomposition of the vector data, each data has a unique dimension basic training. The two-dimensional data can support a two-order tensor to express this three-dimensional data, and both of the two have relationship through the data sequence. When various sequence data changes, the temperature of the tensor will appear a variety of changes. The decomposition can be constructed through the varying process in latitude of the sea level changing and

the specific perspective, which is favorable to construct and discuss the parameters structure, so as to implement the dynamic data expression. Based on this, the exploration and analysis can be carried out.

## 5. Conclusion

The expression of multidimensional data is an important direction for the development of the subject. Through the analysis of the tensor structure, various data models can be established, which is conducive to the establishment of a unified expression and calculation framework. Taking into account the characteristics of high dimension, new solutions can be proposed to analyze the dynamic expression and analysis method. Through the study in this paper, it can be seen that the tensor structure is good for the multi-dimensional statistics and calculation, the tensor structure established is conducive to the plenty decomposition and the decomposition of the feature data, and the data expression can also be realized. At the same time, according to India ocean satellite data, the original pattern and dynamic characteristics can be reproduced. The results of this paper are conducive to a better use of the tensor structure extension, which is supportive to the intrinsic computer. Based on the construction method of the tensor, the significant of the high-dimensional data analysis and expression can be analyzed. In addition, the tensor analysis is unceasing developing and progressing, which is conducive to go deep into the comprehensive use of related fields. In this research, the related demands were analyzed on the basis of the existing problems, and the calculation process was optimized by using tensor operation characteristics, so as to analyze and deconstruct the irregular data through various multi-dimensional data representation methods, instead of the data values.

## References

- [1] J. LUO, C. ZHOU, Z. SEHN, X. JANG, C. QIAO, Q. CHEN, D. MING: *Theoretic and methodological review on sensor information tupu computation*. Journal of Geo-Information Science 11 (2009), No. 05, 664–669.
- [2] G. ANDRIENKO, N. ANDRIENKO, U. DEMSAR, D. DRANSCH, J. DYKES, S. I. FABRIKANT, M. JERN, M. J. KRAAK, H. SCHUMANN, C. TOMINSKI: *Space, time and visual analytics*. International Journal of Geographical Information Science 24 (2010), No. 10, 1577–1600.
- [3] H. SU, Y. WEN, M. CHEN, H. TAO, J. SHEN: *VGEs-oriented multi-sourced heterogeneous spatial data integration*. Proc. Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Advanced Spatial Data Models and Analyses, 28–29 June 2008, Guangzhou, China, paper 71460D.
- [4] A. H. MONAHAN, J. C. FYFE, M. P. H. AMBAUM, D. B. STEPHENSON, G. R. NORTH: *Empirical orthogonal functions: The medium is the message*. American Meteorological Society, Journal of Climate 22 (2009), No. 24, 6501–6514.
- [5] L. ZHANG, C. YANG, D. LIU, Y. REN, X. RUI: *A web-mapping system for real-time visualization of the global terrain*. Computers & Geosciences 31 (2005), No. 3, 343–352.
- [6] J. HU, Z. MA, H. WU, M. PAN: *3D visualization of massive terrain data based on*

- grid partition*. Journal of Computer Aided Design & Computer Graphics 16 (2004), No. 8, 1164–1168.
- [7] Y. XIA, R. MO, J. ZHANG, Q. TIAN, Z. CHANG: *Design process saving and reconstruction based on knowledge*. Journal of Computer Aided Design & Computer Graphics 17 (2005), No. 12, 2702–2708.
- [8] Y. G. REN, G. YU: *Research and development of the data visualization techniques*. Computer Science 31 (2004), No. 12, 92–96
- [9] A. ZELI: *The financial distress indicators trend in Italy: An analysis of medium-size enterprises*. Eurasian Economic Review 4 (2014), No. 2, 199–221.
- [10] L. D. LATHAUWER, B. D. MOOR, J. VANDEWALLE: *On the best Rank-1 and Rank-( $R_1, R_2, \dots, R_N$ ) approximation of higher-order tensors*. Siam Journal on Matrix Analysis & Applications 21, (2006), No. 4, 1324–1342.
- [11] M. BACHMAYR, W. DAHMEN: *Adaptive near-optimal rank tensor approximation for high-dimensional operator equations*. Foundations of Computational Mathematics 15 (2015), No. 4, 839–898.
- [12] D. LEIBOVICI, R. SABATIER: *A singular value decomposition of ak-way array for a principal component analysis of multiway data, PTA-k*. Linear Algebra and its Applications 269 (1998), No. 1, 307–329.
- [13] P. DRINEAS, M. V. MAHONEY: *A randomized algorithm for a tensor-based generalization of the singular value decomposition*. Linear Algebra and its Applications 420 (2007), Nos. 2–3, 553–571.

Received June 29, 2017



# Semantic analysis of English vocabulary based on random feature selection

JUFANG GONG<sup>1</sup>

**Abstract.** The false statements of English fact on the Internet have seriously affected people's effective access to information, and how to determine whether the English fact statement is semantically credible becomes an urgent problem to be solved. A kind of Multi-answer English fact Statements Verification model (MFSV) based on random feature selection for the English vocabulary is proposed in this paper. In view of the characteristics of English fact statements, this model collects the English vocabulary information that is related to the English fact statements to be verified from the Internet, and measures the random feature selection corresponding to the English fact statements. At the same time, this model takes into consideration the difference in the semantic credibility of the relevant English vocabulary information, measures the semantic credibility of the relevant English vocabulary information source from the two aspects of popularity and importance, and obtains the semantic credibility ranking of the relevant English vocabulary information. According to the random feature selection and the semantic credibility ranking, the contribution of the relevant English vocabulary information to the semantic credibility verification of the corresponding English fact statements is measured, based on which the verification of the semantic credibility of the English fact statements to be judged is realized. And a series of experiments have verified the rationality and accuracy of the semantic verification of the model.

**Key words.** English fact statements, random feature selection, semantic verification, semantic credibility ranking.

## 1. Introduction

The Internet is an important source of information for people to obtain information; however, false statements of English fact on the Internet have seriously affected people's effective access to information. Therefore, how to determine whether the English fact statement is semantically credible becomes an urgent problem to be solved. On the Internet, the English fact descriptive information is mainly expressed in the form of English Fact Statement, which only expresses the true semantics of

---

<sup>1</sup>Department of Foreign Languages of Xiangsihu College, Guangxi University for Nationalities, Nanning, Guangxi, 530008, China

the information delivered, and such English fact statement is a semantically credible English fact statement; on the contrary, the English fact statement is not credible. According to the characteristics of the English fact statements, it can be seen that for any negative English fact statement, there is always a corresponding affirmative English fact statement, and the semantic credibility verification of the negative English fact statements can be realized through the semantic a credibility verification of the corresponding affirmative English fact statement. In this paper, only the credibility verification on the affirmative English fact statements is conducted.

When determining whether an English fact statement is semantically credible, the unit of doubt in the English fact statement will be specified [1], that is, the part in the English fact statement that the users need to verify. In this case, the English fact statements can be regarded as the answer to a subject. When the subject corresponding to the English fact statements has the only correct answer, the English fact statement is the only answer to the English fact statement; on the other hand, it is called the multi-answer English fact statement. For example, in the specified English fact statement "Obama is American president", "Obama" is the unit of doubt, and then the subject of the English fact statement corresponds to: "Who is American president", and the answer to the subject is unique. Therefore, the English fact statement "Obama is American" is the only answer to the English fact statement. When the unit of doubt is designated as "American president", the corresponding subject is "What is Obama", and the subject has multiple correct answers, including "Obama is a Christian", "Obama is American president" and so on. In this case, the English fact statement is a multi-answer English fact statement.

The basic idea of the information semantic credibility verification is to realize the semantic credibility verification of the information by acquiring and analyzing the English vocabulary information related to the information to be judged. Literatures [2-3] put forward the verification method for the news information. Through obtaining the relevant information of the news information to be verified from the news website, the consistency and objectivity of the content of the relevant information and the news information to be verified is taken into consideration, so as to realize the credibility verification of the news information semantics. Literature [4] puts forward an event semantics credibility verification method. Through obtaining the relevant information to be verified from the relevant website where the event occurs, the relevance of the relevant information and the event to be verified is analyzed from the three dimensions of time, space and characters. At the same time, the semantic credibility of the information source is taken into account, to achieve the verification function of the auxiliary users for the authenticity of the event. Literature [5, 6] put forward an English fact statement semantic credibility verification system Verify that is irrelevant to the direct verification domain. Through the search engine, Verify obtains the English fact relevant information to be verified, from which the English fact that can be compared is identified; then, the relevant information of the English fact that can be compared is obtained, respectively, and the rating is conducted for the comparable English facts selected from the aspects of the English vocabulary feature of the relevant information and the source features, and so on. The one with the highest rating is the semantic credible English fact statement.

In view of the defects of the aforementioned research work, this paper puts forward a kind of new multi-answer English fact statement verification (MFSV) that is irrelevant to the new domain. The model obtains the relevant English vocabulary information of the English fact statements to be verified through the search engine. In the process of verification of the English fact statements, the model takes into account the supporting relationship between the relevant English vocabulary information and the corresponding English fact statement, as well as the difference in the semantic credibility of the relevant English vocabulary information, thus making up for the defect of the first category of verification method [7]. In addition, the process of using this model to conduct English fact statement verification, it is not necessary to specify the unit of doubt of the English fact statements, or look for and analyze the English fact statement that can be compared. Therefore, this model is also applicable to the multi-answer English fact statement semantic credibility verification, so as to make up for the limitation of the second category of the verification [8-9] in the multi-answer English fact statement verification.

## 2. Multi-answer English fact statements verification model (MFSV)

English fact statements semantic credibility verification model MFSV is shown in Fig. 1. The model consists of four modules: The relevant English vocabulary information acquisition; random feature selection measurement; English vocabulary information semantic credibility ranking; and English facts statements semantic credibility verification. The input of the model is a statement of the English fact to be verified, and the output is the result of the semantic credibility verification of the English fact statement.

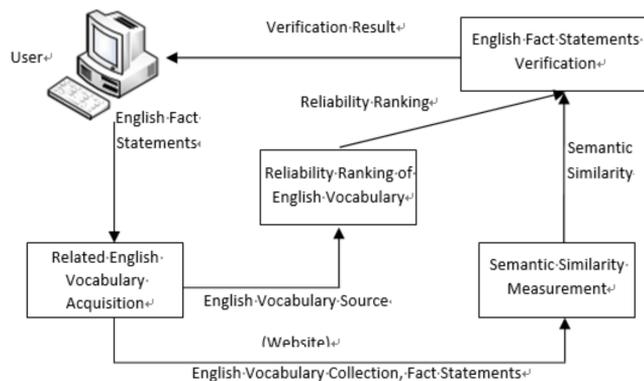


Fig. 1. Multi-answer English fact semantic credibility verification model MFSV

The related English vocabulary information acquisition module obtains the related English vocabulary information of the English fact statement to be verified. As the model is to make verification based on the English fact statement that is

irrelevant to the domain, the acquisition of the relevant English vocabulary information is completed by the search engine. The input of the module is the English fact statement to be verified. And the outcome is the relevant English vocabulary information collection corresponding to the statement. The random feature selection measurement module is used to calculate the random feature selection between the relevant English vocabulary information and the corresponding English fact statement. Firstly, the sentences that make sense for the English fact statement verification is extracted from the English vocabulary information; secondly, the similarity between the sentences extracted and the corresponding English fact statement is measured, so that the English vocabulary information and the random feature selection of the corresponding English fact statement is obtained. The input of this module is the relevant English vocabulary information and the corresponding English fact statement. The output is the English vocabulary information and the random feature selection of the corresponding statement. The semantic credibility ranking module realizes the semantic credibility ranking of the relevant English vocabulary information. In this module, through obtaining the Page rank corresponding to the relevant English vocabulary information source (website) and its position in the Alex ranking, the importance ranking and the popularity ranking of the source of the relevant information is realized. From the combination of these two ranking, the relevant English vocabulary information semantic credibility ranking can be obtained. The input of this module is the relevant English vocabulary information source (website), and the output is the credibility ranking of the English vocabulary information semantic. The English fact statement semantic credibility verification module has realized the semantic credibility verification of the English fact statements. In this module, according to the similarity of the English vocabulary information and the corresponding English fact statement, as well as the English vocabulary information semantics credibility ranking, the contribution of other English vocabulary information to the corresponding English fact statement verification is measured. Combined with the measurement on the contribution made by the English vocabulary information, the semantic credibility verification of the English fact statement is realized. The input of this module is the random feature selection and the semantic credibility ranking of the relevant English vocabulary information. And the output is the verification result.

### *2.1. Random feature selection measurement*

The similarity between the relevant English vocabulary information and the corresponding English fact statements is the basis for the measurement of the supporting relationship between the relevant English vocabulary information and the corresponding English fact statement. And this section has described the calculation process of the similarity of the relevant English vocabulary information with the corresponding English fact statements.

The random feature selection of the English vocabulary information  $r_i$  and  $fs$  is the similarity of the sentence of  $st_i$  and  $fs$ . This paper has extended the sentence similarity calculation method based on the semantics and word order [10], so as to

realize the similarity measurement of  $st_i$  and  $fs$ . The traditional sentence similarity calculation method based on the semantics and word order does not take into account the influence of the adjacent word of the influence of the adjacent word to the target word on the acquisition of the matching word when generating the semantic vector and work order vector through searching for the matching word of the target word.

In the sentence similarity calculation method described in this paper, it is considered that the semantics of the words in a sentence will be affected by the adjacent words. Therefore, in the process of acquiring the optimal matching word for the target word, the adjacent word to the target word is taken as an important factor for measurement. And the optimal matching word acquisition algorithm is put forward. And on this basis, the corresponding semantic vector and word order vector is obtained. At the same time, in the process of the similarity calculation, the case of the English vocabulary information negative corresponding English fact statement is taken into consideration.

The similarity between the words is the basis of generating the semantic vectors and word order vectors. The formula to calculate the similarity between the words is shown in equation (1). Equation (1) calculates the similarity between the calculated words  $w_1$  and  $w_2$ . Symbols  $l$  and  $h$  represent the shortest distance between  $w_1$  and  $w_2$  in the Wordnet, and depth of the common category that both  $w_1$  and  $w_2$  belong to in the Wordnet, respectively. When  $\alpha = 0.2$ ,  $\beta = 0.45$ , the similarity between the words can be measured properly through equation

$$S_w(w_1, w_2) = \begin{cases} e^{-\alpha l} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}}, & w_1 \neq w_2, \\ 1, & w_1 = w_2. \end{cases} \quad (1)$$

In equation (1), when  $w_1 = w_2$ , it is considered that the relevance is 1; in addition, as in the actual situation, WordNet cannot completely cover all the words that appear in the information, when  $w_1$  or  $w_2$  is not covered by the WordNet, there is  $S_w(w_1, w_2) = 0$ .

Assuming that  $s_1$  is the sentence  $st_i$  that is extracted from  $r_i$ , and  $s_2$  is the corresponding English fact statement  $fs$  of  $r_i$ . The next section describes the similarity calculation process taking the calculation of the similarity of  $s_1$  and  $s_2$ , for example.

Conduct the semantic vector correlation calculation. Through the generation of the corresponding semantic vector of the sentence  $s_1$  and  $s_2$ , the cosine similarity between the semantic vectors is calculated, so as to realize the calculation of the semantic vector correlation. Assuming that the word set after the suspension word is removed from  $s_1$  and  $s_2$  is  $W_1 = \{w_{11}, w_{12}, \dots, w_{1n}\}$ ,  $W_2 = \{w_{21}, w_{22}, \dots, w_{2n}\}$ ,  $W = W_1 \cup W_2$  and  $W = \{w_1, w_2, \dots, w_k\}$ , respectively. Assuming the corresponding semantic vector to  $s_1$  is  $V_1 = \{v_{11}, v_{12}, \dots, v_{1k}\}$ , the process to obtain the component  $v_{1i}$  is as follows:

- (1) Suppose  $w_i \in W$ . If  $w_i \in W_1$ , then  $v_{1i} = 1$ .
- (2) Let  $w_i \in W$ . If  $w_i \notin W_1$ , search for the optimal match word  $w_{bm}$  for  $w_i$  (target word) in  $s_1$ . If  $w_{bm}$  exists,  $v_{1i} = S_w(w_i, w_{bm})$ . Otherwise,  $v_{1i} = 0$  is the process to obtain the optimal match word  $w_{bm}$ .

Similarly to the method of obtaining  $V_1$ , the semantic vector  $V_2$  corresponding

to  $s_2$  can be obtained. The semantic vector correlation of  $s_1$  and  $s_2$  can be obtained by calculating the cosine similarity of  $V_1$  and  $V_2$ , as shown in the formula

$$S_s(s_1, s_2) = \frac{V_1 \bullet V_2}{\|V_1\| \bullet \|V_2\|}. \tag{2}$$

Calculate the word order vector relevance. Through generating the word order vector corresponding to the sentence, according to equation (3), and the word order vector similarity between the sentences is calculated.  $O_1$  and  $O_2$  in Equation (3) represent the word order vector of  $s_1$  and  $s_2$ , respectively. The process to generate the word order vector  $O_1 = \{o_{11}, o_{12}, \dots, o_{1k}\}$  is as follows:

- (1) Let  $w_i \in W_1$ . If  $w_i \in W_1$ ,  $o_{11}$  is the position of  $w_i$  in  $s_1$ .
- (2) Let  $w_i \in W_1$ . If  $w_i \notin W_1$ , according to Algorithm 2, search for the optimal match word  $w_{bm}$  to  $w_i$  in  $s_1$ . If  $w_{bm}$  is present,  $o_{1i}$  is the position of  $w_{bm}$  in  $s_1$ . Otherwise,  $o_{1i} = 0$ . In the process of obtaining the word order vector, the optimal value of the parameter  $\zeta$  involved in Algorithm 2 is 0.4.

$$S_{re}(s_1, s_2) = 1 - \frac{\|O_1 - O_2\|}{\|O_1 + O_2\|}. \tag{3}$$

Conduct the random feature selection calculation. According the semantic vector correlation and word order vector correlation, the random feature selection of  $s_1$  and  $s_2$  can be calculated by equation (4). Since the sentences  $st_i$  and  $s_2$  are the English fact statements  $fs$  (corresponding to  $r_i$  extracted from  $r_i$ ,  $s_1$  and  $s_2$ ) are expressed using  $st_i$  and  $fs$  in equation (4), respectively. And the optimal value of the parameter  $\theta$  in equation (4) is 0.85.

$$S(st_i, fs) = \begin{cases} \theta S_s(st_i, fs) + (1 - \theta) S(st_i, fs) & (r_i \text{ does not have a negative tendency for } fs), \\ -(\theta S_s(st_i, fs) + (1 - \theta) S(st_i, fs)) & (r_i \text{ has a negative tendency for } fs). \end{cases} \tag{4}$$

Whether  $r_i$  has the negative tendency to  $fs$  is verified according to whether the process to obtain  $st_i$  involves the negative grammatical dependency relationship representation, and whether the negative adverb in  $c_i$ , such as hardly, rarely, few, seldom and so on. As can be known from Stanford Parser, the sentence corresponding grammatical dependency between words can represent the negative tendency that is clearly existent in the sentence. For example, when a negative word not appears in a sentence, the corresponding grammatical dependency relationship is neg. And when the negative conjunction, such as rather than, and so on, occurs, it is reflected in the grammar dependency as conj\_negcc, and so on. Therefore, it is possible to determine whether there is negative tendency in the sentence through the grammatical dependency relationship. In addition, when the negative adverbs, hardly, rarely, few, seldom and so on appear in the sentence, it is also the basis to

verify whether there is negative tendency in the sentence. Therefore, whether  $r_i$  has a negative tendency to  $fs$  can be determined according to the following rules:

(1) Whether the grammatical dependency relationships when  $st_i$  is extracted from  $c_i$  includes the grammatical dependency relationship that represents the negative, such as `neg` and `conj_negcc`, etc.

(2) Whether  $c_i$  includes negative adverbs, such as `hardly`, `rarely`, `few`, `seldom`, `scarcely`, `never`, `little`, etc. If one of these aforementioned rules is met, it is considered that the corresponding English vocabulary information  $r_i$  has the negative tendency to  $fs$ . As this paper is to conduct the semantic verification for the affirmative English fact statement, in the aforementioned negative tendency verification process, the case that  $fs$  is the negative English fact statement is not considered.

## 2.2. English fact statement semantic credibility verification

According to the similarity of the relevant information to the corresponding statement, the threshold value  $k$  is introduced, and the relevant English vocabulary information is divided into three categories:

(1) For the supportive English vocabulary information of the corresponding statement, its set is expressed with  $R_{\text{pos}}$ , if  $S(r_i, fs) \geq k$ ,  $r_i \in R_{\text{pos}}$ .

(2) For the objection to the corresponding statement English vocabulary information, and its set is expressed with  $R_{\text{neg}}$ , if  $|S(r_i, fs)| \geq k$ , and  $S(r_i, fs) < 0$ , then  $r_i \in R_{\text{neg}}$ .

(3) For the neutral English vocabulary information, its set is expressed with  $R_{\text{neu}}$ , if  $|S(r_i, fs)| < k$ ,  $r_i \in R_{\text{neu}}$ . The contribution of the English vocabulary information to the English fact statement to be determined is determined by the corresponding semantic relevance and the ranking of the English vocabulary information in the semantic credibility ranking. The contribution of  $r_i$  to the semantic credibility verification of  $fs$  is  $S(r_i, fs) / \text{Crank}$ . Symbols  $\Delta_{\text{pos}}$ ,  $\Delta_{\text{neg}}$  and  $\Delta_{\text{neu}}$  represent that the optimal value of the summary  $k$  of the contribution of the English vocabulary information in  $R_{\text{pos}}$ ,  $R_{\text{neg}}$  and  $R_{\text{neu}}$  to the English fact statements is determined by the experiment.

The basic verification method suggests that if the contribution of the English vocabulary information supporting the statement is much greater than the contribution of the information against the statement in the relevant English vocabulary information corresponding to the English fact statement  $fs$ , the semantics of  $fs$  is credible and vice versa,  $fs$  is the English fact statement with unconfirmed semantic credibility. Through introducing the threshold value  $\delta$ , according to the sum of  $\Delta_{\text{pos}}$  and  $\Delta_{\text{neg}}$  the sum and the relationship with the size of  $\delta$ , the semantic credibility verification of  $fs$  can be achieved. Input the relevant English vocabulary information set  $R$  to the English fact statement  $fs$  to be verified, the semantic credibility ranking corresponding to  $R$  is  $\text{Crank}$ , and the similarity of the English vocabulary information  $r_i$  and the corresponding English fact statement  $fs$  is  $S(r_i, fs)$ , and the threshold value is  $\delta$ ; the output of the algorithm is the result of the verification on  $fs$ . Firstly, the relevant English vocabulary information  $r_i = (i = 1, \dots, n)$  is processed one by one, and  $\Delta_{\text{pos}}, \Delta_{\text{neg}}$  are calculated; then the sum of  $\Delta_{\text{pos}}$  and  $\Delta_{\text{neg}}$

is calculated. If its value is greater than or equal to the threshold value  $\delta$  (the value of  $\delta$  is determined by the experiment), then  $fs$  is the semantic credible English fact statement (return True); on the contrary,  $fs$  is the English facts statement with unconfirmed semantic credibility.

### 3. Experiments

In this paper, there is not public recognized data set existing in the English fact statement semantic verification research field. Therefore, the English fact statements are obtained from TREC2007 to constitute experimental dataset. 30 semantic credible unique answers to the English fact statements and 20 semantics credible multi-answer English fact statements are randomly selected from TREC2007 to constitute the dataset semantics credible English fact statement part. As in the TREC2007, any semantic credible English fact has a corresponding English fact statement that is close to it and without credible semantics, 50 corresponding non-semantically credible English fact statements are selected as the part of the non semantic credible English fact statements in the dataset. In this paper, two related English vocabulary information acquisition methods are provided as follows: To use the English fact statement as the search engine to query and access to the relevant English vocabulary information, which is referred to as FQ method for short; to use the keyword collection of the English fact statement as the search engine to query and access to the relevant English vocabulary information, which is referred to as the KQ method for short. For any English fact statement, through yahoo boss 2.0, two methods of FQ, KQ are used to obtain the first 150 search results (relevant English vocabulary information). In addition, 13 graduate students with long-term Internet experience annotate the English vocabulary information according to the relationship between the obtained English vocabulary information and the corresponding English fact statement. According to the annotation, the English vocabulary information can be divided into three categories, including supporting corresponding statements, objection to the corresponding statements and irrelevant to the corresponding statements.

The rationality and accuracy of the MFSV, English semantic credibility verification model are verified by carrying out a series of experiments.

(1) Through the experimental analysis, in the different English vocabulary information acquisition methods (FQ and KQ), the distribution of the English vocabulary information with the semantics containing the corresponding English fact statement.

(2) The value of the threshold  $k$  affects the classification of the relevant English vocabulary information, thus affecting the accuracy of the English fact statement verification, and the optimal value is obtained through the experiment.

(3) Analyze the ranking distribution of the English vocabulary information semantic credibility in different relevant English vocabulary access modes.

(4) Analyze the influence of English vocabulary information quantity  $n$ , threshold value  $\delta$ , semantic credibility ranking and English vocabulary information acquisition mode on the accuracy of basic verification method.

(5) The influence of  $n$ , semantic credibility ranking and the relevant English vocabulary information acquisition mode on the accuracy of the SVM verification

method.

(6) Analyze the difference between the basic verification method and the SVM verification method in determining the accuracy.

The experiment was conducted on the platform of Intel Core 2Quad2, 66 GHz processor and 2 GB memory Windows 7.

### 3.1. Semantic credibility ranking of English vocabulary information

This experiment analyzes the distribution of the semantic credibility ranking in two acquisitions modes FQ and KQ. Figures 2 and 3 show in FQ and KQ mode, when the number of the English vocabulary information is 150 ( $n = 150$ ), the distribution of CBrank, CBGrank, CFrank and CFGrank, respectively. And the horizontal coordinates represent the position of the English vocabulary information in the English vocabulary information set. The vertical coordinates represent the semantic credibility ranking of the English vocabulary information in the corresponding position.

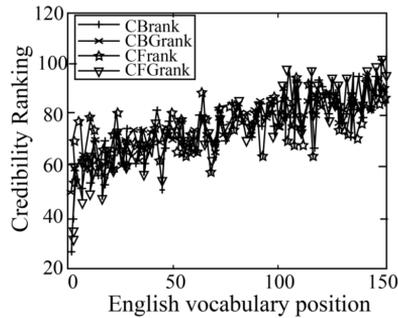


Fig. 2. Semantic credibility ranking in FQ mode

It can be seen from Fig. 2, that the semantic credibility ranking is not significantly related to the position of the English vocabulary information in the English vocabulary information set. The semantic credibility of the English vocabulary information in the English vocabulary information set is not always lower than the semantic credibility of the top ranking English vocabulary information. As CBGrank and CFrank take the Alexa ranking interval into consideration, CBGrank and CFGrank have the characteristics of large span compared with CBrank and CFrank. As can be seen from Fig. 3, in the KQ mode, the semantic credibility ranking shows the similar trend to Fig. 2; compared with the FQ mode, the relevant English vocabulary information semantic credibility distribution in the FQ mode is more concentrated (the span in the FQ and KQ under CFGrank is 26 ~ 102 and 24 ~ 96, respectively), and the reason is the influence of the semantic information contained in the search when the search engine returns the search result in the FQ method compared with the KQ mode.

*3.1.1. Verification case analysis* The validity of the method described in this paper is illustrated taking the English fact statement "English is the primary lan-

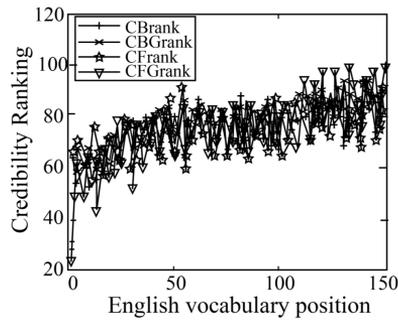


Fig. 3. Distribution of the semantic credibility ranking in the KQ mode

guage of the Philippines" as an example. As "English is the primary language of the Philippines" is a multi-answer English fact statement (when "English" is specified as the unit of doubt, the statement is multi-answer English fact statement). In order to show that the method described in this paper is still valid for the multi-answer English fact statements, another English fact statement "Filipino is the primary language of the Philippines" is verified at the same time; to illustrate the method in this paper can also correctly verify the non-semantically credible statement, the third English fact statement "Chinese is the primary language of the Philippines" is verified. Using the SVM verification method described in this paper to verify the above-mentioned English facts, and the verification result is that the first second English fact statements are true, and the third English fact statement is not true. This case illustrates the validity of the method proposed in this paper in the verification of the English fact statements.

## 4. Conclusion

This paper proposes a multi-answer English fact statements verification model MFSV that is irrelevant to the domains selected based on random features. This model realizes the semantic credibility verification of the English fact to be verified by acquiring and analyzing the relevant English vocabulary information corresponding to the English fact statement. In the process of semantic credibility verification of the English fact statements, the random feature selection between the relevant English vocabulary information and the English fact statements, as well as the semantic credibility of the relevant English vocabulary information is taken into account. And the relevant English vocabulary information semantic credibility and other factors are considered as well, based on which the contribution of the relevant English vocabulary information to the English fact statement semantic credibility verification, and realize the English fact semantic credibility verification. The verification model does not require specifying the English fact statement unit of doubt, which makes the verification model applicable for the unique answer to the English fact and multi-answer English fact semantic verification. As the English fact statements do not include the emotion and degree description, for the negative English fact

statement semantic credibility verification, its semantic credibility verification can be conducted through the corresponding affirmative English fact statements, so as to help make the semantic credibility verification for the negative English fact statement. Therefore, in this paper, verification is conducted mainly on the affirmative English fact statements. The quality of the related English vocabulary information is the prerequisite for the accurate determination of the English facts. When the English facts to be verified are relatively complicated, the corresponding information quality is relatively low, and it is difficult to make the correct verification to such English facts. Therefore, in the future work, it is expected that the English fact statements decomposition, rewriting and other technologies should be adopted to access to the relevant high-quality English vocabulary information, so that the semantic credibility verification for the English fact statements is more accurate.

## References

- [1] B. AGARWAL, S. PORIA, N. MITTAL, A. GELBUKH, A. HUSSAIN: *Concept-level sentiment analysis with dependency-based semantic parsing: A novel approach*. *Cognitive Computation* 7 (2015), No. 4, 487–499.
- [2] Y. W. HUNG, Y. H. CHIU, Y. C. JOU, W. H. CHEN, K. S. CHENG: *Bed posture classification based on artificial neural network using fuzzy c-means and latent semantic analysis*. *Journal of the Chinese Institute of Engineers* 38 (2015), No. 4, 415–425.
- [3] P. MUNRO, S. SIYAMBALAPITIYA: *Improved word comprehension in Global aphasia using a modified semantic feature analysis treatment*. *Clinical Linguistics & Phonetics* 31 (2017), No. 2, 119–136.
- [4] J. U. HEU, I. QASIM, D. H. LEE: *FoDoSu: Multi-document summarization exploiting semantic analysis based on social Folksonomy*. *Information Processing & Management* 51 (2015), No. 1, 212–225.
- [5] J. WANG, J. PENG, O. LIU: *A classification approach for less popular webpages based on latent semantic analysis and rough set model*. *Expert Systems with Applications* 42 (2015), No. 1, 642–648.
- [6] M. GAMBHIR, V. GUPTA: *Recent automatic text summarization techniques: a survey*. *Artificial Intelligence Review* 47 (2017), No. 1, 1–66.
- [7] T. RUETTE, K. EHRET, B. SZMRECSANYI: *A lectometric analysis of aggregated lexical variation in written Standard English with Semantic Vector Space models*. *International Journal of Corpus Linguistics* 21 (2016), No. 1, 48–79.
- [8] L. VAN MAASTRICHT, E. KRAHMER, M. SWERTS: *Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa*. *Language Learning* 66 (2016), No. 1, 124–158.
- [9] G. DIAS, R. MORALIYSKI, J. CORDEIRO, A. DOUCET: *Automatic discovery of word semantic relations using paraphrase alignment and distributional lexical semantics analysis*. *Natural Language Engineering* 16 (2010), No. 4, 439–467.
- [10] M. GRANROTH-WILDING, M. STEEDMAN: *A robust parser-interpreter for jazz chord sequences*. *Journal of New Music Research* 43, (2014), No. 4, 355–374.

Received July 12, 2017



# Wireless multimedia network feasible path routing algorithm

CAILIN LU<sup>1</sup>

**Abstract.** The congestion routing algorithm based on the minimal set covering theory is adopted to carry out inference only on the shared bottleneck path. When there are multiple path congestions in the congested path, the inference performance of the algorithm decreases drastically. Aiming at this problem, an improved Moore Relaxation Sub-gradient algorithm based on KMP (hereinafter referred to as MRSKMP for short) on the basis of Kalman Maximum Posterior (hereinafter referred to as KMP for short) is put forward. In view of the influence of the path coverage in the algorithm on the inference performance of the algorithm, the cost problem is taken into consideration on the basis of the guaranteed path coverage, so as to ensure the inference performance of the algorithm. The experiment has verified the accuracy and robustness of the proposed algorithm.

**Key words.** Congestion path inference, tomography, Kalman network model, Moore relaxation, Kalman maximum posteriori (KMP) criterion.

## 1. Introduction

With the increasing scale of the wireless multimedia network and the rapid growth in the number of network terminal accesses, the number of routers/switches is increasing, in addition to the network congestion caused by the physical path cut-off, complex network structure and unreasonable routing principles will all lead to the occurrence of network feasible path congestion, causing the sharp decline in the overall network performance and quality of service. The high network latency and high packet loss rate of the wireless multimedia network caused by the feasible path congestion may also be caused by the violation of the related service level agreements (SLAs) such as Service Level Agreement (hereinafter referred to as SLA for short) [1–2]. Therefore, the network manager needs to locate and handle the congestion in the network in timely and accurate manner.

At present, domestic and foreign wireless multimedia network internal path performance inference is mainly through two methods including the active detection and

---

<sup>1</sup>School of Information Technology and Engineering , Jinzhong University, 030619, Jinzhong, Shanxi, China

passive detection. In this paper, the active detection of path performance based on a small number of end-to-end (E2E) path detection methods [3–5] is proposed, as it has the advantages of not involving user privacy, real-time performance, small cost, and little impact on the network performance and other advantages, it is preferred by network operators and research scholars at home and abroad. At present, by the application of Boolean algebra, the method of congestion path inference based on the Smallest Consistent Failure Set (SCFS) theory [6] is adopted. When the proportion of the congested path of wireless multimedia network is increasing, in particular, when there are other congestion paths in addition to shared bottleneck paths in a certain congestion path, the performance of algorithmic inference will degrade due to the defects of algorithm theory. In addition, there are some literatures on the active detection method in the probe deployment point [7–8] and E2E contract path optimization [9–10] and other aspects, on the basis of the minimization of cost, to cover as much path range in the wireless multimedia network to be measured as possible. But it does not study the influence of path coverage changes on the performance of algorithm inference. In addition, Boolean tomography makes use of the priori probability of path congestion and CLINK algorithm of congestion path inference based on Kalman theory, which can effectively avoid the dependency of single time slot E2E path detection on time strong correlation. However, in the case of large-scale wireless multimedia networks, it is difficult to solve the problem due to the sparseness of the system matrix coefficient matrix to be solved by the path prior probabilities, it tends to lead to the failure of the solution, and no good solution has been proposed in the literature so far.

In view of the aforementioned problems, based on the practicality of Boolean tomography, this paper proposes an improved Moore Relaxation Sub-gradient algorithm based on KMP (MRSKMP) based on Kalman Maximum A-Posterior (KMP) for large-scale wireless multi-media network feasible path congestion scenario is proposed. Considering the degree of network user and manager's tolerance to the wireless multimedia network congestion, the Path Congestion Time (PCT) parameter is introduced in the process of path prior probabilistic learning, and the E2E path with number of congestion less than PCT in the process of  $N$  times of E2E performance detection is regarded as the normal path. By removing the normal path and the transit path, the congested routing matrix and the congestion Kalman network model in the wireless multimedia network to be measured are constructed during the learning process of path congestion priori probabilities. In the process of congestion path inference, firstly, the normal path and transit path are removed from the congestion Kalman network model which is constructed during the learning process of path congestion priori probabilities, and the remnant congestion routing in the process of congestion path inference is obtained. Finally, based on the KMP criterion, the wireless multimedia network congestion path inference problem is transformed into the Set Cover Problem (SCP), and the MRSKMP algorithm proposed in this paper is used to solve the SCP iterative solution within the polynomial time.

## 2. MRSKMP algorithm

In this paper, a feasible routing algorithm MRSKMP for the large-scale wireless multimedia network is proposed, which mainly consists of three parts:

(1) E2E path and probe deployment optimization. Based on the guaranteed path coverage, according to the wireless multimedia network topology to be measured, the E2E contract detection path and contract routing probe deployment location optimization is conducted.

(2) Path congestion priori learning. According to the result of  $N$  times of E2E path performance measurement, the learning algorithm covered congestion priori probabilities of each path are obtained.

(3) Current time congestion path inference. According to the congestion status of each E2E path at the current inference time, the set of path where congestion is most likely to occur in the current wireless multimedia network is deduced based on the KMP criterion. And the algorithm block diagram is shown in Fig. 1.

## 3. Congestion Kalman network model and congestion routing matrix construction

The Kalman network model is a directed acyclic graph (hereinafter referred to as DAG for short), which can be expressed by the equation

$$G = (v, \varepsilon), \quad (1)$$

where  $v$  represents the node, and  $\varepsilon$  represents the directed edge of the connected node. In the Kalman network, each node stores a conditional probability table. When the node is a known evidence node, the condition probability table is the priori probability distribution of the node. According to the causality in the graph and the consistent conditional probability and prior probability, the unknown hidden node state can be inferred through the evidence node. When constructing the Kalman network model for the wireless multimedia network, the set of state variables  $Y = (y_1, \dots, y_i, \dots, y_{n_p})$  of each E2E path is the observation nodes in the Kalman network. The state variables  $X = (x_1, \dots, x_j, \dots, x_{n_c})$  of each E2E path's transit path are hidden nodes. In order to carry out the congestion path inference, it is necessary to construct the congestion Kalman network model of the wireless multimedia network to be measured at the time of inference.

*Definition 1.* E2E path  $P_i$  congestion, its state variable  $y_i = 1$ ; normally,  $y_i = 0$ . Similarly, in path congestion, its state variable  $x_j = 1$ ; normally,  $x_j = 0$

The Kalman network inference model constructed by the wireless multimedia network is shown in Fig. 2.

When the congested path inference is carried out in the wireless multimedia network, as the path where the congestion path is located must be the congested path, in order to simplify the inference process, the normal path and the transit path in the wireless multimedia network can be omitted in the consideration.

*Definition 2.* Remove each of the E2E probed normal paths (observation nodes)

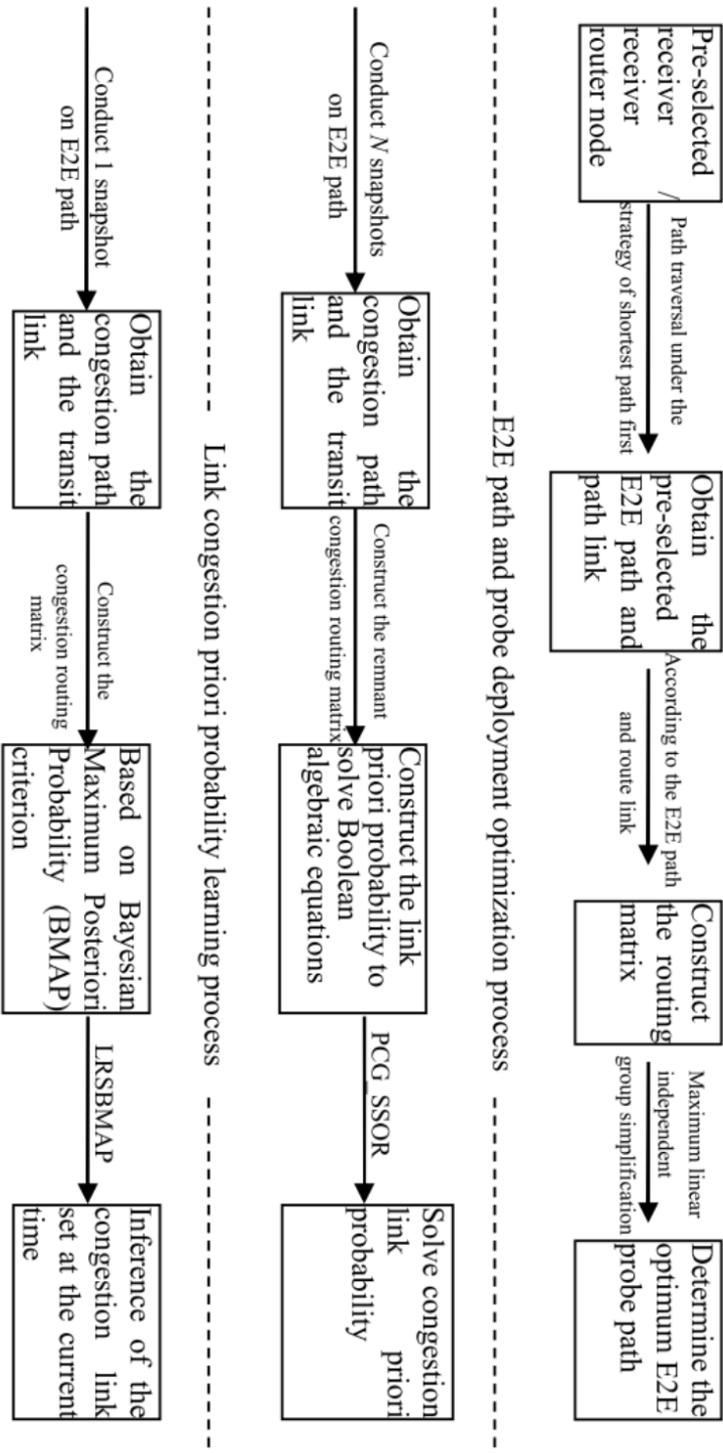


Fig. 1. MRSKMP algorithm block diagram

and the transit paths (hidden nodes) and connect the directed edges in the Kalman network model constructed in the wireless multimedia network to be measured, then the congestion Kalman network model of the wireless multimedia network to be measured is obtained.

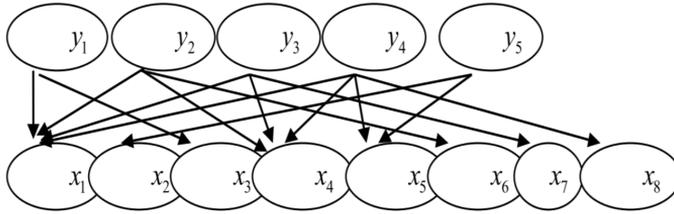


Fig. 2. Kalman network inference model of wireless multimedia network

The process of congestion path inference in this paper includes the construction process of two congestion Kalman network models, which are divided into the construction of the congestion Kalman network model in the wireless multimedia network path congestion priori probabilities learning process and the construction of the congestion Kalman network model in the congestion path inference process.

**3.1. Construction of congestion routing matrix in the learning process**

In the path congestion priori probability learning process, the congestion routing matrix is used as the coefficient matrix in the system of linear equations. Therefore, it is necessary to construct the Kalman network model and the congestion routing matrix in the learning process.  $N$  times of snapshots are performed for each E2E path of the wireless multimedia network to be measured. When the path congestion number does not exceed the set threshold PCT (Path Congestion Time), the path is normal and the transit path is also normal. On the contrary, the path is congested. The size of the parameter PCT can also be set according to the degree of congestion tolerance of the wireless multimedia network to be measured according to the network user or manager. If the network performance requirement is high,  $PCT = 0$  can be set. That is, in  $N$  times snapshot, as long as there is path congestion in one time of detection, then the path is congested. Remove the normal path and transit path from the Kalman network model of the wireless multimedia network as shown in Fig. 3, then the congestion Kalman network model in the learning process of the path congestion priori probabilities can be obtained.

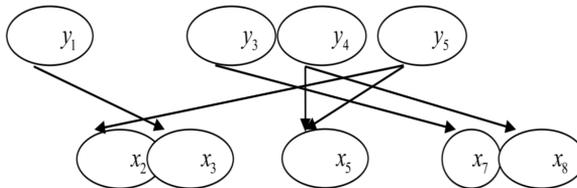


Fig. 3. Congestion Kalman network model of wireless multimedia network

Remove the normal path and the transit route path corresponding matrix rows and columns in  $N$  times of E2E path snapshots from the linearly independent simplified path matrix  $D'$ , and linearly independent simplification is performed again to obtain the path congestion routing matrix  $D''$  of the wireless multimedia network to be measured in the path congestion priori probability learning process.  $N = 30$  times of E2E path detection is performed for the wireless multimedia network as shown in Fig. 2, and the path  $P_2$  remains normal all the time, then the path  $P_2$  and the state variables  $x_1, x_4, X_6$  corresponding to the transit path  $L_1, L_4$  and  $L_6$  as well as the connected directed edges can be removed from the wireless multimedia network Kalman model as shown in Fig. 3, and the congestion Kalman network model in the solving process of the priori probabilities can be obtained after the removal.

Similarly, remove the matrix rows and columns corresponding to the congestion path in the decorrelation reduced matrix  $D'$ , and the matrix after the removal is  $D'_1$  in the form

$$\begin{aligned}
 D'_1 &= \begin{pmatrix} L_1 & L_2 & L_3 & L_4 & L_5 & L_6 & L_7 & L_8 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \end{matrix} \Rightarrow D'_1 = \\
 &= \begin{pmatrix} L_2 & L_3 & L_5 & L_7 & L_8 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{matrix} P_1 \\ P_3 \\ P_4 \\ P_5 \end{matrix} . \tag{2}
 \end{aligned}$$

For the matrix  $D'_1$ : after de-correlation and simplification  $D''$  can be obtained, where  $D'' = D'_1$ , as shown in the equation

$$D'' = \begin{pmatrix} L_2 & L_3 & L_5 & L_7 & L_8 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{matrix} P_1 \\ P_3 \\ P_4 \\ P_5 \end{matrix} . \tag{3}$$

In the inference of the congestion path, such as path  $P_1$  congestion, it is caused by the congestion path  $L_3$ , similarly, path  $L_3$  congestion is caused by the path  $L_7$ . The construction of the congestion Kalman network model can effectively reduce the complexity of the congestion path inference.

### 3.2. Construction of the congestion routing matrix in the inference process

In the process of the congestion path inference, it is necessary to construct the remnant congestion routing matrix  $D_d$  to perform one time of E2E performance

detection snapshot for the corresponding congestion path in the congestion Kalman network model during the learning process of path congestion prior probabilities, to obtain the performance results of each E2E path, and the remnant congested Kalman-net model of the current congestion chain inference can be obtained by removing the normal path in the detection results and the corresponding node as well as the directed edge of the transit path in the learning process. In the same way, the normal path at the time of inference and the transit path corresponding matrix lines and matrix columns are removed from the congestion routing matrix  $D''$  constructed during the learning process of path congestion priori probabilities. After the linearity-independent simplification, the remnant congestion matrix  $D_d$  of the wireless multimedia network to be measured in the congestion path inference can be obtained. As shown in Fig. 2, for the wireless multimedia network, in the inference process, if the measured path P4 is a normal path, the remnant congestion routing matrix is shown in the following equation

$$D_d = \begin{pmatrix} L_2 & L_3 & L_5 & L_7 & L_8 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{matrix} P_1 \\ P_3 \\ P_4 \\ P_5 \end{matrix} \Rightarrow D_d = \begin{pmatrix} L_2 & L_3 & L_7 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{matrix} P_1 \\ P_3 \\ P_5 \end{matrix}. \quad (4)$$

#### 4. Experimental verification

In order to validate the effectiveness and accuracy of the routing algorithm, three different types and scales of Waxman, BA and GLP are generated respectively by Brite topology generator. Among them, the Waxman model is the representative based on the random graph model, and the node degree value in the model increases with the number of nodes, but the random graph model cannot generate a network with many nodes but the one with small node average value. As the scale of the wireless multimedia network continues to expand, new router nodes tend to connect to the "Big nodes" with high values, when they join the Internet. Based on these two features, the scale-free network model BA and GLP with the power distribution of degree distribution are constructed so as to better verify the performance of the proposed algorithm in different Internet environments, and experiment is conducted to compare the algorithm performance under the three kinds of topological network models.

The topology model is introduced to complete the construction of the network to be measured through the Eclipse platform, and all the routing algorithms are used to verify the congestion path inference experimental verification. In the experiment of congestion path inference, the shortest path first principle of the wireless multimedia network routing algorithm is simulated, and the ICMP protocol is used to perform the snapshots (including Traceroute and Ping) respectively, and E2E path and route path and E2E path performance measurement values are obtained. The RNM (Random Number Model) in this paper is applied to simulate the congestion events generated by the path covered by the algorithm in each snapshots of the

wireless multimedia network to be measured.

In this paper, the parameters of the proposed algorithm mainly include:

When  $DTV$ -router degree  $\leq DTV$ , the router is taken as the pre-selected transceiver router, and automatically optimized according to the value of  $\rho$  for  $DTV$ ;

In the  $PCT$ - $N$  times E2E path detection, the path with number of congestion less than  $\leq PCT$  is normal. The algorithm defaults to set  $N = 30$  and  $PCT = 0$ ;

LCR (Link Congestion Ratio) - parameters set in the MRSKMP algorithm simulation experiment. That is: the ratio of the congestion path to the path covered by the algorithm, with the value taking range of  $[0, 1]$ . By selecting the path random number assigned from large to small according to LCR to obtain the congestion path of each snapshots. The detection rate DR and the false positive rate (FPR) is used to evaluate the congestion path inference result of the MRSKMP algorithm proposed in this paper. In order to reduce the effect of the random number model on the inference performance of the algorithm, the DR and FPR in each experiment are the results obtained after averaging the 10 experimental results under the same parameters.

The calculation formula of DR and FPR is shown in equation

$$DR = \frac{F \cap X}{F} \quad FPR = \frac{X \setminus F}{X}, \quad (5)$$

where  $F$  is the actual congestion path and  $X$  is the congestion path deduced by the algorithm. The simulation experiment process is shown in Fig. 4 as the following.

In order to verify the effectiveness and accuracy of the proposed MRSKMP algorithm in congestion path inference, Brite topology generator is adopted to simulate wireless multimedia network models Waxman, BA and GLP with different types and sizes by the default parameters, and compared with the CLINK algorithm on the inference performance.

For the wireless multimedia network model with the scale of 150 nodes, CLR changes from 0.05~0.6, and DR and FPR of the two algorithms under optimal  $DTV$  are shown in Fig. 5.

In different types of wireless multimedia network model, MRSKMP algorithm inference performance is superior to CLINK algorithm. With the increase of CLR, DR shows a decreasing trend. The DR of the two algorithms is the highest under the GLP model, followed by the BA and Waxman models. As Waxman is a stochastic model, the path is relatively long, while BA and GLP are power-rate models, in which some routers have larger values and share more paths than Waxman's model in the wireless multimedia network model topology. Therefore, in the Waxman model, DR has decreased significantly compared with GLP and BA model. When  $CLR < 0.2$ , the inference performance of MRSKMP and CLINK algorithm in GLP and BA model is not very different. However, when  $CLR > 0.2$ , the inference performance of MRSKMP algorithm is better than CLINK algorithm in Waxman, BA and GLP models, and when CLR increases, the inference performance advantage is more significant, demonstrating the performance advantage of MRSKMP algorithm under feasible path congestion. As the CLR increases, the performance degradation of the MRSKMP algorithm is slower than that of the CLINK algorithm. CLRR is

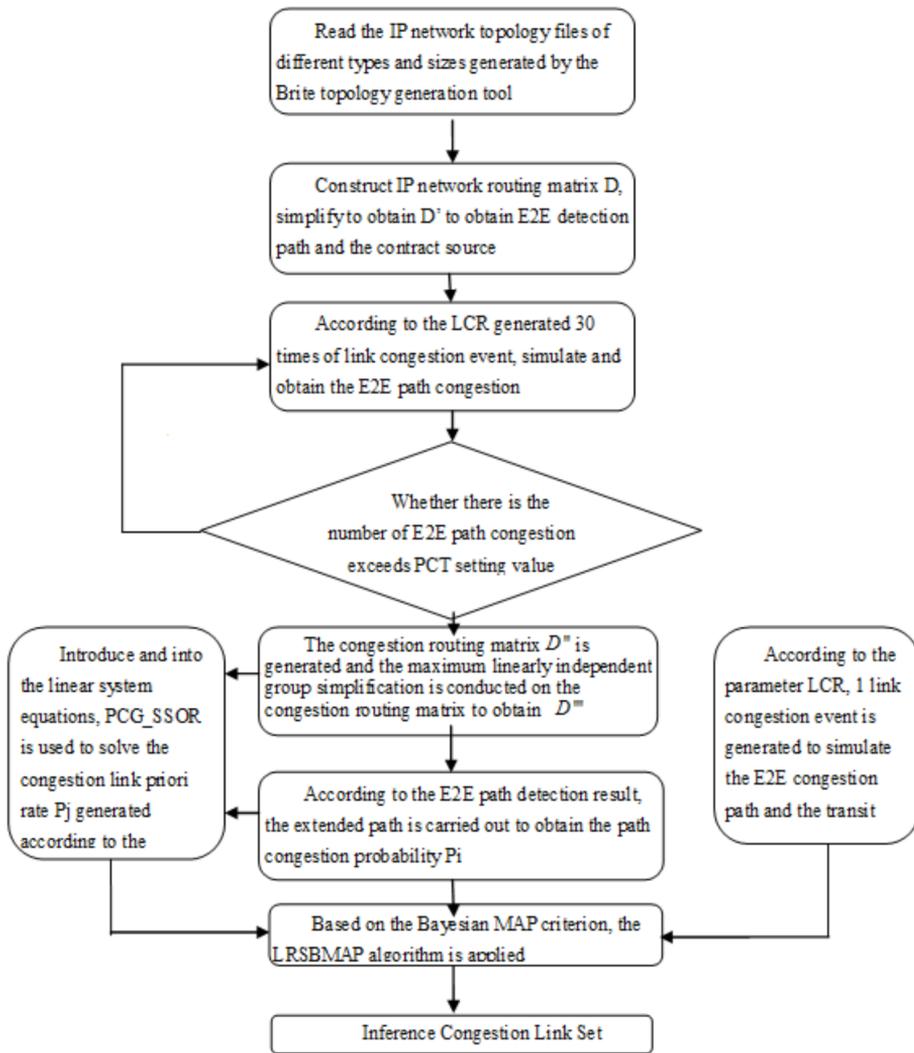


Fig. 4. Simulation experiment process

less than 55% in GLP and BA model, and only 40% in Waxman model; and DR still remains about 75 in DRP model, and is 65% and 55% and above in the BA and Waxman model respectively. Both algorithms have the lowest FPR in the GLP model, followed by the BA and Waxman models. With the increase of CLR, FPR first shows a slowly rising trend, and when CLR reaches a certain percentage, FPR shows a downward trend.

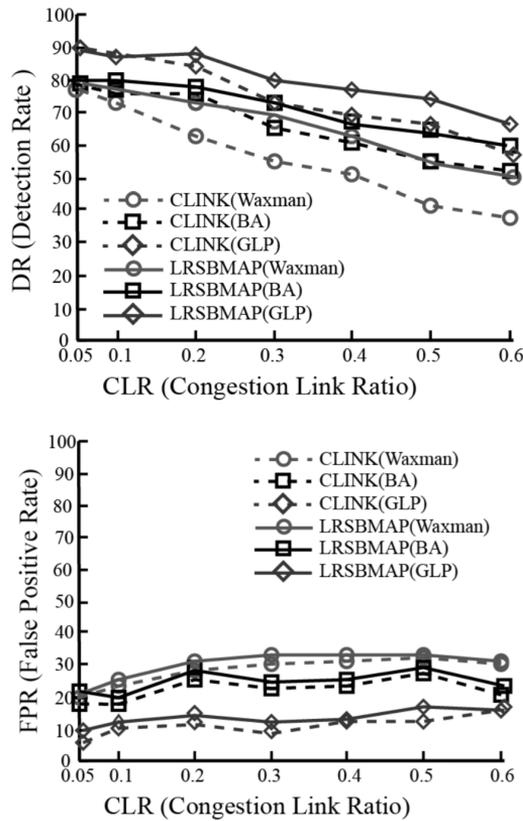


Fig. 5. Comparison of the inference performance of two algorithms under different CLR (number of nodes = 150)

#### 4.1. Influence of different network scale on the algorithm

In order to verify the inference performance of the algorithm in different wireless multimedia network types and scales, the Waxman, BA and GLP network topological model with the node number 50~500 generated by Brite is adopted. Set the feasible path congestion scenario,  $CLR = 0.5$ . DR and FPR of MRSKMP algorithm and CLINK algorithm are shown in Fig. 6.

From Fig. 6, the inference performance of the two algorithms in different types and scales of wireless multimedia network models decreases slowly with the increase of network scale. Among them, MRSKMP algorithm is superior to CLINK algorithm in the reference performance for the Waxman, BA and GLP model, and DR is highest in the GLP model, followed by BA and Waxman model. In the GLP model, FPR is the lowest, followed by the BA and Waxman model. The inference FPR of the two algorithms in the Waxman, BA and GLP models remain basically stable with the increase of the wireless multimedia network scale. And FPR in the GLP model is lower than that in the BA and Waxman models. Under the three different network

models, the difference of the congestion path inference FPR between the MRSKMP and the CLINK algorithm is small. When  $CLR = 0.5$ , the average FPR of the MRSKMP algorithm is slightly higher than that of the CLINK algorithm.

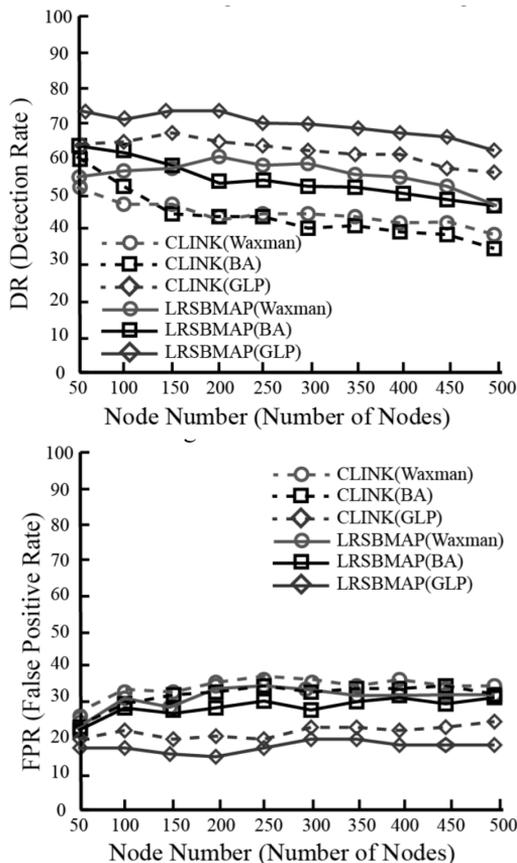


Fig. 6. Comparison of the inference performance of two algorithms  $CLR = 0.5$  under different network scale

## 5. Conclusion

This paper proposes a congestion path routing algorithm MRSKMP in the scenario of a large scale wireless multimedia network with feasible path congestion. The path coverage and the number of E2E probe paths and the probe deployment overhead are taken into account through degree threshold optimization, so as to cover as many paths to be measured as possible; based on the remnant congestion routing matrix and KMP criterion at the time of inference, the improved Moore relaxation sub-gradient algorithm is adopted to infer the set of paths where congestion is most likely to occur. And the experiment has verified the accuracy and robustness of the algorithm proposed in the paper.

## References

- [1] S. L. PAN, X. R. YANG, Z. Y. ZHANG, F. QIAN, G. M. HU: *Congestion link identification under multipath routing for single-source networks*. Journal of Electronics & Information Technology *37* (2015), No. 09, 2232–2237.
- [2] B. BAKHSHI, S. KHORSANDI: *Complexity and design of QoS routing algorithms in wireless mesh networks*. Computer Communications *34* (2011), No. 14, 1722–1737.
- [3] N. G. DUFFIELD: *Network tomography of binary network performance characteristics*. IEEE Transactions on Information Theory *52* (2006), No. 12, 5373–5388.
- [4] X. CAO, R. C. WANG, H. P. HUANG, L. SUN, F. XIAO: *Multi-path routing algorithm for video stream in wireless multimedia sensor networks*. Journal of Software *23* (2012), No. 1, 108–121.
- [5] M. COATES, A. HERO, R. NOWAK, B. YU: *Internet tomography*. IEEE Signal Processing Magazine *19* (2002), No. 3, 47–65.
- [6] L. BRABANT, J. VLASSENBROECK, Y. DE WITTE, V. CNUUDE: *Three-dimensional analysis of high-resolution X-ray computed tomography data with morpho+*. Microscopy and Microanalysis *17* (2011), No. 2, 252–263.
- [7] T. HE, A. GKELIAS, L. MA, K. K. LEUNG, A. SWAMI, D. TOWSLEY: *Robust and efficient monitor placement for network tomography in dynamic networks*. IEEE/ACM Transactions on Networking *25* (2017), No. 3, 1732–1745.
- [8] L. SHAN, Z. LI, H. HU: *Converged mobile cellular networks and wireless sensor networks for machine-to-machine communications*. KSII Transactions on Internet and Information Systems *6* (2012), No. 1, 147–161.
- [9] T. MATSUDA, M. NAGAHARA, K. HAYASHI: *Link quality classifier with compressed sensing based on L1-L2 optimization*. IEEE Communications Letters *15* (2011), No. 10, 1117–1119.
- [10] K. NAKANISHI, S. HARA, T. MATSUDA, K. TAKIZAWA, F. ONO, R. MIURA: *Synchronization-free delay tomography based on compressed sensing*. IEEE Communications Letters *18*, (2014), No. 8, 1343–1346.

Received July 6, 2017

# Fault diagnosis of smart grid based on improved immune optimization algorithm<sup>1</sup>

HONGNA LI<sup>2,3</sup>, HUI ZHAO<sup>2,4,5</sup>

**Abstract.** Aiming at the problem of fault diagnosis of smart grid, an improved immune optimization algorithm was proposed in this paper. Firstly, the idea of chaos algorithm was integrated into the immune optimization algorithm to make the algorithm have stronger global optimization ability; secondly, an improved mutation operator was proposed to make the algorithm be in the vicinity of the current optimal solution for local search, so as to strengthen the local search capabilities. Simulation results show that the performance of the improved immune optimization algorithm is superior to that of the traditional immune optimization algorithm, and it has better stability and search ability, and is more suitable for applications requiring high stability and accuracy.

**Key words.** Smart grid, fault diagnosis, immune optimization, chaos algorithm.

## 1. Introduction

With the increase in the type of electricity, the levels of distribution network voltage continue to increase, and the structure of distribution network is also more complex. Electrical wiring aging, man-made operational errors or natural disasters and other reasons have caused the occurrence of the distribution network. Therefore, how to improve the speed of fault diagnosis and the efficiency of distribution network has become a hot research topic.

In order to meet the goal of actual fault diagnosis and realize the goal of fault location and fast location in the distribution network, many experts and scholars

---

<sup>1</sup>This research was performed with 2012 Colleges and Universities in Tianjin Science and Technology Development Fund Program (Project No. 20120801) and The Teaching Quality and Teaching Reform of Undergraduate Course of Common Colleges and Universities in Tianjin Research Projects (Project No. B01-0818).

<sup>2</sup>School of Electrical Engineering & Automation, Tianjin University, 300072, Tianjin, China

<sup>3</sup>Maritime College, Tianjin University of Technology, 300222, Tianjin, China

<sup>4</sup>College of Engineering and Technology, Tianjin Agricultural University, 300384, Tianjin, China

<sup>5</sup>Corresponding author

who engage in the electric power research and fault location research at home and abroad have put forward many methods. In reference [1], the binary group intelligent algorithm was used to optimize the fault diagnosis of the smart grid. In reference [2], the binary particle swarm optimization algorithm was used to optimize the fault diagnosis of the smart grid. In reference [3], an improved genetic algorithm was used to optimize BP-NN to optimize the fault diagnosis of smart grid.

The immune optimization algorithm is an intelligent optimization algorithm developed in recent years [4, 5], which draws on the immune system to deal with the performance of the pathogen. As with the genetic algorithm, the immune optimization algorithm has been optimized according to the law that “it is not the strongest of the species that survive, but the one most responsive to change” in body system. Immune evolutionary algorithm, as an intelligent algorithm with global optimization ability, generates the offspring population based on the optimal solution of the previous generation, and then uses the convergence of the best individual instead of the population’s convergence, which has good adaptability and distribution. Therefore, immune evolutionary algorithm has been applied in many optimization problems because of the excellent effect. But the algorithm is easy to fall into the local optimum, and the convergence is slow.

In this paper, an improved immune algorithm was proposed to solve the problem of intelligent fault diagnosis. A new mutation operator was proposed, which not only has the ability of global optimization, but also improves the convergence speed. The simulation results show that the improved algorithm is effective.

## 2. Methodology

The idea of establishing the optimal model for fault diagnosis of smart grid is to establish a model to minimize the optimization goal based on the logical relationship between the faulty elements and the protection and switching action. Then the fault diagnosis problem is transformed into the 0–1 programming problem, so that the corresponding optimization algorithm can be optimized to obtain an optimal failure hypothesis.

### *2.1. Factors affecting the smart grid fault diagnosis*

The factors that affect the fault diagnosis of the smart grid can be summarized into four categories: protection and error actions in circuit breaker, error reporting in information transmission, incomplete model building, uncertainties in time and space dimensions.

When there is a problem in the related equipment in smart grid, the device with the protection device can quickly make the action, in order to jump the corresponding circuit breaker, so that the problem of equipment will be isolated, and other normal equipment can continue to work without affected. However, when the protection or circuit breaker malfunction causes errors in other normal equipment operation, it will lead to power failure in grid area that should not have been affected, thus making things more serious.

During the process of transmission of information in smart grid system, the important information transmission distortion in the smart grid system is caused by the errors of the relevant equipment in the sampling process, the unexplained interruption of the channel or other reasons, which will cause the problem of failure to report information in a timely manner, or to report a delay, thus leading to false positives and making troubleshooting more difficult.

In recent years, there has been lots of research on smart grid fault diagnosis, and research on diagnostic model has also made great progress. Most of the major models consider the error reporting in the smart grid equipment, protection and circuit breaker action and transmission of information, but these factors are not perfect with several problems: the possibility of expansion faults caused by power grid equipment are not considered. The consistency of protection action is not taken into account in the measurement function. Sampling data is mainly taken from fault diagnosis, which lacks the overall consideration.

When the fault occurred in the normal work of the smart grid, there is no way to know in advance that how long the failure occurred and how many times it occurred during malfunction. The fault problem may be found in 10 seconds if problem is quite simple. Once the fault problem becomes serious, the time will be extended to a few minutes, and different time period will result in difference in the number of fault events. So the time and space dimensions in the fault diagnosis of the smart grid have uncertainties.

Based on the idea of reference [6], the fault diagnosis problem of the smart grid is expressed as the minimization of the objective function shown in the equation

$$E(S) = \sum_{k=1}^{n_r} |r_k - r_k^*(S)| + \sum_{j=1}^{n_c} |c_j - c_j^*(S, R)|, \quad (1)$$

where  $S$  is an  $n$ -dimensional vector representing the state of the elements in the system ( $n$  is the number of elements in the system). Equality  $S_i = 0$  represents that the  $i$ th element is normal;  $S_i = 1$  means that the  $i$ th element is faulted. Symbol  $n_r$  denotes the total number of protection,  $n_c$  is the total number of circuit breakers,  $r_k$  is the  $k$ th element in  $R$  representing the real state of the  $k$ th protection (non-action state or action state), in which  $R$  is the  $n_r$ -dimensional vector representing the actual state of the  $n_r$  protection. Quantity  $r_k^*(S)$  is the  $k$ th element in  $R^*(S)$  representing the desired state of the  $k$ th protection; if the  $k$ th one protects this action, then  $r_k^*(S) = 1$ , or  $r_k^*(S) = 0$ ;  $R^*(S)$  is the  $n_r$ -dimensional vector representing the expected state of the  $n_r$ th protection. The value of  $R^*(S)$  was decided by the state of  $S$ ;  $c_j$  is the  $j$ th element in  $C$  representing the real state of the  $j$ th circuit breaker. Equality  $c_j = 0$  means that the  $j$ th circuit breaker is in the un-tripped state,  $c_j = 1$  means that the  $j$ th circuit breaker is in the tripped state. Symbol  $C$  denotes the  $n_c$ -dimensional vector representing the actual state of the  $n_c$ th circuit breaker and  $c_j^*(S, R)$  is the  $j$ th element in  $C^*(S, R)$  representing the expected state of the  $j$ th circuit breaker. If the  $j$ th circuit breaker ought to trip, then  $c_j^*(S, R) = 1$ , or  $c_j^*(S, R) = 0$ . Symbol  $C^*(S, R)$  denotes the  $n_c$ -dimensional vector representing the expected state of the  $n_c$ th circuit breaker. The value of  $C^*(S, R)$  was decided

by the states of  $S$  and  $R$ .

Based on the above model, it can be found that the model exhibits some flaws, leaves out of consideration on switch and wrong action in the circuit breaker, such as malfunction and refusal. Therefore, new factors can be added into the modeling on the basis of the above model [7]. The model is as follows:

$$E(G) = \sum_{i=1}^Z \|r_i = r'_i\| + \sum_{i=1}^K \|c_i = c'_i\| + \\ + w_1 \sum_{i=1}^{Z+K} \|d_i\| + w_2 \sum_{i=1}^{Z+K} \|m_i\| + w_3 \sum_{i=1}^{2Z+2K} \|F_i(S, R, C, M, D)\| \quad (2)$$

In the above model expression,  $\|\cdot\|$  represents the norm of the logical variable; on the right side of the above equation, the first two items are the false action logic of the protection and circuit breaker representing the error report of the warning message. On the right side, the third and fourth items represent abnormal action of protection of a circuit breaker, and the last one represents the constraints of the model;  $w_1$ ,  $w_2$ , and  $w_3$  are the model weights. Particularly,  $w_1$  and  $w_2$  represent the weights of abnormal action of protection and circuit breaker, while  $w_3$  is the model of the protection factor.

The weight value of the model has great influence on the practicability of the model. The correct weights represent the model more practical, therefore, the weight of the model should be chosen. Weight  $w_3$  as the guaranteed coefficient of the model has greater value than 1, and  $w_1$  and  $w_2$  represent the weights of abnormal action of protection and circuit breaker. We set the sum of  $w_1$  and  $w_2$  to 1, and their further identification is carried out by the AHP method.

Analytic Hierarchy Process divides the factors related to decision-making into target layer, scheme layer and criterion layer. The decision-making mode is quantitative and qualitative analysis for relevant factors, which was proposed by Suttie, a famous American operational research expert, in the early 1970s of 20th century.

Analytic Hierarchy Process refers to regard a complex decision-making problem with multiple decision objects as a multi-layer system. It divides the goals of decision into multiple targets or criteria, and then divides them into several levels with certain indexes and constraints. The qualitative index fuzzy quantification method is used to calculate the single rank and the total ranking of the hierarchy, which can be used as the multi-objective and multi-scheme system method to optimize the decision [8]. Since it can be turned into simple-weight polynomials, this method has been applied to many practical optimization problems [9].

The steps for the determination of weight are:

(1) The first step is to establish hierarchical structure, the top layer is abnormal action of the protection and circuit breaker. The bottom layer is the object layer, namely, the rejecting action of protection and circuit breaker, and malfunction of protection and circuit breaker.

(2) The second step is to construct a 6-person decision-making group, the individual can judge two factors according to the relative importance of the scale method,

so as to obtain pairwise comparison matrix according to the judgment results.

(3) The third step is to calculate the maximum eigenvalue of each comparison matrix and its corresponding eigenvector.

(4) The fourth step is to use consistency indicators for testing, if they pass the test, the normalization of the feature vector can be obtained after the weight vector.

(5) The final weight of each factor is the average of the six groups of weight vector.

Finally, the analytic hierarchy process can be drawn:  $w_1 = 0.64$ ,  $w_2 = 0.36$ .

## ***2.2. Improvement of the algorithm***

The artificial immune optimization algorithm is an optimization algorithm for simulating the immune system of the human body and a new intelligent optimization algorithm based on the immune system. The immune algorithm is characterized by the diversity of the population and the maintenance mechanism of the algorithm. Compared with other algorithms, it avoids dealing with more difficult precocious problem. And it performs well in global optimization. The occurrence time of algorithm is later than some of the classic intelligent algorithm. The theory has developed just over a decade. Farmer and other scholars in the 1980s proposed basic framework of the immune system based on the immune system theory to explore the connection between immune system and other artificial intelligence methods, thus creating the immune system.

General steps of immune algorithm is as follows:

(1) Analyze problems that needed to be resolved. Figure out the solution of the mathematical model.

(2) Generate the first generation population. Randomly generate  $N$ th individuals, and then compose the first generation solution population with the use of  $m$  individuals taken from Memory Bank, where  $m$  is the number of individuals inside the memory.

(3) Evaluate the above individuals within the population. The evaluation of the individual in the immune optimization algorithm is based on the expected reproduction rate  $P$  of the individual.

(4) Generate the parent population. The initial population is sorted in descending order by the expected reproduction rate  $P$ , and then the previous individual is extracted to form the parent population; and the previous individual is placed in the memory bank.

(5) If the maximum number of iterations is reached, the algorithm terminates; if the maximum number of iterations is not reached, the algorithm continues to be optimized.

(6) Generate new populations. According to the optimization result obtained in the step (4), the antibody is crossed, selected and mutated to generate new population, and then the individuals in the memory are taken out to form new population.

(7) Go back to perform step (3).

Some definitions of immune algorithm include:

The affinity between the antibody and the antigen in the algorithm is used to indicate the recognition degree of the antigen by the antibody, the affinity function is  $A_V = \frac{1}{F_V}$ , where  $F_V$  is the objective function, namely, the optimization model:

$$E(G) = \sum_{i=1}^Z \left\| r_i = r'_i \right\| + \sum_{i=1}^K \left\| c_i = c'_i \right\| + \\ + w_1 \sum_{i=1}^{Z+K} \|d_i\| + w_2 \sum_{i=1}^{Z+K} \|m_i\| + w_3 \sum_{i=1}^{2Z+2K} \|F_i(S, R, C, M, D)\| .$$

The affinity between the two antibodies represents the degree of similarity between the two antibodies, and the immune optimization algorithm expresses the affinity by the equation

$$S_{v,s} = \frac{k_{v,s}}{L}, \quad (3)$$

where  $k_{v,s}$  is the same number of bits between antibody  $v$  and antibody  $s$ ;  $L$  is the length of the antibody. Finally,

$$C_V = \frac{1}{N} \sum_{j \in N} S_{v,s}, \quad (4)$$

where  $N$  is the total number of antibodies. Quantity  $S_{v,s}$  is given as

$$S_{v,s} = \begin{cases} 1 & S_{v,s} > T, \\ 0 & \text{others} \end{cases},$$

where  $T$  is a preset threshold value.

In a population, the expected reproductive probability of an individual is determined on the basis of both the antibody-antigen affinity  $A_V$  and the antibody concentration  $C_V$  namely:

$$P = \alpha \frac{A_V}{\sum A_V} + (1 - \alpha) \frac{C_V}{\sum C_V} \quad (5)$$

where  $\alpha$  is constant. According to the above formula, it can be seen that the greater the adaptability value of the corresponding adaptability, the greater the expected reproductive probability. The larger the value of the individual corresponding concentration, the smaller the expected reproductive probability. It not only strengthens the individual with high adaptability, but also weakens the individual with high concentration, and then maintains the individual diversity.

Because the clonal selection in immune algorithm is mainly judge on the basis of the size of the affinity. Only the larger affinity and stronger combining capacity of antibody can be maintained to the next step optimization. Mutation operation is relatively antibody with strong binding capacity. Therefore, the algorithm is easy to fall into the local optimal solution, and the convergence speed is slow, so the immune

algorithm is needed to be improved accordingly.

### 2.3. Introduction of chaos algorithm

The chaos algorithm has the characteristics of randomness and ergodicity, so it can be searched completely within the set range, so it can jump out of the local optimal solution. So many optimization algorithms and chaos algorithm are combined to optimize the effect. The chaos model used in this paper is Logistic mapping model and its equation is:

$$x_{k+1} = \lambda x_k(1 - x_k) \quad x_k \in [0, 1] \quad (6)$$

In the above formula,  $\lambda$  is the control parameters, and its value is between 0 and 4. Logistic map is an irreversible range between 0 and 1. When  $\lambda = 4$ , the system enters the so-called chaotic state. The initial point is set in any position, and can generate points between 0 and 1. A logistic map is used to obtain a chaotic point sequence, which is then transformed into a variable of the solution space of the problem to be solved, and the optimal solution for problem can be searched.

In this paper, the chaotic algorithm is used to optimize the current optimal solution, and the current optimal solution is mapped into the chaotic variable [0,1]. According to  $x'_i = c_i + d_i \beta_i^{(\mu+1)}$ ,  $r$  chaotic variables are selected into the optimal variables of the  $r$ th equation (4) to transfer the optimal variables into chaotic variables  $x'_i$ , which is the variation range of the chaotic variable corresponding to the range of the optimal variable. Here,  $c_i d_i$  are the changed constant,  $i = 1, 2, \dots, r$ . Let

$$X = (x_1 x_2 \dots x_r), \quad X' = (x'_1 x'_2 \dots x'_r), \quad (7)$$

and then, chaotic variables are encoded.

### 2.4. Improved mutation operator

In some traditional optimization algorithms, the mutation operator can realize the searching in the whole solution space by the random position transformation. The search ability of the algorithm is greatly enhanced, but the convergence speed is slow due to the lack of regular search. The mutation operator which is the same as the traditional mutation operator is proposed when the normal algorithm is optimized. However, if necessary, the mutation operator can also search the near optimal solution to strengthen a certain local search capabilities.

Definition: let  $N$ -dimensional optimal problem feasible region  $[l_i, u_i]$ ,  $i = 0, 1, \dots, N$ , Suppose the parent is  $a_1 = [a_{11} a_{12} \dots a_{1N}]$ . After mutation of the offspring, improved mutation algorithms are obtained  $T_{mm}(a_1) = [a'_{11} a'_{12} \dots a'_{1N}]$ , which makes it possible for different mutation operators to act on individual components accord-

ing to different probabilities or remain unchanged:

$$a'_{1i} = \begin{cases} a_{1,i} + N(0, 1), rand < p_i, \\ a_{1,i}, p_i \leq rand \leq p_j, \\ l(i) + rand \times u(i) - l(i), rand \geq p_j, \end{cases} \quad (8)$$

where  $i = 1, 2, \dots, N$ ,  $N(0, 1)$  is a random number obeying the standard normal distribution; rand is a random number between  $[0,1]$  with uniform distribution.

In the equation,  $a_{1,i}$  in the antibody  $a_1$  is the probability of  $p_i$  in Gaussian variation. That is, variation around the individual can enhance the local search ability of the algorithm and improve the local search precision of the algorithm; the probability of  $1-p_j$  was used for the variation in global scope, which can maintain the diversity of the population, improve the global search ability of the algorithm, and make the whole group easily jump out of the local optimal.

### 3. Experiment simulation

In order to verify the effectiveness of this algorithm, taking the actual distribution network system as an example. A simple scheme of the grid system is shown in Fig. 1. In Fig. 1,  $S_1 - S_8$  are eight elements,  $c_1 - c_7$  are circuit breakers. Finally,  $r_1 - r_{20}$  are 20 protection devices.

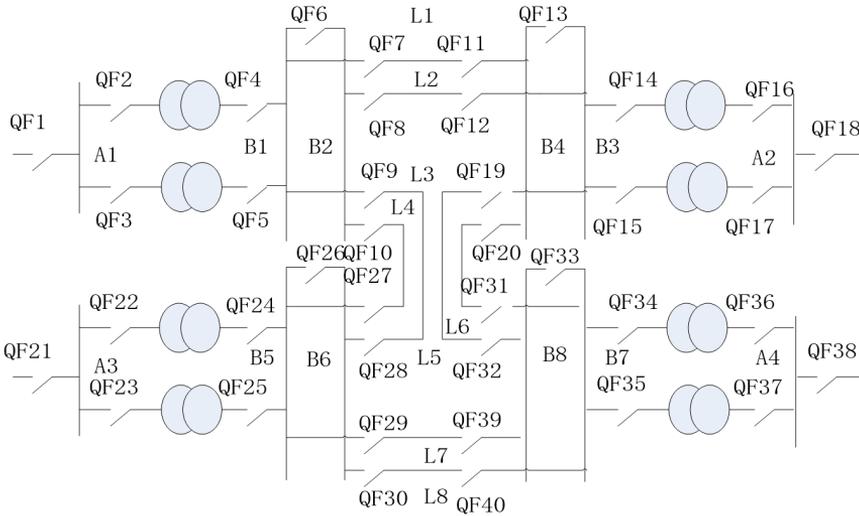


Fig. 1. Actual distribution of network system

Taking the action information of 7 grid protections and circuit breakers as examples, algorithm in this paper is used to optimize

- a. protect  $B_{im}, L_{2Rs}, L_{4Rs}$  action, switches  $QF_4, QF_5, QF_7, QF_9, QF_{12}, QF_{27}$  were tripped;
- b. protect  $B_{1m}, L_{1Sp}, L_{1Rm}$  action, switches  $QF_4, QF_5, QF_6, QF_7, QF_9, QF_{11}$  were

tripped;

c. protect  $B_{1m}, B_{2m}, L_{1Sm}, L_{1Rp}, L_{2Sp}, L_{2Rm}$ , action, switches  $QF_4, QF_5, QF_6, QF_7, QF_8, QF_9, QF_{10}, QF_{11}, QF_{12}$  were tripped;

d. protect  $T_{3p}, L_{7Sp}, L_{2Rp}$  action, switches  $QF_{14}, QF_{16}, QF_{29}, QF_{39}$  were tripped;

e. protect  $T_{5s}, T_{6s}$  action, switches  $QF_{22}, QF_{23}, QF_{24}, QF_{25}$  were tripped;

f. protect  $T_{7m}, T_{8p}, B_{7pm}, B_{8m}, L_{5Sm}, L_{5Rp}, L_{6Ss}, L_{7Sp}, L_{7Rm}, L_{8Ss}$  action, switches  $QF_{19}, QF_{20}, QF_{29}, QF_{30}, QF_{32}, QF_{33}, QF_{34}, QF_{35}, QF_{36}, QF_{37}, QF_{39}$  were tripped;

g. protect  $L_{1Sm}, L_{1Rp}, L_{2Sp}, L_{2Rp}, L_{7Sp}, L_{7Rm}, L_{8Sp}, L_{8Rm}$  action, switches  $QF_7, QF_8, QF_{11}, QF_{12}, QF_{29}, QF_{30}, QF_{39}, QF_{40}$  were tripped.

As can be seen from Table 1, the optimization results in this paper are the same as those in reference [9], which proves that the proposed algorithm can effectively solve the fault diagnosis problem of the smart grid, which can be found in the convergence algebra. This algorithm can converge to the global optimal solution more quickly when solving the same problem.

Table 1. Optimized results

Sequence of grid action information	Optimized results of reference [11]	Mean convergence generations	Optimized results of algorithm in this paper	Mean convergence generations
a	$B_1$	23	$B_1$	36
b	$B_1L_1$	53	$B_1L_1$	145
c	$B_1B_2L_1L_2$	69	$B_1B_2L_1L_2$	98
d	$T_3T_7$	53	$T_3T_7$	79
e	$A_3$	59	$A_3$	82
f	$L_5L_7B_7B_8T_7T_8$	97	$L_5L_7B_7B_8T_7T_8$	148
g	$L_1L_2L_7L_8$	87	$L_1L_2L_7L_8$	135

In order to directly prove the performance of the improved algorithm, b was taken as an example to simulate the experiment. Figure 2 shows the simulation results. The results of the optimal adaptability value and the average adaptability value were compared. And the average adaptability value is higher. In the later period, the algorithm can converge to the optimal solution nearby.

In order to further verify the optimization performance of the algorithm, the algorithm in this paper was compared with the immune optimization algorithm, b. and c. were taken as examples to optimize, and then the performance of the proposed algorithm and the traditional immune algorithm in the different specifications of the smart grid fault optimization was compared. The results are shown in Figs. 3 and 4.

From the simulation in Fig. 2, it can be found that, compared with the traditional immune algorithm, the integration of the chaotic algorithm has the following advantages:

- (1) It can strengthen the global search ability of the algorithm.
- (2) It can improve the mutation operator, which can make the mutation operation

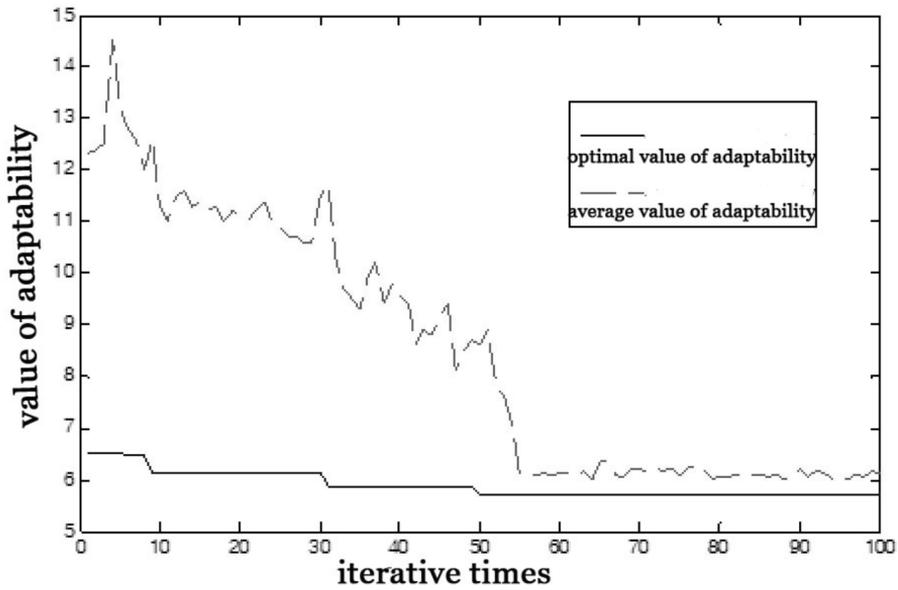


Fig. 2. Convergence curve of the algorithm

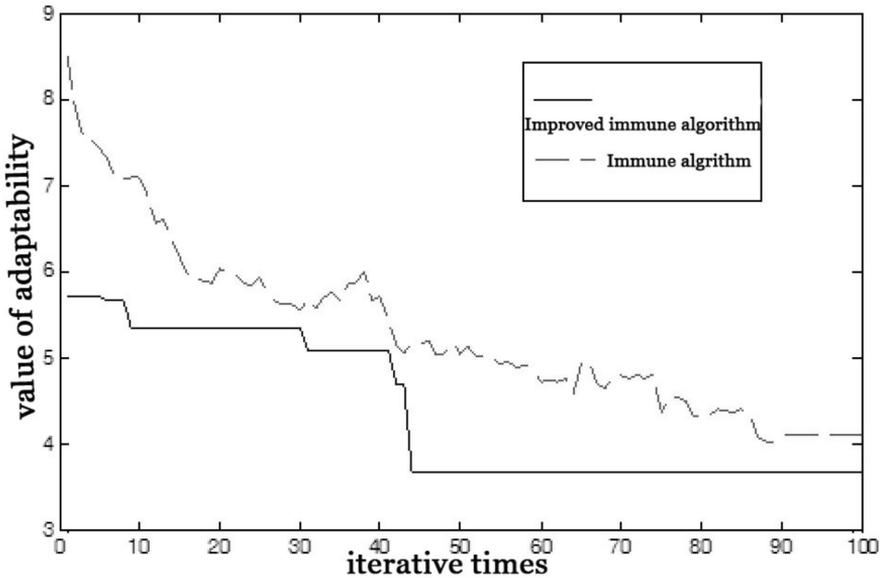


Fig. 3. Convergence of algorithm (example b.)

in the vicinity of the current optimal solution speed up the convergence rate, thus improving immune algorithm in the optimization efficiency.

Compared with the traditional immune algorithm, the integration of the chaotic

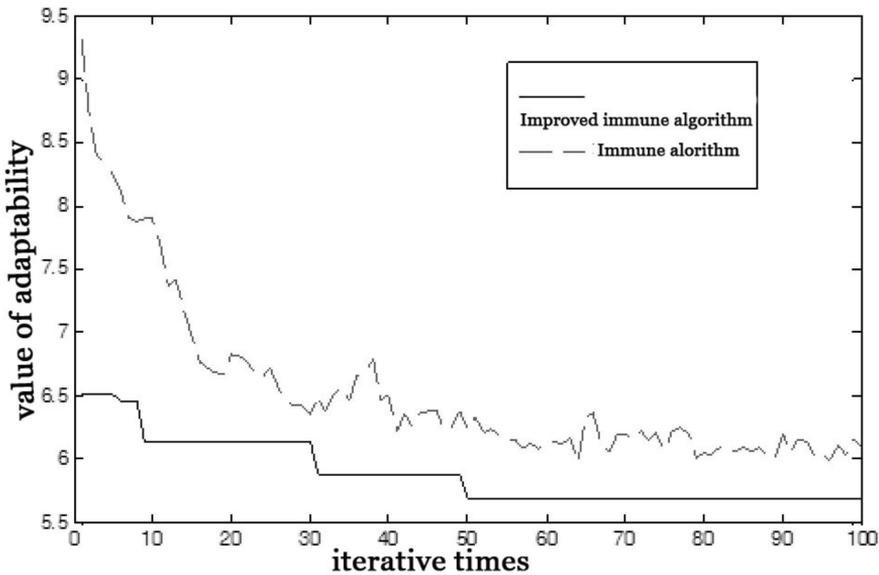


Fig. 4. Convergence of algorithm (example c.)

algorithm can exhibit a good performance, which can ensure a faster convergence to the optimal solution, so as to find a better solution.

### 4. Conclusion

In this paper, an improved immune algorithm was proposed to solve the problem of smart grid fault diagnosis, and the idea of chaos algorithm was put forward. The current optimal solution of the iteration of the algorithm will not fall into local optimization and enhance the global optimization ability of the algorithm, and a new mutation operator can make the mutation operation in the vicinity of the current optimal solution, so that the improved algorithm not only has the ability of global optimization, but also improves the convergence speed. Finally, in this study, the optimization of the power network diagnosis problem was compared with the traditional immune algorithm, and the effectiveness of the algorithm was proved.

### References

- [1] T. L. ZANG, Z. Y. HE, C. W. LI, Q. Q. QIAN: *Fault section estimation in transmission network based on binary swarm intelligence algorithm*. Power System Protection and Control 38 (2010), No. 14, 16–22.
- [2] J. H. WEN, S. H. LEE, Y. F. HUANG, H. L. HUNG: *A suboptimal PTS algorithm based on particle swarm optimization technique for PAPR reduction in OFDM Systems*. EURASIP Journal on Wireless Communications and Networking (2008), No. 14, paper 601346.

- [3] X. F. LIU, M. F. PENG: *Grounding grids fault diagnosis based on PCA-BP neural network*. Advanced Materials Research 516–517 (2012), 1774–1778.
- [4] J. B. XIAHOU, Q. Q. WEI, X. N. DENG, X. W. LIU: *Research and optimization of materialized views selection algorithm based on the data warehouse*. Advanced Materials Research 926–930 (2014), 3165–3170.
- [5] W. LU, B. LIANG, Y. CHENG, D. MENG, J. JANG, T. ZHANG: *Deep model based domain adaptation for fault diagnosis*. IEEE Transactions on Industrial Electronics 64 (2017), No. 3, 2296–2305.
- [6] W. U. NING, X. U. YANG, L. U. YUPING: *New fault section location algorithm for distribution network with DG*. Automation of Electric Power Systems 33 (2009), No. 14, 77–82.
- [7] D. LIU, X. GU, H. LI: *Complete analytic model for fault diagnosis of power systems*. Proceedings of the CSEE 31 (2011), No. 34, 85–92.
- [8] X. LIU, X. H. ZHANG, J. YUAN: *Relevance vector machine and fuzzy system based multi-objective dynamic design optimization: A case study*. Expert Systems with Applications 37 (2010), No. 5, 3598–3604.
- [9] G. SONG, Z. MA, G. LI, X. JIN: *Phase current fault component based single-phase earth fault segment location in non-solidly earthed distribution networks*. International Transactions on Electrical Energy Systems 25 (2015), No. 11, 2713–2730.

Received July 6, 2017

# Leakage monitoring algorithm of water supply pipe network based on information entropy difference<sup>1</sup>

ZHIXIA HAN<sup>2</sup>

**Abstract.** In order to improve the performance of leakage monitoring and reduce the constraint of single discrimination parameter in unit process, this paper proposes a leakage monitoring algorithm for water supply pipe network based on information entropy difference. Firstly, the Kalman set is used to predict the fault quickly. Secondly, the parameter fault coverage is introduced. The information entropy difference is used to filter the fault. Finally, the entropy difference of the parameter fault information is defined to complete the source leakage monitoring. The simulation results show that the fault set predicted by this algorithm has compressibility, and the set of faults after screening retains the real fault, and has higher fault detection rate and lower false alarm rate.

**Key words.** Information entropy difference, leakage monitoring, Kalman set, information entropy.

## 1. Introduction

As the amount of data carried in the network increases, the survivability of the network is increased and the loss caused by the network fault is reduced. This is of great significance. In order to realize the timely recovery of network fault-missing service, fast and accurate leakage monitoring and monitoring mechanism is needed. Therefore, as an essential part of survivability research, high performance monitoring algorithm has been a hot research topic both at home and abroad [1, 2].

The alarm obtained by the monitoring module in the network can be regarded as an external symptom of the fault. According to the collected symptom set, the most likely water supply network leakage collection can be predicted, and the root

---

<sup>1</sup>This work was support by the Industrial Research Project of Shaanxi Science and Technology Agency (2014K05-47). The Research of Leakage Monitoring Application Technology of Water Supply Pipe Network Based on Complex System Theory.

<sup>2</sup>Electronics & Electric Engineering College, Baoji University of Arts and Sciences, Baoji, Shaanxi, China, 721016

cause fault can be detected. In order to improve the monitoring performance, a fault propagation model must be introduced to represent the causal relationship between the fault and the symptom, because one symptom may correspond to multiple faults.

Based on the traditional monitoring methods such as monitoring loop, monitoring trace and monitoring tree [3–7], good monitoring performance can be achieved. However, the need to use additional detection wavelength to achieve leakage monitoring, monitoring costs are high. Therefore, in order to reduce the consumption of network resources for leakage monitoring, the leakage monitoring algorithm using traffic capture symptom in the network has been extensively studied. In this kind of algorithm, the leak detection algorithm based on Kalman set can realize the accurate leakage monitoring in the network even if the symptom is not complete, which has great advantage. The fault propagation model based on Kalman set is used in literature [8, 9]. The leakage detection is realized by the approximate reasoning algorithm. The monitoring performance is better, but the computation complexity is higher. In the paper [10], a simplified Kalman set is used as the fault propagation model. A discriminating parameter is defined as the criterion of fault diagnosis, which reduces the computational complexity and achieves better monitoring performance. Because of the limited parameters, it is difficult to achieve more accurate leakage monitoring using a single discriminating parameter.

In order to improve the performance of the monitoring algorithm in the bipartite fault propagation model, an information entropy difference based on water supply pipe network leakage detection (IWLD) is proposed in this paper. The algorithm divides the monitoring process into three modules: fault prediction, screening and monitoring. First, the prediction module quickly finds the largest possible fault set corresponding to the symptom. Then, the filter module introduces parameter fault coverage to convert the maximum possible fault set into the signal, and uses the information entropy difference (CS) method to eliminate the redundancy of the signal. Finally, the entropy difference (ED) of the parameter information is defined to identify the root cause fault. The fault detection and monitoring are completed by several parameters respectively, and the monitoring performance is improved.

## 2. Fault information entropy difference monitoring algorithm

### 2.1. *Fault prediction module*

Probabilistic Weighted Bipartite Graph (PWBG) is chosen as the fault propagation model in order to achieve fast fault prediction and accurate leakage monitoring, as shown in Fig. 1. According to PWBG, it is possible to quickly and accurately identify all possible faults associated with the sign in the symptom set, get the maximum possible fault set.

The maximum possible failure set is the set of all possible faults associated with the symptom set.

Define Redundancy rate to represent percentage of the number of redundant

faults in the largest fault collection, which is defined as shown in expression

$$R(H_{\text{MaX}}) = \frac{|H_{\text{R}}|}{|H_{\text{MaX}}|} = 1 - \frac{|F_{\text{C}}|}{|H_{\text{MaX}}|}. \tag{1}$$

Here,  $H_{\text{MaX}}$  represents the largest possible set of failures,  $H_{\text{R}}$  represents the set composed of redundant faults in  $H_{\text{MaX}}$ , and  $F_{\text{C}}$  represent the set of real failures in the network.

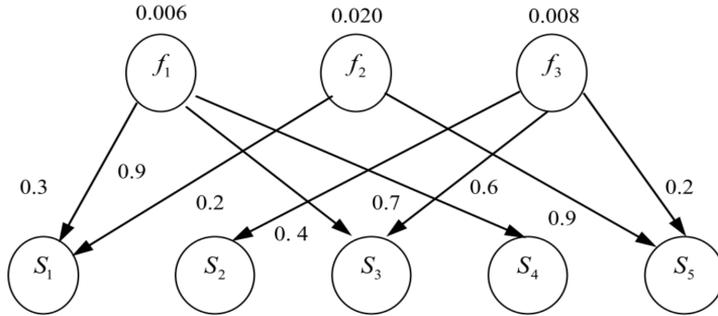


Fig. 1. Probabilistic Weighted Bipartite Graph (PWBG)

The maximum number of possible fault sets is always greater than the number of faults actually occurred on the network, i.e., most of the failures are not actually occurring. Therefore, it is necessary to filter the failures in the set, filter out the possibility of less likely failure, get fewer possible number of elements of the fault set, the maximum possible elimination of redundant faults on the monitoring algorithm to achieve more accurate leakage damage monitoring.

**2.2. Information entropy difference fault screening algorithm**

In order to reduce the redundancy and reduce the influence of redundancy fault on fault judgment, we must eliminate redundant faults in the largest fault collection as much as possible, and select the maximum fault set, screening out the fault of larger possibility, taking the failure as a signal, the introduction of coverage as the signal strength of the fault, the fault screening problem will be converted into signal processing problems. Information entropy difference method can be used as a signal processing method to achieve important information to retain the signal and remove part of the signal or all the redundant information purposes, to achieve fault collection screening.

Workflow of information entropy difference is shown in Fig.2. Define the signal strength of the corresponding fault signal as the important information, and the signal strength of the redundant fault corresponding signal as redundant information.

(1) The first step of the entropy difference is to verify the compressibility of the signal, i.e. to show that the processed signal contains redundant information. The maximum fault set contains redundant faults with high redundancy, and the

corresponding signals contain more redundant information and are compressible. Therefore, it is reasonable to use the information entropy difference method to deal with the fault signal.

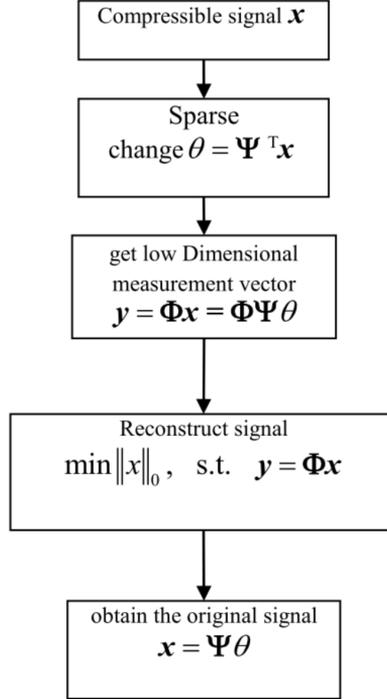


Fig. 2. Information entropy difference workflow

(2) The second step of the information entropy difference is to obtain the observation vector value of the signal according to the observation model. The observation model mainly uses the perceptual matrix  $\Phi$  to project the signal and obtain the observed vector value of the signal. It is the key to improve the filtering performance of the perceptual screening method by designing the appropriate sensing matrix so that the observed vector values contain both the important information of the original signal and the redundant information as much as possible.

Set the signal intensity threshold  $\alpha_{SI}$ , the design of the sense matrix, to retain fault signal of the original signal ( $x = (x_1, x_2, \dots, x_n)$ ), whose signal strength is higher than the threshold. Here,  $\alpha_{SI}$  is obtained from (2), in which,  $\mu$  ( $0 \leq \mu \leq 1$ ) is the scale factor, the size of  $\alpha_{SI}$  being under flexible control.

$$\alpha_{SI} = \mu \cdot \text{Max} \{x_1, x_2, \dots, x_n\}. \quad (2)$$

Since the diagonal matrix is multiplied by the target matrix, the size of the elements within the target matrix can be scaled. Thus, a diagonal matrix  $\mathbf{A}$  ( $\mathbf{A} = \text{diag}(a_1, a_2, \dots, a_n)$ ) may be introduced as the perceptual matrix  $\Phi$  of the original signal  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ . The diagonal matrix element values are given by equation

(3) below. According to the designed perceptual matrix  $\Phi = \mathbf{A}$ , the observation vector  $\mathbf{y}$  of the original signal  $\mathbf{x}$  can be obtained by equation (4).

$$a_i = \left\{ \begin{array}{l} 1, x_i \geq \alpha_{\text{SI}} \\ 0, x_i < \alpha_{\text{SI}} \end{array} \right\}. \quad (3)$$

(3) The third step of the information entropy difference is to reconstruct the original signal value according to the observation vector  $\mathbf{y}$ . As shown in equation (2), in all cases which meet  $\mathbf{y} = \Phi \mathbf{x}$  to find the sparsest characteristics of the signal  $\mathbf{x}'$  is the demand, which meet the requirements of the collection,  $\mathbf{x}'$  has the minimum number of non-zero elements.

The set of fault components corresponding to non-zero elements in the reconstructed signal  $\mathbf{x}'$  is called the filtered fault set  $H_S$ . Use the filtered fault set  $H_S$ .

The true fault coverage  $\eta(H_S)$  and redundancy of the set  $R(H_S)$  is used to observe the performance of the screening algorithm. The calculation of  $\eta(H_S)$  and  $R(H_S)$  is shown in equation (4) and equation (5),  $H_{RS}$  being a set of redundancy fault components in  $H_S$  and  $F_{CS}$  being a set of real faults in  $H_S$ .

$$\eta(H_S) = \frac{|H_S| - |H_{RS}|}{|F_C|} = \frac{|F_{CS}|}{|F_C|}, \quad (4)$$

$$R(H_S) = \frac{|H_{RS}|}{|H_S|} = 1 - \frac{|F_{CS}|}{|H_S|}. \quad (5)$$

The larger the number  $\eta(H_S)$ , the more important information that  $H_S$  represents the original fault collection  $H_{\text{Max}}$  will contain. Expression  $\eta(H_S) = 1$  means that all important information of  $H_{\text{Max}}$  is retained in  $H_S$ . In contrast to  $R(H_{\text{Max}})$ , the smaller  $R(H_S)$  is, the less number of redundant faults, the better the performance of the filtering algorithm, and when  $R(H_S) = 0$ , there is no redundant fault in  $H_S$ .

### ***2.3. Leakage monitoring module based on fault information entropy difference monitoring algorithm***

Because there may be multiple faults with maximum coverage, the possible fault set after filtering contains multiple possible failures. At this point, the use of coverage as a discriminating parameter for leakage monitoring is highly likely to lead to misjudgment of the situation. Therefore, it is necessary to introduce new reasonable parameters for leakage monitoring. This module defines the information entropy difference  $\Delta H(f)$  of the fault  $f$  and uses it as the discriminating parameter. A leakage monitoring algorithm based on information entropy difference is proposed. The following theoretical analysis shows that the use of information entropy difference  $\Delta H(f)$  as a discriminating parameter is reasonable.

When a symptom occurs in the network, it will provide a certain amount of information about the fault associated with it. Information Entropy is the average amount of information provided by a number of indications associated with a failure. The greater the amount of information is, indicating that the greater the uncertainty

of the variable is the failure to determine the probability of failure that will be less likely.

In order to analyze the reasonability of fault information entropy difference as the parameter of fault judgment, the following two parameters are defined: ideal information entropy and actual information entropy.

The ideal information entropy of the fault  $f$  is the information entropy of the fault  $f$  in the ideal case, that is, all the symptoms associated with the fault  $f$  occur. The calculation of the ideal information entropy  $H_1(f)$  is shown in the equatio.

$$H_1(f) = - \sum_{s_i \in S(f) \cap S_O} p(f|s_i) \log p(f|s_i). \quad (6)$$

Here,  $S(f)$  is a corresponding set of signs to  $f$ , the network can all be shown outside the signs of the set  $S_O$ , that  $p(f|s_i)$  can be obtained through the Bayes formula

$$p(f|s_i) = \frac{p(f)p(s_i|f)}{\sum_{f_j \in F} p(f_j)p(s_i|f_j)}. \quad (7)$$

The actual information entropy of the fault  $f$  is the information entropy gathered by  $f$  the symptom set  $S_N$  in the real situation. The calculation of the actual information entropy  $H_2(f)$  is shown in the expression

$$H_2(f) = - \sum_{s_I \in S(f) \cap S} p(f|s_i) \log p(f|s_i). \quad (8)$$

The smaller the information entropy difference  $\Delta H(f) = H_1(f) - H_2(f)$  of the fault  $f$  is, the closer the greater the probability of occurrence of the fault  $f$  to the ideal condition is, therefore, the information entropy difference  $\Delta H(f)$  of the fault  $f$  can be used as a parameter in the leakage monitoring judgment.

#### 2.4. Simulation and result analysis

In this paper, a fault prediction module and a fault screening module are added to the proposed algorithm. Among them, the failure prediction module quickly predicts the maximum possible fault set  $H_{\text{Max}}$ , and calculates the redundancy. Fault filtering module outputs possible filtered fault sets  $H_S$  and filters algorithm performance. When the scale factor  $\mu = 0$ , that  $H_{\text{Max}}$  is not actually screening for treatment; at that time  $0 < \mu \leq 1$ ,  $H_{\text{Max}}$  was the corresponding screening to be filtered after the possible failure of the collection  $H_S$ . The algorithm is IWLD, and the comparison algorithm is MCA [10] and BSD [11].

In order to observe the performance of the algorithm in different random networks, 10 random networks are generated, and 50 valid single fault cases are generated in each network. The input of each case algorithm is  $S_i$ , the output is the fault hypothesis collection  $H$ , the detection rate  $\text{DR}(S_i)$ ,  $\text{DR}(S_i) = |H \cap F_C| / |F_C|$ , and the false positive rate  $\text{FPR}(S_i)$ ,  $\text{FPR}(S_i) = |H - F_C| / |H|$ . In addition, in order to observe the performance of the filter module, this algorithm outputs the maximum

possible fault set redundancy  $R_{Si}(H_{MaX})$ , possible fault set redundancy  $R_{Si}(H_S)$  and fault coverage  $\eta_{Si}(H_S)$  under different scale factors.

Suppose the number of cases in each network is  $n$ , the output fault detection rate  $DR(Net_i) = \sum_{j=1}^n DR(S_j)/n$ , the error detection rate  $FPR(Net_i) = \sum_{j=1}^n FPR(S_j)/n$ , the algorithm outputs the maximum possible fault set redundancy  $R_{Net_i}(H_{MaX}) = \sum_{j=1}^n R_{S_j}(H_{MaX})/n$ , the probability of failure aggregation under different scale factors  $R_{Net_i}(H_S) = \sum_{j=1}^n R_{S_j}(H_S)/n$ , fault coverage  $\eta_{Net_i}(H_S) = \sum_{j=1}^n \eta_{S_j}(H_S)/n$ .

Suppose that the number of random networks is  $m$ , the final output is: fault detection rate  $DR = \sum_{i=1}^m DR(Net_i)/m$ , fault detection rate variance  $VDR = \sum_{i=1}^m \{DR(Net_i) - DR\}^2/m$ , fault false detection rate  $FPR = \sum_{i=1}^m FPR(Net_i)/m$ . In addition, the proposed algorithm can output the maximum possible fault set redundancy  $R(H_{MaX}) = \sum_{i=1}^m R_{Net_i}(H_{MaX})/m$ , the possible failure set redundancy under different scale factors  $R(H_{MaX}) = \sum_{i=1}^m R_{Net_i}(H_{MaX})/m$  and fault coverage  $\eta(H_S) = \sum_{i=1}^m \eta_{Net_i}(H_S)/m$ .

**2.5. Simulation results and analysis**

Figure 3 shows the filtered fault set redundancy degree under different scale factors. For  $\mu = 0$ , the value of  $R(H_{MaX})$  is 83.34% ~ 94.67%, the mean 90.84%. For  $\mu = 0.2$ , the value of  $R(H_{MaX})$  is 80.79% ~ 88.98%, with an average of 86.70%. For  $\mu = 0.4$ , the value of  $R(H_{MaX})$  is between 70.10% and 79.49%, the mean value is 76.77%. For  $\mu = 0.6$ , the value of  $R(H_{MaX})$  is 55.01% ~ 67.53%, the mean is 63.50%. For  $\mu = 0.8$ , the value  $R(H_{MaX})$  was between 42.22% and 56.26% with the mean value of 50.56%. Finally, for  $\mu = 1$ , the value of  $R(H_{MaX})$  is in the range 34.41% ~ 41.35%, the mean is 37.97%.

It can be seen that  $H_{MaX}$  has a high redundant fault, the corresponding signal contains more redundant information is with compressibility. In the screening algorithm, with the increase of the scale factor, the redundancy of the possible fault set is reduced.

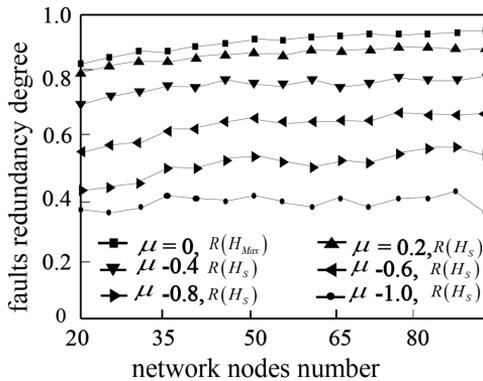


Fig. 3. Faults set redundancy degree after filtering at any scale factor

Table 1 shows the true fault coverage  $\eta(H_S)$  for the possible fault sets  $H_S$  after filtering. It can be seen from Table 1, in different network sizes, the value of  $\eta(H_S)$  is 1 and  $H_S$  can keep true failure. This is because the real fault always has the largest coverage, and at any scale factor, the real fault can always be preserved.

Table 1.  $H_S$  set true fault coverage ( $\eta(H_S)$ )

Network nodes	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90
$\eta(H_S)$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

### 3. Leakage monitoring performance analysis

Because when  $\mu = 1.0$ ,  $H_S$  not only has the smallest redundancy, but also can keep the true breakdown, has the best screening performance. Therefore, the algorithm in this paper under  $\mu = 1.0$ , conduct fault screening and monitoring, and get monitor the results.

Figure 4 comparisons for the three kinds of algorithms fault detection rate. As shown in the figure, the proposed IWLD algorithm fault detection rate from 94.2% ~ 97.4%, mean is 96.17%. BSD algorithm fault detection rate from 91.6% to 97.4%, mean is 94.6%. The failure detection rate of MCA algorithm is between 73.2% and 89.8%, mean value is 79.39%. It can be seen that the IWLD algorithm has the highest fault detection rate, the BSD algorithm has a lower fault detection rate and the MCA algorithm has the lowest detection rate.

Figure 5 shows the comparison of the three algorithms' fault detection rate variance in different networks, and shows the stability of the three algorithms' fault detection in different random networks. As shown in the figure, IWLD algorithm fault detection rate variance is between 0.00040 ~ 0.00276, mean 0.00095. BSD algorithm fault detection rate variance is between 0.00038 ~ 0.00264, mean 0.00138. MCA algorithm fault detection rate variance is between 0.00008 ~ 0.00546, mean 0.00299. IWLD algorithm and BSD algorithm are almost equal to zero. Overall, IWLD algorithm is more stable than BSD algorithm, that is, the stability of fault detection rate is higher in different random networks. In contrast, MCA algorithm stability is relatively low.

Figure 6 presents the three kinds of algorithm error rate comparison. As shown in the figure, IWLD algorithm fault error detection rate of 2.6% to 5.8%, mean value 3.83%. BSD algorithm error rate is between 16.16% ~ 23.25%, mean value is 20.65%. The error rate of MCA algorithm is between 10.2% ~ 27.2%, and the mean is 20.61%. IWLD algorithm has the lowest false alarm rate, BSD and MCA algorithms have higher false alarm rate. This is because IWLD algorithm has taken the initial screening of faults, to a large extent reduced the redundant fault on the impact of leakage monitoring.

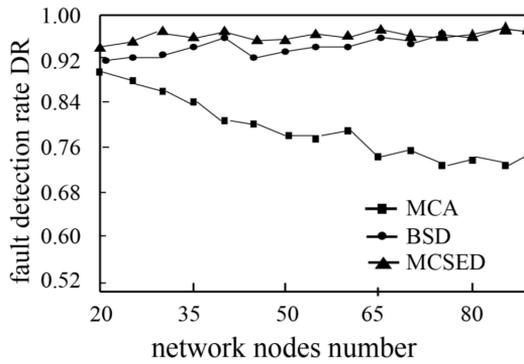


Fig. 4. Fault detection rate

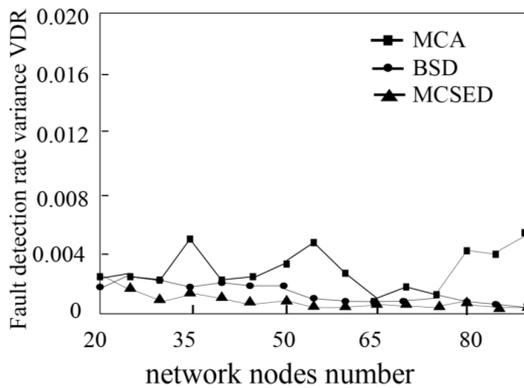


Fig. 5. Fault detection rate variance

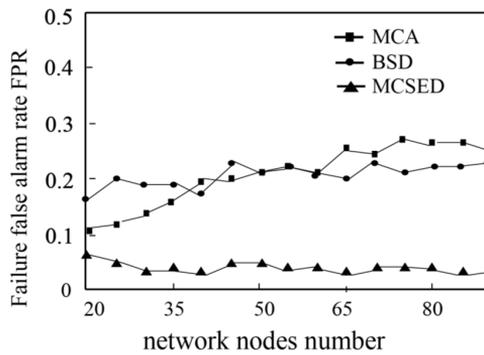


Fig. 6. Failure false alarm rate

## 4. Conclusions

In this paper, a water supply pipe network leakage monitoring algorithm based on information entropy difference is proposed. Firstly, the Kalman set of the algorithm is used in the prediction module to get the prediction result. Secondly,

the parameter fault coverage is introduced into the filtering module. The fault set is transformed into the signal, and the redundancy of the signal is eliminated by the information entropy difference. Lower redundancy fault set is got; finally, the parameter information entropy difference is defined as the criterion of fault detection, and the root fault collection is monitored. The simulation results show that the proposed algorithm can stably show high fault detection rate in different random networks and greatly reduce the false detection rate. In order to obtain more accurate monitoring performance, it is necessary to find a better combination of parameters in the monitoring algorithm of information entropy, which will be a further step of future research work.

## References

- [1] J. TAPOLCAI, P. HO, L./, RONYAI, P./, BABARCZI, B./, WU: *Failure localization for shared risk link groups in all-optical mesh networks using monitoring trails*. *Journal of Lightwave Technology* 29 (2011), No. 10, 1597–1606.
- [2] Y. XIONG, Z. Y. XIONG, D. P. WU, Y. LI: *Multi-fault aware parallel localization protocol for backbone network with many constraints*. *Photonic Network Communications* 24 (2012), No. 3, 210–218.
- [3] R. Y. WANG, J. F. CHANG, K. P. LONG, X. L. YANG, W. L. ZHU: *A fault detection and location mechanism for optical burst switching network*. *Journal of Optoelectronics Laser* 17 (2006), No. 12, 1477–1481.
- [4] R. Y. WANG, J. F. CHANG, K. P. LONG, H. CHANG, P. Z. ZHANG: *Performance and fault monitoring scheme of data channel for optical burst switching networks*. *Optical Communication Technology* 30 (2006), No. 08, 8–11.
- [5] A. HADDAD, E. E. DOUMITH, M. GAGNAIRE: *A fast and accurate meta-heuristic for failure localization based on the monitoring trail concept*. *Telecommunication Systems* 52 (2013), No. 2, 813–824.
- [6] E. HOSSZU, J. TAPOLCAI, L. RONYAI, P. SOPRONI, P. BABARCZI, P. H. HO: *Fast failure localization in all-optical networks with length-constrained monitoring trails*. *International Congress on Ultra Modern Telecommunications and Control Systems*, 3–5 Oct. 2012, St. Petersburg, Russia, *IEEE Conference Publications* (2005) 677–683.
- [7] J. TAPOLCAI, L. RONYAI, P. HO: *Link fault localization using Bi-directional M-trails in All-optical mesh networks*. *IEEE Transactions on Communications* 61 (2013), No. 1, 291–300.
- [8] R. BARCO, P. LÁZARO, V. WILLE, L. DIEZ, S. PATEL: *Knowledge acquisition for diagnosis model in wireless networks*. *Expert Systems with Applications*, Part: 1 36 (2009), No. 3, 4745–4752.
- [9] M. STEINDER, A. S. SETHI: *Probabilistic fault diagnosis in communication systems through incremental hypothesis updating*. *Computer Networks* 45 (2004), No. 4, 537 to 562.
- [10] X. H. HUANG, S. H. ZOU, L. W. CHU: *Internet Services Fault Management: Layering Model and Algorithm*. *Journal of Software* 18, (2007), No. 10, 2584–2594.
- [11] C. ZHANG, J. X. LIAO, X. M. ZHU: *Heuristic fault localization algorithm based on Bayesian suspected degree*. *Journal of Software* 21 (2010), No. 10, 2610–2621.

Received July 12, 2017

# Human resource data location privacy protection method based on prefix characteristics

YULONG QI<sup>1,2</sup>, ENYI ZHOU<sup>1</sup>

**Abstract.** With the arrival of the big data era, a large amount of human resource data location information is implicitly collected. They cross reference with the spatial and temporal data the user initiatively published, which has caused new human resource privacy disclosure problem in the big data era. The existing location privacy protection mechanism cannot effectively protect the privacy of the users due to the fact that it does not take into consideration that the implicitly collected spatial-temporal data can cross reference with the human resource location data that is initiatively distributed by the user. The privacy protection problem in the implicitly collected temporal and spatial data is defined and studied for the first time, and the privacy protection framework that takes the characteristics of the prefix into consideration is put forward. In particular, a nested loop algorithm which takes the prefix filter into account is proposed to discover the records in the implicitly collected temporal and spatial data that may disclose the privacy of the human resource data. And the dummy filling method based on the frequent moving prefix is put forward to eliminate these records. In addition, a more efficient reverse aprior algorithm and graph based dummy filling algorithm are put forward respectively. Finally, the proposed algorithm is fully tested in a number of real data sets. The experimental results show that these algorithms have relatively high protective effect and performance.

**Key words.** Implicit privacy, temporal and spatial data, privacy protection, human resource data.

## 1. Introduction

In the era of big data, with the development of the location technology, the location-based services are becoming increasingly popular, and the user's temporal and spatial data is distributed through all types of services. While the users take the initiative to publish their own temporal and spatial behaviors through the sign-in and other mobile social network services, a large amount of temporal and spatial data

---

<sup>1</sup>School of Management, Xi'an University of Architecture and Technology, Xi'an, Shaanxi, 710055, China

<sup>2</sup>Corresponding author

that records people's behaviors when they use mobile phones to make phone calls, send and receive short messages is implicitly collected by the mobile communication operators at the same time [1–2]. As the temporal and spatial data between the mobile phones and mobile communication operators is collected by the mobile phone base station automatically, these implicitly collected data is characterized by large data volume and containing the human behaviors, which has played a key role in the social issues such as the human resource distribution, as well as the important social applications such as the human resource referral [3–6], human resource level [7] and so on. However, these implicitly collected temporal and spatial data will exposes the sensitive privacy information of the user such as the personal identity, purpose of action, health status, interests and hobbies and many other aspects through the cross reference to the temporal and spatial data initially published by the user [8–9]. In recent years, with the enhancement of the human resource privacy concept and the soundness of the laws and regulations in the publishing and using of the data, it is necessary to first eliminate the records which may expose the privacy of the human resource data before the implicitly collected spatial and temporal data is used for scientific research and data mining.

In order to ensure that the human resource sensitive information is not compromised, a large amount of work for the protection of the temporal and spatial data privacy is committed to anonymize the temporal and spatial data that may expose the human resource sensitive information. For example, the  $k$  anonymity on the location and trajectory data and other methods can generalize the location records of the user at the specific time range and spatial region, so that the attacker cannot recognize the specific user in a certain time range and spatial region [10–12]. However, these methods do not take into account that the attacker can refer to the temporal and spatial data that is initially published by the user and find the records that can reveal the human resource data privacy from the implicitly collected temporal and spatial data. Therefore, the temporal and spatial data sets which are protected by these methods can still disclose the data of the human resource data privacy.

In order to meet the aforementioned challenges, in this paper, the implicit privacy which is independent of the data initially published by the user is defined on the implicitly collected data set, so as to ensure that no matter how much data initially published by the users is collected by the attacker, the spatial and temporal data sets after the privacy protection will not reveal the additional information.

In this paper, the first section mainly introduces the related technology of the privacy protection in the publishing of the temporal and spatial data. The second section introduces the implicit privacy problem and its prefix feature protection framework. The third section introduces the corresponding discovery and elimination algorithm. And the fourth section is the demonstration of the experimental results.

## 2. Algorithm and analysis

This section describes the algorithms of discovery and elimination of  $(\varepsilon, k)$  privacy disclosure, respectively.

## 2.1. Discovery of $(\varepsilon, k)$ privacy disclosure

Firstly, we introduce the nested loop algorithm of the discovery of  $(\varepsilon, k)$  privacy disclosure that takes the prefix filter into consideration, and then point out its defect in lack of efficiency, and put forward a more efficient reverse aprior algorithm.

According to the privacy protection parameters  $\varepsilon$  and  $k$ , the basic idea of the temporal and spatial data of the discovery of the  $(\varepsilon, k)$  privacy disclosure is the set composed of all  $k$  temporal and spatial points by enumeration, and check whether each combination is associated with a unique user. To this end, the prefix filter based nest loop (referred to as PF-NL for short) is used to implement  $k$  nested loop for the temporal and spatial points by enumerating the  $k$ -bit non-repeating numerical numbers, and conducts pruning in the enumeration using the prefix filter method. Firstly, we introduce the important properties of the  $(\varepsilon, k)$  privacy.

**Property 1.** Denote the set of all the temporal and spatial points that expose the  $(\varepsilon, k)$  privacy as  $U_k$ , and denote the set of all the temporal and spatial points with the size of  $k$  that can uniquely associate with moving prefix as  $u_k$ , then  $U_k = u_1 \cup \dots \cup u_k$ .

We skim the trivial proof of Property 1, which shows that, we can start from the set of temporal and spatial points with the size 1, until we enumerate all the sets of temporal and spatial points with the size not exceeding  $k$ . To this end, we number  $n$  temporal and spatial points from 1, and each temporal and spatial set with the size  $k$  can be regarded as a numerical number with  $k$  bits and scale  $n$ , and each of the same set of the temporal and spatial points has the unique representation method after sorted according to the sequence.

Thus, the set of temporal and spatial points with the size  $k$  ( $k > 1$ ) has prefixes with the length  $\{1, \dots, k-1\}$ . For example, we can denote the six effective temporal and spatial points  $p_{A,T_1}$ ,  $Sp_{A,(T_1+\varepsilon_1/2)}$ ,  $Sp_{B,T_1}$ ,  $Sp_{C,T_2}$ ,  $Sp_{D,T_2}$  and  $Sp_{merge}$  in Table 1 with the number 1 to 6, respectively. Considering the  $(\varepsilon = 0, k = 3)$  privacy, for the temporal and spatial point set  $\{1, 2, 3\}$  with the size 3, it has the prefix  $\{1, 2\}$ . It is known that  $|Sp_1 \cap Sp_2| = 0$ , therefore, the set of temporal and spatial points with this prefix will definitely not expose the  $(0,3)$  privacy. In the prefix filter method, we avoid enumerating the temporal and spatial point set containing such prefix.

Combined with the aforementioned basic ideas and prefix filter optimization method, we put forward the basic algorithm PF-NL that discovers the implicit privacy violation. In the algorithm PF-NL, we enumerate the temporal and spatial point set  $num$  with the size  $k$  in turn, let the maximum value that each corresponds to be denoted by *bound* array (Line 1). When the size of each moving prefix set included in a certain set of temporal and spatial points after intersection is 1, we add it into the set  $R$  of all the temporal and spatial points that violate the privacy requirements (Line 3 and Line 4). Thus, the process of generating a new set of new temporal and spatial points is the process of adding an element in the array (Line 8~Line 13). It is worth noting that, when we are checking whether a certain temporal and spatial point has exposed the  $(\varepsilon, k)$  privacy, the prefix that it may exist is calculated (Line 6), and the prefix is filtered out when a new set of temporal and spatial points is generated (Line 8). When the new temporal and spatial points cannot be generated, the algorithm PF-NL ends (Line 10).

Algorithm PF–NL uses the prefix filter to try to filter out the temporal and spatial point set with empty intersection in the enumeration, which has accelerated the searching process of the temporal and spatial point set that violates the  $(\varepsilon, k)$  privacy. But as many prefixes can be used in the filter, we cannot record each one, it has still conducted a lot of unnecessary work. Next, we will introduce the reverse aprior algorithm, making use of the breadth-first search idea, to avoid containing the relatively small temporal and spatial point set that does not violate the  $(\varepsilon, k)$  privacy in the search for a larger set of temporal and spatial points.

Algorithm 2 shows the reverse aprior (referred to as RA algorithm for short), in which we use  $u_i, z_i$  and  $c_1$  to represent the temporal and spatial point set that violates the  $(\varepsilon, k)$  privacy with the size  $i$ , cannot violate the  $(\varepsilon, k)$  privacy, and may violate the  $(\varepsilon, k)$  privacy in the larger set of temporal and spatial points.

In the RA algorithm, we first check all the temporal and spatial points, and add the temporal and spatial points that violate the  $(\varepsilon, k)$  privacy into  $u_1$ ; as the temporal and spatial points must contain the prefix, they are all added into  $c_1$  (Line 1).

Next, we consider the set composed of more temporal and spatial points. The relatively large temporal and spatial point set consist of the temporal and spatial point set with relatively smaller probability to violate the  $(\varepsilon, k)$  privacy (Line 5). As the set with the size  $i + 1$  has multiple combination methods, different combinations may lead to different sizes of the temporal and spatial point sets to be tested. In order to reduce the time overhead, we choose two sets of temporal and spatial points with the smallest Cartesian product to generate it (Line 4). In the same time, it is used to check whether the temporal and spatial point set that violates the  $(\varepsilon, k)$  privacy should not include smaller temporal and spatial point set that cannot violate the  $(\varepsilon, k)$  privacy (Line 6). In the end, violate the moving prefix sets contained in each temporal and spatial point in the temporal and spatial point set to solve the intersection, and return its size. The temporal and spatial point set with the size 1 is in violation of the  $(\varepsilon, k)$  privacy, and added into  $u_{i+1}$  (Line 6), while the set of temporal and spatial points with the intersection size of 0 will not violate the  $(\varepsilon, k)$  privacy, therefore is added into  $z_{i+1}$ , and removed from  $c_{i+1}$  (Line 8 and Line 9).

Example 1: For Table 1, we first check whether the set of temporal and spatial points which are composed of six effective temporal and spatial points is in violation of the  $(\varepsilon, k)$  privacy. Firstly, as  $Sp_1$  and  $Sp_2$  contains a set each,  $u_1 = \{1, 2\}$ ,  $c_1 = \{3, 4, 5, 6\}$ . Next, according to the temporal and spatial point set to be tested with the size 2 which is generated by the Cartesian product of  $c_1$  and  $c_1$ , that is  $\{\{3,4\}, \{3,5\}, \{3,6\}, \{4,5\}, \{4,6\}, \dots, \{5,6\}\}$ . After checking, the four pairs of temporal and spatial point set that violates the  $(\varepsilon, k)$  privacy is  $u_2 = \{\{3,4\}, \{3,5\}, \{4,6\}, \{5,6\}\}$ , while the temporal and spatial point set  $\{3, 6\}$  and  $\{4, 5\}$  does not contain the same moving prefix. Therefore, the set of temporal and spatial points containing them cannot violate the  $(\varepsilon, k)$  privacy. Hence they are removed from  $c_2$ . Finally,  $c_2 = \emptyset$ ,  $z_2 = \{\{3,6\}, \{4,5\}\}$  is obtained. Therefore, algorithm RA ends. It is worth noting that, in general, the larger the set of temporal and spatial points is, the more temporal and spatial point sets we need to check. However, in Example 1, when the size of the set of empty points increases, there is no increase in the set of

temporal and spatial points to be examined, which proves the effectiveness of the RA algorithm intuitively.

## 2.2. Elimination of the $(\varepsilon, k)$ privacy violation temporal and spatial data

In this section, we introduce the method of dummy filling to eliminate the temporal and spatial data that violates the  $(\varepsilon, k)$  privacy. In particular, we introduce the frequent moving prefix based dummy filling method that does not need to look for the temporal and spatial point set that violates the  $(\varepsilon, k)$  privacy and the graph based privacy disclosure elimination method that can realize relatively high data effectiveness. Regardless of the set of the temporal and spatial points that have been found to violate the  $(\varepsilon, k)$  privacy, a simple method of privacy protection is to perform dummy filling with the same user id to each user for each temporal and spatial point where it exists, for example, to protect the (0,2) privacy disclosure caused in Table 1. After such dummy filling, the following can be obtained  $Sp_{A,T_1} = \{u_1, u_5\}$ ,  $Sp_{A,(T_1+\varepsilon_1/2)} = \{u_2, u_6\}$ ,  $Sp_{B,T_1} = \{u_3, u_4, u_7, u_8\}$ ,  $Sp_{C,T_2} = \{u_1, u_3, u_5, u_7\}$ ,  $Sp_{D,T_2} = \{u_2, u_4, u_6, u_8\}$ ,  $Sp_{merge} = \{u_1, u_5, u_2, u_6\}$ . At this point, the temporal and spatial data for user  $u_5 \sim u_7$  has added 8 entries of dummy, and no longer in violation of (0, 2) privacy. However, this method requires filling 100% (or more) dummy. As an improvement, we add two moving prefix that occur the most frequently for each temporal and spatial point.

Algorithm 3 (frequent moving object, referred to as the FMO algorithm for short) shows the process of the dummy fixing based on the frequent moving object. Given the privacy parameter  $(\varepsilon, k)$  with uniqueness, we first find the most frequently occurred two moving prefix, and in this process there are a lot of fast algorithms; secondly, we add these moving prefix to each set of the temporal and spatial points

As the set of the temporal and spatial points of unique privacy disclosure found in Section 3.1 is not taken into consideration, the algorithm FMO performs dummy filling indiscriminately to the temporal and spatial points, resulting in the filling of a lot of data that is not necessary.

Example 2: For the following four temporal and spatial points:  $\{u_1, u_6\}$ ,  $\{u_2, u_3\}$ ,  $\{u_3, u_4\}$ ,  $\{u_7, u_8\}$ , according to the FMO algorithm, perform dummy filling and obtain  $\{u_1, u_6, u_2, u_3\}$ ,  $\{u_2, u_3\}$ ,  $\{u_2, u_3, u_4\}$ ,  $\{u_7, u_8, u_2, u_3\}$ , and 62.5 % of the data has been filled. However, the first temporal and spatial point clearly does not need to fill any dummy.

Taking the set of temporal and spatial point set that violates the  $(\varepsilon, k)$  privacy found in Section 3.1, algorithm 4 shows the graph based dummy filling (referred to as G-DF for short) process. In algorithm G-DF, we regard the human resource location data as a graph, in which each temporal and spatial point represents a node in the graph. If two temporal and spatial points exist in a set of temporal and spatial points that disclose the unique privacy, we add an edge to them in the graph (the second line). We only run algorithm 3 for each connected component in the graph, that is, looking for the most frequent two moving prefixes in each connected component, and then adding them to each node in the connected component (Line 3~Line 5).

Example 3: It is noted that in Example 2, only the set of temporal and spatial points  $\{\{u_2, u_3\}, \{u_3, u_4\}\}$  with the size 2 has violated the  $(\varepsilon, k)$  privacy, and Fig.1 is the case that Example 2 is converted into graph. In this way, according to algorithm G-DF, the data in Example 2 is converted into  $\{u_1, u_6\}, \{u_2, u_3\}, \{u_2, u_3, u_4\}, \{u_7, u_8\}$ , which has increased 62.5% more data compared with the FMO algorithm, while the G-DF algorithm has only added 12.5% of the data.

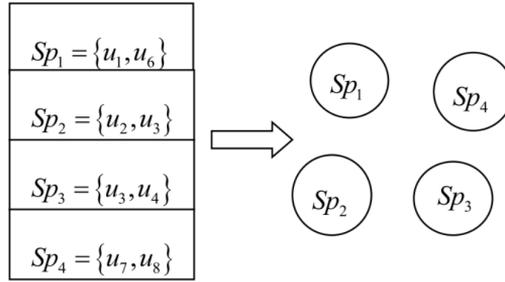


Fig. 1. Convert the temporal and spatial points into graph

### 3. Experiments

We use two real data sets to compare the performance and effectiveness of the two algorithms that discover the temporal and spatial data violating the  $(\varepsilon, k)$  implicit privacy and two algorithms that eliminate such violation in the prefix characteristic framework. In particular, we will answer the following questions through the experiment:

- (1) How the privacy protection parameter  $(\varepsilon, k)$  affect the  $(\varepsilon, k)$  privacy in the temporal and spatial data.
- (2) The influence of the privacy protection parameter  $(\varepsilon, k)$  on the performance of the privacy violation discovery algorithm.
- (3) The effect and performance of the privacy protection parameter  $(\varepsilon, k)$  on the privacy protection algorithms.

#### 3.1. Experimental environment and data set description

We use Java 1.7 to implement Algorithm 1 ~ Algorithm 4 in this paper. The experimental environment is a Linux server, Intel Xeon E5645 2.4 GHz processor, 128 G RAM, and 1T SATA hard disk.

In addition to the PF-NL method and RA method applied to discover the temporal and spatial set that violates the  $(\varepsilon, k)$  privacy in this paper, as well as the FMO and G-DF methods in this paper that are used to eliminate the temporal and spatial point set that violates the  $(\varepsilon, k)$  privacy. And other comparative methods include the following:

YCWA[3]: This method is the latest method that adopts the trajectory

anonymization technology to protect the privacy of the temporal and spatial data. It divides the trajectories into dwell point, and protects the privacy information by anonymizing these dwell points. This method mainly focuses on the performance of the trajectory privacy protection.

This method focuses on the data availability of the trajectory anonymity technology. While anonymizing the trajectory, it minimizes the distance between the changed temporal and spatial points and the original temporal and spatial points at the same time.

We use two public data sets, GeoSocial [5] and GeoLife [6] as the implicitly collected temporal and spatial data sets to conduct experiment. Their data size and the number of users are shown in Table 2.

Table 1. Data set

Data set	Number of record entries	Number of moving prefix
GeoSocial	4 M	18 K
GeoLife	20 M	17 K

### 3.2. Comparison of the privacy protection effect

Figure 2 shows the effects of the spatial and temporal data privacy protection for each method on the GeoSocial and GeoLife data sets under relatively stringent  $(\varepsilon, k)$  privacy condition ( $\varepsilon_1 = 10$  min,  $\varepsilon_2 = 1$  km,  $k = 10$ ). Our method RA \* G-DF uses the RA method that has the best performance in the discovery phase, and the G-DF method that has the best performance in the elimination phase. We can see that, there are still a large number of temporal and spatial sets that violate the  $(\varepsilon, k)$  privacy in the trajectory after the YCWA and SQL-ANON treatment. This is because these methods do not take the privacy disclosure caused by the cross-reference of the implicitly collected temporal and spatial data and the user initially published temporal and spatial data into consideration.

### 3.3. Performance of the algorithm that discovers the $(\varepsilon, k)$ privacy

We compare and verify the running efficiency of the two proposed  $(\varepsilon, k)$  implicit privacy discovery algorithms with the real data sets GeoLife and GeoSocial. Figures 3–5 show the run time of the two algorithms in the GeoSocial data set to find all the  $(\varepsilon, k)$  privacy violation in different  $(\varepsilon, k)$ . It can be seen from the comparison of Fig. 3 and Fig. 4, the RA algorithm is 1~2 orders of magnitude faster than the PF-NL algorithm under the same privacy parameter with the same privacy parameter  $(\varepsilon, k)$ . Particularly, PF-NL algorithm cannot even calculate the situation when  $k > 3$ .

For the RA algorithm, we have adjusted the GeoLife data set size and the number of the moving prefix and tested its performance. Figure 3 (c) shows that, the size of GeoLife data set has the greatest impact on the algorithm, as it has changed the number of the temporal and spatial points in the data set; however, the number of

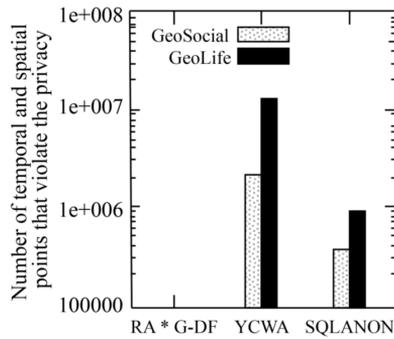


Fig. 2. Comparison of privacy protection effect

prefix has little effect on the algorithm.

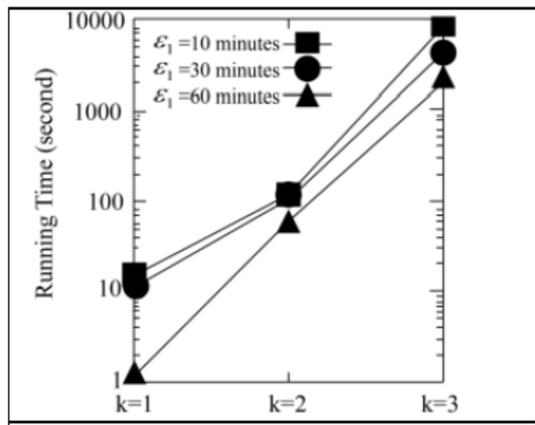


Fig. 3. Performance of PF-NL algorithm on GeoSocial data set ( $\epsilon_2 = 1$  km)

## 4. Conclusion

In this paper, the definition of the  $(\epsilon, k)$  implicit privacy in the temporal and spatial data is proposed for the first time in view of the situation of the cross-reference of the data set initially published by the user and the temporal and spatial data set implicitly collected. And the prefix characteristic privacy protection framework is put forward. In particular, two highly efficient algorithms are proposed in this paper to discover the temporal and spatial point set that violates the  $(\epsilon, k)$  privacy. In addition, the dummy filling anonymous protective method is put forward in this paper. In order to improve the effectiveness of data, this paper further proposes the graph based dummy filling method. The full experiment on the real data set shows that, the proposed algorithm is highly efficient. In the future work, we will further improve the performance of the proposed method in the paper.

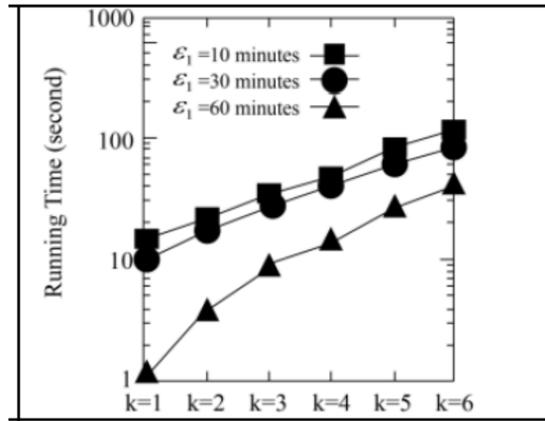


Fig. 4. Performance of RA algorithm on GeoSocial data set ( $\epsilon_2 = 1$  km)

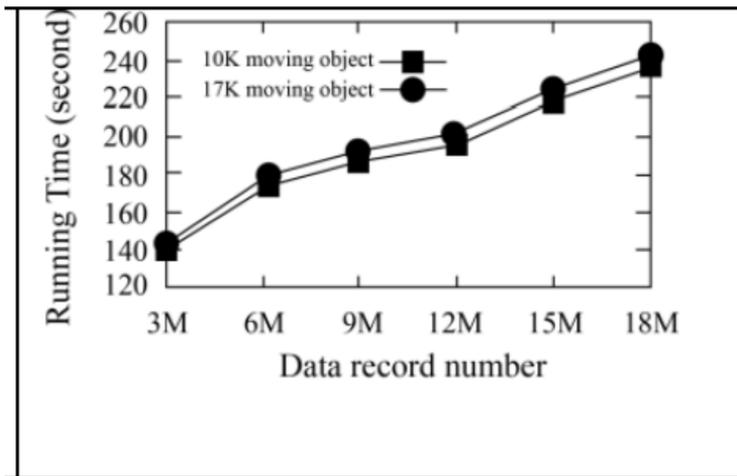


Fig. 5. Performance of RA algorithm on GeoSocial data set ( $k = 3$ )

## References

- [1] A. PENTLAND, D. LAZER, D. BREWER: *Using reality mining to improve public health and medicine*. Studies in Health Technology & Informatics 149 (2009), 93–102.
- [2] A. L. MENACH, A. J. TATEM, J. M. COHEN, S. I. HAY, H. RANDELL, A. P. PATIL, D. L. SMITH: *Travel risk, malaria importation and malaria transmission in Zanzibar*. Scientific Reports 1 (2011), Srep. No. 93, 1–7.
- [3] A. WESOLOWSKI, N. EAGLE, A. J. TATEM, D. L. SMITH, A. M. NOOR, R. W. SNOW, C. O. BUCKEE: *Quantifying the impact of human mobility on malaria*. Science 338 (2012), No. 6104, 267–270.
- [4] J. D. LILLY, D. A. GRAY, M. VIRICK: *Outsourcing the human resource function: Environmental and organizational characteristics that affect HR performance*. Journal of Business Strategies 22 (2005), No. 1, 55–72.
- [5] B. L. TATIBEKOV, J. S. ADAMS, N. A. PROCHASKA: *Characteristics of the labor mar-*

- ket and human resources management in the Republic of Kazakhstan*. *Advances in Competitiveness Research* 12 (2004), No. 1, 44–56.
- [6] M. WARNER: *Reassessing human resource management ‘with Chinese characteristics’: An overview*. *International Journal of Human Resource Management* 19 (2008), No. 5, 771–801.
  - [7] N. J. YUAN, Y. ZHENG, L. ZHANG, X. XIE: *T-Finder, a recommender system for finding passengers and vacant taxis*. *IEEE Transactions on Knowledge and Data Engineering* 25 (2013), No. 10, 2390–2403.
  - [8] S. B. WICKER: *Cellular telephony and the question of privacy*. *Communications of the ACM* 54 (2011), No. 7, 88–98.
  - [9] B. W. XU, L. XU, X. F. MENG, G. YU, Z. D. LU, Y. X. HE, J. Y. SHEN: *Web-based information systems and applications, A survey*. *Proc. Conference on Web Information System and Applications (WISA)*, 29–31 October 2004, Wuhan, China.
  - [10] Y. A. DE MONTJOYE, C. A. HIDALGO, M. VERLEYSSEN, V. D. BLONDEL: *Unique in the crowd, the privacy bounds of human mobility*. *Scientific Reports* 3, (2013), Article No. 1376.
  - [11] I. OZALP, M. E. NERGIZ, M. E. GURSOY, Y. SAYGIN: *Privacy-Preserving Publishing of Hierarchical Data*. *ACM Transactions on Privacy and Security* 19 (2016), No. 3, paper 7.
  - [12] A. E. CICEK, M. E. NERGIZ, Y. SAYGIN: *Ensuring location diversity in privacy-preserving spatio-temporal data publishing*. *The VLDB Journal* 23 (2014), No. 4, 609 to 625.

Received July 12, 2017

# High-throughput computing optimization of cloud computing platform based on Markov jump

DAFEI WU<sup>1</sup>

**Abstract.** Cloud computing platform transport (CCPT) is a new transport protocol based on Stream Control Transmission Protocol (SCTP), which uses the multi-destination feature of the SCTP protocol, and multiple paths in an end-to-end coupling. The parallel data transmission is carried out in parallel, and the bandwidth of the single-path transmission can be improved, and the end-to-end throughput can be enhanced. In this paper, the data transmission of the cloud computing platform is optimized according to whether the amount of transmission is affected by the received cluster value. The CCPT throughput model based on Markov jump association is analyzed in this paper, and the congestion window of the two phases is analyzed, and the congestion window is analyzed in the second stage of the timeout period, the slow start stage and the congestion avoidance stage. Secondly, we get a high-throughput function of RTT, RTO and packet loss in the data transmission process of a cloud computing platform, which can estimate the high throughput of data transmission protocol in the cloud computing platform. Finally, the result analysis verifies the validity and accuracy of the data transmission throughput model of the cloud computing platform.

**Key words.** Cloud computing platform data transmission, throughput model, transmission volume, Markov jump.

## 1. Introduction

Internet-based end-to-end transmission mainly uses TCP protocol. The new end-to-end transmission protocol SCTP (Stream Control Transfer Protocol) [1–2] uses the multihake (multihoming) and multi-flow (multistreaming) technology. TCP has a high end-to-end path fault tolerance and other advantages. Based on the SCTP cloud computing platform data transmission CCPT (Concurrent Multipath Transfer) [3] that is relative to SCTP end-to-end bandwidth aggregation and many other advantages [4], it can make full use of bandwidth resources, and it also presents a new challenge to the research and application of the basic theory of the Internet transport layer. In recent years, the research has also attracted the attention of

---

<sup>1</sup>Hunan University of Science and Engineering, Yongzhou, Hunan, 425199, China

scholars [5] because of its unique advantages, which has become the object of concern for the next generation in the network transport layer protocol [6].

The next generation of Internet in the cloud computing platform data transmission requirements are mainly reflected in the protection of a wide range of applications. It can be on the network resources for dynamic and reasonable scheduling and can use of the need to transfer the association of multiple paths as a whole. To take full account of the relevance of the path, for each path such as the bandwidth, delay, packet loss rate and other network attributes, especially the throughput, this paper should give high attention to how to establish a highly efficient and reasonable throughput transfer model in this paper as the focus.

In recent years, many scholars have given different throughput optimization methods for SCTP single path transmission characteristics, which are of some inspiration for establishing the multi-path transmission throughput model. As in [7], with the reference to traditional TCP throughput model [8], I first consider the mathematical expectation of the three phases of the slow start, the congestion avoidance and the overtime retransmission, and then I establish the throughput model of the SCTP [9]. The establishment of the model gives a more detailed discussion. In the literature [10–12], the whole model is divided into two parts from the perspective of Markov chain analysis.

Based on the analysis of the above research results, this paper proposes a CCPT high throughput optimization method in accordance with the Markov transition and the influence of the amount of transmission or the value of the received cluster. With the CCPTHOMJ (CCPT high-throughput optimization based on MarkTH jump, CCPTHOMJ), the TCTH data transfer time, the slow start phase and the congestion avoidance phase data transmission, the CCPTHOMJ is established in detail. Secondly, by comparing the calculation results of the model and the simulation results in NS2, it verifies the accuracy of the model and compares it with the existing model to verify the advanced nature of the model. Finally, the work is summarized and forecast.

## 2. CCPT's high-throughput algorithm optimization based on Markov jump

Based on the CCPT transmission model established in Section 2, the high-throughput calculation process of CCPT is derived here. For the description of Step 3 in the CCPT transmission model, the calculation process is based on the concept of Table 1 to give a description of the variables involved in the calculation and the related meaning.

### 2.1. Division of CCPT data transmission phase

The symbol  $S_i$  shows the  $i$ th first stage of the data transfer process, from the beginning of the first  $i$ th timeout period to the end of the  $i + 1$ th first timeout period. When denoting the time-out stage as  $Z_i^{TO}$ , the slow start phase as  $Z_i^{SS}$  and the congestion avoidance phase as  $Z_i^{CA}$ , the  $i$ th stage of the data transmission

process can be expressed by the formula

$$S_i - Z_i^{TO} + Z_i^{SS} + Z_i^{CA} \quad (1)$$

Table 1. Variable names and related meanings in CCPT high throughput models

Variable name	Meaning
<i>Sund</i>	The size of sender window
<i>q</i>	The probability for the successful transmission in each Data Chrnik [2]
<i>P</i>	Packet loss probability
<i>b</i>	Cumulative response factor
<i>SS</i>	Slow start
<i>CA</i>	Congestion avoidance
<i>TO</i>	Overtime
<i>B</i>	Throughput capacity
<i>QDP</i>	The stage between two packet dropping instructions in the congestion avoidance phase
<i>QD</i>	4 times repeated confirmation
<i>Q</i>	When the packet loss occurs, the probability of packet loss produces due to timeout
<i>SST</i>	Slow start threshold
<i>Cund</i>	Sender congestion window (CW)
<i>RTO</i>	Timeout retransmission time
<i>RTT</i>	Round-trip time

Symbol  $Y_i$  represents the number of packets transmitted in the  $S_i$  phase. This symbol also consists of three parts: the number of packets transmitted in the timeout phase denoted as  $Y_i^{TO}$ , the number of packets transmitted in the slow start phase denoted as  $Y_i^{SS}$ , and the number of packets transmitted in the congestion avoidance stage, which is denoted as  $Y_i^{CA}$ . The number of packets transmitted in the  $i$ th stage can be then expressed as

$$Y_i - Y_i^{TO} + Y_i^{SS} + Y_i^{CA} . \quad (2)$$

The handling capacity  $B$  can be expressed as

$$B - E[Y] / E[S] . \quad (3)$$

The change condition for data forwarding phase of CCPT is shown in Fig. 1.

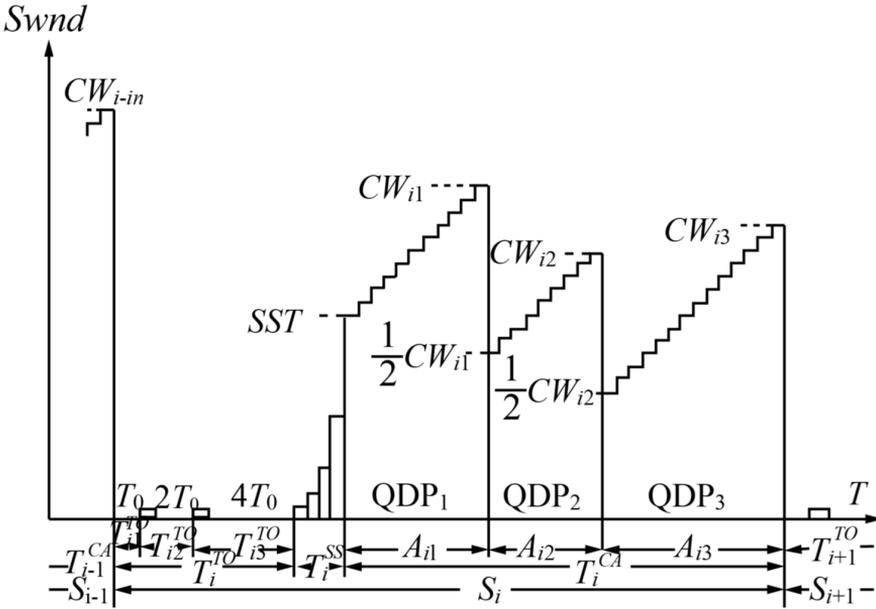


Fig. 1. Description of the window change in the CCPT data transmission phase

Because the amount of transmission ( $Swnd$ ) depends on the minimum value of both the congestion window ( $Swnd$ ) and receiver's cluster value ( $Rwnd$ ) at the end-to-end transmission environment, the derivation of the model is divided into two parts: Part 1: when the congestion window is less than the reception cluster value, the throughput model is not affected by the received cluster value. Part 2 is the throughput model when the congestion window is larger than the received cluster value. The throughput is affected by the received cluster value.

**2.2. Throughput model when the CCPT data transmission is not affected by the received cluster value (CCPTHOMJ)**

According to the Markov jump method, from the beginning denoted as  $t = 0$ , the sender has to send the data and define  $N_t$  as the number of packets transmitted as a  $[0, t]$  period of time. Quantity  $B_t$  is the throughput for that time and also for time  $t > 0$ . For the latter case

$$B_t = N_t/t. \tag{4}$$

We define the throughput expression for the long-term stability as follows:

$$B = \lim_{t \rightarrow \infty} B_t = \lim_{t \rightarrow \infty} \frac{N_t}{t}. \tag{5}$$

Figure 2 shows the window change in the congestion avoidance phase. The size of the dispatched volume of CCPT is  $Swnd$ . In the congestion avoidance

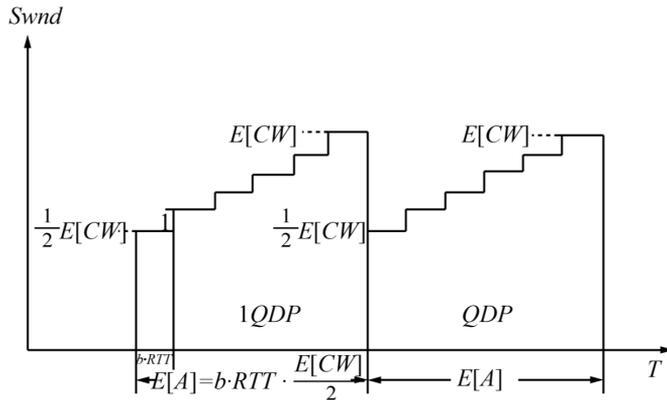


Fig. 2. Schematic diagram of the average window change in the congestion avoidance

phase, the sending end will increase the size of the SACK. If there are  $Swnd$  packets that are sent successfully in the  $i$ th time, the sender will receive  $Swnd/b$  SACK, assuming that the size of the window is  $Swnd'$  in the next round, so that

$$Swnd' = Swnd + \frac{1}{b}. \tag{6}$$

Therefore, if there is no packet loss,  $Swnd$  will grow linearly. The slope is  $1/b$ , and if the packet is detected, the window value will be reduced, depending on whether the packet loss is caused by repeated confirmation, or the overtime. If the packet is caused by a duplicate acknowledgment, it remains in the congestion avoidance. If the packet loss is caused by a timeout, the timeout period enters.

If  $H_{ij}$  is the number of  $j$  packets of  $QDP$  transmitted in the  $i$  phase, the congestion avoidance stage consists of  $j$   $QDP$ s:  $\Lambda_{ij}$  is the time of the  $j$ th phase and  $QDP$  stage and  $CW$  is the size of the congestion window at the end of each  $QDP$  phase. Quantity  $X_{ij}$  is the the packet loss occurred in the round in the  $i$ th phase and the  $QDP$  stage. Symbol  $b$  is the cumulative response factor and it is the confirmed number of SACK data packets. The congestion avoidance phase of the throughput is as follows:

$$B_{CA} = \frac{E[H]}{E[A]}. \tag{7}$$

In order to derive the  $B_{CA}$  expression, we first derive the expressions for  $E[H]$  and  $E[A]$ .

The  $i$  of the  $QDP$  stage is shown in Fig. 3.

The  $QDP$  stage begins from the instruction on packet loss. When the current window value is  $CW_{ij-1}/2$ ,  $QD$  occurs in half of the window. The value of the window increases in each time. When 1 increases in time  $b$ , we define  $a_{ij}$  as the lost data packet of  $QDP$ . Then, a packet is sent before the packet is detected, and  $CW_i - 1$  packet is sent on the  $a_{ij}$  time of occurrences and subsequent times, so that



Here,  $r_n$  is the random variable, and its size has nothing to do with the window size. It can be concluded that

$$E[A](E[X] + 1)E[r]. \quad (14)$$

In the following section, I can use the expression  $RTT = E[r]$  at the average time of  $RTT$ . We derive the  $E[X]$ th expression. Quantity  $CW_i$  is the function of number of rounds in order to simplify the model derivation. We assume that  $CW_{i-1/2}$  and  $X_i/b$  are integers. Firstly, we observe the  $i$ th  $QDP$  stage. The window value changes between  $CW_{i-1/2}$  and  $CW_i$ , the  $\frac{1}{b}$  is the linear slope. We can conclude that

$$CW_i = \frac{1}{2}CW_i + \frac{X_i}{b}, \quad i = 1, 2, \dots \quad (15)$$

### 3. Simulation results and model validation

To verify the accuracy and advancement of the proposed model (CCPTHOMJ), we use the CCPT module provided by the Protocol Engineering Laboratory of Delaware University in USA. Figure 4 shows the topological structure diagram of the simulation environment.

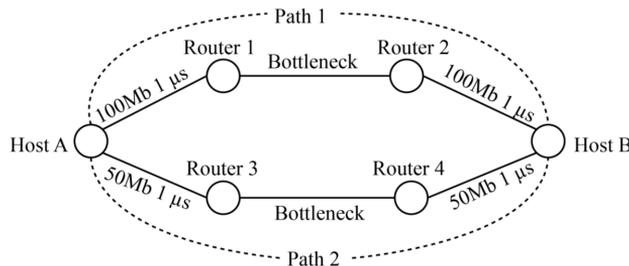


Fig. 4. Simulation topology

As shown in Fig.4, Host A and Host B represent CCPT terminals with dual NICs, respectively. To simulate various network scenarios, different bandwidths, latency, and packet loss rates are set here, with the design of [3] CUC, SFR and DAC algorithm. The data block is set to 1468 Bytes and is used the traditional file transfer protocol FTP for the application layer protocol. The following for this article is not affected by the received cluster value and received by the cluster value of the two cases. The high throughput model CCPTHOMJ of the CCPTH data transmission carries on the simulation experiment and the result analysis.

#### 3.1. Comparison with the simulation results and calculation results of CCPTHOMJ model affected by the received cluster value

Figure 5, upper part, depicts the packet loss rate of 1% and 5%, CCPTHOMJ model of the simulation results and the results of the calculation change by  $CW_{\max}$ .

The cumulative response factor is 1, the transmission round trip delay is the fixed value of 100 ms. From  $E[CW] = \sqrt{8/3bp}$  of expression (19), the average of the window size is  $E[CW] = 51.64$ . We set the  $CW_{max}$  of simulation results and calculation results of the 5 MSS (Maximum Segment Size), 10 MSS, 20 MSS, 30 MSS, 40 MSS and 50 MSS.

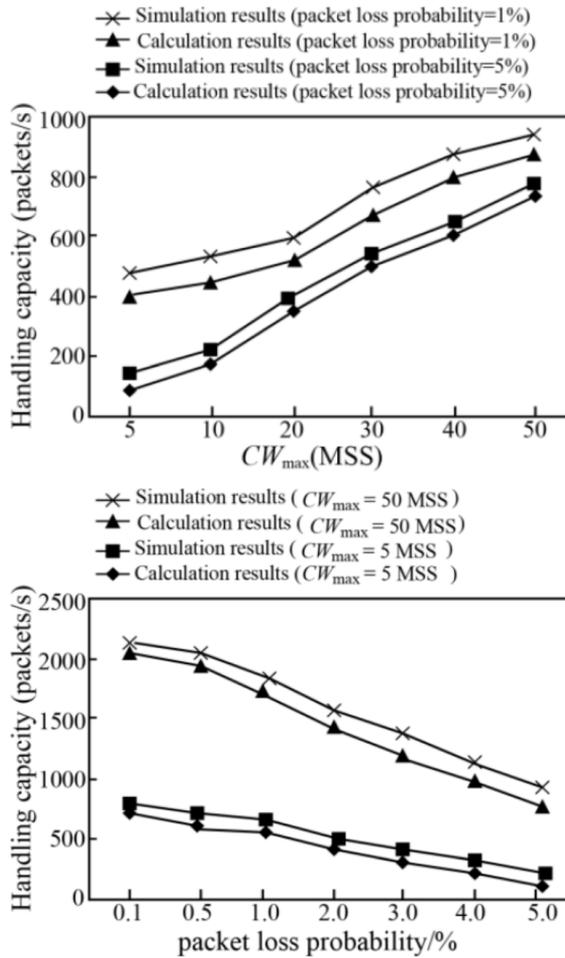


Fig. 5. Comparison of the simulation results and model results

It can be seen from Fig. 5, bottom part, that, when  $CW_{max}$  is 5, the simulation results of the model are basically the same as those of the calculated results, and the model can predict the handling capacity. The handling capacity increases with the growth of the packet loss rate. The simulation results are in good agreement with the calculated results, and the different packet loss rates also has a certain influence on the accuracy of the model. When the packet loss rate is 1%, the simulation

results are close to the calculation results.

When the packet loss rate increases to 5%, the separation distance of the two curves becomes larger and the accuracy of the model decreases, but on the whole, the calculation result of CM-ORM model can well reflect the change of handling capacity.

In order to verify the accuracy of CCPTHOMJ model in multi-angle, we analyze the fixed value and the delay of path transmission. The calculation result and simulation result of the model change with the packet loss rate. When the rates are 0.1%, 0.5%, 1%, 2%, 3%, 4%, and 5%, the approximate values of the path were respectively 50 MSS, 40 MSS, 30 MSS, 20 MSS, 10 MSS, and 5 MSS (but without losing the general meaning). The model simulation results and calculation results are compared, as shown in Fig. 5, bottom part.

Figure 5, bottom part, shows that when the packet loss rate is 0.1%, the model can better predict the throughput, but with the packet loss rate increases, the model simulation results and the results show a certain deviation. The  $CW_{\max}$  of different paths have the effect on the model accuracy. When  $CW_{\max}$  is 5 MSS, the simulation results of the model are not the same as those of the calculated results. When  $CW_{\max}$  increases to 50 MSS, the spacing between the two lines increases to a certain extent, and the accuracy of the model reduces to a certain extent. On the whole, the model of the calculation results can still be a good response to the changes in the trend of handling capacity.

The above analysis shows that the CCPTHOMJ model has high accuracy regardless of whether the amount of transmission is affected by the received cluster value or the amount of transmission is not affected by the received cluster value.

## 4. Conclusion

In this paper, it is firstly established a model of CCPT data transmission, and on the basis of this model we consider the influence of CCPT on handling capacity in different stages of congestion control, slow start and timeout retransmission, and propose a kind of cloud computing platform through the optimization of the handling capacity. According to whether the amount of transmission is affected by the value of the received cluster, the model is divided into two cases to obtain a high-throughput relationship with the packet loss rate. At the same time, we can verify the model accuracy, time delay, packet loss rate and simulation experiment. The comparison between the simulation results and the calculated results shows that the model designed in this paper can predict the trend of handling capacity and has a high accuracy. With the advanced model, the comparison is optimized in handling capacity of data transmission in the cloud computing platform.

## References

- [1] Y. JARARWEH, O. AL-SHARQAWI, N. ABDULLA, L. TAWALBEH, M. ALHAMMOURI: *High-throughput encryption for cloud computing storage system*. International Jour-

- nal of Cloud Applications and Computing 4 (2014), No. 4, 1–14.
- [2] F. ZHANG, J. CAO, K. HVANG, K. LI, S. U. KHAN: *Adaptive workflow scheduling on cloud computing platforms with iterativeordinal optimization*. IEEE Transactions on Cloud Computing 3 (2015), No. 2, 156–168.
  - [3] M. MINERVINI, C. RUSU, M. DAMIANO, : *Large-scale analysis of neuroimaging data on commercial clouds with content-aware resource allocation strategies*. International Journal of High Performance Computing Applications 29 (2015), No. 4, 473–488.
  - [4] J. R. IYENGAR, P. D. AMER, R. STEWART: *Performance implications of a bounded receive buffer in concurrent multipath transfer*. Computer Communications 30 (2007), No. 4, 818–829.
  - [5] J. LIU, L. BAI, C. YU: *Partial paths concurrent transfer for CMT with transport layer multihoming*. Journal of Computational Information Systems 10 (2014), No. 16, paper 7007.
  - [6] M. YABANDEH, S. ZARIFZADEH, N. YAZDANI: *Improving performance of transport protocols in multipath transferring schemes*. Computer Communications 30 (2007), No. 17, 3270–3284.
  - [7] J. HOFFMANN, I. WEBER, G. GOVERNATORI: *On compliance checking for clausal constraints in annotated process models*. Information Systems Frontiers 14 (2012), No. 2, 155–177.
  - [8] F. M. ANJUM: *TCP algorithms and multiple paths: Considerations for the future of the internet*. Information Systems Frontiers 6 (2004), No. 1, 91–104.
  - [9] F. PAGANINI, E. MALLADA: *A unified approach to congestion control and node-based multipath routing*. IEEE/ACM Transactions on Networking 17 (2009), No. 5, 1413 to 1426.
  - [10] P. KEY, L. MASSOULIÉ, D. TOWSLEY: *Multipath routing, congestion control and load balancing*. Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 15–20 April 2007, Honolulu, HI, USA, IEEE Conference Publications 4, (2007), IV 1341–IV 1344.
  - [11] J. Y. TEO, Y. HA, C. K. THAM: *Interference-minimized multipath routing with congestion control in wireless sensor network for high-rate streaming*. IEEE Transactions on Mobile Computing 7 (2008), No. 9, 1124–1137.
  - [12] C. L. AURIEMMA, C. A. NGUYEN, R. BRONHEIM, S. KENT, S. NADIGER, D. PARDO, S. D. HALPERN: *Stability of end-of-life preferences: A systematic review of the evidence*. JAMA Internal Medicine 174 (2014), No. 7, 1085–1092.

Received July 12, 2017

# Virtual reality remote control system based on image<sup>1</sup>

LIN GAN<sup>2</sup>, JUN DU<sup>2</sup>, XINGCONG FU<sup>3</sup>

**Abstract.** In order to improve the effectiveness of remote control, we designed a virtual reality remote control system based on image. We implanted virtual reality technology to improve the remote control performance of unmanned vehicles. The system supports viewpoint transformation and overcomes the limitation of observation field restriction in traditional vehicle remote control. The simulation results show that the system provides a theoretical and technical basis for the application and development of 3D visualization and remote display system of unmanned vehicle. Also, the new tracking system based on image can support scene roaming at any point of view, which overcomes the restriction of the limited observation field of vision in the traditional vehicle remote control mode, and meets the demand of unmanned vehicle remote display. The system we designed solved problems of the camera limited field view and cannot obtain 3D environment information.

**Key words.** Unmanned vehicle, virtual reality, 3D terrain, texture mapping, image stitching

## 1. Introduction

With the rapid development of artificial intelligence, computer and sensor technology, a mobile robot, which uses sensors to sense its own state and external environment, has gradually entered a substantive application phase. The unmanned vehicle is one of the special intelligent mobile robots. It can complete the continuous and independent movement under a variety of complex roads or environmental conditions. It has been widely used in people's daily life, as well as in the national military operations and many other fields [1]. Essentially, the intelligent system of unmanned vehicle is a complex whole with many functions, such as integrated route planning, environment perception and moving direction control. At present, the related research has not carried on the thorough analysis to the remote-control op-

---

<sup>1</sup>The authors acknowledge the National Natural Science Foundation of China (Grants: 51578109 and 51121005).

<sup>2</sup>North China University of Science and Technology, Tangshan, Hebei 063210, China

<sup>3</sup>Beijing Huilong Technology Co., Ltd., BeiJing, 101200, China

erating system based on the scene and the image. In order to realize remote control of unmanned vehicle, all scene information of traffic environment need to be obtained in time. The three-dimensional images of virtual reality are displayed in front of the control personnel, so that they can read and make remote control instructions quickly, and improve the remote-control performance of unmanned vehicles.

The remote control of the traditional unmanned vehicle mainly uses the camera to obtain the image. However, due to the limitations of the camera's visual field and the high latency and lack of data caused by environmental interference, remote control of unmanned vehicle becomes very difficult. Dąbrowska found that the vibration amplitude of the camera had a significant effect on the efficiency of the remote unmanned vehicle. At the same time, many research institutes apply lidar to the collection of scene images. Di Gennaro uses a multi-sensor method and remote sensing technique to obtain thermodynamic images through lidar mounted on unmanned aerial vehicles. Now, with the help of computer graphics and image processing technology, 3D visualization software has been widely used in many fields such as cultural relic protection, robot navigation and building manufacture [2]. It can be divided into three categories: modeling, platform and application. The modeling part is the core and foundation of the technology. For mobile robot remote control experiment, Kelly proposed to create a highly realistic 3D scene model based on virtual reality technology. First, a 3D terrain model is constructed by using 3D visualization technology based on virtual reality. Then, based on the registration images and radar data, realistic terrain rendering is implemented. Finally, the 3D terrain remote display system of unmanned vehicle remote control is designed and implemented, which has positive significance for the actual application of unmanned vehicles in the military field [3].

## 2. Experimental procedure

### 2.1. *Virtual reality and 3D visualization*

Virtual reality (VR) is a five-dimensional simulation of the real world, that is to say, besides the simulation of one dimension and three-dimensional space, it also includes the simulation of natural interaction. It is generated by a computer. By means of sensory modalities such as vision, hearing and touch, the system generates immersive interactive scene simulation. It's a computer system that creates and experiences the virtual world [4]. Virtual reality technology is a product of computer networks, graphics, artificial intelligence, sensors, information processing and other technologies. As shown in figure 1, it has three basic characteristics of immersion, interaction, and imagination, that is, the three "I". Immersion, also known as telepresence, means that in a virtual reality environment, users feel they become a "discoverer" and "actor". At the same time, it can sense the reality of being a protagonist in a simulated environment. Interactivity refers to the degree to which the participants are able to operate within the virtual environment and receive feedback from the environment. That is, the user is the subject of interaction and is multi aware. Users can operate directly on objects in a simulated environment and get

information or feel from the environment. These feelings can come from olfactory, gustatory, visual, and other sensory pathways [5]. Conceptual refers to the fact that users are immersed in multidimensional information spaces and acquire knowledge in a wide range of ways based on their perceptions and cognitive abilities. In the process of perception, they will exert their subjective initiative, get inspiration, seek answers, and form new concepts. In essence, the construction of virtual reality system is to create an information environment that enables participants to be in an immersive immersion, perfect interaction and inspiring ideas.

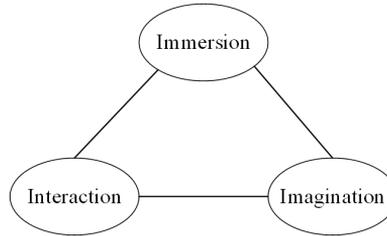


Fig. 1. Three "I" features of virtual reality

Three-dimensional visualization is the most important form of virtual reality technology. After a series of transformations, it converts the original analog data of the computer into the image that can be displayed. The idea is to convert abstract information into a format that can be understood by the human perception system. 3D terrain visualization is to build a 3D terrain model based on computer graphics and image processing technology. It is a reappearance of the real world's geographical environment with multi-levels and high fidelity. The realistic terrain model combines 3D spatial data with real image information. It gives the image information obtained by the camera to the terrain surface in the form of texture mapping. At the same time, it can display complex terrain, geometric structure and surface properties, and satisfy people's demand for visual effects very well. A full realistic 3D modeling includes modeling objects, rich geometric information, and complete texture information. The specific process is shown in Fig. 2.

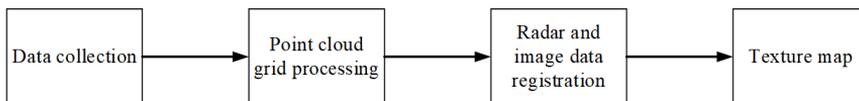


Fig. 2. 3D model building process

Firstly, lidar is used to obtain the point cloud data of the surface and depth information of the modeling object, and the corresponding image information is captured at the same time as the camera capture. In order to avoid the phenomenon of image information discarding, triangular meshes are used to process the point cloud data. After the lidar data are processed, the geometric model is obtained, and the camera image and the geometric model are corresponding to each other in space. Each point in the model is visible on the image, and the corresponding relation between the geometric model and the camera image is obtained [6]. After the

registration, the corresponding coordinates of each vertex of the geometric model in the texture image can be determined. The texture map can be completed by drawing the engine to calculate the texture coordinates corresponding to the pixels in each mesh unit.

## 2.2. Texture mapping and image stitching

The camera projected the 3D scene in the real world onto the two-dimensional plane of the camera sensor, and completed the mapping from 3D space to two-dimensional plane. The accurate mapping relation is an important condition to assign correct texture information to the geometric surface of the modeling object in three-dimensional reconstruction. The image captured by the camera is converted into a digital image in a computer, and is represented by an array of  $M \times N$  dimensions. Each element in the image is called a pixel, and the corresponding value is the brightness or RGB value of the image point. The coordinates of each pixel are represented by two values  $(U, V)$ , representing the number of rows and columns of the pixel in the array of images, respectively. The origin  $O_1$  is defined at the intersection of the camera's optical axis and the image plane, and the  $X$  and  $Y$  axes are parallel to the  $U$  and  $V$  axes, respectively. The coordinates of  $O_1$  in the  $U$  and  $V$  coordinate systems are set to  $(u_0, v_0)$ . The physical dimensions of each pixel in the  $X$  direction and the  $Y$  direction are  $dx$  and  $dy$ . The coordinates of any pixel in the two coordinate systems are as follows:

$$u = \frac{x}{dx} + u_0, \quad v = \frac{y}{dy} + v_0. \quad (1)$$

The internal and external parameters can be obtained by the calibration of the camera [7]. When the position of radar and camera is fixed, the calibration relation between lidar and camera can be determined by the process shown below, and the unique corresponding point of radar data point in the camera image can be obtained. The known  $[xyz1]^T$  coordinates are regarded as coordinates of a point in the radar coordinate system. They are the corresponding coordinate of the point in the image coordinate system, which can be seen by the direct calibration method of the lidar and the camera:

$$\begin{bmatrix} u & v & 1 \end{bmatrix}^T = G \begin{bmatrix} x & y & z & 1 \end{bmatrix}^T. \quad (2)$$

According to formula (2), the equation  $G = K[R|T]M$  can be obtained. The transformation matrix  $G$  can be used to find the only pixels corresponding to the lidar data points in the image. The matrix  $K$  contains the intrinsic parameters of the camera. The rotation matrix  $R$  and the translation vector  $T$  are the external parameters of the camera, and the orientation and position of the camera are determined.

The general process of texture mapping is to associate the multiple deformation vertices with their texture coordinates in the texture space during the modeling phase. Through the calibration of camera and radar, the elevation map triangle mesh vertex is mapped to the newest image acquired by camera. Each LIDAR point has a unique pixel corresponding to the data [8]. At this point, the task of the

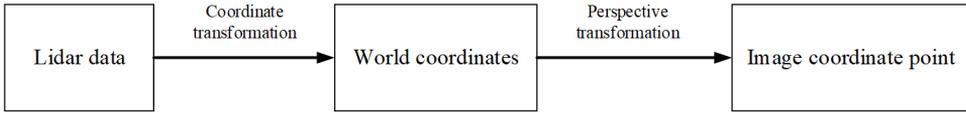


Fig. 3. Direct calibration principle of lidar and camera

rendering engine is to find out the corresponding texture coordinates for each pixel in each polygon. In the mainstream polygon rendering technology, the properties of the pixels inside the polygon are obtained by interpolating the properties of the vertices of the polygon. The triangular surface of the texture image is mapped to the target image one by one, so that the triangle in the source image is transformed into a triangle in the target image. Linear interpolation formula is shown as follows:  $P_1(x_1, y_1)$ ,  $P_2(x_2, y_2)$  and  $P_3(x_3, y_3)$  are the three vertexes of a triangle. Symbols  $P_a(x_a, y_k)$  and  $P_b(x_b, y_k)$  are the two points over the  $P_k$  sweeping line

$$\begin{aligned}
 P_a &= \frac{1}{y_1 - y_2} [P_1(y_k - y_2) + P_2(y_1 - y_k)] , \\
 P_b &= \frac{1}{y_1 - y_3} [P_1(y_k - y_3) + P_3(y_1 - y_k)] , \\
 P_k &= \frac{1}{x_b - x_a} [P_a(x_b - x_k) + P_b(x_k - x_a)] .
 \end{aligned} \tag{3}$$

The process of image mosaic is to transform images with overlapping to the same coordinate system, then the greater picture was synthesized. Image acquisition takes into account the focal length, type, position and motion state of the camera. Image preprocessing is the basic operation of digital image filtering, distortion correction, histogram processing and so on. The image matching process is to align the multiple images acquired at different times, locations and different cameras in space. Image transformation is a model transformation of the established image based on the matching point. The image to be spliced is converted to a unified reference image coordinate system. The process of image fusion is to eliminate the stitching seam and the matching error in the coincident area of the mosaic image.

### 2.3. Terrain geometry modeling and rendering methods

In order to make 3D realistic scene model meet the remote control requirement of unmanned vehicle, we must pay attention to the efficiency of data processing and the effect of 3D rendering of scene. In order to ensure the modeling effect, we need to preprocess the original cloud data, such as filtering, denoising, data segmentation and so on, before we build the 3D terrain model based on lidar data. In order to achieve the effect of photo realistic terrain visualization, fast terrain rendering, background region rendering and terrain visualization modeling are carried out on the basis of geometric model and image data obtained from registration.

For the 3D terrain construction method based on lidar data, the interpolation technique is first used to obtain continuous and smooth point cloud data after pre-

processing the original point cloud data. Then the triangulation algorithm is used to obtain the terrain surface. The terrain mesh obtained by interpolation and subdivision can only represent the geometry information of the far end scene. Texture information is also a necessary part of the remote scene model to achieve realistic. The camera image is simultaneously acquired by a plurality of cameras loaded by an unmanned vehicle [9]. These images are used for stitching, and the stitching results are applied to Billboard vision. This can show greater vision space, realize the realistic sense of the single frame terrain visualization.

#### ***2.4. Three-dimensional remote display system of unmanned vehicle***

A 3D terrain remote display system for unmanned vehicle remote control is designed and implemented. The Basler SC A1400-17gm model camera and Velodyne HDL-64E S2 64 line lidar are used as sensors. The sensor is fixed on an unmanned vehicle platform, and simultaneously generates 3D point data and corresponding image data in the radar coordinate system. The data is transmitted to the remote control platform by wireless link, and the wireless network is used to transmit the information. The virtual simulation environment is used as visual feedback, so that the operator can switch the viewpoint from the driving position in the reconstructed scene and the position near the unmanned vehicle. This process realizes the observation of the environment from the virtual viewpoint. In the integrated development environment, a series of terrain and ground modeling experiments are carried out by using the programming program, and the 3D visualization of the lidar image and the camera image data is realized. A visualized 3D terrain model of unmanned vehicle remote control is established, and the validity and feasibility of the system are verified.

### **3. Results and discussion**

#### ***3.1. A telepresence scheme for realistic scenes***

This paper starts with two aspects of modeling realism and high efficiency. The high precision 3D data of the measured scene is acquired by lidar, and the 3D triangle mesh model of the scene is rapidly built. Through the calibration of radar and camera, combined with the image information or coloring scheme of camera, a 3D realistic scene model with rich surface information and remote control operation can be reconstructed. The data processing flow of the design scheme. The overall design scheme includes two approaches, common cloud point data preprocessing and 3D terrain geometry modeling. The realistic terrain rendering method based on projection texture mapping and the fast terrain rendering method based on point coloring are implemented. The remote scene expresses the Billboard part and the ground surface model representation of non ground points.

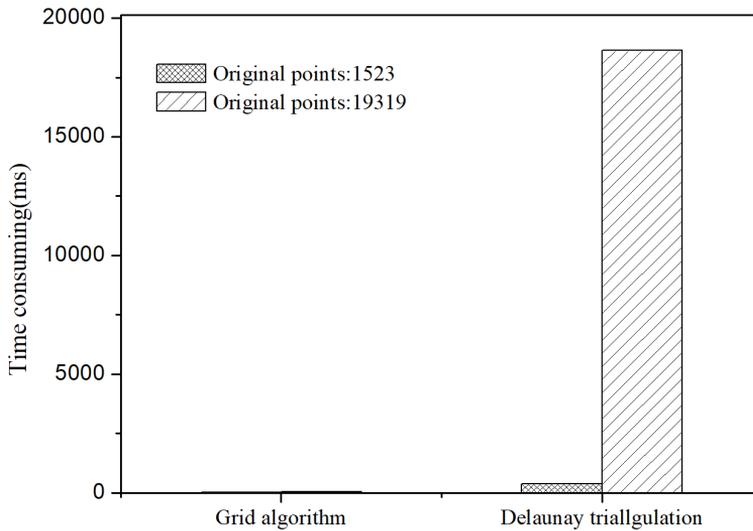


Fig. 4. Time consumption contrast between two algorithms

### 3.2. Three-dimensional terrain rendering results

The 3D mesh surface model of the system is constructed by elevation map with a uniform square grid. However, the traditional Delaunay algorithm combines the similar points in the original discrete point set data according to certain criteria to form the optimal triangle. In order to make the point centralized, each data point becomes a vertex of the triangular mesh, and the original scattered point sets are used to generate continuous irregular triangulated networks that can cover the whole point set region. The time-consuming comparison results are shown in Fig. 4. The Delaunay algorithm can produce better segmentation effect, but it takes too long. In order to improve the computational efficiency of terrain subdivision and make it applicable to the unmanned vehicle simulation environment with high real-time requirement, the terrain grid is verified by experiments. The algorithm is simple and time-consuming. And the segmentation effect can approximately express the relief of the terrain, satisfy the demand of mapping and remote control, and have the whole optimum property.

Texture information can not only give a realistic visual effect, but also make up for the lack of geometric modeling accuracy. Therefore, taking eight-line lidar data and camera image data as an example, the texture information obtained by using some technical means is added to the camera. This verifies the validity of the rendering. Texture mapping and terrain visualization for eight-line radar data are shown in Table 1.

Compared with traditional video remote manipulation images, the texture of realistic terrain model after texture mapping has abundant texture information, and can realize the observation of environment from different viewpoints [10]. With

the help of the terrain model of 3D terrain and image information, the geometry information of the scene can be enriched, and the exit path and the accessible road can be easily judged. As shown in Table 1, interpolation time and texture coordinate computation time are increased as the number of meshes increases, while rendering time is almost zero. In the texture mapping of  $400 \times 400$  mm mesh size, good visual effects have been achieved and the time consuming is about 55 ms/frames.

Table 1. Experimental parameters

Line No.	Mesh size (mm)	Mesh number	Interpolation time (ms)	Texture coordinate calculation time (ms)	Render time (ms)
5292	$400 \times 400$	$25 \times 45$	16	31	8
10260	$300 \times 300$	$34 \times 60$	15	47	14
14688	$250 \times 250$	$40 \times 71$	32	78	16

Compared with texture mapping, point coloring method avoids the process of texture interpolation, and the rendering efficiency is much higher than that of texture mapping method. However, the resolution of radar data is far below the image resolution. A large amount of image information is discarded if it is directly colored by the point cloud. This effect cannot meet the remote-control requirements. Therefore, the patch blocks are used instead of point coloring points to fill the gap of the point cloud, so as to compensate for the loss of the information when the resolution of the radar is lower than the resolution of the image. Under the condition that the number of interpolation points is 5292, the mesh size and the number are  $400 \times 400$  mm and  $25 \times 45$  mm, respectively, the two methods of terrain rendering based on projection texture mapping and point based shading are compared. The result is shown in Fig. 5. When using the same size mesh, the interpolation time of the two rendering methods is different from that of the texture/color coordinate. But for the rendering time, it is more efficient to use the texture mapping based on realistic terrain modeling method.

In order to show the special effects of 3D scene realistically, billboard method is used to display the background area. The billboard is the projection of the camera image, the surface must be parallel to the camera plane, and located on the camera's optical axis. Its size covers the projection area, and the image moves with the motion of the vehicle. When the virtual viewpoint is the same as the camera's actual location, the billboard is barely perceptible. As the virtual viewpoint moves farther away from the camera, the deformation becomes apparent. Billboard performs better when viewed from a distant point of view, but when the point of view approaches, its effect becomes dramatically worse. In the process of image matching, NCC operator is used as the similarity measure operator to match the two images. Image mosaics are performed using an affine transformation model with the parameter 6. Finally, in order to eliminate the stitching difference between two images, a smooth and natural stitching result is obtained. The pixel values of the pixels on the two sides of the

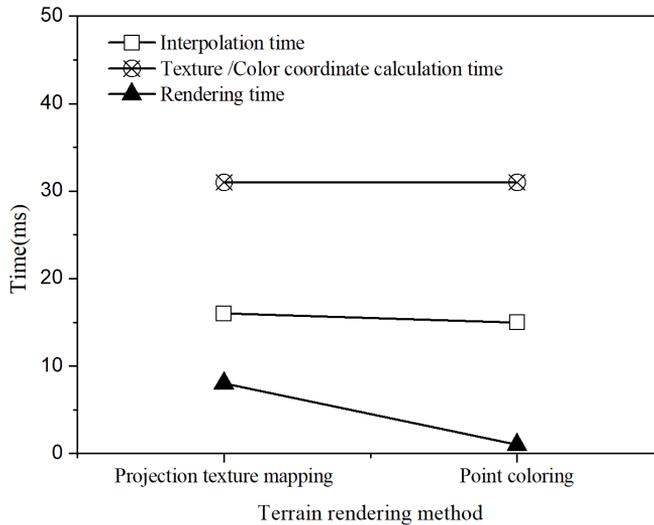


Fig. 5. Time consumption contrast between two rendering methods

joint and the distance between the pixels and the joint are weighted by the method of linear fade. The pixels of the overlapped region are mixed to obtain the final image results. Integrating the terrain model with the object model, the complete remote scene expression is obtained.

### *3.3. Design and implementation of 3D remote display system for unmanned vehicle*

Virtual reality uses digital graphics models to recreate scenes in the real world. The effects and the fidelity of the scene we see are related to the model. The basic hardware settings at the system level include the remote control station, the on-board unit equipped with the lidar, the camera, and the wireless communication module for transmitting instructions and video. In the Visual C++ integrated development environment, the application of Open Scene Graph is applied to carry out a series of experiments on terrain and ground features modeling. The 3D visualization of imaging lidar images and camera images is realized, and a 3D visualization model of unmanned vehicle remote control is established.

Figures 6 and 7 are time consumption comparison of terrain modeling between real terrain rendering module and point shading terrain rendering module. The original LIDAR point cloud number is about 11966, and the interpolation points are 11484 and 44814 respectively. The corresponding mesh sizes and numbers are  $400 \times 400$  mm and  $91 \times 32$  mm,  $200 \times 200$  mm and  $194 \times 87$  mm, respectively.

Using  $400 \times 400$  mm mesh size to interpolate texture mapping, the time consumed during modeling is approximately 141 ms/frames. Texture mapping using  $200 \times 200$  mm interpolation is similar to it. But the time consumption of interpolation and texture coordinate calculation is obviously increased. In addition, the

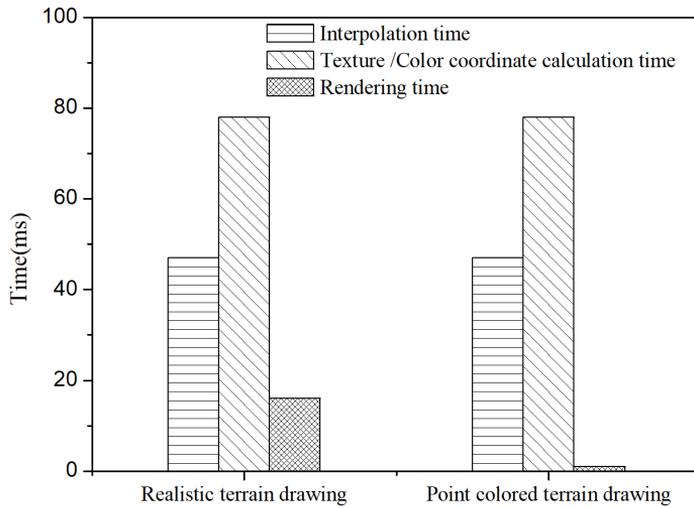


Fig. 6. Time consumption comparison results (mesh size  $400 \times 400$ )

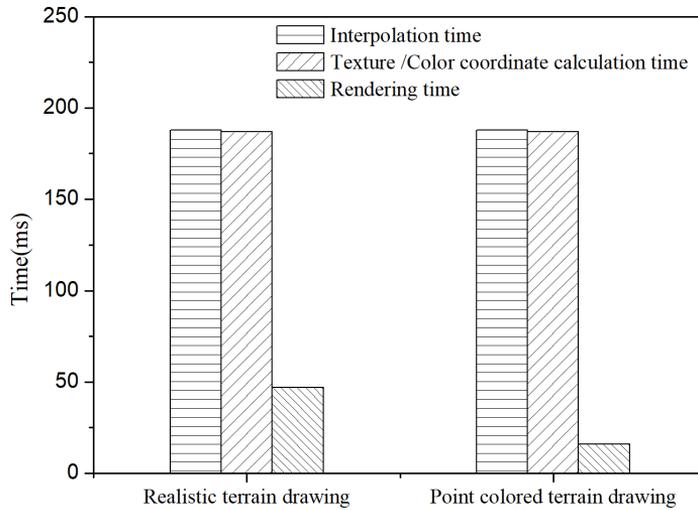


Fig. 7. Time consumption comparison results (mesh size  $200 \times 200$ )

$400 \times 400$  mm mesh size interpolation of point cloud coloring models has large gaps and is inconvenient for operator driving observation. After shrinking the mesh to  $200 \times 200$  mm, the coloring effect basically meets the needs of remote display, but the time consumption reaches 391 ms/frames. The computation of texture coordinates of interpolation and point clouds takes a great deal of time, while the time consumption of rendering is significantly lower than that of realistic terrain rendering. The modeling results of the two functional modules can meet the remote control requirements of unmanned vehicle remote control. In contrast, photorealistic ren-

dering module works better and is more realistic. After fine interpolation, the point shading terrain rendering module can achieve the approximate realistic shading effect. But the time consumption of interpolation processing is relatively large. In time processing, the two methods have greater room for improvement.

## 4. Conclusion

The remote visualization system of unmanned vehicle remote control is used as an application background, and the 3D visualization terrain modeling method is studied from two aspects of theory and practice. Firstly, the basic theories and algorithms involved in the data fusion techniques of lidar and camera images are analyzed. On this basis, a rapid terrain geometry modeling scheme is proposed. Considering the two aspects of modeling effect and rendering efficiency, realistic rendering based on projection texture mapping is applied to realize realistic rendering of terrain model with rich scene information. Then, the point rendering terrain visualization method with adjustable resolution is used to improve the rendering efficiency of the model. At the same time, the point size control function makes the interpolation time consumption and the modeling effect reach a good compromise. Experiments show that the two methods can realize the observation of the virtual viewpoint environment without the sensor data angle. High precision 3D data of measured terrain are acquired by lidar. A 3D terrain model with abundant surface information is reconstructed by combining the image information of the camera or the coloring scheme. The system can support scene roaming at any point of view, which overcomes the restriction of the limited observation field of vision in the traditional vehicle remote control mode, and meets the demand of unmanned vehicle remote display.

The system only realized the terrain modeling of unmanned vehicle remote control, and did not study the modeling of non-surface objects. At the same time, the system does not have the Billboard vision display function based on image mosaic. Based on the visualization of single frame terrain, the future research can provide location and status information by means of navigation system, and integrate and map multi frame radar data. In this way, more abundant information can be expressed, and a realistic offline 3D scene map can be set up, which can better meet and meet the needs of unmanned vehicle applications.

## References

- [1] A. J. MATHEWS, J. L. R. JENSEN: *Visualizing and quantifying vineyard canopy LAI using an unmanned aerial vehicle (UAV) collected high density structure from motion point cloud*. Remote sensing 5 (2013), No. 5, 2164–2183.
- [2] A. DĄBROWSKA, M. B. JASKÓŁOWSKI, A. RUBIEC: *Cameras vibrations influence on efficiency of teleoperated unmanned ground vehicle*. IEEE International Conference on Methods and Models in Automation and Robotics (MMAR), 29 August–1 September 2016, Miedzyzdroje, Poland, IEEE Conference Publications, (2016), 772–777.
- [3] S. F. DI GENNARO, A. MATESE, B. GIOLI, P. TOSCANO, A. ZALDEI, A. PALLIOTTI,

- L. GENESIO: *Multisensor approach to assess vineyard thermal dynamics combining high-resolution unmanned aerial vehicle (UAV) remote sensing and wireless sensor network (WSN) proximal sensing*. *Scientia Horticulturae* 221 (2017), 83–88.
- [4] A. KELLY, E. CAPSTICK, D. HUBER, H. HERMAN, P. RANDER, R. WARNER: *Real-time photorealistic virtualized reality interface for remote mobile robot control*. *Robotics Research, Part: Springer Tracts in Advanced Robotics, Book series STAR 70* (2011), 211–226.
- [5] G. M. E. SANCHES, T. VAN RENTERGHEM, K. SUN, B. DE COENSEL, D. BOTTELDOOREN: *Using virtual reality for assessing the role of noise in the audio-visual design of an urban public space*. *Landscape and Urban Planning* 167 (2017), 98–107.
- [6] M. KULAWIAK, M. KULAWIAK: *Application of web-GIS for dissemination and 3D visualization of large-volume LiDAR data*. *The Rise of Big Spatial Data, Part: Lecture Notes in Geoinformation and Cartography, Springer Nature* (2017) 1–12.
- [7] S. IWASAKI, T. NARUMI, T. TANIKAWA, M. HIROSE: *Guidance method to allow a user free exploration with a photorealistic view in 3D reconstructed virtual environments*. *Proc. International Conference on Distributed, Ambient, and Pervasive Interactions (DAPI), 9–14 July 2017, Vancouver, BC, Canada, Conference proceedings DAPI, Part: Lecture Notes in Computer Science, Springer Nature 10291* (2017), 347–357.
- [8] J. LIU, J. LI, W. XU, Y. LIU, M. ZHANG: *Mosaicing of fisheye images through conformal projection and sphere matching*. *Proc. International Conference on Graphic and Image Processing (ICGIP), 29 October 2016, Tokyo, Japan, Processing SPIE 10225* (2016).
- [9] Y. Y. NIE, Z. TANG, J. F. YU, Y. R. ZHU, J. CHANG, J. J. ZHANG, S. H. GUO, Y. SU: *Image-based 3D scene reconstruction and rescue simulation framework for railway accidents*. *Proc. IEEE International Conference on Virtual Reality and Visualization (ICVRV), 24–26 September 2016, Hangzhou, China, IEEE Conference Publications, (2016), 335–340*.
- [10] G. ZHONG, H. DENG, Y. KOBAYASHI, H. WANG: *Theoretical and experimental study on remote dynamic balance control for a suspended wheeled mobile manipulator*. *Non-linear Dynamics* 79, (2015), No. 2, 851–864.

Received July 12, 2017

# Research on data acquisition and fusion system based on wireless sensor

WANG BOHUAI<sup>1</sup>

**Abstract.** In recent years, wireless sensor network as a new information technology has been subject to scientific and industrial sessions at home and abroad with the rapid development of Internet of things. Based on this, the data acquisition and fusion system based on wireless sensor was studied. In this paper, the structure of the wireless sensor was introduced, and the design of the data acquisition system based on ZigBee wireless sensor network was studied, then the design of the monitoring software of the data acquisition and fusion system based on wireless sensor was carried out and the actual application was implemented. Experiments show that the data acquisition system based on wireless sensor network can be used for collecting and realizing all kinds of information data.

**Key words.** Wireless sensor network, data acquisition, temperature and humidity, data fusion.

## 1. Introduction

In recent years, new short-range wireless communication technology has emerged with the rapid development of computer networks, wireless communication technology, microcomputer and sensor technology. Wireless sensor network makes a large number of sensor nodes form a network in a self-organizing manner, and integrates information awareness through wireless short-range communication technologies. The embedded system and wireless communication technology is one of the main research subjects in the world. The wireless sensor networks reviewed by the Massachusetts Institute of Technology as one of the ten technologies that changes the future of the world. Traditional data acquisition system usually collects data of equipment by wiring and manual mode, industrial production equipment appears in the dispersed areas with the development of productivity technology. High temperature and high pressure equipment for on-site data collection and maintenance are difficult and dangerous, which needs to invest a lot of manpower and financial resources. Wireless sensor network based data acquisition system can not only solve the problems in the

---

<sup>1</sup>School of Information Engineering, Yulin University, Yulin, Shaanxi, 719000, China

artificial cable system, but also can improve the market competitiveness of enterprises. Extensive use of sensor network data acquisition system can be made on the basis of power consumption of industrial production. On the one hand, temperature and humidity control effects can be improved by wireless monitoring of some process temperature and humidity, so as to improve product quality; on the other hand, the power consumption of the power plant is monitored in real time, and the equipment maintenance and replacement can be carried out, so as to realize the purpose of saving cost and reducing energy consumption.

## 2. State of the art

The United States first used wireless sensor network in military areas. The application of wireless sensor network in the United States expands from the military field to civilian areas such as home automation, environment, energy monitoring and building automation [1]. Intelligent dust engineering of branch of University of California at Berkeley achieved a cubic millimeter application platform in July 2010 [2]. According to the project in 1999, the development of ultra-small operating systems was still the first choice for wireless sensor network operating systems [3]. In addition to the University of Berkeley, Massachusetts Institute of Technology was also engaged in the research of wireless sensor network with very low-power [5]. While Auburn University was engaged in self-organizing sensor network [4]. In addition, the University of Pennsylvania and Cleveland State University were also engaged in related research. In addition to universities and research institutions, major foreign manufacturers have also conducted a wireless sensor network research [6]. Philips and Motorola and other well-known international companies set up the ZigBee Union in 2001, the alliance was committed to the study of ZigBee technologies such as short distance, low power consumption and low cost. ZigBee has become one of the best technologies in the field of wireless sensor network [7]. All kinds of commercial organizations represented by academicians of the Chinese Academy of Sciences as well as representatives of major scientific research institutions and operators have explored and researched the theory and application of wireless sensor network [8]. The Institute of computer science of the Chinese Academy of Sciences has developed a wireless sensor network node GAINZ, which has been compatible with the mainstream 2.4G wireless sensor network nodes in the market, and transmission distance and power consumption have had certain advantages. The Institute of information engineering of the Chinese Academy of Sciences has done a great deal of work in wireless sensor network security protocols [9].

## 3. Methodology

Wireless sensor is defined according to three kinds of nodes from the perspective of network logic: PAN coordinator, coordinator and terminal node; the PAN coordinator is unique across the network, and it is usually used as a network setup and initialization to implement a more powerful [10]. The coordinator is determined by

the number in the network, which assumes the function of network maintenance. The terminal node is generally the terminal node of the network, the function is simple and the cost is the lowest. In this paper, the convergence nodes on the PAN coordinator, the temperature and humidity sensing points, and the instrument nodes of the end nodes play the important role. MAC structure is divided into star, topology and peer structure. In topology, all communications are built on the PAN coordinator and other devices. There are three transport modes in peer to peer topology: data transformation from the terminal device to the coordinator, data transformation from the coordinated to the terminal, and data transformation between the coordinator and coordinator. Physical map of wireless sensor is shown in Fig. 1.

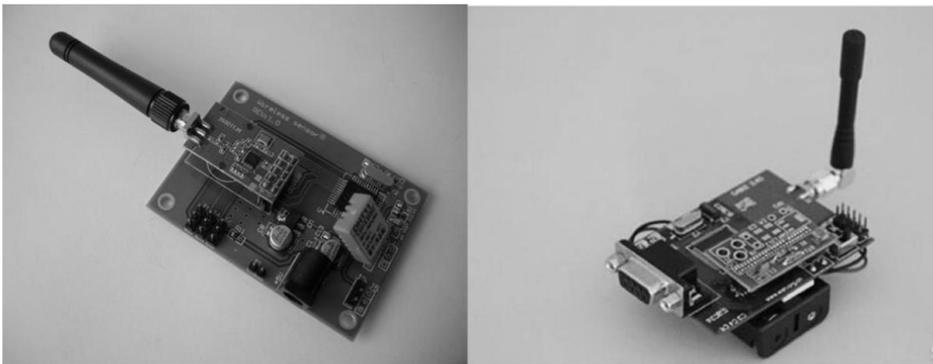


Fig. 1. Wireless sensor

The physical layer includes home automation, PC peripherals, games and personal care, the use of peer-to-peer network nodes can form more complex networks, the network topology has a wider range of applications and higher security, the mesh network topology used in this paper is also based on peer-to-peer network topology. The main function of the network layer is routing addressing and establishment and maintenance of network. ZigBee network is a kind of wireless self-organizing network, the nodes in the network are communicated and organized each other through single hop or multi-hop, and are distributed flexibly throughout the network. ZigBee network topology is rich and diverse, such as star network, tree network and mesh network topology, the network based on peer-to-peer network and mesh network. Network topology control can automatically generate a good network topology through topology control, which can improve the efficiency of routing protocols and MAC protocols, lay the foundation for data management, and save node energy to prolong network lifetime. The sensor network protocol is responsible for the formation of a multi-hop data transmission network for individual nodes; the sensor network is compatible with the IEEE 802.15.4 standard. The standard physical layer is used O-QPSK coded modulation, and radio frequency signaling is transmitted by differential means to form a peer to peer network (hierarchical routing). CSMA/CA access channel is used at the data link layer, two addressing modes: 16 bit short addressing and 64 bit IEEE addressing are used, at the same time, a cache mechanism

with reduced power consumption is adopted to allow the energy devices to sleep in most of the time, whether or not a message is waiting for processing by periodically listening to the radio channel to balance energy consumption and message latency and ensure low power consumption management, the transport layer uses a complete handshake protocol.

The design of data acquisition system based on ZigBee network mainly includes three parts: terminal node, data acquisition node and monitoring software, of which the terminal node and the data acquisition node are implemented in the two part of the ZigBee network. The terminal node corresponds to the temperature and humidity sensing nodes and instruments of the ZigBee network, and the terminal node and the data collection point form a mesh network, the main function of monitoring software PC is to monitor the ZigBee network in real time and end the node perception data. The terminal node-aware information data is sent to the GPRS DTU through a string at the convergence point of the data collection point. GPRS DTU will receive TCP data packets, and finally sent to the monitoring software for real-time display, the structure of data acquisition system based on ZigBee network is designed as in Fig. 2.

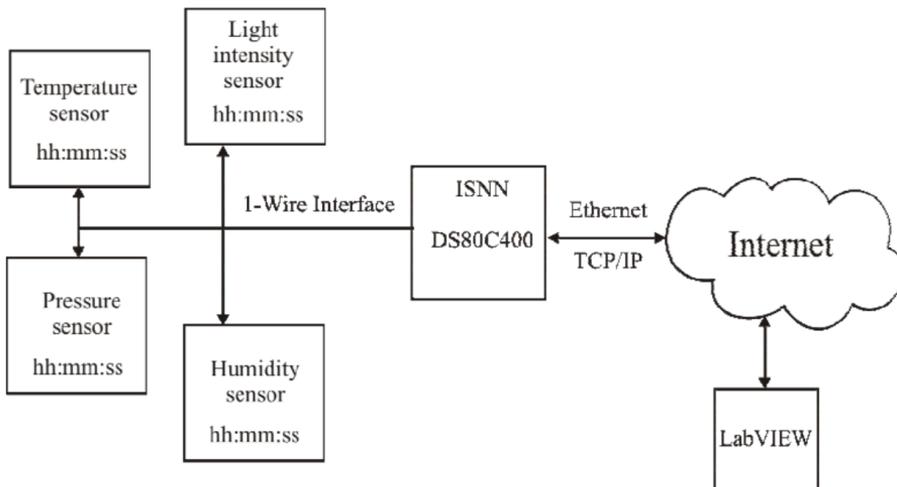


Fig. 2. Structure of data acquisition system based on ZigBee network

The ZigBee network includes both terminal nodes and data acquisition contacts. There is only one data node and the rest is the routing point in all data collection points. The data sensing point is powered by the battery, and the data acquisition node requires a stable power supply, and all nodes use a mesh network and tree network as an alternative. Endpoint nodes are sensed by temperature and humidity data and instrument data, ZigBee transmits wireless power to monitor point processing. The monitoring point will receive data through three types of nodes to the network node, as shown in Table 1.

WPF is a technology developed by Microsoft, which uses the new presentation layer framework technology and integrates many framework technologies, including user GDI, GDI + and HTML. WPF has done the separation of the front desk inter-

face designer and the daemon developer. It has also flexible layout system, powerful data binding, advanced graphics animation support, and easy-to-use template styles. ZigBee monitoring software needs to dynamically display changes in each terminal node, such as the terminal node to join the network or leave the network, and it needs to dynamically display the value of each terminal node acquisition. WPF can not only easily realize the 3D effect display of each terminal node, but also can realize the animation effect of each terminal node easily. WPF will automatically handle these details, and we just need to explain how we move the animation, other things are completed by the WPF system.

Table 1. Nodes of ZigBee network

Converged nodes	In the initial processing phase, the ZigBee sink node is used to initialize the network structure, select the channel, PAN ID and extend PAN ID. After the completion of the network, on the one hand, the sink node acts as the role of the ZigBee router and is responsible for forwarding data; on the other hand, the sink node will also be responsible for network channel management and handle the binding requests of the new terminal nodes. If the security protocol is used, it also acts as a trust center.
Routing nodes	The ZigBee routing node is used to transfer information sent by other nodes in the network. In default, the network layer is used in tree addressing and meshes networks. ZigBee routing nodes select routes to forward messages through training. When the node's memory is not enough to run a mesh network, a tree network with lower memory requirements is automatically selected.
Terminal nodes	The terminal node is the node that realizes temperature and humidity sensing and meter reading. The first one is needed to join the ZigBee node or the ZigBee coordinator node, and then managed by the ZigBee sink node and the routing node.

In this paper, there were 10 terminal nodes and 1 receiving nodes in ZigBee network and the terminal node had four temperature and humidity acquisition nodes and six meter reading nodes. The circle represented a node, and 11 circles were drawn. Each terminal node dynamically joined the network or left the network, which needed to be embodied in the graphical monitoring interface program, so the timer was set for each terminal node. After the terminal node joined the network, the monitoring data and the 64-bit long address were sent to the monitoring software, and the terminal node graph automatically displayed the animation in 5 seconds. The terminal node temperature and humidity and meter reading value were sent once every 1 second, but packet loss was sometimes taken into account in the network state, accordingly, each terminal node corresponded to a timer count of 10 seconds,

that was, if there was no terminal node sending the successful data in 10 seconds to the monitoring software, the timer would cause the terminal node to disappear within 5 seconds.

In order to reflect the relationship between the terminal node and the starting node, each terminal node and the destination node were connected to the network through a line segment. When the terminal node joined or left the network, the corresponding rows would also display or disappear animation. The data received from the TCP was the network type that contained the network address, and the terminal node sent both numeric and numeric fields. After parsing the data, the network address needed to be stored in an array, and each item in the array corresponded to the terminal node pattern. This was because when a terminal node left the network and rejoined the network, the array would be able to find which graphics the terminal node should display. If the 64 bit address was indicated, the ID of the terminal node would become very long, so the last 16 bits in 64 bits were displayed in hexadecimal numbers. The sink node was connected to the monitoring terminal, it could be used to configure and manage the wireless sensor network node, issue the monitoring task, and receive and process the monitoring data. The tone generator was just an exception alarm and added a serial communication interface to the computer. The components of the terminal, SQL, Postgres, database, and user interface were monitored through the user interface, and the user interface provided the interface between the monitoring terminal and the wireless sensor network through the sink node. The Postgres SQL database stored various types of sensor data for querying, scheduling, and analysis. The user interface was based on the Java application interface, which was consisted of a set of Java classes and some applications, including the following contents: user interface, graphical interface, display sensor, experimental data tables and change charts, analysis of monitoring results, visual dynamic network topology, and warning the abnormal work environment on node.

#### 4. Result analysis and discussion

The system can regularly collect data of temperature and light intensity and data storage. The main program flow chart of the system is shown in figure 2. When the system was running the acquisition program, first of all, the sensor node determined whether the collected temperature or illuminance was greater than the threshold value or not, and the environment anomaly near the node was indicated if the condition was satisfied (this node was treated as an exception node in this paper), sink node alarmed, monitor terminal displayed alert; system automatically stored data and maps at that time, and sensor nodes continued to acquire. At this point, the monitoring terminal could first be used to calculate the abnormal cause of the node according to the saved distribution map at that time, then all the data collected that were deployed with 6 sensor nodes inside the room were placed next to one device, meanwhile, the fluorescent lamps just above the nodes 2, 6, 5 and 3 were in the open state. Node 0 was the sink node, which was connected with the monitoring terminal. Firstly, the sink node sent the commands to the sensor nodes

for continuous collection of ambient temperature and illuminance, and then each node sent data to the sink node and displayed it at the monitoring terminal in real time. The interval of data sampling was 2s, and the temperature threshold was set to 30 °C, and the illuminance threshold was set to 650l. The data collected by the 6 nodes temperature sensor at a certain time is shown in Table 1. There was a gap of 2 °C between the maximum and the minimum, which was much more accurate than the single point acquisition, and also showed that the temperature values in each region weren't the same. The experimental results show that the wireless sensor network technology can achieve fine sampling in the monitoring area.

Each terminal node was represented by a text box. The default template for a text box was rectangular, but each rectangle could be changed to a circle by replacing template. Data sent by temperature and humidity sensing nodes was displayed by analyzing three rows of data: the first line of data was the 64 bit long address after 16 bits and sixteen decimal numbers; the second line of data was the temperature, the unit was centigrade; the third line was the relative humidity, the unit was the degree. For power meter nodes, the data was displayed by analyzing the two rows of data sent: the first line of data was 64 bit long address with sixteen hexadecimal numbers; the second line of data was numeric. There were 4 temperature and humidity sensing nodes, 6 power meter reading nodes, and a sink node consisting of ZigBee networks. The GPRS DTU simulation software was sent to the display interface of monitor software, as shown in Fig. 3.

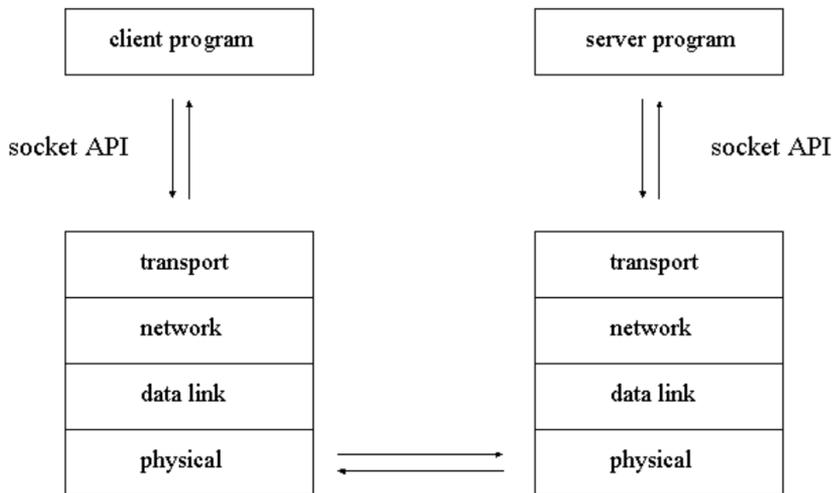


Fig. 3. Program of monitoring software interface

In Fig. 3, the API was used to enter the IP address of the data monitor. The “listen port” text box allowed us to enter the TCP port number that the data listener had. Listening of data was monitored by starting TCP sockets, the server stopped listening to data through "Stop listening" button. Sink nodes and information collected from each terminal node are as shown in Table 2.

Table 2. Information collected by terminal node

Sequence of nodes	Types of nodes	Value of information
1	Temperature and humidity nodes	Temperature: 26.5; humidity: 39
2	Temperature and humidity nodes	Temperature: 25.5; humidity: 37
3	Table lookup node	Meter reading value: 0.23
4	Table lookup node	Meter reading value: 0.25
5	Temperature and humidity nodes	Temperature: 25.8; humidity: 36
6	Table lookup node	Meter reading value: 0.26

As can be seen from the above table, the maximum temperature was 25.47 °C and the minimum temperature was 25.040 °C. The difference between the highest and lowest temperatures was less than 0.6 °C. The meter reading value was the same as the meter reading of each instrument, so the collected instrument read the data accurately. For temperature and humidity data, the monitoring value of the monitoring software was changed every second. For meter reading data, when the power meter was connected to blower with power supply 1 kWh, monitoring meter readings of monitoring software would increase by 0.01 kWh every 36 seconds, therefore, the effect of real-time monitoring was also realized.

## 5. Conclusion

The data acquisition accuracy of traditional single point data acquisition system is not high enough to achieve the purpose of accurate monitoring. Multiple sensors can monitor the environment in many directions. Therefore, a scheme of environment data acquisition system based on wireless sensor network was proposed, in this paper, the data acquisition system based on ZigBee wireless sensor network was designed and implemented, and the data types collected included air temperature, humidity, and meter reading data. Based on the full study of ZigBee network, the networking of ZigBee mesh networks and data transmission were realized, and network layer security related interface was called to realize data security transmission of ZigBee network layer. In addition, high-precision temperature and humidity data and accurate meter reading data were collected. Finally, it is found that the difference between the highest and lowest temperatures is less than or equal to 0.5 °C through experiment; in addition, the monitoring software developed in this paper can be used for wireless meter reading. The monitoring software developed in this paper has a simpler function, the number of ZigBee network terminal nodes displayed can't exceed 10, and monitoring software cannot send information to the terminal node because of the limited time and effort. Therefore, the further improvement of the function of monitoring software is needed.

## References

- [1] L. Q. YANG, J. C. ZHAO, Y. ZHOU, S. F. MA: *Design of data acquisition system based on wireless sensor network*. *Microelectronics & Computer* 24 (2007), No. 09, 68–71.
- [2] M. TANG: *Design and development of data acquisition system based on Zig Bee wireless sensor network*. *Computer Knowledge and Technology* 42 (2015), No. 14, 1552–1553.
- [3] Z. HANDLEY: *Is text-to-speech synthesis ready for use in computer-assisted language learning?*. *Speech Communication* 51 (2009), No. 10, 906–919.
- [4] N. K. SURYADEVARA, S. C. MUKHOPADHYAY: *Wireless sensor network based home monitoring system for wellness determination of elderly*. *IEEE Sensors Journal* 12 (2012), No. 6, 1965–1972.
- [5] S. LI, L. D. XU, X. WANG: *Compressed sensing signal and data acquisition in wireless sensor network and internet of things*. *IEEE Transactions on Industrial Informatics* 9 (2013), No. 4, 2177–2186.
- [6] P. LI, L. SUN, X. FU: *Security in wireless sensor networks*. *Wireless Network Security*, Springer, Berlin, Heidelberg, 2013.
- [7] J. DEMONGEOT, G. VIRONE, F. DUCHÊNE, G. BENCHETRIT, T. HERVÉ, N. NOURY, V. RIALLE: *Multi-sensors acquisition, data fusion, knowledge mining and alarm triggering in health smart homes for elderly people*. *Comptes Rendus Biologies* 325 (2002), No. 6, 673–682.
- [8] D. F. LARIOS, J. BARBANCHO, G. RODRÍGUEZ, J. L. SEVILLANO, F. J. MOLINA, C. LEÓN: *Energy efficient wireless sensor network communications based on computational intelligent data fusion for environmental monitoring*. *IET Communications* 6 (2012), No. 14, 2189–2197.
- [9] A. SCHMIDT, M. BEIGL, H. W. GELLERSEN: *Transverse vibrations of orthotropic non-uniform rectangular plate with continuously varying density*. *Computers & Graphics* 23 (1999), No. 6, 893–901.
- [10] H. W. GELLERSEN, A. SCHMIDT, M. BEIGL: *Multi-sensor context-awareness in mobile devices and smart artifacts*. *Mobile Networks and Applications* 7, (2002), No. 7, 341 to 351.

Received June 6, 2017



# Analysis of key technologies of distributed file system based on big data

ZHOU JUNPING<sup>1</sup>

**Abstract.** With the advent of the big data era, it is becoming more and more urgent to provide a stable and efficient distributed file system. The performance of traditional distributed file systems is becoming increasingly difficult to meet the growing needs of the Internet in terms of scalability, stability, and access to multiple users. The key technologies of big data oriented distributed system - Clover cluster system were designed in this paper. In this system, the two-stage mapping relation and two segment decision protocol were adopted to ensure the operability, extensibility and availability of metadata. The performance test shows that the Clover cluster has good operability, scalability and availability in metadata, and can be used as the key technology for big data oriented distributed systems.

**Key words.** Big data analysis, distributed file systems, Clover cluster.

## 1. Introduction

With the development of information technology, there have been more and more kinds of information in data center and Internet, and semi-structured and unstructured data has become explosive increase in geometric quantity [1]. These are signs of the era of big data. Compared with the previous data, the information in the big data era presents the following characteristics:

(1) With the rapid growth of massive data files, the size of the file has expanded rapidly. In large-scale Internet Co, such as Alibaba, Tencent and so on, the size of their data has already exceeded the PB level, and the amount and size of documents to be processed is beyond measure [2].

(2) User access: for some large Internet Co in China, it is common to say that more than one million or ten million people live online at the same time. The large amount of random read and write data caused by so many users accessing online at the same time is fatal to the entire system, whether it is storage or access [3].

(3) Data structures and processing are varied. With the steady development of

---

<sup>1</sup>Ulanqab Vocational College, Wulanchabu City in Inner Mongolia, 012000, China

Internet applications, the types of Internet applications have become multifarious [4]. For both offline data transmission and online data analysis, the system needs to provide uninterrupted high-quality services, so that the system security requirements are getting higher and higher [5].

With the advent of the big data era, the demand for storage systems has become even severer. The storage system should not only support data processing and analysis of data center and Internet applications, but also should store a large amount of data in a timely and effective way, which is embodied in the following aspects:

(1) In the face of large amounts of data and huge files, the storage of the system needs better scalability and concurrent processing capability.

(2) The system can support different kinds of access requirements in different ways.

(3) The sharing of data and the security of data are becoming more and more important. Under the premise of providing full freedom and sharing, the system should also ensure the safety and reliability of data [6].

## 2. State of the art

The distributed file system refers to the physical storage resources that are managed by the local file system and are connected to the nodes through the Internet network rather than directly connected to the local nodes [7]. The design of a distributed file system comes from the client server model. In a full network, the server can provide services that many users access online simultaneously [8]. At the same time, the network also allows some systems to act as both users and servers [9]. For example, a user releases data that other users can also access, and when other users access the data, the data is local drivers for the client.

## 3. Methodology

Specifically, all kinds of distributed file systems can adapt to a certain application environment and play a superior performance and well meet the needs of the computing system for storage system at the various stages of storage technology development [10]. The design and implementation of the distributed file system are based on a certain storage structure, which is organized according to the storage medium, and the storage structure adopted in the distributed file system is mainly divided into virtualization storage structure and object storage structure [11]. The more successful distributed file storage systems currently used are GFS, HDFS, PNUTS, etc.

- Research on GFS architecture: GFS supports massive data block storage. For file updates, GFS is done by adding new data rather than modifying the raw data [12]. In the GFS architecture, the stream mode of a large number of data is the read operation, and the random mean of a small amount of data is the read operation [13]. As can be seen from the above, GFS performs better in large-scale search services, but it also limits GFS's business applications in

other areas.

- Research on HDFS architecture: HDFS supports user or user programs to create folders and stores data and documents [14]. Users can use the permissions to create, delete, move, and rename the folder, but HDFS does not support user disk quotas and access restrictions. In data storage, HDFS also supports data block storage, while the disk data errors, heartbeat detection and re-replication are features of HDFS [15].

### 3.1. Clover file system

The Clover distributed file system uses the multiple data server architecture, which can be completely compatible with the HDFS interface by separating the data stream and the control flow. While ensuring the functionality of the file system, it also enables the use and scalability of the very high meta data, and the operation of metadata in the no data server can also be unified.

A typical Clover cluster is shown in Figure 1, the underlying layer consists of several data servers (datanode, DN), and metadata servers are linked (metadata, server, MDA), a cluster is formed by storing, analyzing, and storing data and meta-data with a shared storage pool, and the services are provided to clients primarily with client nodes.

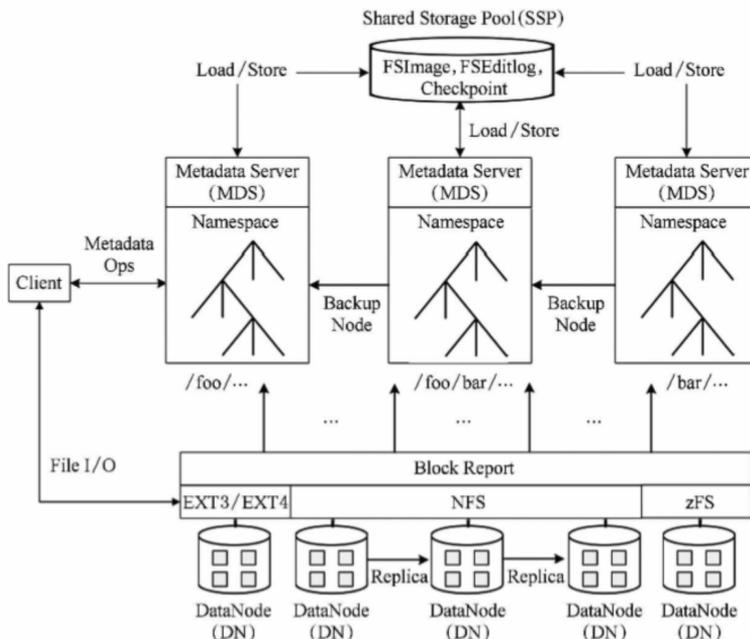


Fig. 1. Typical structure of Clover cluster

In the Clover cluster, the files are stored on a specified number of data servers

in block form, and the data servers are placed in duplicate copies of the blocks of files, and the data is checked on the data server. When the metadata server sends a command to the file, the relevant data server sends the file of the genus to the corresponding node according to the block information, so as to reduce the memory occupancy of metadata servers and optimizing the system. At the same time, similar to the HDFS disk data error and heartbeat detection function, the Clover cluster also periodically detects nodes on the metadata server and uploads the test files corresponding to the information and content claims to the shared storage pool. The node detection of the metadata server can be directed to a single metadata server node, or to all metadata server nodes within the cluster, at the same time, the corresponding nodes can be backed up and saved, so that when the fault occurs, the backup can be replaced quickly, and the functions of the cluster can be quickly restored.

When the metadata server issues a request to process metadata, the corresponding metadata server will form a corresponding cluster, then store metadata related information, and provide corresponding functions, such as file retrieval, file attributes, file names, and so on. At the same time, the cluster stores the corresponding contents written in the shared storage pool, so that the next operation can be done quickly and effectively. While writing content to a shared storage pool, the metadata server also writes the content back to the local server, and thus the reliability and convenience of the next operation are guaranteed by comparing the two contents in the shared memory and the local server.

### *3.2. Extensibility of metadata*

By reasonably dividing and mapping the spatial information of the names, the division of namespaces can be self-adaptable and extended, and can adapt to the rapidly growing amount of data and files, so as to save space and optimize the system. In the Clover cluster, the Hash is used to partition namespaces.

When data is stored into a number of metadata servers, the operations such as renaming, copying, pasting can affect more than just a metadata server, which can cause a large amount of metadata to be transmitted in the metadata server, and make system pressure. Through the global distributed directory Hash table (global, Hash, table, GDT), the mapping relationship between directory and metadata server is established. In this mapping relationship, there are two levels in the Clover cluster. The mapping of the first level is the storage path name of the file or data, and the Hash is calculated to find the GDT mapping relation. The second level is based on the first level, and the mapping relation of one to one is obtained. Through the consistency Hash algorithm, the nearest metadata server node is searched in the Clover cluster and mapped to the corresponding metadata server.

GDT likes a huge combo cabinet, each metadata service is a drawer on the cabinet with different maps. GDT is composed of these mapping tables, and each mapping table is stored on different metadata servers in blocks. Figure 3 shows the process of metadata manipulation by using the GDT principle. The dotted line represents the corresponding files generated during the process, but when the

operation is completed, the system is deleted. Starting with the mapping of the first level, the GDT on the metadata server 2 where the `/bin/foo7` is stored is searched. Then, by using the second level algorithm, the relevant metadata information is directly operated on the metadata server 2. The advantage of the process is to operate metadata directly on the metadata server, so as to avoid space occupation caused by metadata migration, greatly save the space of the system and optimize the structure of the system.

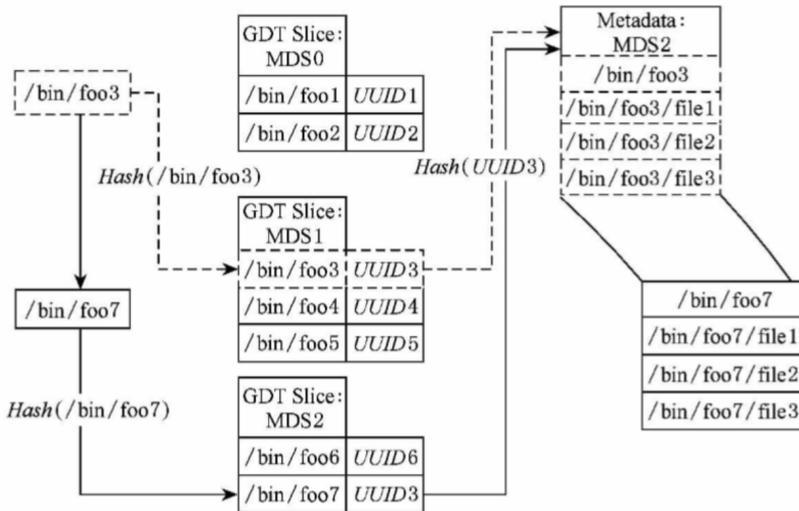


Fig. 2. Operation using GDT

### 3.3. Shared storage pool

The shared storage pool of Clover mainly solves the problem of consistency in distributed metadata analysis processing. Because of the mapping structure of the Clover two levels, Clover metadata is not stored on the corresponding metadata server, which is possibly from another metadata server. Storing data on a metadata server while operating it on another metadata server are likely to cause metadata differences. Therefore, in order to ensure consistency of metadata in the file system, Clover reaches two protocols with the shared storage pool in response to the two level mapping structure of Clover, so as to ensure that the system can recover quickly when the metadata failed or varied.

There are the preparatory stage and the execution phase of the two paragraphs of the agreement. In the preparation phase, the client proposes requirements, and the system selects the metadata server according to the mapping structure of the two levels and sends the ready commands to them. When the selected server receives the ready command, the censoring itself is carried out to ensure that it meets the requirements and can complete the commands. At this point, in the shared storage pool, the status of these metadata servers is generated, namely, the original data files.

When the selected metadata server reviews itself, the system is ready to complete the instructions, while the metadata server that does not pass the audit will not send ready to complete commands. When the selected metadata server sends the prepared instruction and the system receives it, the preparation phase is complete. When the system does not receive the ready and complete command of the selected metadata server within a specified time, the preparation is not completed. At this point, in the shared storage pool, records are made to metadata servers that are not ready for replies, at the same time, the system sends the failed records to the metadata server.

When the preparation phase is complete, the system enters the execution phase. As the execution phase begins, the system assigns roles to the underlying metadata server and issues different execution commands. At the same time, the shared storage pool will back up the relevant metadata server records. After the metadata server executes the relevant commands, the result is returned to the client and the execution of the command is resumed to the system. When the execution fails, the metadata server sends the canceled command to the system, and in the shared storage pool, a failed record is added to the metadata server's record, and the system sends a failed record to the metadata server.

### ***3.4. Metadata recovery mechanism***

In the face of huge amounts of data, it is inevitable that the data server will fail. With the expansion of Clove cluster size, the number of metadata increases with the geometric index, and metadata is used more frequently in the process of metadata processing and analysis. However, with the frequent use of metadata, there are failures, errors and other phenomena. How to recover metadata, optimize the system, and avoid the metadata server failures and node failures because of the storage errors or invalid metadata has become the focus of the research.

The Clover cluster is in the shared storage pool, and the metadata server is used for each time, backups and records are automatically generated. When the metadata server fails or fails to respond, backups from the shared storage pool play a role and operate on the associated records. In this way, both the system can be optimized and the metadata for invalidation or error can be recorded. In Fig. 3, when the 1 command execution system metadata server fails, the backup stored in a shared storage pool plays a role in the system and avoids no response.

## **4. Result analysis and discussion**

### ***4.1. Operational performance of metadata***

When metadata was operated under a Clover cluster, its performance was tested. These operations such as file information capture, creation, deletion, renaming, and subdirectory creation, may occur on a metadata server, or may be linked in series with several metadata servers and completed through multiple nodes. When the test was carried on, loads were added to the Clover cluster, files were created on

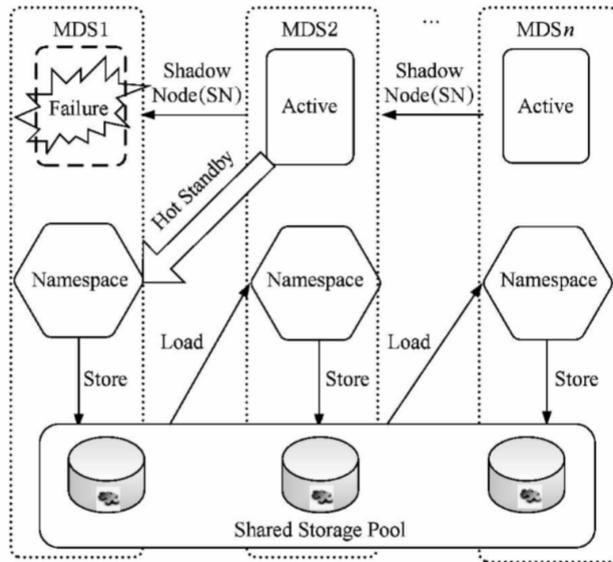


Fig. 3. Backup based on shared storage pool

each metadata server, and then the properties of the file were searched, and a series of operations such as creating sub files, renaming, deleting and so on were carried out, so as to enable the system to begin processing large amounts of data. For each of these operations, each operation was performed one hundred thousand times. The performance test results are shown in Fig. 4. Vertical coordinates showed the time taken by operations, which was used to measure the performance of metadata. Clover-nMDS stood for the number of metadata servers in the Clover cluster.

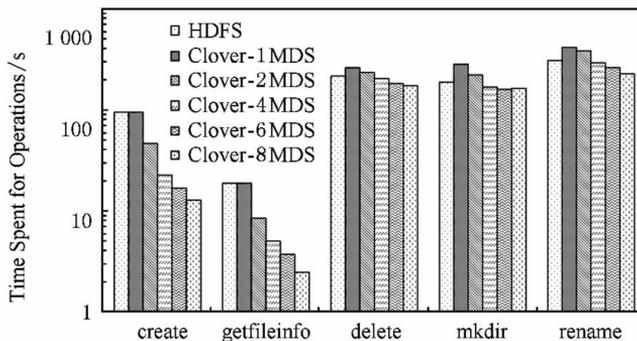


Fig. 4. Results chart of metadata operational performance testing

As can be seen from Fig. 4, under the condition of a metadata server for HDFS and Clover, there was no difference between the time required for file creation and information acquisition. However, with the increasing number of metadata servers, the time required for file creation and information acquisition became shorter and

shorter. Because with the increasing number of metadata servers in the Clover cluster, metadata transfers did not occur between each other, so that the time required to complete the operation became shorter and shorter. For file deletion, renaming, and the creation of sub files, the 1–2 metadata servers did not have any advantages over HDFS, but consumed more time instead. However, as the number of metadata servers increased, the time required for operations began to decrease.

#### *4.2. Metadata extension performance testing*

The extended performance of metadata is a key performance of large data processing and analysis. In the case of using a constant number of clients to operate together, HDFS was taken as a reference index, the metadata extension performances of Clover clusters with different metadata servers were compared in this paper. The number of clients should be more than 10, and twelve clients were selected for simultaneous operation.

Table 1. Metadata extension performance of different Clover metadata servers

Number of MDS Added	Create	Getfileinfo	Delete	Mkdir	Rename
1	159.32	156.45	-3.92	-9.35	-7.73
3	112.38	147.44	5.13	6.33	5.06
5	103.56	136.92	6.51	8.41	8.39
7	100.73	116.59	7.7	9.63	9.17

As can be seen from Table 1, as a large number of 10 clients operated simultaneously, the number of metadata servers in the Clover cluster increased, the performance of the element boards became better and better. For file creation and file information access, adding a metadata server had maximum scalability, which was the same as the previous metadata performance test results. However, for the metadata extension of file deletion, renaming and the creation of sub files, the more the number of metadata servers increased, the better the scalability was achieved. Because in the presence of multiple metadata servers, load metadata servers became the same, and were no longer focused on one or several metadata servers, the extensibility of metadata was improved.

#### *4.3. Metadata usability testing*

The availability of metadata refers to the time which is taken to recover a system when the metadata server fails. HDFS itself has a large number of spare nodes, and can detect the metadata server through periodic disk data detection, heartbeat detection and so on. When the metadata server fails, it is necessary to manually load the relevant files, check the metadata server and node, and restore the system through the standby nodes. The Clover cluster supports the backup and recovery of the metadata server through the backup data of synchronous metadata server in the shared storage pool and through the system failure mechanism. Table 2 shows the system recovery time required by the HDFS and Clover clusters for different size

dependent recovery files under the condition of a metadata server.

Table 2. Size and recovery time of the file

Size of the Checkpoint File/MB	HDFS	Clover
16	3.334	0.728
32	6.472	0.871
64	12.975	0.649
128	25.832	0.732
256	52.187	0.806

As can be seen from the table, the recovery time required by the Clover cluster was significantly smaller than that of HDFS, and the size of the recovery file was not related to the size of the Clover cluster as compared with the HDFS. The reason is that in the Clover cluster, the shared storage system can also backup metadata servers, so that no matter how large the file size is restored, the Clover cluster can always recover the system quickly and keep the service of the file system uninterrupted.

## 5. Conclusion

With the advent of the big data era, the processing and analysis of mass data become more and more common. In this paper, the key technologies of distributed system design for big data were studied, and the Clover cluster system based on the two-level mapping structure of the global distributed directory Hash table was proposed. With the backup and record of the shared storage pool as the metadata server, the system was optimized by using the two section decision method, and a stable and efficient file distributed system was obtained. Through related performance testing, it can be found that the Clover cluster system has better metadata operation, metadata expansion is also improved, the system stability is high, and the system optimization is better, thus providing a critical technology for distributed file systems for big data analysis.

With the above advantages, Clover cluster can ensure the operability of metadata, improve its scalability and availability, and lead to more stable and effective distributed file system. With the future optimization and improvement of Clover cluster, Clover cluster technology will occupy a more important position in the key technology centers of big data oriented distributed files.

## References

- [1] Y. WU, F. YE, K. CHEN, W. ZHENG: *Modeling of distributed file systems for practical performance analysis*. IEEE Transactions on Parallel and Distributed Systems 25 (2014), No. 1, 156–166.
- [2] J. S. KIM, K. Y. WHANG, H. Y. KWON, I. Y. SONG: *PARADISE: Big data analytics*

- using the DBMS tightly integrated with the distributed file system.* World Wide Web 19 (2016), No. 3, 299–322.
- [3] P. MALI, O. KULKARNI, V. SHAH, K. PASTE, M. BHANUSHALI: *Public sharing file system using google drive.* International Journal on Recent and Innovation Trends in Computing and Communication 3 (2015), No. 3, 1550–1553.
  - [4] R. MALHOTRA, J. M. MELLICHAMP: *Intelligent simulation code generator: A relational database management approach.* Journal of Intelligent Manufacturing 8 (1997), No. 2, 125–136.
  - [5] J. DEBRABANT, A. PAVLO, S. TU, M. STONEBRAKER, S. ZDONIK: *Anti-caching: A new approach to database management system architecture.* Journal Proceedings of the VLDB Endowment 6 (2013), No. 14, 1942–1953.
  - [6] S. A. GEBREZGABHER, M. P. M. MEUWISSEN, A. G. J. M. O. LANSINK: *A multiple criteria decision making approach to manure management systems in the Netherlands.* European Journal of Operational Research 232 (2014), No. 3, 643–653.
  - [7] X. HUA, H. WU, S. REN: *Enhancing throughput of Hadoop distributed file system for interaction-intensive tasks.* Journal of Parallel and Distributed Computing 74 (2014), No. 8, 2770–2779.
  - [8] N. CAO, Z. H. WU, H. Z. LIU, Q. X. ZHANG: *Improving downloading performance in Hadoop distributed file system.* Journal of Computer Applications 30 (2010), No. 08, 2060–2065.
  - [9] Y. S. JEONG, Y. T. KIM: *A token-based authentication security scheme for Hadoop distributed file system using elliptic curve cryptography.* Journal of Computer Virology and Hacking Techniques 11 (2015), No. 3, 137–142.
  - [10] W. W. SMARI, S. FIORE, C. TRINITIS: *Recent developments in high-performance computing and simulation: distributed systems, architectures, algorithms, and applications.* Concurrency and Computation, Practice and Experience 27, (2015), No. 9, 2191–2195.
  - [11] J. ZHOU, W. WANG, D. MENG, C. MA, X. GU, J. JIANG: *Key technology in distributed file system towards big data analysis.* Journal of Computer Research and Development 51 (2014), No. 02, 382–394.
  - [12] H. C. HSIAO, H. Y. CHUNG, H. SHEN, Y. C. CHAO: *Load rebalancing for distributed file systems in clouds.* IEEE Transactions on Parallel & Distributed Systems 24 (2013), No. 5, 951–962.
  - [13] A. MEI, L. V. MANCINI, S. JAJODIA: *Secure dynamic fragment and replica allocation in large-scale distributed file systems.* IEEE Transactions on Parallel & Distributed Systems 14 (2003), No. 9, 885–896.
  - [14] S. MILTCHEV, J. M. SMITH, V. PREVELAKIS, A. D. KEROMYTIS, S. IOANNIDIS: *Decentralized access control in distributed file systems.* ACM Computing Surveys 40 (2008), No. 3, 1–30.
  - [15] C. K. LEE, H. A. EAGLES, J. R. CARADUS, K. F. M. REED: *Investigation of yield and persistence of white clover using cluster analyses.* Euphytica 72 (1993), No. 3, 219–224.

Received July 12, 2017

# Research on application of DDoS attack detection technology based on software defined network<sup>1</sup>

GUO JING<sup>2</sup>

**Abstract.** Distributed denial of service (DDoS) attack is one of the main threats to network security. Although the principle and mechanism of DDoS attack has been widely understood and researched, the detection and prevention of the attack behavior is still very arduous, because this form of attack has high degree of concealment and the dynamic distributivity. In this paper, the characteristics of software defined network and the characteristics of DDoS attacks are analyzed and researched. An improved DDoS attack detection method is proposed, which is different from the traditional method that only detects the single link and network for the victims. Based on the software defined network (SDN), classification learning algorithm is used to detect DDoS attacks online through the construction of an efficient global network flow table feature values, and combined with the feature selection algorithm that can optimize the flow table feature sequence. The experimental results show that for the DARPA 99 and CAIDA DDoS 2007 two typical real data sets, the method can improve the detection rate of the DDoS attacks and significantly reduce the false alarm rate, and it has good comprehensive performance.

**Key words.** Software defined network, DDoS attack detection, feature selection algorithm.

## 1. Introduction

The Internet has become an important information infrastructure in modern society. The rapid popularization and wide application of the Internet have profoundly changed the way of human life, and the Internet has become an indispensable part of daily affairs. A short service interruption may jeopardize people's normal work and life, the financial order, economic stability, and even national security, bringing immeasurable loss. Therefore, it is of great significance to ensure the continued applications of network for the maintenance of network security. Distributed denial of service attack (DDoS) [1] sends a large number of data packets to the target host by

---

<sup>1</sup>The authors acknowledge the Special research project of Shaanxi Provincial Department of Education(17JK2038).

<sup>2</sup>School of Information & Engineering, Shaanxi Institute of international Trade & Commerce, Xi'an, 712046, China

multiple hosts, making the target host resources unable to provide normal services because of excessive consumption. DDoS shows the characteristics of strong attack intensity, strong source of attack, attack range and strong concealment on attack means, characteristics and effects [2]. Compared with other attacks, DDoS attack is simple, the destructiveness is strong, and it is difficult to detect and defend.

## 2. Related technology

### 2.1. *OpenFlow protocol*

The OpenFlow protocol is the communication interface standard between the controller and the switch controller [3]. The configuration and management of controller to switch is through the regulation message types of protocols, including symmetric message, the controller-to-switch and asynchronous message, and each kind of news has many types of sub-news. Among them, the symmetric message can be initiated by either the switch or controller, and is used to establish or maintain the connection between controller and switch; the controller-to-switch messages is initiated by the controller, and is used to acquire or manage the state of the switch for the controller. The asynchronous message is initiated by the switch, and is used for the exchange to notify the state changes or network event to the controller.

According to the OpenFlow protocol, the flow table in the switch is the basis for transmitting of the data packets. Each of which consists of a plurality of flow table entries [4]. The flow table represents the data forwarding rules, including the matching domain, operation, counters. The matching domain is used for matching packets, so as to use rich rules in forwarding a packet, including most of the key identifications of the link layer, network layer and transport layer. Each address identification can be a certain value or any value, but also can be used to achieve a more accurate matching. The operation shows the actions to the data packet that the matching is successful, such as forwarding to a port, packet loss or modify the packet header information. For the data package that does not match the flow table, the switch will encapsulate it to sent to the controller through a secure channel, and the controller will decide what kind of operations on it. The counter is used for the the statistical of the basic information of data stream, including the number of packets and the number of bits that match the flow table.

### 2.2. *DDoS detection technology*

According to the DDoS attack detection strategy, the detection technology can be divided into two kinds: anomaly detection and misuse detection. Misuse detection can match the known attack characteristics with the collected and observed user behavior characteristics [5]. The system can quickly judge the attack behavior. At present, the most representative misuse detection methods include expert system, pattern matching, state transition analysis and so on. Different from other network attacks, the content of the DDoS attack is legal without any vicious data code, so it is difficult to extract and detect the inherent characteristics of the attacks.

Therefore, the detection efficiency based on misuse DDoS attack detection method is limited, and the missing report rate is higher. Anomaly detection method is based on the establishment of the normal behavior model of the target user and the system to compare the deviation between the measured behavior characteristics and the normal behavior model to identify the attack [6]. This method not only can detect DDoS attacks effectively, but also can detect new attacks that is similar to attack features. This paper mainly studies the anomaly detection for DDoS

### 3. Attack detection

In this chapter, we design and implement a DDoS attack detection method based on the characteristics of flow table. This method is applicable to software defined network environment [7]. Through the analysis of the characteristics of DDoS and Open flow protocol, combined with mutual information to select the optimal flow surface features, the classification algorithm of attack detection is used to realize the comprehensive and effective attack behavior judgment.

#### *3.1. Attack detection method based on the characteristics of flow table*

The basic idea of attack detection method based on flow table feature is that the flow table information in the software defined network switching equipment is extracted to convert into feature vectors, and the optimal feature is extracted to build the attack detection classifier [8], and finally the new network flow is classified. In this section, we introduce an attack detection method which can be applied to software defined network from the flow table feature matrix and feature selection algorithm.

#### *3.2. OpenFlow flow table features*

The characteristics can be used to characterize some known attacks, is a description for the attack behavior [9]. For each kind of network attack, we should extract the characteristics of the attack, in the ideal state, we should always be able to detect and identify malicious attacks through these features. The feature vector of the flow table is a high-level abstraction of the attack and normal data, which is the basis for the identification of the attack.

## 4. Experiment and analysis

### *4.1. Experimental environment*

In this paper, DARPA 99 is used as normal data set Trace normal, CAIDA DDoS2007 data set is used as a base of the generation of abnormal flow Tracel, and the attack in DARPA 99 data set ia taken as Trace2 abnormal flow, the ab-

normal flow of Tracel and Trace2 are mixed as the attack data sets Trace DARPA 99 is attack attack. DARPA 99 is used for the laboratory evaluation of intrusion detection system by MIT Lincoln laboratory and the U.S. Air Force Research Laboratory, and each data of the data set has recorded the detailed information of data packet. CAIDADDoS 2007 contains approximately one-hour DDoS attack flow data, in which the flow data contains only the attack flow rate and the response of the destination host, which is stored in pcap format.

Firstly, the normal data set and the attack data set are sampled periodically, and the sampling periods are recorded as  $t$ , and  $T$ , respectively. At the same time, the flow set is converted to a set of flow table items. The average packet number, average number of bits, flow table rate, single stream item rate, request flow ratio, source address entropy, source port entropy, and destination port entropy 8 kinds of attributes are obtained to generate the experimental samples and the signature sequences are signed after the flow table feature of the generated flow table item set are counted. The characteristic attributes are 8 categories, which are labeled as 1–8, and the classification attributes are classified into 2 categories, namely: 1 and -1.

KNN, SOM and SVM algorithms are used to learn and test data sets. The analysis tool used by SVM classifier is the LIBSVM software package. KNN and SOM use the analysis tools in Matlab. In this paper, the detection rate, false alarm rate and the total error rate these 3 evaluation indicators are used to assess the effectiveness of the test, which are expressed as  $DR$ ,  $FR$ ,  $ER$ , respectively, as shown in the following formulae:

$$DR = \frac{TN}{TN + FN}, \quad (1)$$

$$FR = \frac{FP}{TP + FP}, \quad (2)$$

$$ER = \frac{FN + FP}{TP + FP + TN + FN}. \quad (3)$$

Here,  $TN$  is the number of attack samples labeled in the attack samples to be tested,  $FN$  is the normal samples number labeled in the attack samples to be tested,  $TP$  is the normal samples number labeled in the normal samples to be tested,  $FP$  is the number of attack samples labeled in the normal samples to be tested.

## 5. Experimental results

### 5.1. Feature selection experiment

*5.1.1. Influence of selected feature sets on different classifiers.* In this paper, we propose a feature selection method based on mutual information, which can stop the search of the feature subset after getting the characteristic attributes of the setting number, thus generating 8 different feature subsets. KNN, SOM and SVM three kinds of classification algorithm can test the different obtained characteristics, in which the SVM function (RBF) is the kernel function. The classification effect

test includes detection rate, false alarm rate, and index. The distribution of feature selection sample set is shown in Table 1.

Table 1. The number distribution of feature selection samples

		Training set	Test set
Normal sample Trace normal		500	2000
Attack samples Trace attack	Trace1 attack	200	1400
	Trace2 attack	200	400

Figures 1 and 2 show the detection effect of classifier on  $k$  ( $1 < k < 8$ ) kinds of sets, which are detection rate, false alarm rate and error rate. When  $k \geq 4$ , the classification detection rate of SVM algorithm is higher than the KNN and SOM algorithm. Only when the feature number is 3, the error rate of SVM algorithm is lower than the SOM algorithm, and in other cases, it is higher than the other two algorithms. When the classification error of SVM algorithm  $K$  is 4,5, it is lower than the other two algorithms. Through the analysis of test results of the classifier, it can be seen that the SVM algorithm is better than KNN. SOM algorithm in the detection rate, but the false alarm rate and error rate are relatively low in some cases, but higher than the other two algorithms in other cases.

The KNN algorithm achieves the highest detection rate when  $k=6$ , and has the lowest false alarm rate and error rate. When  $k = 3$  and  $k = 4$ , the SOM algorithm and the SVM algorithm achieve good detection results, respectively. The characteristic attributes and the time consumption corresponding to the three kinds of classification algorithms are shown in Table 1. The characteristic attributes have 8 categories, respectively, labeled as a 1–8.

It can be seen from Table 1, the optimal characteristics number of the SOM algorithm is the least, but the training time is the most. On the other hand, the KNN algorithm has the largest number of characteristics and the shortest classification time. The SVM algorithm is between the two, and its classification time is far lower than that of SOM algorithm, but by about 0.0156 seconds higher than KNN algorithm, so it has better classification efficiency.

This paper is mainly based on feature selection algorithm, the classifier of the classification efficiency of the optimal characteristics sample on DDoS attack detection effect in the number of different characteristics options is considered. The selected feature number  $k = 4$  and SVM algorithm are taken as the classification algorithm in subsequent tests, which has a higher detection rate and lower error rate.

*5.1.2. Distribution of the selected feature attributes.* Table 2 shows that when  $k = 4$ , the selected feature categories of the SVM algorithm are 3, 4, 6,8, respectively representing the feature types: flows rate, single flow rate, source address entropy and destination port entropy, which can be expressed as FR, SFR, H (sip) and H (deport). In order to better reflect the distribution of the characteristics value, each column of characteristics in the sample set are normalized, at the same time, the

feature sequences after standardized have not units.

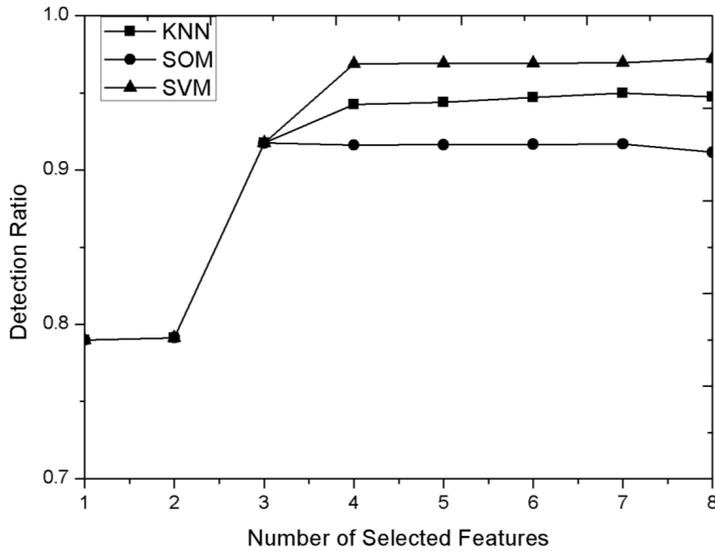


Fig. 1. Detection effect of classifier on  $K$  subset–Detection Ratio

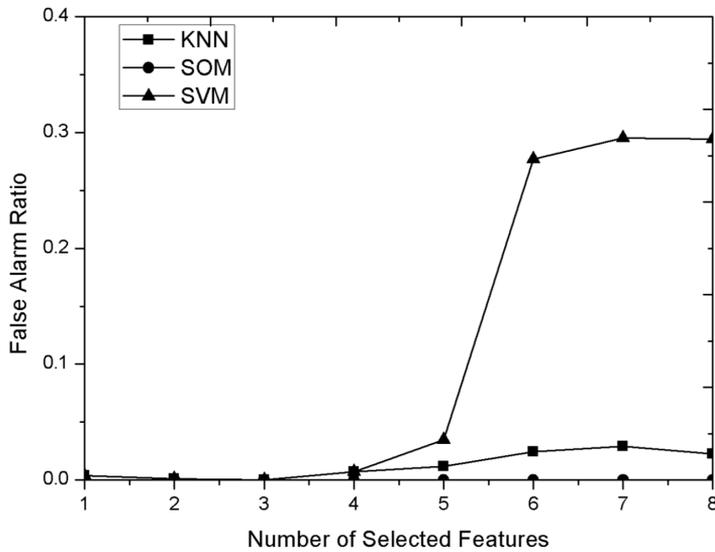


Fig. 2. Detection effect of classifier on  $K$  subset–False Alarm Ratio

These two kinds of feature samples have a clear interface when the sampling frequency of the attack characteristics is 100 times of the normal characteristics. It can be explained that the feature attribute of the optimal feature subset obtained according to the characteristics selected algorithm can better reflect the difference

between the normal flow and attack flow, and enhance the ability to recognize attack.

Table 2. Classification efficiency of best feature selection samples

Classification algorithm	Characteristic attribute	Classification time (seconds)
KNN ( $k = 6$ )	2, 3, 4, 5, 6, 8	0.0468
SOM ( $k = 3$ )	4, 6, 8	38.8910
SVM ( $k = 4$ )	3, 4, 6, 8	0.0624

## 5.2. Attack detection experiment

According to the feature selected experiment, we can know that the classification algorithm of attack detection is the support vector machine, and the optimal flow table number is  $k = 4$ . First of all, the obtained attack data set is converted to the flow table set with normal data set, and the flow table set is sampled periodically, while the statistics of its flow table characteristics is completed, including flows rate, single flow rate, source address entropy and destination port entropy. The number distribution of characteristic samples in this experiment is shown in Table 3.

Table 3. Classification performance of SVM sample data with different kernel functions

	Kernel functions	Detection rate
Liner kernel function	$K(x_i, y_j) = x_i^T x_j$	99.76 %
Polynomial kernel function	$K(x_i, y_j) = (yx_i^T x_j + b)^d$ $y = 0.001, b = 0, d = 3$	99.73 %
RBFkernel function	$K(x_i, y_j) = \exp(-y \ x_i - x_j\ ^2)$ $y = 0.001$	93.76 %
Sigmoid kernel function	$K(x_i, x_j) = \tanh(yx_i^T x_j + b)$ $y = 0.1, b = 2.1$	92.49 %

After determining the normal and attack samples, the detection rate of the classifier is taken as the evaluation index, and the kernel function of the SVM is determined to be  $K(x_i, y_j) = h(x_i) \cdot h(x_j)$ , realizing the mapping from input space and feature space. As shown in table 3-S, the kernel function with higher detection rate is selected for further analysis by comparing the classification performance of SVM classifier with different kernel functions on sample data.

By comparing the classification results of four kinds of kernel functions in table 4, we finally choose the linear kernel function with higher detection rate as the  $k(x_i, y_j)$  of SVM, and then proceed to the next classification experiment. The selected flow table feature training set is used as the input data of the classifier, and the classification model is constructed. The test set is used to test the classification model.

Table 4. Classification effect before and after feature selection

	Detection results		Classification time (s)
The flow table feature after selection ( $k = 4$ )	DR	99.76 %	0.0368
	FR	0.3 %	
	ER	0.27 %	
The flow table feature after selection ( $k = 8$ )	DR	95.56 %	0.0368
	FR	0.6 %	
	ER	2.54 %	

Table 4 is the comparison of the classification performance of the SVM classifier to the feature set after selected and the feature set before selected. It can be seen that the detection rate of the former is higher than that of the latter, and has lower false alarm rate and error rate. At the same time, the input test feature sequence has different dimensions, which makes the flow table after selected feature set have faster classification efficiency.

### 5.3. Comparative analysis

The detection rate, the false alarm rate and total error rate are selected as the evaluation index, and the detection after selected feature is taken as the contrast method, as shown in Table 5.

It can be seen in table 5 that compared with the traditional network attack detection method, the detection algorithm in this paper has higher detection rate and lower error rate. Compared with OpenTAD method of software defined network, this method has high detection rate and low false alarm rate. It can be concluded that the attack detection method based on the flow table features in this paper has good comprehensive performance, and can effectively identify DDos attacks.

## 6. Conclusion

This paper mainly introduces the method and experiment of DDos attack detection based on the characteristics of flow table. First of all, the characteristics of OpenFlow flow table and attack flow are analyzed, and the characteristic matrix of flow table for all communication IP is constructed. Secondly, the correlation between the self and the categories and the features is considered, the feature selection algorithm is designed based on the comprehensive correlation which is taken as the evaluation function of feature selection. Next, the DDos attack detection method and its implementation based on the flow table characteristics is described. The feature selection algorithm is used to preprocess the flow table features, and the optimal characteristic attribute is chosen to build the classification model on the basis of the treated samples so as to identify the attack. Finally, the realization of attack detection method is based on the Matlab, and the experimental results of this

method are verified and analyzed. The data shows: the classification model after the feature selection has a good classification effect and shorter classification time. At the same time, compared with the previous attack detection method, it has good comprehensive performance.

Table 5. Detection results of different algorithm in SDN networks and traditional networks on DDoS attacks

	Detection results	
	Software defined network (the flow table feature after selection, $k = 4$ )	DR
FR		0.3 %
ER		0.27 %
Software defined network (Open TAD)	DR	91.7 %
	FR	4.2 %
	ER	...
Traditional network (TFCE)	DR	97.4 %
	FR	0.1 %
	ER	1.4 %

## References

- [1] Q. YAN, F. R. YU, Q. GONG, J. LI: *Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: A survey, some research issues, and challenges*. IEEE Communications Surveys & Tutorials 18 (2016), No. 1, 602–622.
- [2] S. SCOTT-HAYWARD, S. NATARAJAN, S. SEZER: *A survey of security in software defined networks*. IEEE Communications Surveys & Tutorials 18 (2016), No. 1, 623–654.
- [3] K. GIOTIS, G. ANDROULIDAKIS, V. MAGLARIS: *A scalable anomaly detection and mitigation architecture for legacy networks via an openflow middlebox*. Security & Communication Networks 9 (2016), No. 13, 1958–1970.
- [4] R. LATIF, H. ABBAS, S. ASSAR: *Distributed denial of service (DDoS) attack in cloud-assisted wireless body area networks: A systematic literature review*. Journal of Medical Systems 38 (2014), No. 11, 1–10.
- [5] R. YAN, Q. ZHENG, H. LI: *Combining adaptive filtering and IF flows to detect DDoS attacks within a router*. KSII Transactions on Internet and Information Systems 4 (2010), No. 3, 428–451.
- [6] A. MANIMARAN: *An extemporized confidence based filtering technique to mitigate Ddos attack in cloud environment*. International Journal of Control Theory and Applications 8 (2015), No. 5, 2405–2413.
- [7] P. S. MANN, D. KUMAR: *A reactive defense mechanism based on an analytical approach to mitigate DDoS attacks and improve network performance*. International Journal of Computer Applications 12 (2011), No. 12, 43–46.
- [8] N. A. SURYAWANSHI, S. R. TODMAL: *DDoS attacks detection of application layer for web services using information based metrics*. International Journal of Computer Applications 117 (2015), No. 9, 22–30.

- [9] N. JEYANTHI, N. C. S. N. IYENGAR: *An entropy based approach to detect and distinguish DDoS attacks from flash crowds in VoIP networks*. International Journal of Network Security *14* (2012), No. 5, 257–269.

Received July 12, 2017

# Research on the application of digital media art in animation control based on Maya MEL language

HONGJI CHEN<sup>1</sup>

**Abstract.** Digital media industry is a modern industry produced by the media, network technology, intelligent terminal technology and joint cultural industry, and has been applied to many fields at present. Thus, in this paper, in order to study the application of digital media technology based on Maya MEL language in animation control, from the point of view of the form of digital media art work, and combining Maya production practice, the technology of controlling character animation in MEL language was selected, and the solution to the problems encountered in manual animation control was studied. The final experimental results show that digital media can show the effective representation of digital media art in Maya design, so as to meet the target audience's cognitive law, and to achieve the effective dissemination of knowledge in the era of digital communication.

**Key words.** Digital media art, 3D animation, Maya MEL language.

## 1. Introduction

Digital media art is not only a technology but also an art, which is formed by the integration of many disciplines, covering the computer image processing technology, multimedia technology, art design and other technologies [1]. It is mainly divided into three categories: firstly, from the perspective of its overall connotation, computer technology runs through the whole design process. And the software and hardware facilities are integrated to carry out the creation, with the form of art to show the quality of the data to create a sense of beauty. Secondly, from its application, huge data is optimized, which is depicted by computer technology in a more appreciative manner, breaking through the traditional mode of creation. From the perspective of application trends, the third one is to use computer technology, network technology, 3D printing technology, intelligent identification technology, and mobile communications software and hardware equipment to simulate and com-

---

<sup>1</sup>College of Information Engineering, Chongqing Industry PolyTechnic College, Chongqing, 401120, China

pound works of art. Therefore, even if a large amount of data is missing in the process of making works of art, the original form of the artistic expression can be restored through the integration of multi-information data and multi-dimensional technology [2]. The combination of computer science and art design has produced the subject of digital media art design, which has many important effects on human beings. With the help of digital media technology or electronic media technology, industrial design products and art works can be displayed perfectly. Through the emerging media such as mobile phones, CD-ROM, the Internet, people can watch works easily and quickly. Multimedia web pages and interactive packaging art bring people good visual experience [3]. More than a result of technical creation or the repeated superposition of some forms of artistic creation, the creation of digital media art is built on the basis of artistic creation through a series of forms such as rewriting, reconstruction and renewal. In the aspect of art, it not only has its own unique significance, but is also reflected in the modernization, basic form and basic theory of aesthetics [4]. The digital media is becoming more and more popular in our daily life, which gradually integrates the mass culture and has the function of serving the society. The design of digital media art covers many subjects such as computer technology, media technology, art, design and so on [5].

## 2. State of the art

### *2.1. Present situation and development trend of digital media at home and abroad*

Through the use of information technology and digital technology, and with the help of the carrier, digital media can spread the information through audio, video, text, pictures and other forms of media expression. With the rapid development of science and technology, the development of computer science and digital media should not be underestimated. Nowadays, the mass media has changed a lot [6]. The new digital media has gradually replaced the traditional media, and is rapidly moving towards digital. The development of digital video, audio, film and so on is also very fast. At present, digital media technology has penetrated into every field of life. Culture, education, economy, management and so on need the use of digital media, especially in the field of art. The appearance of digital media has greatly increased the form of art diversification. The process of digitalization of communication information refers to the processing of information collected, so that the form and content of communication can be digitized. Therefore, the digital media has become the latest information carrier after the language, the text, the picture, the electronic technology and so on [7].

With the development of the culture and communication industry, and their demand for high-tech applications, in recent years, the development of digital TV technology, communication technology and computer network technology has promoted the development of multimedia industry and cultural media industry. The multimedia industry has gone deep into cultural education, art design, entertainment and publishing, economic management and other aspects. And the combination of

digital media and computer, film and television, education has promoted the growth of knowledge economy. The multimedia industry in China starts relatively late, but the development speed is very fast. In particular, the development prospects of creative companies with high technology and culture are very optimistic [8]. The multimedia industry has also made a great contribution to the national economy. In the past ten years, the registered animation and related industries in China have expanded from a few hundred to more than 5 thousand, gradually forming the animation industry chain, providing jobs for nearly ten thousand people. The rapid development of small and medium sized enterprises in China and the introduction of transnational capital have injected fresh blood into the information economy of our country. The multimedia industry has played an important role in the sustainable development of economy and the adjustment of economic structure in China.

## *2.2. Application of Maya and MEL language*

The most influential 3D animation software Maya is designed by the American Autodesk company, which is suitable for the design of movie special effects scene, character animation and video advertisement. The software is simple and easy to use [9]. Maya is more perfect than ordinary 3D visual effect. Combined with the forefront of cloth simulation digital, hair rendering, modeling and other technologies, realistic rendering can be enhanced. It can be said that Maya is the most professional film making software. In today's tools for digital and three-dimensional production, Maya can be run in a common operating system to become the first choice for the development of three-dimensional animation film and television works. There is an enhanced Script called MEL. And Command language is the core engine of Maya software, which is a kind of embedded language in Maya, belonging to a kind of command and script language. The purpose is to enable the user to control the Maya commands as well as working procedures and processes [10].

The basic architecture of Maya is provided by MEL, and all of the key points in the software are from MEL commands and scripts. MEL extension and Maya functions can be used to further develop the software to meet the needs of users more practical. In addition, MEL can also directly control the Maya command, work procedures and processes. Because each function and operation of Maya are written in MEL program, every time a Maya operation is equivalent to the implementation of a MEL [11]. MEL is the foundation of the whole Maya software. By using the features of MEL and Maya, the designer can define and extend the function of Maya software. The MEL language in Maya is a platform that contains a number of built-in functions and commands. These functions and commands belong to the upper level [12]. Because each function in Maya is achieved through the MEL language, all command buttons in the Maya interface are controlled by the MEL language. Through the MEL language, the menu command in the Maya interface can be used more conveniently, and more properties of Maya can be accessed. MEL language is a very flexible language, and thus animators can even use the MEL language to edit their own widgets or menus.

### 3. Methodology

#### *3.1. MEL language features*

In the Maya software, the instruction set and the function are the essential components in the creation process. At the same time, MEL language will also be used to write and define the collection of procedures required. There are five main functions.

Firstly, the process controller of the scene state and the scene modifier is created to provide the corresponding method for the control of the scene animation and the modification of the object. Secondly, the realization of interactive control of the program depends on the command line window and the monitoring window [13]. Thirdly, the scroll bar of the working interface is mainly to implant the preparation of MEL language, so as to provide a more standard Maya user interface for MEL. Fourthly, its user input/output tools are created primarily by inserting the I/O file. Finally, it has many features of the Maya software, such as the construction model, lamp settings, animation editing, texture, rendering imaging.

#### *3.2. Construction of 3D animation model based on Maya MEL language*

It is a very important work and technology to build 3D model in Maya, which is the basic condition of 3D animation. Maya software modeling technology includes the following: paint (brush) modeling, Nurbs (surface) modeling, Polygon (polygon) modeling, Subdiv (subdivision) modeling and modeling of some special technologies. And material, lighting, rendering and other operations will be based on this [14]. In general, the creators will rely on Maya software modeling technology to complete the physical entities, such as people, animals, objects, scenes and even natural landscapes and other shapes of animation, then which must be translated into a digital computer language. In fact, the scene of Maya is a database of some computer languages based on numbers. The above information will be interpreted by the Maya script interpreter, and can be displayed on the computer screen with a friendly and visual interface. It is intuitive and convenient to build a model in a general interaction in the Maya software. However, when multiple objects are repeated, the use of MEL language to create a model will have a unique advantage, which is not only more convenient, but also can be used to create, edit and modify the properties of 3D objects when the precise control of the model is strictly required.

#### *3.3. Analysis of the principle of dynamic animation with MEL language system*

Another important function of the MEL language in the Maya software is the establishment of the property map (Dependency graph/DG) node. In order to achieve the special effect, the scene playback must depend on the connection between the nodes. At the same time, MEL language and animation elements are closely linked in Maya software. Then, the main effect of MEL language is to control the properties of the object, using the equal sign to assign the object attributes [15].

In the creation of animation, nodes establishment, node connections and other operations were carried out on the back end. In the case analysis, the monkey was chosen as the experimental model, as shown in Fig.1. Each monkey in the scene would appear in the three interrelated nodes, make Nurb Baby1, NURB Baby Shape1, and NURB Baby1.

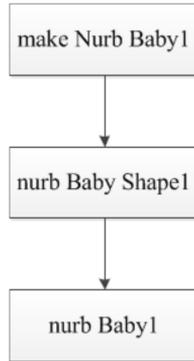


Fig. 1. Node of the monkey

As can be seen in the Maya software, the creation of attributes and prototype animation operations were equipped with more complex meaning. The MEL language can also simplify the work of node network. The definition of MEL in Maya was more complex than the simple attribute value connection. Its working principle was as follows: a new MEL expression node was created by using the input value of the channel bar or the expression editor, including the expression script. At the same time, the calculation method of the data can be output/input. Instead of making a big change on the whole, only changing the first node and the second node can rebuild the calculation method.

#### 4. Result analysis and discussion

The use of MEL language control can simplify the transformation of the role model handle. Similarly, it is very effective to use the MEL to simulate the motion of the individual in the process of dynamic individual expression. Language was used to simulate dynamic process, and MEL was used to identify the movement process of the individual, including the change of the movement rules and the influence of the mass on the movement. MEL language was adopted to simulate the dynamic process of monkey tail skeleton, so as to realize the control of a certain characteristic image in many images.

As shown in Fig.2, the monkey tail skeleton model was constructed. It was assumed that the tail skeleton was a six segment active connector, and six linked modules were built to simulate the skeleton. The monkey's body was constructed, displayed as a square script and connected to the skeleton module.

Before the measurement and control, the value of the acceleration of the square script in the direction of  $x$ ,  $y$  and  $z$ , as well as the value of the square script itself

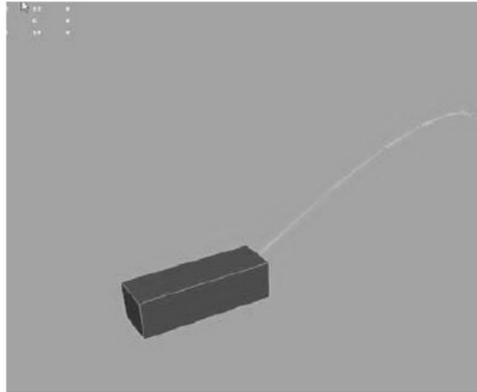


Fig. 2. The model of monkey skeleton

were determined.

The control of monkey tail was realized by the acceleration of the six modules connected with the square body. Therefore, the acceleration of the square body along the  $x$ ,  $y$ , and  $z$  directions was represented by a name. In addition, a sensitivity parameter  $K$  was added. The parameter  $K$  was used as an independent variable and was expressed in MEL design language. Then the  $K$  property was bound to somewhere in the six module. Through the change in the size of the  $K$  value, the actual value of the control speed was increased, and the adjustment and control of the tail sensitivity were realized. Figure 3 showed the effect of applying it to a monkey model.

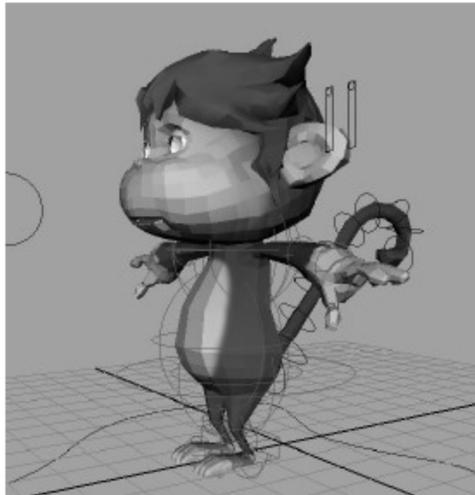


Fig. 3. The effect of applying the control to a monkey model

The numerical values that required to be expressed in MEL language were mainly the numerical values of the dynamic characteristics, including the velocity, angular

velocity, direction and speed of the body. The way of numerical transformation was to express partial direction change of body by the relationship among the value, direction, and velocity of angular velocity. Moreover, the value of the direction change of the angular velocity can indicate changes in this direction. At the same time, the increase of velocity in the  $x$  direction and  $y$  direction can be calculated by the accumulation of body velocity. Through the calculation of the change of velocity and the value of the direction, the expressions of  $x$  direction and  $y$  direction were realized.

As shown in Fig. 4, other bones were basically the same, but the difference was the correction of individual parameters. It should be noted that the above mentioned 1.5 of the formula was the length of the first skeleton of the monkey. And with the actual situation of different bones, other joints can make different changes accordingly.

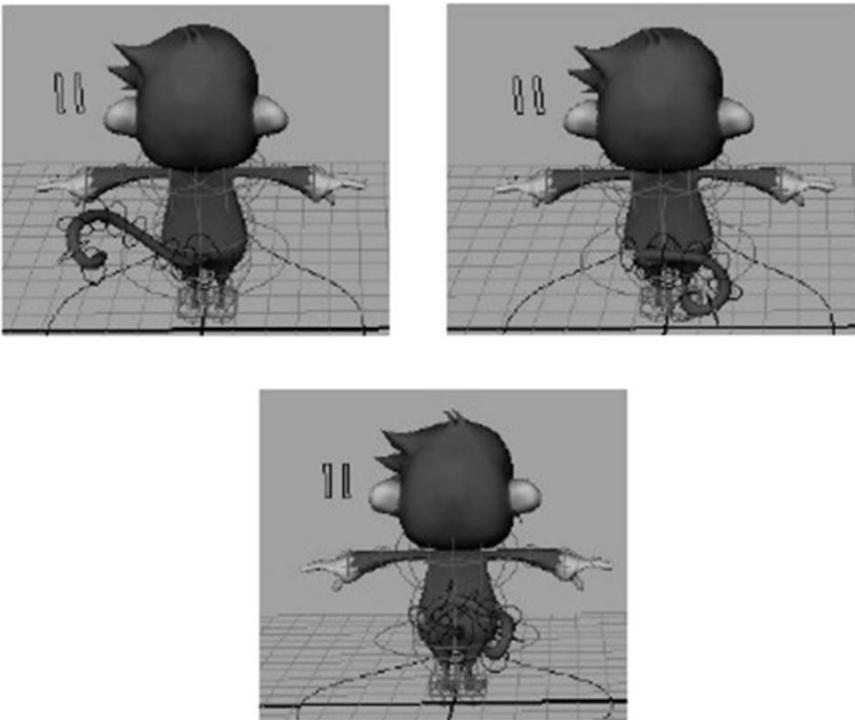


Fig. 4. The effect of applying the control to a monkey model

When using MEL language to simulate the dynamics, different precision requirements should be determined according to different simulation objects. The six modules connected by the square represented the monkey tail bones. Due to the special nature of the creature, the monkey will often roll tail, and swing tail. Therefore, the MEL language model was only suitable for expressing the dynamic motion of the tail with small amplitude. In order to realize the dynamic simulation of high curvature, the MEL precision control was needed.

## 5. Conclusion

With the development of digital media, the relationship between the Internet and the IT industry is getting closer and closer, which will become an important force in the development of new media industry and promote the development of the whole industry in the near future. Based on this background, MEL language was used as the development language design, and the three-dimensional animation was developed in this paper. Moreover, combined with specific examples, further study and research on the strategies and methods of 3D character animation design were carried out.

The functions and features of MEL language were analyzed in this paper. The method of producing animation with MEL language system was written. And MEL language simulation and animation control were described in detail. On this basis, the method of animation control in MEL language was summarized, and the advanced theory of MEL language control animation was obtained. Through these methods, the problems existing in the traditional animation were solved effectively. Then the basic theory of 3D animation and the architecture of the whole system were clearly recognized. Based on the function and structure of computer 3D animation, the analysis was carried out from the aspects of character animation, motion editing and so on. When the model was created in Maya software, the MEL language was used to control the action. Through the MEL language programming, Maya character animation was produced, and group animation control system was also set up. Finally, all previous research results and processes were summarized, and character animation control system was designed and implemented.

However, due to the limitations of professional standards, there are still many problems in the MEL programming language and Maya animation, thus needing further studying in the future.

## References

- [1] R. HARRIS: *Design on the Go: How African American youth use mobile technologies for digital content creation*. Design and Technology Education 15 (2010), No. 2, 9–17.
- [2] F. CARROLL, R. KOP: *A learning, research and development framework to design for a 'Holistic' learning experience*. E-Learning and Digital Media 8 (2011), No. 4, 315–326.
- [3] P. HU: *Subjective choice and active participation: The application characteristics of digital media designing on network business platform*. Zhuangshi Chinese Journal of Design (2012), No. 226, paper 143.
- [4] V. M. BARABASH, V. I. BEGICHEV, M. A. BELEVITSKAYA, N. N. SMIRNOV: *Problems and trends in the development of the theory and practice of mixing of fluid media*. Theoretical Foundations of Chemical Engineering 41 (2007), No. 2, 130–136.
- [5] C. BECKSTEIN, G. GOERZ, D. HERNÁNDEZ, M. TIELMANN: *An integration of object-oriented knowledge representation and rule-oriented programming as a basis for design and diagnosis of technical systems*. Annals of Operations Research 16 (1988), No. 1, 13–31.
- [6] J. C. HOLLMANN, K. CRAUSE: *Digital imaging and the revelation of 'hidden' rock art: Vaalekop Shelter, KwaZulu-Natal*. Southern African Humanities 23 (2011), No. 1, 55 to 76.

- [7] M. NITSCHKE: *Performance art and digital media*. Digital Creativity 24 (2013), No. 2, 93–95.
- [8] X. J. BAN, Q. SHEN, M. WANG, X. LIU: *An optimization to parallelization of large-scale artificial life animation system based on the Olfactory-Visual Perception model*. Journal of Computational Information Systems 8 (2012), No. 16, 6825–6845.
- [9] X. J. BAN, G. P. ZENG, X. Y. TU: *Design and implementation of perception system based on self-learning for artificial fish*. Acta Electronica Sinica 32 (2004), No. 12, 2041–2045.
- [10] M. NESHAT, A. ADELI, G. SEPIDNAM, A. N. TOOSI: *A review of artificial fish swarm optimization methods and applications*. International Journal on Smart Sensing and Intelligent Systems 5, (2012), No. 1, 107–148.
- [11] X. J. BAN, L. PENG, X. H. WANG, G. P. ZENG: *Multi-agent-based algorithm research on a self-organization behavior of artificial fish colony*. Computer Science 34 (2007), No. 7, 193–196.
- [12] K. ZHANG, J. LUO: *Research on flatness errors evaluation based on artificial fish swarm algorithm and Powell method*. International Journal of Computing Science and Mathematics 4 (2013), No. 4, 402–411.
- [13] J. YIN, Q. Z. LI, H. E. LIANG: *Analysis of the interactive interface design for rich media advertising*. Packaging Engineering 33 (2012), No. 12, 21–24.
- [14] Q. CHEN, J. CUI, G. FANG, S. E. BRAUTH, Y. TANG: *Acoustic analysis of the advertisement calls of the music frog, Babina daunchina*. Journal of Herpetology 45 (2011), No. 4, 406–416.
- [15] E. EDMONDS: *The man—computer interface: A note on concepts and design*. International Journal of Man-Machine Studies 16 (1982), No. 3, 231–236.

Received June 6, 2017



# Computer intrusion factor decomposition detection based on biotechnology

TANG GUOQIANG<sup>1</sup>

**Abstract.** Computer monitoring system has been widely used in hydropower plants, which has provided a guarantee for the safe and economical operation of hydropower plants. In order to study the computer intrusion factor decomposition of biotechnology, the biological immune mechanism was used to establish an intrusion detection system based on artificial immune as a research hotspot. The intrusion detection system, biological immune system, artificial immune system, risk theory and other related research were adopted. The characteristics and limitations of the cellular algorithm were studied. The dendritic cell algorithm was improved and the dendritic cell algorithm was applied to the intrusion detection system. The final experimental results show that this method can better detect the computer intrusion factor and obtain less error rate, and the dendritic cell algorithm is suitable for the field of vibration anomaly detection.

**Key words.** Dendritic cell algorithm, intrusion detection, anomaly detection.

## 1. Introduction

As a proactive security technology, intrusion detection is considered to be the second safety valve after the security measures such as firewalls, it can monitor the network system in real time without affecting the network, thus providing real-time defense against external attacks, internal attacks and misoperation. Traditional intrusion detection techniques basically start with defining intrusion patterns or normal behavior patterns, and then compare detected data with these known patterns to detect intrusions. There is a large number of missing and false positives in the result, which lacks diversity, real-time and scalability. The biological immune system is naturally endowed with magical powers of self-preservation, which is not only able to resist a variety of foreign known or unknown pathogens threat, but also to repair the body's own damage and maintain the stability and balance of the body. This is what the current intrusion detection research institute expects to achieve, which is

---

<sup>1</sup>Department of Information Engineering, Guangzhou Institute of Technology & Engineering, 510925, China

conductive to overcome the shortcomings of the traditional intrusion detection system. There are many similarities between computer network security defense system and biological immune system in system function and architecture, which protect the system's security and maintain a stable balance of the system in complex environments. The specific immune defense, immune surveillance, immune recognition, immune tolerance, immune memory, immune regulation and immune homeostasis of the biological immune system, and stronger adaptability, diversity, distribution, dynamic and robust characteristics are the dream of the current intrusion detection system.

## 2. State of the art

The development of modern science cannot be separated from each other. In the cross field of information science and life science, artificial immune system developed by mimicking the immune system of the biological system has gradually become the focus of research, and has also provided new ideas and methods for intrusion detection [1]. As the advantage and weakness of the classifier, the dendritic cell algorithm indicated the future development direction of the dendritic cell algorithm, and discussed the applicability and limitations of the network intrusion detection system based on the dendritic cell algorithm. In view of the danger theory, some suggestions were given for the need for modification, and an improved intrusion detection method was proposed to apply the fuzzy set theory to the dendritic cell algorithm [2]. The blunt boundaries between mature and semi-mature dendritic cells in the original dendritic cell algorithm were changed by fuzzy decision. The experimental results show that the accuracy of the improved decision has been improved [3]. More and more researchers draw on and simulate the information processing ability of the immune system to design intelligent algorithms, establish artificial immune model, and apply to solve engineering and scientific problems, so as to make a lot of progress. But relative to the development of artificial neural networks, fuzzy systems and evolutionary computation, this is only the beginning [4]. Dendritic cell algorithm has been taken as the latest research in the theory of risk in artificial immunology, but the combination of many immune mechanisms and application problems of dendritic cell algorithms has not yet been fully understood. Whether it is a dendritic cell algorithm itself or its popularization, there are still many areas worth exploring and improving [5].

## 3. Methodology

As a full-time antigen-presenting cell in the innate immune system, dendritic cells (DC) can fuse a variety of environmental signals and correlate the signals with antigens to analyze the abnormalities of the antigen [6]. Inspired by the DC function, the dendritic cell algorithm (DCA) is designed by abstract modeling the antigen presentation behavior, which creates a new immune algorithm.

DC is currently known as the strongest antigen-presenting cells, as shown in

Fig. 1, when it matures, the cell membrane sticks out many long dendritic processes, so it gets its name. It belongs to an immune cell in the innate immune system, which plays a major role in perceiving the risk of the organism and controlling the immune response. DC collects antigen by endocytosis, or uses its dendrites to capture or retain antigens to process the collected antigens, presents antigens for cell recognition, and stimulates or inhibits cellular immune responses to antigens according to environmental signals [6].

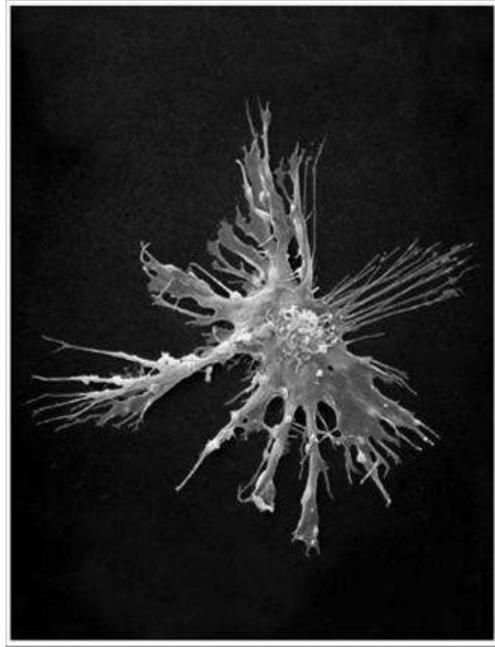


Fig. 1. Dendritic cell

DCA is the latest research result of dangerous theory in artificial immunology. Its basic principle is to simulate the state transition process of DC in biological immune system, abstract it into a data structure similar to the signal processor, and calculate the input signal and obtain the state of the output signal through the signal processing process simulated by the linear signal processing model [7].

As shown in Table 1, the names and functions of the biological signals and the corresponding abstract signals and meanings are summarized.

When the immature DC processes a set of input signals, it sums the three output signals, respectively. When  $\sum csm$  in DC reaches the migration threshold (MT), it compares  $\sum semi$  and  $\sum mat$ : if  $\sum semi > \sum mat$ , DC is converted to semi-mature state, the antigenic environment value = 0, which indicates that the cell environment is safe, and which means that the antigen is collected in the normal state; otherwise the DC is transformed into a mature state, and the presenting antigenic environment value = 1, the cellular environment is dangerous, which means that the antigen is collected in an abnormal state [8].

Table 1. Function of biological signals and their abstract signals and meanings

Biological signal	Features	Abstract signal	Meaning
PAMP	Indicating the presence of pathogens	PAMP	Indicating abnormal features
Necrosis signal	Indicating tissue damage	DS	Indicating high likelihood of abnormality
Apoptotic signal	Indicating tissue health	SS	Indicating a high likelihood of normal
Proinflammatory cytokine	Indicating that the tissue is damaged in general	IC	Zoom in other input signals
Costimulatory signal	Co stimulatory molecule	csm	Determines whether the iDC is converted
Mature output signal	Cytokines secreted by mDC	mat	Abnormal signal
Semi mature output signal	Cytokines secreted by smDC	semi	Normal signal

The basic principle map of DCA is shown in Fig. 2.

DCA can combine the treatment of a variety of environmental signals associated with the antigen, and analyze the anomalous index of antigen, which has the advantages of small scale calculation, fast response speed and strong recognition ability. In particular, there is no difference between known and unknown data for DCA, it recognizes invasion as long as it is aware of danger, even if it is the first encounter, there are no extensive training and centralized control [9]. Therefore, DCA is suitable for the distributed real-time intrusion detection between the internal network and the external network, the subnets and the nodes in the computer monitoring system of the hydropower plant. Of course, DCA will also produce false positives and omission. In order to meet the stringent requirements of network security for computer monitoring systems in hydropower plants, an intrusion detection system model combining with innate immunity and adaptive immunity is designed and the intrusion detection system which is connected to DCA is tested through the KDD Cup99 data set [10].

The system mainly includes antigen and signal acquisition module, DCA detection module, detector module, intrusion comprehensive evaluation module, administrator confirmation module [11]. Since this paper focuses on DCA-based intrusion detection, the adaptive immune part is simplified.

Antigen and signal acquisition module is mainly to simulate the collection of antigen and various signals in the system. On the one hand, it provides DCA with antigen flow and signal flow; on the other hand, it provides the antigenic flow directly to the detector module. Because intrusion often leads to host and network anomalies,

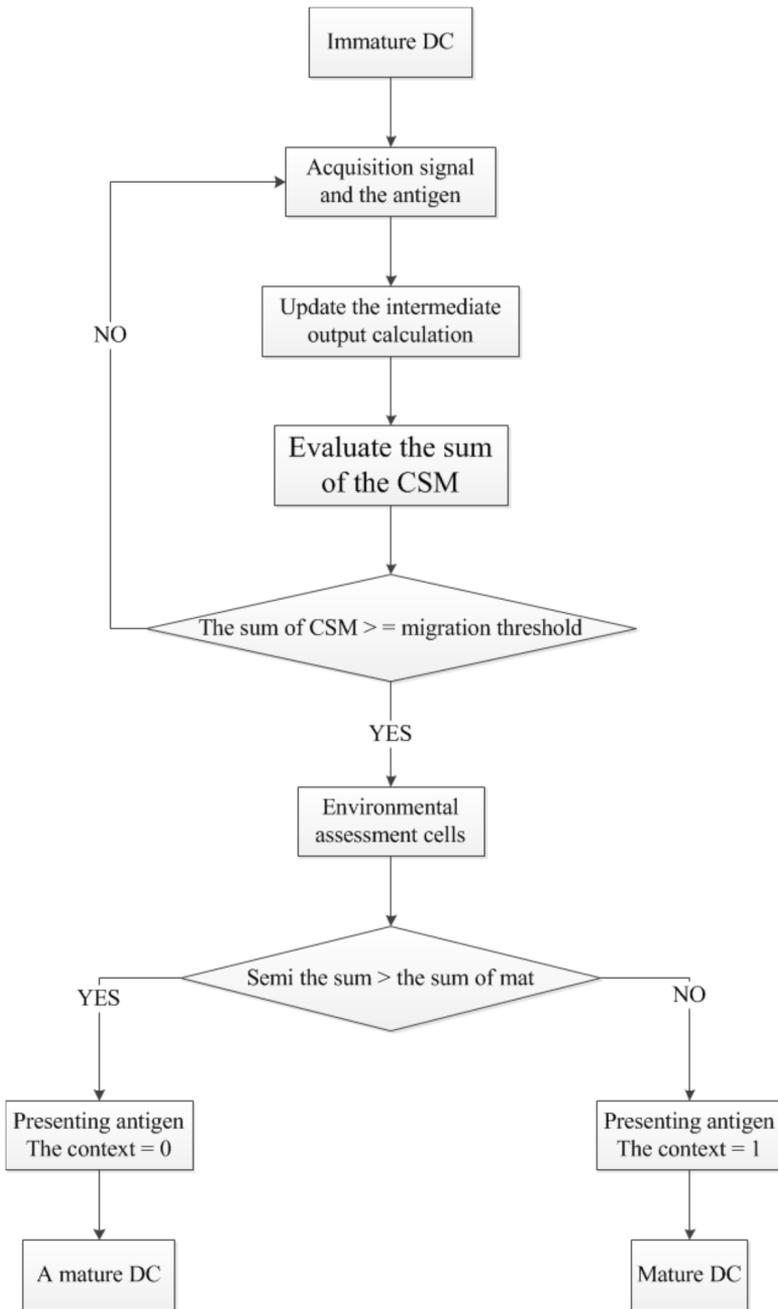


Fig. 2. DCA basic schematic diagram

this can be used as a basis for DCA signal extraction. PAMP indicates that a clear exception can be described by repeated login failures, connection errors, and so on in

the intrusion detection. Danger signal DS indicates that the possibility of anomaly is very large. In the intrusion detection, it can be expressed by the high utilization ratio of CPU and memory, the rapid change of network traffic, the exhaustion of bandwidth, the abnormal data receiving and sending and the high request of host connection and so on [12]. The security signal SS indicates that the organizational environment is in a normal state and can be represented by stable network traffic, low memory usage in intrusion detection.

Step 1: An antigen cell with the length of  $m$  is initialized, and the first  $m$  entries of the dataset are put into the pool.

Step 2: The DC is initialized, and the random migration threshold and life span in a certain range are set.

Step 3: DC randomly takes samples from the antigen signal pool within the lifetime, accumulates three output values: csm, semi, mat, DC which is over lifetime is re-initialized and transferred to Step2.

Step 4: Whether the cumulative csm reaches the migration threshold is determined, if it is not, which needs to be transferred to Step3 to continue sampling.

Step 5: The antigens used in DC are counted, the number of times and the number of exceptions for each antigen are recorded.

Step 6: Whether there is an antigen that has reached the number of times of determination  $N$  is determined, if so, the antigen of the MCAV is calculated, the abnormality of the antigen and output it in real time is evaluated, and then the evaluated antigen is removed from the antigen cell, otherwise, it is needed to turn Step2.

Step 7: Whether there is a new antigen is determined, if so, the new antigen will be added to the sampling pool to fill the vacancies removed, turn Step2.

Step 8: Whether any antigens in the pool have not been evaluated is determined, if so, then it is needed to turn Step2, otherwise the algorithm terminates.

The whole system model is based on the combination of congenital and adaptive immunity, and congenital immunity and adaptive immunity are not isolated, but are interrelated. Innate immunity is instructive for more accurate body non-auto recognition [13]. In this module, in addition to the realization of DCA detection function, the detector set is dynamically updated according to the detection results.

The detector module consists of two sets of detectors; one is the general set of detectors, which contain a large number of detectors, and can cover as large a detection space as possible. The other is a collection of memory detectors that contain relatively few detectors, but are aimed at typical invasion features. The number of memory detectors is relatively small, which is designed to reduce computational overhead and expect fast response. For typical intrusion features, it fully reflects its specificity and strengthens the detection accuracy. The antigenic flow provided by the antigen collection module first passes through the memory detector module and the antigen detected by the memory detector has a high confidence in the results [14]. The detected antigen detected by DCA will pass through the general detector module, and the results of the test and the results of the DCA test are sent to the comprehensive evaluation module for intrusion. The detector set in the module is dynamically updated. On the one hand, it comes from the assistance of DCA

test results; on the other hand, it is adaptive development through its own cloning, mutation, life cycle, elimination mechanism.

The general responsibility of the integrated evaluation module is to determine the final test results, the general principle is: firstly, the detection results of memory detectors are used as standard, unless the detector set is changed by the results of the test and the confirmation of the administrator, therefore, the detection results of memory detectors are generally not processed and can be output directly; secondly, the determination detection of DCA is taken as the primary. DCA has the strong real-time and high detection accuracy. The determination of the DCA test results provides assistance for the dynamic update of the detector with the participation of the administrator; thirdly, the antigen in the interval of DCA is detected by the general detector and then evaluated with the results of DCA test.

The administrator confirmation module is set up for the decision of the special case, for example, the detection result updates the control of the detector as well as the final verdict of the intrusion synthesis evaluation.

KDD99 data set is a benchmark database used in intrusion detection field. So far, it is the most used and most authoritative data set for intrusion detection researchers. The experiment used a subset of the KDD99 data set of 10% (kdd-cup.data\_10\_percent.gz), including a total of 494021 data, of which 97278 were normally connected and 396743 attacks. The 10% subset had similar statistical properties to the complete data set, maintained a similar normal connection and attack ratio. Each connection in the KDD99 data set had 42 data items, and the last item was a sign to mark the journey. Anomaly detection based on DCA was a two classification algorithm, which only detected whether it was intrusion or not, and could not tell which was the invasion. Therefore, all normal connected tags were set at 1 (Class1), which indicated normal; and all attack markers were set at 2 (Class2), which indicated an exception. Then, the remaining 41 items were examined, and some of the data items had little effect on the detection and could be removed to reduce the dimension of the data set.

The information gain of the property shows the statistical correlation of the attribute to the classification result. In the information gain, if the feature can bring more information for the classification system, then it is more important. The information gain of each attribute was calculated and analyzed, and the attribute of the low information gain was deleted from the data set. Eventually, 10 data items were selected to extract the three signals required for DCA:

PAMP:error\_rate?srv\_error\_rate;same\_srv\_rate;dst\_host\_error\_rate?dst\_serror\_rate?  
 Dangerous signal: count;srv\_count?  
 Safety signal: logget\_in;srv\_diff\_host\_rate;dst\_host\_count?

The specific information of these 10 data items is shown in Table 2. The value of each attribute was normalized to the [0,100] interval according to the linear function conversion method. The average value of each type of signal was the value of such signal.

Table 2. Data item information for KDD99 data

Features	Description	Types	Ranges
error_rate	Over the past 2 seconds, the percentage of connections with "SYN" errors occurred in the connection with the same target host as the current link.	continuous	[0.00,1.00]
srv_error_rate	Over the past 2 seconds, the percentage of connections with "SYN" errors in the connection with the same service as the current link	continuous	[0.00,1.00]
same_srv_rate	In the last 2 seconds, in the connection with the same target host as the current link, the percentage of connections with the same service as the current connection	continuous	[0.00,1.00]
dst_host_error_rate	In the first 100 connections, the percentage of connections with "SYN" errors in the connection with the same destination host in the current connection	continuous	[0.00,1.00]
dst_serror_rate	In the first 100 connections, the percentage of connections with "REJ" errors in the connection with the same destination host in the current connection	continuous	[0.00,1.00]
count	The number of connections for the target host that has the same current connection as the current connection over the last 2 seconds	continuous	[0,511]
srv_count	The number of connections that have the same service as the current connection over the last 2 seconds	continuous	[0,511]
logget_in	Successful landing is 1, otherwise 0	dispersed	0 or 1
srv_diff_host_rate	In the last 2 seconds, the percentage of connections with different target hosts that connect to the current connection in a connection that has the same service as the current connection	continuous	[0.00,1.00]
dst_host_count	The number of connections in the first 100 connections that have the same target host as the current connection	continuous	[0,255]

#### 4. Result analysis and discussion

The experimental data was randomly selected from the entire data set according to the attack type.

Experiment 1: first of all, DCA experienced a wave attack after a stable normal environment. 100 normal connections +100 smurf attacks were taken as the experimental data. The experimental results are shown in Table 3.

Table 3. Statistics of results of DCA running 10 times

Numbering	Number of false positives	The number of missed	False alarm rate (%)	False negative rate (%)	Accuracy (%)
1	2	0	1.00	0.00	99.00
2	1	0	0.50	0.00	99.50
3	4	0	2.00	0.00	98.00
4	0	0	0.00	0.00	100.00
5	4	0	2.00	0.00	98.00
6	3	0	1.50	0.00	98.50
7	2	0	1.00	0.00	99.00
8	5	0	2.50	0.00	97.50
9	2	0	1.00	0.00	99.00
10	1	0	0.50	0.00	99.50
average	2.4	0	1.20	0.00	98.80

Table 3 shows the results of the experiment run statistics. It can be seen that the average detection accuracy was 98.80 %, missing was 0, false positives were 0–5, and all of them occurred in the transition period of the two types of environmental change, which conformed to the characteristics of DCA. The experimental error threshold was set to 0.6, if it was set to 0.8 or 0.9, the experimental results were better. The detection accuracy was more than 99.50 %, which showed that DCA had good classification effect on normal connections and smurf attacks. If the abnormal threshold was set to 0.5, the detection accuracy would be reduced to 97.5 %.

It is necessary to note that since the exception threshold is set to 0.8 or 0.9 to achieve the best detection results, why the exception threshold is set to 0.5, the reason is that in addition to detecting smurf attacks, it is necessary to face other more attack types, Some of the attack threshold is low, for example, it can be detected and only it is set to 0.3, otherwise the false negative rate will be high. Experiments with multiple abnormal thresholds are prepared for the final DCA's comprehensive experiment against multiple attacks.

Experiment 2: the order of the experimental data used the normal connection data + attack data alternately placed form: 100 normal connection +100 smurf attack +100 normal connection +100 neptune attack + .... The test results are shown in Table 4.

Table 4 shows the comprehensive detection data of the different anomaly threshold. As can be seen from the experimental results, when the abnormal threshold was

set to 0.5, the detection accuracy was the highest, which reached 88.8%. It should be noted that the multiplication mechanism for the disordered data set is not used in the experiment because the data stream with a length of 100 can form a stable environment compared to an antigen signal pool with a length of 20. But in the actual environment, it will appear intermittent and brief attack. In order to detect such attacks, MMDCA with multipath merge mechanism can be used.

Table 4. Comprehensive detection data for different abnormal thresholds

Numbering	Abnormal threshold	Average false positive rate (%)	Average false negative rate (%)	Average inspection accuracy (%)
1	0.8	0.3	15.2	84.5
2	0.7	0.6	13.0	86.4
3	0.6	1.2	10.9	87.9
4	0.5	1.9	9.3	88.8
5	0.4	3.3	8.1	88.6
6	0.3	4.8	6.7	88.5

## 5. Conclusion

The computer monitoring system of hydropower plant is responsible for the supervision and control of the main auxiliary equipment of hydropower plants. In order to ensure its safety, the traditional DCA off-line analysis process was improved, and an online analysis mechanism parallel to the detection process was designed, an antigen that was evaluated for a sufficient number of times was exported, thus achieving the goal of real-time or near real-time analysis. The conclusions were drawn as follows: the distributed real-time self-protection mechanism of biological immune system provides a new idea for the study of intrusion detection. DCA is a congenital immune algorithm based on the dangerous theory, and it is not based on pattern matching of antigenic features, but rather associates antigen with signals and assesses the degree of abnormalities of the antigen according to the degree of environmental risk, which has the characteristics of small operation scale, fast response speed and strong recognition ability. For DCA, there is no difference between known and unknown data, so it does not need a lot of training and centralized control, which is suitable for the distributed real-time intrusion detection in the computer monitoring system of hydropower plants. However, the study still has some limitations, for example, intrusion detection is only the first step in intrusion prevention, which should combine intrusion detection with firewall, vulnerability scanning, antivirus software and other security products to establish a more complete intrusion prevention system.

## References

- [1] J. C. NI, Z. S. LI, J. R. SUN, L. P. ZHOU: *Research on differentiation model and application of dendritic cells in artificial immune system*. Acta Electronica Sinica 36 (2008), No. 11, 2210–2215.
- [2] Z. CHELLY, Z. ELOUEDI: *Hybridization schemes of the fuzzy dendritic cell immune binary classifier based on different fuzzy clustering techniques*. New Generation Computing 33 (2015), No. 1, 1–31.
- [3] T. SCHULZE, S. GOLFIER, C. TABELING, K. RÄBEL, M. H. GRÄLER, M. WITZENRATH, M. LIPP: *Sphingosine-1-phosphate receptor 4 (S1P<sub>4</sub>)*. FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology 25 (2011), No. 11, 4024–4036.
- [4] I. HANSENNE, C. RENARD-CHARLET, R. GREIMERS, V. GEENEN: *Dendritic cell differentiation and immune tolerance to insulin-related peptides in Igf2-deficient mice*. Journal of Immunology 176 (2006), No. 8, 4651–4657.
- [5] S. LIU, J. KE: *Method of locating anomaly source in software system based on dendritic cell algorithm*. Applied Mechanics and Materials 556–562 (2014), 6255–6258.
- [6] A. A. AL-HASAN, E. S. M. EL-ALFY: *Flexural vibrations of non-homogeneous elliptic plates*. Procedia Computer Science 52 (2015) 244–251.
- [7] J. Y. ZHAO, H. GUO, L. F. WANG: *Chinese spam filtering algorithms based on case base reasoning*. Microelectronics & Computer 26 (2009), No. 12, 64–66.
- [8] Z. H. OUYANG, X. Y. ZHANG, S. TU, J. J. GU: *Research on problem of application of "Cloud security" in computer anti-virus*. Computer and Modernization (2010), No. 12, 72–75.
- [9] S. K. CHOI, J. H. SONG, S. H. KIM, Y. D. LEE: *Characteristics of the load of small hard body used for impact resistance test of the lightweight wall*. Journal of the Korea Institute of Building Construction 14 (2014), No. 4, 350–358.
- [10] D. GUO, X. L. HONG, X. HU, X. WU: *A bit-parallel algorithm for sequential pattern matching with wildcards*. Cybernetics and Systems 42, (2011), No. 6, 382–401.
- [11] M. ASAKA, T. ONABUTA, T. INOUE, S. OKAZAWA, S. GOTO: *A new intrusion detection method based on discriminant analysis*. IEICE Transactions on Information and Systems E84-D (2001), No. 5, 570–577.
- [12] E. F. FERNÁNDEZ, F. ALMONACID, N. SARMAH, P. RODRIGO, T. K. MALICK, P. PÉREZ-HIGUERAS: *A model based on artificial neuronal network for the prediction of the maximum power of a low concentration photovoltaic module for building integration*. Solar Energy 100 (2014), 148–158.
- [13] A. SANO: *Generating novel memories by integration of chaotic neural network modules*. Artificial Life and Robotics 4 (2000), No. 1, 42–45.
- [14] J. SU, P. L. QIAO, Y. H. LIU: *Distributed intrusion detection method based on immune evolution computing*. IJ Solids and Structures 36 (2010), No. 6, 163–165.

Received July 12, 2017



# Optimization design of incineration processing of radioactive solid waste

JİYUN XU<sup>1,2</sup>

**Abstract.** With the continuous development of nuclear power projects, solid radioactive waste are increasing year by year. The inventory of solid waste is overloaded and can't be handled in time, thus bringing great pressure on radioactive waste management. In order to solve the above problems, the design and research of radioactive solid waste incineration were carried out in this paper. Field investigation of solid waste management in a nuclear power station in China was carried out. Then the optimized scheme of incineration intensification was put forward. And combined with the feasibility evaluation results, the reliability of the treatment scheme was verified. There is theoretical significance for the incineration of radioactive solid waste.

**Key words.** Radioactive solid waste, incineration treatment, optimization design.

## 1. Introduction

Domestic research investment on nuclear power has increased year by year, and the number of nuclear power plants has been increasing steadily, as a result, the amount of radioactive material has increased. The annual output of nuclear waste of a basic nuclear power unit will be in the range of 500 square meters, and the output of solid waste will be higher [1]. If the generating capacity of nuclear power generation units in China reaches 10000 kilowatts, the accumulation of radioactive solid materials will be considerable. Therefore, under such circumstances, it is very important to explore the comprehensive treatment plan, introduce advanced management program and reduce the treatment cost for the storage and optimal management of solid waste for nuclear power plants [2]. China's relevant laws show that the processing of radioactive waste must be synchronized with the construction process, and it is subject to rigorous review and approval prior to commissioning. Some nuclear power stations have set up many relatively independent waste disposal systems to simplify approval procedures. Such an approach is likely to lead to the repeated use of equipment and human resources management [3]. In addition, the

---

<sup>1</sup>College of Engineering, Peking University, Beijing, 100871, China

<sup>2</sup>China Everbright Group Post-Doctoral Scientific Research Workstation, Beijing, 100033, China

investment and use of nuclear power plants are different, therefore, the independent management of radioactive solid waste in nuclear power plants lacks unified operation and management, when the resources can't be shared, the operation of the nuclear power plant will suffer serious security threats [4]. In this paper, an optimized scheme for incineration of radioactive solid material was put forward, so as to improve the comprehensive utilization ratio of equipment and personnel and the management system of waste treatment, and minimize waste disposal. The research conclusions can provide management experience for the treatment of new radioactive solid wastes.

## 2. State of the art

Under the current situation, China advocates an intensive economic growth situation, and controls the extensive economic development situation as the intensive transformation of the development situation, and the main focus of the work is to control the growth mode of the national economy as a benign trend of development. The development of the enterprise is the key point of the control, and the regular theory control of the operation can achieve certain fruitful results [5]. The implementation of intensive economic development mode is mainly to increase production efficiency for enterprises, integrate a certain amount of manpower and material resources, and then carry out the unified deployment, so as to realize the concrete and practical management of specific affairs in the actual enterprise management, reduce the cost, and improve efficiency and achieve the sustainable development [6]. The main research direction of this paper is the process analysis of the pollution source of the radioactive waste from the resource enterprises, in particular, the optimization and design of the management is carried out for the incineration of solid radioactive waste, at the same time, the research combines the development of the enterprise and the reengineering theory of the new method of sustainable development to get the management innovation in technology, so as to achieve intensive management of radioactive waste and enhance the efficiency of waste incineration treatment. For the process intensive management of enterprise development, the main business process is to meet the core competitiveness of enterprises [7]. The key process is to eliminate the process of delivering value chains of environmental pollution, rather than change the core value chain [8]. Managers need to carry on the transformation from the angle of process and evaluation of incineration of radioactive solid waste, so as to achieve the intensive management of business results. In this study, a standardized model of intensive management of radioactive solid waste incineration was put forward in light of actual conditions, the theory of reengineering for enterprise management was innovated, and the intensive management of the scheme was evaluated with the combination of economic budget growth theory, so as to open up an optimized treatment approach with sustainable economic operation, safety management and scientific development.

### 3. Methodology

#### 3.1. Treatment status analysis

The arranging mechanism of radioactive waste selected in this study is a nuclear power station in the Yangtze River delta, the nuclear power station in the region has many staging projects, and the average capacity of the power station unit is 700 thousand kilowatts. Since its establishment, nuclear power plant has established independent departments for the treatment of radioactive pollutants. The classification and management of radioactive waste are carried out with different technical standards. The management of solid emission waste is achieved in a division of responsibility system, senior engineer or deputy general manager of the management carries on the global control, and supervises the responsibilities of each department as head of the department, specific site management departments are responsible for daily solid material handling, equipment maintenance, etc. The overall storage method of radioactive solid waste is realized through the preliminary handling and transportation and the post shipment package, the main core problem is to convert the previously used waste into solid form, and some process waste also should be compressed and packaged. In addition to the compression and packing of solid wastes, the filter waste conditioning concrete way is also a commonly used process mode for a lot of radioactive solid waste factories. Table 1 shows the changes of the radioactive solid material before and after the overall treatment.

Table 1. Changes of radioactive solid waste before and after treatment

Scrap type	Technology	Before processing /m <sup>3</sup>	After treatment /m <sup>3</sup>	Volume reduction ratio
Technology	Pre compression	2397	799	3
Concentrate	Cement solidification	95	213	0.45
Silt	Cement solidification	0.9	3.1	0.33
Resin	Cement solidification	61.4	252	0.25
Filter	Cement solidification	14.9	97.7	0.15

In the study, three phases of management model are used for radioactive nuclear power companies. However, although the regions are slightly different, the waste treatment processes are basically the same, and the waste treatment problems faced are basically similar, the only difference is that the operators belonging to the enterprises are slightly different, the management's technical reserves are slightly different, but the technical and human resources can't be integrated completely. Be-

cause most of the solid wastes in reactors which have been used are not stored as a whole, the temporary storage has many security risks. If the overall optimization of managers can be carried out according to hierarchical classification, there is still a possibility of benefit improvement.

Figure 1 shows the annual waste production of a unit in a radioactive solid waste generating plant. As can be seen from Fig. 1, the annual solid waste reserves of the plant are roughly  $70 \text{ m}^3$ , although reserves of solid waste are decreasing year by year, which can remain at least  $60 \text{ m}^3$ , such solid waste reserve capacity should be taken seriously, as time goes on, these solid wastes will be the major source of radioactive solid waste disposal in the future. Radioactive pollution sources will continue to decrease as the equipment improves and the level of operation increases. However, there are still some gaps in the technology level between our country and foreign countries, and the utilization rate of foreign waste is still much higher than the domestic utilization rate. Our country can draw lessons from the management hierarchy of foreign countries and reduce the overall allocation of human resources.

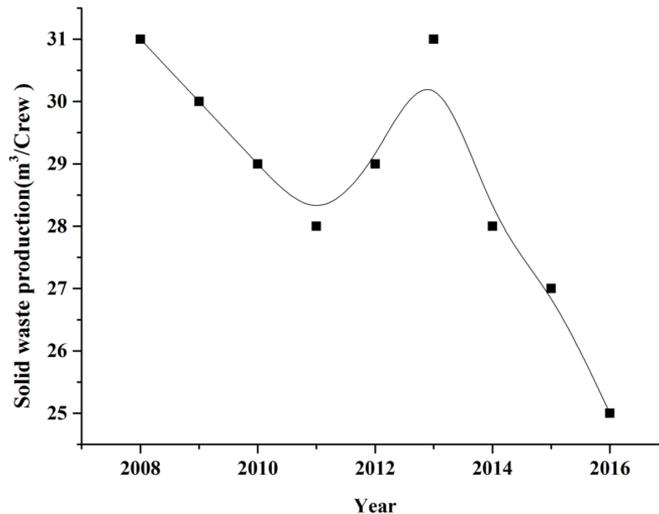


Fig. 1. Annual emissions of solid waste

### 3.2. Optimization design scheme

The solid waste in the plant was packed and disposed, the solid waste of the unit was setup, and the pre-compression and the post pyrolysis incineration were carried out, then, the conditioning treatment was made again after the cement was solidified, so as to achieve the post processing stability of solid wastes [9]. Because the total production of radioactive solid waste is gradually decreasing, the location of solid waste in that area can't always be in the same place, and it remains to be handled throughout the area. Although incineration is currently a more suitable method of handling solid radioactive pollutants, the metal scrap has a certain use value and

can also be smelted after incineration [10]. In this study, the preliminary plan for the incineration of radioactive solid wastes and the treatment scheme for the later stage are shown in Fig. 2. The compressed solid waste was processed by compression, and the metals were cleaved, broken down, and finally sorted and packed. The reduction volume processing was made in the factory, the compression barrel management for the burning ashes produced after the waste material incineration was carried out. The smelting decontamination was done in the region away from the plant, and the fixed transportation processing for unqualified waste was carried out, while the qualified materials were re-used when the conditions and the technology were mature.

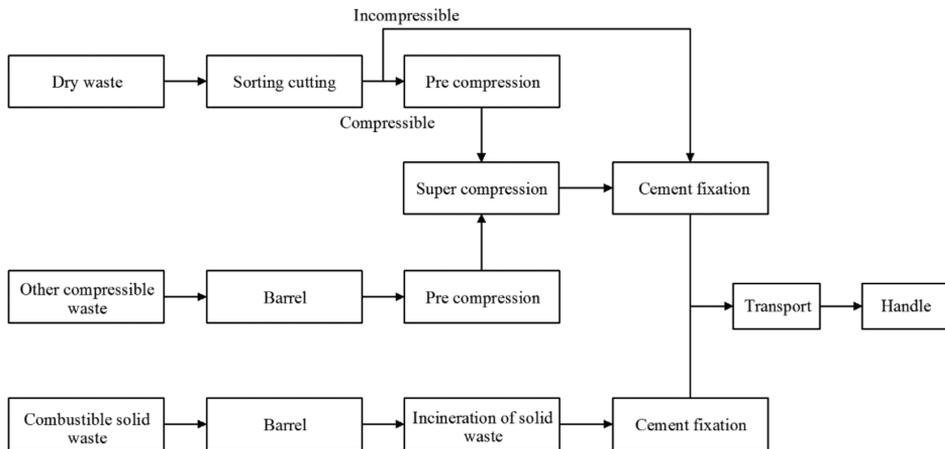


Fig. 2. Flow chart of solid radioactive waste treatment

After optimizing the disposal plan of radioactive solid waste, the mechanical and electrical groups in the factory were sorted, and the existing solid waste was fixed with concrete, and treated in a centrally fixed area [11]. The pyrolytic incineration of compression and waste materials was used to carry on the after-treatment and solidifying for the waste that had been treated once. The existing overall reserves were reduced, and the concentration of solid waste was setup furthest. The high pressure digestion for waste ashes after the incinerator was carried out, so as to reduce operators and improve technology and personnel reserves, thus, the waste was closer to the minimum requirement. Table 2 shows the technical standards and results of the overall disposal of solid waste in the plant area.

Table 2. Technical standards and results of solid waste treatment

Processing system	Pack	Pre-compression	Cement solidification	Concrete curing	Pyrolysis incineration
Before processing	4	4	4	4	0
After treatment	3	3	1	1	1

### 3.3. Evaluation method

The technology of high reducing capacity ratio was added in the intensive treatment scheme of radioactive solid waste incineration, including solid waste incineration technology, and the compression remelting technology for waste incineration [12]. The compression incineration technology was optimized and upgraded, and the level of radioactive solid waste treatment technology was enhanced as a whole. The promotion of technical level drove the promotion of production efficiency, and also drove the saving and control of production cost. By optimizing the selection and equipment, the entire incineration system can be optimized as an operation level with unified technical indicators [13]. Through the intensive management of the system, the irrational equipment is eliminated and the solid waste incineration and later cement solidification treatment can be carried out, which is the core point of the process plan, so that the enterprise's operation funds can be invigorated, and the running time and the daily maintenance cost can be reduced.

After a super burning compression device is installed in the treatment plan, and the pyrolysis and incineration of radioactive waste materials can be carried out after the equipment is debugged. The following is the evaluation for the pyrolysis of radioactive solid wastes with synergistic mechanisms and the return on the investment using incineration equipment. The capacity of pyrolysis incineration treatment of the installed treatment equipment is 30 kg per hour, and an annual handling capacity is about 145 t, which can meet the requirement of radioactive solid waste treatment in the plant area. The construction costs are about 41 million, and the purchase of equipment was about 22 million.

The return on investment for projects and equipment requires a certain amount of time to accumulate, and the main economic characteristics need a certain evaluation method for reference. Without considering the allocation of funds, the recycling standard configuration can be carried out through the time management of income and the construction of cycle return characteristics, the calculation method is:

$$\sum_{T=1}^{T_j} (C_0 - C_i)_T = 0. \quad (1)$$

Here, the term  $(C_0 - C_i)_T$  indicates that the cash outflow during the year  $T$  is the difference between the cash inflow and the cash inflow  $C_i$ . Symbol  $T_j$  represents the static payback period of construction investment.

The future dynamic payback period requires the budget of the recovery value of the time spent, the investment scheme is discounted at the base income of each year's capital flow, so as to predict the recovery investment and the dynamic payback period, and calculate the payback period of investment returns. The capital for incineration and compression processing equipment is obtained from bank loans, the expression of the net present value of the dynamic recovery fund is:

$$NPV = \sum_{T=1}^{T_d} (C_i - C_0)_T (1 + i)^{-T}. \quad (2)$$

Here,  $T_d$  represents the value of the cumulative payback period of the investment process. Quantity  $NPV$  describes the net cash value of the dynamic return on investment, and the interest rate for bank loans is expressed as  $i$ .

The process of technological innovation needs to consider the feasibility of the economy. Emphasizing the independent development of technology is not consistent with the law of sustainable development, emphasizing the catalytic role of innovation and using the treatment technology of the pre-compressed solid radioactive waste and the late sorting compression under the circumstances of economic permission can reduce the cost effectively [14]. In the process of the introduction of equipment, combustible solid radioactive waste should be strongly supported by remote sensing technology because of the influence of remote control. After entering the new century, the maintenance technology and remote sensing technology have made great progress, and the innovation technology has stridden forward while saving the economy, which has effectively improved the management level and economic operation state.

#### 4. Result analysis and discussion

The combination of technological innovation and economic growth theory holds that the combination of knowledge accumulation and human resources will be the driving force for the cumulative economic growth of solid waste incineration in the future, the progress in technology and knowledge is also the driving force behind economic growth in the future, which can be put into the economic growth model budget for technological growth conditions. Nuclear power is a major trend in the development of clean energy in the future, which can promote national economic growth. The management of remaining radioactive solid waste during the use of nuclear power will be the stress and challenge in the future. In order to effectively manage the radioactive waste from nuclear power and achieve the optimization and allocation of resources, the extensive management mode and the intensive management method are used to optimize the management mode, and the innovation is carried out in technology, so as to further promote the optimization of radioactive solid incineration treatment program. Therefore, this study considers that the intensive treatment scheme of radioactive solid waste incineration not only can meet the needs of specialized management concepts and the optimization requirements of radioactive waste management mode, but also can save resources and improve the utilization ratio of equipment. It is considered that the scheme is feasible in theory.

In the disposal plan of radioactive solid waste incineration, the adjustment and optimization of management organization structure are defined. The human resource allocation scheme is also an important link of optimizing the solution, so that the lean operation goal can be realized. The basic composition of management personnel setting method is shown in Table 3. Through the remuneration of human resources, it can be seen that the salary level of managers is relatively high, the technicians are in the middle reaches, and the operators' income is in the middle and lower levels.

The treatment of radioactive solid waste can be divided into low and medium positions, and the methods for calculating the costs are different [15]. Outward

delivery charges include the cost of the recovery of waste after combustion and the delivery within the plant; transportation expenses refer to the expenses of unloading, radiation monitoring and so on. The service life of disposal sites needs to be greater than 280 years to degrade radioactive solid incineration materials. Waste generated by the solid radioactive waste which is treated by intensive incineration is treated by compression, the volume is greatly reduced, and the cost of transportation is reduced. Table 4 shows the waste retention situation after the use of the scheme proposed in this study.

Table 3. Basic composition of management personnel

Personnel classification	All hands	Management	Artisan	Operator
Original number	265	28	22	214
New arrivals	123	23	20	80
Difference in number	141	5	2	134

Table 4. Retained waste after disposal

Waste type	Technology	Before processing /m <sup>3</sup>	After treatment /m <sup>3</sup>	Volume reduction ratio
Incinerating	Pre compression	96	32	3
liquid	Cement solidification	5.6	13.7	0.41
compress	Pre compression	33	13	2.5
Other	Concrete fixation	3.1	6	0.5

The intensive reform plan of the incineration of radioactive solid waste materials is based on the radioactive waste management template in the factory area for intensive engineering transformation, and the employees should be greatly reduced. The salary of employees will be increased, and 20 million of the salary can be saved every year on the basis of guaranteeing the salary level of the basic staff. In the study, the degradable pyrolysis incineration method is used for the waste which is compressed into solid, so the effect of volume reduction will be more obvious in the later stage, the volume of solid radioactive waste and the funds for disposal waste site can be reduced, the disposal area reduced is 200 m<sup>3</sup>, and the disposal costs reduced is about 8 million.

The combination of flammable solid added and compression equipment for the incineration in the scheme can greatly reduce the freight. But the total consideration

is that incineration covers a larger area, requiring the overall shutdown of equipment in the plant and the running production line, and thus spare plant resources will also be wasted. Therefore, the centralized processing plant is located outside the factory area, and the cost will not exceed more than 1 million 400 thousand per annum, the static recovery period is set to 8 years, and the dynamic recovery period of the development is 12 years. Therefore, in economic analysis, the optimized design scheme for incineration of radioactive solid waste studied in this paper is feasible as a whole.

## 5. Conclusion

Nuclear materials bring clean new energy to human beings, however, the radioactive waste after using brings great difficulties to later processing, which requires a great deal of manpower and material resources to manage and optimize treatment. In this study, the optimal management model of radioactive waste materials, the solutions and staffing and other problems at the present stage in China were analyzed and summarized, and a centralized and intensive optimization scheme was put forward. Taking a nuclear power station in an area of our country as an example, the remanufacturing of the enterprise was studied from the aspects of the theory of treatment, the method of disposal, the technology used, the economic return of the treatment and the corresponding expectation. It can be seen from the treatment plan that starting from the theory of economic growth, the incineration of radioactive solid wastes requires the pre-curing treatment and a unified management of the compressed equipment after incineration, the intensive disposal of radioactive solid waste requires rational human resource allocation and advanced technology, at the same time, the whole process and equipment need to maintain a reasonable design concept, reduce human resource costs, reduce return cycle and reduce processing budget. The optimized disposal plan of radioactive solid waste incineration proposed in this study is feasible in theory, and the economic budget evaluation meets the theoretical requirements. However, many problems may be encountered in the process of implementation, and so the program should be improved and perfected in practice.

## References

- [1] K. OSHITA, H. AOKI, S. FUKUTANI, K. SHIOTA, T. FUJIMORI, M. TAKAOKA: *Behavior of cesium in municipal solid waste incineration*. Journal of Environmental Radioactivity 143 (2015), 1–6.
- [2] H. A. VAN DER SLOOT, D. S. KOSSON, O. HJELMAR: *Characteristics, treatment and utilization of residues from municipal waste incineration*. Waste Management 21 (2001), No. 8, 753–765.
- [3] O. V. MILLA, H. H. WANG, W. J. HUANG: *Feasibility study using municipal solid waste incineration bottom ash and biochar from binary mixtures of organic waste as agronomic materials*. Journal of Hazardous, Toxic, and Radioactive Waste 17 (2013), No. 3, 187–195.

- [4] A. LIU, F. REN, W. Y. LIN, J. Y. WANG: *A review of municipal solid waste environmental standards with a focus on incinerator residues*. International Journal of Sustainable Built Environment 4 (2015), No. 2, 165–188.
- [5] S. V. STEFANOVSKY, O. I. STEFANOVSKAYA: *EPR of radiation-induced paramagnetic centers in irradiated sodium silicate glass for immobilization of solid radioactive wastes synthesized in various conditions*. Inorganic Materials: Applied Research 5 (2014), No. 3, 271–277.
- [6] S. E. A. SHARAF EL-DEEN, G. E. SHARAF EL-DEEN: *Adsorption of Cr(VI) from aqueous solution by activated carbon prepared from agricultural solid waste*. Separation Science and Technology 50 (2015), No. 10, 1469–1479.
- [7] WWW.UKRINFORM.NET: *South Ukraine NPP to improve radioactive waste treatment system*, 2013.
- [8] M. KOTZASSARLIDOU, I. MONE, K. GIANNOPULOU, O. KIRILIDOU, A. PIPINTAKOU, M. CHATZIMARKOU, A. SIDIRADI: *Measurement of iodine (I125) radioactive solid waste derived from radio immune analysis (RIA-IRMA) performed annually in "THEAGENIO" nuclear medicine department*. Physica Medica: European Journal of Medical Physics 30 (2014), No. 1, e89–90.
- [9] H. ECKE, H. SAKANAKURA, T. MATSUTO, N. TANAKA, A. LAGERKVIST: *State-of-the-art treatment processes for municipal solid waste incineration residues in Japan*. Waste Management & Research Mathematics 18 (2000), No. 1, 41–51.
- [10] J. LIU, F. X. WEI, Z. WANG: *Environmental risk of nuclear power and policy proposal on disposal of solid radioactive waste*. Advanced Materials Research 726–731, (2013), 2894–2897.
- [11] V. V. KRUPSKAYA, S. V. ZAKUSIN, E. A. TYUPINA, M. S. CHERNOV: *Features of cesium adsorption on bentonite barriers in solid radioactive waste disposal*. Ore & Metals, Mining Journal (2016), No. 2, 79–85.
- [12] A. N. BOBRAKOV, A. A. KUDRINSKII, A. V. PERESLAVTSEV, V. L. SHIRAYEVSKII, A. V. ARTEMOV: *Development of plasma techniques for solid radioactive waste processing*. Russian Journal of General Chemistry 84 (2014), No. 5, 1031–1040.
- [13] S. MOBBS, G. SHAW, S. NORRIS, L. MARANG, T. SUMERLING, A. ALBRECHT, S. XU, M. THORNE, L. LIMER, K. SMITH, G. SMITH: *Intercomparison of models of  $^{14}\text{C}$  in the biosphere for solid radioactive waste disposal*. The University of Arizona Libraries, Radiocarbon 55 (2013), Nos. 2–3, 814–825.
- [14] S. YAPICI, H. EROGLU: *Batch biosorption of radioactive thallium on solid waste of oleum rosea process*. Journal of Chemical Technology and Biotechnology 88 (2013), No. 11, 2082–2090.
- [15] T. ITO, S. Y. KIM, Y. XU, K. HITOMI, K. ISHII, R. NAGAISHI, T. KIMURA: *Study on separation of platinum group metals from high level radioactive waste using solid state adsorbent*. CYRIC Annual Report 2012–2013 (2013), 101–106.

Received July 12, 2017

# Application of automatic control technology of drilling machine based on PLC

NING GUANGQING<sup>1</sup>

**Abstract.** At present, the rocker shaft produced by combined drilling machine is still in the condition of small quantity supply, which has the problems of low production efficiency and high processing cost. In order to solve these problems, in this paper, a variety of composite components were selected to formulate I/O point table and design the PLC control system that can meet the requirements through the analysis of the control requirements of modular machine tools, and the PLC program was written. After the control combination completed a process, the rotary table scale division was controlled, and then the next process was completed, until all processing was completed. The final experimental results proved the feasibility of PLC advanced control technology in modular machine tools. In addition, this technology can realize automatic production and solve the problems of low production efficiency and high processing cost.

**Key words.** Combination drilling machine, PLC control technology, multi station.

## 1. Introduction

With the advent of industry 4.0, the level of automation and intelligence is getting higher and higher. As the pillar industry of the national economy, the automobile industry is the benchmark of advanced technology. However, China is only a big country in automobile manufacturing, not a powerful country, and lacks in its own intellectual property rights and innovative capabilities. Therefore, the development of advanced manufacturing technology is a requirement for the new normal state of industrial transformation and upgrading, and it is a dream of a new generation of auto makers. The auto parts industry is an important part of the automobile industry chain, which provides the standardized, universal and serialized components for the automobile factory. At present, China's manufacturing industry is facing a situation that: high-end manufacturing industry of Europe and the United States in China appears reflux, and the low-end manufacturing industry moves to Southeast Asia, at the same time, the aging of the population makes the enterprises in China

---

<sup>1</sup>Zhengzhou Railway Vocational and Technical College, Zhengzhou, Henan, 450052, China

with backward traditional machinery and equipment pay a high cost to recruit labor. Facing the trend of the reduction of population bonus, China is at the cusp of technological transformation and upgrading. It is the theme of the times to develop new high technology and cultivate the creative talents. And along with the world economic downturn, the revival of manufacturing will lead to economic growth. Now, many private machinery processing enterprises have personally felt such recruitment difficult pain. They are trying to reduce the reliance on labor through equipment modification or the introduction of highly automated equipment to meet mechanized production.

## 2. State of the art

As the preferred product of discrete control, PLC developed rapidly from 1980s to 1990s. The worldwide annual growth rate of PLC controllers was maintained at 20~30% [1]. With the continuous improvement of factory automation and the expansion of PLC controller market capacity base, the growth rate of PLC in industrial developed countries has slowed down in recent years. However, the PLC growth in China and other developing countries is very rapid [2]. According to the comprehensive information, the world's PLC sales revenue in 2004 was about \$10 billion, which occupied a very important position in the field of automation. The PLC controller is developed by imitating the principle of the original relay control. In 1970s, the PLC controller had only on-off logic control, and the first application was the automobile manufacturing industry [3]. It uses storage, execution, logical operations, sequential control, timing, counting, and arithmetic operations, and controls all kinds of machinery or production process through the digital input and output operation. The user's control program expresses the process requirements of the production process, which is stored in the user program memory of the PLC controller [4]. The operation is executed one by one according to the contents of the stored program to complete the operation of the process. In the past 10 years, with the continuous decrease of the price of PLC controller and the expansion of user's demand, more and more small and medium-sized equipment has begun to adopt PLC controller to control, and the application of PLC controller in our country has been increasing very fast [5]. With the rapid development of China's economy and the continuous improvement of the basic automation level, the PLC controller will maintain high growth momentum for a period of time.

## 3. Methodology

The main movement of combination drilling machine is the rotational movement of the cutter, and the feed motion is the axial movement of the cutter and the indexing movement of NC rotary table [6]. The schematic diagram of drilling a hole is shown in Fig.1. The speed of the power head is controlled by adjusting the hydraulic damper of the pneumatic power head.

PLC technology is used to control the rotary table and pneumatic power head,

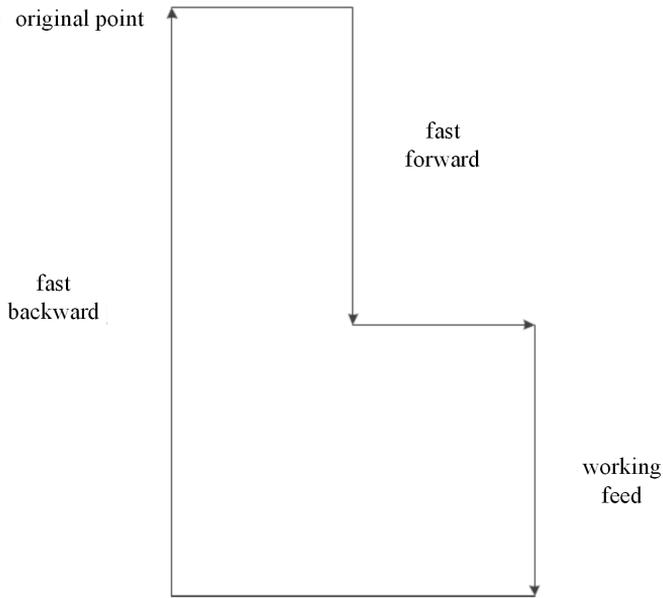


Fig. 1. Sketch map of drilling motion

so as to complete the workpiece processing. After each process, the workpiece completes automatically indexing rotation to the next working position [7]. Combination drilling machine needs 10 divisions to finish a rocker shaft processing. Control requirements in machining are as follows. The main movement is driven by a three-phase AC asynchronous motor, which turns forward and does not change speed. The stroke of the feed motion is controlled by the travel switch. The transposition of the workpiece is controlled by the indexing of the rotary table. The clamping and releasing of the workpiece is realized by pneumatic device. All actuating elements are controlled by the PLC [8]. The machining position of the combination machine tool is shown in Fig. 2.

The working process of the equipment: After the system is started, the three-phase asynchronous motor action of the pneumatic power head realizes the rotation of the main movement of the cutter. Meanwhile, the cylinder action of the power head is fast forward. When the hydraulic damper touches the iron block to move forward, it achieves rewind after processing. After the tool is pulled out of the upper limit of the workpiece, the slewing table is shifted with 36 degrees, and the second order is processed. In turn, reciprocal cycles are conducted [9].

Motion control system is mainly composed of motion controller, electrical servo mechanism, mechanical device and detection device. The block diagram of the system is shown in Fig. 3.

The operating panel is the part used by field operators to provide a complete interface between the control system and the operator. The electric servo mechanism is the control motor. AC servo motors are often used in the servo system of mechatronic equipment. Because AC servo motor adopts detecting device to realize

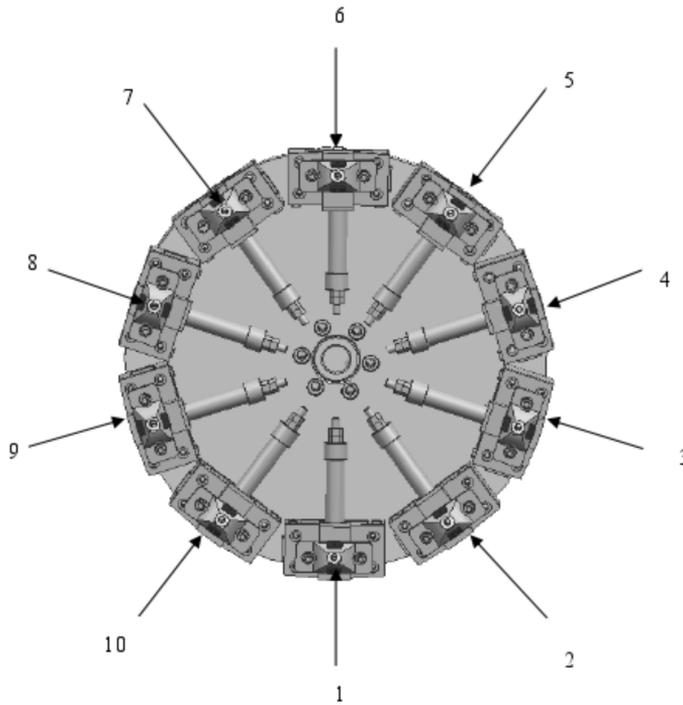


Fig. 2. 1–No.1 position; 2–No.2 position; 3–No.3 position; 4–No.4 position; 5–No.5 position; 6–No.6 position; 7–No.7 position; 8–No.8 position; 9–No.9 position; 10–No.10 position

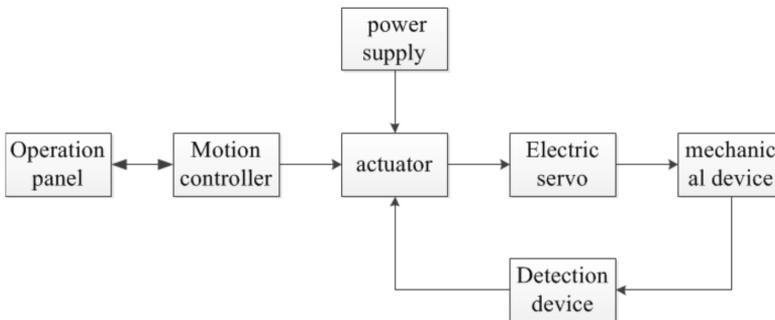


Fig. 3. System diagram

closed loop control, the precision is very high. In order to ensure that the control mechanism has the accurate location of the origin, the servo motor has the advantage in the precision range [10]. For normal operation, servo drive must be the functions as follows. First, it receives control signals, such as position and speed of the controller, and processes them. The servo motor is driven by the processed signal and operates at a certain speed. The second is to receive signals or other displacement signals from the encoder of the servo motor to determine the rotation speed and

displacement of the servo motor, so as to form a speed closed loop or a position closed loop control [11]. The combination parameters of the BCH servo motor and the Lexim2323Plus driver of the rotary table are shown in Table 1.

Table 1. Combination parameters

BCH Servo motor output power (KW)	Rated torque (NM)	Peak to stop torque (NM)	Maximum mechanical speed (rpm)	Rated speed (rpm)	Type of servo driver	Type of servo motor
4.5	28.26	71.62	3000	2000	LXM23-U45M3X	BCH1803M21C

In the I/O servo drive position control mode of operation, Lexium 23 servo drives the pulse train to position control through the controller, which is also compatible with high differential multi pulse input. The PTO PLC generator outputs high speed pulses through digital output points Q0.0 and Q0.1 [12]. Here, we use the PLC Q0.1 output terminal as the digital output of the PTO generator, and the Q0.1 will not have other digital output functions.

The hardware of the electrical system includes sensors, cylinders and solenoid valves.

A sensor is a detection device. Its working principle is that it can be sensitive to the information of the measured object, and convert the detected signal source into electrical signals or other forms of signal output according to the specific rules, so as to achieve the requirements of information transmission, processing, display, record and control [13]. Sensors include electrochemical sensors, electrical sensors, resistance sensors, temperature sensors, displacement sensors, and pressure sensors [9]. In the combination drilling machine control system, the main sensor switches the amount of power to the electrical signal to the PLC, which serves as a basis for the judgment of the PLC control system CPU. Each pneumatic power head has two proximity switch sensors as signal bits for resetting and cutting. Proximity switch sensor adopts small square proximity sensor KJT-Y8S. The inductance proximity switch is composed of a LC high frequency oscillator and an amplifying circuit. It uses metal objects to produce eddy currents in the object near the oscillating induction head that generates electromagnetic fields. This vortex reacts on the proximity switch, oscillating capability is weak, and internal circuit parameters change. Thus, it is recognized that whether the metal object is approaching, thus controlling the on-off of the switch.

The cylinder transforms the pressure energy of the compressed air into mechanical energy, and the drive mechanism acts as a linear reciprocating motion, rotation, or swing motion [14]. Solenoid valve is an electromagnetic control of industrial equipment, which is an automatic basic component for controlling fluids. It is an actuating element that adjusts the direction, flow, speed, and other parameters of the medium. The gas source is connected by the main air pipe through the triple parts (air filter, pressure reducing valve, and oil sprayer), the electromagnetic valve

and one-way throttle valve to the actuating cylinder. No.1 pneumatic power head cylinder turns on the positive coil by reversing the two-sit five-pass (double control solenoid valve), and the positive air path is connected (positive movement makes air in stoma). Even if the positive action coil is switched off, the positive air path is still connected and will remain until the reverse action coil is energized. If the reverse action coil is energized, the reverse action gas circuit will be switched on (reverse movement makes air in stoma). Even if the reverse action coil is switched off, the reverse action gas path is still connected and will remain until the positive action coil is energized. This is equivalent to "self-locking".

Based on the analysis of the combination drilling machine control system, we can get the following control program design method.

After the PLC is powered on, the hardware and software are initialized and scanned. It will read the input, execute the user program, and process the communication request. It automatically checks whether the firmware, program memory, and extension module work correctly and overwrites the output. If the program uses an interrupt program, an interrupt event will occur. CPU stops the normal scan and executes the interrupt program.

If the program uses the immediate I/O directive, the value of the I/O point can be read and written directly.

According to the I/O number estimation, storage capacity, control function and model type, we select SIEMENS S7-200 CPU 226; programming software is selected as: STEP 7-Micro/WIN.

According to the actual processing requirements, PLC system will be equipped with 3 modes of work: automatic mode, manual mode and return to the origin. The selector switches SA1 SA3 indicate the control of the mode of the automatic operation, manual operation, and return points operation. The selector switches SA4 SA10 mean switching the power head 1, the power head 2, the power head 3, the power head 4, the power head 5, the power head 6, and the power head 7 in manual operation. The oil pump push button switch and the oil pump button switch (SB1, SB2) are used to control the opening and disconnection of the oil pump motor, which provide power for the hydraulic system. The water pump push button switch and the water pump button switch (SB3, SB4) are used to control the opening and disconnection of the water pump motor, and provide power for the cooling system. The clamping and loosening button switches (SB5, SB6) are used to control the clamping and loosening of the two states in manual operation. A rotary push-button switch (SB8) is used to control the rotation of the rotary table. The start and stop button switches (SB9, SB10) are used to control the opening and stop of actions in manual mode, automatic mode, and return to origin mode. The power head starter and the power head stop button switch (SB11, SB12) are connected in series with the contact of the selector switch SA4 SA10 to control the starting and stopping of the power head in the manual operation mode. Load power and emergency stop button switch (SB13, SB14): Press the "load power" button so that the KM coil is electrically self-latching. The KM's main contact is switched on, and the AC power is supplied to the external load. In case of emergency, disconnect the power with the emergency stop button. Manual work is mainly to debug machine tools, and

test whether the working position can meet the actual processing requirements. It is easy to fine tune the position of the fixture and sensor.

The PLC control program consists of four parts, the main program, the public program, the manual program and the automatic program.

The main program can call various subroutines to facilitate the switching of various working modes.

Public programs are used to handle the switching conditions between tasks that are performed in different ways and different modes of work.

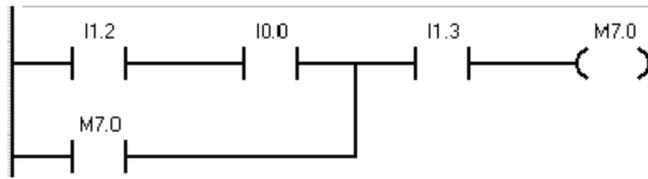
In the manual program, when the system is in manual operation, the control requirements are: the separate work and stop of 7 power head motors; the feed movement of fast forward, together forward and rewind of 7 power head motors as well as the clamping and loosening of the cylinder valve. Manual operation: press the "relax" button, the cylinder valve control fixture in a relaxed state, and we can manually put the upper part. After we have finished manually pressing the clamping button, the workpiece is clamped in the fixture. Press the power head start button, then the power head motor starts the spindle rotation. Then press the "start" button, the feed movement will begin to complete the fast forward, together forward, and rewind backward and stop in situ. At this time, control motor power off through the "power head stop". Then through the selection switch, adjust the required power station head motor to get power. Control the transposition by means of the "turn" button. This is done manually. The purpose of manual operation is to test the manufacturability of the system and test whether it can meet the process requirements. If there is a deviation, proper debugging is required.

An automatic program is a way of controlling the work of the system cycle. Automatic working mode mainly controls the automatic feeding, machining, indexing and feeding of the system. The automatic program describes the control process of the automatic program with the actions of the power head, the feeding mechanism and the No. 1 fixture.

#### 4. Results analysis and discussion

In order to prove the feasibility of PLC control system, an example is given to verify the feasibility of the control system. A continuous flag bit M7.0 is set in the program, which is used as a continuous or stop conversion condition. The normally open contact of the start button is connected in series with the normally closed contact of the stop button. When the "start button" is pressed, the continuous flag M7.0 is ON, and the M7.0 coil remains self-locked, so as to cycle in cycles. When the stop button is pressed, the continuous flag bit M7.0 becomes OFF, and the system completes the cycle and stops at the initial step. Fig. 4 is the automatic program *a*.

When the cycle is stopped, the stop button is pressed to break the normally closed contact of the stop button switch. At this point, the successive flag bit M7.0 is reset to OFF. When M7.0 is OFF, the system reset the initial step M0.0 and performs the previous step of stopping the loop. M5.5, M5.6, M5.7, M6.0, M6.1, M6.2, M6.3 steps are reset, and the power head motors (Q1.1, Q1.2, Q1.3, Q1.4, Q1.5, Q1.6, and Q1.7) are reset. The system is in a stopped state. Fig. 5 depicts the

Fig. 4. Automatic program *a*

automatic program *b*.

Fig. 5. Automatic program *b*

Fixture cylinder clamping and relaxation of the cycle control requires that after feeding, the workpiece enters into the fixture and clamped the workpiece. The 10 clamp cylinder is also a cycle of clamping and releasing. When it is at the No.1 position, the feeding of the feeding mechanism is in place. The contact sensor I8.0 of the front limit sensor is ON, and the C1 counter is ON. The normally open contact of the C1 is ON, and it triggers the Q2.1 to ON and clamps the workpiece. When the normally closed contact M8.1 is broken (OFF), the Q2.1 is OFF. The cylinder is in a relaxed state, and the material ejecting mechanism ejects the workpiece. The other No.2 to 9 cylinders works in the same way, but the trigger conditions are different. The counter of No. 1 cylinder is triggered by the loosening of cylinder 1, and the Q2.1 is on the falling edge of OFF to trigger the reset of the counter. The No. 2 cylinder clamping Q2.2 is the rising edge of ON, the trigger circuit is disconnected to control the No. 1 cylinder Q2.1OFF to relax, and the workpiece is ejected into the clamp. Fig.6 is the automatic program *c*.

The work processes are as follows. The process that workpiece is introduced from the charging mechanism is called "feeding". The return of the finished material is called "feeding returns". The upward movement of the clamp cylinder is called "release". The downward movement of the clamp cylinder action is called "clamping". When feeding the material and the feeding mechanism triggers the "front limit" sensor, the switch quantity signal becomes ON, and the clamp cylinder is "tensioned". When the lower limit of the cylinder is triggered, the switch signal is changed to ON, and the feed mechanism cylinder is "returned". When the rear limit sensor is triggered, the rotary table is lifted. When the upper position sensor of the rotary worktable is triggered, the rotary table rotates 36 degrees. When the rotary

trigger rotary position sensor, power head start to fast forward, together forward and rewind. The projectile is used to trigger the loosening of the clamp cylinder by 10 counters, and then the ejection mechanism ejects the workpiece. With the set program, the system goes round and round until the "stop" button is pressed.

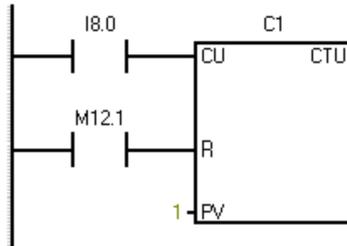


Fig. 6. Automatic program c

## 5. Conclusion

In order to solve the problems in actual production, the automatic drilling machine control technology based on PLC and its application were studied in this paper. The parts of the combination drilling machine were selected according to the requirements. Then through the formulation of the I/O point table, the PLC control system which could meet the requirements was designed. The practical verification was carried out to achieve theoretical linkage practice finally. Practical validation of the theoretical results demonstrated the feasibility of the PLC control system. Some conclusions were obtained as follows: in this paper, through the analysis of the control requirements of modular machine tools, a variety of composite components were selected to formulate I/O point table and design the PLC control system to meet the requirements. PLC's "automatic" mode of operation, "manual" mode of operation and "return to the origin" work were in line with the actual production needs. In the "manual" mode of work, the machine tools could be debugged, and the quality of product processing could be tested. In the "automatic" mode of work, the system can achieve periodic work. The "return to the origin" mode can provide initialization conditions for the "automatic" mode of operation. Although this paper has achieved good results, there are still some problems that need further study. For example, through the study of this paper, it can be seen that the design process of PLC control is tedious, and how to implement a more convenient PLC design is a key point of research in the future.

## References

- [1] H. GAO, Y. J. WU: *The research status and development trend of electric vehicle power supply technology*. Advanced Materials Research 1008–1009 (2014), 381–384.

- [2] C. Q. DENG, D. Q. XIONG: *The application and development of the PLC in the control system for NC machine tool*. Manufacturing Automation 31 (2009), No. 11, 108–110.
- [3] S. VOSOUGH, A. VOSOUGH: *PLC and its Applications*. International Journal of Multi-disciplinary Sciences and Engineering 2 (2011), No. 8, 41–46.
- [4] J. WAN, S. TANG, Z. SHU, D. LI, S. WANG, M. IMRAN, A. V. VASILAKOS: *Software-defined industrial internet of things in the context of industry 4.0*. IEEE Sensors Journal 16 (2016), No. 20, 7373–7380.
- [5] J. SCHLECHTENDAHL, M. KEINERT, F. KRETSCHMER, A. LECHLER, A. VERL: *Making existing production systems Industry 4.0-ready*. Production Engineering 9 (2015), No. 1, 143–148.
- [6] A. F. ROMERO-LÓPEZ, M. L. ORTEGA: *Pneumatic auto-control model for breathing regulators*. IFAC Proceedings Volumes 23 (1990), No. 8, Part 2, 209–214.
- [7] Y. XING, S. JIN, J. A. ZHOU: *Assembly sequence planning for auto-body parts based on liaison graph*. Automotive Engineering 28 (2006), No. 2, 190–193.
- [8] E. D. KUNT, A. T. NASKALI, I. S. M. KHALIL, A. SABANOVIC, E. YÜKSEL: *Design and development of workstation for microparts manipulation and assembly*. Turkish Journal of Electrical Engineering & Computer Sciences 19 (2011), No. 6, 973–992.
- [9] S. FATIKOW, J. SEYFRIED, S. FAHLBUSCH, A. BUERKLE, F. SCHMOECKEL: *A flexible microrobot-based microassembly station*. Journal of Intelligent and Robotic Systems 27 (2000), Nos. 1–2, 135–169.
- [10] S. P. LONG: *Design on control system of automatic sole edge grinding machine based on PLC*. Advanced Materials Research 998–999, (2014), 741–744.
- [11] C. YU, X. HUANG, C. FANG: *Optimizing dynamic characteristics of NC rotary table based on electromechanical-hydraulic coupling*. Journal of Mechanical Science and Technology 27 (2013), No. 4, 1081–1088.
- [12] G. Q. WANG: *Design based on PLC programming control de-dust system*. Applied Mechanics and Materials 644–650 (2014), 3643–3646.
- [13] M. TRIGUI, R. BENHADJ, N. AIFAOU: *An interoperability CAD assembly sequence plan approach*. International Journal of Advanced Manufacturing Technology 79 (2015), Nos. 9–12, 1465–1476.
- [14] X. ZHOU, P. DU, Y. ZHOU: *A model-based approach to assembly sequence planning*. International Journal of Advanced Manufacturing Technology 39 (2008), Nos. 9–10, 983–994.

Received June 6, 2017

# Application of PLC technology in fire-fighting control system of oil field united station

CHANG CHUNYANG<sup>1</sup>

**Abstract.** In recent years, many hidden dangers have appeared in the process of using energy, especially the fire safety in the joint station of oil field. Fire control used to be controlled by hand, and should be updated. In this paper, PLC technology was introduced into the fire protection system of the oil field joint station, and a fire control model was established and applied to the oil field united station. The model results show that PLC technology can help the whole oilfield joint station achieve integrated fire control, simplify operation flow, reduce operation difficulty and improve its reliability and stability. Therefore, PLC technology can effectively improve the automation technology of oilfield joint station, and ensure its operation more security.

**Key words.** Oil field united station, fire-fighting control system, PLC technology.

## 1. Introduction

Oil is a very important source of energy, which occupies a very important position in human daily life and production, and can effectively promote the development of social economy [1]. Therefore, the dependence of mankind on oil is getting higher and higher, the efficiency of oil use is also increasing, and the oil field united stations are built up, which also occupy a very important position in the oil and gas gathering and transportation system [2]. In this context, the safety of the oil field united station has received increasing attention [3]. With the continuous development of information technology, all walks of life are beginning to introduce information technology into the industry to apply [4]. Then, the oilfield oil and gas gathering and transportation system also introduces the PLC technology, and it is widely applied to all aspects of the oil field united station [5]. At present, the fire control safety of many oil field united stations is still the manual operation, and this backward control method is often prone to errors, so there is an urgent need to improve the management level [6].

When there is a fire in the oil field united station, people need to promptly detect

---

<sup>1</sup>Langfang Polytechnic Institute, Langfang, Hebei, 065000, China

and analyze the fire situation and make a rapid response to the sudden situation, while it is more prone to security risks through the antiquated manual operation [7]. Therefore, it is necessary to build a series of advanced, scientific fire-fighting control safety system, so as to ensure the safety of the oil field united station. Therefore, the PLC technology was introduced into the fire-fighting control system of the oil field united station in this paper, thus improving the control level of its fire-fighting safety [8].

## 2. State of the art

Because the oil field united station has a pivotal position in the whole system, it is necessary to take strict control on its safety control, so as to guard against the safety accident [9]. Since the fire safety part is the focus of the safety control system, it is necessary to strictly control its control mode [10]. As the past oil field fire-fighting control system is controlled by manual operation, so its control mode must be reformed. Therefore, this paper applies the PLC technology to the oil field united station fire-fighting control system and makes an in-depth discussion on its application effect [11].

PLC technology has more prominent advantages, for example: the faster transmission speed of this technology, the stable and high transmission quality, which is very important for the application of the oil [12]. Till now, the production control and safety control of many oil field united stations have been done manually, which have required a larger area for the oil field united station [13]. Thus, it is currently imminent to improve the information level of the operation mode of oil field united station [14]. Therefore, in recent years, our country has also begun to update the fire safety control operation of the oil field united station through PLC technology [15].

## 3. Methodology

### *3.1. Overview of fire-fighting control in oil field united station*

In the design process of the fire-fighting control system of the oil field united station, the PLC technology is adopted as the core technology of its control system, furthermore, an engineering station is set up in the control office of the fire safety system, so that the engineers can control the fire safety of the oil field united station. In addition, two operator stations are also set up to enable staff to effectively control the safe operation of the oil field united station at all times.

PLC technology can quickly transfer fiber to the system and retain the Ethernet interface, so that the higher management personnel of the control center can more easily check and monitor the production process of all aspects of the oil field united station, and the data and information in the system can be transmitted to the superior management office, in addition, the orders issued by superior can also be

received, so as to make a quick response. Through the above-mentioned work links, the degree of automation of management and control work of various aspects of oil field united station is rapidly improved. Figure 1 shows the oil field operations site, and the oil field safety must be fully protected.



Fig. 1. Oil field operation field

When the sudden fire of the oil field united station happens, the fire safety control center will find the fire location in a fastest speed and issue an alarm. When staffs of the oil field united station hear the alarm, they can respond timely and go to the fire control center to control the electric valve and fire pump, and the fire situation. Furthermore, the cupping foam mixture is opened through the remote control system, so that the fire can be quickly controlled. In addition, the fire-fighting control center can also view the main line of fire water in real time, and timely warning. When the pressure in the fire-fighting system is less than 1 MPa, the control center will automatically open the circulating pump, and when the pressure in the fire-fighting system is greater than 1 MPa, the control center will delay for ten seconds and suspend the circulating pump. Subsequently, according to the water level in the fire pool, the control center will control it, and take a high level alarm mechanism, so that staff can find the security risks existing in the oil field united station, quickly remove these risks, so as to ensure the safe operation of the oil field united station.

### ***3.2. Description of fire-fighting interlocking in oil field united station***

The initiation of oil field united station is usually equipped with detectors and fire dike, which can detect the temperature of this region. When the temperature exceeds a certain extent, the fire alarm button can be opened. When the area bursts fire, a warning can be sent at a very fast speed through the PLC technology, then, the fire situation can be confirmed quickly and the staff can be informed. After the staff confirming the fire, the fire electric valve and water valve, as well as the pump can be quickly opened through the remote control system, so as to control the fire. After the above links, the staff can also close the electric valve in this area, thereby guarding against the fire risks caused by power leakage.

When there is an oil field united station fire in the reality, a fire signal will be quickly issued. And when the staff in the control center quickly finds the fire, the first step is to immediately confirm it and quickly open the foam tube and fire water pipes. The second step is to open the first set of fire pumps and foam pumps after 15 seconds, and open second fire pumps and foam pumps after another 10 seconds. The third step is to send the control status to PLC in time when the fire water pumps and foam pumps of the above two groups are opened for 15 seconds, and to analyze the feedback information in depth through PLC technology, so as to make accurate judgments about the process and causes of the fire. If the fire water pumps and foam pumps of the above two groups are already open, then, there is no need to open the fire pumps and foam pumps in the third group.

Then, the control center is also capable of conducting the real-time detection on the fire-fighting water main line pressure. When the detected pressure is less than 1 MPa, the first group of fire pumps and foam pump will automatically open, and its operation situation is detected, then, the detected information is sent to the PLC center. After 10 seconds, if the operation is relatively smooth, there is no need to open the second fire pump and foam pump. However, when the first group of fire pump and foam pump suspends its operation, the second group of fire pump and foam pump must be opened. When the detected pressure is greater than 1 MPa, the first and second groups of fire pumps and foam pumps will be suspended after 10 seconds.

### ***3.3. Hardware configuration of fire-fighting control system in oil field united station***

PLC technology is introduced into the fire-fighting system of the oil field united station, and then, a fire control model is established. According to the prescribed process, the series S7-300C PLC technology is selected and the fire-fighting duty office is set up in the control center. In addition, the engineer and operation rooms are also set up in the control center, so as to facilitate the operation of staff. Among various equipment of the oil field united station, the transmission and exchange of information can be carried out through the switch. Moreover, the fire-fighting system capacity adjustment is conducted through these control equipment. Figure 2 shows the hardware configuration of the fire-fighting control system of the oil field

united station.

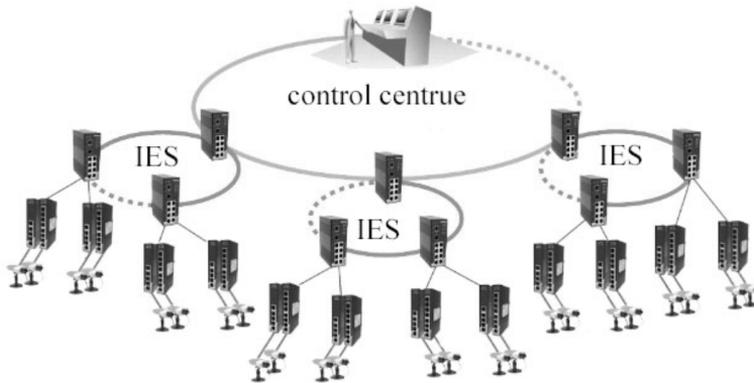


Fig. 2. Hardware configuration of fire-fighting control system in oil field united station

At the same time, a power supply is installed and three digital input and output modules are configured. In addition, in the fire-fighting control center, the industrial-grade switches are also purchased for information exchange. In order to allow the system's information exchange to maintain long-term stability of the state, three microcomputers are also configured to help the data transmission work, so as to make the whole system can run perfectly and smoothly.

### ***3.4. Software design of fire-fighting control system in oil field united station***

After the above design, the software design of the fire control system of the whole oil field joint station was also carried out in this paper, so that the system can have perfect performance and practicality. Firstly, the PLC technology was selected as the basis of the system, and STEP7 was regarded as the system's development environment. In the design of the system software, the programming was achieved in accordance with the ladder language, which can fully ensure that the system can be more efficient and smooth in the use process, so that users can more easily grasp the use methods. At the time of using PLC technology to conduct programming, the fire-fighting safety control process of the oil field united station should be comprehensively designed.

At the aspects of the design of man-machine interface and the choice of hardware, based on facilitating users' application, the engineer station and operator station set by the control center were also constructed through high performance computers. In the software selection, the SIMATIC WinCC V6.0 algorithm was selected, and furthermore, the system interface was devised more concise and the entire system parameters were designed more reasonably. Figure 3 showed the basic structure of the fire-fighting control system of the oil field united station.

The basic architecture in Fig. 3, is divided into three layers. These three layers

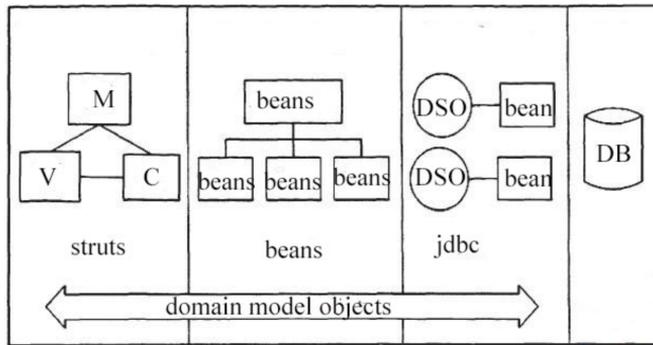


Fig. 3. Basic structure of fire-fighting control system in oil field united station

are data layer, liaison layer and transport layer. The first layer is responsible for collection, analysis and data storage. The second is responsible for the various functional modules communicating with each other, and the third layer is responsible for all kinds of information sharing and transmission functions.

In terms of function settings, the system was enabled to perfectly support the fire-fighting control, report production and printing, network information release and other functions. In the functions that the system can implement specifically, all operations of the oil field united station were explained through the main screen, and the comprehensive control was also carried out. Furthermore, all parameters, such as pressure, temperature and other data, were set in advance. While at the interface of the control center, an emergency device was displayed. When an emergency occurs, the staff can react quickly. In addition, a report production and display interface were also set up to clearly show the data and information on the report to staffs. At the same time, the information management of the file can be also achieved. The most important thing was that the alarm interface can display alert messages, so that the staff can respond quickly after receiving an alarm.

#### 4. Result analysis and discussion

In this paper, PLC technology was introduced into the oil field united station, and through these technologies, the fire safety aspects of the oil field united station were strictly controlled, thus establishing the fire-fighting control system of the oil field united station. After the establishment of the fire-fighting control system of oil field united station, it was necessary to apply the system to the oil field united station, so as to verify the safety control ability of system in the operation process of the oil field united station. After the above research, the fire-fighting control system of the oil field united station was applied into the oil field operations, and a complete collection for the resulting data was carried out. Moreover, functions and effects of the system on the fire safety control of the oil field united station were determined through these data. The following Table 1 the data obtained by applying the oil field united station fire-fighting control system to the oil field united station.

Table 1. Experimental data collection of fire-fighting control system in oil field united station

Project	Oil field a	Oil field b	Oil field c	Oil field d
Experiment 1	8.11	3.67	6.65	6.95
Experiment 2	5.30	0.63	3.84	4.15
Experiment 3	4.58	4.91	4.88	3.05
Experiment 4	1.27	5.66	3.46	6.10

The data collected above was the level of fire safety in four stages of the experiment of the oil field united station fire-fighting control system applied by the oil field united station, so the data would be analyzed deeply after collecting data. Then, the above data were input to the system and calculated, and after getting the results, the data were analyzed, so that the fire-fighting safety role of the oil field united station fire-fighting control system on the oil field united station was obtained. The following Table 2 shows the results calculated by the oil field united station fire-fighting control system.

Table 2. Calculation results of fire-fighting control system in oil field united station

Project	Oil field a	Oil field b	Oil field c	Oil field d
Experiment 1	8.31	7.63	6.48	5.88
Experiment 2	2.09	8.64	5.98	1.87
Experiment 3	5.49	2.95	7.13	1.13
Experiment 4	2.19	1.22	6.77	9.96

Then, the above data was analyzed, and the impact to the fire safety control level after applying the oil field united station fire-fighting control system to the oil field operation was summarized. Figure 4 shows the role and effect of the fire-fighting control system of the oil field united station.

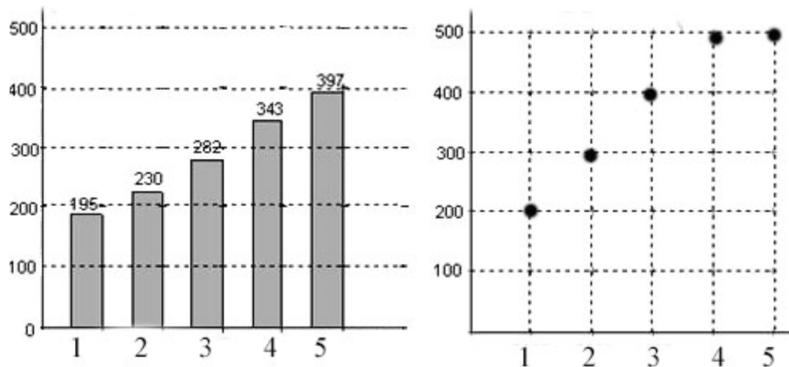


Fig. 4. Application effect of fire-fighting control system in oil field united station

From Fig. 4, it could be clearly seen that according to selection of PLC technology to complete the entire oil field united station fire-fighting control, the entire fire-fighting system could make the operation easier and the maintenance could be more convenient. At the same time, the system had higher reliability and security, which greatly improved the level of automatic production of the oil field united station, ensured the safe, efficient and stable operation of the oil field united station. And then, after the system was put into operation, the good results were achieved.

The fixed temperature fire detector and manual fire alarm button are set up at the top of the tank. When the fire occurs in the tank field, the system can quickly sends the warning to the oil field united station fire-fighting control system in the control room and the fire duty room, meanwhile, the start button of the fire confirmation is set up in the control room and the fire duty room, after the operator on duty determines the fire, the fire confirmation button is started. In addition, oil field united station fire-fighting control system can also bring good economic benefits to the oil field united station. At the same time, the system takes the Ethernet as its communication way, making the whole system very rhetoric, which can upgrade the system all the time in accordance with the actual size of the united station, support remote download and program modification, so as to make the whole system easier to maintain.

## 5. Conclusion

As the human's development degree on the oil is getting higher and higher, the control on the production and security is the focus of the entire united station in the whole production operation of the oil field united station. In this paper, PLC technology was used to control the fire safety in the oil field united station, so as to expect to make fire-fighting control operation more convenient. Therefore, based on the PLC technology, in this paper, the fire-fighting control system of the oil field united station was established, and the system to the actual operation of the oil field united station was applied, so that the fire safety control level of the oil field united station was effectively improved. This not only greatly improved the safety of the oil field united station, but also fully improved its level of information. Furthermore, its production level was also effectively improved, which brought great economic benefits. Therefore, the application of PLC technology in the oil field united station fire-fighting control work can effectively improve its fire safety control level.

## References

- [1] J. HAN, C. S. CHOI, W. K. PARK, I. LEE, S. H. KIM: *Smart home energy management system including renewable energy based on ZigBee and PLC*. IEEE Transactions on Consumer Electronics 60 (2014), No. 2, 198–202.
- [2] J. HAN, C. S. CHOI, W. K. PARK, I. LEE, S. H. KIM: *PLC-based photovoltaic system management for smart home energy management system*. IEEE Transactions on Consumer Electronics 60 (2014), No. 2, 184–189.

- [3] A. B. FUKS, S. GAVRA, A. CHOSACK: *Long-term followup of traumatized incisors treated by partial pulpotomy*. *Pediatric Dentistry* 15 (1993), No. 5, 334–336.
- [4] V. GRUBER, R. MARCELINO, J. B. DA SILVA: *Technology PLC - Power line communication, used in monitoring systems online*. *International Journal of Online Engineering* 9 (2013), No. S1, 22–25.
- [5] X. WANG, P. SHAO, H. SUN: *Application of PLC monitoring system to fire protection engineering of large-scale tank area*. *Petrochemical Safety and Environmental Protection Technology* (2011), No. 2, 219–221.
- [6] J. S. GAO, Y. HUANG, H. WANG: *Application of fuzzy control in aircraft fire fighting system*. *Microcomputer Information* 25 (2009), No. 22, 35–36.
- [7] P. HALLINGER, M. LEE, J. KO: *Exploring the impact of school principals on teacher professional communities in Hong Kong*. *Leadership and Policy in Schools* 13 (2014), No. 3, 229–259.
- [8] R. BHAISSWAR, A. A. SALODKAR, P. KSHIRASAGAR: *Power Management using PLC and SCADA*. *International Journal of Engineering Innovations and Research* 1 (2012), No. 1, 32–35.
- [9] K. SUZUKI, T. YAMADA, O. MORIWAKI, H. TAKAHASHI, M. OKUNO: *Polarization-insensitive operation of lithium niobate Mach-Zehnder interferometer with silica PLC-based polarization diversity circuit*. *IEEE Photonics Technology Letters* 20 (2008), No. 10, 773–775.
- [10] M. G. IOANNIDES: *Design and implementation of PLC-based monitoring control system for induction motor*. *IEEE Transactions on Energy Conversion* 19, (2004), No. 3, 469–476.
- [11] S. J. HU, H. X. ZHANG, Y. J. CHANG, J. LI, C. J. LI, L. L. ZHANG: *The embedded soft PLC development system research of numerical control machine*. *Applied Mechanics and Materials* 401–403 (2013), 1724–1727.
- [12] A. SENDIN, I. BERGANZA, A. ARZUAGA, X. OSORIO, I. URRUTIA, P. ANGUEIRA: *Enhanced operation of electricity distribution grids through smart metering PLC network monitoring, analysis and grid conditioning*. *Energies* 6 (2013), No. 1, 539–556.
- [13] T. A. PAPADOPOULOS, C. G. KALOUDAS, A. I. CHRYSOCHOS, G. K. PAPAGIANNIS: *Application of narrowband power-line communication in medium-voltage smart distribution grids*. *IEEE Transactions on Power Delivery* 28 (2013), No. 2, 981–988.
- [14] R. WANG, Y. GUAN, L. LUO, X. LI, J. ZHANG: *Component-based formal modeling of PLC systems*. *Journal of Applied Mathematics* (2013), No. ID 721624.
- [15] A. PIĘTAK, M. MIKULSKI: *On the adaptation of CAN BUS network for use in the ship electronic systems*. *Polish Maritime Research, Journal of Gdansk University of Technology* 16 (2009), No. 4, 62–69.

Received July 12, 2017



# Improvement research of genetic algorithm and particle swarm optimization algorithm based on analytical mathematics

SHUAI MAN<sup>1</sup>

**Abstract.** Through its own evolution, the optimization method makes many problems that seem highly complex to be solved more perfectly, so a new intelligent calculation method that is different from the classical optimization method is generated. In this paper, the algorithm mechanism, algorithm improvement and application of two kinds of biomimetic intelligent calculation methods of genetic algorithm and particle swarm optimization algorithm were studied deeply. Besides, in view of the constrained optimization problems, two different improvement strategies were adopted, two different improved evolutionary algorithms were proposed respectively and their time and spatial complexity were analyzed. The final experimental results proved that the evolutionary algorithm that integrates the improvement strategies is feasible and effective, and the uniformity and diversity of the solution set are ideal.

**Key words.** Genetic algorithm, particle swarm optimization, optimization problem.

## 1. Introduction

The essence of all human activities is nothing more than "knowing the world and building the world". Understanding the world depends on establishing the model, constructing the world relies on the optimal decision-making, and the purpose of optimization is to find a set of parameter values under meeting certain constraint conditions, so as to make some of the performance indicators of the model reach maximum or minimum. The application of the optimization problem can be said everywhere, which always runs through the process of all human activities. In a sense, all human knowledge is nothing more than the phenomenon and process understanding model of human beings to a field. The purpose of knowing the world is to build the world better, similarly, and the purpose of modeling is to optimize. Assuming that the world must first understand the world, similarly, all optimization

---

<sup>1</sup>Langfang Health Vocational College, Langfang, Hebei, 065000, China

cannot be separated from the model. However, with the continuous improvement of social productivity, the ability of humans to understand and build the world is also growing, followed by optimization problems, which also show the features of high-dimensional, strong nonlinear, strong constraints, difficult to model.

## 2. State of the art

Bionics was founded in the middle 1950s. People seek a new method for solving all kinds of complex problems in real world from the biological evolution mechanism, and then, the biological simulation becomes an important part of computer science. For example: the early theory was to assume that the machine was composed of basic elements in which they were similar to the neuron [1]. Under the influence of this biological simulation thought, in the early 1970s, a professor in the University of Michigan first proposed the mathematical framework of the genetic algorithm [2]. The idea of genetic algorithm came from Darwin's evolutionism, Weismann's theory of species selection and Mendel's theory of population genetics in biological sciences. In terms of GA, it was a probabilistic search algorithm method that used natural selection and evolutionary mechanisms to find the optimal point in  $N$ -dimensional space [3]. From the evolutionary thought, "the survival of the fittest" makes the individual quality of the population has been improved. And the random exchange theory uses the existing information in the original solution to speed up the search process to the optimization. Since more than a decade of the proposition of the particle swarm optimization algorithm, it has attracted many researchers and research institutions at home and abroad to conduct various aspects of exploration to its theory and application. In addition, the research results of the PSO algorithm are increasingly published in high-level publications [4]. Then, the famous conference IEEE CEC in the field of evolutionary computing has set up a special discussion of PSO algorithms. And the important international conference PPSN and GECCO related to computational intelligence have made PSO algorithms as one of the key themes of the conference [5]. In 2001, in view of the PSO theory research and the emergence of applied monograph Group Intelligence, in 2003, the first swarm intelligence symposium IEEE Swarm Intelligence Symposium was held in the United States. Then, PSO algorithm was regarded as one of the main bodies in each year of Symposium. In 2004, the top academic journals in the field of evolutionary computing IEEE Transaction on Evolutionary Computation published PSO algorithm special issue, till now, PSO algorithm has become an important research topic in computational intelligence field [6].

## 3. Methodology

Genetic algorithm is a kind of self-organizing and adaptive probabilistic search algorithm which simulates the natural evolution process and mechanism to solve the optimization problem. It does not depend on the specific model of the problem, and has strong robustness to all kinds of complex optimization problems. The basic

idea of genetic algorithm constructs a fitness function according to the objective function of problems that wait to be solved. Then, according to certain rules, the initial population after the gene encoding is generated, and the evaluation, genetic operations (crossover and mutation), selection and other operations are carried out to the group [7]. After several generations of evolution, one or several optimal individuals with the best fitness are obtained as the optimal solution of the problem. Figure 1 shows the basic GA flow chart. It can be seen from Fig.1 that the steps in the genetic algorithm implementation include coding strategy, initial population generation, fitness function design, selection strategy, genetic operation and stop criterion.

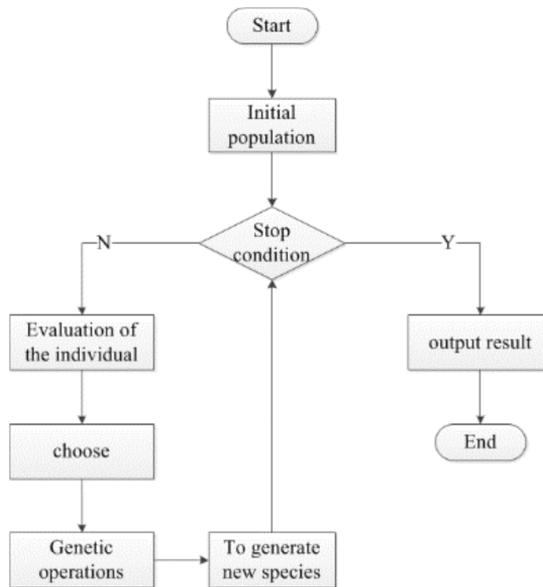


Fig. 1. Genetic algorithm flow chart

It is found that under the situation of the absence of centralized control, the bird group composed of a large number of individuals can make collective acts in flight, such as changing direction, spreading or reorganizing the formation. According to the further observation to the predatory behavior of birds, people feel that there must be some potential capacity or rules to ensure these intelligent behaviors. As a kind of bionic evolutionary algorithm, PSO is similar to genetic algorithm, and it is a kind of optimization technology based on iteration. However, there is no cross mutation operation in the algorithm implementation process [8]. At present, many improved algorithms have been proposed, such as adaptive PSO, hybrid PSO, cooperative PSO [9]. These improved algorithms are mostly based on the standard particle swarm optimization algorithm, and Fig. 2 describes the implementation flow chart of the particle optimization algorithm. The basic principles of the particle swarm optimization algorithm are described as follows.

A group composed of  $m$  particles (Particles) flies at a certain speed in  $D$ -dimensional

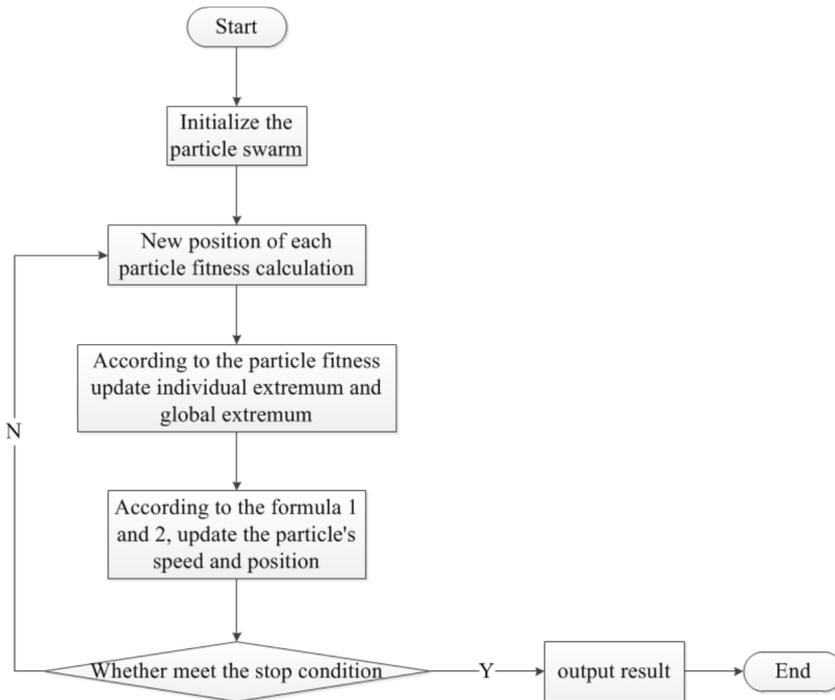


Fig. 2. Particle swarm optimization algorithm flow chart

search space, each particle representing a bird in the search space. For an optimization problem that waits to be solved, a particle is a potential solution [10]. Each particle also has a speed to determine the distance and direction of its flight. All particles have a fitness value that is determined by the optimized function. In the course of the flight, the particles will use their own flight experience and group flight experience to dynamically adjust themselves, after several iterations of the search, and ultimately, the optimal solution is obtained [11]. PSO is initialized as a group of random particles, and then, the optimal solution is found by iteration. The particles update themselves by tracking two "extremes" in each time of iteration. One is that the optimal solution found by the particle itself is called the individual extremum  $pbest$ , and the other is the optimal solution currently found by the whole population. This extreme is the global extremum  $gbest$  [12]. Figure 3 depicts the trajectory of the particle flight, and each particle updates its speed and new position by formulas 1 and 2 when the two optimal values are found:

$$v_{k+1} = c_0 v_k + c_1 (pbest_k - x_k) + c_2 (gbest_k - x_k), \quad (1)$$

$$x_{k+1} = x_k + v_{k+1}. \quad (2)$$

Here,  $v_k$  is particle's velocity vector,  $x_k$  is the position of the current particle,  $pbest_k$  represents the position of the optimal solution found by the particle itself,

$qbest_k$  represents the position of the optimal solution currently found by the whole population,  $c_0$  is the random number that is generally between (0, 1), which is called the inertia coefficient or contraction factor. Symbols  $c_1$  and  $c_2$  are called as the "self-cognition factor" and "social cognition factor" of the particle, which are respectively used to adjust the effect intension of  $pbest_k$  and  $qbest_k$  to the particle attraction. Then, the values of  $c_1$  and  $c_2$  are the random number between (0, 2). Finally,  $v_{k+1}$  is the sum of vectors  $v_k$ ,  $pbest_k - x_k$  and  $qbest_k - x_k$ . The velocity of each dimension of the particle will be limited by a maximum speed  $v_{max}$ .

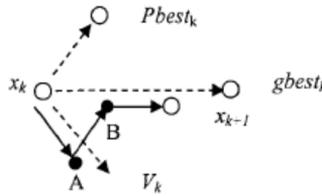


Fig. 3. Particle flight path map

In the PSO algorithm, if all the particles in the group are used as the neighborhood members, at this time, it is known as the global version of the PSO algorithm; if part of members in the group constitute the neighborhood, then, this is called the local version of the PSO algorithm [13]. In the local version, there are two ways to form the neighborhood, one is that the particles in which the index numbers are adjacent constitute the neighborhood, and the other is that the particles that are adjacent according to the spatial distance constitute the neighborhood. The neighborhood definition strategy of the particle swarm optimization algorithm is also called the neighborhood topology structure of the particle swarm.

Multi-objective optimization is a common problem in the field of engineering. Its main characteristic is that there is a conflict between the targets, furthermore, all the targets cannot obtain the optimal value at the same time, only to find a set of compromised Pareto non-inferior solution [14]. The traditional method of solving MOP is to convert MOP into a number of different single objective optimization problems, and then solve them. For the case where the front of the Pareto is non-convex, all Pareto optimal solutions cannot be obtained and the computational complexity is large. Evolutionary algorithm is an adaptive global optimization probabilistic algorithm of the simulated creatures formed in the genetic and evolutionary processes in the natural environment. It is characterized by the multi-directional and global nature of the search, which can process large-scale search space in parallel, and moreover, it has a good adaptability to complex MOP.

Based on the Pareto principle's evolutionary algorithm elitist mechanism, preserving the non-inferior solution obtained by constructing the Pareto candidate set and maintaining the solution diversity in this solution set is an effective means for the multi-objective evolutionary algorithm to obtain non-inferior solution [15]. Based on the criterion of neighborhood function, this paper proposes the construction and maintenance mechanism of Pareto candidate set. The process is described as follows:

Firstly, if the scale of the Pareto candidate solution set does not reach the specified size, the obtained non-inferior solution is added directly to the Pareto candidate solution set.

Secondly, if the new non-inferior solution dictates the individual of the Pareto candidate solution set, the new individual is added to the Pareto candidate solution set and the individuals dominated are placed into another independent external set (IES). Otherwise, the new non-inferior solution is directly added to the Pareto candidate solution set, then, the neighbor function criterion is used to maintain the diversity and population size of the Pareto candidate solution set.

Thirdly, the neighborhood local search is carried out to the individuals of IES, and the newly obtained individual is compared with the Pareto candidate solution set again. If the new individual dominates an individual of the Pareto candidate solution set, this individual is replaced directly. Then, empty the IES.

The algorithm flow is as follows:

Step 1: Setting algorithm parameters: evolutionary algebra is  $N_p$ , the length of the population is  $I_s$ , Pareto candidate solution set length is  $E_s$ , local search length is  $L_s$ , crossover probability is  $P_c$ , probability of mutation is  $P_m$ , and evolutionary algebraic indexer is  $t$ .

Step 2: Initializing the population  $p_{init}$ , assuming  $t = 1$ .

Step 3: A set of non-inferior solution is obtained from the group  $p_{init}$ , and the number of the non-inferior solution is  $u$ , then, the elite retention mechanism is used to add non-inferior solutions into the Pareto candidate solution set.

Step 4: In Pareto candidate solution set, the discrete crossover operator and Gaussian variation operator are used to generate  $u_f \times v$  individual,  $u_f$  is the individual number in Pareto candidate solution set, and  $v$  is the proportion coefficient of the sum of descendants and offspring.

Step 5: The population obtained by the cross and variation in the above steps is combined with the Pareto candidate solution set, and then, they are combined into a new population  $p_{new}$ , next is to execute  $p_{init} \leftarrow p_{new}$ .

Step 6: Symbol  $I_s$  is the number of individuals which are selected in the population  $p_{init}$ . Here, NSGA-II's non-inferior solution scheme is adopted, and one selection is carried out from low to high (the lower the level is, the higher the rank is), thus, the limited different "grade" groups  $\{f_1, f_2, \dots, f_n\}$  can be obtained, similarly, in the last grade of population  $f_i$ , ( $1 \leq i \leq n$ ) of the individual, the neighbor function criterion is used to select the remaining individuals, and if the termination condition is satisfied, the algorithm stops, otherwise, it moves to Step 3.

## 4. Result analysis and discussion

Experimental environment: Intel Pentium 4, 2.26 GHz, 512 MB memory, Windows Xp Professional, Matlab 7.0.

The experiment was carried out in two groups G1 and G2, and the algorithm used the real number coding. Then, two groups of typical multi-objective optimization function were selected, the first group of optimization problem was a single target minimum optimization function with high-dimensional constraint conditions, the

variable constraints included the test questions of inequality constraints and equality constraints, and the objective function contained up to 10 decision variables. In the second group, the performance comparison of the algorithm in this paper and two commonly used multi-objective evolutionary algorithms NSGA-II and SPEA were conducted by using graphing method.

In the first group of tests, the initial parameters of the algorithm were set as: population individuals were  $I_s = 200$ ,  $N_p = 300$ , Pareto candidate solution set length was  $E_s = 30$ , local search length was  $L_s = 5$ , crossover rate was  $P_c = 0.9$ , variation rate was  $P_m = 0.5$  and each issue ran 20 times independently under the same conditions. Then, all of the implementation of the algorithm was completed on the same computer, and when the algorithm was running, the calculation accuracy was set as  $10^{-4}$ . In order to compare the solution performance of the proposed algorithm on the high-dimensional single objective optimization problems, the comparison was carried out with the other three algorithms: random sort method, homomorphic mapping method and Pareto intensity value evolution algorithm. And the algorithms were respectively denoted as RY, KM and ZW. After each operation, the performance test of the algorithm used the best result (Best), the worst result (Worst) and the average result (Mean) respectively, and it was compared with the optimal solution experimental data. Table 1 shows the correlation comparison result between new algorithm MP and algorithms KM, RY and ZW. The results are listed in Table 1.

It can be seen from Table 1 that for the question  $g_1$ , the optimal solution comparison obtained by MP algorithm and ZW algorithm was relatively close, which was close to the real optimal solution, furthermore, it was better than the optimal solution RY. On the mean value and worst solution such two tests, the optimal solution was inferior to ZW, but better than RY. Then, in each performance test, question  $g_2$  and  $g_4$  were better than the other three algorithms, moreover, the proposed MP algorithm found the optimal solution for 13 times in 20 times of the independent operation. While for the question  $g_3$ , the MP algorithm was inferior to the ZW algorithm in the three performance tests, but better than other two algorithms KM and RY. In question  $g_5$ , although the optimal solution obtained by MP was inferior to ZW, the gap between the solutions was smaller, which was close to the true solution. In addition, MP was superior to the other three algorithms in mean value and worst solution performance. And the reason was that in 20 independent operations, MP found the optimal solution for 16 times.

In the second group of tests, in order to verify whether the proposed algorithm MP can handle two-dimensional multi-objective optimization problems, it was often necessary to design some test functions to evaluate the algorithm. Then, the performance measure standard was used, and two kinds of performance evaluation standards between different algorithms were given, which was respectively the convergence: the convergence of the algorithm could be measured through the actually obtained non-inferior optimal target domain and the minimum distance average value between the theoretically non-inferior optimal target domain; diversity: diversity was used to describe the spread coverage between non-inferior solutions in a population. The performance measure  $M_1$  was used to evaluate the performance of the non-inferior solution comparison in the Pareto candidate solution set of use

algorithm in  $M_1$  under the same conditions. The performance measure data of the algorithm are given in Tables 2–4.

Table 1. Correlation comparison between new algorithm MP and algorithm KM, RY and ZW

Problems		$g_1$	$g_2$	$g_3$	$g_4$	$g_5$
Optimal solution		5126.49811	0.0539498	680.6300573	7049.3307	23.3062091
Best solution	MP	5126.498162	0.053949821	680.631127	7049.33122	24.30630327
	KM	-	0.054	680.91	7147.9	24.620
	RY	5126.4965	0.053957	680.630	7054.316	24.307
	ZW	5126.49811	0.053949831	680.6300573	7049.2480205	24.306209068
Mean solution	MP	5126.52765	0.053940235	680.6312233	7050.153672	24.315356306
	KM	-	0.064	681.16	8163.6	24.826
	RY	512.881	0.057006	680.656	7559.192	24.374
	ZW	5126.52654	0.053950257	680.6300573	7051.2874292	24.325487652
Worst solution	MP	5139.2522	0.0539677331	680.63123678	7055.233735	24.350563206
	KM	-	0.557	683.18	8659.3	25.069
	RY	5124.472	0.216915	680.763	8835.665	24.642
	ZW	5127.15641	0.053972292	680.6300573	7058.2353585	24.362999860

Table 2. Convergence performance measure  $\gamma$

Algorithm	SCH	ZDT1	ZDT2	ZDT3	ZDT4	ZDT6
MP	0.003401	0.001079	0.000856	0.001189	0.001137	0.320625
	0	0.000101	0.000056	0.000069	0.000153	0.042030
NSGA-II	0.003389	0.033480	0.072389	0.114499	0.513053	0.296566
	0	0.004751	0.031688	0.007938	0.118460	0.013138
SPEA	0.003403	0.001799	0.001339	0.047520	7.340299	0.221138
	0	0.000001	0	0.000049	6.572416	0.000448

Table 3. Diversity performance measure  $\Delta$

Algorithm	SCH	ZDT1	ZDT2	ZDT3	ZDT4	ZDT6
MP	0.566835	0.325136	0.328062	0.408365	0.358806	0.750632
	0.0236635	0.001460	0.001406	0.002352	0.016002	0.005562
NSGA-II	0.0477899	0.0390307	0.430776	0.738540	0.702612	0.668025
	0.003471	0.001876	0.004721	0.019706	0.064648	0.009923
SPEA	1.021110	0.784525	0.755148	0.672938	0.798463	0.849389
	0.004372	0.004440	0.004521	0.003587	0.014616	0.003916

Table 4. Comparison of the performance measures  $M_1$  and SP

Algorithm	SCH	ZDT1	ZDT2	ZDT3	ZDT4	ZDT6
MP	0.002635	0.002312	0.002640	0.001355	0.039862	0.040362
	0.032651	0.002682	0.003919	0.001561	0.402513	0.056230
NSGA-II	0.006321	0.002906	0.003368	0.0015626	0.043115	0.036250
	0.010265	0.005613	0.007227	0.0089013	0.336522	0.033621
SPEA	0.009632	0.008653	0.140322	0.0026354	0.044752	0.040356
	0.010263	0.006739	0.010654	0.02956665	0.35032	0.0625103

As can be seen from Table 2, on the SCH, the convergence of the proposed algorithm MP was closer to the other two algorithms. In ZDT6, it was worse than the other two algorithms. The convergence of ZDT1 to ZDT4 indicated that the algorithm in this paper was superior to the other two algorithms. The data of diversity performance measures  $\Delta$  in Table 3 showed that the diversity of MP in the SCH solution was inferior to the other two algorithms, but superior to the other two algorithms on ZDT1 to ZDT4. Then, the diversity of the three algorithms' solution on ZDT6 was closer. Besides, in the performance measurements on  $M_1$  and SP, Table 4 showed that for the non-inferior performance difference generated under the same conditions and the algorithm performed the same number of target calculation each time, the performance measure of MP in SCH, ZDT1 to ZDT3 was better than other two algorithms, on ZDT4 and ZDT6, the performance measures of the three algorithms were similar.

In this paper, from above experimental results and performance analysis, it can be seen that under the single-dimensional and two-dimensional situation, the algorithm in this paper performs better, and the Pareto front obtained by the algorithm is relatively close to the real Pareto front. On the basis of the proposed algorithm MP, the improvement is carried out by combining with the characteristics of practical engineering applications, which will have a good engineering application prospect.

## 5. Conclusion

Although the intelligent optimization method has achieved many remarkable results, the combination of it with the specific practice areas still has many problems to be solved. In this paper, the popular hot spot methods in the field of intelligent optimization, the genetic algorithm and particle swarm optimization algorithm, were studied from the perspective of bionics, then combined with the biology basis of the two methods, some in-depth researched were carried out and some achievements were obtained: for the multi-objective optimization problem, the method of maintaining the population diversity by traditional evolutionary algorithm mainly depends on the shared function, however, its niche radius is difficult to set effectively. Then, the neighborhood function criterion can be introduced to the selection process, so as to select good individuals from the population and ensure the diversity of the population. In addition, a candidate set maintenance method based on neighborhood function criterion can be integrated into the new algorithm, and the use of this

method can effectively maintain the diversity of individuals in candidate solution sets. Then, the proposed algorithm is analyzed theoretically from time and space complexity. The test of a group of typical optimization problem shows that the proposed algorithm has relatively high search performance, and the diversity and convergence of solution set distribution are ideal. Of course, there are many places in this paper that need further study, such as an improved algorithm for nonlinear programming problems.

## References

- [1] Y. WANG, Z. CAI, G. GUO, Y. ZHOU: *Multiobjective optimization and hybrid evolutionary algorithm to solve constrained optimization problems*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 37 (2007), No. 3, 560–575.
- [2] M. KANAKUBO, M. HAQIWARA: *Parameter-free genetic algorithm using pseudo-simplex method*. Systems and Computers in Japan 36 (2005), No. 13, 35–44.
- [3] T. LI, D. W. CUI: *Genetic algorithm selection and parameter setting based on rule induction*. Computer Engineering 36 (2010), No. 3, 218–220.
- [4] M. S. GIBBS, H. R. MAIER, G. C. DANDY: *Comparison of genetic algorithm parameter setting methods for chlorine injection optimization*. Journal of Water Resources Planning and Management 136 (2010), No. 2, 288–291.
- [5] C. C. Y. DOREA, J. A. GUERRA JR., R. MORGADO, A. G. C. PEREIRA: *Multistage Markov chain modeling of the genetic algorithm and convergence results*. Numerical Functional Analysis and Optimization 31 (2010), No. 2, 164–171.
- [6] J. F. WANG, B. Q. DU, H. M. DING: *A genetic algorithm for the flexible job-shop scheduling problem*. Advanced Research on Computer Science and Information Engineering. Proc. International Conference, CSIE 2011, May 21–22, 2011, Zhengzhou, China, Part I, 332–339.
- [7] Y. PENG, X. LUO, W. WEI: *A new fuzzy adaptive simulated annealing genetic algorithm and its convergence analysis and convergence rate estimation*. International Journal of Control, Automation and Systems 12 (2014), No. 3, 670–679.
- [8] F. HERRERA, M. LOZANO, J. L. VERDEGAY: *Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis*. Artificial Intelligence Review 12 (1998), No. 4, 265–319.
- [9] F. Y. XIAO, P. K. LI: *An exact schema theorem for adaptive genetic algorithm and its application to machine cell formation*. Expert Systems with Applications 38 (2011), No. 7, 8538–8552.
- [10] H. MIN, H. J. KO, C. S. KO: *A genetic algorithm approach to developing the multi-echelon reverse logistics network for product returns*. Omega 34, (2006), No. 1, 56–69.
- [11] C. S. ZHANG, D. T. OUYANG, N. YUE, Y. G. ZANG: *A hybrid algorithm based on genetic algorithm and Levenberg-Marquardt*. Journal of Jilin University(Science Edition) 46 (2008), No. 4, 675–680.
- [12] S. ZHUO, C. LIU, K. Q HE: *A software pattern of the genetic algorithm*. Wuhan University Journal of Natural Sciences 6 (2001), Nos. 1–2, 209–217.
- [13] K. DEEP, K. N. DAS: *Quadratic approximation based hybrid genetic algorithm for function optimization*. Applied Mathematics and Computation 203 (2008), No. 1, 86 to 98.
- [14] D. H. KIM, A. ABRAHAM: *Hybrid genetic algorithm and bacterial foraging approach for global optimization and robust tuning of PID controller with disturbance rejection*. Hybrid Evolutionary Algorithms, Part: Studies in Computational Intelligence, Springer-Verlag Berlin Heidelberg 75 (2007).
- [15] K. VAISAKH, L. R. SRINIVAS: *Evolving ant colony optimization based unit commitment*. Applied Soft Computing 11 (2011), No. 2, 2863–2870.

Received July 12, 2017