CONTENTS

Acta Technica *62* No. 2A/2017, 1–14

Reverse logistics network for waste in ecotourism area based on quad tree classification

Liu Jing[1]

Abstract. With the rapid development of economy, people's awareness of environmental protection has also been improved, and the eco-tourism area has been put forward to protect natural resources. Local residents and tourists produce waste, which affects the local environment and has a major impact on the disposal of these wastes. Therefore, the solid wastes which are difficult to collect were selected in this paper; based on the idea of quad tree classification, the reverse logistics network of ecotourism waste area was designed, and relevant examples were also used to verify it; and then the appropriate address was selected to build the waste reverse logistics network. The results show that the network can locate correctly.

Key words. Environmental protection, ecotourism, quadtree, reverse logistics.

1. Introduction

With the development of the world economy, people's awareness of environmental protection is also increasing [1]. Eco tourism zone, based on natural resources, is a sustainable development that can utilize the landscape advantages of scenic spots to pursue the ecological environment, the basic purpose of which is to protect the ecological environment and realize the harmonious and balanced development between man and nature [2]. This kind of tourism means that all the pollution caused by tourism activities is limited within the scope of self-purification of natural ecology and sustainable development [3]. In order to maintain sustainable development, the pollution is controlled within a certain range according to the possible self-purification capacity of various pollutants and the environment [4]. Among them, the solid waste produced by travelers is one of the main pollutants, and the collection and disposal of solid waste in ecotourism area is one of the important means to control pollution [5]. However, the collection and management of solid waste are the most difficult and complex [6]. In the collection, transportation, treat-

[1]Hebei Vocational Art College, Shijiazhuang, Hebei, 050011, China; E-mail: liujing6835286@163.com

ment, and regeneration of solid waste management, about 60–70 % of the cost is spent on the collection, because the total cost of governance is large. If there is a slight improvement in collection, the total cost can be saved [7]. Therefore, in the design and research of waste reverse logistics network based on the idea of quad tree classification, the emphasis was on the collection of solid waste in this paper.

2. State of the art

With the gradual increase of environmental awareness, domestic and foreign scholars have conducted a multi-party research on reverse logistics network design. For example, in 1989, some scholars proposed a multi-objective mixed integer programming model to select the mode of transportation of recycled goods [8]. In 1993, a multi-objective mixed integer programming model and a heuristic algorithm were proposed to solve the problem of location and allocation of municipal solid waste facility services [9]. In 1998, some scholars developed a network structure system for regenerating sand from building waste, and established a three-level location model with capacity constraints [10]. In 2000, some scholars used the establishment of multiple target models to study the waste management system in Taiwan. In 2003, some scholars proposed a common model for the recovery and reuse of waste items for a two-tier site model with only one type of facility [11]. The waste was collected from the user's area and was sent to the recycling center for processing as reusable material before entering the market. According to the quantity of recycled materials, compared with the market demand, the model larger than market demand and the model smaller than the market demand were put forward, and heuristic algorithm was designed. In 2007, some scholars studied the problem of collecting discarded products from consumers and setting up recycling centers, and proposed a tabu search algorithm to solve them [12].

3. Methodology

Among the methods of operation of various waste recycling logistics systems, the more advanced method is to use the transfer station to separate waste collection and transportation, and to collect waste with a collection vehicle with a compression device and transport the waste to a treatment plant using a large capacity vehicle [13]. The most important thing in this model is the selection of the relay station and the planning of the vehicle path [14]. Most of the research at this stage is a straight line, and some scholars have studied the collection of waste on the route, taking into account the fact that the amount of waste collection is constant.

If the transfer station is built within the scope of the study, the points produced are divided, and the number of waste generated by each point is also known, after investigation, the optional site suitable for setting up transit stations can be selected [15]. At present, the following problems should be solved: (1) the number and location of the transfer station should be established; (2) the driving route of the vehicle should be collected, that is, site selection-path selection problem. Site selection -

path selection problem, it is generally considered that there are some known demand points and candidate facilities in the plane of the study, the location of the facility is selected in the candidate facility, and the route to the demand point is given at the same time. According to the requirements of the objective function, the tour path is determined, and usually the objective function is the minimization of cost.

Site selection - path selection problem, the key point is to establish a mixed location - path model with good number and location of facilities. Under certain constraints, the best transportation route can be selected, so that the total cost is lowest. The total cost generally includes fixed investment in facilities, operating costs, vehicle capacity and quantity limits. Site selection - path selection problem is generally divided into two algorithms: precision algorithm and heuristic algorithm.

The production of solid waste in the tourist area is caused by: (1) aboriginal residents, fixed production; (2) foreign tourists, uncertainty production. In this paper, it was assumed that the production of solid waste in tourist attractions was divided into three periods: season, off-season and flat season, and the method and route of garbage collection were considered.

In this paper, the disposal plant was not in the tourist area. In the selection of solid waste collection-routing problem, only the location routing problem from point of arrival to transfer station was considered. All vehicles started from the parking lot, passed through the transit station and returned to the parking lot, and all the vehicles were the same. Each vehicle was responsible for collecting only one route. The main objective was to minimize the cost of fixed investment and operating costs of the whole system.

The size of the transfer station was determined according to the amount of garbage transfer. The amount of litter transport should be based on the actual data of the daily output of the garbage in the service area of the tourism service area. Therefore, the transfer route should be changed according to the situation of the high season and off-season peace season.

The model must first meet the following conditions:

$$x_j = \begin{cases} 1, & \text{If you build a transfer station at } J \\ 0 & \text{otherwise} \end{cases}$$

$$y_{jk} = \begin{cases} 1, & \text{The vehicle } K \text{ passes through the transfer station } j \\ 0 & \text{otherwise} \end{cases}$$

$$z_{jk} = \begin{cases} 1, & \text{The vehicle } K \text{ goes straight to the node } j \text{ from a road node } i \\ 0 & \text{otherwise} \end{cases}$$

The best site selection model of solid waste collection in the tourist area is:

$$\text{Min} \sum_{j \in J} f_j x_j + \sum_{k \in K} \left[\frac{C}{2} \left(\sum_{j \in N} \sum_{i \in N} Z_{ijk} d_{ij} \right) \times \sum_{j \in N} \sum_{i \in I} Z_{ijk} h_i^I \right] . \tag{1}$$

$$\sum_{k\in K}\sum_{j\in N} Z_{ijk} = 1, \qquad (i \in I)\,, \tag{2}$$

$$\sum_{i\in N} Z_{ijk} - \sum_{i\in N} Z_{jik} = 0, \qquad (j \in N, k \in K)\,, \tag{3}$$

$$\sum_{i\in S}\sum_{j\in S}\sum_{k\in K} Z_{ijk} \le |S| - 1, \qquad 2 < \ |S| \le |I|\,, \tag{4}$$

$$y_{jk} \le x_j, \qquad (j \in J, k \in K)\,, \tag{5}$$

$$\sum_{j\in J} y_{jk} = 1, \tag{6}$$

$$\sum_{i\in I} h_i^I Z_{ijk} \le Q_{\mathrm{C}}, \qquad (j \in N,\ k \in K)\,, \tag{7}$$

$$\sum_{i\in I} Z_{ijk} = y_{jk}\,, \tag{8}$$

$$z_{j0k} = y_{jk}, \qquad (j \in J,\ k \in K)\,, \tag{9}$$

$$N_1 \le \sum_{j\in J} x_j \le N_2, \tag{10}$$

$$x_j \in \left\{\ 0, \quad 1\ \right\}, \qquad (j \in J)\,, \tag{11}$$

$$y_{jk} \in \left\{\ 0, \quad 1\ \right\}, \qquad (j \in J,\ k \in K)\,, \tag{12}$$

$$z_{ijk} \in \left\{\ 0, \quad 1\ \right\}, \qquad (i, j \in J,\ k \in K)\,. \tag{13}$$

Here, the parking lot is Q, the solid waste collection point is I, the collection point of the parking lot, the solid waste collection point and the transit station is N, the candidate site of the transit station is J, and the amount of solid waste produced at point i is $i \in I, \quad I = 1, 2, 3$ high season, shoulder season, low season.

The objective function (1) is the sum of the location cost of the transfer station and the transportation cost of the waste. The double sum $\sum_{j\in N}\sum_{i\in N} Z_{ijk} d_{ij}$ is the total distance on the kth road, while the double sum $\sum_{j\in N}\sum_{i\in I} Z_{ijk} h_i^I$ is the total amount of traffic on the kth road. For ease of calculation, it was assumed that the waste ran at half the total distance on the Kth circuit. Formula (2) means that each collection point only appeared on one road, and formula (3) represents the network stream conservation constraint. Formula (4) represents the standard branch elimination constraint, formula (5) means that only when the transfer station was built, the collection vehicle could pass through. Formula (6) means that a vehicle

only passed through the transfer station, formula (7) indicates the limit of vehicle loading capacity and formula (8) means that the vehicle passing through the transfer station must have a line from the collection point to the transfer station. Formula (9) shows that when the vehicle passed through the transfer station, the transfer station and the parking lot being connected and formulas (8) and (9) also ensured that the collection vehicle was returned to the parking lot after unloading at the transfer station. Formula (10) represents the number of transfer stations, while formulas (11), (12) and (13) give the decision variables.

Location–path selection problem was divided into 2 steps, namely, the partition of the collecting points and the determination of the route and location.

(1) Collection point partition. The method of the quad tree was used to select the collection points that were relatively concentrated at the geographic location, and the total production of the collection points in the candidate area was equal to or close to the total capacity of the vehicle.

(2) Location and route determination. Heuristic algorithms were used to locate and determine the routing routes for each partition.

Quad tree can continuously divide the geographical space into four parts, so as to form a grid of $2^k \times 2^k$, which can better describe the relationship between point and line in the geographical space. The quad tree method has the advantages of quick and concise technology. The quad tree classification can solve the problem of location partitioning. Therefore, in this paper, the quad tree principle was used to determine the location and the candidate partition.

The idea of the quad tree is to divide the geographical area into four parts. The four regions are regarded as the first layer of the division, and then, the four branches of the first layer are divided into four equal branches until the area satisfies the set condition. Specific steps are as follows:

Firstly, the coordinate system is established, the parking area is the origin, and the relative coordinates $v_m = (a_m,\ b_m)$ are used at the point of need.

Secondly, quad tree is generated, the maximum network length of quad tree is established with the maximum absolute value of a_m and b_m. The criterion of generating quad tree is that the amount of each partition is less than the market demand. If the amount of the partition is greater than the market demand, it will be split until the amount of each partition is less than the market demand.

Thirdly, the candidate partitions are merged and formed, merged from the bottom layer, and the principle is that their upper layers belong to the same branch. The minimum distance from the last two points in the grid is merged, and the distance between the two points in the two grids cannot exceed the distance from the nearest point to the transit station and ensure that the collection of the grid after the merger does not exceed the market demand, and then followed by the merger, the lower that cannot participate in the merger can participate in the merger of the upper, the merging principle is the same. The mesh is as minimal as possible, and a grid of candidate points is removed to form a candidate partition.

Tour path and location: after the partition becomes a small TSP problem, the lingo8.0 (related interface, shown in Fig. 1) is adopted to solve the routing and distance of the partitions in the peak season to different candidate stations, and

the full enumeration method is used to find the objective function (1) based on the results, so as to achieve the optimal number of transfer stations and distribution of transfer stations. During off-season and peaceful seasons, the transfer station which is determined by the peak season can be used to change the zoning and tour routes.

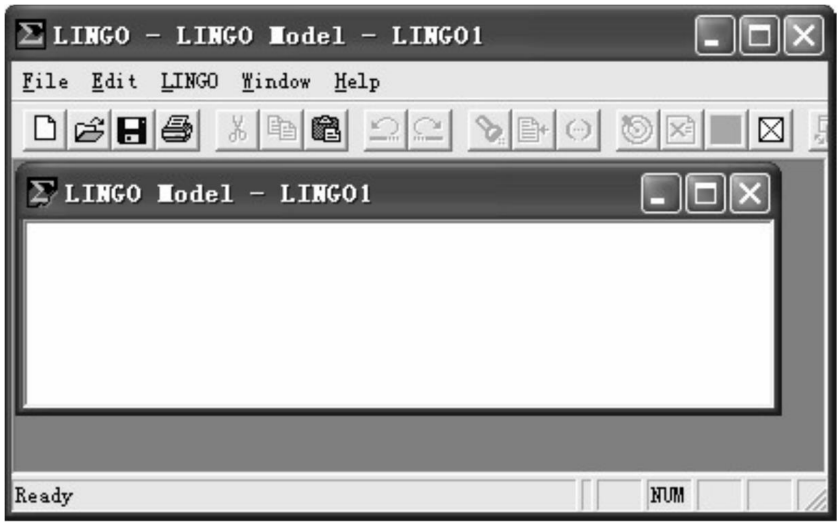


Fig. 1. Lingo8.0 operation interface

There were 5 candidate transfer stations and station costs, as shown in Table 1. The amount of waste generated at 20 different waste generation points and during the unified collection cycle is shown in Table 2. The quantity of resident location was unchanged, and the location of tourism was the point of change. The collection cycle was five days. 3 months were in the busy season. The cost per ton of waste was 11 yuan, 100 meters per kilometer. The cost and transportation cost of the transfer station were converted to the monthly cost calculation. Transportation costs were calculated according to the peak season.

Table 1. Transit station candidate point location and construction costs

Transfer station number	Abscissa system	Vertical coordinate system	Cost (million yuans)
1	11.76	14.06	55
2	4.39	3.98	57
3	13.63	8.58	47
4	7.59	6.09	50
5	16.64	3.79	51

From the above table, it can be seen that the calculation result was 1, 2 and 4, and the establishment of 3 transfer stations made the cost reach the optimum.

Table 2. Collection point coordinates and collection in different situations

Collection point number	Abscissa	Ordinate	Collecting in different cases (kg)
1	19.01	1.16	4300 4300 4300
2	4.62	7.06	3100 1200 1000
3	12.14	16.26	2900 1100 600
4	9.72	0.20	2200 2200 2200
5	17.83	2.78	1900 1900 1900
6	15.24	4.06	2900 1000 800
7	9.13	3.97	3700 3700 3700
8	7.37	9.08	4000 2000 600
9	16.43	5.44	3700 800 600
10	2.50	1.98	3000 1000 600
11	12.31	10.31	2000 2000 2000
12	15.84	8.90	2100 2100 2100
13	14.76	18.63	2600 900 400
14	3.53	9.32	2700 1000 300
15	3.53	9.32	2700 1000 300
16	8.11	8.37	3200 3200 3200
17	18.71	16.92	3400 800 400
18	18.34	10.50	3400 3400 3400
19	6.21	4.05	2900 2900 2900
20	17.87	13.44	1200 1200 1200

4. Result analysis and discussion

On the basis of the above, the amount of solid waste produced in the tourist area was changed into the reverse logistics network of solid waste recycling in ecotourism area. The establishment of reverse logistics network was based on the relevant policies and regulations of our country. According to the actual situation of our country, the teaching model of waste network design was established, which referred to the total capacity limit of the transfer station, and the waste was undergone the process from the point of production to the transfer station and then to the processing plant. If the points are divided, and the number of wastes generated by each generating point is identified, firstly, it is necessary to find out the candidate sites which are suitable for setting up processing stations, and then deal with the location of stations so as to minimize the negative impacts on the residents. And

then under the condition of satisfying the total capacity limit, the freight of the transfer station should be the lowest.

The site selection is the first step in the reverse network engineering of solid waste. Its rationality affects the project cost, the operation stability and the secondary pollution control, which mainly follows the following principles: 1. the principle of safety in the view of pollution prevention; 2. the economic and reasonable principle considered from the economic point of view. Safety principle is the basic principle of processing station location. The economic principle requires that the best economic results can be achieved by using reasonable technical and economic schemes and minimizing investment, so as to achieve the goal of environmental protection. It should meet the following requirements:

(1) The site should meet the overall planning of the local construction.

(2) There are appropriate geological conditions.

(3) There are appropriate natural conditions.

(4) The impact on the environment is small.

(5) The economic costs are as low as possible, but achieve the best effect.

(6) The site selection should be in line with national policies, laws and regulations, as well as the consent of the majority of residents.

There are many solid waste generation points in the ecotourism area, including local residents' residences and various attractions. These points will be set near the transfer station, the production of solid waste will be placed near the transfer station, while establishing a processing station, its main function is to transfer the solid waste from the transfer station to classify and compress the waste. The processing station may affect the surrounding residents and tourists in the scenic area, resulting in discomfort. As a result, residents and tourists will want to stay away from them. In the site selection, both the construction cost and the wishes of the residents and tourists should be considered. Under the premise of meeting the demand, in order to determine the number and location of establish transit station, in this paper, it assumed that the AHP method was used to select candidate treatment station, and the processing station location was determined, so as to maximize the minimum distance between processing stations to residential areas and scenic spots, and make the construction site cost, operation cost and transportation cost lowest.

The traveling season was divided into the peak season, the low season and the peaceful season. In different circumstances, the number of tourists changed, so that the amount of solid waste varied. The scale of the transfer station should be determined based on the amount of refuse transfer, and the amount of refuse transfer should be determined according to the actual quantity of daily average production of garbage in the service area. Thus, the number of transfer stations was established on the basis of the peak season, while in the off-season or flat season, some completed transit stations may be closed.

The model contained the following assumptions:

(1) Waste can only be transported directly to the transfer station, transported by the transfer station to the processing station, but cannot be sent directly to the processing station.

(2) The network discrete location defined the alternative relay stations and pro-

cessing stations.

(3) The distance and waste transportation costs were simply linear.

(4) The processing station and a number of transit stations were established in the study area.

(5) If the peak season, off-season and peace season each had 3, 3 and 6 months, then $a = \{\ 3 \quad 3 \quad 6\ \}$, and at the same time, the following conditions were met:

$$x_j = \begin{cases} 1, & \text{If you build a transfer station at } J \\ 0 & \text{otherwise,} \end{cases}$$

$$y_{jk} = \begin{cases} 1, & \text{If you build a transfer station at } k \\ 0 & \text{otherwise,} \end{cases}$$

$$z_{ij} = \begin{cases} 1, & I\text{th point of waste to the } J\text{th point transfer station} \\ 0 & \text{otherwise.} \end{cases}$$

A model where the amount of solid waste is constant was:

$$\min \sum_j f_j x_j + \sum_i \sum_j a_l p_j x_j +$$

$$+ \sum_k G_k y_k + C \left(\sum_i \sum_j \sum_k h_i^l z_{ij} y_k d_{jk} + \sum_l \sum_j h_i^l z_{ij} d_{ij} \right), \tag{14}$$

$$\min p\,, \tag{15}$$

$$\sum_j z_{ij} = 1 \qquad \forall i \in I\,, \tag{16}$$

$$z_{ij} < x_j \qquad \forall i \in I, \quad \forall j \in J\,, \tag{17}$$

$$d_{ik} y_k \geq P \qquad \forall i \in I, \quad k \in K\,, \tag{18}$$

$$\sum_i h_i^l z_{ij} \leq Q_j\,, \tag{19}$$

$$\sum_k y_k = 1\,, \tag{20}$$

$$x_j, y_k, z_{ij} \in \{0, 1\} \qquad \forall i \in I,\ j \in J, k \in K\,. \tag{21}$$

The objective function (14) represented the minimization of construction cost, operation cost and transportation cost. The objective function (15) represented the

maximum distance from the point at which the waste was generated to the nearest processing station, and was defined in the constraint (18). Condition (16) ensured that each point was only shipped to a transfer station and (17) ensured that only in the construction of jth transfer station, waste in i can be shipped to$j$. The condition (19) represents the total capacity limit of the transfer station, (20) only built a processing station and (21) indicates that the decision variable was 0 or 1 for the variable.

There were four processing stations, five transit station candidate sites and construction costs, as shown in Table 3. Fig. 2 shows the amount of waste generated at 20 different waste generation points during a collection cycle. SPSS20.0 software (the operation interface shown in Fig. 3) was used for analysis and finishing, and the results are shown in Table 4, of which, 5 days was a collection cycle. The freight cost per ton was 100 yuans per kilometer, representing 1 kilometer per unit distance. A reasonable distance was 4.1 kilometers, $\delta_p = 0.5$. The cost per month included reasonable costs, station and transfer costs of processing station and transfer station, so the reasonable cost was converted into reasonable cost per month, $T = 65\delta_T = 6.5$.

Table 3. Processing station candidate point location and station cost

Processing station number	Abscissa system	Vertical coordinate system	Cost (million yuans)
1	8.12	14.06	200
2	12.09	5.89	190
3	5.93	13.88	215
4	15.03	13.88	215

When the maximum satisfaction degree of the objective function was 85 %, the processing station was set at (8.12, 14.05) with number of 1, and the transfer station was established at (11.76, 14.06), (7.59, 6.09) and (16.63, 3.79).

5. Conclusion

The main purpose of the eco-tourism area is to use the advantages of the scenic area to minimize the damage to the environment to carry out tourism. In recent years, with the economic improvement, people's awareness of environmental protection has gradually increased. Although ecotourism has reduced pollution to the environment as much as possible, local residents and tourists still produce waste. How to deal with waste has an important impact on environmental protection, of which, the collection and management of solid waste are the most complex. Thus, the solid waste was selected as the object of study, the present situation of waste reverse logistics network was briefly introduced. Based on the idea of quad tree classification, the optimal site selection model of solid waste collection in tourism area was established and a model for the constant production of solid waste was established. And then the research on the reverse logistics network design for waste

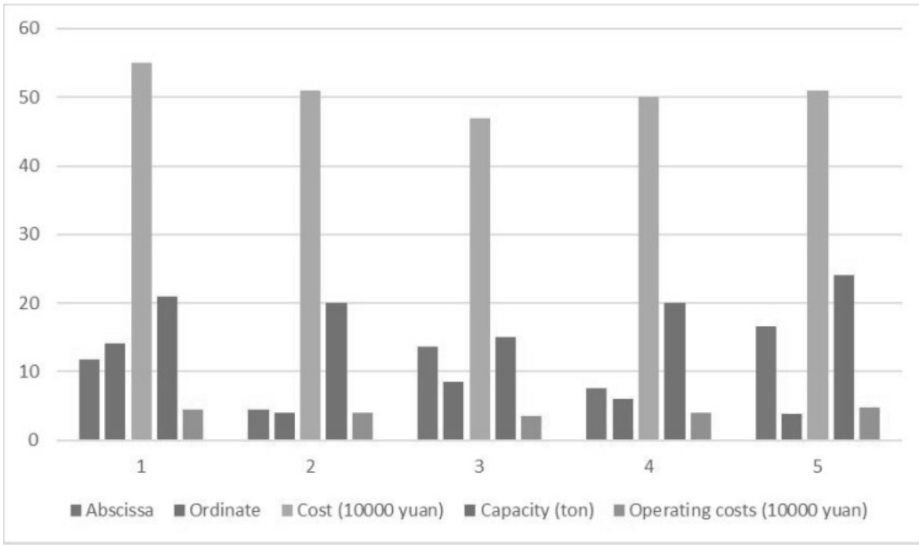


Fig. 2. Location of transit point, station cost and monthly operating expenses

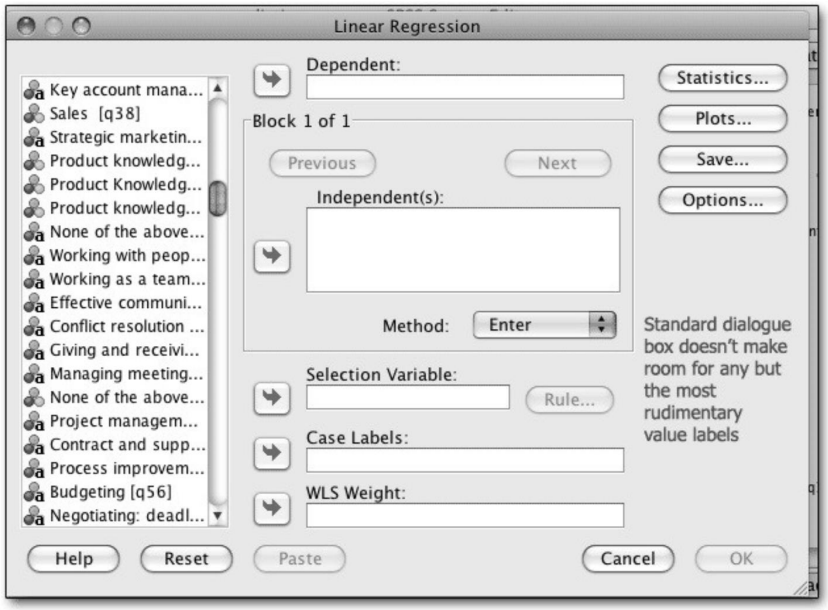


Fig. 3. SPSS 20.0 operation interface

in ecotourism area based on quad tree classification was carried out and verified by the relevant examples. The results show that the network can be located correctly.

Table 4. Collection point coordinates and collection in different cases

Collection point number	Abscissa	Ordinate	H_i^l (kg), $(i = 1, 2, 3$	δ_i (kg)
1	19.01	1.16	4300 4300 4300	500
2	4.62	7.06	3100 1200 1000	200
3	12.14	16.26	2900 1100 600	200
4	9.72	0.20	2200 2200 2200	300
5	17.83	2.78	1900 1900 1900	300
6	15.24	4.06	2900 1000 800	200
7	9.13	3.97	3700 3700 3700	400
8	7.37	9.08	4000 2000 600	300
9	16.43	5.44	3700 800 600	200
10	2.50	1.98	3000 1000 600	200
11	12.31	10.31	2000 2000 2000	300
12	15.84	8.90	2100 2100 2100	200
13	14.76	18.63	2600 900 400	300
14	3.53	9.32	2700 1000 300	200
15	3.53	9.32	2700 1000 300	200
16	8.11	8.37	3200 3200 3200	400
17	18.71	16.92	3400 800 400	200
18	18.34	10.50	3400 3400 3400	400
19	6.21	4.05	2900 2900 2900	300
20	17.87	13.44	1200 1200 1200	200

References

[1] M. Heydari, H. Poorbabaei, M. Bazgir, A. Salehi, J. Eshaghirad: *Earthworms as indicators for different forest management types and human disturbance in Ilam oak forest, Iran.* Folia Forestalia Polonica *56* (2014), No. 3, 121–134.

[2] A. Eraydin, B. Armatli-Köroglu, C. N. Uzun: *Importance of social capital in coping with and benefiting from new economic conditions.* Tijdschrift voor economische en sociale geografie, Journal of Economic and Social Geography *103* (2012), No. 2, 222–239.

[3] Y. T. H. Chiu, W. I. Lee, T. H. Chen: *Environmentally responsible behavior in ecotourism: Antecedents and implications.* Tourism Management *40* (2014), 321–329.

[4] T. H. Lee, F. H. Jan, G. W. Huang: *The influence of recreation experiences on en-*

vironmentally responsible behavior: The case of Liuqiu Island, Taiwan. Journal of Sustainable Tourism *23* (2015), No. 6, 947–967.

[5] D. Datta, S. Banerji: *Local tourism initiative in an eastern Himalayan village: Sustainable ecotourism or small-scale nature exploitation?.* Bulletin of Geography. Socio-economic Series (2015), No. 27, 33–49.

[6] L. A. Guerrero, G. Maas, W. Hogland: *Solid waste management challenges for cities in developing countries.* Waste Management *33* (2013), No. 1, 220–232.

[7] M. S. Pishvaee, R. Z. Farahani, W. Dullaert: *A memetic algorithm for bi-objective integrated forward/reverse logistics network design.* Computers & Operations Research *37* (20010), No. 6, 1100–1112.

[8] W. Hare, Y. Lucet, F. Rahman: *A mixed-integer linear programming model to optimize the vertical alignment considering blocks and side-slopes in road construction.* European Journal of Operational Research *241* (2015), No. 3, 631–641.

[9] M. S. Pishvaei, F. R. Zanjirani: *A memetic algorithm for integrated forward/reverse logistics network design in a supply chain.* Physical Review Letters *87* (2009), No. 25, paper 257601.

[10] P. Luathep, A. Sumalee, W. H. K. Lam, Z. C. Li, H. K. Lo: *Global optimization method for mixed transportation network design problem: A mixed-integer linear programming approach.* Transportation Research Part B: Methodological *45*, (2011), No. 5, 808–827.

[11] R. Archer, G. Nates, S. Donovan, H. Waterer: *Wind turbine interference in a wind farm layout optimization mixed integer linear programming model.* Wind Engineering *35* (2011), No. 2, 165–175.

[12] B. Yao, P. Hu, M. Zhang: *A support vector machine with the tabu search algorithm for freeway incident detection.* International Journal of Applied Mathematics and Computer Science *24* (2014), No. 2, 397–404.

[13] S. A. Naghibi, H. R. Pourghasemi, B. Dixon: *GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran.* Environmental Monitoring and Assessment *188* (2016), No. 1, paper 44.

[14] A. M. Youssef, H. R. Pourghasemi, Z. S. Pourtaghi, M. A. Al-Katheeri: *Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia.* Landslides *13* (2016), No. 5, 839–856.

[15] S. A. Naghibi, H. R. Pourghasemi: *A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping.* Water Resources Management *29* (2015), No. 4, 5217–5236.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 15–26

Application of network model in coal mine safety management strategy[1]

Gao Xiaoxu[2,3,4], Shao Guangtong[2,3]

Abstract. In order to study the strategy of coal mine safety management and the application of network model in management, in this paper, through consulting a large number of documents and expert interviews, and combining genetic algorithm and fuzzy neural network model, the deep cause of coal mine safety accident was analyzed, the cause model of coal mine safety production was constructed, and the neural network was optimized. The results showed that the main factors of safety production in coal mine enterprises mainly included 5 aspects: management system, man, machine, environment, and information. And the dynamic monitoring of security management and the better convergence speed of network and the generalization mapping ability of network were guaranteed. A coal mine early-warning model was established, laying a foundation for the application and extension of fuzzy neural network in coal mine safety management.

Key words. Fuzzy neural network, security management, early warning management, genetic algorithm.

1. Introduction

Coal industry is the basic industry of our country and plays an important role in the development of national economy. By 2015, coal resources accounted for 65 % of China's energy consumption. And it is expected to reach 50 % by 2025. Coal is the main energy consumption in our country and makes great contribution to the economic construction of our country for a long time. However, safety production of coal mine has always been a major problem. In 2015, 1863 coal mines in China were closed because they did not have safe production conditions, and the coal mine accidents happened frequently and caused huge economic losses and casualties. In recent years, China has formulated and promulgated a series of safety measures and

[1]The study was supported by National Natural Science Foundation of China (51504183).

[2]School of Energy Science and Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi, 710054, China

[3]Key Laboratory of Western Mine & Hazard Prevention, China Ministry of Education, Xi'an, Shaanxi, 710054, China

[4]Corresponding author

policies to strengthen the safety management of coal mines, and the domestic safety of coal mine has been improved, but the situation has been unoptimistic. Therefore, the coal enterprises should more fully utilize modern technology and combine with computer technology and network communication management to accelerate the upgrading of coal mine safety management, reduce and avoid the occurrence of coal mine accidents, and reduce economic losses and personnel casualties.

2. State of the art

Safety evaluation refers to the method of systematic engineering to evaluate the possible risks and the consequences, and formulate corresponding measures to ensure the safety of the system according to the size of the predicted risk. The most widely used method of safety assessment in China is fuzzy comprehensive evaluation method [1]. In addition, early warning management should be carried out. In foreign countries, foreign scholars have begun to carry out early warning management research at the beginning of the 20th century. Early warning management is originated in the economic field, but the study has found that early warning management is applicable to not only the economic field, but also other areas [2]. The United States first introduced the theory of successful early warning management into the micro field, and then introduced it into the emergency management of enterprises. Since then, countries such as the United States, Japan and Russia have carried out more in-depth research and application of early-warning management theory. In China, the application of early warning management theory is mainly based on the theory of economic early warning, and other industries have relatively little initial application [3]. First of all, the theory of early-warning management is applied to the macro-economic field, and then it is slowly extended to various industries and goes deep into the micro field of the industry. In recent years, early warning management theory has been widely used in various sectors outside the economic sector, and mainly in high-risk industries, such as aviation disasters and early warning of coal mine safety. Scholar Xie Kefan has put forward the theory of "survival risk of enterprises", and enterprise crisis and risk warning have been systematically studied. Hu Huaxia has studied the enterprise forewarning management system from the angle of enterprise's survival crisis. Wang Shuai has put forward the system reengineering of coal mine safety management based on early warning management theory [4]. Li Wen has proposed a control and avoidance process based on the risk of coal mine development, and formulated a concrete method for the assessment and early warning of coal mine development risk [5]. Zhang Haifeng has proposed in his research that there are four main parts of the safety early-warning analysis of coal mine development, namely monitoring, identification, diagnosis and evaluation, and constructed the forecasting system [6].

3. Methodology

With the development of China's economy, China's various industries and enterprises have made great progress and development degree, and the demand for coal and other industries has been increasing with the development of the industry rising speed. China's coal mining has also made some progress, coal mine safety issues have further become one of the important links of China's mining industry [7]. In recent years, China's coal industry has developed to a certain extent. China's major coal mining industry distribution areas are shown in figure 1. The development and progress of China's coal industry have made more coal resources gradually mined and excavated, and have further provided a certain energy support for the development of other industries and provided a certain positive impact for the development of China's industry [8]. However, there are some defects and problems in the coal mine safety management in China; so some security problems appear constantly in the process of coal mine industry development, which has some negative influences on the development of the coal industry; and the emergence of certain insecurity factors makes the process of industry development and the country's sustained stability have a certain constraint [9]. However, the development of computer technology has played a technical support role for coal mine safety management, especially the emergence of network model has made the safety management of coal mine get sustained development, and has had certain positive influence and promotion function to enterprise development [10].

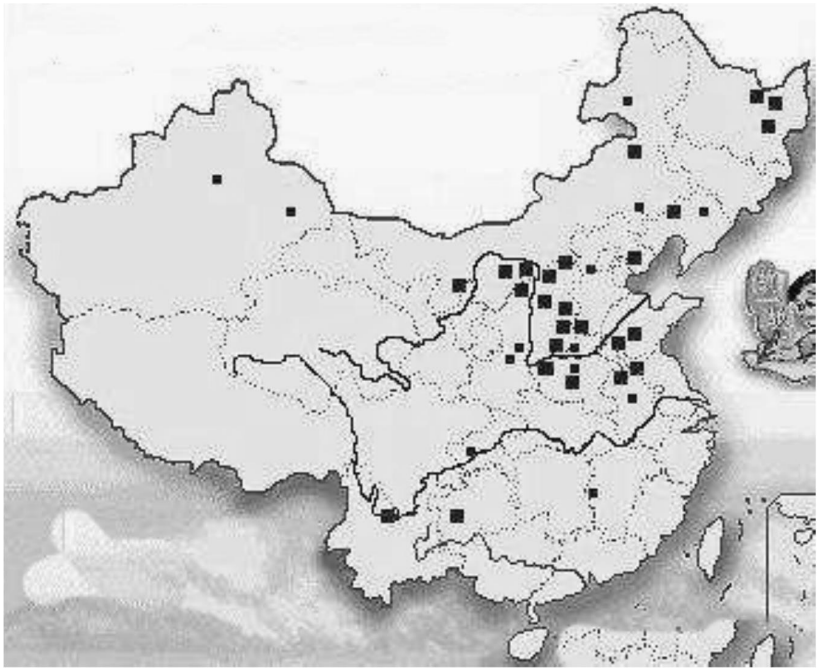


Fig. 1. Distribution of coal mines in China

(1) Information entropy method. The present situation of coal mine safety production management was taken as the starting point, the qualitative and comprehensive analysis of the main influencing factors in the process of safety production was carried out, 24 influencing factors were obtained through comprehensive analysis and qualitative analysis, among which the relevant influencing factors were summarized as shown in Table 1. Then the quantitative analysis method was used to optimize the index, and the scientific evaluation system was constructed.

Table 1. Main evaluation indicators for the impact of coal mine safety production management

Index of the first level	Index of the second level	Index of the first level	Index of the second level
Human factor	The three violation rates of employee	Management factor	Perfect rate of safety system
	Monthly average training time of employees		Perfect rate of safety measures
	Average educational level		Management efficiency
	Average age		Timeliness of management
	Average length of service		Perfect rate of emergency mechanism
	Proportion of migrant workers		Improvement rate of coal mine access management
Device factor	Improvement of the reliability of equipment	Environmental factor	The outstanding times of tons of gas
	Reliable consideration rate of transport equipment		Dip angle of coal seams
	Reliability of power supply equipment		The normal water inflow of mine
	Reliability of mechanical and electrical equipment		Maximum dust concentration on the face
	Reliability of ventilation equipment		Spontaneous combustion period of coal seam
	Reliability of drainage equipment		Reliability of top and bottom plate
	Improvement rate of water proof and fire-fighting facilities		
Information factor	Informatization degree		
	Information identification		
	Ability of processing		

Through the information entropy method, each coal mine safety evaluation index was selected. The relevant model is

$$e(i) = -k \sum_{j=1}^{m} \frac{x_{ij}}{E_i} \ln \frac{x_{ij}}{E_i} \,. \tag{1}$$

Here, E_i is the sum of the main indexes of coal mine safety evaluation in this study, k represents the coefficient of computation (usually 0.2), m is the total number of indicators, and x_{ij} is the evaluation index for all the research.

The overall entropy of the relevant evaluation matrix is

$$E^{'} = \sum_{i=1}^{n} e(i) \,, \tag{2}$$

where $e(i)$ is the evaluation value of all indexes and n represents the total quantity of all assessment indicators.

The measure of all the evaluation indicators used for this study is

$$w_i = \frac{1}{N - E^{'}} [1 - e(i)], \ i = 1, 2, \ldots, N \,, \tag{3}$$

where N represents the number of all assessment indicators.

(2) Questionnaire: in this paper, a questionnaire survey was conducted among 20 employees of 200 coal mines with different specifications in different parts of China, the safety training situation of coal mine workers was investigated and analyzed, and the safety engineering technology and management staffing rate were statistically analyzed.

(3) Statistical analysis: in this study, a large amount of data was obtained through consulting relevant documents about coal mine safety management and conducting questionnaire survey among 20 employees of 200 coal mines with different specifications as well as the consultation with experts concerned. And the SPSS method was used to collect and analyze a large amount of data, the main influencing factors of coal mine safety accidents were determined according to the accurate and scientific analysis results, and application of the fuzzy neural network model provided the scientific theory safeguard in the coal mine safety management.

4. Result analysis and discussion

Coal is the main consumption energy in our country and contributes greatly to the economic construction of our country [11]. Especially since the reform and opening up of China, China's various industries and industries have been greatly improved, and the demand for energy has been also increasing, therefore, China's coal enterprises have been greatly developed, the number of coal has been increasing at the same time, and coal production has also been greatly developed, which has provided certain positive impetus for the development of our country's other

industries and even the whole country's comprehensive economic level directly or indirectly. However, coal mine safety production has always been a major problem. In 2015, 1863 coal mines in China were closed because they did not have safe production condition; coal mine accidents happened frequently and caused huge economic losses and casualties. Figure 2 shows the scene of rescue in the accident.

Fig. 2. Rescue after the gas explosion of coal mine in 2016

China has also developed and introduced a series of safety measures and policies to strengthen the safety management of coal mines, the domestic coal mine safety situation has improved, but the situation has been still unoptimistic and coal mine accidents have been frequent [12]. In 2016, there were 197 coal accidents, coal accidents killed more than 20,000 people during 2011 to 2016; an average of more than 3000 people died each year from coal mine accidents, with nearly 10 deaths a day on coal accidents. If the death of one person per death causes economic losses of 50,000 Yuan, China's annual direct losses caused by coal mine accidents reach 150 million Yuan, and the actual economic losses are much higher than this number. Table 2 shows a partial coal mine accident that counts more than 10 deaths from 2011 to 2016.

Coal mine accidents frequently occur, although the mortality rate of millions of tons of coal production in China is declined in recent years, there is still a big gap compared with some advanced coal mining countries. For example, in recent years, the United States has a mortality rate of millions of tons about 0.019, India is 0.45, and Russia is only 0.42. While the mortality rate of millions of tons in China has remained above 2.0 for a long time, and has maintained between 0.5-1.0 in recent years. Table 3 shows the number of coal mine accidents, the number of deaths and the death rate of millions of tons in China during the years from 2007 to 2016.

Based on the current status of coal mine safety, after consulting a large number of documents, research on the coal mine safety management of experts at home and abroad was comprehensively studied and a thorough study of coal mine safety production management was made, and the mechanism of coal mine accidents, the

development of early warning management system and fuzzy network model were mainly included to evaluate the safety of the system. Safety evaluation refers to the method of systematic engineering to evaluate the possible risks and the consequences, and formulate corresponding measures to ensure the safety of the system according to the size of the predicted risk.

Table 2. More than 10 deaths caused by recent coal mine accidents in China

Time	Name of the coal mine that occurs accident	Types of accident	Death toll
2011-10-04	Anping coal mine in Libo county, Qiannan city, Guizhou province	Coal and gas outburst	17
2011-10-11	Jindi coal mine in Dong county, Jixi city, Heilongjiang province	Permeable	13
2012-10-16	Hetian coal mine in Yaozhou district, Tongchuan city, Shaanxi province	Gas accident	11
2012-11-06	Fufa coal mine of Dashu town, Fengjie county, Chongqing province	Gas mine	13
2013-10-27	Jiulishan coal mine of Jiaozuo coking coal group in Jiaozuo city, Henan province	Coal and gas outburst	18
2014-11-10	Xialiuchong coal mine of Changjiang town, Hengyang county, Hunan province	Gas explosion	29
2014-11-03	Qianqiu coal mine of Yima coal group in Sanmenxia city, Henan province	Rock burst accident	10
2015-12-07	Liuyi coal mine of Sandu town, Zixing city, Binzhou county, Hunan province	Gas explosion	11
2015-11-18	Shizhuang coal mine of Shizong county, Qujing city, Yunnan province	Gas explosion	43
2016-08-21	Xiaojiawan coal mine of Panzhihua city, Sichuan province	Gas explosion	46
2016-09-02	Gaokeng coal mine of Pingxiang coal industry group in Jiangxi province	Gas explosion	13

Fuzzy theory and neural network technology have received extensive attention and thorough research in recent years [13]. The fuzzy system corresponds to the artificial neural network model (Fig. 3), the model is to simulate the structure of the human brain thinking mode, and it has a strong self-learning ability, high accuracy, but it cannot really handle the fuzzy information like the human brain, so the requirement of the sample is relatively high [14]. The fuzzy neural network requires a lot less samples, and can make better use of the expert's knowledge system, so that the whole reasoning process is easier to understand, and then it also has a long reasoning time and is vulnerable to manual intervention. Therefore, two theories were combined organically to obtain the complementary evaluation effect in this study [15].

Then, all the main impact index values were collected through the questionnaire survey of A, B, C, and D four coal mines of a province. The results are shown in Table 4. Further, the membership degree of each index value of each coal mine was analyzed by formula. The analysis result is shown in Table 5. The results show that the factor of equipment is the main influencing factor for each coal mine, therefore, only the equipment factor is constantly improving, the safe operation can be ensured in the process of using 5BP network model in coal mine safety management in coal industry, and a scientific support can be further provided to the development of other industries.

Table 3. Coal mine accidents and deaths in China from 2007 to 2016

Year	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Total number of accidents	3082	4344	4143	3641	3341	2945	2421	1954	1616	1201
Total death toll	5670	6995	6434	6027	5986	4746	3786	3214	2640	1973
Number of major accidents	75	65	58	49	58	38	28	38	24	21
Mortality rate of millions of tons	5.07	4.94	3.71	3.08	2.81	2.041	1.485	1.182	0.749	0.564

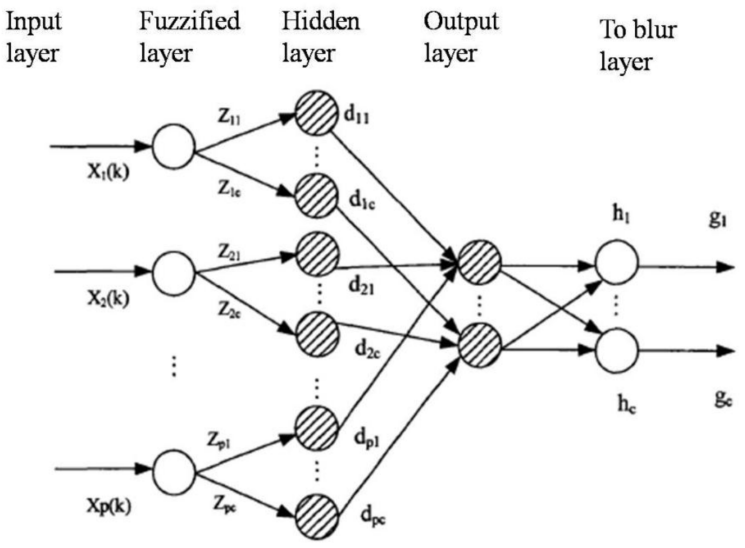


Fig. 3. Analysis of network model development

Table 4. Collection of raw data of A, B, C, D coal mines

Index of the first level	Index of the second level	A	B	C	D	E_i	e_i	w_i
Human factor	Rate of three violations of personnel	2.5 %	5.4 %	3.7 %	9.0 %	21 %	0.992	0.143
	Monthly average training time of employees	7.5	5.6	6.1	4.5	23.70	0.957	0.021
	Average educational level	2.8	1.9	2.1	1.7	8.50	0.931	0.024
	Average age	37.2	38.9	36.4	38.1	150.60	0.921	0.000
	Average length of service	5.4	4.2	6.1	5.0	20.70	0.988	0.011
	Proportion of migrant workers	66 %	72 %	75 %	67 %	280 %	0.987	0.002
Device factor	Improvement of the reliability of equipment	95 %	89 %	93 %	87 %	364 %	0.993	0.003
	Reliable consideration rate of transport equipment	93 %	92 %	93 %	85 %	363 %	0.999	0.002
	Reliability of power supply equipment	96 %	95 %	96 %	94 %	381 %	0.999	0.000
	Reliability of mechanical and electrical equipment	85 %	82 %	89 %	81 %	337 %	0.999	0.002
	Reliability of ventilation equipment	89 %	82 %	89 %	71 %	331 %	0.997	0.005
	Reliability of drainage equipment	77 %	84 %	89 %	80 %	285 %	0.998	0.002
	Improvement rate of water proof and fire-fighting facilities	80 %	70 %	62 %	73 %	330 %	0.997	0.005

Continuation of Table 4 is on the following page

Management factor	Perfect rate of safety system	0.095	0.014	0.011	0.028	0.15	0.732	0.486
	Perfect rate of safety measures	19	24	14	22	79.00	0.986	0.025
	Management efficiency	1830	433	735	1431	4429.0	0.905	0.171
	Timeliness of management	5	7	6	8	26.00	0.902	0.019
	Perfect rate of emergency mechanism	5	4	7	7	23.00	0.989	0.034
	Improvement rate of coal mine access management	77 %	84 %	89 %	80 %	330 %	0.981	0.002
Environmental factor	The outstanding times of millions of tons of gas	82 %	67 %	75 %	59 %	283 %	0.998	0.009
	Dip angle of coal seam	79 %	71 %	74 %	62 %	286 %	0.994	0.005
	The normal water inflow of mine	81 %	74 %	70 %	69 %	294 %	0.997	0.003
	Maximum dust concentration on the face	89 %	81 %	87 %	80 %	337 %	0.998	0.001
	Spontaneous combustion period of coal seam	92 %	90 %	94 %	84 %	360 %	0.999	0.001
	Reliability of top and bottom plate	82 %	83 %	81 %	84 %	330 %	0.999	0.000
Information factor	Informatization degree	62 %	53 %	68 %	49 %	232 %	0.994	0.011
	Information identification	60 %	57 %	70 %	52 %	239 %	0.995	0.008

Finally, the situation of training status of coal mine safety management personnel was investigated. The results of the survey are shown in Fig. 4. The results show that the coal mine enterprises in China begin to strengthen the training for the related staff, and this will also provide certain technical support and guarantee for the safe operation of China's coal mining enterprises, and provide some support of energy for the development of other industries.

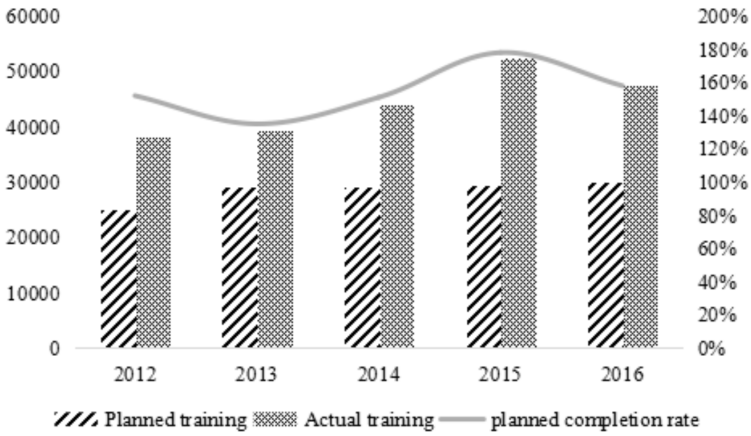


Fig. 4. Analysis of the related personnel training in China's coal mine safety management

Table 5. Membership degree of influence index of each coal mine

Index of the first level	Coal mine A	Coal mine B	Coal mine C	Coal mine D
Human factor	0.9667	0.7000	0.9000	0.6667
Device factor	1.0000	1.0000	1.0000	0.8000
Management factor	0.4000	0.1000	0.6000	0.0000
Environmental factor	0.3333	0.2333	0.6667	0.0667
Information factor	0.6333	0.4667	0.5667	0.6000

5. Conclusion

The coal industry is the basic industry of our country and plays an important role in the development of the national economy. Coal is the main consumption energy in our country and contributes greatly to the economic construction of our country. However, the safety production of coal mine has always been a major problem and has caused huge economic losses and casualties. Therefore, coal mine enterprises should make full use of modern technical means, combine computer technology and network communication management to accelerate the improvement of safety management instructions of coal mine, reduce and avoid coal mine accidents. The deep causes of coal mine safety accidents were first analyzed, and the main factors of coal mine safety production were comprehensively evaluated, including management system factor, human factor, machine factor and environment information factor; and the cause model of coal mine safety production was constructed. On the basis of the study, the dynamic monitoring of coal mine safety management was carried out by using modern scientific theory, and genetic algorithm and fuzzy neural network model were combined to optimize the neural network and guarantee the better convergence speed and the generalization mapping ability of the network; in addition,

the early-warning model was established for coal mines, and the key aspects of coal mine safety management were determined, so as to lay the foundation for the application and popularization of fuzzy neural network in coal mine safety management.

References

[1] C. C. Hogue: *Injury in late life: Epidemiology.* Journal of the American Geriatrics Society *30* (1982), No. 3, 183–190.

[2] J. Beugin, J. Marais: *Simulation-based evaluation of dependability and safety properties of satellite technologies for railway localization.* Transportation Research Part C: Emerging Technologies *22* (2012), 42–57.

[3] W. G. Johnson: *MORT: The management oversight and risk tree.* Journal of Safety Research (1975), No. 7, 4–15.

[4] G. Stafford, L. Yu, A. K. Armoo: *Crisis management and recovery how Washington, D.C., hotels responded to terrorism.* The Cornell Hotel and Restaurant Administration Quarterly *43* (2002), No. 5, 27–40.

[5] J. G. U. Adams: *Risk and freedom: The record of road safety regulation.* Transport Publishing Projects, 59 Park Place Cardiff, Wales (1985), Accident Analysis & Prevention (Elsevier), *18* (1986), No. 6, 505–506.

[6] E. Hollnagel: *The reliability of man-machine interaction.* Reliability Engineering & System Safety *38* (1992), Nos. 1–2, 81–89.

[7] Y. Himeno, T. Nakamura, S. Terunuma, T. Furubayashi: *Improvement of man-machine interaction by artificial intelligence for advanced reactors.* Reliability Engineering & System Safety *38* (1992), Nos. 1–2, 135–144.

[8] W. Graf, S. Freitag, M. Kaliske, J. U. Sickert: *Recurrent neural networks for uncertain time-dependent structural behavior.* Computer-Aided Civil and Infrastructure Engineering *25* (2010), No. 5, 322–323.

[9] D. M. Siegel, V. H. Frankos, M. A. Schneiderman: *Formaldehyde risk assessment for occupationally exposed workers.* Regulatory Toxicology and Pharmacology *3* (1983), No. 4, 355–371.

[10] R. A. Bare: *Decision making and probabilistic risk assessment.* Nuclear Engineering and Design *93*, (1986), Nos. 2–3, 341–348.

[11] J. B. Bowles, C. E. Peláez: *Fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis.* Reliability Engineering & System Safety *50* (1995), No. 2, 203–213.

[12] W. Hatton, M. K. G. Whateley: *Risk assessment applied to coal tonnage estimation in the United Kingdom.* Transactions — Institution of Mining & Metallurgy, Section A (1995), No. 104, A79–A86, International Journal of Rock Mechanics and Mining Science & Geomechanics *32*, (1995), No. 6, paper 276.

[13] H. A. Gabbar, K. Suzuki, Y. Shimada: *Design of plant safety model in plant enterprise engineering environment.* Reliability Engineering & System Safety *73* (2001), No. 1, 35–47.

[14] K. S. Murray, D. T. Rogers: *Groundwater vulnerability, brownfield redevelopment and land use planning.* Journal of Environmental Planning and Management *42* (1999), No. 6, 801–810.

[15] F. Tödtling, A. Kaufmann: *SMEs in regional innovation systems and the role of innovation support–the case of Upper Austria.* Journal of Technology Transfer *27* (2002), No. 1, 15–26.

Received July 2, 2017

Acta Technica *62* No. 2A/2017, 27–38

Research on intelligent home energy scheduling algorithm[1]

Chuanwei Qi[2]

Abstract. Aiming at the trade-off between energy utilization and user comfort in smart home, an intelligent home energy dispatching algorithm based on energy consumption prediction is proposed. Then, the concept of "dynamic comfort" is expounded, which describes the relationship between user comfort and energy savings in each time period. According to the prediction of the equipment operation and the energy information provided by outer family, the scheduling algorithm makes reasonable allocation and scheduling of energy and equipment. At the same time, aiming at the deviation between forecasting information and real-time information, the real-time correction scheduling algorithm is proposed. The prediction algorithm and the scheduling algorithm are corrected by collecting the feedback information of the user. The results show that the proposed algorithm can improve the energy utilization rate while protecting the user's comfort. It achieves a trade-off between user's comfort and energy savings.

Key words. Smart home, energy scheduling, energy utilization, users' comfort index.

1. Introduction

In recent years, with the development of science and technology, social life has become increasingly close ties with the information [1]. The standard of living is improved, and the living habits are changed. People are no longer satisfied with the most basic living environment. Therefore, smart home is very popular. Smart home will become a major change in people's way of life. In China, the irrational energy structure and environmental degradation have become the focus of current social concern [2]. Reasonable energy development strategy and smart grid construction have become very urgent. Smart grid is a new type of intelligent power network. By creating an open information platform, it promotes the linkage of power flow, information flow, and business flow [3]. The operation and management of the infrastructure and power supply of the power industry are optimized. Based on

[1]Funding program: 2017 Henan science and technology research plan 172102210433, the autonomous learning intelligent platform for Higher Vocational Education based on Bot Framework.

[2]The Mathematics and Information Engineering Department of Puyang Vocational and Technical College, Puyang, Henan, 457000, China

these advantages, many countries and regions around the world have taken the smart grid as an important development target for the future power grid [4]. Smart grid is widely researched and practiced in various fields. As a smart grid users, intelligent home integrated Internet of Things technology to build efficient residential facilities and family planning business management system [5]. It can enhance and protect the user's needs for the living environment, and realize the green living environment through energy management. At present, energy resources is becoming scarce, household electricity equipment is increasing, and users' demand for energy conservation is improving [6]. Therefore, intelligent home energy scheduling has become an important research direction. As an important part of smart grid, smart home has a good application prospect for smart grid technology [7].

In terms of energy, due to the deterioration of the environment and the increase in the greenhouse effect, energy conservation has become one of the development trends of all walks of life [8]. The popularity and development of energy-saving technologies in smart homes will play a key role in energy issues around the world. Through smart grid technology, users can reduce household energy costs by scheduling home appliances. It not only meets the interests of users, but also provides further theoretical support and technical support for energy conservation. Due to low energy consumption and the number of users, smart home can effectively achieve energy-saving and emission-reduction. Its size and efficiency are considerable. In this paper, an intelligent home energy scheduling algorithm based on the user's energy consumption prediction is used to schedule the energy use in smart home, which largely alleviates the energy pressure. In addition, this algorithm takes into account the user's comfort, and ensures the safety, convenience and comfort of the home. It helps to optimize people's way of life. Intelligent home control system has a very bright future, which is the significance of this research.

2. State of the art

With the development of electronic information technology, there has been an electronic home machine in the world since 1980 [9]. At this time, there is a residential electronic. Since 1985, residential electronics have been further developed [10]. The household appliances integrate the communication function, so as to realize the automation of the residence. Since 1988, with the further development of electronic information technology, all kinds of large companies began to develop residential automation system. The earlier commercial appliance automation system is the US Smart Home system. The system uses bus technology to achieve the electrical monitoring and management. It becomes the master and origin of the modern smart home system. Before the residential automation, the concept of smart home has emerged. However, after residential automation, smart home is gradually applied to the building. The earliest smart home is appeared in the United States Hartford City, known as City Place Building, that is, intelligent building. The building was built in 1984 by the United States of America Joint Technology Company [11]. The building opened the prelude of smart homes around the world. Smart home has become an integrated building system integrating information technology and home

appliances.

The emergence of the world's first smart home case caused the reform of home appliances in major countries, such as Europe and the United States. Various countries have introduced policies to promote the development of smart home. So far, in the United States and Japan and other developed countries, intelligent home has become an important part of people's daily lives. When the smart home is still a new concept, the prospects for its development have been widely endorsed by the world. With its market sensitivity, all large companies have launched smart home products. In the smart home industry, in order to occupy a place in the ranks of smart home development, these companies invest huge sums of money. Motorola, IBM, Panasonic and other home appliance giants have established branches for the smart home production [12]. These companies have launched a development and production of smart home systems and smart home mode. As a high-tech followers and creators, US Silicon Valley also carried out research and development of intelligent home [13]. It integrates information technology and intelligent home systems, and takes it as an important direction for the development of its home technology. In the smart home market, capital has become one of the most important factors limiting the development of the industry. However, it does not hinder the development of the industry because of the lack of capital. Various consortia and venture capital institutions have seen the future development of smart home, and invested in the industry. With venture capital, a variety of emerging small and medium enterprises also joined the smart home production and research and development. In the process of capital allocation, it not only reflects the support of small and medium-sized enterprises, but also reflects the combination of capital and large enterprises. Large household electrical appliance enterprises are working with the consortium and venture capital continue to improve the technology of smart home. They combine smart home with web browsers. Smart home network browsing allows it to carry out their own indicators of the collection and diagnosis, and can monitor their own performance. The strong research and development of these companies not only makes some intelligent products continue to emerge, but also makes the existing electronic products gradually intelligent and humane. TV, cameras and other home appliances can be remote control. People can choose a variety of functions. It shows more intelligent from the operation of electronic products to the use of home appliances.

At present, the development and application of intelligent home control system in our country are still in the initial stage. However, the smart home market has shown a prosperous development of the scene at this stage. "Twelfth Five-Year Plan" also made it clear that the smart home is taken as one of the directions of nurturing new industries. In the strong support of the Internet of Things, China's intelligent home control system will usher in a new period of prosperity. However, in the past decade, China's smart home market has encountered bottlenecks. In most homes, smart home systems are still a concept that does not form a consumer. China's smart home industry development status is not good, and it has a lot of problems. For example: China's smart home system technical standards are not uniform. Intelligent home system is poor operability and expensive. Over the past ten years, although the smart home system makes a lot of conceptual products

become a reality, it cannot meet all the needs of people. The most important reason is that the stability of the smart home control system products is not strong, and the compatibility is still very immature. Therefore, the actual effect is very different, and it cannot achieve the desired effect.

At this stage, the development of China's real estate industry strongly supports the smart home industry, which has created a new direction and development goals for the development of the real estate industry. IT companies began to study the intelligent home control system from the practical point of view. Statistics show that 70 % of households in China have broadband networks [14]. In the process of urbanization, about 50 % of households achieve intelligent control. Although the smart home control system started late, some well-known traditional home appliance industry giants have entered the smart home industry, such as Haier, TCL and so on. These companies have invested a lot of money to study smart home technology. Basically, most home appliance companies are trying to launch their own flagship smart home products, such as Haier group launched the "U-home", and TCL launched the "mihome my intelligent housekeeper". Millet and developers cooperate to study the support of home appliances projects, which forms a smart home family experience closed-loop. It has been applied to a certain extent [15].

3. Methods

3.1. Overview of intelligent home energy scheduling algorithms

In the intelligent home energy dispatching algorithm, the user side of the renewable energy efficiency, low electricity costs, user comfort and the power grid side of the peak ratio and the avoidance of the rebound peak both are the main needs of intelligent home system development. It is also an important performance indicator for the merits of smart home systems. Their relationship is close and mutual influence. In the case of only considering the energy efficiency and electricity consumption, this may lead to a reduction in user comfort or the emergence of the peak of the grid side rebound. However, the optimization of user comfort will increase the peak to average ratio. Therefore, how to balance multiple objects of smart home energy scheduling algorithm is the current research hotspots. Based on the comprehensive consideration of multiple influencing factors, and from the general research methods that have been studied, this paper summarizes the technology and their characteristic related to the energy dispatching mechanism from different aspects, and analyzes its limitations and advantages. A reasonable method of energy dispatching algorithm is designed and applied.

In the scheduling area, there are multi-user energy scheduling algorithm and single-user energy scheduling algorithm. The energy scheduling algorithm focuses on how to coordinate the behavior of multi-user electricity to protect the stability of the power grid. User privacy issues are an urgent need to address the problem, and it lacks user comfort considerations. The indoor energy dispatching algorithm focuses on how to coordinate the operation of home appliances in the home to reduce the

cost of the family and how to ensure comfort. However, it will affect the stability of the grid in the actual grid environment. Therefore, in the scope of scheduling, we need to achieve the combination of indoor scheduling. On the basis of ensuring the user comfort and optimizing the electric charge, we should take into account the stability of the grid. This will be the main direction of the future research.

In the aspect of data information providing technology, the basic information of scheduling algorithm is mainly provided by real-time information acquisition technology and prediction algorithm. Among them, the energy scheduling algorithm based on real-time information is more flexible and real-time, but it has a great influence on the stability of grid side. By using the prediction algorithm, the energy dispatching algorithm obtains the basic information needed for scheduling, so as to realize the large-scale pre-scheduling of the home equipment. According to the different forecast data, it is divided into electricity price prediction and energy consumption prediction. The energy dispatching algorithm based on electricity price forecast has some effect on reducing the peak-to-peak ratio of the grid and preventing the rebound peak. The energy dispatching algorithm based on energy consumption forecast has some effect on the comfort of the user and the stability of the grid side grid. Because of the inevitability of the difference between the predicted information and the real-time information and the unavoidability of the occurrence of an unexpected event (artificially open the equipment without the task of the task) during the scheduling process, there is a disadvantage of the predictive energy scheduling algorithm, that is, the processing capacity of real-time information. Therefore, on the basis of the forecast, it realizes the pre-scheduling repair through real-time scheduling, thus combining the advantages of the two scheduling. This is the only way to the development of smart home energy management strategy.

In the aspect of user participation, the existing intelligent home system embedded in the energy dispatching algorithm mainly includes simple functions such as setting the parameters of the scheduling range and the user's sudden intervention operation. Most of the energy scheduling algorithms have less consideration for user interaction. Under the premise of simple operation, the effective use of user interaction technology can enhance the observation and protection of user comfort. At the same time, after the user feedback information is dataized, it is applied to the feedback of the intelligent algorithm. According to the user feedback information, the algorithm can be adjusted in time, so that the algorithm is more close to the actual needs of the user, to achieve intelligent user feedback information processing unit. These are not only an important part of user interaction, but also an integral part of the future smart home system.

Therefore, the multi-objective trade-offs and the effective combination of multiple technologies will further develop the energy dispatching algorithm. Intelligent and humanized design is the key to solve the demand of energy scheduling algorithm.

3.2. The concept of dynamic comfort

In order to achieve a dynamic balance between energy efficiency and user comfort, this paper puts forward the concept of dynamic comfort. This concept refers

to the dynamic relationship between energy efficiency and user comfort within 24 hours. Through the analysis of the household energy consumption data in the REDD database, we can get the operation time of the equipment within 24 hours and the frequency of use is different in the working day and the non-working day. Therefore, we set the dynamic comfort in 24 hours on weekdays and non-working days respectively. As shown in Table 1, on weekdays, in these two time periods of 09:00–18:00 and 00:00–06:00, the user's comfort requirements are relatively low, and the energy utilization rate is relatively high, so it can be set to low comfort and high energy saving state. Because in the time period of 09:00–18:00, the user is usually at work or in school, the family is less or not, so the requirements of the home environment are not very high. In the time period of 00:00–06:00, the user is in a state of sleep, and they do not have too much activity, so there is not much demand for the operation of most equipment. In other time periods, the users have higher requirements for the comfort of the home environment, and the promotion of energy efficiency is carried out under the premise of satisfying the user's comfort. Therefore, in this period of time, the dynamic comfort is high comfort and low energy state. In this period of time, the user is more active in the use of equipment, and the requirements for the home environment are high. In non-working days, in the time period of 06:00–24:00, the users have higher requirements for the comfort of the home environment, and the promotion of energy efficiency is carried out under the premise of satisfying the user's comfort. Therefore, in this period of time, the dynamic comfort is high comfort and low energy state. In the non-working days, in this period of time, the user's activities are more frequently, and the use of equipment is higher.

Table 1. The distribution of the dynamic comfort

	Time	Comfort requirements	Energy-saving requirements
Working day	06:00-09:00	high	low
	18:00-00:00	high	low
	00:00-06:00	low	high
Non-working day	09:00-18:00	low	high
	06:00-00:00	high	low
	00:00-06:00	low	high

3.3. The users' feedback

In high-level entity management system (HEMS), the equipment operation display module provides a platform for users to evaluate the performance of the equipment. According to their own habits, users can evaluate the operation of today's equipment. They can point out that they are not satisfied with the equipment, and the satisfaction of the equipment running status, including the best opening time, run time and open point range. According to the difference between the feedback information and the actual operation information, the prediction algorithm and the

scheduling algorithm are used to correct the algorithm. If the proposed opening range does not include the predicted optimal opening point, the feedback information is fed back to the prediction algorithm to retrain the network in the prediction algorithm. If the proposed opening range includes the predicted optimal opening point, the feedback information is fed back to the scheduling algorithm to modify the corresponding information.

4. Simulation results analysis

In this paper, we use MATLAB simulation software. The household energy data in the REDD database, the renewable energy data in IRENA-Global-Atlas, and the real-time electricity price in the Com Ed's RRTP program are used to simulate the proposed algorithm. Then, its feasibility is verified. In the case of using the scheduling algorithm and without the use of scheduling algorithm, we can verify the effectiveness of the proposed scheduling algorithm in improving the energy efficiency by comparing the daily electricity expenditure for one week, the total energy consumption of the day, and the power consumption of the grid of the day. By calculating the average offset of the device's turn-on time deviation from the user's satisfaction time, we validate the effectiveness of the proposed algorithm in improving user comfort. Through a comprehensive comparison, we verify that the proposed algorithm can achieve a good tradeoff between energy efficiency and user comfort.

In the case of using the scheduling algorithm and without the use of scheduling algorithm, the daily cost of the week is compared. The results are shown in Fig. 1.

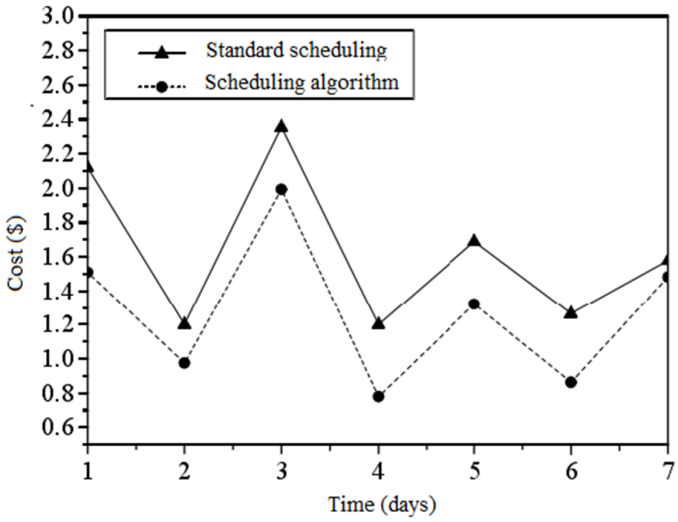


Fig. 1. The daily cost in a week

From Figure 1, it can be seen that the daily electricity bill is less when using the scheduling algorithm. The daily electricity bill is great when not using the scheduling

algorithm. Through the calculation, we can conclude that the average daily 22 % of electricity bills are saved. It meets the requirements of users to reduce electricity costs. As a result, it verifies the effectiveness and feasibility of the algorithm in reducing the electricity consumption of users.

On weekdays, the energy consumption is compared for each time period by using the scheduling algorithm and not using the scheduling algorithm. The results are shown in Fig. 2. On non-working days, the energy consumption is compared for each time period by using the scheduling algorithm and not using the scheduling algorithm. The results are shown in Fig. 3.

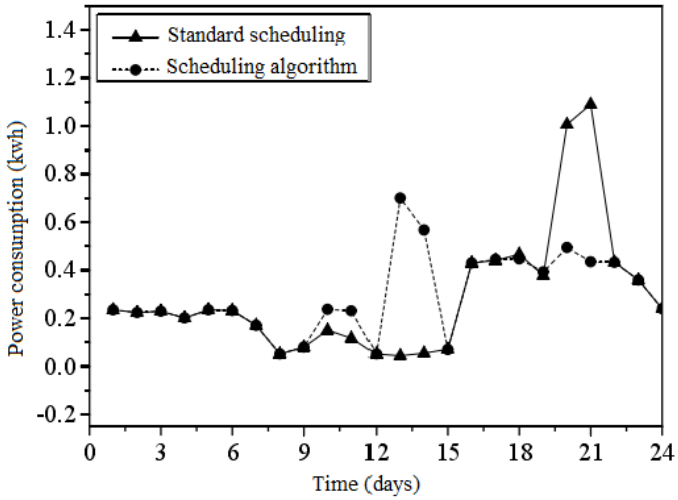


Fig. 2. The hourly electricity consumption in a workday

From Figs. 2 and 3, it can be seen that the total energy consumption is not reduced every day, but the peak of electricity has changed. The concentration of electricity in the family effectively avoids the peak price or uses renewable energy to reduce the amount of electricity consumed during peak hours, which reduces carbon emissions and household expenses.

Figure 4 shows the use of grid power under the condition of using the scheduling algorithm and not using the scheduling algorithm on weekdays.

Figure 5 shows the use of grid power under the condition of using the scheduling algorithm and not using the scheduling algorithm on non-working days.

From Figs. 4 and 5, it can be seen that the use of the scheduling algorithm for grid energy usage is significantly less than that without the use of scheduling algorithms. Among them, in the working day, an average of 38 % of the grid power usage is reduced. In non-working days, an average of 44 % of grid power usage is reduced. Combined with Figures 4 and Figure 5, it is verified that the scheduling algorithm can help users to reduce the cost of energy utilization during periods of low comfort and high energy efficiency requirements.

Figure 6 shows the average offset per hour for the turn-on time of the device after the dispatching algorithm deviates from the user's happy opening time in the

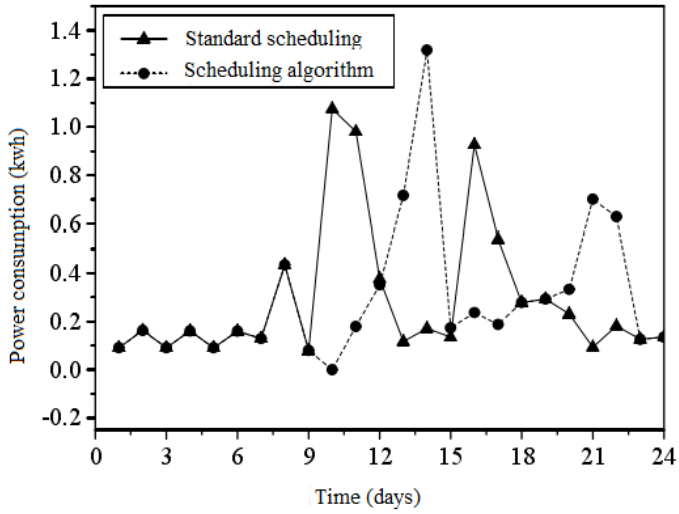


Fig. 3. The hourly electricity consumption in a weekend

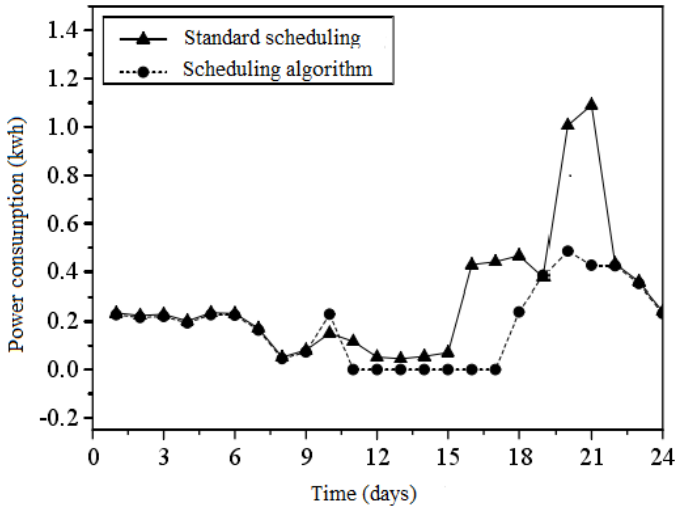


Fig. 4. The hourly grid's electricity consumption in a workday

working and non-working days. The average offset represents the average opening time of the device and the average time that the user is most satisfied with the opening time or the delay.

As can be seen from Figure 6, the offset of the device opening time is small in the period of high comfort requirement. This ensures the comfort of the user. It proves that the scheduling algorithm can effectively guarantee the user's requirements for the operation of the equipment while saving energy.

To sum up, the proposed algorithm meets the requirements of users to reduce the household power expenditure. In the period of low comfort requirements, it

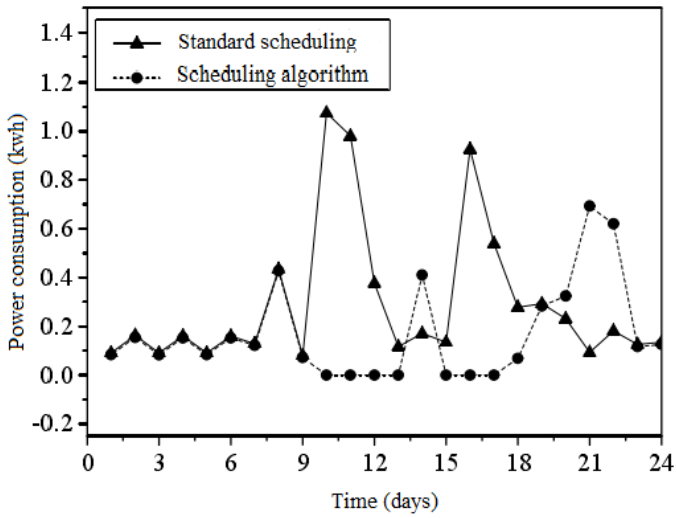


Fig. 5. The hourly grid's electricity consumption in a weekend

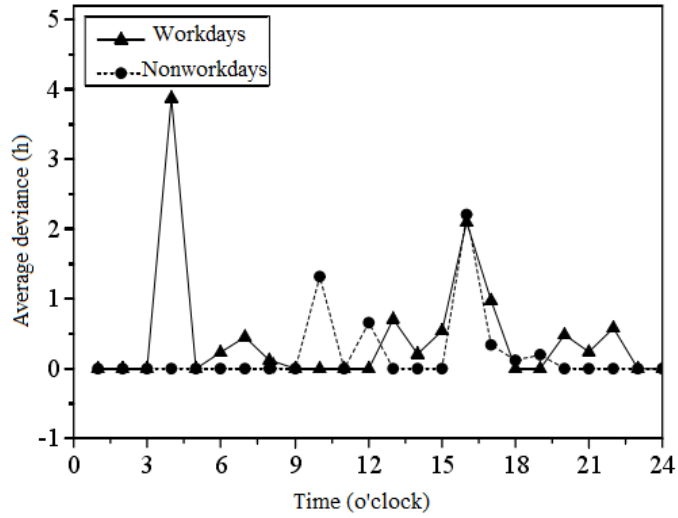


Fig. 6. The average offset of the opening time of the appliances

improves the energy efficiency. In the period of high comfort requirements, it ensures and improves user comfort. The scheduling algorithm achieves a trade-off between energy utilization and user comfort.

5. Conclusion

The scheduling algorithm distributed energy and equipment reasonably according to forecast and real-time information. The user feedback link provided the user with

feedback information for the prediction algorithm and the scheduling algorithm, which can be used to correct the algorithm. The results show that the scheduling algorithm achieves a trade-off between energy utilization and user comfort. When the demand of comfort is high, it improves energy efficiency while protecting the user's comfort. When the demand of energy-saving is high, it ensures the user's comfort while improving energy efficiency.

References

[1] S. Althaher, P. Mancarella, J. Mutale: *Automated demand response from home energy management system under dynamic pricing and power and comfort constraints.* IEEE Transactions on Smart Grid *6* (2015), No. 4, 1874–1883.

[2] K. A. Moornani, M. Haeri: *Robust stability check for fractional PID-based control systems.* Transactions of the Institute of Measurement and Control *35* (2013), No. 2, 236–246.

[3] A. Anvari-Moghaddam, H. Monsef, A. Rahimi-Kian: *Optimal smart home energy management considering energy saving and a comfortable lifestyle.* IEEE Transactions on Smart Grid *6* (2015), No. 1, 324–332.

[4] M. Collotta, G. Pau: *An innovative approach for forecasting of energy requirements to improve a smart home management system based on BLE.* IEEE Transactions on Green Communications and Networking *1* (2017), No. 1, 112–120.

[5] S. Saponara, T. Bacchillone: *Network architecture, security issues, and hardware implementation of a home area network for smart grid.* Journal of Computer Networks and Communications (2012), paper 534512.

[6] O. A. Saraereh, M. A. Smadi, A. K. S. Al-Bayati, Q. H. Alsafasfeh: *A novel low-profile coupled-fed printed twelve-band mobile phone antenna with slotted ground plane for LTE/GSM/UMTS/WIMAX/WLAN operations.* World Academy of Science, Engineering and Technology *7* (2013), No. 8, 1063–1072.

[7] Y. Fu, M. Liu, L. Li: *Multiobjective stochastic economic dispatch with variable wind generation using scenario-based decomposition and asynchronous block iteration.* IEEE Transactions on Sustainable Energy *7* (2016), No. 1, 139–149.

[8] N. G. Paterakis, O. Erdinç, I. N. Pappi, A. G. Bakirtzis, J. P. S. Catalão: *Coordinated operation of a neighborhood of smart households comprising electric vehicles, energy storage and distributed generation.* IEEE Transactions on Smart Grid *7*, (2016), No. 6, 2736–2747.

[9] M. E. M. B. Gaid, A. S. Cela, Y. Hamam: *Optimal real-time scheduling of control tasks with state feedback resource allocation.* IEEE Transactions on Control Systems Technology *17* (2009), No. 2, 309–326.

[10] H. T. Nguyen, D. T. Nguyen, L. B. Le: *Energy management for households with solar assisted thermal load considering renewable energy and price uncertainty.* IEEE Transactions on Smart Grid *6* (2015), No. 1, 301–314.

[11] U. S. Premarathne, I. Khalil, M. Atiquzzaman: *Trust based reliable transmissions strategies for smart home energy consumption management in cognitive radio based smart grid.* Ad Hoc Networks *41* (2016), No. C, 15–29.

[12] T. Shin, H. T. Jeon, J. Byun: *Developing nontrivial standby power management using consumer pattern tracking for on-demand appliance energy saving over cloud networks.* IEEE Transactions on Consumer Electronics *62* (2016), No. 3, 251–257.

[13] F. Salah, C. M. Flath: *Deadline differentiated pricing in practice: Marketing EV charging in car parks.* Computer Science - Research and Development *31* (2016), Nos. 1–2, 33–40.

[14] X. Wu, X. Hu, X. Yin, S. Moura: *Stochastic optimal energy management of smart home with PEV energy storage.* IEEE Transactions on Smart Grid *PP* (2016), No. 99, 1–1.

[15] J. Vanus, P. Valicek, T. Novak, R. Martinek, P. Bilik, J. Zidek: *Utilization of regression analysis to increase the control accuracy of dimmer lighting systems in the smart home*. IFAC-PapersOnLine *49* (2016), No. 25, 517–522.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 39–50

Study on video vehicle detection based on improved mean-shift algorithm

Geng Liwei[1], Dong Tingting[1]

Abstract. In allusion to defects of traditional Mean-Shift algorithm such as missed detection and object overlap, this paper introduces blob detection into Mean-Shift algorithm, puts forward vehicle video tracking algorithm of an improved Mean-Shift algorithm and achieves accurate and fast positioning of single object and multiple objects. The experimental result shows that Mean-Shift algorithm based on blob can not only achieve multi-object tracking, but also have more accurate tracking effect. Its tracking effect is obviously better than that of traditional Mean-Shift algorithm.

Key words. Mean-Shift algorithm, similarity, blob detection, Kalman filtering.

1. Introduction

Object tracking is a key and hotspot issue in the field of computer vision [1]. How to achieve object tracking in video sequence accurately is a difficult issue in the current research. As an important algorithm in object tracking research, Mean-Shift algorithm was first put forward by Fukunaga and Hostetler in 1975. Yizong Cheng [2] conducted extending study on it based on Mean-Shift theory and applied it to the field of computer vision. It started to arouse wide concern of domestic and foreign research scholars. Dorin Comaniciu and Peter Meer [3] boiled down nonrigid object tracking issue to optimization problem of Mean-Shift algorithm approximately. Mean-Shift algorithm has so far applied to object tracking formally. Nummiaro [4] combined Mean-Shift algorithm and particle filter and put forward multi-object tracking algorithm of an improved Mean-Shift algorithm. In allusion to infrared small object tracking problem, Chen Jiang et al. [5] combined Mean-Shift algorithm and cascade gray space and achieved good effect through application. Tracking and positioning accuracy has been greatly improved compared to Mean-Shift algorithm.

As kernel bandwidth of search window of traditional Mean-Shift algorithm remains unchanged in the tracking process, good tracking effect can be achieved for objects with minor scale change. However, stable tracking cannot be achieved and

[1]Agricultural University of Hebei, Baoding, Hebei Province, 071000, China

objects are often lost or tracking objects are missed for detection in the case of significant change of spatial scale of objects. To describe the change of object scale with time, Zhu Shengli et al. [6] used affine transformation for description, obtained transformation parameter with maximum correlation coefficient of match window in Mean-Shift and two continuous frames and calculated the change of bandwidth and the initial position of Mean-Shift. However, affine model is easily influenced by illumination intensity. Moreover, it is required to conduct correlation operation of point N in angle and scale. The calculation is complicated and real-time tracking cannot be achieved.

In allusion to defects of Mean-Shift algorithm, this paper introduces blob detection into Mean-Shift algorithm based on literature [6], puts forward vehicle video tracking algorithm of an improved Mean-Shift algorithm and achieves accurate and fast positioning of single object and multiple objects.

2. Methodology

Mean-Shift video tracking algorithm is a semi-automatic algorithm [7–8]. In starting frame, moving object is generally determined through manual or other recognition algorithms and meanwhile histogram distribution of target window is calculated. Then, histogram distribution of the corresponding window is calculated in the N^{th}frame of video sequence, which is used as candidated object model. The similarity of the two models is calculated and maximum similarity of the two models is regarded as the principle for moving tracking window so as to determine the position of object accurately.

2.1. Modeling

Mean-Shift algorithm generally uses probability distribution of weighting of characteristic value to determine initial object model [9]. In starting frame of video sequence, probability distribution of characteristic value at each point within target window area is counted with kernel function weighting method. To improve the efficiency of operation, characteristic space of object is transformed into multiple characteristic values, in which value q_u of the uth feature is [10–11]:

$$\hat{q}_u = C \sum_{i=1}^{n} k \left(\left\| \frac{x_i - x_0}{h} \right\|^2 \right) \delta \left[b\left(x_i\right) - u \right] . \tag{1}$$

In formula (1), x_0, x_i, respectively, correspond to the coordinate vectors in target window center and any point in the window, $k||x||^2$ refers to the kernel function, h refers to the bandwidth of object model, $b(x_i)$ refers to the characteristic value at point x_i, C denotes the standard coefficient of characterization and the sum of probabilities of all characteristic values is 1. Candidated object model can be obtained according to feature distribution of search window located in y in frame N of computing center of video sequence [12].

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{x_i - y}{h}\right\|^2\right) \delta\left[b\left(x_i\right) - u\right] . \tag{2}$$

2.2. Similarity comparison

So called $\hat{p}_u(y)$ similarity with $\hat{q}_u$ is described through the Bhattacharyya coefficient. Quantity $\hat{\rho}(\mathbf{y}) = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\mathbf{y})\hat{q}_u}$ is used for measurement and maximum similarity is regarded as the selection principle. Quantity $\hat{\rho}(\mathbf{y})$ is subject to Taylor expansion at $\hat{p}_u(\mathbf{y}_0)$ of the last frame

$$\hat{\rho}(\mathbf{y}) \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\mathbf{y}_0)\hat{q}_u} + \frac{1}{2}\sum_{u=1}^{m} \hat{p}_u(\mathbf{y})\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y}_0)}} . \tag{3}$$

The first item on the right-hand side of formula (3) is a constant value. Therefore, the value of the second item must be maximum in order to achieve the maximum of $\hat{\rho}(\mathbf{y})$.

2.3. Object positioning

After further concretization, formula (3) can be transcripted in the following way,

$$\hat{\rho}(\mathbf{y}) \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\mathbf{y}_0)\hat{q}_u} + \frac{C_h}{2}\sum_{i=1}^{n_h}\left\{k\left(\left\|\frac{\mathbf{y} - \mathbf{x}_i}{h}\right\|^2\right)\sum_{u=1}^{m}\delta\left[b\left(\mathbf{x}_i\right) - u\right]\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y}_0)}}\right\} . \tag{4}$$

According to Mean-Shift algorithm, the mean shift vector can be obtained through the formula

$$\mathbf{m}_{h,G}(\mathbf{x}) = \hat{\mathbf{y}}_1 - \mathbf{y}_0 = \frac{\sum_{i=1}^{n_h} \mathbf{x}_i w_i g\left(\left\|\frac{\mathbf{y}_0 - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{\mathbf{y}_0 - \mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{y}_0 . \tag{5}$$

In formula (5), $w_i = \sum_{u=1}^{m} \delta\left[b\left(\mathbf{x}_i\right) - u\right]\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\mathbf{y}_0)}}$, where $g\left(\bullet\right)$ is the shadow kernel function of $k\left(\bullet\right)$. After a new position is obtained, Mean-Shift iteration is conducted with the new position as base point till the optimal central position of the object is obtained.

3. Improved Mean-Shift vehicle detection algorithms

3.1. Blob definition

Blob segmentation in target scene first came from human visual system. While observing the whole scene of an object, observers divide sceneries observed within the

visual scope into several independent regions with similar features roughly according to comprehensive features of the observed object such as texture, outline and color, neglect minor details between regions, form several interested regions in vision and call each interested region as a blob.

Blob segmentation splits foreground object according to some rules under the support of such theory and makes arithmetic processing no longer pertinent to the whole foreground image and instead aim at local regions of image with comprehensive features such as color, position and size.

3.2. Blob segmentation

Blob segmentation is an important foundation of new blob detection. The exterior rectangle of a moving object is regarded as the shape of blob in the foreground. If the distance between the center of two adjacent blob rectangles, (i.e. two moving objects), is less than the sum of their width, the two moving objects represent a blob. Figure 1 shows its detection diagram.

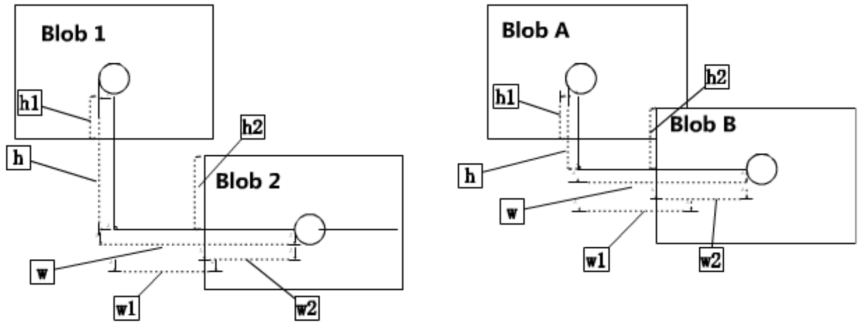


Fig. 1. Blob detection diagram

As shown in the right part of Fig. 1, if width w or height h between the center of blob 1 and blob 2 meets formula (6), blob 1 and blob 2 are two independent blobs; on the contrary, they belong to the same blob.

$$w > (w_1 + w_2) || h > (h_1 + h_2)\,. \tag{6}$$

Blob segmentation is the first step of blob detection. Based on the definition of blob, all blobs (moving vehicles) within detection region are put into a blob chain table. When a new vehicle enters the detection region, the new blob detected is added to blob chain table.

3.3. Location prediction

Kalman filtering can achieve location prediction of moving objects, reduce the calculating amount of Mean-Shift algorithm and meanwhile solve the problem of failure to extract features of target vehicle or wrong extraction caused by occlusion

effectively. During actual tracking algorithm, this paper integrates location prediction of moving vehicle into Mean-Shift algorithm.

The dimension of blob is also required to set the size of Mean-Shift search box. Formulas for calculating blob dimension are as below.

$$Ratio = \frac{BlobH_{t0}}{BlobW_{t0}}\,, \tag{7}$$

$$BlobW_t = (LenTop + LenBtm) \times \frac{yPos_t - yBtm}{yTop - yBtm}\,, \tag{8}$$

$$BlobH_t = BlobW_t \times Ratio\,. \tag{9}$$

4. Multi-object tracking

Though Mean-Shift algorithm has many advantages in object tracking, it has two fatal deficiencies:In blob segmentation module, active blobs detected are put into a blob list and then the list is transmitted to tracking module. The number of blobs $BlobNum$ in the blob list is obtained before entry into tracking module. If $BlobNum > 0$, there is an object to be tracked currently. Formula (10) shows a loop computation:

$$BlobNum = \begin{cases} BlobNum - 1, \text{ if } BlobNum > 0\,, \\ 0 \text{ otherwise}\,. \end{cases} \tag{10}$$

Thread is launched for each loop computation; tracking algorithm is executed in the thread and multi-object tracking problem is achieved through multithreading.

5. Experimental simulation

5.1. Sing-object tracking

The test videos in Figs. 2 and 3 are in PETS2000 format and video object is an automobile with constantly reducing scale.

According to Figs. 2 and 3, traditional Mean-Shift algorithm can realize accurate object positioning and tracking in the process of tracking automobile with constantly reducing scale. However, when object scale changes, it is difficult to make automatic adjustment. The sequence has a total length of 299 frames. In video sequence before about frame 266, traditional Mean-Shift algorithm can track the object very well. Fig. 4 shows its tracking result, respectively the tracking result in frames 10, 19, 54, 252, 266 and 270.

According to Fig. 4, Mean-Shift algorithm has good tracking effect in the video before frame 266. The improved Mean-Shift algorithm can solve the problem of lost object tracking effectively. Fig. 6 shows its tracking result.

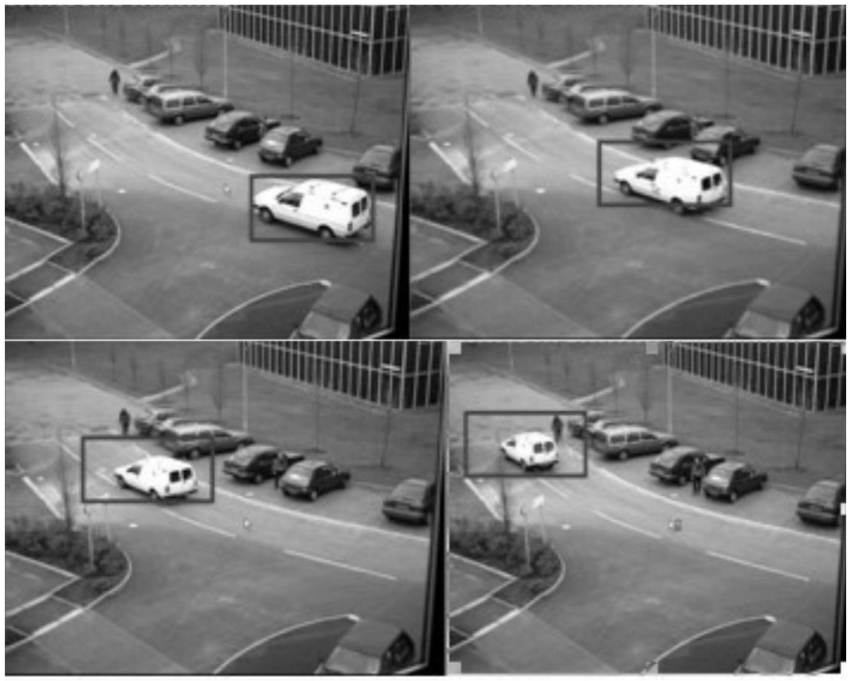


Fig. 2. Automobile tracking effect of traditional Mean-Shift algorithm

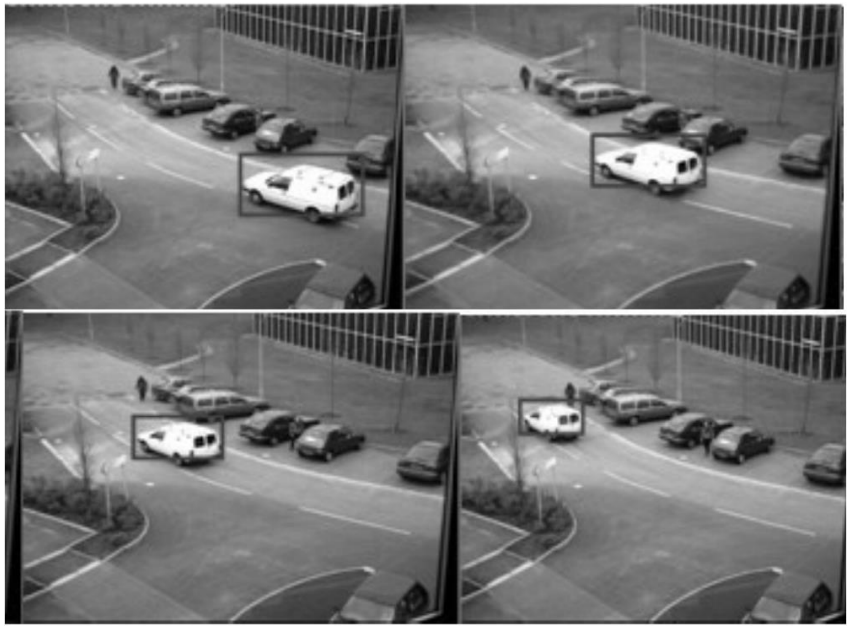


Fig. 3. Automobile tracking effect of improved Mean-Shift algorithm

Fig. 4. Video tracking result of Mean-Shift algorithm

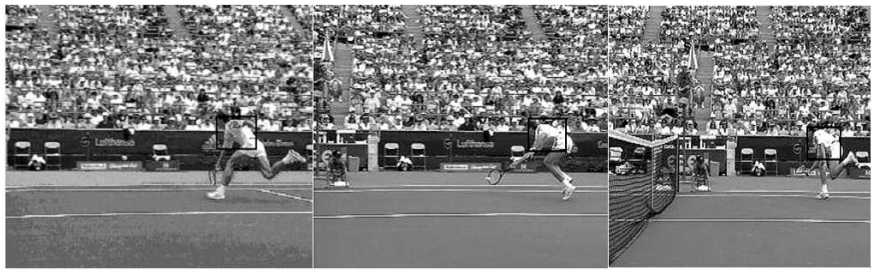


Fig. 5. Video tracking result of improved Mean-Shift algorithm

5.2. Multi-test tracking

Test videos in in Figs. 6 and 7 are in PETS2000 format and video objects are multi-object moving automobiles.

Fig. 6. Automobile tracking effect of traditional Mean-Shift algorithm

Fig. 7. Automobile tracking effect of improved Mean-Shift algorithm

According to Fig. 6, traditional Mean-Shift algorithm has some errors in multi-object tracking, which have a high frequency of occurrence. Fig. 6, left part, shows missed detection in the case of blob overlap. Fig. 6, middle part, refers to the identification of multiple vehicles into single vehicle in the case of unstable foreground and increase of search region of the algorithm. Figure 6, right part, shows object overlap when the tracking object is much larger than the actual object.

According to Fig. 7, compared to traditional Mean-Shift algorithm, problems in Fig. 6 are well solved. The improved Mean-Shift algorithm based on blob can not only achieve multi-object tracking, but also have more accurate tracking effect.

In allusion to object occlusion problem, football video sequence is used as the research object in order to verify the effectiveness of the algorithm in this paper. Figure 8 shows the tracking result of the algorithm in this paper which can solve the problem of number occlusion effectively.

According to the comparison of analysis results before and after improvement in Tables 1 and 2, the algorithm in this paper can reduce the number of iterations of Mean-shift in each frame and the total computation time of each frame is reduced relatively. The average running time of first 22 frames is 0.0185 s and the corresponding average time of Mean-Shift algorithm is 0.0284 s.

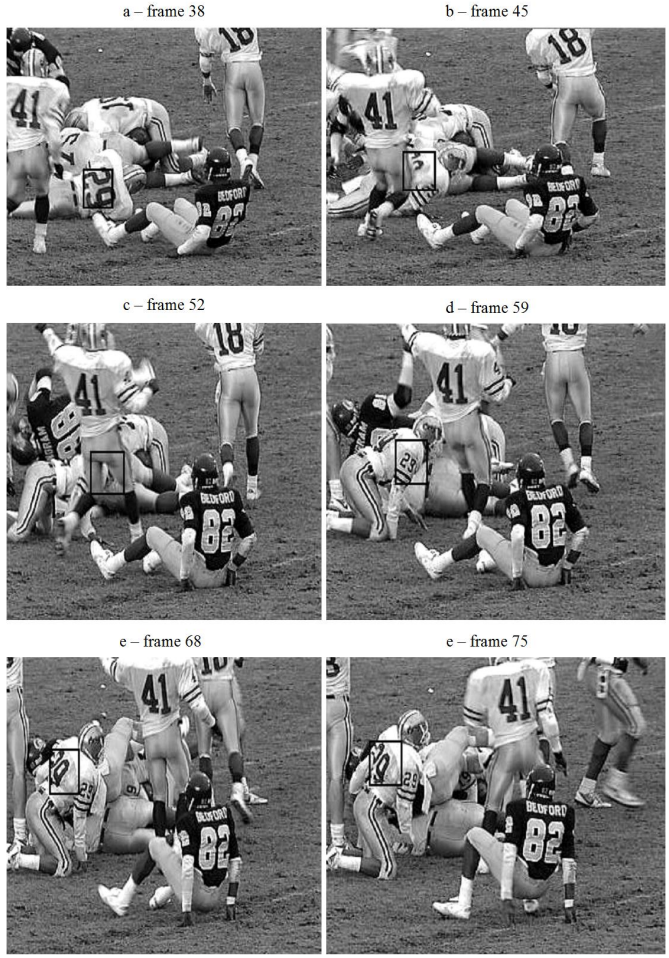


Fig. 8. Result of occlusion tracking of improved Mean-Shift algorithm

Table 1. Analysis result of improved Mean-Shift algorithm

Frame	2	3	4	5	6	7	8	9	10	11
Number of iterations	2	4	8	10	3	3	3	4	3	3
Time	0.016	0.016	0.032	0.046	0.016	0.015	0.016	0.015	0.016	0.015
Frame	12	13	14	15	16	17	18	19	20	21
Number of iterations	3	3	3	2	2	4	3	3	3	6
Time	0.016	0.015	0.016	0.015	0.016	0.017	0.016	0.016	0.016	0.022

Table 2. Analysis result of Mean-Shift algorithm

Frame	2	3	4	5	6	7	8	9	10	11
Number of iterations	3	3	18	4	3	3	6	5	3	4
Time	0.032	0.015	0.094	0.031	0.016	0.032	0.047	0.031	0.015	0.031
Frame	12	13	14	15	16	17	18	19	20	21
Number of iterations	3	3	4	3	2	5	4	4	2	3
Time	0.015	0.015	0.032	0.016	0.016	0.032	0.031	0.031	0.016	0.015

6. Conclusion

In allusion to defects of traditional Mean-Shift algorithm, this paper introduces blob detection into Mean-Shift algorithm, puts forward vehicle video tracking algorithm of an improved Mean-Shift algorithm and achieves accurate and fast positioning of single object and multiple objects. The experimental result shows that Mean-Shift algorithm based on blob can not only achieve multi-target tracking, but also have more accurate tracking effect. Its tracking effect is obviously better than that of traditional Mean-Shift algorithm.

References

[1] Z. Lv, A. Tek, F. Da Silva, C. Empereur-mot, M. Chavent, M. Baaden: *Game on, science - how video game technology may help biologists tackle visualization challenges.* PloS One *8* (2013), No. 3, e57990.

[2] W. T. Pan: *A new fruit fly optimization algorithm: Taking the financial distressmodel as an example.* Knowledge-Based Systems *26* (2012), 69–74.

[3] Y. Chen, Z. Cao: *An improved MRF-based change detection approach for multitemporal remote sensing imagery.* Signal Processing *93* (2013), No. 1, 163–175.

[4] G. F. Lu, Z. Lin, Z. Jin: *Face recognition using discriminant locality preserving projections based on maximum margin criterion.* Pattern Recognition *43* (2010), No. 10, 3572–3579.

[5] L. Ong, M. Motani: *Optimal routing for decode-forward in cooperative wireless networks.* IEEE Transactions on Communications *58* (2010), No. 8, 2345–2355.

[6] J. Yao, S. Feng, X. Zhou, Y. Liu: *Secure routing in multihop wireless ad-hoc networks with decode-and-forward relaying.* IEEE Transactions on Communications *64* (2016), No. 2, 753–764.

[7] K. Wang, K. Qiao, I. Sadooghi, X. Zhou, T. Li, M. Lang, I. Raicu: *Load-balanced and locality-aware scheduling for data-intensive workloads at extreme scales.* Concurrency and Computation Practice and Experience *28* (2016), No. 1, 70–94.

[8] Y. Lin, J. Yang, Z. Lv, W. Wei, H. Song: *A self-assessment stereo capture model applicable to the internet of things.* Sensors (Basel) *15* (2015), No. 8, 20925–20944.

[9] Y. Xie, Y. He, A. Cheng, J. Zhang: *Study on medical image enhancement based on*

IFOA improved grayscale image adaptive enhancement. Multimedia Tools and Applications *75* (2016), No. 22, 14367–14379.

[10] H. Li, H. S. Yang: *Fast and reliable image enhancement using fuzzy relaxation technique.* IEEE Transactions on Systems, Man, and Cybernetics *19*, (1989), No. 5, 1276 to 1281.

[11] H. R. Tizhoosh, G. Krell, B. Michaelis: *Knowledge-based enhancement of megavoltage images in radiation therapy using a hybrid neuro-fuzzy system.* Image and Vision Computing *19* (2001), No. 4, 217–233.

[12] S. K. Pal, R. A. King: *Image enhancement using smoothing with fuzzy sets.* IEEE Transactions on Systems, Man, and Cybernetics *11* (1981), No. 7, 494–501.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 51–62

The analysis and optimization strategy of network failure recovery

Changtian Ying[1,2], Weiqing Wang[1], Jiong Yu[2], Hong Jiang[2], Lei Qi[2]

Abstract. In the traditional network failure mechanism, the programmer takes the responsibility to select the network checkpoints, which may lead to the problem of unavailable service and much more recovery time. To address this issue, we first analyzed the architecture of network failure in this paper, established the fault tolerance model on the basis of network characteristics, and then proposed the optimization strategy for network failure including the checkpoint algorithm and the recovery algorithm. The checkpoint algorithm chose the appropriate checkpoints based on the analysis of the network, and the recovery algorithm took advantage of the latest checkpoints to recover the lost data. Finally, we conducted the experiments to evaluate, and both of the two datasets had the less recovery time and better recovery ratio. The experiment results verified the validity of the failure recovery strategy.

Key words. Network optimization, unavailable service, failure recovery.

1. Introduction

The past few years have seen a major change in computing systems, as growing data volumes and stalling processor speeds require more and more applications to scale out to distributed systems. To face the challenges brought by the big data [1, 2], data processing technology has come under heated discussion among domestic scholars in recent years. Spark has won more and more recognition and support in the new generation of large data processing framework. It is a general-purpose high-performance parallel computing framework. Spark uses flexible distributed datasets that are called RDD (resilient distributed datasets) as the data structure. If a partition of an RDD is lost, Spark read the checkpoint data, and uses the linage to re-compute that partition.

The checkpoint/recovery strategy is fault-tolerant technology, which has been

[1]Post Doctoral Research Station of Electrical Engineering, Xinjiang University, Urumqi, Xinjiang, 830046, China
[2]School of Software, Xinjiang University, Urumqi, Xinjiang, 830046, China

widely used in cluster computing. Checkpoint is a common traditional strategy in the domestic and foreign research [3]. The classifications of checkpoint technology are the system level, the application level and the user level [4]. Literature [5] analyzed the characteristics and advantages of different examination method, and proposed an application level check pointing technology. And another method is presented to implement a system level checkpoint [6]. This literature proposed automatic fault tolerance technology based on high performance computing system, which can complete the automatic check pointing/recovery [7]. Literature [8] proposed checkpoint/restore library, which could support the different ways of preserving the checkpoint data in the memory. Literature [9] used mixed strategy to fulfill incremental checkpoint, thereby reducing the amount of checkpoint data. Fault tolerance strategy of literature [10] used multilevel storage and encoding redundancy checkpoints, so as to reduce the fault tolerance overhead.

However, in the Spark checkpoint mechanism, the programmers make the decision for selecting the object and time of checkpoints. The checkpoint is set up only when the checkpoint instruction is executed. This implementation increases the uncertainty of checkpoint strategy, and it is much difficult to maximize the checkpoint performance. If the checkpoint strategy is unsuitable, it may not only reduce the application efficiency, but also increase the risk of application exception. Therefore, if the failure optimization algorithm can efficiently implemented, it can ease the burden of programmers, and improve system efficiency and availability.

In this paper, we have conducted a sophisticated theoretical and technical study on checkpoint technology. First we established the task scheduling and recovery efficiency model, then analyzed relevant factors and established the RDD weight model. Then we proposed optimization strategy to max-relieve checkpoint bottleneck and optimize the performance. The experiments verified the effectiveness of the strategy.

2. Problem analysis

Due to the lazy scheduling mechanism, the task is compiled into multiple DAGs when performs action. And each RDD is divided into several partitions to be calculated by the cluster nodes.

Definition 1 - Partition failure rate. RDD partitions are computed in parallel. When considering the hardware failure, partition failure rate is determined by worker failure rate. If the worker failure rate of w_m is wf_m, then the failure rate of the partition PT_{ijk} is denoted as

$$\mathrm{FR}_{\mathrm{PT}_{ijk}} = pb_{ijkm} wf_m \,. \tag{1}$$

Definition 2 - RDD failure rate. If the RDD partition is unavailable, the entire RDD cannot be used. So the failure rate of RDD_i is defined as

$$\mathrm{FR}_{\mathrm{RDD}_{ij}} = \max(\mathrm{FR}_{\mathrm{PT}_{\mathrm{ij1}}},\ \mathrm{FR}_{\mathrm{PT}_{\mathrm{ij2}}},\ \mathrm{FR}_{\mathrm{PT}_{\mathrm{ij3}}},\ \cdots,\ \mathrm{FR}_{\mathrm{PT}_{\mathrm{ijk}}}) \,. \tag{2}$$

Definition 3 - Task failure rate. When the task is executed, it is compiled into a RDD DAG, and the failure rate of the task is denoted as

$$\mathrm{FR}_{\mathrm{Task}_i} = 1 - \prod_{j=1}^{m} (1 - \mathrm{FR}_{\mathrm{RDD}_{ij}}) \,. \tag{3}$$

Definition 4 - Partition recovery cost. Recovering of missing partition needs the parent partitions and checkpoints. If the ancestor PT_{ijp} has been set as checkpoint, then the recovery cost of the partition PT_{ijk} can be defined as

$$R(\mathrm{PT}_{ijk}) = \alpha_i + \mathrm{read}(\mathrm{PT}_{ij(k-1)}) + \mathrm{proc}(\mathrm{PT}_{ijk}) \,,$$

$$R(\mathrm{PT}_{ij(k-1)}) = \alpha_i + \mathrm{read}(\mathrm{PT}_{ij(k-2)}) + \mathrm{proc}(\mathrm{PT}_{ij(k-1)}) \,,$$

$$R(\mathrm{PT}_{ij(p+1)}) = \alpha_i + \mathrm{read}(\mathrm{PT}_{ijp}) + \mathrm{proc}(\mathrm{PT}_{ij(p+1)}) \,.$$

Assume $\mathrm{T}_{\mathrm{PT}_{ijk}}$ as the time cost of PT_{ijk} and $T_{\mathrm{PT}_{ijk}} = \mathrm{read}(\mathrm{PT}_{ijk}) + \mathrm{proc}(\mathrm{PT}_{ijk})$. Then

$$R(\mathrm{PT}_{ijk}) = \alpha_i + \sum_{l=p}^{k} \mathrm{T}_{\mathrm{PT}_{ijk}} \,. \tag{4}$$

Here, α_i denotes the fault detection overhead. If the checkpoints are not set, all RDDs are calculated from the beginning.

Definition 5 - RDD recovery cost. Suppose the checkpoints as the set $C_i = (c_{i1},\ c_{i2},\ ...,\ c_{ip})\}$, where C_i is a subset of Task_i, c_{ip} is the newest checkpoint, and RDD_{ik} is the kth RDD of Task_i. The worker is failed while computing RDD_{ij}, then the recovery cost can be denoted as

$$\begin{aligned} R_{\mathrm{RDD}_{ij}} &= \alpha_i + \mathrm{read}(\mathrm{RDD}_{i(k+1)}) + \mathrm{proc}(\mathrm{RDD}_{i(k+1)}) + \mathrm{read}(\mathrm{RDD}_{i(k+2)}) + \\ &+ \mathrm{proc}(\mathrm{RDD}_{i(k+2)}) + \cdots + \mathrm{read}(\mathrm{RDD}_{ij}) + \mathrm{proc}(\mathrm{RDD}_{ij}) = \\ &= \alpha_i + T\mathrm{RDD}_{i(k+1)} + T\mathrm{RDD}_{i(k+2)} + \cdots + T\mathrm{RDD}_{ij} = \\ &= \alpha_i + \sum_{q=k+1}^{j} T\mathrm{RDD}_{iq} \,. \end{aligned} \tag{5}$$

where α_i denotes the fault detection overhead.

$$R_{\mathrm{RDD}_{ij}} = \alpha_i + \sum_{q=1}^{j} T_{\mathrm{RDD}_{iq}} \,. \tag{6}$$

Definition 6 - task recovery cost. In the failure case, the task recovery cost during the task execution process is the completion overhead for the task, and the overhead for the lost RDDs to recovery. If the number of failure times is 0, R_{task_i} is the recovery overhead of Task_i. If the number of failure times is k, then fault recovery

overhead is used by the lost RDDs.

$$R_{\text{task}_i} = \sum_{j=1}^{k} R_{\text{RDD}_{ij}} = \sum_{j=1}^{k} (\alpha_i + \sum_{q=1}^{j} T_{\text{RDD}_{iq}}) \,. \tag{7}$$

Definition 7 - RDD life cycle. For RDD_{ij}, the maximum life cycle is from the starting time of RDD_{ij} to the completion time of task. Denote $ST_{\text{RDD}_{ij}}$ as starting computing time of RDD_{ij}, FT_{task_i} as the completion time for the task, then the maximum life cycle of the RDD_{ij} can be defined as

$$\max LC_{\text{RDD}_{ij}} = FT_{\text{task}_i} - ST_{\text{RDD}_{ij}} \,. \tag{8}$$

For RDD_{ij}, the minimum life cycle is from the start time of RDD_{ij} to the last used completion time of RDD_{ij}. Denote $UT_{\text{RDD}_{ij}}$ as RDD_{ij} of the last used completion time

$$\min LC_{\text{RDD}_{ij}} = UT_{\text{RDD}_{ij}} - ST_{\text{RDD}_{ij}} \tag{9}$$

So the range of the RDD life cycle is expressed as

$$\min LC_{\text{RDD}_{ij}} \leq LC_{\text{RDD}_{ij}} \leq \max LC_{\text{RDD}_{ij}} \,. \tag{10}$$

By analyzing the availability, it is easy to know that the failure probability, the selection of the checkpoint time and objects are important factors influencing the recovery efficiency. Suppose the fault probability does not change the situation, then the recovery overhead of the task R(task) is smaller, the task availability is greater. Therefore, the goal of automatic checkpoint strategy is to minimize task recovery overhead while meeting the requirement of system resources. And it can be formalized as

Object: $\min(R_{\text{task}})$, s.t. $\sum_{i \in \text{Tasks}} A_{im} \leq r_m$.

3. Optimization strategy

3.1. Relevant proof

Theorem 1. It is difficult to choose proper period time according to the time.

Proof: select the period checkpoint time need considering the user experience judgment and the prediction task execution time. For the checkpoint time T_i, there are three kinds of states.

1. T_i is too small, $T_i < \min T_{\text{RDD}}$ (RDD being the minimum running time). It may cause frequently storing data to disk and reducing system throughput.

2. T_i is too large, $T_i > \max T_{\text{RDD}}$ (RDD maximum running time). The time interval of the storing data is too long, if the worker power down, it may lead to a large number of data RDD necessary for recalculation.

3. T_i is moderate, $\min T_{\text{RDD}} < T_i < \max T_{\text{RDD}}$. The time interval is moderate, but with the current state of the system, may not be a good fit with the task.

Theorem 2. The long lineage principle. The longer the RDD lineage is, the

greater its recovery overhead is. And it should be gave higher priority.

Proof: Task_i has two RDDs: $\text{RDD}_{i(j-1)}$ and RDD_{ij}. $\text{RDD}_{i(j-1)}$ is father RDD of RDD_{ij}. Suppose the lineage depth of $\text{RDD}_{i(j-1)}$ is k, the lineage depth of RDD_{ij} is $k+1$, then RDD_{ij} has a longer lineage. Denote RDD_{ip} the latest checkpoint, then recovery overheads of two RDDs are

$$R_{\text{RDD}_{i(j-1)}} = \alpha_i + \sum_{q=p+1}^{j} T_{\text{RDD}_{iq}} ,$$

$$R_{\text{RDD}_{ij}} = \alpha_i + \sum_{q=p+1}^{(j-1)} T_{\text{RDD}_{iq}} . \tag{11}$$

As $R_{\text{RDD}_{ij}} - R_{\text{RDD}_{i(j-1)}} = T_{\text{RDD}_{ij}}$, therefore $R_{\text{RDD}_{ij}} > R_{\text{RDD}_{i(j-1)}}$.

That is, the RDD recovery overhead with longer lineage is greater. When the checkpoint is set, the RDD with long lineage should be selected to reduce the recovery overhead.

Theorem 3. Wide dependency principle. The recovery overhead of narrow dependency is larger than that of the wide dependency.

Proof: Recovery overhead from the parent $\text{RDD}_{i(j-1)}$ to RDD_{ij} is

$$R_{\text{RDD}_{ij}} = \alpha_i + \sum_{l=p}^{m} T(PT_{ijl}) . \tag{12}$$

When the parent data RDD_{ij} is fixed, the difference is that the operation is of wide or narrow dependency. If the operation is narrow dependency and the lth partition PT_{ijl} is lost, only the parent of PT_{ijl} is calculated.

$$R_{\text{RDD}_{ij}}(\text{narrow}) = R(PT_{ijl}) = \alpha_i + \text{read}(PT_{ij(l-1)}) + \text{proc}(PT_{ijl}) . \tag{13}$$

If the operation is wide dependency and the lth partition PT_{ijl} is lost, the partition is calculated by all the parent partitions:

$$R_{\text{RDD}_{ij}}(\text{wide}) = R_{\text{RDD}_{ij}} = \max(R(PT_{ij1}), \cdots, R(PT_{ijk}) . \tag{14}$$

Therefore

$$R_{\text{RDD}_{i}j}(\text{wide}) \geq R_{\text{RDD}_{i}j}(\text{narrow}) .$$

We should preferred RDD with wide dependency as checkpoints, thereby reducing the recovery overhead.

Theorem 4. High computation cost principle. If the RDD with higher computation cost is not stored, it may lead higher recovery overhead. So it is necessary to give priority to the RDD with higher cost.

Proof: For Task_i, in the case of the same parameters, when the cost of the RDD calculation is not the same size, the impact on the recovery of the RDD overhead is different.

When the restore is required and the latest checkpoint is RDD_{ip}, the recovery of RDD_{ij} overhead is

$$R_{\mathrm{RDD}_{ij}} = \alpha_i + \sum_{q=1}^{j} T_{\mathrm{RDD}_{iq}} = \alpha_i + \sum_{q=1}^{j-1} T_{\mathrm{RDD}_{iq}} + T_{\mathrm{RDD}_{ij}} \,. \tag{15}$$

While the other parameters are fixed, if the computation cost of RDD_{ij} is greater, the recovery overhead is greater. Therefore, RDD with greater computation cost should be preferred as checkpoint, thereby reducing the recovery overhead.

3.2. RDD weight model

Based on the above analysis, the recovery overhead key factors are: 1) RDD lineage length, 2) the complexity of operation, 3) RDD computation cost.

Definition 8 - depth of RDD. Depth(RDDij) denotes as the RDD lineage length, which is the layer number for RDD_{ij} in the directed acyclic graph (DAG). Assumed the first layer RDD, depth is defined as 1, and the depth of the last RDD is the maximum depth m of directed acyclic graph (DAG).

Assume two RDD_{ix} and RDD_{iy} have different depths, the depth as $\mathrm{Depth}(\mathrm{RDD}_{ix})$, $\mathrm{Depth}(\mathrm{RDD}_{iy})$, respectively. If $\mathrm{Depth}(\mathrm{RDD}_{ix})$>$\mathrm{Depth}(\mathrm{RDD}_{iy})$, then the lineage of RDD_{ix} is longer than that of RDD_{iy}.

Definition 9 -the operation complexity. RDD operation divides into two kinds, narrow dependency and wide dependency. Recovery overhead of wide dependency is larger, and is relevant to the partitions number. Denote the operation complexity as $\mathrm{OC}_{\mathrm{RDD}_{ij}}$, and RDD_{ij} has k partitions. When the RDD_{ij} operation is narrow dependency, the operation complexity is defined as 1. On the contrary, wide dependency occurs, when the operation complexity is defined as k.

Definition 10 - computation overhead. According to definition, PT_ijk computation cost requires a comprehensive assessment of the data acquisition cost, the data processing cost, and the evaluation of algorithms. It is difficult to predict it. But we could easily get the start time and completion time of RDD_{ij}. So the computation cost can be expressed as:

$$\mathrm{Cost}_{\mathrm{RDD}_{ij}} = \mathrm{FT}_{\mathrm{RDD}_{ij}} - \mathrm{ST}_{\mathrm{RDD}_{ij}} \,. \tag{16}$$

Definition 11 - RDD weight. The weight of RDD is expressed as follows:

$$\mathrm{CR}_{\mathrm{RDD}_{ij}} = \alpha \times \mathrm{Depth}_{\mathrm{RDD}_{ij}} + \beta \times \mathrm{OC}_{\mathrm{RDD}_{ij}} + \gamma \times \mathrm{Cost}_{\mathrm{RDD}_{ij}} \,. \tag{17}$$

Here, $0 \leq \alpha, \beta, \gamma \leq 1$, and $\alpha + \beta + \gamma = 1$. When $\alpha = 1$, the weight is determined by RDD lineage depth. When $\beta = 1$, the key is up to the operation complexity. When $\gamma = 1$, the key is decided by computation overhead.

4. Optimization processes

Before performing the task, traverse the DAG of the task. Then get the operation and the properties of each RDD. After analyze the DAG and the current implementation of the progress, calculate the RDD weight.

According to the RDD weight, select the RDDs as checkpoint to perform. Checkpoint time began from the first generation of RDD to the latest generation of RDD. During the task execution, comparison of the new generation of multiple RDD, select the largest RDD as the checkpoint.

After downtime, the spark performs the recovery operation, and recovers from the latest checkpoint RDD. The latest checkpoint is read into memory, thereby reducing the recovery and execution overhead. When a RDD need recover, re-execute the lineage and recovery through its parent node.

To complete the processing tasks, the recovery strategy steps are:

1) Choose free worker, if the current has no idle worker, waiting for the worker assignment.

2) Receive a checkpoint sequence.

3) It has been lost RDD sequence.

4) Determine the need to use what RDD by lineage.

5) If the operation is of wide dependency, or lost the RDD all partitions, you need to read the RDD all partitions to memory.

6) If it is the loss of RDD partition, and the lineage without wide dependency, only need to read the lost partition checkpoint into memory.

7) Read the checkpoint into memory, and compute by the lineage.

5. Experimental and evaluation

5.1. Experiment environment

This section will be compared and evaluated by experiments, which verify the validity of the checkpoint automatic selection algorithm and the recovery algorithm.

The experiments perform with a master server and eight computing nodes as the master and Hadoop namenodes of Spark, and the nodes configuration is shown in Table 1.

5.2. The optimization strategy

5.2.1. Different failure rate. Figure 1 indicates that the iteration time is 1–10, when the failure rate is $fr = 0.375$ and 0.5, the algorithm uses different data sets, and compared the failure recovery algorithm.

As shown in Fig. 1, with the increase of failure rate, task execution time also increases. This is because of the failure rate is high, which means the more node failure. Therefore, in order to restore the RDD, it is need more time overhead to recalculate the corresponding. Comparison of different data sets, and Web-Google Wiki-Talk under different algorithms, Wiki-Talk has bigger time overhead. This is

due to the size difference computation.

Table 1. Configuration parameters

Parameters	Values
CPU	Intel CORE i7/2.2 GHz
RAM	1 GB
Hard Disk	200 GB/SATA3.0
OS	CentOS 6.4
Spark	Apache Spark 1.4.1
Hadoop	Apache Hadoop 2.6
Scala	Scala-2.10.4
JDK	OpenJDK 1.8.0 25

Comparing the number of iteration, it can be seen that when the iteration number is 1, the task execution time of failure recovery algorithm or system recovery algorithm is basically the same. This is because the check pointing algorithm is set to complete the checkpoint, so the failure recovery algorithm does not affect task. And with the increasing in the number of iterations, the execution speed of failure recovery is obviously better than that of system recovery algorithm. The system recovery algorithm recomputed RDD from the beginning. Then the bigger the iteration time is, the longer recovery time is. Therefore, if the number of iterations is small, the user can use the lineage to recover and does not need to set up checkpoints to improve the efficiency.

5.2.2. Accelerate the recovery ratio. Figure 2 indicates that the algorithm uses different data sets, the average recovery time and recovery speedup ratio of the failure recovery algorithm.

Figure 2 shows that the comparison of different iterations, with the increasing in the number of iterations, the original Spark by lineage to recover the time cost is larger. When the node fails, it will result in the loss of RDD partitions. Spark task will perform these tasks concurrent on other machines. The task reread the input data, and reconstruction RDD based on the lineage. The longer the iteration and computing time is, the greater recovery time overhead is. Comparing the different data sets, the recovery of Wiki-Talk speedup more, because it needs more computation, The execution time is longer, so the overhead of the recovery is larger, and the recovery cost can be significantly reduced by using checkpoint recovery algorithm.

Then there were compared Fig. 1 and Fig. 2 comprehensively, analyzed task execution time, recovery time and recovery speedup ratio. Although the algorithm may increases the amount of time overhead, compared to the traditional policy uncertainty and even abnormal risk, the extra overhead is worth it. On the basis of automatic checkpoint algorithm, failure recovery algorithm considers not only the RDD lineage length, but also the computation cost and operation complexity. The greater the weight of RDD is, the higher checkpoint priority is. It will make the task to minimum the overall computational cost, so as to improve the recovery efficiency

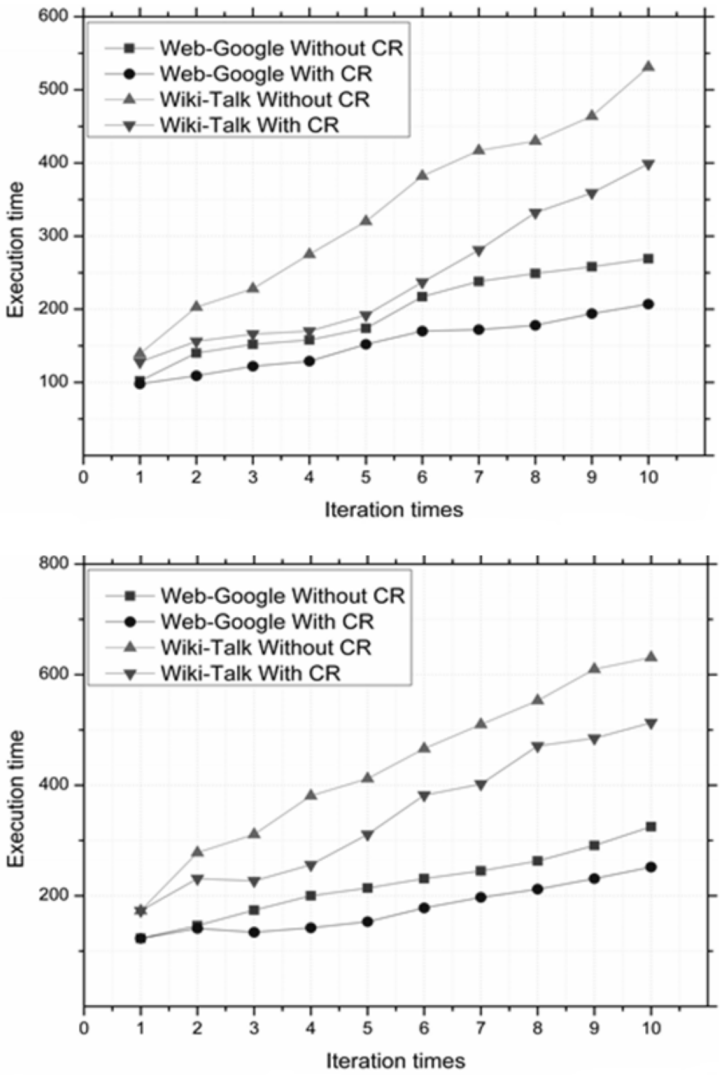


Fig. 1. Execution time (s) of CR(s): up–$fr = 0.375$, bottom–$fr = 0.5$

of tasks. Therefore, checkpoint selection algorithm cannot significantly affect the performance of the Spark system under the condition of enhancing system stability and reliability.

6. Conclusion

The traditional network failure recovery strategies set the checkpoints depending on the experiences of the programmers, which may cause larger execution time and

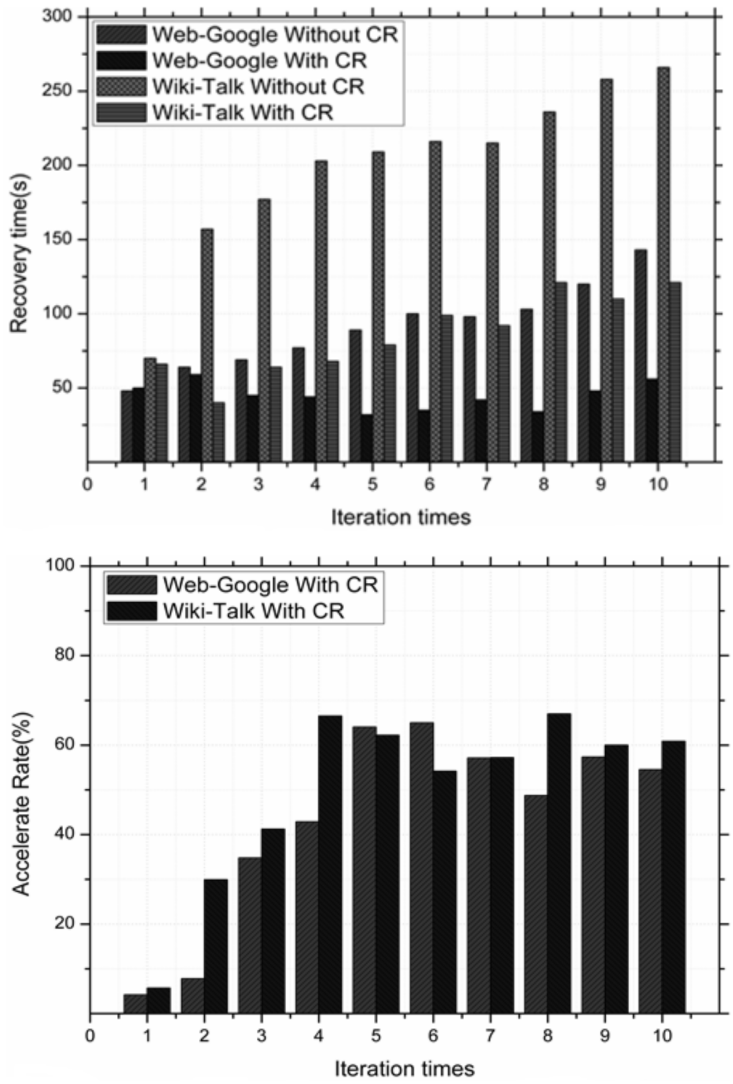


Fig. 2. Average recovery efficiency of CR: up–recovery time (s), bottom–accelerate rate (%)

lower efficiency. To address the issue, we analyzed task execution mechanism, and established the task execution efficiency model. Then put forward the RDD weight model, which provided a theoretical basis for the strategy presented. Experiments were conducted with different data sets and failure rate. And the results demonstrated that the strategy can improve the recovery efficiency and the utilization of system resources at the same time. Especially with the rising in large data analysis, network optimization has become increasingly prominent, and network failure recovery is one of the key problems to be addressed.

However, due to the limited capacities, the research should further be conducted in the future. Our work will mainly focus in the following aspects:

(1) Analyze in multiple checkpoints failure and different recovery strategy for network efficiency.

(2) With the decreasing cost of network devices, using new medium to enhance recovery efficiency becomes feasible.

(3) By constructing a multi-level failure tolerance network to improve the performance of the system is a future research direction.

References

[1] Z. Zheng, P. Wang, J. Liu, S. Sun: *Real-time big data processing framework: challenges and solutions.* Applied Mathematics & Information Sciences *9* (2015), No. 6, 3169–3190.

[2] O. Shamir, N. Srebro, T. Zhang: *Communication efficient distributed optimization using an approximate Newton-type method.* International conference on machine learning (ICML), 21–26 June 2014, Beijing, China (2014), 1000–1008, arXiv:1312.7853v4 [cs.LG] 13 May 2014.

[3] E. Yildirim, T. Kosar: *End-to-end data-flow parallelism for throughput optimization in high-speed networks.* Journal of Grid Computing *10* (2012), No. 3, 395–418.

[4] Z. Kozhirbayev, R. O. Sinnott: *A performance comparison of container-based technologies for the cloud.* Future Generation Computer Systems *68* (2017), 175–182.

[5] Z. Xiao, W. Song, Q. Chen: *Dynamic resource allocation using virtual machines for cloud computing environment.* IEEE Transactions on Parallel and Distributed Systems *24* (2013), No. 6, 1107–1117.

[6] H. Shen, G. Liu: *An efficient and trustworthy resource sharing platform for collaborative cloud computing.* IEEE Transactions on Parallel and Distributed Systems *25* (2014), No. 4, 862–875.

[7] I. P. Egwutuoha, D. Levy, B. Selic, S. Chen: *A survey of fault tolerance mechanisms and checkpoint/restart implementations for high performance computing systems.* The Journal of Supercomputing *65* (2013), No. 3, 1302–1326.

[8] I. Cores, G. Rodríguez, M. J. Martín, P. González, R. R. Osorio: *Improving scalability of application-level checkpoint-recovery by reducing checkpoint sizes.* New Generation Computing *31* (2013), No. 3, 163–185.

[9] S. H. Lim, S. Lee, B. H. Lee, S. Lee, H. W. Lee: *Stochastic method for power-aware checkpoint intervals in wireless environments: Theory and application.* Journal of Industrial and Management Optimization *8* (2012), No. 4, 969–986.

[10] K. B. Ferreira, R. Riesen, P. Bridges, D. Arnold, R. Brightwell: *Accelerating incremental checkpointing for extreme-scale computing.* Future Generation Computer Systems *30*, (2014), 66–77.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 63–74

Research on rumors of law and countermeasures based on big data analysis[1]

Liu Xiaoming[2]

Abstract. The spread of rumors in the era of large data transmission shows the new features with diversification of communication channels, the speed of transmission, the interference of noise and the complexity, which has a wide range and broad harm. In order to avoid a series of sudden group events which can be triggered by rumors and avoid the social harm and economic loss caused by a wider range or longer, this paper analyzes the media reports and rumors based on the data analysis technology. Based on this, the rumor spread model used for considering the role of media is established. And then the simulation is carried out to study the influence of media reports on the spread of rumors. The results show that the interaction mechanism between media reports and rumor spread is highly complex. Finally, the strategic suggestion for the spread of rumors is proposed from different aspects by adjusting the parameters under different factors

Key words. Big data, spreading of rumors, media report.

1. Introduction

Rumors are generally defined as unexplained statements or annotations spreading through various channels. They are usually things or problems that can arouse public interests. It is an important form of social interaction without difference between right and wrong, which has greatly affected people's daily work and life. Currently, China has already into the background of large data dissemination, technical development and changes in the way the audience participation, which has a big impact on the dissemination of information. With the changes in the way of communication, the authority of the mainstream discourse is weakened. The most direct result is that the cycle of rumors processing is shorter and shorter, and the controllability is getting lower and lower, which becomes a big challenge for the government to guide the healthy development of public opinion. In this paper, we

[1]The author acknowledges the National Natural Science Foundation of China (Grant 51578109 and Grant 51121005).

[2]Yangtze University College of Arts and Science, Jingzhou, Hubei, 4340201, China

explore the law of rumor spread based on the large data analysis and put forward specific countermeasures to provide reference and suggestions for our response to emergencies.

2. Experimental procedure

2.1. Study on the influence of media report on rumor transmission

Media coverage plays a different role in different stages of rumor spread. In rumors spread latency, it can arouse public attention and increase public knowledge, which plays an early warning role. The public's emotions are often easy to intensify when the spread of rumors into the outbreak and the climax period. During this time, the public is in urgent need of official media coverage of the real information, and media reports play a guiding role in public opinion. In the spread of rumors dissipation, the public's emotional change, instability and media coverage play a role in regulating, stabilizing and appeasing the public uneasiness. However, media coverage is not always positive. When the media credibility is not high, and the release frequency is too high, it can easily lead to panic. For example, during the SARS, the government response is not fast enough, and emergency measures are not strong. In addition, the expert voice is too weak, which leads to a variety of rumors and finally triggers a number of snapping up event [1].

At a macro level, media coverage may change the state of the public. New populations with self-protection and people who are skeptical of rumors are appeared. These transmission rates are significantly lower than those without self-protection blind communicator. Misra and others think that the warning role of media is mainly expressed on the group of "easy to dye" state. The people with awareness will no longer contact with the infected state. Obviously, this assumption is too idealistic. Funk and other experts pointed out that media coverage can enhance public scepticism and raise public awareness of self-protection and rational consciousness. These suggestions can enhance public recognition and remind them no longer blindly believing in the rumors. Meanwhile, the public will be treated with caution to avoid economic losses. These findings provide a solid theoretical basis for further study of the impact of media coverage on rumors [2].

2.2. Research on rumor spread patterns considering the media reporting

First, model assumptions and generation rules will be mentioned. Based on the research results of Misra and Funk, this paper explores the interaction mechanism between media coverage and rumor transmission, and establishes the rumor spread dynamics model considering media coverage. The whole population is divided into four categories according to their state: ignorant (I), communicator (S), skeptics (H) and immunized (R). According to the strength of self-protection consciousness, the ignorant and the communicator are divided into two categories: the ignorant I_{m}

with self-protection consciousness and the ignorant I_{u} without self-protection consciousness, the communicator S_{m} with self-protection consciousness and the communicator S_{u} without self-protection consciousness. Symbol H represents the number of suspects, and R represents the number of immunized persons. The variable M_{e} indicates the number of relevant events reported by the media. These media reports can improve self-protection awareness and rational awareness. It also can reduce crowd panic and herd behavior. Table 1 gives the meaning of the basic parameters in the model [4].

Table 1. Meaning of parameters in the model

α	Doubt rate coefficient
γ	Immunization rate coefficient
v	Self-protection awareness coefficient
β_{m}	Media report infectious rate coefficient
β_{u}	Rumor transmission rate coefficient
u_{i}	Ignorant alertness loss rate coefficient
u_{s}	Communicator's alertness loss coefficient
δ_{i}	Effective contact rate coefficient of ignorant with self-protection awareness
δ_{s}	Effective contact rate coefficient of communicator with self-protection awareness
p_{s}	The impact coefficient of the communicator on the media coverage
p_{h}	The impact coefficient of the reporter on the media coverage
p	Media report consumption rate

The basic assumptions about the dynamics of media coverage and rumors are as the follows:

(1) Media coverage is affected by the number of communicators and suspects. On the one hand, the rate of change in the number of media reports is proportional to the number of communicators $(S_{\mathrm{u}} + S_{\mathrm{m}})$. The influence coefficient is p_{s}. On the other hand, the rate of change in the number of media reports is proportional to the number of skeptics H with the speed of proportionality coefficient p_{h}. With the pass of time and the release of emergency information, the rate of change in the number of media reports has gradually become slow with a speed of the consumption rate p. Therefore, p_{s} and p_{h} can be considered as the influence of communicators and suspects on the media, and P is the consumption of media coverage caused by the weakening of vigilance.

(2) In the people of I_{m} group, the contact with the communicator will be reduced. The contacts include mobile phones, text messages, word of mouth and other means. This paper assumes that the contact rate is reduced to the ratio of δ_{i} $(0 < \delta_{\mathrm{i}} < 1)$, so that the number of ignorant persons who are effectively contacted with the communicator is reduced to $\delta_{\mathrm{i}} I_{\mathrm{m}}$. Similarly, in the people of the S_{m} group, the effective contact rate between the communicator and the ignorant with the self-protection will reduce [3]. We assume that the ratio is reduced to δ_{s} $(0 < \delta_{\mathrm{s}} < 1)$, then the number of communicators in effective contact with the ignorant is $\delta_{\mathrm{s}} S_{\mathrm{m}}$.

(3) In the people of I_{u} group, due to the impact of media coverage, the ignorant without self-protection consciousness will become an ignorant I_{m} with self-protection consciousness at β_{m} transport rate. The communicator S_{u} without self-protection awareness will become a communicator S_{m} with self-protection awareness by a rate of v part. The change of the ignorant with self-protection consciousness is influenced by the media coverage, and the communicator's self-protection consciousness is mainly influenced by the degree of self-protection and the knowledge of the popular science knowledge. With the pass of time, the flow of real information and the impact of human self-forgotten mechanism, self-protection consciousness often lose their vigilance. After a certain period of time, an ignorant I_{m} with self-protection awareness changes into the ignorance I_{u} without self-protection awareness by the proportion u_{i} part. The communicator S_{m} with self-protection awareness changes into the communicator S_{u} without self-protection awareness by the rate of u_{s} part.

The members of I_{m} group can transform a communicator S_{m} with protection awareness through connecting with the two communicators S_{m} and S_{u}. The members of I_{u} group can transform a transmitter S_{u} without protection awareness through connecting with S_{m} and S_{u}. Quantity I_{u} will not directly become a communicator S_{m} with protection awareness. It will become I_{m} first, and then will become S_{m}, or directly will become S_{u}.

Based on the above four rules, Fig. 1 shows the interaction schematic diagram between the media reports and rumors. This paper does not take into account the flow of migratory factors in the discussion.

In conjunction with the model framework, we can see that the rumors of media coverage are as follows:

$$\begin{cases} \frac{\mathrm{d}I_{\mathrm{m}}}{\mathrm{d}t} = -\beta_{\mathrm{u}}\delta_{\mathrm{i}}I_{\mathrm{m}}(S_{\mathrm{u}} + \delta_{\mathrm{s}}S_{\mathrm{m}}) + \beta_{\mathrm{m}}I_{\mathrm{u}}M_{\mathrm{e}} - u_{\mathrm{i}}I_{\mathrm{m}}\,, \\ \frac{\mathrm{d}I_{\mathrm{u}}}{\mathrm{d}t} = -\beta_{\mathrm{u}}I_{\mathrm{u}}(S_{\mathrm{u}} + \delta_{\mathrm{s}}S_{\mathrm{m}}) - \beta_{\mathrm{m}}I_{\mathrm{u}}M_{\mathrm{e}} + u_{\mathrm{i}}I_{\mathrm{m}}\,, \\ \frac{\mathrm{d}S_{\mathrm{m}}}{\mathrm{d}t} = \beta_{\mathrm{u}}\delta_{\mathrm{i}}I_{\mathrm{m}}(S_{\mathrm{u}} + \delta_{\mathrm{s}}S_{\mathrm{m}}) + vS_{\mathrm{u}} - (a + u_{\mathrm{s}})S_{\mathrm{m}}\,, \\ \frac{\mathrm{d}S_{\mathrm{u}}}{\mathrm{d}t} = \beta_{\mathrm{u}}I_{\mathrm{u}}(S_{\mathrm{u}} + \delta_{\mathrm{s}}S_{\mathrm{m}}) + u_{\mathrm{s}}S_{\mathrm{m}} - (a + v)S_{\mathrm{u}}\,, \\ \frac{\mathrm{d}H}{\mathrm{d}t} = a(S_{\mathrm{u}} + S_{\mathrm{m}}) - \gamma H\,, \\ \frac{\mathrm{d}M_{\mathrm{e}}}{\mathrm{d}t} = p_{\mathrm{s}}(S_{\mathrm{u}} + S_{\mathrm{m}}) + p_{\mathrm{h}}H - pM_{\mathrm{e}}\,. \end{cases} \tag{1}$$

Sudbury and others consider that the population density who never heard the rumor is 0.203 when the total population tends to infinity. That is, due to the limits of transmission and other conditions, in the process of rumor spread, the number of ignorant is relatively stable. Therefore, this paper assumes that the number of ignorant in the whole process is relatively fixed. The total number of ignorant can be approximated as a constant in a short time. This assumption is also very similar to that given by Liu and others. To keep it simple, the symbol $I = I_{\mathrm{m}} + I_{\mathrm{u}}$ and $S = S_{\mathrm{m}} + S_{\mathrm{u}}$ is introduced, which represents the total population of the ignorant and the total number of communicators, respectively. The model of rumor spread

considering the media coverage can transform the following form:

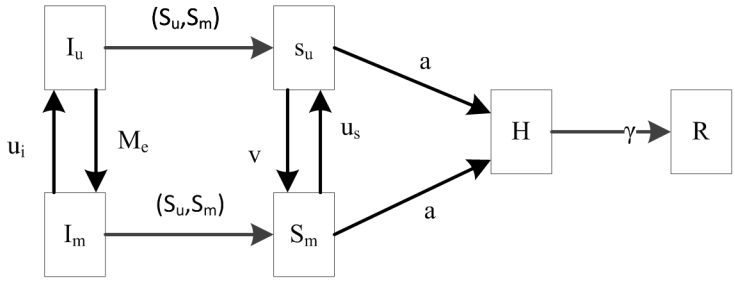


Fig. 1. Schematic diagram of rumor spread considering the media coverage

$$\begin{cases} \frac{\mathrm{d}I_{\mathrm{m}}}{\mathrm{d}t} = -\beta_{\mathrm{u}}\delta_{\mathrm{i}}I_{\mathrm{m}}(S_{\mathrm{u}} + \delta_{\mathrm{s}}S_{\mathrm{m}}) + \beta_{\mathrm{m}}(I - I_{\mathrm{m}})M_{\mathrm{e}} - u_{\mathrm{i}}I_{\mathrm{m}}\,, \\ \frac{\mathrm{d}S_{\mathrm{m}}}{\mathrm{d}t} = \beta_{\mathrm{u}}\delta_{\mathrm{i}}I_{\mathrm{m}}(S_{\mathrm{u}} + \delta_{\mathrm{s}}S_{\mathrm{m}}) + vS_{\mathrm{u}} - (a + u_{\mathrm{s}})S_{\mathrm{m}}\,, \\ \frac{\mathrm{d}S_{\mathrm{u}}}{\mathrm{d}t} = \beta_{\mathrm{u}}(I - I_{\mathrm{m}})(S_{\mathrm{u}} + \delta_{\mathrm{s}}S_{\mathrm{m}}) + u_{\mathrm{s}}S_{\mathrm{m}} - (a + v)S_{\mathrm{u}}\,, \\ \frac{\mathrm{d}H}{\mathrm{d}t} = a(S_{\mathrm{u}} + S_{\mathrm{m}}) - \gamma H\,, \\ \frac{\mathrm{d}M_{\mathrm{e}}}{\mathrm{d}t} = p_{\mathrm{s}}(S_{\mathrm{u}} + S_{\mathrm{m}}) + p_{\mathrm{h}}H - pM_{\mathrm{e}}\,. \end{cases} \tag{2}$$

In this paper, the model of rumor spread proposed considering media coverage is more general than previous studies.

2.3. Model simulation

The simulation results show that the total number of virtual systems is 5446105, and the number of the initial communicators is one. Generally, the basic number of reproduction is the key threshold value of the rumors spread. Based on the above analysis, this paper calculates the basic regeneration number of the model $R_0 = 1.106$, agreeing with the basic reproductive number data reported by official. In addition, the local equilibrium of the model $E_1 = (763920,\ 1169.6,\ 1218.4,\ 13134,\ 17909)$ and 5 eigenvalues ($\lambda_{12} = \pm 0.0261,\ \lambda_3 = -0.0746, \lambda_4 = -0.3374,\ \lambda_5 = -0.2985$) also can be calculated.

Table 2 shows further evaluated parameters.

Table 2. Meaning of parameters in the model

Symbol	α	γ	δ_{i}	δ_{s}	v	p_{h}	p_{s}	p	u_{i}	u_{s}
Evaluation	0.33	0.05	0.50	0.50	0.27	0.08	0.01	0.06	0.01	0.001

(Note: the evaluation of parameters $a,\ \gamma,\ p_h, \delta_{\mathrm{i}},\ \delta_{\mathrm{s}},\ , p$ comes from reference, and the evaluation of others comes from survey experiment.)

3. Results and discussion

3.1. Simulation results analysis

Using the Runge-Kutta method, Fig. 2, left upper part, shows the trend of the number of communicators with self-protection and self-protection. This paper simulates the variation of the number of communicators with time changes. It is clear that the number of communicators with self-protection and without self-protection is showing the characteristics of amplitude-cycle oscillations with the evolution of time. In addition, the number of communicators without self-protection awareness shows a large scatter [5]. Fig. 2, left upper part, also shows that the number of people without self-protection awareness is relatively large. These people are more likely to believe in and spreading rumours. During this time, the emergency information released of official can play a better role in public opinion. Fig. 2, right upper part, we can see that the number of suspects and the number of media reports also show the phenomenon of periodic oscillation. The number of rumors and suspects is far greater than the number of communicators, which means that most of the communicators began to think rationally and doubt the original blind rumors after the official release of emergency information. From Fig. 2, bottom part, we can see that the number of people with self-protection awareness also appeared in the trend of cyclical swings with time changes [6].

In order to further study the equilibrium point of the model, this paper uses MATLAB software to give a phase diagram of the rumor propagation model considering the influence of the media. Fig. 3 shows the impact of media reports on the number of skeptical on the plane $M_{\mathrm{e}}H$ and relationship of the number of people with self-protection awareness and suspects on the plane $I_{\mathrm{m}} - H$. From the two limit cycles in the figure, it can be easily seen that due to the intervention of the media emergency information, the number of skeptical has cyclical shock with the changes of the number of media reports and ignorant with self-protection awareness. During this time, the system is in the emergency management of the turbulent state. If the rumor is handled properly, it will gradually be clarified. However, if it is handled in a wrong way, then the situation is easy to get out of control [7].

Then, this paper discusses the countermeasures of rumor spread from the perspective of rumors, the contact patterns, the self-protection consciousness and the media coverage. By adjusting the parameters under different influencing factors, the effective coping strategies for rumor spread are extracted. This paper gives two numerical results to illustrate this problem. Fig. 4 simulates the effects of media exposure rate, rumor spread rate, and the communicator's effective contact rate on rumors. Fig. 4, left part, simulates the variation of the number of suspects under the two media transmission rate coefficients satisfying the periodic shock conditions, in which the larger media attention rate reduces the number of suspects. Fig. 4, right part, simulates the impact of the rumor transmission rate coefficient on the number of suspects.

When $\beta_{\mathrm{u}} = 0.47/5446105$, and the parameters meet the conditions of the cycle of shock, then the number of skeptical has a period of oscillation with the increase

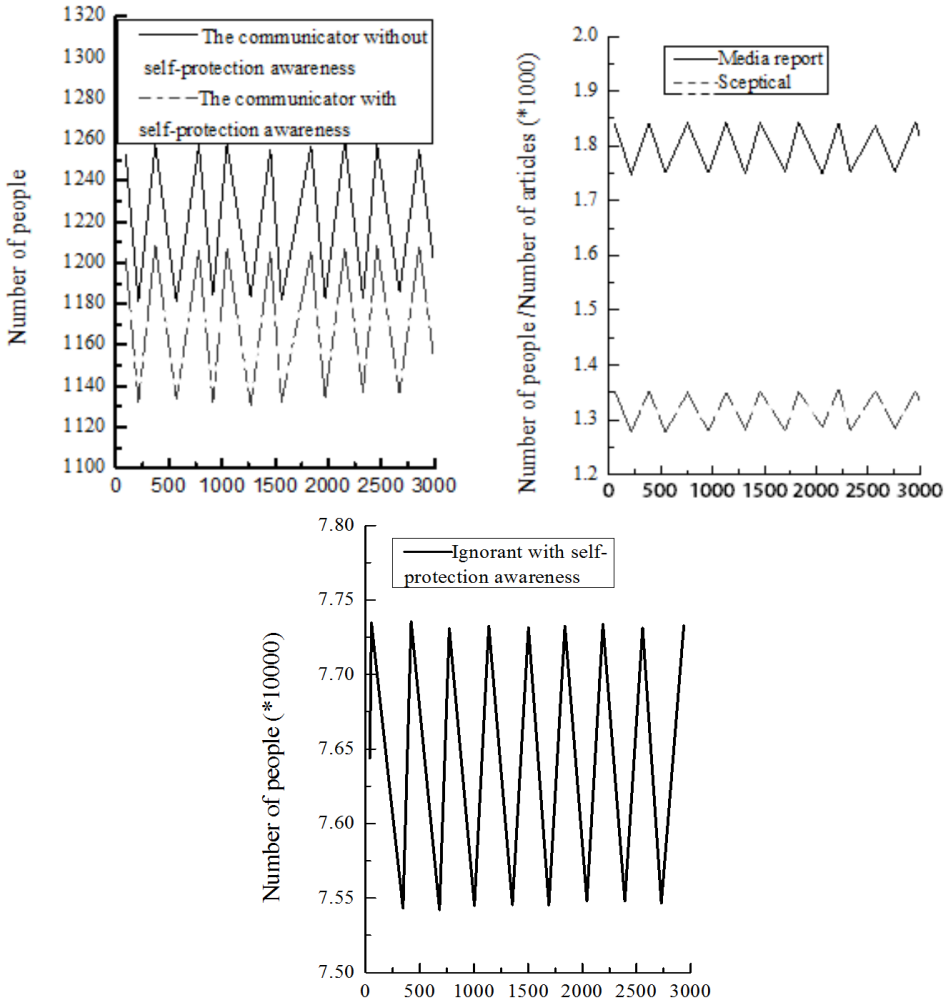


Fig. 2. Changing curve of the number of communicators, the sceptical, ignorant and media reports with time

of time. When the parameters are reduced to 0.45/5446105, and do not meet the conditions of cyclical shocks, that is, the official media reports reduce the spread of rumors at this time and the number of suspects has a sharp decline with the pass of time. This means that media coverage has played a stabilizing role in the spread of rumors. The official media issues emergency information, which can reduce the effective contact rate of rumors between ignorance and communicators. As can be seen from Fig. 4 (c) and (d), the number of skeptics will appear cycle shocks when the contact rate reduce to the half of original [8].

Figure 5 shows that the reduction of the warning awareness and self-protection of the communicator's awareness has a great impact on rumors spread. Figure 5

(a) shows that it is an effective way to deal with rumors by improving the warning awareness of communicators. Figure 5 (b) simulates the effect of the insecure loss rate on the spread of the rumor, which indicates that the improvement of warning awareness for ignorant can reduce the number of skeptics. Based on this, the relevant departments need to grasp the emergency information release time and intensity.

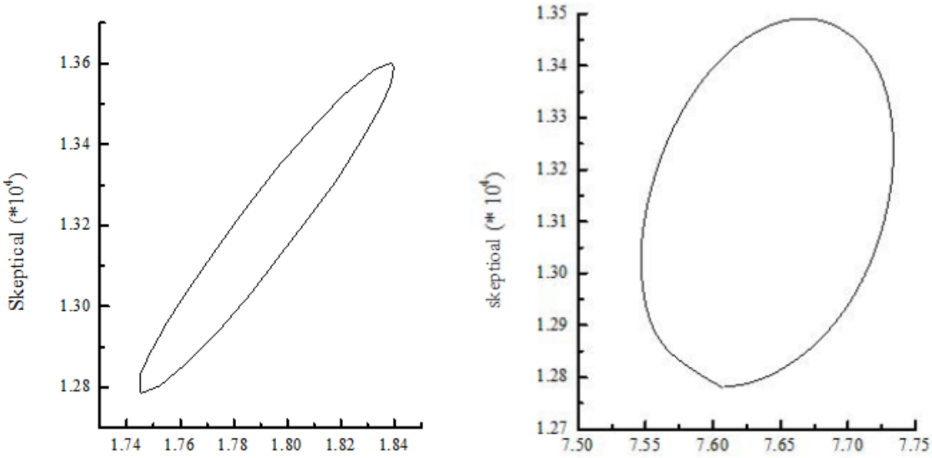


Fig. 3. The phase diagram of rumors spread model considering the media coverage

This article will be from the perspective of Figs. 4 and 5 research, and simulation results can be extracted to deal with the spread of rumors effective measures.

3.2. Suggestions for rumors spread coping strategies

Rumors spread is influenced by a number of factors, such as the nature of the rumor itself, the strength of self-protection awareness of ignorant and communicator, the strength of the rationality awareness, as well as the time, function and strength of official media emergency information and so on.

First, we should grasp the opportunity to effectively deal with rumors spread. In the early period of rumors spread, the appropriate media coverage can play an early warning role. They can effectively eliminate all kinds of suspicion and clarify rumors using the high credibility and high ratings of the media.

Second, we should pay attention to the media type of emergency information. When the rumors produce local negative view, and the skeptical gradually increase, the media with high credibility and emergency response capability should be used to publish information, such as SMS, Tencent QQ, and the official blog and so on. If we miss the best time of releasing the emergency information, and media type selected by the government does not have a high degree of credibility and ratings, then rumors will took the opportunity and enter inside. In addition, information dissemination and the government's voice all can easily lead to a series of emergencies.

Finally, the government should play the role of public opinion leaders, and establish a higher authority to improve public confidence in the eyes of the public. In the

face of the declining credibility of the mainstream media, if the government voice cannot play an immediate effect, then it will breed new rumors, causing anxiety, suspicion and even fear. It is not conducive to the effective response to rumors.

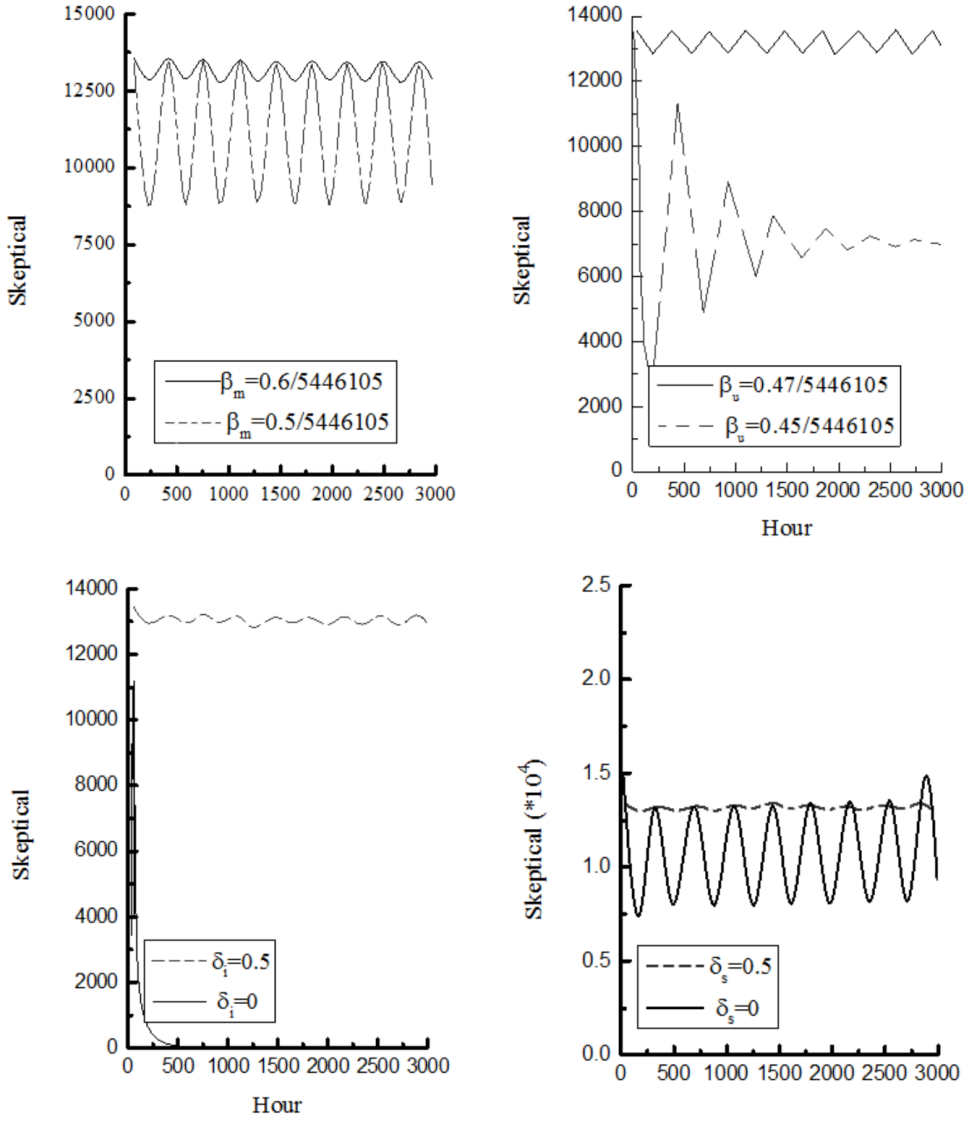


Fig. 4. The influence curve of the transmission rate parameter β_{m}, β_{u} and the effective contact rate parameter δ_{i}, δ_{s} on the number of the skeptics

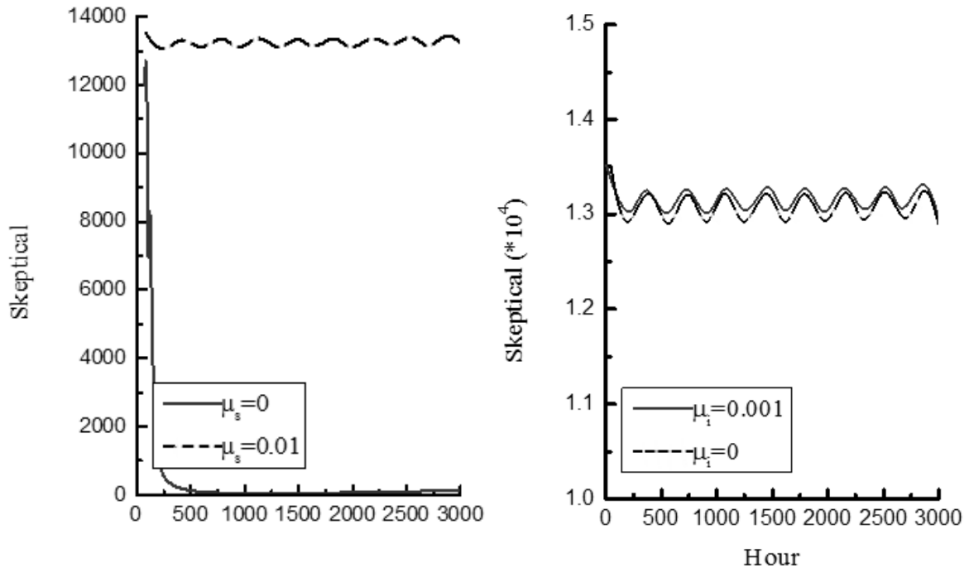


Fig. 5. The change curve of the number of skeptics with time changes under the influence of the weakness rates μ_{i}, μ_{s}

4. Conclusion

After entering the era of large data, everyone can be the disseminator or receiver of information. Currently, the rapid expansion of social information has influenced people's life at all time. At the same time, the proliferation of rubbish and false information, especially the existence of huge rumor, have seriously affected people's life. To this end, we should analyze the law of rumors in the era of large data to formulate the corresponding coping strategies. This paper mainly considers the influence of media reports on the spread of rumors and constructs the corresponding rumor spread model. It also studies the influence rule of the media reports on the rumor transmission mechanism through simulation. Through adjusting the parameters under different influencing factors, many response strategies are proposed. The study points that the media with high credibility and responsiveness should be used to clarify rumors at the early stage. At the same time, the government should improve its credibility in the eyes of the public and play the role of public opinion leaders. This study not only provides useful reference for the government to take effective rumor measures, but also provides objective decision-making basis for policymakers to regulate public opinion orientation.

References

[1] A. Rahmani, A. Amine, R. M. Hamou, M. E. Rahmani, H. A. Bouarara: *Privacy preserving through fireworks algorithm based model for image perturbation in Big Data.*

International Journal of Swarm Intelligence Research *6* (2015), No. 3, 41–58.

[2] J. S. Blumenthal-Barby, H. Krieger: *Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy.* Medical Decision Making *35* (2015), No. 4, 539–557.

[3] M. Tang, Z. Wu: *Research on the mechanisms of Big Data on consumer behavior using the models of C2C e-commerce and countermeasures.* African Journal of Business Management *9* (2015), No. 1, 18–34.

[4] D. Gil, I. Y. Song: *Third international workshop on modeling and management of Big Data (MoBiD'14).* International Workshop on Modeling and Management of Big Data (MoBiD'2014), 27–30 October 2014, Atlanta, GA, USA, Future Generation Computer Systems (2015).

[5] D. Gil, I. Y. Song: *Modeling and management of Big Data: Challenges and opportunities.* Future Generation Computer Systems *63* (2016), 96–99.

[6] K. Gharabaghi, B. Anderson-Nathe: *Big Data for child and youth services?.* Child & Youth Services *35* (2014), No. 3, 193–195.

[7] Z. Wang, Q. Yu: *Privacy trust crisis of personal data in China in the era of Big Data: The survey and countermeasures.* Computer Law & Security Review *31* (2015), No. 6, 782–792.

[8] N. Kshetri: *Big Data's impact on privacy, security and consumer welfare.* Telecommunications Policy *38* (2014), No. 11, 1134–1145.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 75–82

Design of logistics tracking and monitoring system based on internet of things

Qiang Yu[1]

Abstract. To improve the information collection and tracking management in existing logistics, we built a logistics tracking system based on Internet of things. The system is implanted with Java language. We designed the system with server layer and application layer. The design adopts the multi-flow method. The implantation of the system shows that the system automatically displays the work order to be processed on the corresponding label card. The system can automatically flow to the corresponding label card of the work order handler of the next task node for the work order that the current node task is processed. The simulation results indicate that the process of the work order is in the closed-loop state, which greatly improves the efficiency of the work order processing. We improved the logistics tracking system with RFID tags, the system can integrate into the intelligence transportation system.

Key words. Internet of things, logistics tracking, monitoring system, ITI.

1. Introduction

At present, with the rapid development of economic globalization and information technology, the service economy has developed rapidly in the world, and the logistics industry as a new service sector has been rising rapidly in the world [1], too. More and more abundant materials bring convenience to people's lives, but also bring some impact. How to store and transport the large number of goods with various kinds for effective information management has become an important issue. In the process of information management, it needs to track and trace the materials, and can automatically deal with the fault events. The traditional management methods are no longer applicable. The emergence of Internet of things provides the methods and ideas for solving these problems. Internet of things can connect any items to the Internet through RFID, WSN and other information sensing equipment. It can achieve information exchange through the network, and is provided with the fast acquisition speed and high degree of automation. The logistics tracking

[1]Harbin University, Harbin, 150080, China

management system based on Internet of things can realize the whole process visual control of purchasing, transportation, storage, distribution and use in the logistics supply chain, and can automatically deal with the fault events and promote the logistics management information [2].

Because of the time-sensitive nature of the perishable food, it is required that the cold chain be able to perform the real-time low-temperature environmental monitoring function and keep the cold chain's low-temperature transportation environment, therefore keeping the low temperature environment is the core requirement of the cold chain system. If the temperature control is not accurate enough, it will lead to a series of low-quality products [3]. Each link of cold chain, from the picking of the product to being sold out, need to participate in control. Every link in the chain can go wrong and break the cold chain, either on the platform of a warehouse, in transit, in storage, or in a retail supermarket, which can affect the final consumer's demand. At present, China's cold chain monitoring system is relatively sound, and sound monitoring measures have been taken in the production, storage and sales. But the cold chain monitoring in the transport is relatively weak. Cold chain system based on Internet of things makes sure that the food is transported in the specified low-temperature environment to keep the original freshness, color and nutrition of the food and ensure the food quality and safety [4].

2. Literature review

Internet of Things is a series of independent systems built on the part of infrastructure of existing Internet. It has a unique intelligent infrastructure. Radio Frequency Identification (RFID) is one of the key technologies of Internet of things. The Internet of things can collect information fast and accurately with RFID technology, and make the information interconnect with the traditional Internet [5]. Communication modes of Internet of things includes things and things communication, things and people communication, people and people communication and other different types, which makes the virtual network world extend to real life. The practical application of the Internet of things provides a new opportunity for the development of the logistics industry, gives the goods in the process of transport intelligence, and turns the logistics information management into intelligent. Instrumented, networking, perception, automation and intelligence are the basic characteristics of things [6].

The intelligent, location, tracking, monitoring and management of the Internet of Things is embodied in the field of logistics as a centralized data processing and service center based on the network. It uses radio frequency identification devises (RFID) to collect material information, and transmits information to the data service center through the network. It also integrates the material information on the map displayed by GIS system to achieve the visibility control of material [7-8]. In foreign countries, the intelligent logistics starts some earlier. The development of intelligent logistics of some countries in Europe and the United States is in a leading position. In the United States, the federal logistics company developed a set of logistics system based on Internet of things for real-time monitoring of logistics.

The system calculates the data uploaded through the sensor in the server-side, if the logistics has problems, it can alarm in time. And the user can monitor the dynamic of the goods in time, which is easy to transport some dangerous goods, fragile goods and high value goods. Caterpillar has developed a system which is specifically designed for intelligent logistics. The system can simulate, plan, evaluate and forecast the process of warehousing and freight transportation. In Japan, the intelligent logistics develops rapidly, and the intelligent logistics is applied in the logistics transportation of cold chain, fresh chain and chemical industry, [9].

In the country, the national policy attaches importance to Internet of things, and more and more technology companies want to develop on the intelligent logistics. The development of intelligent logistics of some traffic hub cities in China is much higher than other cities. Major electronic business develop rapidly in the field of e-commerce, but also aware the importance of the logistics industry in the supply chain, have increased logistics construction, and Jingdong, Suning and other electricity providers continue to invest in logistics warehousing. With the rapidly development of upstream and downstream industries of supply chain, the logistics industry has also been paid more and more attention [10]. At the same time, the arrival of the era of large data brings opportunities and challenges to the development of intelligent logistics. The storage, processing, query of logistics information uploaded all the time makes the traditional data storage become powerless. Therefore, the new logistics information system of Internet of things should improve the storage way of data.

3. System composition and flow design

3.1. The overall structure of system

The overall structure of logistics tracking system based on Internet of things is shown in Fig. 1. It is mainly divided into server layer and application layer. The server layer is the bottom of the application layer. Server layer will deal with the distribution of goods, invoicing and inventory information collected through a variety of sensors. Application layer mainly turns the information processed by server layer into events, and stores the information in the database in accordance with the event format required by logistics enterprises. And then the different information is showed to each user in the system interface according to the different user rights to query and maintain conveniently.

3.2. System composition

Server layer. Server layer is divided into three layers: sensor abstraction layer, application engine layer and communication layer. The sensor abstraction layer provides a common API that integrates with the LLRP sensor adapter, the Alien sensor adapter, the Borcorde sensor adapter, and the DB sensor adapter to collect various data information from the sensor and generate raw events. The application engine layer filters the collected original events according to event handling rules

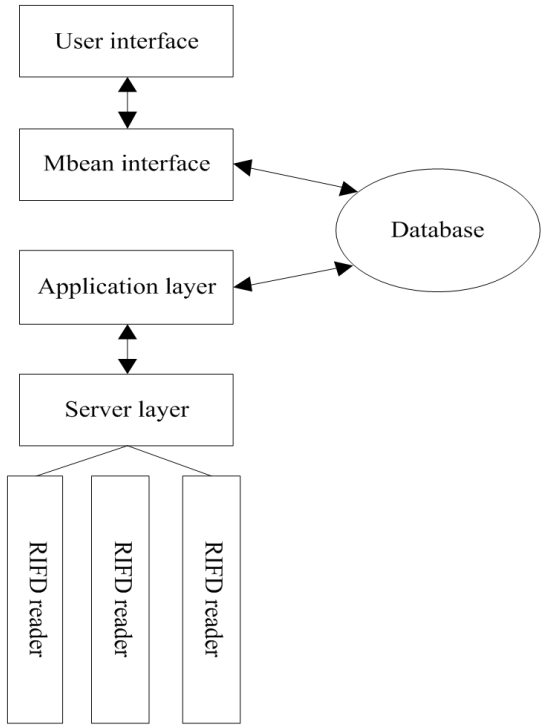


Fig. 1. Overall structure of logistics tracking system based on Internet of things

to get the industrial events needed by logistics enterprises. Through the TCP and SOAP protocol, JMS and RMI technology, communication layer will use the event of the industry in the engine layer to integrate with other systems (such as: database, operation and maintenance systems, etc.). The overall architecture of the server layer is shown in Fig. 2. Sensor abstraction layer: server layer connects to Borcorde reader, RFID reader, terminal mobile devices and wireless sensors and other sensors, from which to collect logistics information for processing. The sensor abstraction layer allows the user to collect all the information needed for the application in a transparent manner. Application Engine Layers: The application engine layer processes a large number of redundant events generated by the sensor abstraction layer and generates meaningful events needed by the logistics industry. The server uses a sophisticated event processing engine Esper that filters and identifies valuable events from a large number of event streams according to rules. Esper allows user to write query statements that are similar to database syntax.

Communication layer: the communication layer will submit the processed data to the application system for integration to realize the information exchange. The design of various application interfaces in this layer achieves effective integration with the existing application system.

Application layer. Application layer mainly turns the event information processed by the server layer into the event information needed by the logistics industry, and

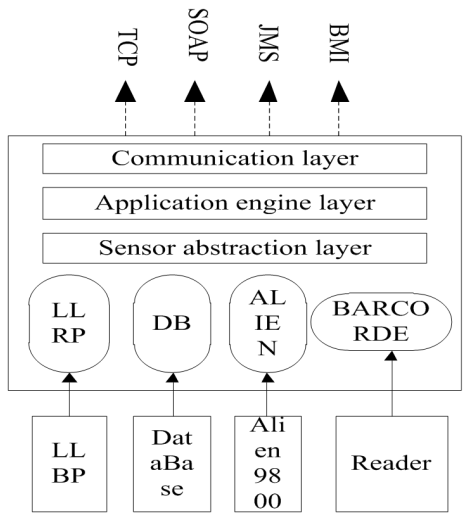


Fig. 2. Overall structure of server layer

then the event information is stored in the database for the display, viewing and dealing with in user interface. In the distribution process, there are two main events: normal events and abnormal events. The normal events include that the designated items reach to the designated location and leave the designated location. Abnormal events indicate that the items are not transported according to the requirements, which includes that items have missed a place, the goods are shipped back, the items don't follow the prescribed route to transport, and there is no corresponding item name and the corresponding location name. The specific process is shown in Fig. 3.

3.3. System flow design

In the logistics business process, the system refers to the best practice and the related standard of the ITIL, and designs the operable, distributed and automatic processing flow according to the event processing in the process of tracking the goods. The design adopts the multi-flow method. The normal events stored in the database by the application layer are handled according to the event work order flow. The abnormal events are processed according to the problem work order flow, and the emergency situation such as line adjustment during the transportation of the goods is processed according to the change work order flow. Event work order can be generated automatically or manually by the failure alarm event during work order processing. Event work order can lead to change work orders and problem work order. Users and customers can query the detail information of the transport process of goods through the system's service desk. Maintenance personnel of logistics operation can deal with the events produced in the transport of goods in accordance with the flow, and manage the assets of logistics enterprises.

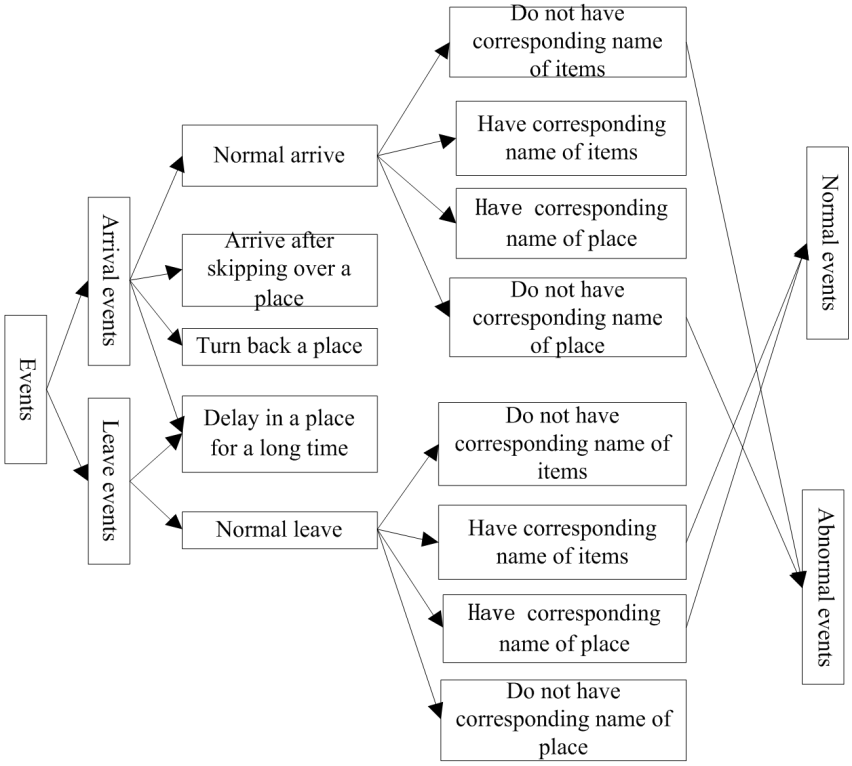


Fig. 3. Application layer events analysis

4. System theory and implementation

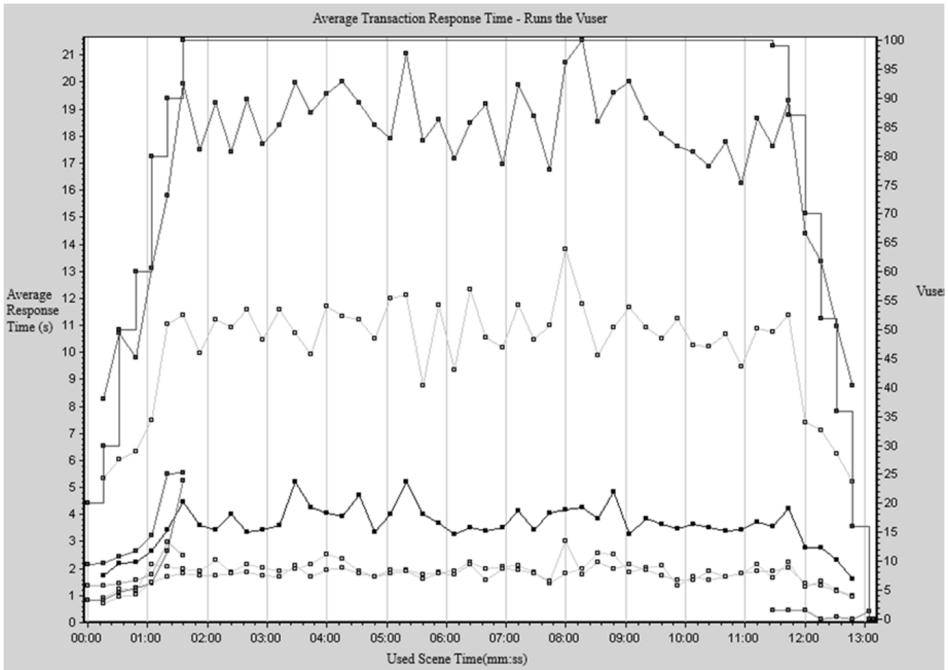
4.1. RFID module

RFID module uses high-frequency RFID devices to collect data, and the goods in transit can be real-time queried. The user can get the information of the arrived items in time through text messages and Email, and can query the historical information.

4.2. Geographic information system

Geographic information system (GIS) is a comprehensive technical system that collects, stores, computes, retrieves, analyzes, displays and describes the geographical information in the whole or part of the Earth's surface space, supported by the computer network system. The geographic information system has the function of displaying the geographical location information of the articles accurately. The system uses this function and the official Chinese map API interface provided by Google to develop to realize the visualization of the geographical information during the transportation of goods, and can improve the accuracy of the display and

maintain the articles line.



Colour	Figure	Proportion	Measure	The Minimum Value of The Graph	The Average of The Graphs	The maximum value of the graph	The Middle of The Graph	Standard Deviation of Graphs
	Response Time	1	Action_Transaction	8.279	17.335	21.55	18.396	3.119
	Response Time	1	vuser_end_Transaction	0.121	0.314	0.468	0.398	0.148
	Response Time	1	vuser_init_Transaction	2.156	3.389	5.526	2.632	1.388
	Response Time	1	Save	0.923	1.867	2.544	1.912	0.371
	Response Time	1	Fill in a form	1.605	3.584	5.201	3.606	0.758
	Response Time	1	Create a new page	5.189	10.192	13.794	10.741	1.92
	Response Time	1	log in	0.808	1.908	5.26	1.266	1.486
	Response Time	1	Enter the work order module	1.343	1.852	2.995	1.573	0.594
	Response Time	1	Select the next link	0.697	1.74	3.032	1.802	0.37
	Run The Vuser	1	Run	0	52.667	100	52	33.486

Fig. 4. Monthly information system response time

4.3. Automatic process of event flow based on ITIL

The event generated by the system is processed through the flow process interface in accordance with pre-designed ITIL process, including the event work order process, the problem work order process and change work order process. The system automatically displays the work order to be processed on the corresponding label card. The system will automatically flow to the corresponding label card of the work order handler of the next task node for the work order that the current node task is processed. In the process of the whole work order, the process of the work order is in the closed-loop state, which greatly improves the efficiency of the work order processing.

5. Conclusion

We design and develop the logistics tracking management system based on Internet of things technology. The system realizes the tracking and monitoring of articles in the process of goods transportation, and adopts the ITIL idea to process the events produced in the process of transportation and realize the automatic management. The realization of the system makes the management of logistics and transportation more intelligent and humanized, and promotes the development of logistics industry information construction. The emergence of Internet of things has brought new opportunities for the development of the logistics industry. With the popularity of RFID tags, information technology development of logistics industry will enter a new stage.

References

[1] L. Da Xu, W. He, S. Li: *Internet of Things in industries: A survey.* IEEE Transactions on Industrial Informatics *10* (2014), No. 4, 2233–2243.

[2] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami: *Internet of Things (IoT): A vision, architectural elements, and future directions.* Future Generation Computer Systems *29* (2013), No. 7, 1645–1660.

[3] V. Turner, J. F. Gantz, D. Reinsel, S. Minton: *The digital universe of opportunities: Rich data and the increasing value of the Internet of Things.* EMC Digital Universe with Research & Analysis by IDC (2014).

[4] J. Jiang, K. Su: *Management platform architecture of modern tobacco logistics based on Internet of Things technologies.* LISS 2012, Springer Nature, Berlin Heidelberg (2013), 1403–1409.

[5] S. L. Chen, Y. Y. Chen, C. Hsu: *A new approach to integrate Internet-of-Things and software-as-a-service model for logistic systems: A case study.* Sensors (Basel) *14* (2014), No. 4, 6144–6164.

[6] D. Singh, G. Tripathi, A. J. Jara: *A survey of Internet-of-Things: Future vision, architecture, challenges and services.* IEEE World Forum on Internet of Things (WF-IoT), 6–8 March 2014, Seoul, South Korea, IEEE Conference Publications (2014) 287–292.

[7] C. N. Verdouw, A. J. M. Beulens, J. G. A. J. van der Vorst: *Virtualisation of floricultural supply chains: A review from an Internet of Things perspective.* Computers and Electronics in Agriculture *99* (2013), 160–175.

[8] A. Whitmore, A. Agarwal, AL. Da Xu: *The Internet of Things—-A survey of topics and trends.* Information Systems Frontiers *17* (2015), No. 2, 261–274.

[9] I. Lee, K. Lee: *The Internet of Things (IoT): Applications, investments, and challenges for enterprises.* Business Horizons *58* (2015), No. 4, 431–440.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 83–96

Mechanism of TDLAS spectral line distortion and application of calibration technique in time division detection system of optical fiber sensor[1]

Dechun Yuan[2]

Abstract. In order to suppress the emission of harmful gases, gas sensors have come into being. With the development of optical fiber technology, the infrared absorption spectrum gas sensor based on optical fiber has become the focus of discussion. In order to study the application of tunable diode laser absorption spectroscopy (TDLAS) technology in optical fiber sensors, the theoretical analysis of selective absorption of gas molecules was carried out in this paper. By theoretical simulation and experimental verification, the spectral distortions of TDLAS system based on three different demodulation circuits (subtraction circuit, division circuit, BRD circuit) under the influence of non-absorption loss of light intensity were compared. The final experimental results show that different demodulation circuits have different suppression effects on the power fluctuation of laser light source, in which the division demodulation circuit can provide higher detection accuracy.

Key words. TDLAS, spectral line distortion, fiber bending loss.

1. Introduction

Infrared absorption spectrum gas sensor is a rising star of sensor family. However, with its advantages, it has become the focus of discussion in the field of sensing. Until now, infrared absorption spectra of gas sensors have been flourishing. Infrared absorption spectroscopy gas sensing technology has shown many branches. Among them, tunable diode laser absorption spectroscopy (TDLAS) technology is the most widely used and the most representative for real-time gas detection in field. The technique has high measuring accuracy and fast response. The utility model can simultaneously detect a plurality of gas parameters, measure and analyze a plurality of

[1]This work is supported by the Fundamental Research Funds for the Central Universities (Project No.040-41416012).

[2]Northeast Forestry University, Harbin, Heilongjiang, 150040, China; E-mail: 8500284@qq.com

gases. It has wide application range, great potential for development of instruments and long measuring distance. The system has low operating cost, which is easy for installation and upgrade. In addition, the continuous updating and development of optical fiber technology has added new advantages to optical fiber based infrared absorption spectrum gas sensors: low transmission loss of optical fiber, small noise and distortion, high transmission quality. It can realize long-distance light propagation, and remote control telemetry. The optical fiber gas sensor system is easy to connect with the computer, and it can realize multi-function and intelligence. In addition, the fiber transmission band is wide, which is easy to form a sensor network.

2. State of the art

TDLAS technology was first proposed by Hinkley and Reid in 1970s. They modulated the laser output wavelength by modulating the injection current or humidity of the tunable diode laser and applying the technique to the detection of contaminated gas [1]. In 1979, Inaba and Chan of Tohoku University in Japan realized differential absorption gas detection by adjusting the dual output wavelengths of the laser [2]. In 1998, Stewart and others of the University of Strathclyde in England used the space division multiplexing technique to realize multi-point gas detection using a distributed feedback semiconductor laser. The detection limit method was ppmv/m. In addition, they believed that the interference noise in the air chamber was the main factor that limits the signal-to-noise ratio of the system [3]. In 2012, Lei Tao of Princeton University used 4.5 m quantum cascade laser to build a simple and portable free space gas sensor for the simultaneous detection of atmospheric N_2O and CO with the detection accuracy of 0.15 ppbv and 0.36 ppbv, respectively [4]. In 2004, Huang Wei of Anhui Institute of Optics and Fine Mechanics of Chinese Academy of Sciences designed a set of tunable near infrared diode laser spectrometer based on wavelength modulation spectroscopy. Various concentrations of CO_2 were measured in a laboratory combined with a long range absorption cell. When the absorption path was 170 m, the CO_2 with pressure of 1.9995 Pa could be detected. In this low voltage case, the signal-to-noise ratio of the two harmonic was still very high [5]. In 2011, Zhang Ruifeng and others from Tianjin University introduced a remote sensing system for mine gas concentration based on wavelength modulation spectroscopy. A method of determining the burning concentration of a light by detecting the harmonic component of the reflected signal was used to eliminate the influence of the background gas and dust on the attenuation of laser intensity. The minimum detection sensitivity of the system was 0.0714 mg/m^3, and the maximum detection distance could reach 10 m [6].

3. Methodology

Direct absorption spectroscopy (TDLAS) is the simplest and most direct detection technique in the field of gas sensing technology. It allows very primitive and intuitive gas absorption contours. Linear fitting based on gas absorption profile can

not only convenient for researchers to get gas concentration information, but also can get other important environmental parameter information, such as gas and temperature [7]. A typical block diagram of the TDLAS gas sensing system based on direct absorption spectroscopy was shown in Fig. 1.

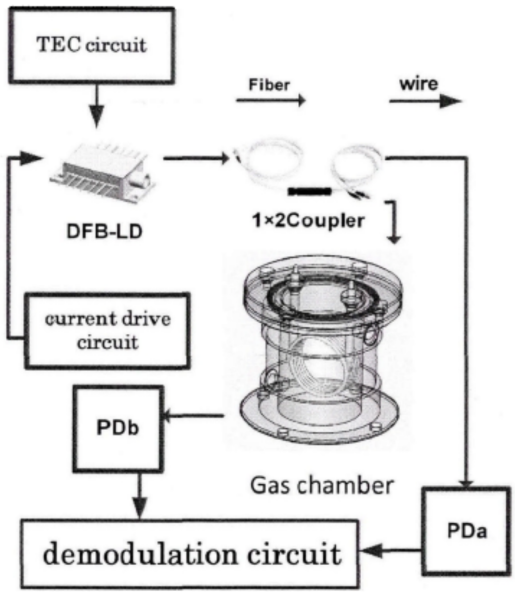


Fig. 1. Detection principle diagram of TDLAS system based on direct absorption spectroscopy

Firstly, the scanning signal (usually sawtooth or triangular wave) is changed by external driving circuit to change the injection current of the laser so as to realize the scanning of the output wavelength of the laser. The output light of the laser is divided into two by an optical fiber coupler. A beam of light passes through the gas chamber and is coupled to the photodetector a, which we call a probe beam. Another beam of light goes directly into the photodetector b as a reference, and we call it a reference beam. Then, the two signals are processed by a differential adjustment circuit to eliminate the scanning baselines brought by the injection current modulation of the laser (current modulation can cause the laser amplitude modulated optical power at the same time), so as to extract the absorption peaks of the tested gas with the zero baseline [8].

There are mainly three kinds of demodulation circuits used in direct absorption spectrum, namely subtraction circuit, division circuit and BRD circuit.

The schematic diagram of the subtraction demodulation circuit is shown in Fig. 2. When the reference beam and the detecting beam are respectively connected by a photoelectric detector, they are converted into a voltage signal through the current and voltage conversion circuit at the same time. Then, after amplifying the proper multiples through the amplifier, the baselines of the two signals are amplified to the same size, and the differential amplifier (subtraction device) is added to the

differential amplifier. The amplified signal finally enters the signal acquisition circuit to be measured (concentration) [9].

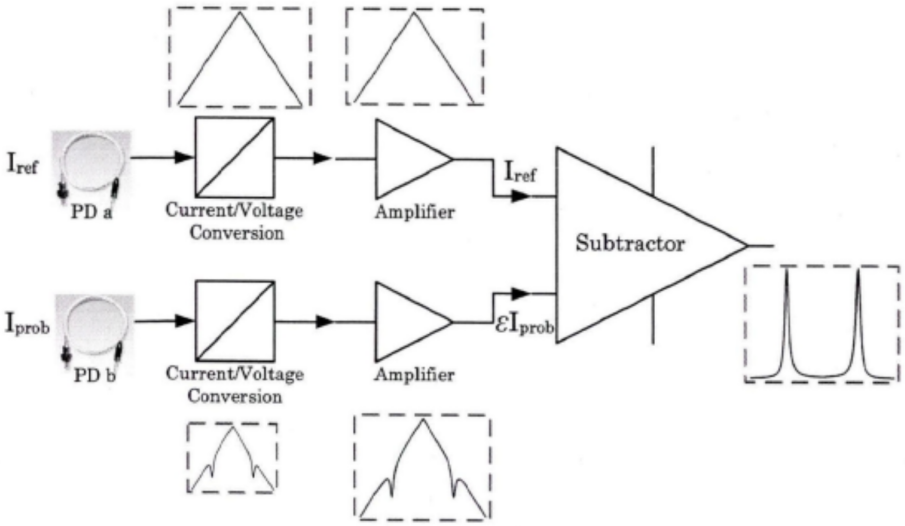


Fig. 2. Schematic diagram of subtraction circuit

According to Fig. 2, it is assumed that the two light intensities coupled to the detector are $I_{\mathrm{ref}}(v)$ and $I_{\mathrm{prob}}(v)$, respectively, as shown in the equations

$$I_{\mathrm{ref}}(v) = I_0(v)\,, \tag{1}$$

$$I_{\mathrm{prob}}(v) = \frac{1}{\varepsilon} I_0(v) \cdot \exp[-\alpha(v)CL]\,. \tag{2}$$

Here, ε is the intensity ratio of reference light to detective light. After the operation amplifier amplifies ε times and the baseline of the probe signal is amplified to the same size as the reference light baseline, as shown in the formula

$$\varepsilon I_{\mathrm{prob}}(v) = I_0(v) \cdot \exp[-\alpha(v)CL]\,. \tag{3}$$

After differential processing, the subtraction device output is shown in the formula

$$I_{\mathrm{out}} = I_{\mathrm{ref}} - \varepsilon I_{\mathrm{prob}} = I_0(v)\,(1 - \exp[-\alpha(v)CL])\ . \tag{4}$$

When $\alpha(v)CL$
$ll1$ and $\exp[-\alpha(v)CL] \approx 1 - \alpha CL$, formula (4) can be converted to formula (5), which is expressed as follows:

$$I_{\mathrm{out}} = I_0(v)\alpha(v)CL\,. \tag{5}$$

When $I_0(v)$ is constant, the output signal $I_{\mathrm{out}}(v)$ is directly proportional to the

concentration of the gas to be measured. Therefore, the concentration of the gas to be measured Ccan be calculated by the peak value of the demodulated absorption peak.

The principal diagram of division circuit is shown in Fig. 3. The principle of division circuit and flow chart of the subtraction circuit are basically the same, but the subtractor replaces the divider. But it should be noted that after the divider, adder subtracter or circuit is needed to adjust the bias. And when the detection light path is as the input of a molecular divider (some division chip output comes out of the inverse function), demodulation absorption peak signal is inverted, which needed to go through the signal processing for reversed phase inverter. Finally, it is sent to the acquisition circuit or computer for signal acquisition [10].

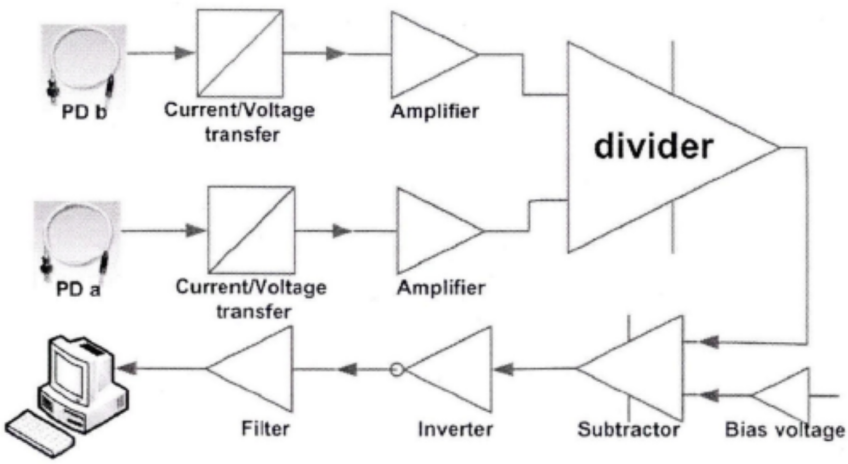


Fig. 3. Principal diagram of division circuit

It is assumed that the two light intensities coupled to the detector remain $I_{\mathrm{ref}}(v)$ and $I_{\mathrm{prob}}(v)$, respectively. After divider treatment, the divider output I_{out} is shown in the formula

$$I_{\mathrm{out}} = I_{\mathrm{prob}}/I_{\mathrm{ref}} = \frac{1}{\varepsilon} I_0(v) \cdot \exp[-\alpha(v)CL]/I_0(v) = \frac{1}{\varepsilon} - \frac{1}{\varepsilon}\alpha(v)CL\,. \tag{6}$$

When the bias component $\frac{1}{\varepsilon}$ is removed from the signal and inverted (used to correct the inverting amplification output of some division chips), the final signal component $\alpha(v)CL/\varepsilon$ is proportional to the concentration of the gas to be measured.

BRD circuit is a method of full electrical noise suppression. It can provide wide band voltage and better linear characteristics in the passband, and have higher CMRR than common mode noise [11]. Its detection function is implemented based on the Ebers-Moll model, as shown in Fig. 4.

The output voltage of the BRD circuit is described by the formula

$$I(v) = G \ln\left(\frac{I_{\mathrm{ref}}}{I_{\mathrm{prob}}} - 1\right), \tag{7}$$

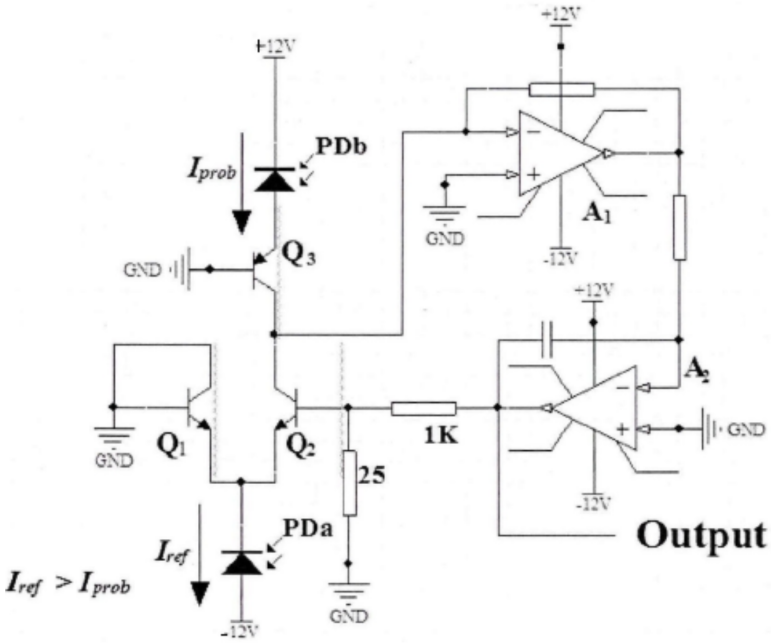


Fig. 4. Schematic diagram of BRD circuit

where G represents the overall gain of the system, and formulas (3)–(7) can be extended as shown in the following two expressions

$$I(v) = G \ln \left[\varepsilon \exp(\alpha(v)CL) - 1\right] , \tag{8}$$

$$\varepsilon \exp(\alpha(v)CL) - 1 \approx \varepsilon + \varepsilon\alpha(v)CL - 1 , \tag{9}$$

For both sides, the derivatives are obtained as follows

$$\mathrm{d}I(v) = \frac{G}{u} \mathrm{div}\ (\varepsilon + \varepsilon\alpha(v)CL - 1) , \tag{10}$$

$$\mathrm{d}I(v) = \frac{G\varepsilon}{u}\, \mathrm{d}(\alpha(v)CL) . \tag{11}$$

Here, $u = I_{\mathrm{ref}}/I_{\mathrm{prob}} - 1$.

As can be seen from formula (11), the output of the BRD circuit is directly proportional to the concentration of the gas to be measured. Therefore, we can directly calculate the amount of gas to be measured through the output voltage.

Compared with the subtraction circuit, the division circuit and the BRD circuit have normalized output, which has better inhibitory effect on the common drift of the system common mode noise and the dual optical power drift. Therefore, the application range is wider.

The scattering effects in the atmosphere include three main categories. When the radius of scattering particles is much smaller than the wavelength of light, the scattering effect is called Rayleigh scattering. When the radius of scattering particles is similar to the wavelength of light, the scattering effect is called Mie scattering. When the radius of a scattering particle grows much longer than that of the optical wave, the scattering effect is called non-selective scattering [12]. The scattering loss in single-mode fibers is generally 0.5–0.2 dB/km. In the double beam optical structure, the scattering effect in optical fiber can be minimized by controlling the difference of two fiber lengths. In this paper, the dual beam structure TDLAS system is used to control the fiber length difference between two beam propagation paths within 5 cm, and the corresponding scattering loss is only 0.000025 dB to 0.00001 dB, which can be neglected [13]. Compared to the dual beam optical structure, the scattering loss of single beam optical structure is relatively large. But because the scattering spectrum is broad spectrum, the scattering power varies very little with wavelength. Therefore, the overall attenuation of light intensity in the scanning range can be approximately considered. This effect is similar to the intensity loss effect of other optical devices that we have described later.

In practical engineering applications, bending deformation of optical fiber under stress is inevitable in optical fiber TDLAS sensing system. Generally, the optical power loss caused by fiber bending is divided into macro bending loss and micro bending loss [14].

Macro bending loss: When the fiber is bent, the transmitted light has to be kept in the same phase as the plane wave in the bent portion of the fiber. The longitudinal propagation velocity of the plane wave front is different from that of the fiber axis. The farther away from the curvature center of the bent fiber, the greater the longitudinal velocity. When the longitudinal velocity exceeds a critical value, a part of the light energy is absorbed into the cladding or passes through the cladding to become a radiation mode, and the leakage is lost, resulting in the loss of optical power. The losses in this case are known as macro bending losses. The macro bend loss of the standard single-mode fiber is given in Table 1.

Table 1. Thermophysical properties of regular fluid and nanoparticles

Bend radius of optical fiber	Bend loss (bend 1 turns)
$R = 15$ mm	<0.03 dB
$R = 10$ mm	0.039 dB
$R = 7.5$ mm	0.065 dB

Relative to macro loss, the influence of micro bend loss and light intensity is small, so the main reason is the influence of macro bending loss.

At present, most of the packaging methods for active or passive optical devices are laser welding, and the gas chamber used in this paper is the same. The two optical fiber collimators are connected together through a copper tube and then fixed by laser welding [15]. Although the laser welding technique has many advantages such as less gas, small heat affected zone, the gas chamber after the laser welding will be affected by the ambient temperature, which makes the fiber support frame

produce slight deformation, thus affecting the collimation of the collimator in the air chamber. The research shows that when the temperature changes, the fiber collimator will cause a certain degree of mismatch due to the slight deformation of the fiber support frame. The mismatch is divided into three types: off axis mismatch, slip angle mismatch, and spacing mismatch. The effect of angle offset on the coupling efficiency is greater than that of axial offset, and the coupling of fiber collimator is not sensitive to the range between them. As shown in Table 2, the insertion loss of the gas cell used in the experiment varies with the ambient temperature is listed.

Table 2. Relationship between insertion loss of air chamber and ambient temperature

Ambient temperature	Insertion loss (dB)	
	Air chamber A	Air chamber B
28 °	0.76	0.30
10 °	0.74	0.28
40 °	1.00	0.36

Through the power meter, the direct output light power of the distributed feedback semiconductor laser pigtail used in the test is measured. The relationship between the light source and the ambient temperature can be obtained, as shown in Table 3.

Table 3. Relation of the output power of distributed feedback semiconductor laser with ambient temperature

Ambient temperature	Output power of DFB semiconductor lasers (μw)
17.3 °	519.9
20.2 °	539.5
20.9 °	554.6
22.5 °	575.4
23.4 °	587.4

Fiber coupler is a very common passive optical device in the field of optical communications. It can realize the distribution and combination of the transmitted optical power between different optical fibers. The fused biconical taper (FBT) method has become the most mature and widely used method for fiber bonding devices because of its advantages of low additional loss, good directivity, and high environmental stability, simple fabrication, low cost and suitable for mass production. The so-called FBT refers that two or more bare optical fibers are put together and heated by a high temperature flame, and both ends of the fiber are stretched to both sides, and finally, a biconical cone structure is formed in the optical fiber flash melting region to form a coupler. Beam splitting ratio of single-mode fused biconical optical fiber coupler depends on the refractive index of the surrounding material of outside the melting zone. The refractive index of matter is directly related to temperature. When the surrounding material is air, the change of humidity will directly affect the refractive index of the air, and then change the beam splitting ratio of the

coupler. As shown in Table 4, the beam splitting ratio of the fiber coupler varies with ambient temperature.

Table 4. Relation of the splitting ratio of single-mode fused biconical optical fiber coupler with ambient temperature

Ambient temperature	Splitting ratio (%)
24 °	48.01
26 °	48.00
43 °	47.904

The transmission power in the optical devices is affected by the ambient temperature changes, which will undoubtedly reduce the detection performance of the TDLAS system. Because of the large number of optical devices in the TDLAS system, the change of temperature is a comprehensive effect for the whole system. When the room temperature changes from 23.8 ° to 24.8 °, the intensity of the detected beam and the reference beam intensity in the TDLAS system are increased to 102.4 % and 102.1 %, respectively. In the fourth part of the experiment and theoretical simulation, we use these two comprehensive change data to verify the detection performance of TDLAS system.

4. Result analysis and discussion

As mentioned above, the TDLAS system based on direct absorption spectrum is realized mostly on the basis of the dual beam optical structure, and the corresponding demodulation circuit has three kinds: subtraction circuit, division circuit and BRD circuit. In this part, the spectral distortion of TDLAS system based on three different demodulation circuits was investigated by theoretical simulation and experimental verification. The detection errors of three systems under the change of environmental factors (fiber bending loss and environmental temperature change) were compared.

The system used a distributed feedback semiconductor laser with a wavelength of 1368 nm. The triangular wave voltage signal with 30 Hz repetition frequency was converted into current modulation by the laser current drive circuit, and the wavelength tuning range was 240 pm. A TDLAS system employing double beam optical structures (branch of direct absorption spectroscopy) was adopted. Three demodulation circuits were used respectively for demodulation.

When the concentration of water vapor in the absorption tank is 270 ppm, the changes of absorption spectrum before and after the temperature fluctuation are shown in Fig. 5.

Because of the different amplification in the three demodulation circuits, the absorption peak of different demodulation circuits is different at the same concentration and temperature. In order to compare, the percentage of peak height change was chosen to evaluate the degree of spectral distortion and measurement error. In division circuit, the peak of water absorption spectrum increased from 520 mV to

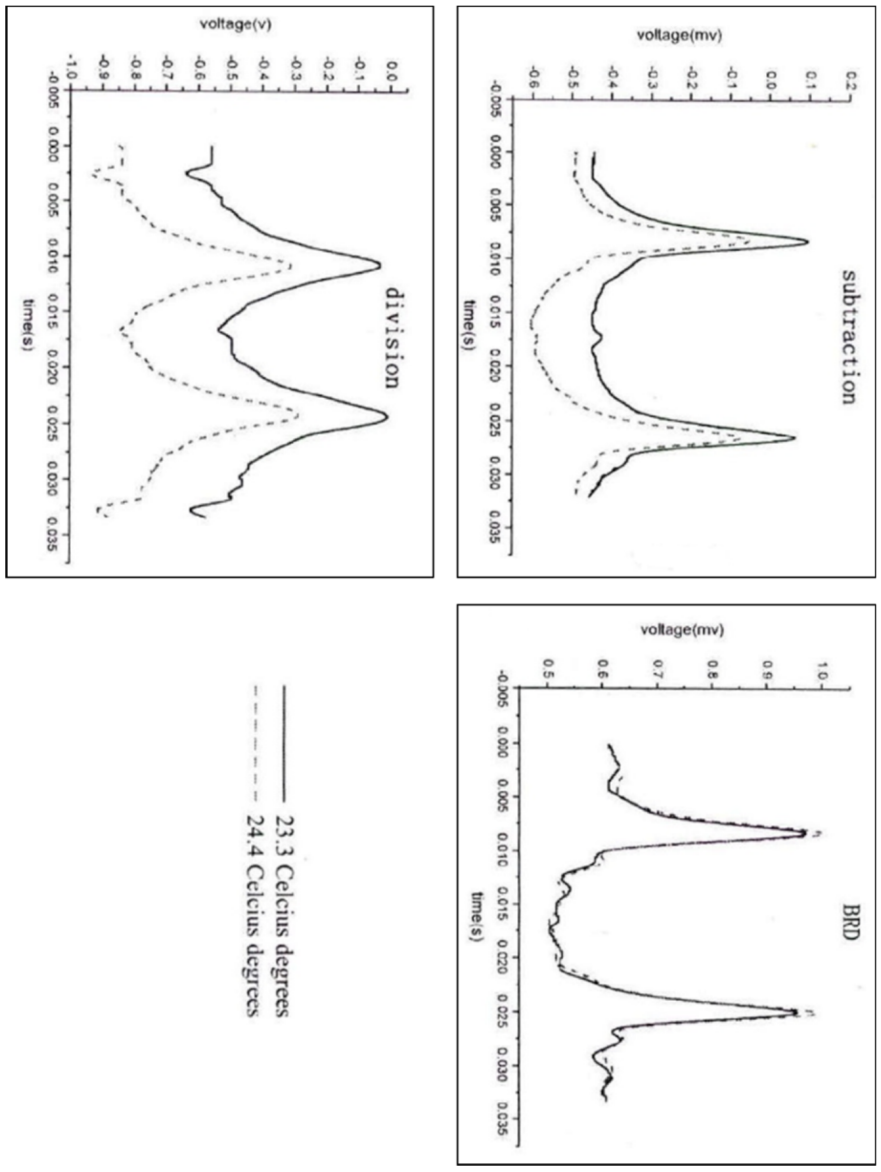


Fig. 5. Demodulation absorption line variation within 1 °C of temperature change

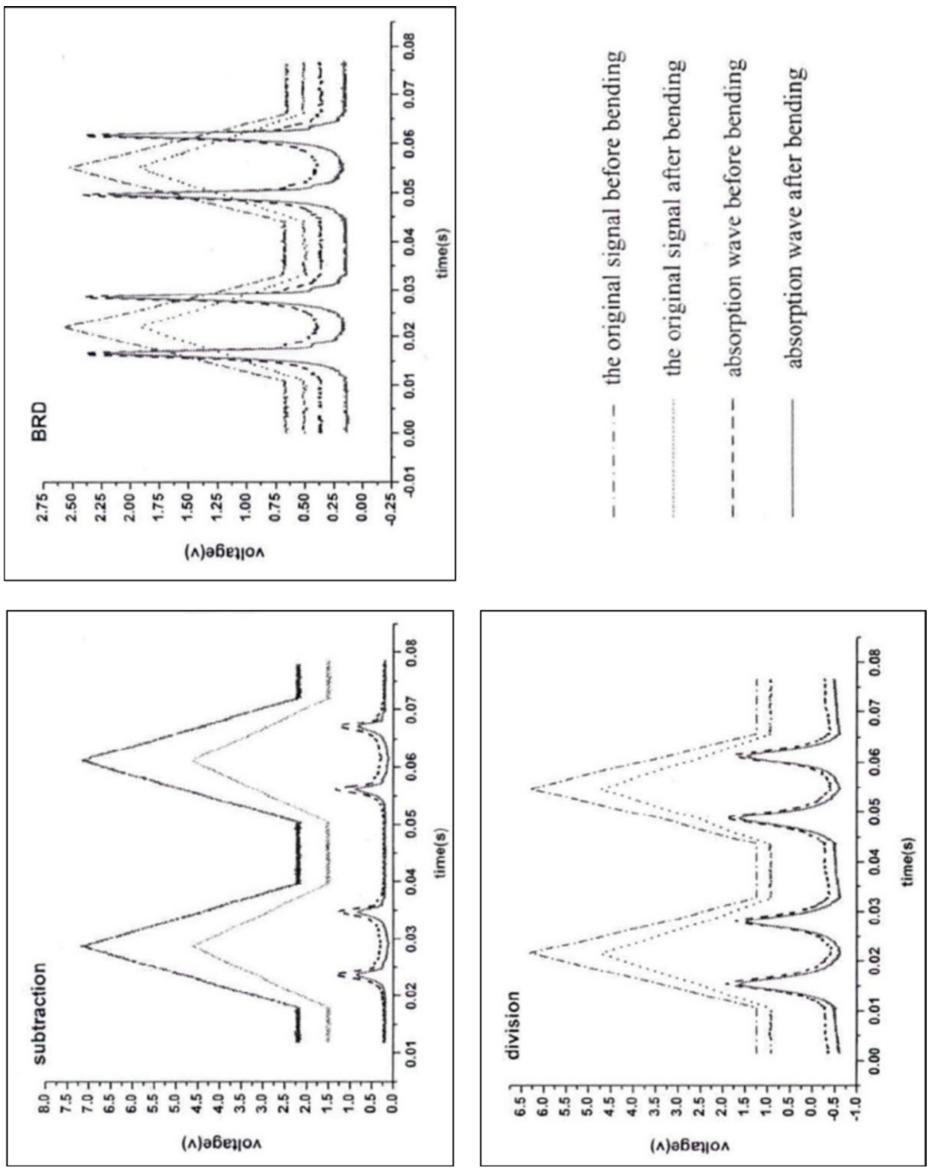


Fig. 6. Change of the absorption spectrum under the power loss of light source

528 mV, and the error was less than 1.54 %. This value was much lower than the error of the 8.82 % of the subtraction circuit (544 m–496 mV) and the error of 8.89 % of the BRD circuit (360 mV–392 mV).

When fiber bending occurred before the fiber coupler that caused the power loss of the laser light source, the water absorption spectrum of the three demodulation circuits was changed, as shown in Fig. 6.

As shown in Fig. 6, in the subtraction circuit, when the background light intensity was lowered from 4.96 V to 3.08 V (caused by the light power loss of light source), the absorption peak height was reduced by 0.44 V (1.10 V–0.66 V), and PFRR was 6.3 dB. In the BRD circuit, when the background light intensity dropped from 1.90 V to 1.42 V, the absorption peak height was decreased by 0.02 V (2.06 V–2.04 V), and PFRR was 13.8 dB. In the division circuit, when the background light intensity was attenuated from 5.08 V to 3.80 V, the absorption peak height was decreased by 0.04 V (2.10 V–2.06 V), and PFRR was 15.1 dB. It can be seen that division and BRD circuits have much higher suppression effect on light source power loss than that of the subtraction circuit. But in practice, the light intensity of two beams cannot be changed strictly. Therefore, even with the normalization, it is impossible to eliminate the measurement error caused by the power change of the light source completely.

5. Conclusion

In this paper, the physical mechanism and correction method of TDLAS spectral line distortion were studied comprehensively. Various factors causing spectral line distortion of TDLAS system were discussed, and the influence of these factors on system detection accuracy was analyzed quantitatively. The corresponding solutions were proposed respectively, and the feasibility of these schemes was verified by theory or experiment. The results show that these schemes can effectively eliminate the influence of various factors on absorption spectral line distortion, reduce the detection error of the system, and improve the detection accuracy and stability of TDLAS system. The conclusions of this study are as follows: when the ambient temperature increases by 1 °C, the detection error of division demodulation circuit is only 0.29 %, which is significantly better than the subtraction demodulation circuit (2.90 %) and BRD demodulation circuit (0.55 %). The concept of power fluctuation suppression ratio was put forward to evaluate the effect of different demodulation circuits on the power fluctuation of laser source. It was pointed out that the division demodulation circuit can provide higher detection accuracy. Through the comparison of fiber bending loss test, it was proved that the system has significant advantages in suppressing the non-absorption loss of light intensity. Although this paper has achieved good results, there are still follow-up works. For example, there has been anhydrous laser on the market, and the various indicators of similar optical devices can be tested and studied in the future.

References

[1] A. G. B. M. Sasse, H. Wormeester, A. V. Silfhout: *New approach for correction of distortions in spectral line profiles in Auger electron spectroscopy.* Surface and Interface Analysis *13* (1988), No. 4, 228–232.

[2] J. Hodgkinson, R. P. Tatam: *Optical gas sensing: A review.* Measurement Science and Technology *24* (2013), No. 1, paper 012004.

[3] A. Karpf, G. N. Rao: *Absorption and wavelength modulation spectroscopy of NO_2 using a tunable, external cavity continuous wave quantum cascade laser.* Applied Optics *48* (2009), No. 2, 408–813.

[4] L. Lathdavong, J. Shao, P. Kluczynski, S. Lundqvist, O. Axner: *Methodology for detection of carbon monoxide in hot, humid media by telecommunication distributed feedback laser-based tunable diode laser absorption spectrometry.* Applied Optics *50* (2011), No. 17, 2531–2550.

[5] M. W. Sigrist, R. Bartlome, D. Marinov, J. M. Rey, D. E. Vogler, H. Wächter: *Trace gas monitoring with infrared laser-based detection schemes.* Applied Physics B *90* (2008), No. 2, 289–300.

[6] L. Tao, K. Sun, M. A. Khan, D. J. Miller, M. A. Zondlo: *Compact and portable open-path sensor for simultaneous measurements of atmospheric N2O and CO using a quantum cascade laser.* Optics Express *20* (2012), No. 27, 28106–28118.

[7] C. G. Zhu, J. Chang, P. p. Wang, W. Wei, S. S. Zhang, Z. Liu, G. D. Peng: *Acquisition of phase-shift fiber grating spectra with 23.5 femtometer spectral resolution using DFB-LD.* Optics Express *21* (2013), No. 25, 31540–31547.

[8] L. A. Wray, J. Li, Z. Q. Qiu, J. Wen, Z. Xu, G. Gu, S. W. Huang, E. Arenholz, W. Yang, Z. Hussain, Y. D. Chuang: *Measurement of the spectral line shapes for orbital excitations in the Mott insulator CoO using high-resolution resonant inelastic x-ray scattering.* Physical Review B *88* (2013), No. 3, paper 035105.

[9] K. Duffin, A. J. McGettrick, W. Johnstone, G. Stewart, D. G. Moodie: *Tunable diode- laser spectroscopy with wavelength modulation: A calibration-free approach to the recovery of absolute gas absorption line shapes.* Journal of Lightwave Technology *25* (2007), No. 10, 3114–3125.

[10] K. Himeno, S. Matsuo, N. Guan, A. Wada: *Low-bending-loss single-mode fibers for fiber-to-the-home.* Journal of Lightwave Technology *23*, (2005), No. 11, 3494–3499.

[11] T. von Lerber, M. W. Sigrist: *Cavity-ring-down principle for fiber-optic resonators: experimental realization of bending loss and evanescent-field sensing.* Applied Optics *41* (2002), No. 18, 3567–3575.

[12] C. Li, Y. M. Zhang, H. Liu, S. Wu, C. A. Huang: *Distributed fiber-optic bidirectional strain–displacement sensor modulated by fiber bending loss.* Sensors and Actuators A: Physical *111* (2004), Nos. 2–3, 236–239.

[13] M. Kashiwagi, K. SaitohK, K. Takenaga, S. Tanigawa, S. Matsuo, M. Fujimaki: *Effectively single-mode all-solid photonic bandgap fiber with large effective area and low bending loss for compact high-power all-fiber lasers.* Optics Express *20* (2012), No. 14, 15061–15070.

[14] M. Napierała, T. Nasilowski, E. Bereś-Pawlik, P. Mergo, F. Berghmans, H. Thienpont: *Large-mode-area photonic crystal fiber with double lattice constant structure and low bending loss.* Optics Express *19* (2011), No. 23, 22628–22636.

[15] C. G. Zhu, J. Chang, P. P. Wang, B. N. Sun, Q. Wang, W. Wei, X. Z. Liu, S. S. Zhang: *Improvement of measurement accuracy of infrared moisture meter by considering the impact of moisture inside optical components.* IEEE Sensors Journal *14* (2014), No. 3, 920–925.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 97–108

A new rotor balancing method based on PTFA theory[1]

Jiang Hong[2], Shi Yongfang[3,4], Ran Xiangfeng[2]

Abstract. Whether the large rotating machinery can run smoothly is closed related to the dynamic balance quality of the rotor, and the dynamic balance technology plays a very important role in the operation and maintenance of the unit. Aiming at the shortage of the current rotor balancing method for processing non-stationary data, the paper applies PTFA (parameterized time-frequency analysis) and holo-balancing method together to determine the unbalance weight and unbalance angle of the rotor. The method rotates the time-frequency characteristics of the rotor run-up vibration signal to extract the rotating frequency component. By adding the trial weight and obtaining the transfer matrix under various speeds, the method can avoids the shortcoming of the traditional rotor balancing method. Experimental results show that this method can reduce unbalanced vibration of the rotor system at working speed and critical speed.

Key words. Balancing method, PTFA, rotor, run-up signal.

1. Introduction

Rotor is the core component of rotary machine, whose vibration problem is inevitable. Rotor imbalance is the most common fault of rotary machine. In order to eliminate or reduce the vibration of the rotor system, we should firstly consider the balance of the rotor.

As early as 1919, Jeffcott raised the need for the dynamic balance of the rotor. Dynamic balance mostly involved rigid rotors alone before the 1950s. But in fact, lots of rotors worked at the first-order critical speed or above the second-order critical speed as their working speed and load continued to increase. The actual needs then promoted the research on methods for the dynamic balance of the flexible rotor. In the past few decades, a series of dynamic balancing methods, techniques and equipment for flexible rotors have been developed at home and abroad. Grobel,

[1]This study is supported by the Science and Technology Supporting of Xinjiang Uygur Autonomous Region(Grant No. 2015211C256)

[2]School of Mechanical Engineering, Xinjiang University, Urumqi, Xinjiang, 8300046, China

[3]School of Medical Engineering, Xinjiang Medical University, 830011,China

[4]Corresponding author

Bishop, Federn and his collaborators laid the foundation for the vibration balancing method, later known as modal balancing [1]. When the modal balancing approach arose in Western Europe, the influence coefficient method [2] came into being in the United States. In 1963, Goodman proposed the use of least squares in the influence coefficient method as its biggest advantage lies in electronic computer-aided dynamic balance which helps to establish a practical multi-plane multi-speed influence coefficient balancing method. In order to further improve this method, a large number of theoretical studies and experiments have been conducted and then many improvement measures have been put forward, contributing to the wide application of the influence coefficient method in practice. Despite constant improvement and development, the above two flexible rotor balancing methods still have their own insurmountable problems.

Based on linear vibration theory, the influence coefficient method works out the calibration quality using the linear relationship between unbalance vector and vibration vector. Modal balancing method relays on the orthogonal principle of rotor vibration mode to keep balance. The vibration of each order rotor vibration mode can only be caused by the corresponding imbalance of vibration mode. If to reduce the vibration of the rotor system, we shall revise the imbalances of each order vibration mode respectively.

Regardless of influence coefficient method or modal balancing method, both of them are based on rotor steady-state response, that is, with the steady-state response data of rotor system under some selected speed, the calibration quality is determined by the steady-state response before and after the trial weight at a balanced speed [3]. Under field balance, we need to carry out many trials to determine the calibration quality. However, the balancing efficiency is low, in addition, we can not ensure the small vibration at other speeds. If the rotor balancing can be rapidly achieved by the transient response data through rotor startup process, it is significantly important to reduce the rotor vibration at all speeds [4].

The rotor startup signals contain rotating frequency component, doubling frequency component and sub-doubling frequency components. Moreover, the vibration of rotating frequency component is mainly related to rotor imbalance. Therefore, it is extremely important for the subsequent rotor balancing to accurately extract the amplitude phase of rotational frequency component during the rotor startup process.

The rotor startup signals are the typical non-stationary signals. In order to effectively study these non-stationary signals, time-frequency analysis has received widespread concern. The basic idea is to construct the joint function of time and frequency through mapping the one-dimensional time domain signal into two-dimensional time-frequency plane. Furthermore, it is also used to describe the energy density and intensity of the signals at different time and frequency [5].

The common time-frequency analysis methods mainly include Short-time Fourier Transform (STFT), Wavelet Transform, Wigner-Ville Distribution and Empirical Mode Decomposition (EMD). When analyzing non-stationary signals, these methods have the following limitations:

1) The time-frequency resolution ratio of STFT is only related to window function. The size and shape of window-field can not be changed with that of the signal

frequency, which lacks self-adaptability.

2) Under the restriction of uncertainty principle, Wavelet Transform can not achieve a high resolution both in time and frequency domains. Once selected, the wavelet can not change throughout the signal analysis process.

3) Wigner-Ville Distribution will be seriously interfered by the cross term when analyzing multi-component signals.

4) EMD has the issues like end effect, mode mixing and so on.

All the time-frequency analysis methods mentioned above have no a priori assumed signal models, and assume the signals to be analyzed as quasi stationary signals to different extent, which is therefore referred to as non-parametric time-frequency method.

Yang Yang and other researchers [6] proposed a new method to analyze non-stationary signal, which is called parameterized time-frequency analysis, whose essential idea is to carry out the time-frequency domain rotation through constructing the transform kernel that matches the signals, so that the concentricity represented by signal time-frequency will be optimal. Based on the parameterized time-frequency analysis, the author decomposes the rotating frequency vibration signals from the startup vibration signals. Combined with the holobalancing method [7], the balance of the rotor under all working conditions is achieved.

2. The extract of rotating frequency component during the rotor startup process

2.1. The basic theory of parameterized time-frequency analysis

The rotor startup signals are the multi-components signals of frequency modulation and amplitude modulation.

$$z(t) = \sum_{i=1}^{N} A_i(t) \exp\left\{ \mathrm{j} \left[2\pi \int f_i(t)\,\mathrm{d}t + \phi_i(t) \right] \right\} . \tag{1}$$

Here, $A_i(t)$, $f_i(t)$ and $\phi_i(t)$ are, respectively, the amplitude, frequency and phase of the ith signal component. The number of signal components is N. All of them are the functions varying with time, and the frequency variation function of rotating frequency component can be fitted by using the key-phase signals.

In view of the above startup signals, and combined with the frequency variation function of rotating frequency signals, the matching rotating operator can be constructed.

$$\Phi^{\mathrm{R}}(t) = \exp\left\{ -\mathrm{j} \left[2\pi \int f(t)\,\mathrm{d}t \right] \right\} . \tag{2}$$

In this way we will obtain a rotation field after rotating the original signals in

the form

$$\mathrm{RTF}_S(t,w,P)=\int_{-\infty}^{+\infty} z(\tau)\Phi^{\mathrm{R}}(\tau)h(\tau-t)\exp(-\mathrm{j}\omega\tau)\,\mathrm{d}\tau\,. \tag{3}$$

The above formula represents the parameterized rotary time-frequency transform. The function of the rotating operator is to rotate the time-frequency feature to a fixed frequency in the time-frequency representation, and to make the signals after the time-frequency rotation pass through the band-pass filter, so as to extract the rotating frequency component during the startup process. The specific decomposition steps are as follows:

(a). Firstly, combined with the frequency variation function of rotating frequency signals, the rotating operator is constructed. Fig. 1(a) and Fig. 2(a) are, respectively, the time domain figure and time-frequency spectrum of the original startup signal.

(b). Rotating the original startup signal, Fig.1 (b) and Fig. 2(b) are, respectively, the time domain figure and time-frequency spectrum of the signal after rotating. According to Fig. 2(b), the time-frequency feature of the rotating frequency component after rotating is completely parallel to the time axis, and is always the initial frequency of the component.

(c). Taking as the center frequency, the zero-phase band-pass filter is designed. After filtering the rotating signal, we will get the filtered signal; Fig. 1(c) and Fig. 2(c) are respectively the time domain figure and time-frequency spectrum of filtered signal.

(d). Rotating the filtered signal through the rotating operator.

Thus, extracting the rotating frequency component, Fig. 1(d) and Fig.2 (d) are, respectively, the time domain figure and time-frequency spectrum of the rotating frequency component.

2.2. The drawing of rotating frequency component Bode diagram

Bode diagram is common method used to analyze the rotor startup information, which can visually display the amplitude and phase of the rotor vibration varying with the speed during the startup process. The traditional drawing method of Bode diagram is discontinuous sampling, which means to collect multiple sets of data during the process of speeding-up. Finally, by connecting all the data points, we will get the Bode diagram, which has a different degree of distortion[8]. This thesis uses the method of achieving the vibration signal complex envelope to draw Bode diagram of the rotor rotating frequency component, which not only reduce the number of signal acquisition, but ensure the accuracy of the Bode diagram.

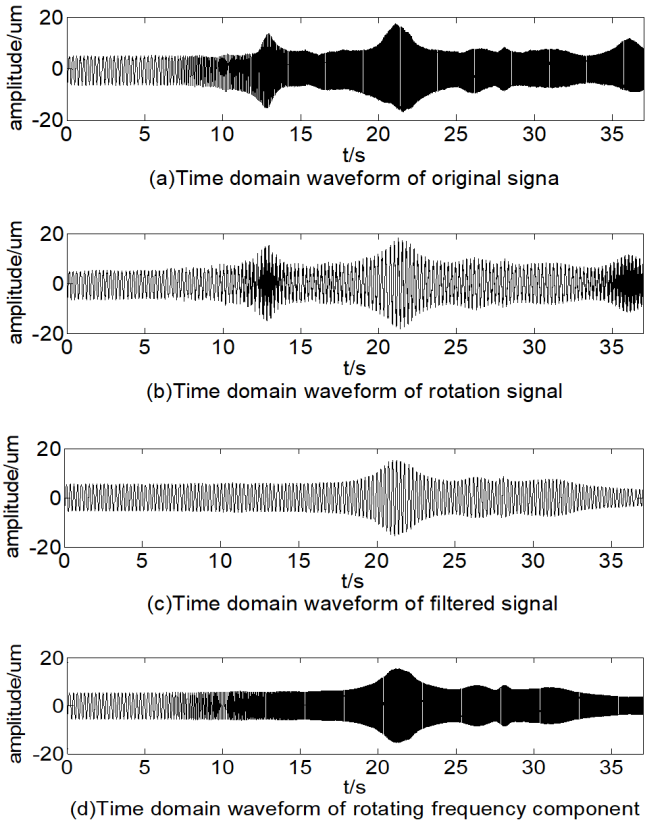


Fig. 1. Time domain waveform of the extraction of rotating frequency component

3. Analysis and experimental verification of simulation signals

3.1. Analysis of simulating signals

In order to verify the accuracy of the parameterized time-frequency analysis method extracting the rotating frequency component, the author has carried out the simulation signal analysis and test bench verification respectively. The simulation signals use Jeffcott model.

Although the coincidence of rotating frequency component extracted by parameterized time-frequency analysis and theoretical rotating frequency component can be observed through the waveform directly, we also have to conduct a quantitative evaluation on the accuracy of the extraction results. This paper uses the root mean square error (RMSE) as the evaluation index

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{n}(s_i - x_i)^2}{n}}\,. \tag{4}$$

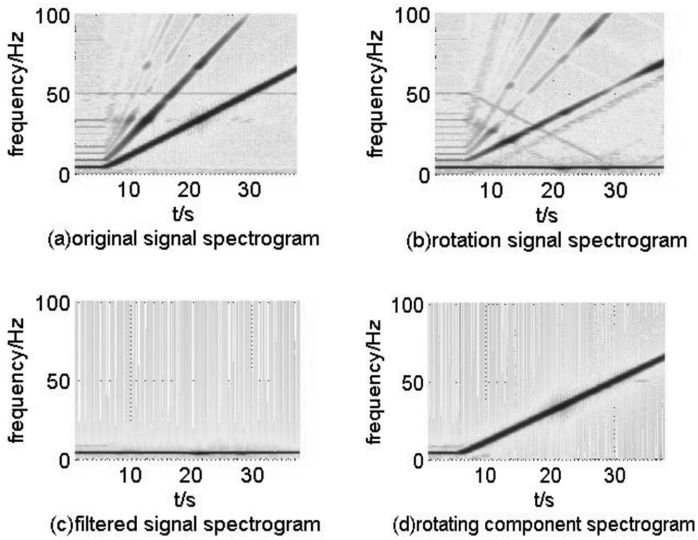


Fig. 2. Time-frequency domain spectrogram of the extraction of rotating frequency component

In formula (4), s_j is the discrete value of rotating frequency component extracted by parameterized time-frequency analysis x_j is the discrete value of theoretical rotating frequency component, an n is the number of discrete points.

After calculating, the root mean square error of rotating frequency component extracted by parameterized time-frequency analysis and theoretical rotating frequency component is 0.0578, which illustrates that the parameterized time-frequency analysis can accurately extract the rotating frequency component.

3.2. The experiment of steady-state startup

The methodology is experimentally verified on Bently RK4 rotor test bench. In order to fully consider the effects of all rotors supporting anisotropy, we respectively install perpendicular eddy current sensors on two measuring planes A and B.

Firstly, collecting a set of steady-state vibration data through the rotor test bench every 20 rpm within the range of speeds between 600 rpm and 3100 rpm, the sampling frequency is 2048 Hz, and the number of sampling points is 2048. After calculating the 126 sets of steady-state startup vibration data collected through Fast Fourier Transform, we will obtain the amplitude and phase values of rotating frequency component. Moreover, through verifying the amplitude and phase values of rotating frequency by spectral correction method, we will calculate more accurate amplitude and phase values of rotating frequency.

Then, collecting the continuous startup data through the rotor test bench every 20 rpm within the range of speeds between 600 rpm and 3100 rpm, and the sampling frequency is 2048 Hz, we will extract the rotating frequency component through parameterized time-frequency method. Moreover, according to its complex envelope,

we also obtain the amplitude and phase values of rotating frequency component.

Taking the root mean square error of formula (4) as evaluation index, the root mean square error of each channel is shown in Table 1.

Table 1. RMSE of rotating frequency components of continuous startup and steady-start startup

	$1-x$	$1-y$	$2-x$	$2-y$
Amplitude (μm)	1.3727	2.0116	1.1386	2.1934
Phase (°)	0.1014	0.0984	0.1107	0.0916

The simulation signal analysis and steady-state startup experiment show that parameterized time-frequency analysis method can accurately extract the rotating frequency component during the rotor startup process, which possess excellent accuracy and reliability.

4. Experiment of balancing

To conduct the balancing experiment on Bently RK4 rotor test bench, we have to set the sampling frequency to be 2048 Hz, and measure to obtain the startup vibration signal of the rotor system within the range of 600 rpm to 3100 rpm.

The rotating frequency component is extracted through using parameterized time-frequency method. Then, according to the complex envelope, we will obtain the amplitude and phase values of rotating frequency component.

Adding 0.6 g ∠135° trail weight to the rotor test bench A correction plane, 0.6 g ∠225° trial weight to the rotor test bench B correction plane, and maintaining the sampling parameters and speeding-up ratio unchanged, we once again measure the startup vibration signal on two planes A and B. We will extract the rotating frequency component through using parameterized time-frequency component, so as to obtain the amplitude and phase values of rotating frequency component.

According to the original startup vibration information and the trial weight added startup vibration information, we can obtain the transfer matrix of the rotor at every speed within the range of 600 rpm to 3100 rpm, that is, the transfer matrix of the rotor under all working conditions. Combined with holobalancing method, we can implement the balancing of rotor under all working conditions.

In order to verify the above balancing method under all working conditions, this thesis respectively chooses the vibration information at 1650 rpm and 2000 rpm to carry out balancing calculation. Moreover, it also chooses 3000 rpm as the working speed of the rotor to analyze its balance effect on the vibrations at the critical and working speed.

Firstly, choosing the vibration information at 1650 rpm to carry out holobalancing, and the weight meter calculation data is shown in Table 2.

Then, choosing the vibration information at 2000 rpm to carry out holobalancing, and the weight meter calculation data is shown in Table 3.

When choosing the vibration information at 1650 rpm to carry out holobalancing, the actual weight added to plane A is 0.9 g∠ 157.5 °, plane B is 1.1 g∠ 225 °. The

vibration amplitude values at the critical and working speeds before and after the balancing are shown in Table 4.

Table 2. The weight meter calculation table of 1650 rpm

Type	Plane A	Plane B
Original vibration $(\mathrm{m}(^\circ)^{-1})$	49.37 ∠ 73.8	45.66∠ 83.4
Vibration of adding trial weight $(\mathrm{m}(^\circ)^{-1})$	25.54∠ 95.3	24.58∠ 100.4
Balance weight $(\mathrm{g}(^\circ)^{-1})$	0.87∠ 159.7	1.08∠ 230.3

Table 3. The weight meter calculation table of 2000 rpm

Type	Plane A	Plane B
Original vibration $(\mathrm{m}(^\circ)^{-1})$	65.38 ∠ 247.78	59.6∠ 254.9
Vibration of adding trial weight $(\mathrm{m}(^\circ)^{-1})$	25.93∠ 267.2	28.6∠ 279.4
Balance weight $(\mathrm{g}(^\circ)^{-1})$	0.4∠ 125.4	1.28∠ 212.5

Table 4. The balancing effects of 1650 rpm

Speed	Position	Amplitude value before balancing (m)	Amplitude value after balancing (m)	Reduction ratio of amplitude value (%)
Critical speed	$1-x$	264.694	33.4107	87.37
	$1-y$	257.5569	44.5457	82.70
	$2-x$	201.858	24.1402	88.04
	$2-y$	258.068	44.4373	82.78
Working speed	$1-x$	26.2071	13.6456	47.93
	$1-y$	23.263	13.0152	44.05
	$2-x$	22.9353	1.5554	93.21
	$2-y$	26.6456	2.4085	90.96

When choosing the vibration information at 2000 rpm to carry out holobalancing, the actual weight added to plane A is $0.4\,\mathrm{g}\angle 135^\circ$, plane B is $1.3\,\mathrm{g}\angle 225^\circ$. The vibration amplitude values at the critical and working speeds before and after the balancing are shown in Table 5.

Figure 3 is the comparison of vibration amplitude value before and after balancing in four channels. According to the experimental results, it can be seen whether to use the vibration information of 1650 rpm or that of 2000 rpm, we can effectively reduce the vibration amplitude values at the critical and working speeds.

Table 5. The balancing effects of 2000 rpm

Speed	Position	Amplitude value before balancing (m)	Amplitude value after balancing (m)	Reduction ratio of amplitude value (%)
Critical speed	$1-x$	264.694	20.317	92.32
	$1-y$	257.5569	34.2177	86.71
	$2-x$	201.858	12.7907	93.66
	$2-y$	258.068	33.8346	86.89
Working speed	$1-x$	26.2071	8.462	67.71
	$1-y$	23.263	8.0702	65.31
	$2-x$	22.9353	8.0465	64.92
	$2-y$	26.6456	9.6458	63.80

5. Conclusion

This paper proposes a new method of controlling the vibration amplitude at the rotor system run-up stage under different speeds effectively, which can ensure the safety operation of rotor systems. In view of the shortage of the current rotor balancing method, this paper constructs appropriate rotation operator by the rotating frequency and rotates the time-frequency characteristics of the run-up signal based on parameterized time-frequency analysis theory, which can accurately extract the rotating frequency component. This balance method can effectively reduced the vibrations at the critical and working speeds, and we have carried out the experimental verification as well. Comparing the proposed and traditional balancing method, we have concluded the following three advantages:

(1) The acquisition of the signals is convenient, because we only need to collect the vibration data during the process of rotor startup, instead of the steady-state data at some certain speed.

(2) The number of startups is less. This method can obtain the transfer matrix of the rotor under full working conditions through only two times of rotor speeding-up startup processes, so as to achieve the balance of the rotor.

(3) The balance effect is excellent, which can effectively reduce the vibrations of the rotor at both working and critical speeds.

The proposed method in this paper can be used at the large-scale rotating machinery balance in practical industrial field. This method need the support of key-phase signal, which made this method may be restricted to use at some rotor without key-phase signal collection device.

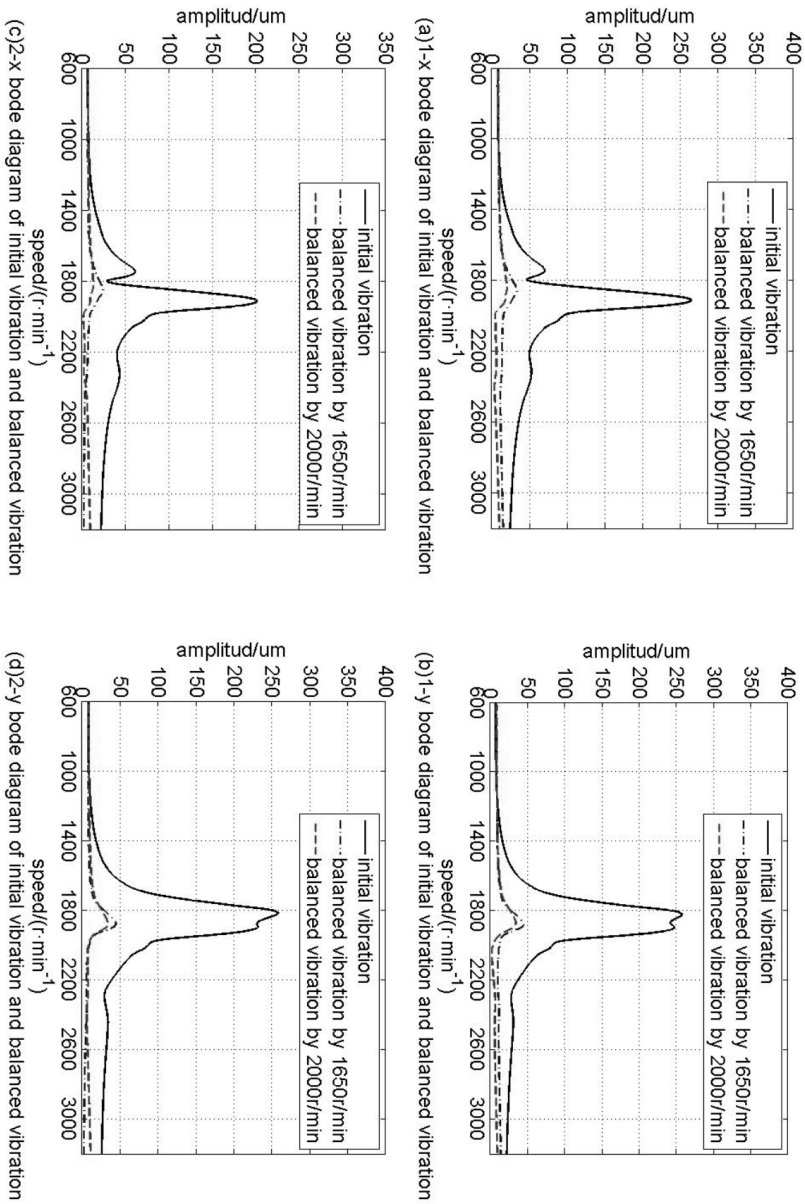


Fig. 3. The amplitude value Bode diagram of rotating frequency before and after balancing

References

[1] S. Liu: *A modified low-speed balancing method for flexible rotors based on holospectrum.* Mechanical Systems and Signal Processing *21* (2007), No. 1, 348–364.

[2] Y. Kang, T. W. Lin, Y. J. Chang, Y. P. Chang, C. C. Wang: *Optimal balancing of flexible rotors by minimizing the condition number of influence coefficients.* Mechanism and Machine Theory *43* (2008), No. 7, 891–908.

[3] M. Paliwal, U. A. Kumar: *Neural networks and statistical techniques: A review of applications.* Expert Systems with Applications *36* (2009), No. 1, 2–17.

[4] X. Li, L. Zheng, Z. Liu: *Balancing of flexible rotors without trial weights based on finite element modal analysis.* Journal of Vibration and Control *19* (2013), No. 3, 461 to 470.

[5] T. Arimitsu: *Non-equilibrium thermo field dynamics and its application to error-correction for spatially correlated quantum errors.* Interdisciplinary Information Sciences *15* (2009), No. 3, 441–471.

[6] Y. Lei, J. Lin, Z. He: *A review on empirical mode decomposition in fault diagnosis of rotating machinery.* Mechanical Systems and Signal Processing *35* (2013), No. 1, 108–126.

[7] S. Ghofrani, D. C. McLernon: *Auto-Wigner–Ville distribution via non-adaptive and adaptive signal decomposition.* Signal Processing *89* (2009), No. 8, 1540–1549.

[8] V. Huard: *L'approche par compétences: Essais de modélisation dans le master « métiers de l'éducation et de la formation ».* Education Permanente (third term 2011), No. 188, 131–146.

[9] D. Jiang, Z. Xu, Z. Chen, Y. Han, H. Xu: *Joint time–frequency sparse estimation of large-scale network traffic.* Computer Networks *55* (2011), No. 15, 3533–3547.

[10] A. W. Jasper, S. J. Klippenstein, L. B. Harding, B. Ruscic: *Kinetics of the reaction of methyl radical with hydroxyl radical and methanol decomposition.* Journal of Physical Chemistry A *111*, (2007), No. 19, 3932–3950.

[11] X. Yang, J. Zhu, L. Qiu, D. Li: *Bioinspired effective prevention of restacking in multilayered graphene films: Towards the next generation of high-performance supercapacitors.* Advanced Materials *23* (2011), No. 25, 2833–2838.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 109–118

Finite element model of high strength concrete joint of weakening type steel

Lihui Shi[1]

Abstract. To improve the seismic performance of steel reinforced high strength concrete frame joints, we built a finite element model of the concrete frame joints to analyze the weak nodes. We established finite element models HSRC-A1 and HSRC-B1 to test the high-strength concrete joints. We used ANSYS program to find the constitutive concrete relations. Comparing the results of HSRC-A1 and HSRC-B1, it can be seen that the weakening of beam-shaped steel flange has a certain influence on the stress-deformation process of high-strength concrete joints. The simulation results are in good agreement with the experimental results, thus the model proposed in this paper has great applicable value.

Key words. Finite element analysis, high strength concrete joint, weakening type steel.

1. Introduction

The steel structure refers to a new type of structure between steel structure and reinforced concrete structure which is formed by rolling or welding steel in concrete. It is also equipped with structural steel bar and a small amount of reinforced steel bar [1]. The reinforced concrete components are equipped with profiled steel, and the steel content of the components has been greatly improved. Therefore, the steel reinforced concrete structure has the characteristics of high bearing capacity, high rigidity, good ductility and strong energy dissipation capacity [2]. The utility model is very suitable for the construction of large span, high rise and heavy load buildings, and it is particularly suitable for use in the construction of the earthquake zone [3].

At present, superplasticizers and highly active mineral admixtures are widely used. Under normal process conditions, high-performance concrete can be easily formulated [4]. High strength concrete is not easy to seepage. The bonding performance is good, and the synergistic ability between the materials is improved. At the same time, the increase of concrete strength improves the bearing capacity of steel reinforced concrete members. Therefore, under the premise that the strength of the component meets the requirements, the section size of the component can

[1]Xijing University, Xian, 710123, China

be further reduced, so as to reduce the weight of the structure, save the steel and increase the construction area. However, with the increase of concrete strength, its seismic performance becomes worse. This will inevitably affect the application of high-strength concrete in seismic engineering [5].

With the development and improvement of the finite element theory, foreign universities and research institutions have made good achievements in finite element analysis of steel reinforced concrete structures. Cornell University has developed a static and dynamic analysis program for 2D and 3D steel frames and steel reinforced concrete frames. In the process of finite element nonlinear analysis, the effect of stiffness degradation of steel members and composite members on the plastic properties of frame joints is considered.

By means of nonlinear finite element analysis of beam column connections in Chiba University, it is pointed out that the nonlinear behavior of nodes can be simulated accurately. It is shown that the finite element analysis is an effective tool to study the performance of nodes. K. Uchiba and H. Noguchi analyzed the finite element nonlinearity of a steel reinforced concrete frame. The results show that the bond behavior has an important effect on the performance of the nodes.

In recent years, a lot of researches have been done on the finite element analysis of steel reinforced concrete. In the process of force, the adhesion between the section steel and the concrete is small, which causes them to slip easily, especially after reaching 80 % of the ultimate strength. Therefore, the finite element model is used to simulate the different elements of the material, and the bonding element is set at the interface of different materials to simulate the bond slip behavior between the two materials. At the same time, it is pointed out that when the finite element analysis of steel reinforced concrete joints is carried out, the node is always in a state of plane stress [6]. The column steel and the flange frame form a good constraint on the node area, and the node zone slip is negligible. Moreover, in view of the difficulty of analysis, the influence of the bond slip between the steel bars, the steel and the concrete is neglected.

2. Finite element analysis of edge joints of high—strength concrete frame

At present, the application of high-rise, super-high-rise buildings and large-span structure medium-sized steel structure and high-strength concrete is becoming more and more common. However, the research on its seismic performance is relatively less. As a new type of seismic construction in recent years, the weakened joint has been widely used in the steel frame, and has gradually been applied to the steel reinforced concrete structure. The seismic test of the reinforced concrete frame joints shows that the dog bone is weakened by the steel frame joints, which can play the role of protecting the core area of the node, and can effectively improve the ductility and energy dissipation capacity of the joints. In the test, cyclic loading is applied to the beam end of the specimen. During the process of force, the steel web of the core area of the node first succumbs and enters the plastic flow stage. With the loading process, the beam near the core area of the specimen joint gradually enters the yield

stage. Finally, when the ultimate load is reached, the concrete in the core area of the node is peeled off and the shear is destroyed.

2.1. Finite element modeling

In this paper, the finite element models HSRC-A1 and HSRC-B1 are established with reference to the test of the high-strength concrete joints of Hunan University. The axial compression ratio of the model is 0.2. The difference between HSRC-A1 and HSRC-B1 is that the shape of HSRC-B1 near beam ends is improved by dog bone. The core area of the joint is designed for the column steel, and the beam is cut off on both sides of the column and welded to the flange of the column steel. At the same time, the horizontal stiffening rib is set at the level of the beam flange of the steel web. The core area of the joint forms a closed flange frame. Beam longitudinal reinforcement and column longitudinal reinforcement link up in the core area of the node. The strength of the steel and the modulus of elasticity of the model are all determined by the measured strength values. The physical properties of steel materials are shown in Table 1.

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

Steel type	Yield strength f_y (N/mm^2)	Ultimate strength f_u (N/mm^2)	Elastic modulus E_s ($\times 10^5$ N/mm^2)	Elongation (%)
I14	293.1	421.9	2.02	33.1
⌀6	415.7	603.7	2.15	23.7
⌀8	434.1	511.3	2.11	17.1
⌀20	500.8	637.9	2.05	12.4
⌀25	419.1	602.3	1.92	16.7

The finite element models HSRC-A1 and HSRC-B1 of high strength concrete joints are established with reference to test specimens HSRC-A1 and HSRC-B1. Model geometry and material strength are the same as the specimen. The steel reinforced high-strength concrete frame joint model is built by ANSYS's APDL parameterized command program. This modeling method can easily simulate and analyze the force and failure process of beam-column joints under different parameters by changing the specific parameters in the command flow, such as axial compression ratio, concrete strength and so on.

The steel reinforced concrete joints are composed of steel, concrete and steel. The materials are modeled by different units. The concrete is SOLID65 unit, the steel is SOLID45 unit, and the longitudinal reinforcement and all the stirrups are LINK8 units. In the actual engineering, the steel reinforced concrete joints are complicated, and the calculation process is also very complicated. So, ANSYS software is used to model the steel reinforced concrete. Before analysis, it should be simplified. This article makes the following assumptions:

(1) The node is always in a plane stress state and follows the small deformation assumption.

(2) In the steel part, the steel web resists shear, and the flange frame is used as a safe reserve. The steel reinforced concrete joints are mainly shear. According to the experimental analysis, it can be seen that the shear force borne by the frame flange is about 5 % of the shearing force of the steel web. It is generally regarded as a safe reserve without considering its shear effect.

(3) The slippage between the sections of steel, concrete and steel is not considered.

(4) The effect of the longitudinal reinforcement pin is not considered.

The elastic modulus E_s, the yield strength f_{y} and the ultimate strength f_{u} of the steel and steel bars can be determined by the steel material test. The Poisson ratio of all the steels in this paper is $\nu = 0.3$. The real constant of the steel element is the section area of its own, and the steel element cannot be set constant or constant is 0.

According to the constitutive relation of each material in the steel reinforced concrete joints, the corresponding stress and strain values of each material are input. The ANSYS program can automatically draw the constitutive relation curve of concrete, steel and steel bar. The constitutive relation of concrete is shown in Fig. 1, the constitutive relation curve of steel is shown in Fig. 2, and the constitutive relation of steel bars is shown in Fig. 3.

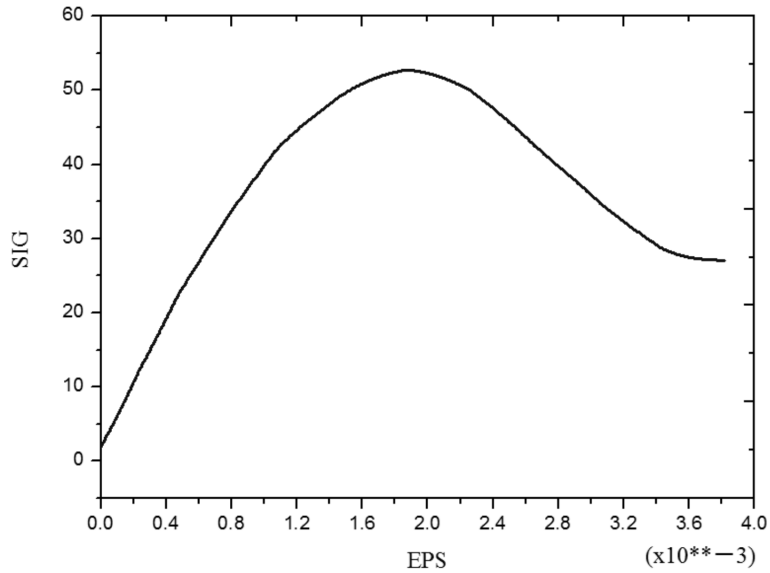


Fig. 1. Constitutive relation curve of concrete

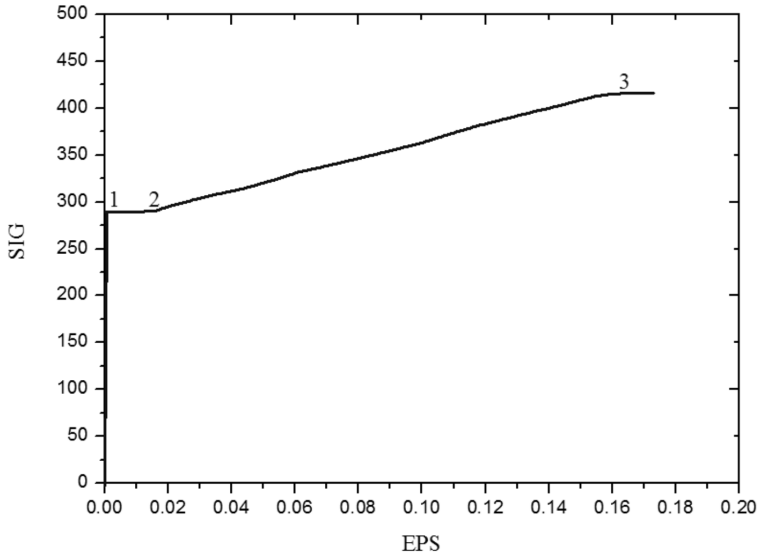


Fig. 2. Constitutive relation curve of steel

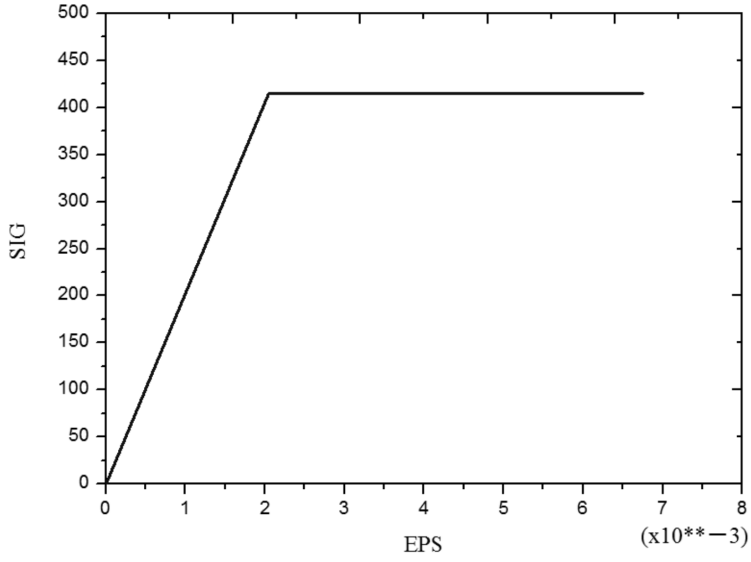


Fig. 3. Constitutive relation curve of concrete iron

2.2. Comparison between the calculated results and experimental results

The model loading point load-displacement curve is captured from the post-processing module, as shown in Fig. 4.

When the area of more than 50 % of the steel web reaches the yield strength,

the load on the beam end is calculated for the yield load, and the corresponding beam end displacement is calculated for the yield displacement. When the model is damaged, the load on the beam end is the ultimate load, and the corresponding beam end displacement is the displacement under the ultimate load. The results of the finite element analysis of HSRC-A1 and HSRC-B1 are shown in Table 2.

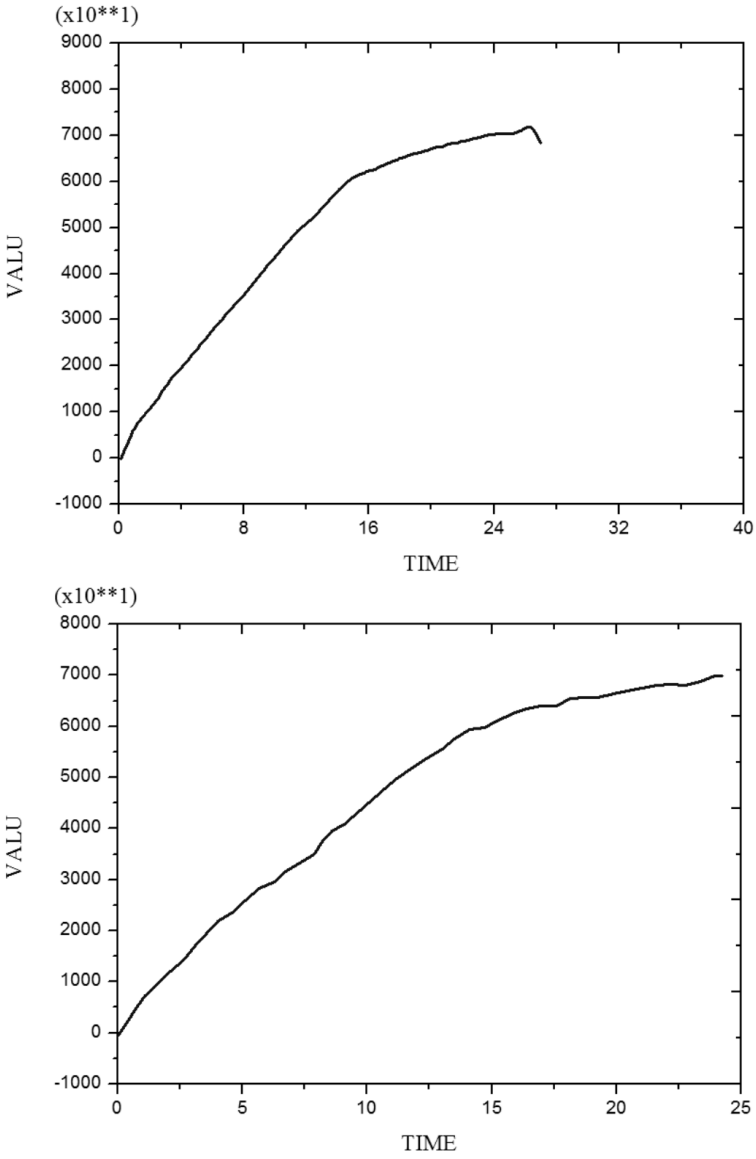


Fig. 4. Model HSRC-A1 and HSRC-B1 load-displacement curve: up–model HSRC-A1, bottom–HSRC-B1

Table 2. The results of finite element method

Number of the model	Yield load (kN)	Ultimate load (kN)	Yield displacement (mm)	Displacement under ultimate load (mm)	Failure form
HSRC-A1	122.35	144.8	17.7	26	Core shear failure
HSRC-B1	116.3	139.8	15.9	23.5	Core shear failure

By comparing the results of HSRC-A1 and HSRC-B1, it can be seen that the weakening of beam-shaped steel flange has a certain influence on the stress-deformation process of high-strength concrete joints. The yield load of the dog bone node is 116.3 kN, which is by 4.9 % lower than the yield of the normal specimen by 122.35 kN. Its ultimate bearing capacity is 139.9 kN. Compared with the ordinary model limit load 144.8 kN, it has no significant change. At the same time, comparing the calculated displacement of the HSRC-A1 and HSRC-B1 with the displacement corresponding to the ultimate load, it can be seen that the dog bone weakening of the nodal beam has little effect on the strength of the component.

The value of the yield loads, ultimate loads, yield displacements and ultimate displacements of HSRC-A1 and HSRC-B1 are shown in Table 3.

Table 3. Experimental results

Number of the model	Yield load (kN)	Ultimate load (kN)	Yield displacement (mm)	Displacement under ultimate load (mm)	Failure form
HSRC-A1	103.7	117.2	18.5	53.7	Core shear failure
HSRC-B1	95.6	114.0	16.8	57.6	Core shear failure

From the experimental results of HSRC-A1 and HSRC-B1, the yield load of dog bone specimen is reduced by about 5 % compared with the average specimen yield value of 114.0 kN. The ultimate load is 117.2 kN. Compared with the limit load of 114 kN, the difference is not big.

ANSYS is difficult to calculate the falling range of the load displacement curve. For convenient for comparing, the displacement corresponding to the ultimate load is compared with the displacement corresponding to the finite element calculation limit load, and assuming that the displacement ductility coefficient is the ratio of the displacement to the yield displacement corresponding to the ultimate load. Comparing the calculated results of ANSYS with the experimental results, the error is the ratio of the absolute value of the difference between the experimental value and the calculated value, as shown in Table 4.

From Table 4 and Table 5, it can be seen that the calculated value of the finite element is not much different from the experimental value, and the error is

controlled within 20 %. The calculated bearing capacity is generally greater than the bearing capacity of the test. The reason for the analysis is that the material is approaching the ideal condition due to the finite element simulation, and the actual material condition, the specimen preparation process and the test process all affect the acquisition of the test result. When the finite element model is calculated, the monotonic load is applied. Therefore, the concrete test of the concrete crack is less, and the bearing capacity of the model is higher. In general, the finite element is well simulated in bearing capacity. The calculated displacement is smaller than the test value. The main reason is that the bond slip between steel and concrete is not taken into account, and the degradation of the overall stiffness of the specimen under repeated loading is not considered. Because the displacement ductility coefficient is the ratio of the displacement and yield displacement of the ultimate load, the displacement ductility coefficient in Table 5 is smaller. It cannot effectively reflect the ductility of steel reinforced high-strength concrete edge joints.

Table 4. Comparison of calculated values of loads with experimental values (kN)

Number of specimen		Yield load		Ultimate load		Failure form
		Experimental value	Calculated value	Experimental value	Calculated value	
HSRC-A1		103.7	122.35	118.2	147.8	Note 1
	Error	15.3 %		19.9 %		
HSRC-B1		95.6	116.3	114.0	139.8	Note 1
	Error	17.8 %		18.45 %		

Note 1: It is mainly shear failure of concrete core area.

Table 5. Comparison of displacement calculated values with experimental values (mm)

Number of specimen		Yield displacement		Diaplacement under ultimate load		The ratio of ultimate load displacement to yield displacement	
		Exper. value	Calcul. value	Exper. value	Calcul. value	Exper. value	Calcul. value
HSRC-A1		18.5	17.7	26.74	26	1.45	1.56
	Error	4.5 %		7 %		7 %	
HSRC-B1		16.8	15.9	24.7	23.5	1.47	1.57
	Error	5.6 %		5.1 %		6 %	

According to Table 5, it can be seen that the displacement ductility coefficient 1.47 of the model HSRC-A1 is not much different from that of the specimen HSRC-B1 displacement ductility coefficient 1.45 in the limit state. According to the ex-

perimental results, it can be seen that the displacement ductility coefficient of the specimen HSRC-B1 is about 15.9 % higher than that of the specimen HSRC-A1, which indicates that the dog bone weakening of the nodal beam can effectively improve the ductility of the node. In the limit state, the influence of the weakening of the girdle of the beam type steel flange on the ductility of the reinforced concrete edge node is not fully displayed. The nodal beam is weakened by dog bone, which can effectively improve the ductility of the "weak node".

The weakening of dog bone can effectively improve the ductility of weak joints, which is due to the nonlinear deformation of the plastic hinge at the end of the common node. However, the non-linear deformation of the plastic hinge can be developed at both ends of the beam at the same time, and the stress concentration is large due to the sudden change of the section. In a word, the use of the dog bone type to weaken the joint construction measures can improve the ductility of steel reinforced high-strength concrete joints, which is consistent with the test results.

3. Conclusion

In 2009, Hunan University conducted a seismic performance test on the edge of the dog bone steel reinforced high-strength concrete frame. On this basis, the finite element analysis software ANSYS was used to analyze the "weak nodes" of the steel reinforced high-strength concrete frame. At the same time, the influence of steel web on the mechanical properties and ductility of "weak node" of high strength concrete was analyzed. The following conclusions are drawn in this paper:

(1) Through the simulation test, the stress change process and the failure mode of the concrete and the steel in the joint are obtained. The core area of the joint steel web first reaches yield. Then, with the increase of the load, the beam steel tensile flange also gradually yield. Finally, the concrete in the core area is crushed, and the joints are destroyed by shear, which is in accordance with the design principle of the weak node.

(2) The finite element model is used as the auxiliary method of the test. The reasonable finite element model, the constitutive relation, the correct boundary condition and the simulation results obtained by the method are in good agreement with the experimental results.

References

[1] Z. Chen, J. Xu, Y. Chen, Y. Su: *Seismic behavior of T-shaped steel reinforced high strength concrete short-limb shear walls under low cyclic reversed loading.* Structural Engineering and Mechanics *57* (2016), No. 4, 681–701.

[2] J. Hegger, S. Görtz, B. Kommer, C. Tigges, C. Drossler: *Prestressed precast beams made of self-compacting concrete.* Betonwerk und Fertigteil-Technik *69* (2003), 40–47.

[3] S. T. Kang, L. Yun, Y. D. Park, J. K. Kim: *Tensile fracture properties of an Ultra High Performance Fiber Reinforced Concrete (UHPFRC) with steel fiber.* Composite Structures *92* (2010), No. 1, 61–71.

[4] C. K. Ma, A. Z. Awang, W. Omar, M. Liang, S. W. Jaw, M. Azimi: *Flexural capacity enhancement of rectangular high-strength concrete columns confined with post-tensioned steel straps: Experimental investigation and analytical modelling.* Structural Concrete *17* (2016), No. 4, 668–676.

[5] E. Grande, M. Imbimbo, E. Sacco: *Bond behaviour of CFRP laminates glued on clay bricks: Experimental and numerical study.* Composites Part B: Engineering *42* (2011), No. 2, 330–340.

[6] L. G. Wang, L. Zhou: *Experimental research on GFRP columns filled with steel-reinforced high-strength concrete subjected to eccentric compression load.* Engineering Mechanics *28* (2011), No. 1, 145–144.

Received July 12, 2017

Acta Technica *62* No. 21/2017, 119–130

Application of RFID technology in the construction and design of quasi automatic warehouse management system for electric power company

Li Yuan[1], Qiu Xiaoping[2,3], Tan Gang[1], Zu Xueying[1]

Abstract. The quasi-automated warehouse management system based on RFID technology plays an important role in improving the turnover efficiency and the comprehensive level of enterprises. At present, China's RFID technology quasi-automated warehouse management system research still has some shortcomings. In order to solve this situation, the research on the relevant theory and concept were clarified and summarized, on this basis, RFID technology was used, and quasi-automated warehouse management system was constructed and designed and applied to the actual power company applications. Finally, the influence factors of quasi-automated warehouse management system of RFID technology and traditional technology were analyzed and compared by gray relational analysis. The results show that the quasi-automated warehouse management system of RFID technology has a positive impact on the development of enterprises.

Key words. RFID technology, quasi-automation, warehouse management system, electricity company.

1. Introduction

With the development of the times, the operation mode and result of every industry begin to be more complicated. The further improvement of the industrial chain structure makes the warehousing management in the industry become the key link of the normal operation of the industry, which has important influence and positive influence on the connection between the main body and the consumer group. In particular, in the context of the gradual increase in the competitiveness of various industries today, reasonable warehouse management system is very important for

[1]State Grid Chongqing Information & Telecommunication Company, Chongqing, 400000, China
[2]Chongqing Xiaomu Technology Co., LTD, Chongqing, 400054, China
[3]Corresponding author

the promotion and development of enterprise competition ability. Under the background of the rapid development of the industry, the traditional warehouse management system has been unable to meet the needs of enterprise development, which urgently requires enterprises and individuals to further advance the overall level of the enterprise through some advanced concepts and technologies. In the high-speed process of the times, many advanced technologies have begun to emerge and combine with the traditional warehouse management system, such as the popularity of computer computing, many companies have begun to apply the technology to its warehouse management system construction process. Because computer technology has the characteristics of large amount of data and large sharing, the reference of this kind of technology effectively improves the management efficiency of warehouse management system and makes the operation period of the whole enterprise be guaranteed. Therefore, the comprehensive level of the whole enterprise or industry has been greatly improved. This research will mainly take the electric power company as the research object, design and study the construction of its quasi automatic warehouse management system, so as to determine the operation advantages of the warehouse management system. The research results are mainly aimed at providing some reference and scientific support for the development of the industry.

2. State of the art

As an important hub for manufacturers and consumers in every industry or industry in today's era, the warehouse management system has a very important impact on the development of a company or business [1]. With the increasing competition level of various industries in recent years, some companies or enterprises have begun to study warehouse management system as an important direction of their future development [2]. Some research suggests that a more reasonable warehouse management system can effectively reduce the cost of manpower during the operation of the company, and can effectively promote the efficiency of product and capital turnover during the development of the industry, and thus have a positive effect on the enhancement of enterprise competitiveness. Many scholars have begun to study more scientific and systematic warehouse management system, and combine the current practical computer technology with the traditional system construction to form a relatively new management system [3]. Some related theories have also been put forward and applied, such as the development of RFID technology and quasi-automated warehouse management system, this technology is based on the traditional enterprise warehouse management system, and introduces more popular computer technology and RFID theory technology [4]. In the perspective of electronic management information, it has a reasonable return on the warehousing products, which has brought some positive effects on the improvement of the final efficiency of warehouse management. It has been found that the warehouse management system based on the technology can directly or indirectly promote the level of comprehensive competition for the enterprise [5].

3. Methodology

Since entering the new century, China's comprehensive national strength and social status have been a certain degree of improvement and progress, various industries have a higher development. In the development of some industries in China, as the key link of the whole industry, the warehouse management system has been continuously improved, which has brought positive effects on the progress of related industries in China. However, due to the scale of development of some enterprises in China is not perfect, the enterprise started relatively late, so the management of the warehouse is only based on the traditional warehouse management system [6]. The system mainly records the relevant data of the warehouse in and out, and uses the paper record method to track the actual running path of the product. However, due to the complete reliance on manual completion, the traditional warehouse management system lacks a certain automation process. With the continuous development of the company and the expansion of the scale, the amount of product storage in a company or enterprise is increasing, which creates some difficulties and obstacles for manual product recording. This is likely to be influenced by certain subjective factors due to simple reliance on artificial methods, resulting in errors in path information tracking of late related products, thus affecting the overall efficiency of the entire process of operation [7]. Therefore, the traditional warehouse management system has begun to seriously hinder the improvement of the competition level of some enterprises or companies in China, and make these companies or enterprises operating costs continue to increase, directly or indirectly caused the company's service quality decline [8]. Although the quasi automatic warehouse management system in some industries in China has a high degree of development and progress, but China's introduction of this technology is relatively late, some theories are still not perfect. Therefore, some of the quasi automated warehouse management systems in some of our enterprises still have some problems, mainly summarized as presented in [9].

The quasi automated warehouse management system under RFID technology is not enough to monitor the related ways of the whole product import and export, because the development of the degree of automation is relatively slow, so there is still the possibility of product storage error in the whole management system.

This has an impact on the increase in the probability of searching and verifying errors in the post-system for product-related information. And with the continuous increase in the scale of production of certain enterprises or companies, in order to better meet the individual needs of customers, the categories of related products produced by the company are also increasing, which makes the storage of the warehouse more pressure. The promotion of people's economic level has also led to the increase of the frequency of the relevant products in and out of the warehouse.

Compared with the traditional warehouse management system, today's warehouse management system began to be more complex and diversified, which requires the development and application of technologies that have a faster and more accurate record of the import and export records of the relevant warehouse products.

In order to better study the construction of quasi-automated warehouse man-

agement system in our country, and to provide reference for improving the comprehensive strength of Chinese enterprises, this study first analyzed and summarized the relevant composition and principle of the quasi-automatic management system of enterprise warehouse in China, and then analyzed and summarized the related characteristics of RFID technology. First of all, through the reading and analysis of relevant documents, the relevant composition of RFID-based technology is shown in Fig. 1 [10].

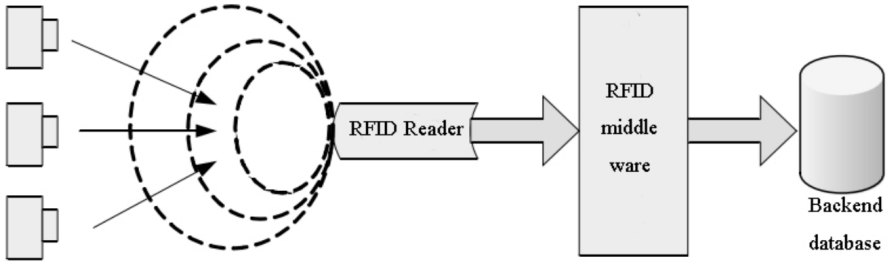


Fig. 1. System composition of RFID technology

On the basis of the relevant theory, the paper summarizes the comparative advantages of the quasi-automated warehouse management system based on RFID technology and the traditional warehouse management system. On this basis, the design goals, construction systems, and the final workflow of the relevant of the automatic management system of electric warehouse based on RFID technology system were designed and discussed, so as to construct RFID technology automatic warehouse management system used in this study.

Then, this study took a power company as the actual test case design, designed the automatic warehouse management system for the power company to introduce FFID Technology. The related RFID technology automatic warehouse management system used the computer language to carry on the programming, finally through the computer programming formed the correlation system construction.

The quasi-automated warehouse management system of the power company designed by the above process was applied to the actual power company warehouse management, and the relevant parameters were statistically and compared, and then the advantages of the quasi-automated warehouse management system were determined. The main influencing factors are shown in Table 1 [11].

Finally, the above parameters were statistical and survey, grey correlation analysis theory was used to synthetically analyze the influence factors of warehouse management system of different technology, so as to determine the influence coefficient of each influence factor in the process of warehouse management of power companies, which provides some scientific reference for the construction and development of the related automation system, and then provides the scientific basis for the optimization and improvement of the existing system. The relevant gray relational analysis model is described below.

Table 1. The main influencing factors required for this study

Primary evaluation parameter	Two stage evaluation parameter	Number
Running cost	The reuse of electronic tags	X1
	Automatic recognition technology	X2
Outbound speed	Outbound speed	X3
	Storage speed	X4
Accuracy	Receiving operation processing	X5
	Picking job handling	X6
	Transport operations processing	X7
	Resource mastery	X8
Stocktaking time	Account alignment	X9
Item query	The state of the object	X10
	Storage status	X11

$$X_b(K) = \frac{X_b^{(0)}(K)}{\frac{1}{n}\sum_{a=1}^{n} X_a^{(0)}(K), \quad b = 1, 2, \cdots, n, \quad K = 1, 2, \cdots, n}. \tag{1}$$

Here, $X_b\,(K)$ represents the dimensionless value of the final survey of each data, $X_b^{(0)}\,(K)$ and $X_a^{(0)}\,(K)$ represent the average values and the actual survey value of the actual survey of each parameter, respectively.

$$\varepsilon_a(K) = \frac{\min\min\left|X_b^{(0)}(K) - X_a^{(0)}(K)\right| + \rho\max\max\left|X_b^{(0)}(K) - X_a^{(0)}(K)\right|}{\left|X_b^{(0)}(K) - X_a^{(0)}(K)\right| + \rho\max\max\left|X_b^{(0)}(K) - X_a^{(0)}(K)\right|}. \tag{2}$$

Here, $\varepsilon_a(K)$ stands for the correlation coefficients between the parameters designed in this study and the automated warehouse management system, while ρ represents the resolution coefficient.

Finally,

$$\gamma_a = \frac{1}{m}\sum_{k=1}^{m} \varepsilon_a(K), \quad a = 1, 2, \cdots, n). \tag{3}$$

Here, γ_a represents the relevance of the relevant parameters involved in this study and quasi-automated warehouse management system. The greater the value is, the greater the effect of the parameter on the construction of the warehouse management system is, and vice versa.

4. Result analysis and discussion

The emergence of computer technology has provided some opportunities for the development of certain industries in China. Many companies in China began to

introduce computer technology into their warehouse management system, and further formed a new warehouse management system which can store and share large amounts of data, brought a more positive impact on the comprehensive development of China's enterprises to a certain extent, made our country's many industrial structure constantly improved and developed, as a result, many of our industry has been able to survive in the growing competition among enterprises (Fig. 2) [12].

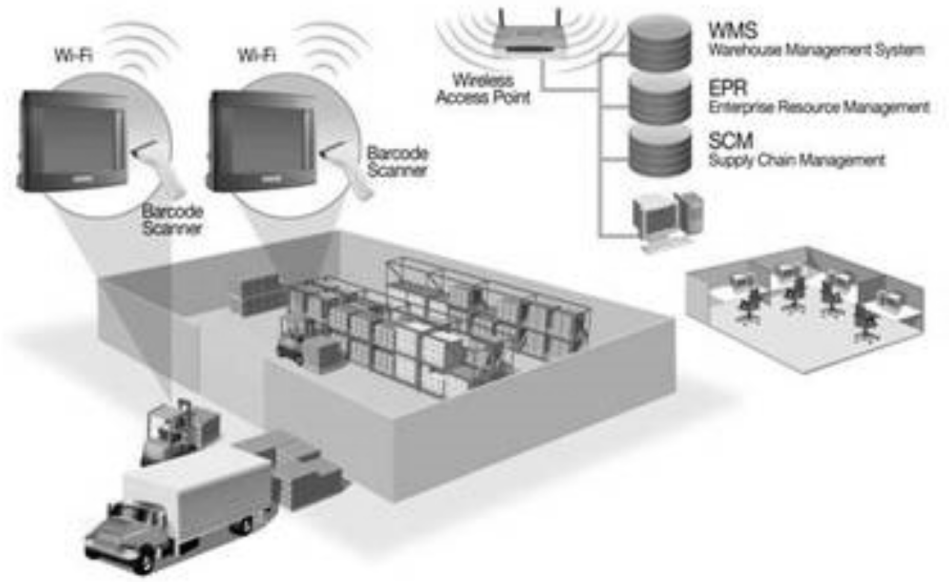


Fig. 2. Application model of enterprise automated warehouse management system in China

Based on the analysis of the application pattern of the automated warehouse management system in our country, the paper further analyzed the advantages and disadvantages of the automated management system under different technologies through the analysis of the relevant data. The analysis results are shown in Table 2 [13]. The results show that the warehouse management system under RFID technology is obviously superior to the traditional warehouse management system in terms of anti-collision technology, storage capacity, inner code recognition and locking of inevitable information, because the warehouse management system of RFID technology has a longer service life, it can adapt to the bad working environment, and the material of relative components is more flexible, so it can be used in different occasions [14].

Based on the clarity of the relevant theory, this study further designed the application of the quasi automated warehouse management system through the relevant programming language, further used the computer to run through the use of the relevant programming language. The interface of the final design inventory management system is shown in Fig. 3 [15]. The performance of the automatic warehouse management system was verified by following up.

Table 2. Comparison and analysis of warehouse storage technologies with different technologies

	RFID technology	Traditional techniques
Anti-collision technology	Without linear scanning, the speed of reading and writing is fast, and it can be used in multi-object identification and motion recognition, with the maximum recognition of 50 at the same time	The scan must be straight, the reading and writing speed is relatively slow, and the recognition target is less
Internal storage space	The production department can produce different types of product storage capacity and the number of bytes per sector according to its needs, and the read and write device can read the memory configuration information, easy to operate in a comprehensive application of different label products	The production department may not be able to obtain more accurate actual demand information, and the storage space is limited, and it can't carry out comprehensive operation on more product information
Identification code	Using the internationally unified and non-repeating 8-byte unique identification of the inner code: the 1-48bit consists of 6 bytes, encoding for product manufacturers, one byte of the 49th to 56th bits is the vendor code	No fixed code recognition, so the relevant statistical information may cause confusion and loss of relevant information
Critical information lock	Read and write again and the sector can be locked independently	Unable to lock information according to user's actual requirement
A sector	4 bytes	2 bytes
Memory	512 bit∼2048 bit	less
Service life	More than 10 years long	short
Read and write times	More than 100,000 times	Less than 50,000 times
Operating temperature	–25 degrees to +70 degrees	Constant temperature state
Material	Flexible package	Simple package
Application	Wide range of applications	Narrow use

The figure for inventory management software Excel file contains the A column for the warehouse only in a product name, B column for the product of production costs, the C column represents the average daily rent expense of certain products, D represents the product transport costs, and E contains the product sales platform rental fee.

On the basis of the design of the related management system, a power company

	A	B	C	D	E
19	Laughing Lumberjack Lager	$ -	$ 518.00	$ 350.00	$ 42.00
20	Longlife Tofu	$ 488.00	$ -	$ -	$ 512.50
21	Louisiana Fiery Hot Pepper Sauce	$ 1,347.36	$ 2,750.69	$ 1,375.62	$ 3,899.51
22	Louisiana Hot Spiced Okra	$ 1,509.60	$ 530.40	$ 68.00	$ 850.00
23	Mozzarella di Giovanni	$ 1,390.00	$ 4,488.20	$ 3,027.60	$ 2,697.00
24	Northwoods Cranberry Sauce	$ -	$ 1,300.00	$ -	$ 2,960.00
25	Ravioli Angelo	$ 499.20	$ 282.75	$ 390.00	$ 984.75
26	Sasquatch Ale	$ 551.60	$ 665.00	$ -	$ 890.40
27	Sir Rodney's Marmalade	$ -	$ 4,252.50	$ 3,061.80	$ -
28	Sir Rodney's Scones	$ 1,462.00	$ 644.00	$ 1,733.00	$ 1,434.00
29	Steeleye Stout	$ 1,310.40	$ 1,368.00	$ 1,323.00	$ 1,273.50
30	Teatime Chocolate Biscuits	$ 943.89	$ 349.60	$ 841.80	$ 851.46
31	Uncle Bob's Organic Dried Pears	$ 1,084.80	$ 1,575.00	$ 2,700.00	$ 3,826.50
32	Veggie-spread	$ 3,202.87	$ 263.40	$ 842.88	$ 2,590.10
33	Grand Total	$ 24,612.91	$ 43,435.04	$ 41,640.74	$ 44,803.26

Fig. 3. Output data table output of automatic warehouse management system of electric power company

was taken as the actual research object, and the main influence parameters of the quasi-automated warehouse management system based on RFID technology and the traditional management system were analyzed statistically. In the study, 100 employees conducted a random survey, and the percentage of the impact of the factors was recorded, the results shown in Table 3.

Table 3. Statistical results of the influencing factors required for this study

Secondary evaluation parameters	Numbering	RFID technology	Traditional technology
Repetitive Usability of Electronic Labels	X1	0.57	0.46
Automatic reading technology	X2	0.84	0.68
The speed of the library	X3	0.77	0.74
Storage speed	X4	0.86	0.82
Receiving operation processing	X5	0.81	0.79
Picking operation	X6	0.58	0.48
Handling of transportation operations	X7	0.63	0.61
Resource Mastery	X8	0.73	0.64
Account alignment	X9	0.82	0.74
The state of an object	X10	0.55	0.49
Storage status	X11	0.45	0.41

Then, the gray correlation analysis model was used to calculate the correlation percentage. The correlation degree of each influencing factor is shown in Table 4. The results show that the main influencing factors of warehouse management system with different application technologies are different. For the RFID technology warehouse management system, automatic reading technology is its most important influencing factors, followed by the handling of receiving operations and further control of some information resources; and for the traditional technology warehouse management system, the influential factors are the statistics of the status of the product, the operation process of product receipt and the control of relevant information resources. Therefore, when using different technologies to build warehouse management system, it is necessary to consider the related factors so as to obtain a more perfect system model.

Table 4. Analysis and sequencing of the correlation degree of the influencing factors of the warehouse management system of RFID technology and traditional technology

Technology factor	RFID technology		Traditional technology	
	Correlation degree r	Sorting	Correlation degree r	Sorting
X1	0.577	4	0.467	9
X2	0.774	1	0.481	8
X3	0.550	5	0.624	4
X4	0.490	8	0.393	11
X5	0.765	2	0.686	2
X6	0.537	6	0.571	5
X7	0.513	7	0.527	6
X8	0.686	3	0.637	3
X9	0.443	9	0.511	7
X10	0.356	10	0.732	1
X11	0.311	11	0.397	10

5. Conclusion

As an important link in the development of modern industry, warehouse management can effectively combine the production of enterprises with the needs of consumers, and provide more accurate development planning for the future development of enterprises. A reasonable warehouse management can effectively improve the efficiency of enterprise operation. The development of RFID technology has provided some positive influence to the improvement of warehouse management system. However, there are still some defects and deficiencies in the warehouse management system of RFID technology in our country, which has restricted the development of the whole enterprise. In this study, the related concepts of RFID technology were sorted out by reading and summarizing the relevant data. On this basis, the quasi-automated warehouse management system of RFID technology was further designed and applied to the actual electric power enterprise. The automatic warehouse management system of RFID technology and traditional technology was contrasted and analyzed, and the gray relational analysis model was used to determine its main influencing factors. The research shows that the quasi-automated warehouse management system of RFID technology has more advantages than traditional technology, and has a positive effect on the development of related enterprises in China. Due to the author's level, there are some shortcomings in this study, but it can provide theoretical basis and reference for the follow-up study.

References

[1] M. Alizadeh, I. Mahdavi, N. Mahdavi-Amiri, S. Shiripour: *A capacitated*

location-allocation problem with stochastic demands using sub-sources: An empirical study. Applied Soft Computing *34* (2015), 551–571.

[2] P. Barge, P. Gay, V. Merlino, C. Tortia: *Item-level radio-frequency identification for the traceability of food products: Application on a dairy product*. Journal of Food Engineering *125* (2014), 119–130.

[3] J. J. Bravo, C. J. Vidal: *Freight transportation function in supply chain optimization models: A critical review of recent trends*. Expert Systems with Applications *40* (2013), No. 17, 6742–6757.

[4] H. Fazlollahtabar, I. Mahdavi, A. Mohajeri: *Applying fuzzy mathematical programming approach to optimize a multiple supply network in uncertain condition with comparative analysis*. Applied Soft Computing *13* (2013), No. 1, 550–562.

[5] M. Hajiaghaei-Keshteli: *The allocation of customers to potential distribution centers in supply chain networks: GA and AIA approaches*. Applied Soft Computing *11* (2011), No. 2, 2069–2078.

[6] J. Dutta, C. Y. Kaya: *A new scalarization and numerical method for constructing the weak pareto front of multi-objective optimization problems*. Optimization *60* (2011), Nos. 8–9, 1091–1104.

[7] M. Ko, A. Tiwari, J. Mehnen: *A review of soft computing applications in supply chain management*. Applied Soft Computing *10* (2010), No. 3, 661–674.

[8] C. J. MacLeod, G. Blackwell, J. Benge: *Reduced pesticide toxicity and increased woody vegetation cover account for enhanced native bird densities in organic orchards*. Journal of Applied Ecology *49* (2012), No. 3, 652–660.

[9] M. Mokhtarinejad, A. Ahmadi, B. Karimi, S. Habib, A. Rahmati: *A novel learning based approach for a new integrated location-routing and scheduling problem within cross-docking considering direct shipment*. Applied Soft Computing *34* (2015), 274–285.

[10] S. Piramuthu, P. FarahaniP, m. Grunow: *RFID-generated traceability for contaminated product recall in perishable food supply networks*. European Journal of Operational Research *225*, (2013), No. 2, 253–262.

[11] S. H. A. Rahmati, V. Hajipour, S. T. A. Niaki: *A soft-computing pareto-based meta-heuristic algorithm for a multi-objective multi-server facility location problem*. Applied Soft Computing *13* (2013), No. 4, 1728–1740.

[12] S. Saavedra, F. Reed-Tsochas, B. Uzzi: *A simple model of bipartite cooperation for ecological and organizational networks*. Nature *457* (2009), No. 7228, 463–466.

[13] M. Soysal, J. M. Bloemhof-Ruwaard, J. G. A. J. van der Vorst: *Modelling food logistics networks with emission considerations: The case of an international beef supply chain*. International Journal of Production Economics *152* (2014), 57–70.

[14] S. Suweis, F. Simini, J. R. Banavar, A. Maritan: *Emergence of structural and dynamical properties of ecological mutualistic networks*. Nature *500* (2013), No. 7463, 449–452.

[15] S. Validi, A. Bhattacharya, P. J. Byrne: *A case analysis of a sustainable food supply chain distribution system—A multi-objective approach*. International Journal of Production Economics *152* (2014), 71–87.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 131–142

Study on prediction and evaluation of Chaohu Lake water quality with MPSO-FSMV

Chi Zhang[1]

Abstract. In consideration of the low prediction accuracy, poor applicability and other weaknesses in traditional water prediction and evaluation models, to realize the high-accuracy water prediction and evaluation, the modified inertia weight was introduced into the particle swarm optimization (PSO) based on fuzzy support vector machine (FSVM) and a MPSO-based optimized FSVM water prediction and evaluation model was proposed. The high-accuracy water prediction and evaluation can be achieved in the condition of optimal parameters by optimizing Gamma and b, FSVM parameters, with MPSO. With the monitoring data of Chaohu Lake water quality from 2010 to 2015 as the object of study, we realized the prediction and evaluation on water quality of Chaohu Lake based on studying the variation trend and time-space relationship of water quality. It is found through comparison of 4 types of water quality evaluation and prediction methods including MPSO-FSVM , PSO-FSVM , FSVM and SVM that the prediction accuracy of MPSO-FSVM reaches up to 95.38 %, which is higher than the prediction accuracies of PSO-FSVM , FSVM and SVM. Therefore, it indicates that MPSO-FSVM has a higher precision and adaptability in water quality evaluation and prediction; the validity and reliability of MPSO-FSVM algorithm was verified, so it can be popularized to other fields of scientific researches and engineering applications.

Key words. PSO, FSVM, water quality evaluation, variation trend, neural network, grey prediction.

1. Introduction

Water quality evaluation and prediction, which are significant research topics and important contents of the foundation of modern environmental science, aim to correctly reflect the quality and pollution of water environment via water quality evaluation, predict the development tendency of water environment quality and thus provide a scientific decision-making basis and method for management, protection and governance of water environment. As the fifth biggest freshwater lake of China, Chaohu Lake, which is located in the central hilly land between the Yangtze River and Huaihe River in Anhui Province, plays an extremely important role in the

[1]Shandong University of Science and Technology, Shandong, 266590, China

economic development of Anhui, so it is of great significance to predict and evaluate the water quality condition of Chaohu Lake.

At present, the methods for water quality prediction mainly consist of the grey prediction method [1], artificial neural network [2] and SVM [3] etc. Although a good result can be obtained by using the grey prediction method for water quality prediction, a high grey level of data will lead to low prediction accuracy; meanwhile, this method is not suitable for long-term prediction of water quality [4–5]. In order to enhance the accuracy in water quality prediction, the literature [6] and literature [7] proposes water quality prediction models respectively based on weighted array, exponential smoothing and GM(1,1) model combination; they all can improve the prediction accuracy to some degree but this method fails to fundamentally solve the problem of prediction error. To enhance the accuracy in water quality prediction, literature [8] puts forward a water evaluation model based on grey metabolism GM(1,1) model. As shown in the empirical result, the prediction accuracy with the method is superior to the traditional GM(1,1) model.

Owing to its accurate nonlinear mapping ability and generalization ability, the artificial neural network has been widely used in scientific research and engineering application. Based on the mapping relation between organic maters in surface river water and main influence factors, literature [9] proposed a water quality prediction model based on BP neural network. The experimental result shows that the method has a high accuracy and generalization ability. Since the small sample data based on BP neural network has low prediction accuracy and is easy to lead local optimum, literature [10] proposed a water quality prediction model based on an improved BP neural network, which improves the prediction accuracy. As for the problem of local optimum of BP neural network, literatures [11–12] optimized the weight, threshold value and structure of BP neural network with the genetic algorithm and established the water quality prediction model of BP neural network improved by the genetic algorithm. It is thus found through comparison that, after optimization with the genetic algorithm, the prediction accuracy and algorithm stability are better than those of the standard BP neural network.

As for the small sample, nonlinear and high-dimension data, the SVM has its unique advantages, so many scholars introduced it into water quality evaluation and made great achievements. However, selection of kernel function and parameter of SVM directly influences the prediction result [13–15]. Since the parameter selection of single SVM prediction model has the disadvantages of low efficiency and dependence on experience, literature extracted the data feature information of water quality with the wavelet analysis and proposed a water quality prediction model based on wavelet transformation and SVRM (Support Vector Regression Machine). Literature brought forward a water quality parameter prediction model based on improved WSVM (weighted support vector machine); the prediction accuracy of this improved method is superior to the unimproved SVM and BP algorithm. As for the small sample and jumping time series data of water quality, literature proposed a water prediction model based on ELPM data pre-processing and LSSVM parameters optimized by simulated annealing algorithm.

Because there are disadvantages such as low prediction accuracy and poor adapt-

ability in the traditional water quality prediction and evaluation model, this paper introduced the modified inertia weight in the PSO based on FSVM and brought forward a water prediction and evaluation model based on FSVM improved by MPSO.

2. FSVM

In 2002, Lin et al. came up with the FSVM algorithm. When the FSVM is used for classification, compared with the training samples of conventional SVM, apart from the features and type identification of sample, the fuzzy membership degree is added to each sample trained with FSVM to reduce the influence of noise point. Let the training sample set be $(x_1, y_1, \mu(x_1))$, $(x_n, y_n, \mu(x_n))$, $x_i \in R^N$, $y_i \in \{-1, 1\}$, $0 < \mu(x_i) \leq 1$. Suppose that $z = \phi(x)$ is the mapping relation by mapping the training sample from the original space RN to the high-dimensional feature space Z. The fuzzy membership degree $\mu(x_i)$ shows the degree of reliability of the sample in a certain type; i signifies the error item in classification in the objective functions of SVM; then, $\mu(x_i)\xi_i$ is the weighted error item. From literature, it is concluded that the optimal classification plane is the optimal solution to the objective function in the formula

$$\begin{cases} \Phi(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}[\mu(x_i)\xi_i], \\ \text{s.t. } y_i\left[\left(w^{\mathrm{T}} z_i\right) + b\right] - 1 + \xi_i \geq 0,\ i = 1, \cdots, n, \\ \xi_i \geq 0, i = 1, \cdots, n \end{cases} . \tag{1}$$

In formula (1), the penalty factor C is a constant; w signifies the weighted coefficient of linear classification function y_i. The discrimination function formula of corresponding optimal face is

$$f(x) = \operatorname{sgn}\left(\sum_{x_i \in SV} w_i K(x_i, x) + b\right). \tag{2}$$

In formula (2), $K(x_i, x)$ signifies the kernel function. The frequently-used Gaussian kernel function is adopted in this paper and now the formula is

$$K(x, z) = \exp\left(-\frac{\Gamma \cdot \|x - z\|^2}{2}\right), \tag{3}$$

where Γ refers to the Gaussian distribution width.

Selection of a proper membership degree in a given problem will directly affect the effect of classification. At present, there are various methods to determine the membership degree, such as linear function, quadratic function, heuristic method and noise distribution. This paper introduces a simple and effective infinitely continuous and differentiable membership function.

If $S^+ = \{x_i \,|y_i = 1\}$ signifies the positive sample set and $S^- = \{x_i \,|y_i = 0$

signifies the negative sample set, the sample novelty membership function is shown below:

$$\mu_i(z)\begin{cases} b\sigma + b\exp\left(\frac{-1}{r^2-\|z-c\|^2}\right) \|z-c\| < r, \\ b\sigma\,(\text{Other})\,. \end{cases} \tag{4}$$

In the formula above, $b = \left[\sigma + \exp\left[\frac{-1}{r^2}\right]\right]^{-1}$ is a sufficiently small positive number; the function is a infinitely continuous and differentiable function, $||z-c||$ signifies the distance between the two points, c is the center of a certain type, and r is the radius of the smallest hypersphere containing the sample set. If the distance from the central point is shorter, the corresponding value is bigger. The value corresponding to the point when $||z-c|| = r$ is the smallest one. As a general rule, there is a long distance between the noise point and the center of such points or $||z-c|| = r$. As for the noise point, the value (fuzzy relation) calculated according to formula (4) is very small, so there is a small influence on correct classification. Thus, the influence of noise on the classification result can be reduced and the FSVM classification accuracy can also be enhanced.

3. MPSO

3.1. PSO

PSO is a swarm intelligence algorithm proposed with the inspiration of bird flock's foraging behavior. In the algorithm, the particle signifies the solution vector and the quality of particle is determined according to the fitness function size. On this basis, continuous renewal of particle position and speed can be realized, and the global optimal searching and optimization can be attained. $X_i = (x_{i1}, x_{i2}, \cdots, x_{id})$ and $V_i = (v_{i1}, v_{i2}, \cdots, v_{id})$ signify the particle positions and speeds, respectively. The updating strategy for particle position and speed is shown in the formulas

$$V_{id}^{(t+1)} = \omega \cdot V_{id}^{(t)} + c_1 r_1 \left(p_{id}^{(t)} - X_{id}^{(t)}\right) + c_2 r_2 \left(p_{gd}^{(t)} - X_{id}^{(t)}\right), \tag{5}$$

$$X_{id}^{(t+1)} = X_{id}^{(t)} + V_{id}^{(t+1)}, \tag{6}$$

and

$$\omega{=}\omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{t_{\max}} \cdot t\,. \tag{7}$$

In the above formulas (5)–(7), $p_{id}^{(t)}$ and $p_{gd}^{(t)}$, respectively, denote the individual optimal solution and global optimal solution of particle at tth moment of iterations, $r_1, r_2 \in (0,1)$ denote the random numbers, c_1, c_2 denote the learning factors; $t_{\max}$ and t, respectively, denote the maximum iterations and current iterations, w signifies the inertia weight, and $w_{\max}$ and $w_{\min}$, respectively, denote the maximum and minimum values of inertia weight.

3.2. MPSO

To avoid local optimization of PSO and accelerate the global searching ability of PSO, this paper introduced the nonlinear dynamic inertia weight coefficient in the inertia weight modified by PSO and proposed a modified particle swarm algorithm (MPSO).

4. MPSO-FSVM-based water quality prediction and evaluation of Chaohu Lake

4.1. Objective function

On the premise that the optimal water quality evaluation result is guaranteed, use MPSO to optimize the Γ and b, parameters of FSVM, so as to achieve the self-adaptive selection of FSVM parameters. If the actual water quality type at tth moment is $y(t)$ and the predicted water quality type is $\hat{y}(t)$, the difference between the actual water quality type $y(t) - \hat{y}(t)$.

As for the evaluation on water quality types, the actual water quality type data is n; use MPSO to optimize Γ and b, parameters of FSVM so as to minimize the quadratic sum of difference between the actual water quality type with FSVM and predicted water quality type.

4.2. Steps of algorithm

The steps of water quality evaluation with the FSVM optimized by MPSO are as follows:

Step 1: Normalize the water quality sample data. Establish training samples and test samples.

Step 2: Set the population size ($popsize$), learning factors c_1 and c_2 and maximum iterations $\max_{\text{gen}}$ of MPSO.

Step 3: Input the established training samples into the FSVM; calculate the functional value of particle fitness according to the objective function formula (7) to search for the individual and global optimal particle position and optimal value.

Step 4: Update the particle speed and position.

Step 5: Calculate the fitness and update position and speed simultaneously.

Step 6: If $\max_{\text{gen}}$, save the optimal solution. Conversely, if gen = gen + 1 and go to Step 4.

5. Experimental analysis

5.1. Water quality evaluation indexes

Water quality evaluation is to calculate and determine the water quality grade of sampling water samples through a certain mathematical model according to the

water quality evaluation standard and all indexes of sampling water samples. There are numerous indexes for water quality analysis. Based on the quality standards for surface water environment, this paper used 6 water quality evaluation indexes including ammonia nitrogen, dissolved oxygen (DO), chemical oxygen demand (COD), permanganate index, total phosphorus and total nitrogen; their corresponding water quality grades are listed in Table 1.

Table 1. Water quality grades and content standards

Type	Type 1	Type 2	Type 3	Type 4	Type 5
Ammonia nitrogen (mg/l) $\leq$	0.15	0.50	1.0	1.5	2.0
DO (mg/l) $\geq$	7.5	6.0	5.0	3.0	2.0
COD /(mg/l) $\leq$	15	15	20	30	40
Permanganate index (mg/l) $\leq$	2.0	4.0	6.0	10	15
hline Total phosphorus (mg/l) $\leq$	0.02	0.10	0.20	0.30	0.40
Total nitrogen (mg/l) $\leq$	0.20	0.50	1.0	1.5	2.0

5.2. Data sources

Water samples of Chaohu Lake were collected as objects of evaluation of water quality. The sampling water intakes were respectively the intersection between Nanfei River and Chaohu Lake, the intersection between Paihe River and Chaohu Lake and dam outlet of Chaohu River. The longitudes and latitudes of sampling sites are listed in Table 2. The sampling time of water quality of Chaohu River was from 2010 to 2015. The sampling frequency was once per quarter. The change tendencies of all indexes of sampling water are shown in Figs. 1 and 2.

Table 2. Longitudes and latitudes of sampling sites

S/N	Sampling site	Longitude	Latitude
1#	Intersection between Nanfei River and Chaohu Lake	117°24′40″	31°42′15″
2#	Intersection between Paihe River and Chaohu Lake	117°18′15″	31°41′30″
3#	Dam outlet of Chaohu River	117°51′46″	31°34′18″

5.3. Diagram for spatial distribution of water quality

To visually observe the relations among various water quality indexes, the diagrams for spatial relations among all water quality indexes were drawn and are shown in Fig. 3.

5.4. Empirical results

To verify the validity and reliability of algorithm proposed in this paper, the water quality data of three sampling sites in Chaohu Lake from 2010 to 2015 were

used as objects of research.

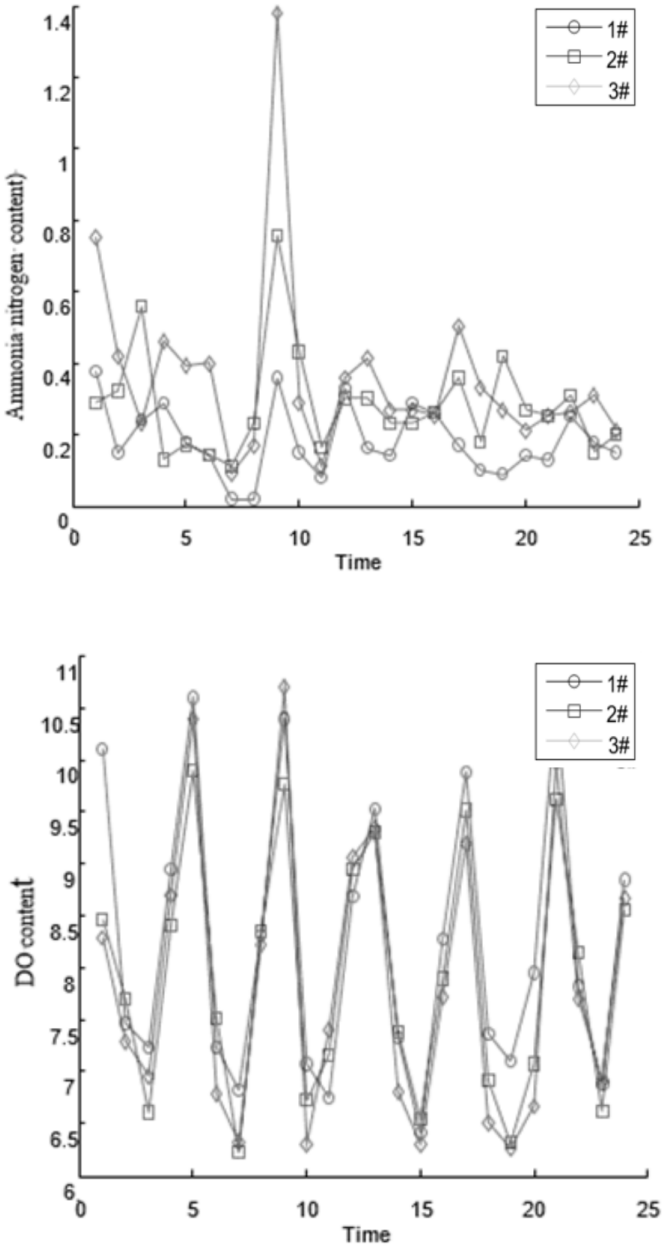


Fig. 1. Tendency chart for changes in water index data: up–Amonnia nitrogen content (mg/l), bottom–DO content (mg/l)

The parameter setting of MPSO is shown below. The population size is 20, the

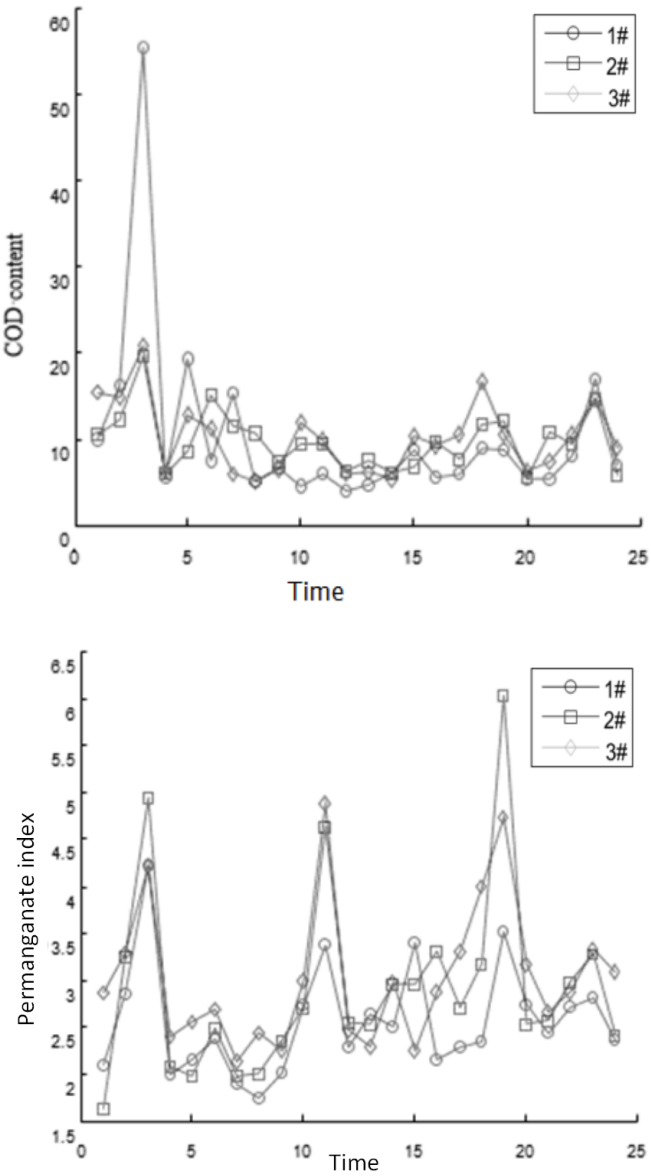


Fig. 2. Tendency chart for changes in water index data: up–COD content (mg/l), bottom–permanganate content (mg/l)

maximum iterations is 100, $c_1 = c_2 = 0.5$, $V_{\max} = 5$, $V_{\min} = -5\,\%$. The prediction and evaluation results of water quality of Chaohu Lake with the FSVM optimized by MPSO are shown in Figs. 4–5 and Table 3 and Table 4.

It can be seen from Table 3 that MPSO-FSVM, PSO-FSVM and FSVMall have

higher accuracies in prediction and evaluation of water qualities of Chaohu Lake than SVM.

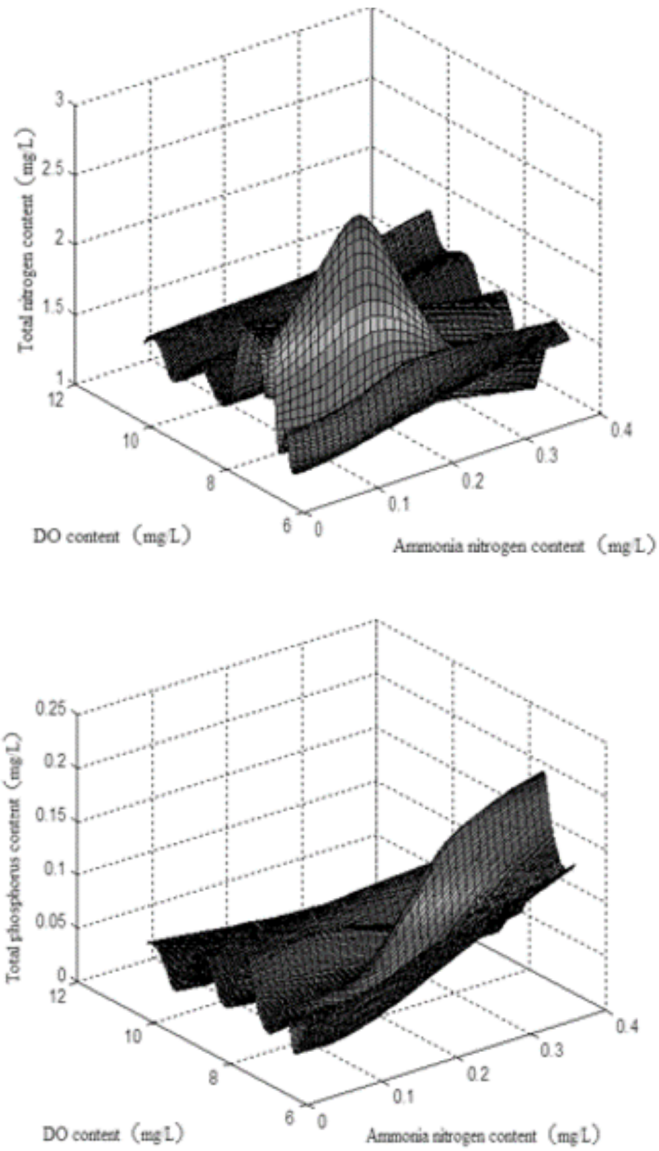


Fig. 3. Diagram for spatial relations among 1#, 2# and 3# sampling sites

There is no big difference in the accuracy rate of water quality grade classification between MPSO-FSVM and PSO-FSVM; however, PSO may easily lead to local optimization and has a poor stability; MPSO, which as strong global optimization ability, can effectively avoid the problem of local optimization. In addition, even

though very high prediction accuracy can also be realized with FSVM, this method is time-consuming and inefficient; besides, it is required to set the searching scope in advance with this method, so it is difficult to control it and to guarantee the accuracy.

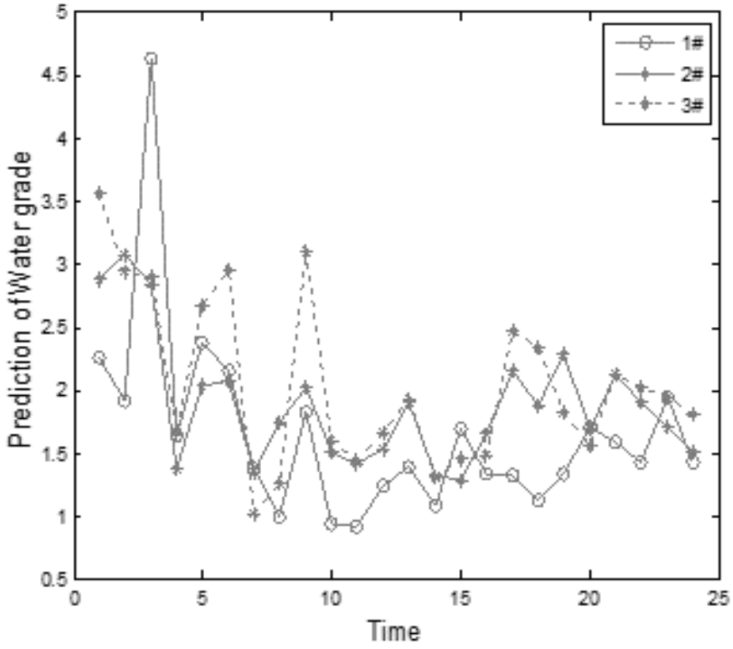


Fig. 4. Water evaluation results

Table 3. Evaluation results of water quality of Chaohu Lake

Time	2010.1	2010.2	2010.3	2010.4	2011.1	2011.2	2011.3	2011.4
1#	3	4	5	3	4	3	2	2
2#	4	4	3	3	3	3	2	3
3#	4	4	3	3	4	3	2	2

Table 4. Comparison of evaluation results of different water qualities

Method	MPSO-FSVM	PSO-FSVM	FSVM	SVM
Quantity of erroneous judgments	3	4	6	9
Accuracy	95.38 %	93.85 %	90.76 %	86.15 %

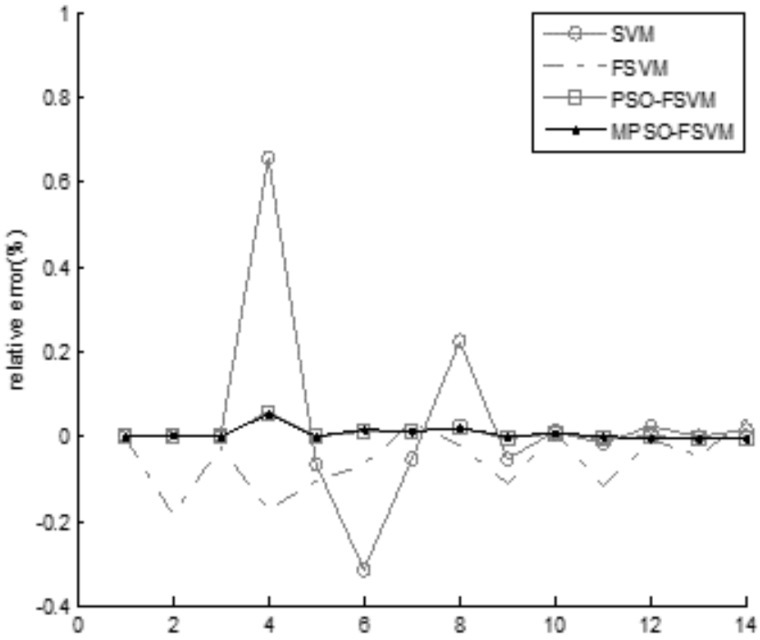


Fig. 5. Error comparison of different evaluation methods

6. Conclusion

In consideration of the low prediction accuracy and poor applicability in traditional water prediction and evaluation models, to realize high-accuracy water quality prediction and evaluation, a water quality prediction and evaluation model based on FSVM optimized by MPSO was put forward on the basis of FSVM. The MPSO was used to optimize Gamma and b, parameters of FSVM and realize the high-accuracy water quality prediction and evaluation under the condition of optimal parameters. The monitoring data of water quality of Chaohu Lake from 2010 to 2015 were selected as objects of research; the water quality prediction and evaluation of Chaohu Lake were achieved based on studying the water quality change trend and time-space relationships. It is found through comparison of 4 types of water quality evaluation and prediction methods including MPSO-FSVM , PSO-FSVM , FSVM and SVM that the prediction accuracy of MPSO-FSVM reaches up to 95.38 %, which is higher than the prediction accuracies of PSO-FSVM , FSVM and SVM. Therefore, it indicates that MPSO-FSVM has a higher precision and adaptability in water quality evaluation and prediction; and it also has a better effect.

References

[1] H. B. Xue, Y. Wei: *Optimize an optimal GM(1,1) based on the discrete function*

with exponential law once again. Mathematics in Practice and Theory *1* (2009), No. 1, 242–246.

[2] X. H. Yao, M. R. Fei, K. Li, H. Kong, B. Zhao: *Recognition of blue-green algae in lakes using distributive genetic algorithm-based neural networks*. Neurocomputing *70* (2007), Nos. 4–6, 641–647.

[3] X. Liu, F. Dong, G. He, J. Liu: *Use of PCA-RBF model for prediction of chlorophyll-a in Yuqiao Reservoir in the Haihe River Basin, China*. Water Science & Technology: Water Supply *14* (2014), No. 1, 73–80.

[4] U. Natarajan, V. M. Periasamy, R. Saravanan: *Application of particle swarm optimisation in artificial neural network for the prediction of tool life*. The International Journal of Advanced Manufacturing Technology *31* (2007), Nos. 9–10, 871–876.

[5] J. G. Yang, S. Y. Wen, H. Zhou, K.z,F. Cen: *An optimized BP network model using genetic algorithm for predicting the ignition-stability index of pulverized coal*. Journal of Power Engineering *26* (2006), No. 1, 81–83.

[6] K. P. Wu, S. D. Wang: *Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space*. Pattern Recognition *42* (2009), No. 5, 710–717.

[7] H. Yoon, S. C. Jun, Y. Hyun, G. O. Bae, K. K. Lee: *A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer*. Journal of Hydrology *396* (2011), Nos. 1–2, 128–138.

[8] D. KarabogaD, B. Akay: *A survey: Algorithms simulating bee swarm intelligence*. Artificial Intelligence Review *31* 2009), Nos. 1–4, 61–85.

[9] H. S. Chang: *Converging marriage in honey-bees optimization and application to stochastic dynamic programming*. Journal of Global Optimization *35* (2006), No. 3, 423–441.

[10] B. Parinet, A. Lhote, B. Legube: *Principal component analysis: An appropriate tool for water quality evaluation and management—application to a tropical lake system*. Ecological Modelling *178*, (2004), Nos. 3–4, 295–311.

[11] L. C. Hsu: *Applying the Grey prediction model to the global integrated circuit industry*. Technological Forecasting and Social Change *70* (2003), No. 6, 563–574.

[12] F. Moatar, M. Meybeck: *Compared performances of different algorithms for estimating annual nutrient loads discharged by the eutrophic river Loire*. Hydrological Processes *19* (2005), No. 2, 429–444.

[13] W. Yan, S. Zhang, P. Sun, S. P. Seitzinger: *How do nitrogen inputs to the Changjiang basin impact the Changjiang River nitrate: A temporal analysis for 1968–1997*. Global Biogeochemical Cycles *17* (2003), No. 4, 1091–1100.

[14] M. Aminian, F. Aminian: *Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor*. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing *47* (2000), No. 2, 151–156.

[15] F Aminian, M. Aminian: *Fault diagnosis of analog circuits using bayesian neural networks with wavelet transform as preprocessor*. Journal of Electronic Testing *17* (2001), No. 1, 29–36.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 143–152

Application of improved Bayesian model based on cosine similarity weighted in prediction of disease classification[1]

Zhaochun Ran[2]

Abstract. With the development of information technology, Bias's classification prediction function has been gradually applied to finance, medical and other fields. Therefore, the application of Bayesian model based on cosine similarity weighted modified classifier in disease classification prediction was studied in this paper. By constructing an improved Bias model based on cosine similarity weighting and performing it on the Spark platform, the accuracy of the improved model was compared with that of the traditional Bias model. Taking hypertension and hyperlipidemia as an example, the prediction results of Bias classification under single machine and cluster mode were compared and analyzed. The results show that under the cluster Spark platform, the improved Bayes model has the highest efficiency in disease prediction classification.

Key words. Cosine similarity, weighted improvement, Bayesian model, disease classification prediction.

1. Introduction

With the continuous development of medical diagnosis, it has fully formed a reciprocal relationship with information technology. The development of computer technology provides a new solution for the development of the medical and health industry, which can be used in medical research, clinical medicine and basic medical treatment [1]. It helps to speed up the development of medicine, promote medical information, reduce medical expenses, and fully protect people's health. At present, many field experts have used computers to build a model for predicting specific diseases, and obtained good prediction results [2].

Classification is the most important part of machine learning research. The model

[1]This work was supported by the Hainan Natural Science Foundation of China, Project number: 20156231, project name: Study on health expert system model based on Bayes algorithm.

[2]Haikou College of Economics, Haikou, 571127, China

established by classification method can analyze unknown input model according to known classification knowledge, and finally determine the attribution category of input model. As an effective and practical forecasting model, Bias has been widely used in people's production and life [3]. In order to improve the classification effect of Naive Bayesian algorithm, scholars have considered the independence assumption condition weakening property. Bias algorithm based on cosine similarity has improved the simple Bias classification algorithm from the point of view of local learning and structure expansion. The frequent item sets mining algorithm is implemented by using cloud framework for large-scale sample data. Finally, the improved algorithm is applied to the problem of disease prediction [4].

2. State of the art

As a classification algorithm, the Naive Bayes algorithm has obvious advantages with a solid theoretical basis, high computational efficiency and high accuracy [5]. In order to reduce the complexity of the model, it is assumed that the Naive Bayes algorithm is independent of each other, which can effectively reduce the complexity of the computation process. In the independent assumption of attributes, the decision attributes for each delegate weight are equal (both are 1) [6]. However, this assumption of independence can be satisfied in very few cases. The conditional attributes are not exactly the same as the weights of the decision attributes, which can lead to lower classification accuracy [7]. To solve this problem, we use weighted Naive Bayes model to assign different weights to each conditional attribute, and relax the independent hypothesis, so as to improve the performance of the classifier on the basis of maintaining the original model. A weighted Bayesian algorithm based on cosine similarity is proposed, which takes different training samples into account to classify the decision weights [8]. Use cosine similarity to measure the distance of samples, and select the best subset of training samples. Moreover, use similarity values as training samples to train and modify Bayesian models.

3. Methodology

3.1. Improved Bayesian model based on weighted cosine similarity

For the traditional Naive Bias classification algorithm (NB), the conditional independent assumption is difficult to satisfy in practical applications. The association between attributes always exists, and it has a bearing on the results. The weighted Naive Bayesian classification algorithm is that the final classification is conducted based on different attributes, and then the corresponding weights are extended for the original algorithm to improve the performance of the Naive Bayes algorithm [9]. The weighted Naive Bias classification algorithm reduces the influence of conditional attribute independence by assigning different weights for different conditional attributes. The accuracy of the posterior probability is calculated as follows

$$P(C_i|H) = P(C_i) \prod P(ak|C_i)\,. \tag{1}$$

Improved Bayesian model based on weighted cosine similarity uses the cosine similarity as the weights to optimize the weights of attributes. In the training set A, each sample A_i contains n conditional attribute fields, each corresponding to a category. The training set can be represented by multiple conditions, attributes, and classes. Treat each condition attribute field as a random variable X_i corresponding to the conditional attribute fields, and treat the class as a random variable Y. The random variables are corresponding to the conditional attributes, and then the distributions of the two values are corresponding, that is to say, for the random variables, all the values contained in them are a_{ij} [10]. Through this transformation, the original training set can be represented as a set of random variables that satisfy a particular probability distribution. Then, the correlation between the training set attributes and categories (A_i, C) is determined by measuring the correlation between the two random variables (X_i, Y). The class attributes C and the characteristic attribute A_i of the training set can be represented by the decision attribute Y and the characteristic attribute X_i. Cosine similarity is introduced as a measure of the correlation degree between two random variables. Cosine similarity is a measure of the difference between two sample vectors. Firstly, the vector space is used to represent the attributes of the sample, and then the similarity between the two vectors is measured by calculating the spatial angle cosine of the two vectors [11]. The smaller the angle of the two vectors (the closer to 0), the greater the cosine of the vector is, which indicates the higher the similarity between them. For random variables X_i and Y, the similarity of two vectors is calculated by the cosine similarity formula.

$$\cos\theta = \cos < Xi, Y >= \frac{\sum_{i=1}^{n} a_{ij} \times y_j}{\sqrt{\sum_{1}^{n} a_{ij}^2} \times \sqrt{\sum_{j=1}^{n} y_j^2}}\,. \tag{2}$$

Correlation analysis is used to measure the degree of correlation between the two variables. Two variables need correlation in the correlation analysis. In machine learning, the correlation of two vectors is usually measured by similarity or distance.

The key to the weighted Bayesian algorithm based on cosine similarity is to find the cosine similarity weights between the conditional attributes and the category, which is denoted as NNB. Firstly, the data sets that need to be processed are discretized and filled with missing values. Secondly, for the classification phase, it is needed to jump to the classification step directly. If it is the training phase, the training sample data set is carried out. The third is the statistical table learning: according to the training sample in the data set, the number of training samples, attributes and their categories is counted.

The fourth is the cosine similarity weight learning: all training samples are traversed, the cosine similarity between each conditional attribute and classification class is calculated according to formula, and the value of formula (2) is taken as the weight coefficient of a_i [12]. Then the attribute weight statistics constitutes, and finally the classification is carried out. The training phase of the Bayesian model for

disease classification prediction is as follows.

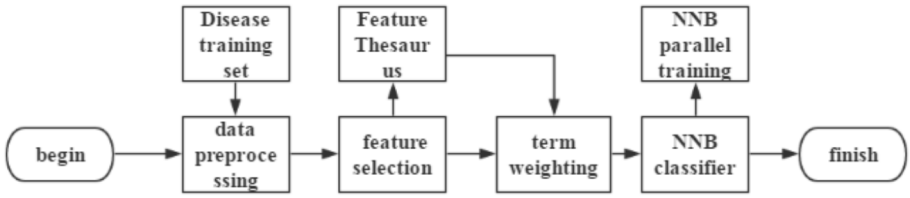


Fig. 1. Bayesian model training phase for disease classification prediction

The application phase is shown in Fig. 2.

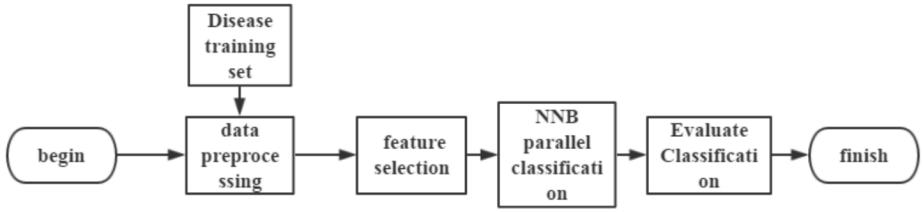


Fig. 2. Application phase of the Bayesian model test for disease classification prediction

3.2. Bias classification algorithm based on cloud computing platform Spark

The parallel Bias classification algorithm is implemented using cloud computing platform Spark. In parallel design ideas, it is similar to the "Map" and "Reduce" ideas in MapReduce. Firstly, we calculate the number of different characters in each category by "Map". Secondly, we calculate the parameters by "Reduce" superposition calculation. The parallelization process of Bias classification algorithm is realized by the Spark kernel scheduling of cloud computing platform. In the process of scheduling, the data sets are automatically allocated according to the number of nodes in the cluster, and then the tasks are executed in parallel to realize the parallelization of the Bias algorithm [13].

In the model training phase, it first reads the file from the HDFS (the file already processed), creates a new RDD, and performs caching operations locally to cache the RDD data. Then, by Map operations, mark each row, stack the value of the StackByKey operation, and count the frequency of each class, the number of classes as well as the frequency of each feature of each class. Finally, the model is trained by these parameters [14]. In the classification test phase, the mapping function is used to predict the classification of each test sample in parallel and calculate the final result. In the model training phase, the file is first read from the HDFS (a data set file that has been processed). Through the set function to obtain classification information, the data file is imported into the program. In the classification prediction phase, the calculation process of each classification can be computed one by one using Map, and then mapped directly to the output of the result, because the calculation of the

sample does not affect each other in the classification prediction stage. In addition to the classification of the prediction process, the training process, query and processing that are generated mainly by two frequency statistics tables [15].

4. Result analysis and discussion

4.1. Comparison and analysis of improved Bias algorithm based on cosine similarity weighting and traditional Bias algorithm

Different data sets (Letter, Lymphography, Segment, Credit-g data sets) were selected, the higher the average accuracy was, and the better the classification performance was. In the contrast experiment, the used classifiers included Naive Bayesian classifier algorithm (NB), Bayesian Network (BayesNet), and the improved Bias classification algorithm (NNB) proposed. The correct rate of each classifier after running was recorded.

Table 1. Accuracy comparison of experimental results based on different algorithm classifiers

Letter	0.7323	0.7393	0.7554
Lymphography	0.8022	0.7814	0.8443
Segment	0.8895	0.9141	0.9046
Credit-g	0.7542	0.7856	0.8343

The improved Bias algorithm based on cosine similarity weighted (NNB) has good performance in accuracy, and has a certain improvement compared with the traditional Bias classification algorithm and Bias network. Credit-g data sets have great dependence on each attribute, so the accuracy of NNB is obviously better than that of the former two algorithms. The improved Bayesian algorithm based on cosine similarity weighted (NNB) adds weights to the conditional attributes to improve the accuracy, but it needs to calculate the time cost of the weights. Therefore, the execution time of the improved algorithm is longer than that of the Naive Bayes algorithm. It can be concluded that the attributing weight based on the traditional Naive Bayesian algorithm can be carried out, so that the correlation between features and categories can be established by cosine similarity, which can alleviate the impact brought by the conditional independence assumption and improve the accuracy of disease classification to some extent.

4.2. Prediction analysis of disease classification under single machine condition

The disease classification prediction phase includes preprocessing, feature selection, classifier classification, and final result output. The classifier is constructed by training phase. The predicted categories of diseases are classified according to the classification and classified into the most relevant categories. Finally, the output is

evaluated according to the classification and evaluation criteria.

The experiment was done by single machine and cluster. A single experiment completed the calculation of the accuracy of the improved algorithm. The cluster mainly completed the cloud computing platform environment to improve the classification speed experiment.

A single NNB classifier was used, and the traditional Decision stump, C4.5, REPtree and Bayesian classification method based on cosine similarity weight prediction model were selected to forecast two kinds of common diseases in the elderly (high blood lipid and hypertension). The classification prediction accuracy and related error statistics are shown in Table 2.

Table 2. Prediction accuracy of classification and related error statistics

Classification method	Accuracy rate	Kappa statistics	Mean absolute error	Root mean square error	Relative absolute error	Relative square root error
Decision stump	58.97 %	0.4019	0.2541	0.3594	69.79 %	84.24 %
C4.5	82.05 %	0.7536	0.0914	0.2739	25.09 %	65.47 %
REPtree	82.05 %	0.7522	0.1326	0.2843	36.40 %	66.63 %
Improved Bayesian prediction model based on cosine similarity weighting	88.72 %	0.8071	0.1682	0.3013	40.76 %	64.38 %

The classification of diseases in this section was realized based on the traditional Bias classification and the improved classifier based on the improved classifier. The accuracy, recall rate and F1 value of disease category prediction under different feature dimensions were investigated. From the above table, the accuracy of Bayesian prediction based on cosine similarity weighted was higher, which reached 88.72 %. The predictive accuracy of Decision stump was low with only 58.97 %, and its predictive performance was poor because of the poor classification performance of hypertension or hyperlipidemia. In addition, the improvement of Bayesian NNB based on cosine similarity weighting was similar to that of NB in parallelization. Additional parallel computation of weighting coefficients was needed to realize NNB parallel algorithm.

It was assumed that the disease data set size was 200 thousand lesion data. In feature selection, the number of feature dimensions from 7000 to 11000 was set at intervals of 500, which was realized respectively through NB and NNB classification algorithm. Accuracy, recall rate, and F1 values are as follows.

Table 3. Classification prediction results of different feature categories of disease category

Characteristic dimension	NB			NNB		
	Accuracy	Recall rate	F1value	Accuracy	Recall rate	F1value
7000	0.8463	0.7935	0.8191	0.8979	0.8310	0.8632
7500	0.8501	0.8021	0.8254	0.9068	0.8367	0.8703
8000	0.8568	0.8095	0.8325	0.9132	0.8425	0.8764
8500	0.8621	0.8136	0.8372	0.9191	0.8487	0.8825
9000	0.8678	0.8186	0.8425	0.9234	0.8531	0.8869
9500	0.8724	0.8213	0.8461	0.9203	0.8501	0.8838
10000	0.8620	0.8203	0.8406	0.9168	0.8456	0.8797
10500	0.8603	0.8186	0.8389	0.9132	0.8413	0.8758
11000	0.8534	0.8103	0.8312	0.9086	0.8326	0.8689

From the above table, in the classification of diseases, the improved Bias classification algorithm based on cosine similarity was superior to the traditional Bias classification algorithm NB in NNB. For feature selection of different feature dimensions, the accuracy, recall rate and F1 value of NNB were higher when the feature dimension was 9000, while NB was higher when the feature dimension was 9500.

The accuracy rate, recall rate and F1 value of NNB and NB in the highest accuracy rate were demonstrated through the histogram form, so as to observe the experimental results more intuitively, as shown in Fig. 3.

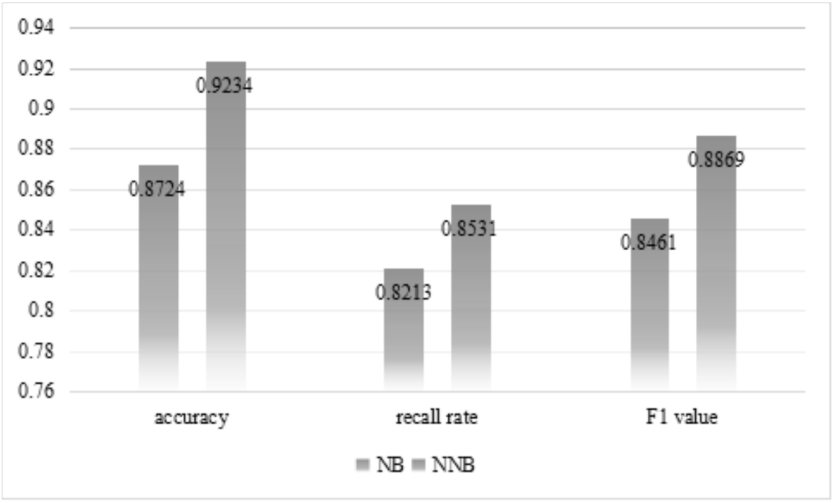


Fig. 3. Accuracy, recall rate and F1 value

Compared with the traditional algorithm NB, the cosine similarity weighted improved algorithm proposed in this paper has certain advantages in accuracy, recall

and F1 value of NNB. The performance of the classifier is improved by means of cosine similarity weighted attributes, and the accuracy of disease classification prediction is improved.

4.3. Prediction analysis of disease classification under cluster condition

The data set size was set, and the training time of different data sets in the cloud computing Spark platform was compared. The running time of the program was observed. Spark adopted 1 main node and 4 slave nodes. Run time results are shown in Table 4.

Table 4. Running time of different data in single machine and Spark

Data quantity (strip)	Single machine running time (s)	Spark run time (s)
5000	2.89	18.56
10000	5.65	23.81
50000	11.46	34.25
100000	24.51	41.39
200000	70.36	52.85
500000	273.75	75.54

The experimental results corresponding to the line chart are shown in Fig. 4.

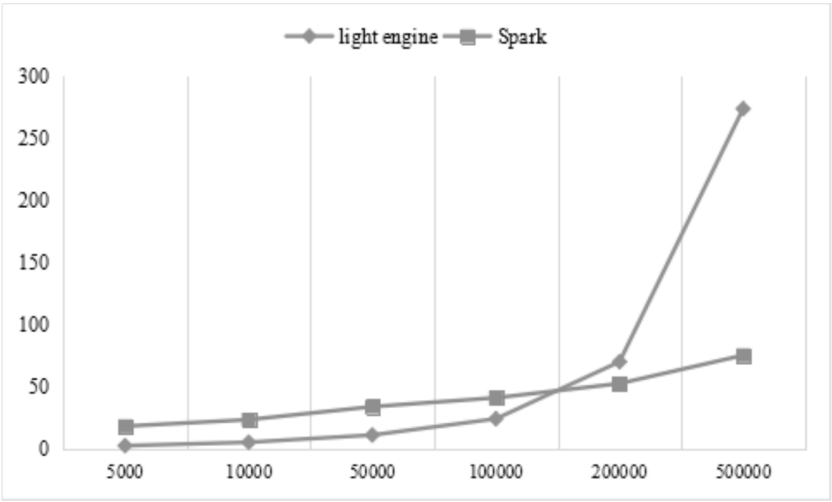


Fig. 4. Line chart of run time results

As can be seen from Fig. 4, in the case of small data (<10 million), the independent running speed of the disease classification program is significantly faster than that of the cloud computing platform Spark. It is mainly because in the cost of node communication, the communication and scheduling costs of the cloud computing platform Spark have relatively large proportion of the total running time with

a small amount of data. When the amount of data increases to some extent, the cloud computing platform Spark is more prominent at runtime than stand-alone. As can be seen from the diagram, when the amount of data reaches 200 thousand, the speed of cloud computing platform Spark is slightly higher than independence. When the amount of data increases to 500 thousand, the running time of program is obviously increased, and the number of cloud computing platform Spark is not increased. Therefore, the cloud computing platform Spark has obvious advantages in dealing with a large number of text classifications. When the data set increases to a certain extent, a single machine environment, such as a memory overflow problem, will appear. And because Spark cache mechanism and RDD conversion operation will greatly accelerate the implementation time, the parallel operation advantages of the Spark cluster are reflected. From the analysis of the current experimental results, it can be seen that the improved Bias classification algorithm based on cosine similarity weighted method has good performance in disease classification based on Spark platform. The improved algorithm is well ported to Spark. In the case of large amounts of data, the Spark based disease classification program can show good running speed, which indicates that the disease classification process works effectively in Spark.

5. Conclusion

Data mining has been widely used in many fields since it was put forward, and its technology is quite mature. As an important technology of data mining, classification technology plays an important role in practice. Nowadays, with the development of information technology, the hospital has gradually realized paperless office, and the database has accumulated a large amount of data. The use of data mining methods for disease classification prediction and clinical decision-making services has a special significance for chronic disease research. In this paper, the application of Bayesian model based on cosine similarity weighted improvement in the prediction of disease classification was analyzed. Firstly, the current situation of medical information mining technology and the research situation of Bayesian prediction algorithm in China were introduced. Secondly, the improved Bias model based on cosine similarity weighting and the Bias classification algorithms based on cloud computing platform Spark were introduced. Finally, the improved Bias algorithm based on cosine similarity weighted was compared with the traditional Bias algorithm, and the disease classification under single machine and cluster condition was predicted. The experimental results show that in the methods of using single classifier and ensemble classifier for disease prediction, the improved Bayesian model ensemble based on cosine similarity weighting has high classification accuracy of classifier, which still needs to promote the application of cloud computing platform in diversity.

References

[1] K. O. Akande, T. O. Owolabi, S. O. Olatunji: *Investigating the effect of correlation-based feature selection on the performance of support vector machines in reservoir characterization.* Journal of Natural Gas Science and Engineering *22* (2015), 515–522.

[2] G. C. Garriga, R. Khardon, L. D. Raedt: *Mining closed patterns in relational, graph and network data.* Annals of Mathematics and Artificial Intelligence *69* (2013), No. 4, 315–342.

[3] J. Nahar, T. Imam, K. S. Tickle, Y. P. P. Chen: *Association rule mining to detect factors which contribute to heart disease in males and females.* Expert Systems with Applications *40* (2013), No. 4, 1086–1093.

[4] F. Nori, M. Deypir, M. Hadi, K. Ziarati: *A new sliding window based algorithm for frequent closed itemset mining over data streams.* International eConference on Computer and Knowledge Engineering (ICCKE), 13–14 October 2011, Mashhad, Iran, IEEE Conference Publications (2011), 249–253.

[5] M. Cusumano: *Cloud computing and SaaS as new computing platforms.* Communications of the ACM, Technology strategy and management *53* (2010), No. 4, 27–29.

[6] S. Maurya, S. K. Shrivastava: *Kalman filter based flexible sliding window algorithm for mining frequent itemset over data stream.* International Journal of Computer Applications *111* (2015), No. 9, 13–19.

[7] M. Deypir, M. H. Sadreddini, M. Tarahomi: *An efficient sliding window based algorithm for adaptive frequent itemset mining over data streams.* Journal of Information Science and Engineering *29* (2013), No. 5, 1001–1020.

[8] J. Chen, B. Zhou, L. Chen, X. Wang, Y. Ding: *Finding frequent closed itemsets in sliding window in linear time.* IEICE Transactions on Information and Systems *E91-D* (2008), No. 10, 2406–2418.

[9] V. B. Wicaksono, R. Saptono, S. W. Sihwi: *Analisis perbandingan metode vector space model dan weighted tree similarity dengan cosine similarity pada kasus pencarian informasi pedoman pengobatan dasar di puskesmas.* Jurnal ITSMART *4* (2015), No. 2, paper 73.

[10] J. Zhu, Y. Ma, Q. Qin, C. Zheng, Y. Hu: *Adaptive weighted real-time compressive tracking.* IET Computer Vision *8*, (2014), No. 6, 740–752.

[11] Y. Wu, N. Jia, J. Sun: *Real-time multi-scale tracking based on compressive sensing.* Visual Computer *31* (2015), No. 4, 417–484.

[12] K. S. Lin: *Fuzzy similarity matching method for interior design drawing recommendation.* Review of Socionetwork Strategies *10* (2016), No. 1, 17–32.

[13] K. Khan, B. Baharudin, A. Khan, A. Ullah: *Mining opinion components from unstructured reviews: A review.* Journal of King Saud University - Computer and Information Sciences *26* (2014), No. 3, 258–275.

[14] A. Triana, R. Saptono, M. E. Sulistyo: *Pemanfaatan metode vector space model dan metode cosine similarity pada fitur deteksi hama dan penyakit tanaman padi.* Jurnal Teknologi Informasi ITSMAR *3* (2014), No. 2, paper 90.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 153–164

Servo control algorithm of handling manipulator based on disturbing observer

Qian Feng[1], Lei Wang[1], Liye Su[1], Wenxue Huang[1], Bing Hu[1]

Abstract. The purpose is to study the servo control algorithm of the handling robot based on the disturbance observer. In order to meet the high precision positioning requirements of the manipulator, the control algorithm of the motor is discussed and studied. The mathematical model of permanent magnet synchronous motor is analyzed. Aiming at the problem of controller tracking performance degradation, a permanent magnet synchronous motor disturbance observer and a speed controller based on fuzzy control are designed. The speed and current double closed loop control system is constructed. By using a disturbance observer, the system disturbance is estimated. According to this, the torque and current compensation are generated, and feed forward control of the speed loop is corrected. The final amount of current is optimized. The results show that the control algorithm has good dynamic performance and steady-state performance. It provides a theoretical basis for high-precision positioning of manipulator. Therefore, it can be concluded that the system achieves the purpose of suppressing the disturbance.

Key words. Manipulator, observer, servo control algorithm, permanent magnet synchronous motor.

1. Introduction

With the rapid development of modern industry and the continuous advancement of machine replacement, the application of manipulators is becoming more and more extensive [1]. It is widely used in marine development, space exploration, aerospace, machining, and civil production. In addition, all kinds of new mechanical hand are still emerging. The robot can complete many repetitive and tedious works. It can also replace mankind in dangerous environments to complete dangerous work [2]. Therefore, the emergence and application of manipulators not only greatly improve the level of human production, but also improve the working environment of workers. Nowadays, many countries begin to pay more attention to the research and

[1]North China University of Science and Technology, Tangshan, 063210, China

exploration of machine hand and robot [3]. It has achieved breakthrough results, while it also has created many practical and innovative robots or robotic devices. In recent years, the equipment manufacturing industry in our country has developed rapidly [4]. However, the cost of employing people is increasing year by year, and a considerable extent has affected the sustained development of China's manufacturing industry. With the gradual promotion of industrial transformation and upgrading, the "China made 2025" project was launched. In order to reduce costs and improve efficiency, the demand for intelligent equipment represented by industrial robots in the domestic manufacturing industry has also ushered in an unprecedented period of explosion [5]. However, compared with the expensive and complex robot, high cost performance, simple structure and convenient manufacturing will be the primary choice for the majority of minor enterprises [6]. To sum up, based on modern mechanical design technology, computer technology and automatic control technology, it is of practical significance to develop a simple, low-cost and high-precision manipulator for small and medium enterprises.

2. State of the art

Servo control technology is a technique that can effectively track and control the position, speed and acceleration of moving object [7]. The servo system is able to track incoming instructions and perform actions. Therefore, it has higher control accuracy. In the process of industrial production, people pay more and more attention to the requirement of accuracy. Today, many electromechanical devices have servo control functions, such as high-precision manipulator, robot, and finishing lathe and so on [8]. The type of servo control system is numerous and the structure is different. From the point of view of automatic control principle, the servo control system usually includes five parts: controller, controlled object, execution link, comparison link and detection link. The component diagram of the servo system is shown in Fig. 1.

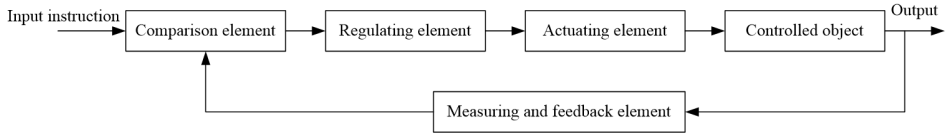


Fig. 1. Schematic diagram of the composition of the servo system

The servo system generally has good dynamic and static performance. Most of the movement of the actuator is driven directly by the motor or indirectly through the transmission mechanism. In order to ensure the accuracy of the positioning of the moving parts, the servo system needs a very high positioning accuracy. In the process of moving the mechanical parts, it is often necessary to control or adjust the speed of the moving unit. Therefore, a high-precision servo system also needs to have better speed and anti-disturbance ability. For the control performance of the servo control system, the speed range and the torque output at low speed need to be considered. The speed range R_N refers to the servo motor which can provide the

ratio of the maximum speed $R_{\max}$ and the minimum speed $R_{\min}$.

In order to simplify the analysis, the permanent magnet synchronous motor is modeled on the following assumptions:

(1) Ignored is the core saturation effect.

(2) Air gap magnetic field has the sinusoidal distribution.

(3) Eddy currents and hysteresis losses are not considered.

The voltage equations of the permanent magnet synchronous motor in the three-phase stationary coordinate $u-v-w$ can be written in the form

$$\begin{pmatrix} u_u \\ u_v \\ u_w \end{pmatrix} = \begin{pmatrix} R_{\mathrm{a}}+PL_u & PM_{uv} & PM_{wu} \\ PM_{uv} & R_{\mathrm{a}}+PL_v & PM_{vw} \\ PM_{wu} & PM_{vw} & R_{\mathrm{a}}+PL_w \end{pmatrix} \begin{pmatrix} i_u \\ i_v \\ i_w \end{pmatrix} + \begin{pmatrix} e_u \\ e_v \\ e_w \end{pmatrix}. \quad (1)$$

In formula (1), u_u, u_v, u_w are the stator voltages. Quantities i_u, i_v, i_w are the phase stator currents. Symbols e_u, e_v, e_w denote the permanent magnet magnetic field in the u, v, w phase armature winding induced electromotive force. Quantity P is the differential operator, and R_{a} is the stator winding resistance. Symbols L_u, L_v, L_w are the stator winding self-inductances and M_{uv}, M_{vw}, M_{wu} are the mutual inductances between the windings.

The d-axis of the two-phase synchronous coordinate system is the same as the u-axis of the three-phase stationary coordinate system, that is, the direction of the d-axis and the direction of the fundamental magnetic field of the permanent magnet are the same. According to the transformation matrix, it can be changed from the three-phase stationary coordinate system $u-v-w$ to the rotation coordinate system $d-q$, and the transformation matrix is

$$[C] = \sqrt{\frac{2}{3}} \begin{pmatrix} \cos\theta & \cos\left(\theta-\frac{2}{3}\pi\right) & \cos\left(\theta+\frac{2}{3}\pi\right) \\ -\sin\theta & -\sin\left(\theta-\frac{2}{3}\pi\right) & -\sin\left(\theta+\frac{2}{3}\pi\right) \end{pmatrix}. \quad (2)$$

By using the above transformation matrix, the voltage equation in the d-q rotating coordinate system can be obtained, as shown in the following equation:

$$\begin{pmatrix} u_d \\ u_q \end{pmatrix} = \begin{pmatrix} R_{\mathrm{a}}+PL_d & -\omega L_q \\ \omega L_d & R_{\mathrm{a}}+PL_q \end{pmatrix} \begin{pmatrix} i_d \\ i_q \end{pmatrix} + \begin{pmatrix} 0 \\ \omega\psi_f \end{pmatrix}. \quad (3)$$

In formula (3), u_d, u_q are the stator currents in axes d and q, i_d, i_q are the phase stator currents and L_d, L_q are the stator winding self-inductances.

The electromagnetic torque can be expressed by the sum of the product of the permanent magnet flux linkage and the armature winding current. According to the coordinate transformation process, the expression of the electromagnetic torque can be obtained in the form

$$T_{\mathrm{e}} = P_{\mathrm{n}}\left[\psi_f i_q + (L_d - L_q)\, i_d i_q\right]. \quad (4)$$

Finally, the equation of motion of the system is

$$\frac{J}{P_{\mathrm{n}}}\frac{d\omega}{\mathrm{d}t} = T_{\mathrm{e}} - T_{\mathrm{L}} - B\omega\,. \tag{5}$$

Here, T_{L} is the load torque, J is the moment of inertia of the system, and B is the viscous friction coefficient.

3. Methodology

3.1. PID control and fuzzy control

In the continuous system control theory, PID controller is the most mature and most widely used control mode [9]. The feedback control used by the general PID plays a controlling role in the ratio, integral and differential of the deviation. The PID controller usually contains three parameters, which are K_{p}, T_{i}, and T_{d}. They represent the scale factor, integral time constant and differential time constant, respectively. By adjusting the K_{p}, T_{i}, and T_{d}, three parameters can change the output of the controller, it allows the output of the system to follow the input changes, in order to achieve the system performance requirements.

In continuous systems, the relationship between the output $u(t)$ of the PID controller and the input $e(t)$ is

$$u\,(t) = K_{\mathrm{p}}\left[e(t) + \frac{1}{T_{\mathrm{i}}}\int e(t)\,\mathrm{d}t + T_{\mathrm{d}}\frac{\mathrm{d}e(t)}{\mathrm{d}t}\right] =$$

$$= K_{\mathrm{p}}\left[e(t) + K_{\mathrm{i}}\int e(t)\,\mathrm{d}t + K_{\mathrm{d}}\frac{\mathrm{d}e\,(t)}{\mathrm{d}t}\right]. \tag{6}$$

In the upper form, $e(t) = r(t) - y(t)$, supposing that the value $r(t)$ is the measured deviation, K_{p} is the proportional gain, T_{i} is the integral time constant, T_{d} is the differential time constant, K_{i} is the integral gain, and K_{d} is the differential gain.

Fuzzy control is a new type of controller, which has the advantage of allowing the mathematical model of the controlled object to be biased against the actual model. As the core of the control system, fuzzy controller can be divided into four parts: fuzzification, knowledge base, fuzzy reasoning and defuzzification.

Fuzzification refers to the precise value of the input quantity of the control system [10]. By fuzzy quantification, the fuzzy variable value is transformed into corresponding fuzzy value. This transformation is done by the corresponding membership function. That is, the input space is divided into several fuzzy domain sets. By defining the membership function on the set of fuzzy sets, the input variables are mapped to the corresponding values in the fuzzy domain, and the exact quantities are transformed into fuzzy quantities.

The knowledge base has specific knowledge about the control domain. It consists of two parts: database and rule base. The database mainly has the membership function and quantization factor of the corresponding fuzzy language and the division

of the fuzzy domain. The rule base mainly includes a series of fuzzy control rules, that is, "if"... then..." form; they represent the embodiment of expert control experience.

The fuzzy reasoning is the core part of fuzzy controller. The reasoning process is based on fuzzy logic and control rules. In general, it is described in the form of "If A and B, then C". According to the fuzzy rules, the input variables are analyzed synthetically. An output in vague language is obtained.

At present, the common method of fuzzy reasoning is Mamdani's max-min synthesis method. The details are as follows: $A_i \times B_i \times C_i$ in the fuzzy rule base is treated as a collection on the X, Y, and Z domains, respectively. The relationship between the control rules is

$$\mu_{R_i}(X,Y,Z) = \mu_{A_i}(x_i) \wedge \mu_{B_i}(y_i) \wedge \mu_{C_i}(z_i) \quad \forall x \in X, \forall y \in Y, \forall z \in Z\,. \tag{7}$$

Then, the fuzzy relation of all fuzzy rules is: $R = \bigcup_{i=1}^{n} R_i$.

The membership function of R is $\mu_{R_i}(X,Y,Z) = \underset{i=1}{\overset{n}{\vee}} \left(\mu_{R_i}(X,Y,Z)\right)$.

When the input variables E and EC become fuzzy sets A and B, respectively, the control quantity U can be obtained according to the lower form $U = (A \times B) \circ R$.

The membership function of U is

$$\mu_U(Z) = \underset{\substack{x \in X \\ y \in Y}}{\vee} \mu_R(X,Y,Z) \wedge [\mu_A(x) \wedge \mu_B(y)] \tag{8}$$

The defuzzification is the output link of fuzzy system. Its function is to convert the fuzzy output from fuzzy inference into accurate value. It contains two parts: scale mapping and defuzzification. The former converts the fuzzy output value into the exact value of the domain by defuzzification. The latter converts the exact values within the domain into actual control values.

3.2. Design of PI controller and fuzzy controller

Speed current double closed loop control is usually used in permanent magnet synchronous motor servo control system. As an inner loop, the action of the current loop is to cause the current to follow a given change. The current loop adopts PI controller. The given input of the current loop is the output of the speed ring. As a result, the transfer functions from the current setpoint i_q^* to i_q can be calculated:

$$G_1(s) = \frac{i_q(s)}{i_q^*(s)} = \frac{\frac{K_{\mathrm{cp}}}{L_{\mathrm{a}}} s + \frac{K_{\mathrm{ci}}}{L_{\mathrm{a}}}}{s^2 + \frac{K_{\mathrm{cp}}+R_{\mathrm{a}}}{L_{\mathrm{a}}} s + \frac{K_{\mathrm{ci}}}{L_{\mathrm{a}}}}\,. \tag{9}$$

The current loop is mainly based on following performance. Therefore, the current loop transfer function is designed as the first order inertial link, that is, $G_1(s) = 1/(T_0 s + 1)$. Symbol T_0 is a time constant. By calculation, the gains of PI

controller $K_{\mathrm{cp}}, K_{\mathrm{ci}}$ can be obtained:

$$K_{\mathrm{cp}} = \frac{L_{\mathrm{a}}}{T_0}, \quad K_{\mathrm{ci}} = \frac{R_{\mathrm{a}}}{T_0}\,. \tag{10}$$

A fuzzy controller with two inputs and one output is designed for the speed loop by fuzzy control. The error E and the error change rate EC are used as inputs to the fuzzy controller, and the U is the output of the fuzzy controller. The error E, the error rate EC and the output U are divided into 7 grades. The scope of the domain is -5.5. The language fuzzy subset definition is NB NM NS ZE PS PM PB. The membership function of input E and EC is triangular, and the membership function of output U is Gaussian. According to the controlled object characteristics, the fuzzy control rules are shown in Table 1.

Table 1. Fuzzy variable rules

	NB	NM	NS	ZE	PS	PM	PB
NB	NB	NB	NB	NM	NS	NS	PS
NM	NB	NB	NM	NS	NS	NS	PS
NS	NB	NM	NS	NS	ZE	ZE	PM
ZE	NM	NS	NS	ZE	PS	PS	PM
PS	NM	ZE	ZE	PS	PS	PM	PB
PM	NS	PS	PS	PS	PM	PB	PB
PB	NS	PS	PS	PM	PB	PB	PB

3.3. The design of disturbance observer

Fuzzy control belongs to nonlinear control. It can achieve better control without the accurate mathematical model of the controlled object. However, the influence of load perturbation TL on the system is not considered before design, and viscous friction is neglected. When the control system is disturbed by disturbances, the steady-state performance of the system will decrease, resulting in greater steady-state error and fluctuation of speed. In this experiment, the disturbance observer is used to compensate the disturbance rejection capability of the fuzzy controller. The disturbance observer can estimate the non measurable disturbance from the measurable information of the system, and then generate a torque current compensation. The i_q^* is modified in advance, so as to improve the anti-interference ability of the system. The block diagram of the disturbance observer is shown in Fig. 2.

In the figure, the upper part of the dashed box is the motor body model. Quantity $G_n(s) = 1/(Js + B)$ is the nominal model. $G_n^{-1}(s)$ is the inverse of the nominal model. The input of the disturbance observer is the measured value of the speed and current signals. When the load of permanent magnet synchronous motor is changed and the disturbance of the system friction torque is affected, the transfer function of the nominal model is

$$G_n(s) = 1/\left(J_1 s + B_1\right)\,. \tag{11}$$

Among them, J_1 is the system equivalent moment of inertia, and B_1 is the

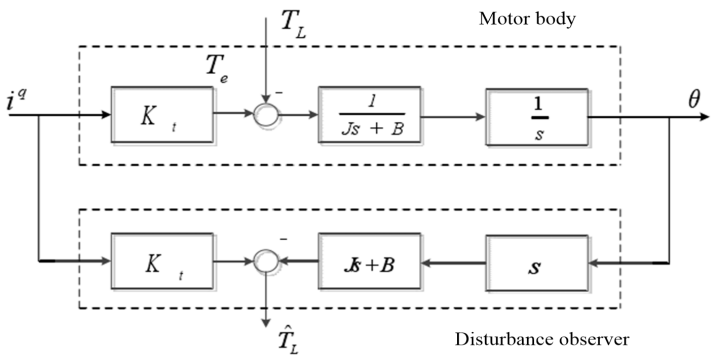


Fig. 2. The principle of disturbance observer

equivalent friction coefficient.

In the process of design, in order to reduce the speed measurement error of the motor, and the influence of the factors such as the current sensor measurement accuracy limit and the CPU calculation error, the filter link needs to be introduced. According to the function and design method of filter, the general form of filter $Q(s)$ can be obtained

$$G(s) = \frac{1 + \sum_{m=1}^{n_q - p_q} f_m s^m}{1 + \sum_{m=1}^{n_q} f_m s^m}. \tag{12}$$

Here, n_q is the order of $Q(s)$, and P_q is the relative degree of $Q(s)$. The relative order of the actual model and the nominal model can be obtained by formula (13). The filter $Q(s)$ must meet the condition of $p_q \geq 1$. For general permanent magnet synchronous motor control systems, the load disturbances can be discretized, which are equivalent to step error disturbances. Therefore, the step disturbance error can be eliminated by only one integral operation. The $Q(s)$ is designed as a first-order low-pass filter form:

$$Q(s) = \frac{\omega}{s + \omega}. \tag{13}$$

In the formula, ω is the cutoff frequency of the low-pass filter. The cutoff frequency of the low-pass filter will directly affect the disturbance suppression capability of the system. In order to minimize the impact of perturbation on the system, and taking into account the stability of the system, the value of ω needs to take into account the appropriate value.

Combining the disturbance compensator, the current loop PI controller and the speed loop fuzzy controller, the permanent magnet synchronous motor servo system block diagram can be obtained, as shown in Fig. 3.

4. Result analysis and discussion

The motor parameters used in the simulation program are presented: the stator winding resistance $R_\mathrm{a} = 2.875\,\Omega$. Stator d phase and q phase winding inductances

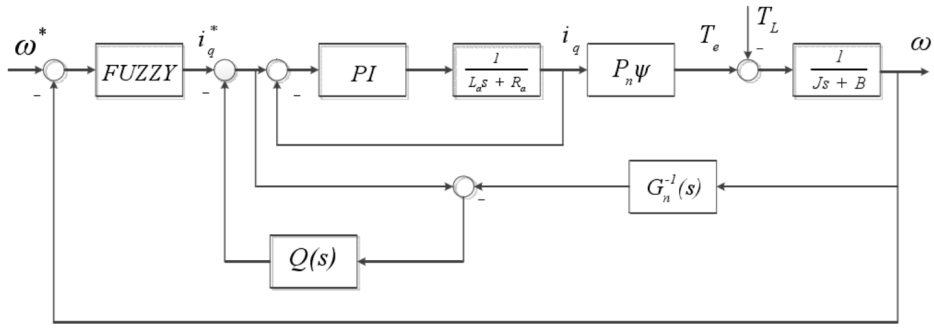


Fig. 3. Block diagram of permanent magnet synchronous motor based on fuzzy control and disturbance observer

$L_d = L_q = L_a = 0.0085\,\mathrm{H}$, $J = 0.0008\,\mathrm{kg\,m^2}$, the flux linkage amplitude $\Psi = 0.175\,\mathrm{Wb}$, the pole-pair number of the motor $P_n = 4$.

To illustrate the effect of FUZZY-DOB control, it is compared with the conventional fuzzy algorithm and the PI algorithm. In the four algorithms, the current loop of PMSM is controlled by conventional PI, and the parameters are the same. According to the principle of the first inner ring, the back outer ring, the current loop is selected, and the desired regulation time is 5 ms. Then, $t_s = 3T_0$, the parameters of the PI current controller $K_{\mathrm{cp}} = 5.1$, and $K_{\mathrm{ci}} = 1275$. The PI speed controller parameter is $K_{\mathrm{cp}} = 0.15$, $K_{\mathrm{ci}} = 0.6$. The speed controller based on the fuzzy algorithm is described as above. The disturbance observer parameter is: $J_1 = 0.001\,\mathrm{kg\,m^2}$, $B_1 = 0.003\,\mathrm{N\,m\,s/rad}$, and the filter cutoff frequency $\omega = 200\,\mathrm{rad/s}$. The simulation results are as follows. Figures 4, 5 and 6 show the comparison of the simulation curves of the four control algorithms. Tables 2 and 3 contain comparisons of specific data.

As shown in Fig. 4, when the given speed is 1000 rpm, the control system based on fuzzy control and disturbance observer is stable within 0.1 second. The steady-state error is 3 rpm. There is no overshoot. The overshoot of the PI controller is 8 %, and the stability time is more than 0.3 seconds.

The system load response is added to the system at 1 second, and the speed response characteristic of the system is shown in Fig. 5. The local zoom diagram is shown in Fig. 6. The speed fluctuation of the control system is 20 rpm based on fuzzy control and disturbance observer. Within 0.05 seconds, it can recover to its initial speed. The disturbance of the control system is 50 rpm based on the PI controller and disturbance observer. The speed fluctuation of the control system without disturbance observer is 140 rpm, and the regulation time is longer.

5. Conclusion

Firstly, the electric servo drive technology is briefly introduced, and the PID control and fuzzy control principle are summarized. On the basis of analyzing the mathematical model of permanent magnet synchronous motor (PMSM), the current

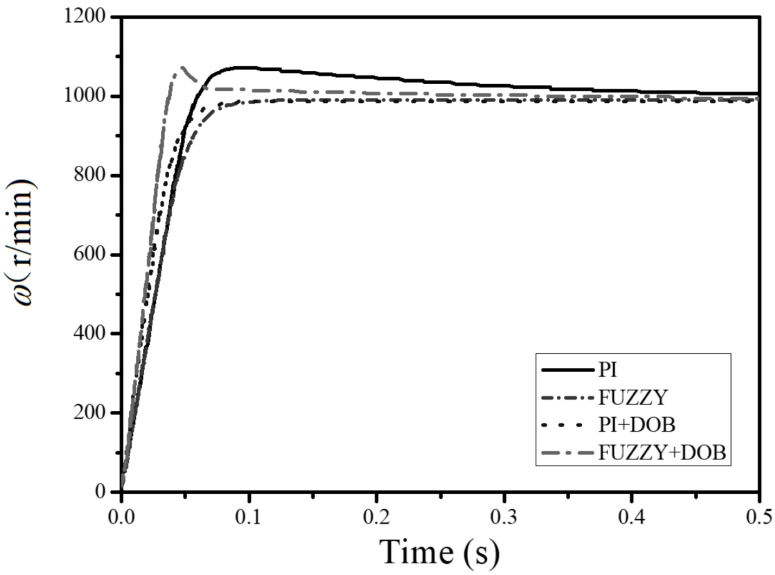


Fig. 4. Comparison of speed characteristic curves of four controllers

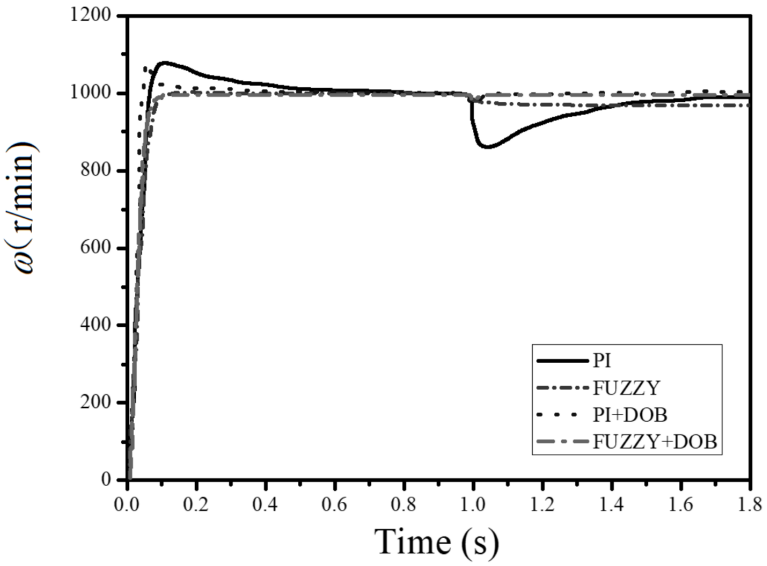


Fig. 5. Comparison of speed characteristic curves under sudden load

loop based on PI control and the speed loop based on fuzzy control are designed. Aiming at the problem of perturbation of servo control system, a double closed loop servo motor controller based on fuzzy control and disturbance observer is proposed. It effectively suppresses the impact of sudden disturbance on the system.

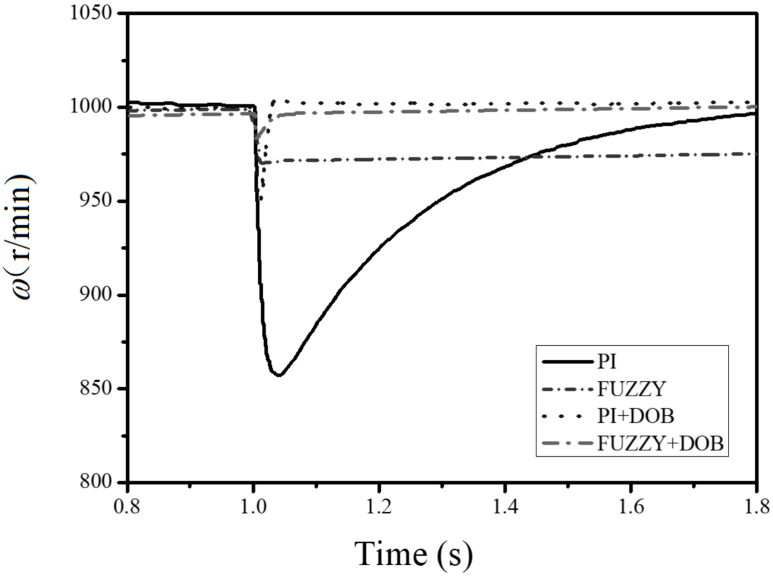


Fig. 6. Partial enlargement of speed characteristic curve during sudden loading

The simulation results show that the proposed control algorithm has good dynamic performance and steady-state performance. It not only provides a theoretical basis for the research of high performance control strategy of PMSM control system, but also provides a theoretical basis for high-precision positioning of robot manipulators.

Table 2. Comparison of four control algorithms

Control method	Rise time (s)	Regulation time (s)	Overshoot	Steady-state error
PI	0.036	0.221	8%	0
FUZZY	0.041	0.056	0	0
PI+DOB	0.018	0.047	8.5%	0
FUZZY+DOS	0.031	0.046	0	3

Table 3. Comparison of four control algorithms in disturbance change

Control method	Maximum velocity fluctuation (r/min)	Regulation time (s)
PI	142	0.94
FUZZY	27	Infinity
PI+DOB	52	0.30
FUZZY+DOS	13	.04

References

[1] T. Winther, H. Wilhite: *An analysis of the household energy rebound effect from a practice perspective: Spatial and temporal dimensions.* Energy Efficiency *8* (2015), No. 3, 595–607.

[2] K. G. Chan, K. Guru, P. Wiklund, J. Catto, B. Yuh, G. Novara, D. G. Murphy, T. Al-Tartir, J. W. Collins, A. Zhumkhawala, T. G. Wilson, Pasadena Consensus Panel: *Robot-assisted radical cystectomy and urinary diversion: Technical recommendations from the Pasadena Consensus Panel.* European Urology *67* (2015), No. 3, 423–431.

[3] P. Backé, M. Feldkircher, T. Slačík: *Economic spillovers from the euro area to the CESEE region via the financial channel: A GVAR approach.* Focus on European Economic Integration (2013), No. 4, 50–64.

[4] K. Salonitis, L. D'Alvise, B. Schoinochoritis, D. Chantzis: *Additive manufacturing and post-processing simulation: Laser cladding followed by high speed machining.* International Journal of Advanced Manufacturing Technology *85* (2016), Nos. 9–12, 2401–2411.

[5] K. Söderholm, P. Söderholm, H. Helenius, M. Pettersson, R. Viklund, V. Masloboev, T. Mingaleva, V. Petrov: *Environmental regulation and competitiveness in the mining industry: Permitting processes with special focus on Finland, Sweden and Russia.* Resources Policy *43* (2015), 130–142.

[6] R. Moreau, M. T. Pham, M. Tavakoli, M. Q. Le, T. Redarce: *Sliding-mode bilateral teleoperation control design for master–slave pneumatic servo systems.* Control Engineering Practice *20* (2012), No. 6, 584–597.

[7] X. Wu, H. Wang, S. Huang, K. Huang, L. Wang: *Sensorless speed control with initial rotor position estimation for surface mounted permanent magnet synchronous motor drive in electric vehicles.* Energies *8* (2015), No. 10, 11030–11046.

[8] S. Hodgson, M. Tavakoli, M. T. Pham, A. Leleve: *Nonlinear discontinuous dynamics averaging and PWM-based sliding control of solenoid-valve pneumatic actuators.* IEEE/ASME Transactions on Mechatronics *20* (2015), No. 2, 876–888.

[9] D. Tang, L. Chen, E. Hu, Z. F. Tian: *A novel actuator controller: Delivering a practical solution to realization of active-truss-based morphing wings.* IEEE Transactions on Industrial Electronics *63* (2016), No. 10, 6226–6237.

[10] R. Errouissi, A. Al-Durra, S. M. Muyeen, S. Leng: *Continuous-time model predictive control of a permanent magnet synchronous motor drive with disturbance decoupling.* IET Electric Power Applicationss *11*, (2017), No. 5, 697–706.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 165–176

3D reconstruction of tree and limb based on aerial image of UAV[1]

Tiebo Sun[2,3], Jianming Kan[2], Jinhao Liu[2], Qingqing[2], Kai Ma[2], Tingting Sun[2]

Abstract. To obtain more accurate and more detailed information, UAV (unmanned aerial vehicle) aerial in the application is inevitable going from two dimension to three dimension. While trees, due to their structure complexity, make the 3D reconstruction have more research significance. UAV aerial images, because of being affected by aerial height, the information contained in the tree crown texture and profile is poor. In terms of this problem, a feature extraction method based on watershed segmentation is proposed, so as to fully extract the feature points that can reflect the crown structure, and the region correlation coefficients are calculated to match the feature regions accurately. The results showed that in the tree modeling method, based on L-system, the structure of the growth of the trees is simulated. At last, it is concluded that in the case of less number of feature points, it well restores the limb structure trees, and constructs a complete model of trees.

Key words. 3D reconstruction, aerial image of UAV, watershed.

1. Introduction

In the UAV image processing, the target 3D reconstruction and localization have become one of the research directions in recent years by means of sequence image analysis [1]. The three-dimensional reconstruction problem based on UAV aerial images can theoretically be attributed to the reduction of the image scene and the 3D structure of the target from the sequence image. Applying the theory and technology of 3D reconstruction of sequence images, the 3D model of scene and target can be obtained by processing of unmanned aerial vehicle sequence images [2]. The three-dimensional structure of the target can be reconstructed and measured accurately based on the intrinsic constraints of the same scene on multiple images. But for

[1]The authors acknowledge the National Natural Science Foundation of China (Grant: 51578109), the National Natural Science Foundation of China (Grant: 51121005).

[2]School of Technology, Beijing Forestry University, 100083, Beijing, China

[3]Department of Mechanical and Electrical Engineering, Jiangsu Food & Pharmaceutical Science College, 223003, Jiangsu, China

the sequence image obtained from the UAV aerial photography, under the effect of height, the color, texture and contour information that the UAV aerial images contained are fuzzy.

In addition, the process of aerial shaking, offset, rotation and so on will cause the camera swing and rotation so that the camera imaging model has much complexity compared with the general camera imaging models. To determine the UAV aerial camera model is the main problem of UAV aerial image 3D reconstruction [3–4]. In the UAV navigation and agricultural applications, it often requires the measurement of a high degree of information on plants and trees, which created a necessity for 3D reconstruction of crown based on aerial images. As a result, 3D reconstruction of trees has become the focus of many scholars for the study [5–6]. However, trees, plants and other objects, influenced by their growth conditions (including photosynthesis, nutrient cycling, energy transmission and forest leaf light environment etc.) and some human intervention (such as pruning branches etc.), the uncertainty of morphology and structure is great. In consequence, it is difficult to establish the experience knowledge of tree morphology [7]. In addition, the surface texture of trees is very similar or almost no texture, so it is difficult to find the corresponding relationship between the surface points of trees in the images, so the difficulty of three-dimensional reconstruction of crown trees is increased. These factors make the extraction of crown feature points and matching and modeling of complex shape and structure of trees have become a key issue for the crown 3D reconstruction. It has very important theoretical guidance and practical significance in accurately using the aerial sequence images for the reconstruction of crown model.

2. Method

Watershed segmentation method is a segmentation method based on mathematical morphology and topology theory. Its basic idea, as the name suggests, is to regard the image as the topological landscape on geodesics. The gray value of each pixel in the image represents the altitude of the point, and every local minimum and the affected region are called the catchment basin, and the boundary of catchment basin forms the watershed. Watershed segmentation usually takes the gradient image as the input so that it can quickly get the closed area, but there is often over segmentation phenomenon [8]. Watershed transformation can locate the edge accurately, and it has the advantages of simple operation and easy parallel processing. In this paper, we hope that the segmentation method can sufficiently distinguish the light and shade areas of the crown image, and the appropriate over segmentation of the watershed segmentation algorithm can meet our requirements to some extent.

The concept and formation of watersheds can be explained by simulating the immersion processes. On the surface of every local minimum, a hole is pierced, and then the whole model is slowly immersed in water. With the immersion deepening, the influence domain of each local minimum slowly expands outward, to build a dam in the two water collecting basin confluence, which forms a watershed. The watershed computation process is an iterative annotation procedure. L.Vincent [9] proposed a more classic watershed segmentation algorithm, and the calculation is

divided into two steps: one is the sorting process, and the other is the flooding process. Firstly, the gray level of each pixel is ordered from low to high, then in the achieving submerged process from low to high, for each local minimum, in the effect domain in h order height, FIFO structure is used for the judgment and annotation.

What the watershed transformation obtained is the catchment basin of the input images, and the boundary between the catchment basin is the watershed. Obviously, the watershed represents the maximum point of the input image. Therefore, in order to obtain the edge information of the image, the gradient image is usually used as the input image, that is

$$g(x,y) = \text{grad}\ (f(x,y)) = \sqrt{[f\,(x,y) - f\,(x-1,y)]^2 + [f\,(x,y) - f\,(x,y-1)]^2}\,. \tag{1}$$

In the above formula, $f\,(x,y)$ represents the original image. Therefore, this paper uses gradient watershed segmentation method, and the algorithm process is as follows: first of all, the Sobel operator is used for longitudinal gradient operation; then, the gradient image is opened and closed for smoothing processing; at last, watershed segmentation is conducted for the smoothed images.

The watershed algorithm has a good response to the weak edge, and the phenomenon of over segmentation is caused by the noise in the image and the slight change of the gray level of the object surface. There are many minimum points in the image in general, and there is usually an over segmentation phenomenon. The gradient threshold segmentation improvement or labeling of watershed algorithm can be applied to connect multi-minimum value areas. But it should be seen that, the watershed algorithm has a good response to the weak edge, which is guaranteed by the closed continuous edge. In addition, the closed catchment basin obtained by the watershed algorithm provides the possibility of analyzing the regional features of the image.

Through the above watershed segmentation, the bright area and the dark area of the crown image can be distinguished well. For the area information recording the segmentation, seen from the image, the pixels the closer to the center, the more it represents the pixel characteristics of that area. By extracting the centroid $(\overline{u}_i, \overline{v}_i)$ of each region as feature points, this paper thinks that for the bright region, the centroid point is approximate to the brightest point; for the dark area, its centroid is similar to the darkest point, and therefore the calculation formulas of $\overline{u}_i$ and $\overline{v}_i$ are as follows.

$$\overline{u}_i = \frac{1}{n_i} \sum_{k=0}^{n_i} u_k\,, \tag{2}$$

$$\overline{v}_i = \frac{1}{n_i} \sum_{k=0}^{n_i} v_k\,. \tag{3}$$

The matching method based on region correlation coefficients only takes into account the local features of the image when matching, so the pixel pairs with high correlation coefficients are not necessarily the correct matching points. That is to say, if the two areas are the matching region pairs, then their correlation coefficient must

be relatively large. On the contrary, RGB regional correlation coefficient calculated by the two regions does not necessarily mean that these two regions are the regions matched to each other. Therefore, the matching algorithm of this paper cannot avoid the existence of some mismatching points.

To further improve the accuracy of the matching algorithm, this paper removes the wrong matching point by calculating the ratio of the most relevant one and sub related one [10]. Among them, the ratio calculation formula of the most relevant one and sub related one is:

$$K = D_{\text{nearest}}/D_{\text{hypo}-\text{nearest}}\,, \tag{4}$$

where D_{nearest} refers to the Euclidean distance between the most relevant matching point and the point to be matched, and $D_{\text{hypo}-\text{nearest}}$ suggests the Euclidean distance between the sub relevant matching point and the point to be matched.

According to the literature [11], the closest distance between the correct matching pairs should be significantly smaller than that of the wrong matching pairs. For the false matching points, due to the impact of high dimension feature space, there may be many other false matching points with similar distance. Thus, this correlation distance is taken as the error matching criteria to determine the individual fuzzy examples. Therefore, for a matching area, if the maximum correlation coefficient and the sub correlation coefficient are more closely, the ratio of them will be greater. It indicates that the matching area pairs are more likely to be the wrong matching examples. On the contrary, if the ratio of the most relevant and sub relevant is smaller, then it means that the matching is more likely to be the correct matching example.

So, a suitable threshold is selected for the proportion. If the proportion of K is greater than a threshold, it indicates that the most relevant matching and the sub relevant matching are very similar. In this paper, we can think the closest Euclidean distance matching might be the wrong matching, and eliminate the matching point.

3. 3D modeling

3.1. L-system

L-system is proposed by the American biologist A. Lindenmayer [12]. At the very beginning of this system is the method of simulating plant biology and cell morphology and growth, and later it is developed into the fractal method effectively simulating scenes of nature (especially vegetation type) in the computer graphics. L-system is a language prompt system consisting of several words, symbols and control parameters. It uses the right hand rectangular coordinate system to define the three-dimensional L-system. The direction is defined by the 3 vectors, H, L and U [13]. The 3D coordinate system of the L-system is shown in Fig. 1.

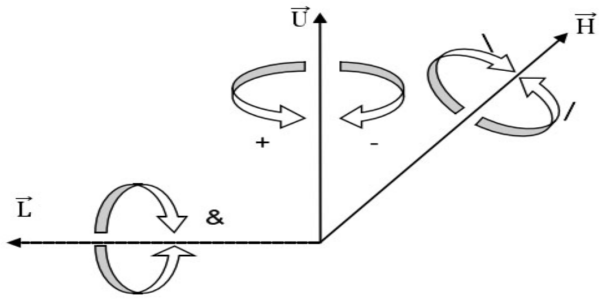


Fig. 1. 3D coordinate system of L-system

3.2. Canopy structure acquisition

In this paper, the skeleton information of the tree canopy is obtained by a certain method. Through the command of the L-system, a three-dimensional model of the crown is created by using Open GL. The modeling process is shown in Fig. 2.

Set a collection Trunk List of tree trunks, and the 3D scatter set contained in these tree blocks is the Include Points collection. The method for obtaining skeleton information is as follows:

(1) To find the highest point of the Z coordinate, and find dz of the highest point and the lowest point. The point dz is taken as the tree height and the vertical axis where the highest point locates is taken as the central trunk of trees. The depth is 1, the trunk set Trunk List is added, and the highest point is added in the tree trunk points set Include Points.

(2) To scan the feature points not in the Include Points collection, and to find the minimum point pi at which the current point is concentrated away from the tree trunk, and the corresponding trunk ti (the depth of the restricted ti is no more than 6, and the branch is no more than 3).

(3)To lead out a new tree trunk $tnew$ with angle of 30 degrees with ti from the tree trunk ti to the point pi. The depth of $tnew$ is the depth of ti (the maximum depth of restricted tree is no more than 4), which is inserted behind ti when adding Trunk List, and the point pi is added to the existing point set Include Points.

4. Results and discussion

Feature points extraction and matching:

In order to obtain the light and shade area in the crown, we make the watershed area segmentation of the crown region, to distinguish the light and shade areas of the crown. The contrast between the watershed segmentation and the original image is shown in Fig. 3. It can be seen from the graph that, the watershed segmentation well achieves the desired effect in this article.

After obtaining the light and shade area of the crown area, the volume of each area is recorded in pixels number, and the statistical region information is shown in

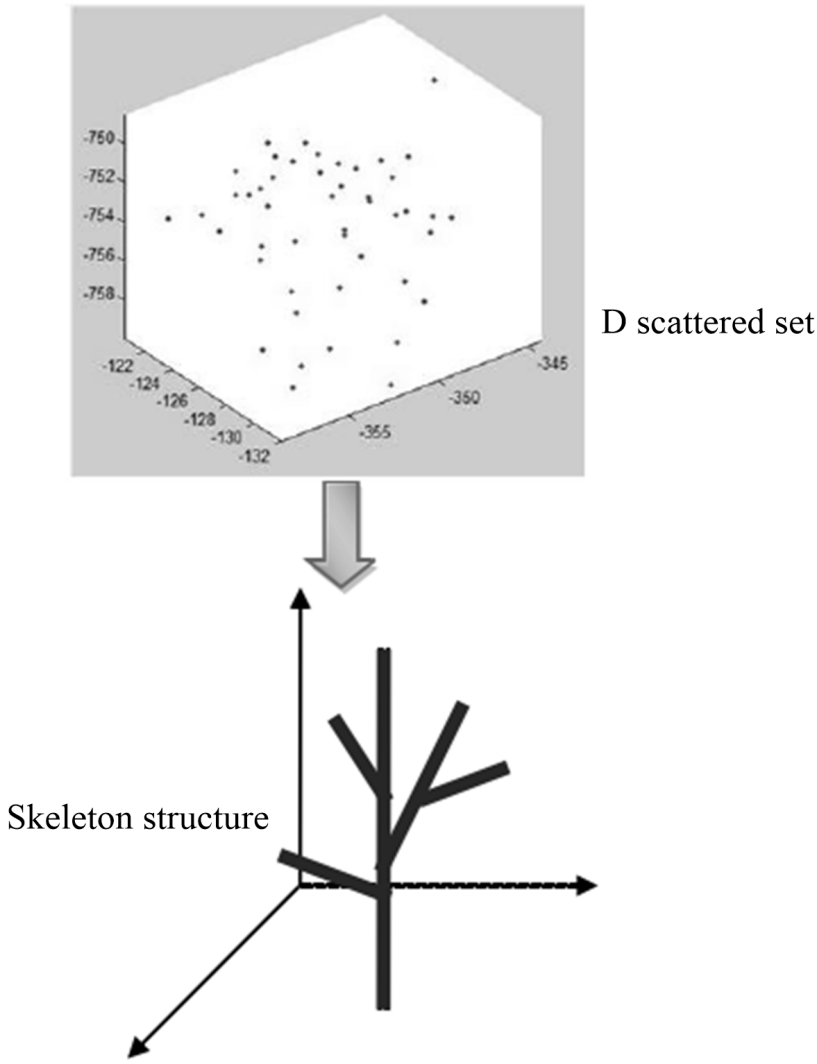


Fig. 2. Modeling process

Table 1. Analyzing the data, it can be found that, in the segmented regions, the size of most pixels is between 40 to 150. Although there is a larger area and a smaller area, it is after all in the minority. The number of areas in the intervals of [1,39] and [150181] has a total of 8. On the whole, the size of the region segmented by watershed is still relatively uniform.

By the statistics of segmentation region information, the size of each region is relatively uniform, so the centroid $(\overline{u}_i, \overline{v}_i)$ of each region is calculated as the feature point of the feature area, and then added to the feature points set of the image. The image feature points extracted are shown in Fig. 4, left part. It can be seen from the

figure that, watershed segmentation can fully extract the brighter and dark points in the tree crown, and these points can well reflect the crown branches information. In the watershed segmentation results, the feature points set position is shown in Fig. 4, right part.

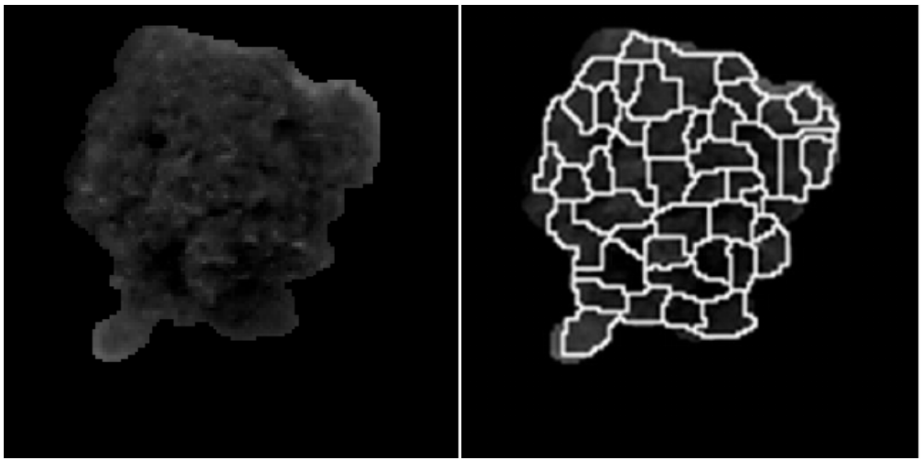


Fig. 3. Watershed segmentation results for tree crowns

Table 1. Results statistics of watershed segmentation

Items	Values	Ratio
The maximum value of areas, unit: pixel	181	-
The minimum value of areas, unit: pixel	1	-
The total number of areas	47	1.000
[1,39]	5	0.106
[40,99]	26	0.553
[100,149]	13	0.277
[150,181]	3	0.064

3D computation and 3D modeling:

In order to obtain the most accurate and complete canopy structure, we take different thresholds in the experiments and compare the corresponding results. Several representative threshold values are selected, the feature points set is extracted, and the skeleton structure and 3D model of the tree are constructed. Figure 5 shows the result when the threshold is 0.93, obtained 28 points. Due to the interference error points, the trunk effect is not beautiful. Figure 6 shows the experimental results when the threshold is 0.91, obtained 23 points, well removing the interference characteristics points, and the effect has been greatly improved. Figure 7 shows the experimental results when the threshold is 0.88. The number of feature points obtained by is only 17. When the false matching is removed, at the same time, the effective information is reduced, and the model constructed is not complete.

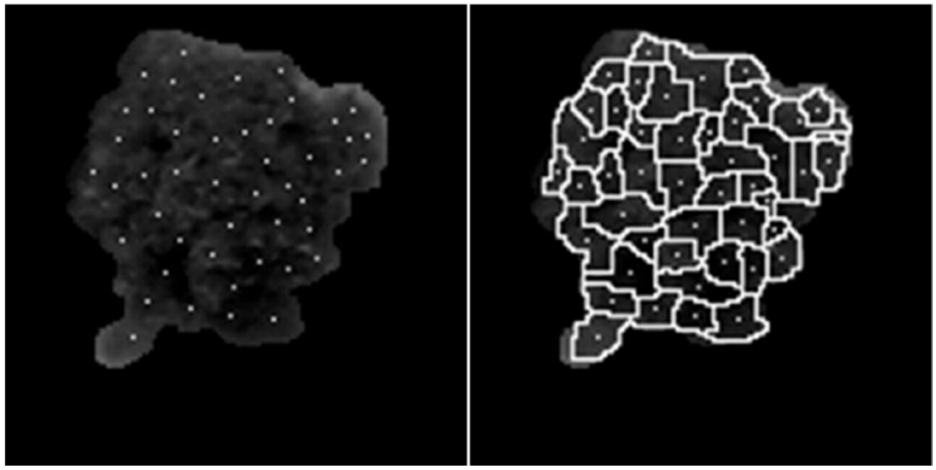


Fig. 4. Watershed segmentation of images and feature points extracted 3D computation and 3D modeling

It can be seen from the above results that, the K threshold is too large, and the removed false matching points are too small, then the left false matching points have interference on the results, thus affecting the 3D modeling effect. If the threshold of K is too small, then in full removal of false matching points, at the same time, it will also remove some correct matching points, and caused the loss of information.

In order to better and more intuitively compare, the proposed feature extraction and matching method are used at the same time for feature extraction and matching [14]. The mean geometric registration errors calculated are compared, and the results are shown in Table 2. As can be seen from the table, this method can match the feature points more accurately, and it takes the appropriate threshold K, and achieves better reconstruction results.

Table 2. Mean geometric registration error

Methods	Mean geometric registration error	Number of feature points
SIFT-KNN	1047.997	5
The proposed method ($K \leq 1$)	1041.99	47
The proposed method ($K \leq 0.8$)	1040.63	14
The proposed method ($K \leq 0.6$)	1039.91	9
The proposed method ($K \leq 0.4$)	1038.23	4

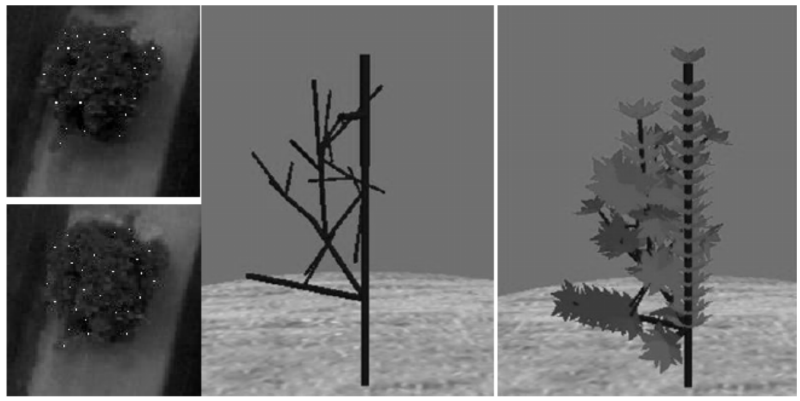


Fig. 5. Experimental results for threshold 0.93

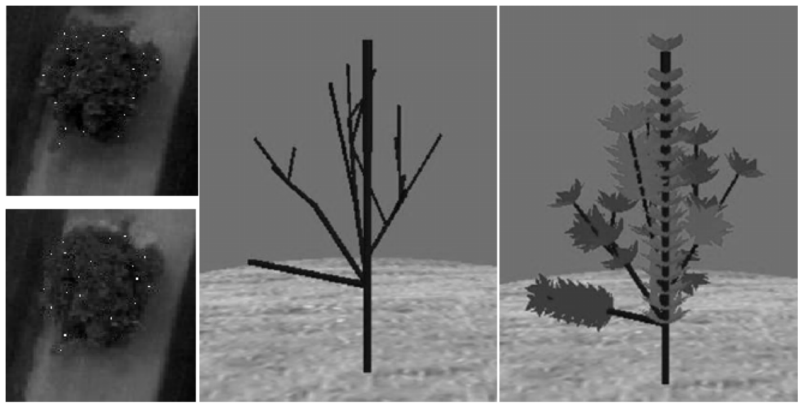


Fig. 6. Experimental results for threshold 0.91

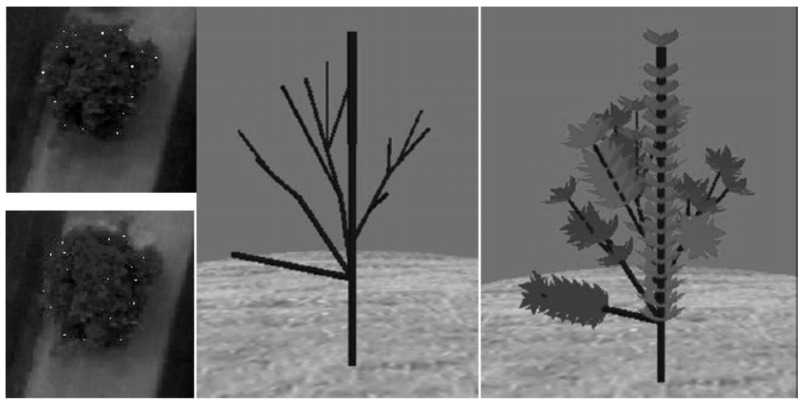


Fig. 7. Experimental results for threshold 0.88

5. Conclusion

In the feature extraction and matching, due to the complex structure of crown and features of UAV aerial image itself, a feature point extraction based on watershed region segmentation completely extracts the feature points that can reflect the crown structure. And the feature matching method based on the RGB region correlation coefficient, in the case of limited image information, can also accurately match the feature points, and then carry out 3D reconstruction. In the tree crown modeling method, based on L-system method, the structure of the growth of the trees is simulated. In the case of less number of feature points, it well restores the limb structure of trees, and constructs a complete model of trees.

References

[1] G. Caroti, I. Martínez-Espejo Zaragoza, A. Piemonte: *Accuracy assessment in Structure from Motion 3D reconstruction from uav-born images: The influence of the data processing methods.* ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences *XL-1/W4* (2015), 103–109.

[2] A. Eltner, D. Schneider: *Analysis of different methods for 3D reconstruction of natural surfaces from parallel-axes UAV images.* Photogrammetric Record *30* (2015), No. 151, 279–299.

[3] M. A. Yucel, R. Y. Turan: *Areal change detection and 3D modeling of mine lakes using high-resolution unmanned aerial vehicle images.* Arabian Journal for Science and Engineering *41* (2016), No. 12, 4867–4878.

[4] S. Harwin, A. Lucieer, J. Osborn: *The impact of the calibration method on the accuracy of point clouds derived using unmanned aerial vehicle multi-view stereopsis.* Remote Sensing *7* (2015), No. 9, 11933–11953.

[5] J. R. Alonso, A. Fernández, J. A. Ferrari: *Reconstruction of perspective shifts and refocusing of a three-dimensional scene from a multi-focus image stack.* Applied Optics *55* (2016), No. 9, 2380–2386.

[6] H. Claes, J. Soete, K. van Noten, H. El-Desouky, M. M. Erthal, F. Vanhaecke, M. Özkul, R. Swennen: *Sedimentology, three-dimensional geobody reconstruction and carbon dioxide origin of pleistocene travertine deposits in the Ballık area (south-west Turkey).* Sedimentology *62* (2015), No. 5, 1408–1445.

[7] J. A. Sereno, H. Lee: *The contribution of segmental and tonal information in Mandarin spoken word processing.* Language and Speech *58* (2015), No. 2, 131–151.

[8] G. Franchi, J. Angulo: *Bagging stochastic watershed on natural color image segmentation.* International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing (ISMM), 27–29 May, Reykjavik, Iceland, Springer Nature LNCS *9082* (2015), 422–433.

[9] C. Duval, M. de Tayrac, K. Michaud, F. Cabillic, C. Paquet, P. V. Gould, S. Saikali: *Automated analysis of 1p/19q status by FISH in oligodendroglial tumors: Rationale and proposal of an algorithm.* PloS one *10* (2015), No. 7, e0132125.

[10] C. Mouelhi, J. Saint-Pierre: *The most relevant value creation indicator under competitive dynamics of the firm.* SSRN (2013), Electronic copy available at: http://ssrn.com/abstract=2266050.

[11] V. Belmonti, A. Berthoz, G. Cioni, S. Fiori, A. Guzzetta: *Navigation strategies as revealed by error patterns on the Magic Carpet test in children with cerebral palsy.* Frontiers in Psychology (2015), No. 6, paper 880.

[12] E. Kwak, S. Ahn, J. Jaworski: *Microfabrication of custom collagen structures ca-*

pable of guiding cell morphology and alignment. Biomacromolecules *16* (2015), No. 6, 1761–1770.

[13] J. L. Junkins, H. Bang: *Maneuver and vibration control of hybrid coordinate systems using Lyapunov stability theory.* Journal of Guidance, Control, and Dynamics *16* (1993), No. 4, 668–676.

[14] N. D. Serej, A. Ahmadian, S. Kasaei, S. M. Sadrehosseini, P. Farnia: *A robust keypoint extraction and matching algorithm based on wavelet transform and information theory for point-based registration in endoscopic sinus cavity data.* Signal, Image and Video Processing *10* (2016), No. 5, 983–991.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 177–186

Network security audit system based on improved neural network

Ding Ding[1]

Abstract. With the rapid development of network technology, the network security has become the focus of the society. Therefore network security audit becomes an important way to protect the computer security. We present an network security audit model based on quantum genetic algorithm and BP neural network, which takes advantage of the global search property of the quantum genetic algorithm and the exact local search characteristics of the BP network. The weight and the thresholds of the BP neural network is optimized by the quantum genetic algorithm. As basic quantum genetic algorithm has week local search ability and is easy to premature, we also propose an improved quantum genetic algorithm. The experiments show that this method can be used to improve the efficiency and accuracy of network security audit system.

Key words. Network security audit, neural network, quantum genetic algorithm.

1. Introduction

With the rapid development of computer technology, microelectronics, communication technology and other kinds of sciences and technologies, especially the development of the Internet with its large amounts of information resources and rapid convenient efficient way to deliver information, the network has been an important tool in people's study and day-to-day life, but computer viruses, hackers, and any other uncertain dangerous problems have been threatening the security of the information on the network, testing people's wisdom to deal with the network danger and protecting the network security at the same time [1]. The network security audit plays an important role in the process of network security management, and it is also a function that the network environment security must support.

At present, the widely used characteristics detection method in security audit system is to define a series of characteristic patterns to identify intrusion by security experts in advance [2]. The problem of this approach is that if the model database is not timely updated, system cannot adaptively detect new attacks in the process of security audit, thus false alarm and alarm failure problem occur frequently. With the

[1]AnHui Audit College, Hefei, 230601, China

popularity of network applications, network data traffic has increased dramatically, some audit records itself contains a large number of irrelevant information. So, the problem of data overload and too low testing speed also appear.

In the process of the genetic algorithm applied to the security audit, it also has some drawbacks. The system can't detect multiple simultaneous attacks, and is unable to realize accurate positioning in the audit records, which makes the results of the detector do not include time information [3–5]. In network security audit based on agent technology, the monitor is the key component of the system. If a monitor stops working, all the repeater controlled by this monitor cannot submit results, and when more than one monitor reporting on the same issue, it may produce inconsistent and repetitive information. Network security audit based on the kernel technology collects data from the operating system kernel to be as basis to detect intrusion or abnormal behavior. This method is mainly used in open source Linux system, its advantage is that it has good detection efficiency and the reliability of data sources, but this method itself has a strong dependence on the safety of the operating system. Network security audit based on rule base is similar to some idea of firewall and anti-virus software, the testing accuracy of which is quite high, and can use the simplest matching method to filter out a lot of invalid audit data information, and is especially effective for to attack using a specific network tools [6]. But its shortcoming is that these rules only corresponds to known attack types or certain attack software. When there is a new attack software or software upgrade, omission of attack alarm is prone to occur, so safety audit method based on rule library has its own limitation. The largest problem based on the mathematical statistics method is how to set statistics threshold, and also is the cut-off point of normal and abnormal value, which often depends on the administrator's experience, and inevitably produce false alarm [7]. Neural network uses adaptive learning technology to extract the characteristics of the abnormal behavior, and obtain normal behavior pattern through training. The neural network dos not have stable network structure and its judgment of abnormal event will not provide any explanation or instruction information, this led to the users cannot confirm the invasion responsibility [8–10]. Introducing BP neural network into security audit system has certain practical significance, which open up new ways for the research of auditing system. Because of its many characteristics, such as adaptability, self-learning ability, in the security audit system based on neural network, we only need to provide the system audit data, then we can extract the characteristic pattern of system activity, without accessing to large amounts of data, thus it simplifies the design of the system. But BP algorithm has also obvious deficiencies. If the network structure and initial weights are not good, it not only makes its convergence speed slow, and may lead to network converge to local optimum. So the neural network is improved by quantum genetic algorithm [11, 12].

In the next section, a kind of network security audit system based on improved neural network is put forward. In section 3, in order to test the performance of network security audit system based on improved neural network, experiments are done. In the end, some conclusions are given.

2. Network security audit system based on improved neural network

Neural network has self-learning and adaptive ability. As long as the audit data or network packets of the system is provided, neural network can extract normal user or activity characteristics mode by self-learning, and detect the abnormal attack mode. These features make it get very good application in security audit detection. One of the most popular neural network learning algorithm is BP algorithm, but it has slow convergence speed and local optimal problem. The major task of the security audit based on neural network is efficiency and accuracy problem. Global search based on quantum genetic algorithm [13–14] and local accurate searching feature of BP thus can be an organically combined. We use quantum genetic algorithm optimization ability to optimize the BP network in audit research, and improve the effectiveness of the detection algorithm for unknown attack detection.

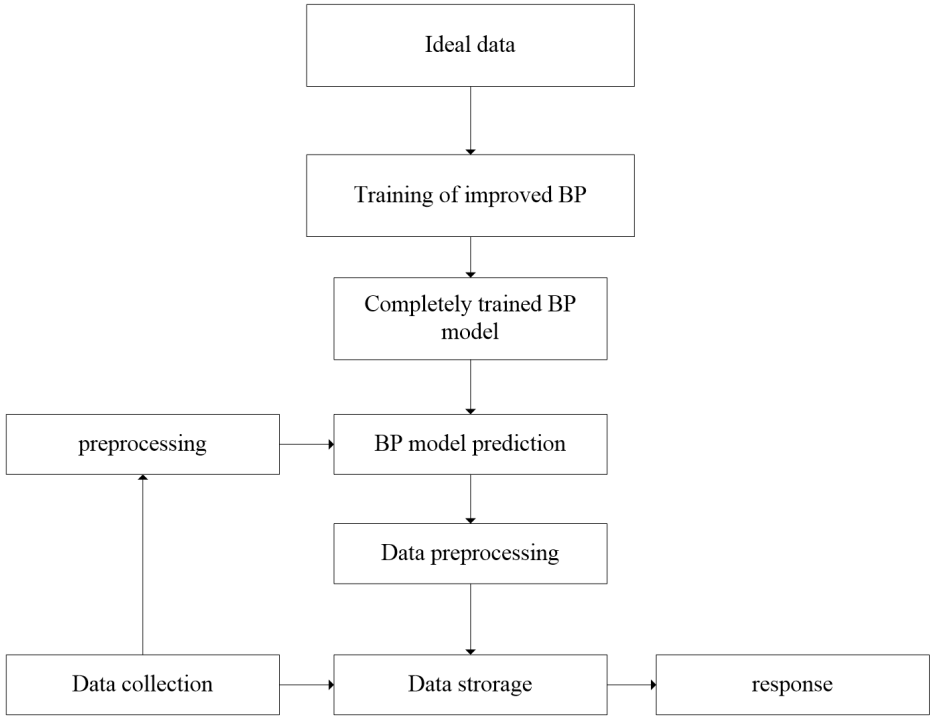


Fig. 1. Network security audit system based on quantum genetic BP

Quantum genetic algorithm simulates the natural evolution process, starting from the random generation of a group of individuals, the evolution strategy of survival of the fittest is adopted to converge to the optimal solution finally. For complex problems, the quantum genetic algorithm has strong search ability and optimization performance. The quantum genetic algorithm is used to optimize neural network weights and threshold value. At the same time, we improve the traditional quan-

tum genetic algorithm, and each link of quantum genetic algorithm is analyzed in detail, including quantum bit coding, fitness function design, quantum door update, mutation and so on. Coding uses quantum bit coding way, the parameters of neural network constitute a kind of chromosome after a certain combination and allocation, and it is converted into binary form further.

$$p_i^j =$$

$$= \left(\underbrace{\begin{array}{cccccccccccc} \alpha_{11}^{'} & \alpha_{12}^{'} & \dots & \alpha_{1l_1}^{'} & \alpha_{21}^{'} & \alpha_{22}^{'} & \dots & \alpha_{2l_2}^{'} & \dots & \alpha_{m1}^{'} & \alpha_{m2}^{'} & \dots & \alpha_{ml_m}^{'} \\ \beta_{11}^{'} & \beta_{12}^{'} & \dots & \beta_{1l_1}^{'} & \beta_{21}^{'} & \beta_{22}^{'} & \dots & \beta_{2l_2}^{'} & \dots & \beta_{m1}^{'} & \beta_{m2}^{'} & \dots & \beta_{ml_m}^{'} \end{array}}_{K_1K_2\dots K_{l_1}, K_1K_2\dots K_{l_2}, K_1K_2\dots K_{l_m}} \right).$$

Symbol m represents gene number of chromosome which corresponds to the number of neural network parameters, k represents the number of quantum bit of each gene. For one parameter, if the parameter is n, $k = \lfloor \log_2 n \rfloor$. Neural network parameters combination scheme is shown in the above equation. In designed code and its mapping sense, we choose a set of all possible neural network parameters. So a certain solution accurately represents a kind of neural network parameter combination scheme.

After selection, crossover and mutation operation, we can use parameter combination for neural network training, then the number of attacks detected correctly and the number of connections that are judged to be attacks mistakenly are worked out. According to the result of training and testing results, we evaluate parameter combination. The first value of chromosome in the first generation of quantum genetic algorithm is taken as parameters, which is assigned to BP neural network. Then BP neural network is used in audit system. Calculate error between experiment result and target result. If the error is greater than error range, quantum genetic algorithm is again used to assign value to the BP neural network parameters. Otherwise, the assignment process stops.

The rotation angle and direction of quantum gate are adjusted by means of evolution equation, and we no longer use look-up table. In this way there are two main advantages. One is to reduce the number of parameters, which simplifies the structure of the quantum genetic algorithm. Another is the evolution equation has characteristic memory, which not only can make use of the individual's own local optimal information, but also can use the optimal information of neighborhood population. It also uses the optimal state information of the whole population, thus it can adjust rotation angle theta more reasonably. It has better ability than traditional quantum genetic algorithm to jump out of local optimum. The evolution equation is

$$\theta = k_1(p_m - x_i) + k_2(p_i - x_i) + k_3(p_j - x_i) + k_4(p - x_i),$$

where k_1, k_2, k_3, k_4 are influence factors, p_i, p_j are population extreme value of the left and right neighborhood, p_m is the extreme value of individual population and p

represents global extreme value. We have designed a kind of quantum mutation in order to improve the performance of the traditional quantum genetic algorithm. In traditional genetic algorithm, the effect of mutation lies in providing algorithm local search ability and prevent premature convergence. Due to the quantum mutation needs to satisfy the mutation requirement of both genetic algorithm and quantum parallelism, we define a simple single quantum bit mutation operation, and the method can be further extended to the situation of more quantum bits. Network security audit system based on quantum genetic BP is shown in Fig. 1 and the specific methods are as follows.

The first step is to choose several individuals from the population randomly according to probability p_i. The second step is to determine one or more mutation bits for the selected individual according to fixed probability. The third step is to carry out exchange operation.The classifier analyzes characteristics of captured data, and then the unknown type is sent into the training sample, after neural network learning, the neural network classifier is used to classify again. According to the controllable network traffic data, we calculate each traffic characteristic statistics data during the period of every minute as the input of the neural network. Through the visual analysis of network traffic anomaly, the network traffic anomaly is divided into two categories. The advantages of quantum genetic algorithm is used to overcome the slow convergence and local convergence of BP algorithm, combined with the BP algorithm at the same time, it also solves the problem that quantum genetic algorithm cannot find the approximate optimal solution in a short period of time, and introducing gradient information of BP algorithm will avoid this kind of phenomenon. So the training of the BP neural network can be divided into two parts. Quantum genetic algorithm is used to optimize the initial weights of network, then BP algorithm is used to train the attack data to get the network model. Three layer of neural network is used. Symbol WI_{ij} represents connection weight value between the ith node of the input layer and the jth node of the hidden layer. Quantity WO_{ji} represents connection weight value between the jth node in the hidden layer and the ith node in the output layer. Symbol H_i represents output of the ith node of hidden layer, O_i represents output of the ith node of the output layer and I_i represents output of the ith node of the input layer. The optimization process is as follows.

Step 1. Initialize population p, including mutation probability and connection weight WI_{ij}, WO_{ji}.

Step 2. Calculate fitness value of each individual and sort it. Here

$$p_s = \frac{f_i}{\sum_{i=1}^{N} f_i,}$$

where f_i represents the fitness value of individual i, which can be measured by the sum of square errors.

Step 3. Use mutation probability p_m to generate the new individual $G_j^{'}$ of G_j.

Step 4. The new individual is inserted into population p and we should fitness of the new individual.

Step 5. Calculate square sum of error of BP neural network. If it achieves preset

value $\varepsilon_{\mathrm{GA}}$, the algorithm stops. Otherwise, it turns to step 3 to go on.

Step 6. The solution of quantum genetic algorithm is taken as initial weight value and BP algorithm is used to train the network until the accuracy meets $\varepsilon_{\mathrm{BP}} < \varepsilon_{\mathrm{GA}}$.

Repeat the above steps, quantum genetic algorithm stops until the network error satisfies the condition. We choose a group of weight value with the smallest network error as initial weights of BP network training. Then we use BP algorithm for training, until the final error requirement is met. The parameters of the neural network are coded into chromosomes directly involved in the genetic operation, which can make the quantum genetic algorithm combined with specific problem. All of the neural network weight value has no limit of the value space, belonging to the unconstrained problem, so that the quantum genetic operation can make global search in the space as possible.

3. Experiment and analysis

We use BP neural network based on genetic and BP neural network based on quantum genetic algorithms for security audit system, so as to verify the application effect of proposed scheme. Experiments data comes from KDD99 CUP data set. We use 10000 piece of data as the training data, and 3000, 5000, 8000, 12000 piece of data is taken as the testing data.

Training process based on quantum genetic BP is shown in Fig. 2, training process based on genetic BP is shown in Fig. 3 and training process based on BP is shown in Fig. 4. The training step of quantum genetic BP is 22, training time is 12.105000 seconds, and the mean square error is 1.2681e–003. The step of genetic BP training is 563, its time is 22.119000 seconds, and the mean square error is 1.2786e–003. The training step of BP is 1132, its time is 25.112000 seconds, and the mean square error is 1.213e–2. It can be seen that quantum genetic BP algorithm exhibits fast convergence speed and low error.

Security audit result based on BP neural network is shown in Table 1, security audit result based on genetic BP neural network is shown in Table 2 and security audit result based on quantum genetic BP neural network is shown in Table 3. It can be seen that network security audit system based on quantum genetic BP has higher detection rate than traditional schemes. Besides, the proposed scheme has lower false alarm rate and missing report rate.

Table 1. Security audit result based on BP neural network

sample number	3000	5000	8000	12000
correctly detected sample number	2709	4533	7284	10926
accuracy	90.3 %	90.66 %	91.05 %	91.05 %
false alarm number	203	362	497	737
false alarm rate	6.77 %	7.24 %	6.21 %	6.14 %
number of missing report	88	105	219	337
rate of missing report	2.93 %	2.10 %	2.74 %	2.81 %

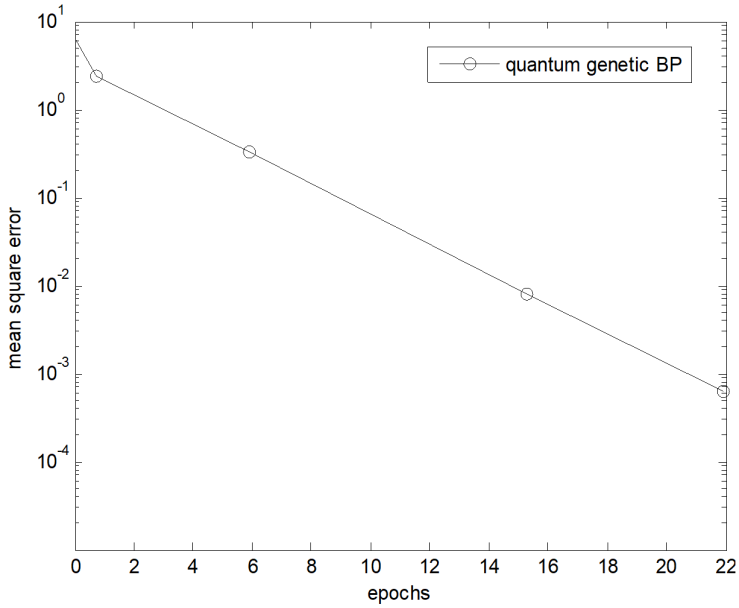


Fig. 2. Training process based on quantum genetic BP

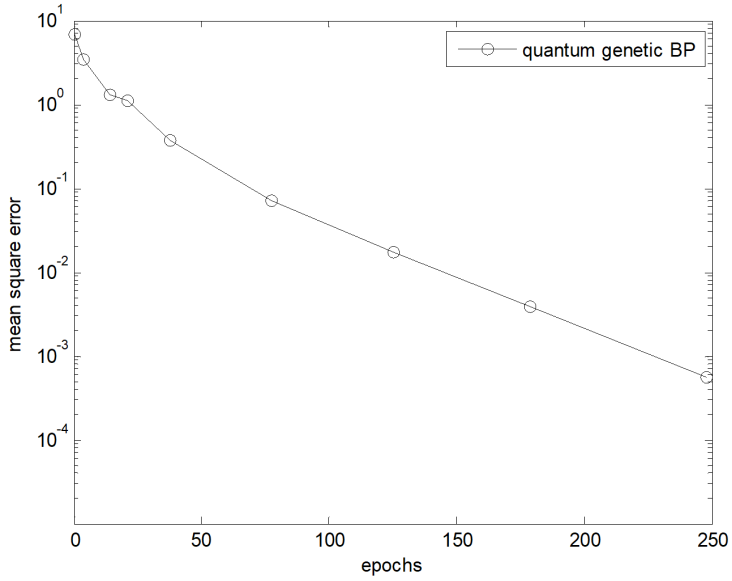


Fig. 3. Training process based on genetic BP

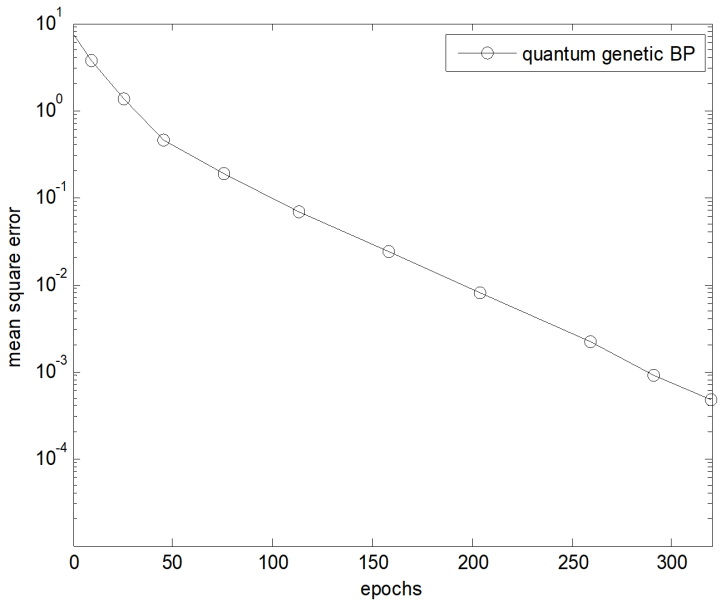


Fig. 4. Training process based on BP

4. Conclusion

BP neural network is applied to network security audit system. Because inherent defects of BP neural network, quantum genetic algorithm is given to optimize the initial weights of BP neural network, and it is applied in security audit system. The experiment results show the proposed scheme has good performance. Based on the practical point of view, the designed network security audit system is not perfect enough, especially the selection of training samples have certain effects on the accuracy of the network identification, which is one of the research direction in the future. With the continuous progress of science and development, interdisciplinary knowledge can be adopted to find out the better solution.

Table 2. Security audit result based on genetic BP neural network

sample number	3000	5000	8000	12000
correctly detected sample number	2827	4705	7518	11296
accuracy	94.23 %	94.10 %	93.98 %	94.13 %
false alarm number	112	221	356	577
false alarm rate	3.73 %	4.42 %	4.45 %	4.81 %
number of missing report	61	74	126	127
rate of missing report	2.03 %	1.48 %	1.58 %	1.06 %

Table 3. Security audit result based on quantum genetic BP neural network

sample number	3000	5000	8000	12000
correctly detected sample number	2871	4796	7618	11426
accuracy	95.70 %	95.92 %	95.23 %	95.22 %
false alarm number	98	156	303	407
false alarm rate	3.27 %	3.12 %	3.45 %	3.39 %
number of missing report	31	48	106	167
missing report rate	1.03 %	0.96 %	1.33 %	1.39 %

References

[1] X. W. Ning, P. Y. Liu: *Log system design in support of linkage analysis of security audit and computer forensics.* Computer Engineering and Design *30* (2009), No. 24, 5580–5583.

[2] M. Liu, Q. Zhang, H. Yhao, D. Yu: *Network security situation assessment based on data fusion.* Proc. International Workshop on Knowledge Discovery and Data Mining (WKDD), 23–24 Januar 2008, Adelaide, SA, Australia, IEEE Conference Publications (2008), 542–545.

[3] J. E. L. de Vergara, A. Guerrero, V. A. Villagrá, J. Berrocal: *Ontology-based network management: Study cases and lessons learned.* Journal of Network and Systems Management *17* (2009), No. 3, 234–254.

[4] J. Lai, H. Wang, X. Liu, Y. Liang: *A quantitative prediction method of network security situation based on wavelet neural network.* International Symposium on Data, Privacy, and E-Commerce (ISDPE), 1–3 November 2007, Chengdu, Sichuan, China, IEEE Conference Publications (2007), 197–202.

[5] Y. Liang, H. Q. Wang, J. B. Lai: *Quantification of network security situational awareness based on evolutionary neural network.* International Conference on Machine Learning and Cybernetics (ICMLC), 19–22 August 2007, Hong Kong, China, IEEE Conference Publications *6* (2007), 3267–3272.

[6] Y. Liang, H. Q. Wang, H. B. Cai, Y. J. He: *A novel stochastic modeling method for network security situational awareness.* International Conference on Industrial Electronics and Applications (ICIEA), 3–5 June 2008, Singapore, Singapore, IEEE Conference Publications (2008) 2422–2426.

[7] H. Wang, J. Lai, X. Liu: *A quantitative forecast method of network security situation basedon BP neural network with genetic algorithm.* International Multi-Symposiums on Computer and Computational Sciences (IMSCCS), 13–15 August 2007, Iowa City, IA, USA, IEEE Conference Publications (2007), 374–380.

[8] J. Li, H. Wang: *A quantification method for network security situational awareness based on conditional random fields.* Proc. International Conference on Computer Sciences and Convergence Information Technology (ICCIT), 24–26 November 2009, Seoul, South Korea, IEEE Conference Publications, (2009), 993–998.

[9] J. Gao, B. Zhang, X. Chen, Z. Luo: *Ontology-based model of network and computer attacks for security assessment.* Journal of Shanghai Jiaotong University (Science) *18* (2013), No. 5, 554–562.

[10] Yang J, Li B, Zhuang Z: *Research of quantum genetic algorith and its application in blind source separation.* Journal of Electronics (China) *20*, (2003), No. 1, 62–68.

[11] K. H. Han, J. H. Kim: *Quantum-inspired evolutionary algorithms with a new termination criterion gate, and two-phase scheme.* IEEE Transactions on Evolutionary Computation *8* (2004), No. 2, 156–169.

[12] X. R. Zhu, X. H. Zhang: *A quantum genetic algorithm with repair function and its application in knapsack question.* Journal of Computer Applications *27* (2007), No. 5, 1187–1190.

[13] P. C. Li, S. Y. Li: *Quantum genetic algorithm based on real-coded and objective function's gradient.* Journal of Harbin Institute of Technology *38* (2006), No. 8, 1216–1218,1223.

[14] S. H. Xu, C. Xu, X. Hao, Y. Wang, P. C. Li: *Improved quantum genetic algorithm with double chains and its application.* Application Research of Computers *21* (2010), No. 6, 2090–2092.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 187–198

Virtual technology of cache and data real time allocation in cloud computing data center[1]

Wenrong Bai[2]

Abstract. In order to develop a new technology used in data analysis and caching, a virtual technology of cache and data real time allocation is proposed. As a new commercial computing service model, users can obtain information services anytime and anywhere through cloud computing. Cloud computing data center has thousands of physical nodes, the application of virtualization technology allows the system to complete the dynamic migration between nodes and data resources on-demand distribution, thus supporting the data center resources, cache and distribution. However, with the expansion of cloud computing, the inherent dynamics of the system and the complexity of management are gradually increasing. A heuristic cloud computing data center resource management approach is proposed to reduce energy costs and ensure customer service quality. The minimum migration strategy is used to keep the number of VMS transferred at a minimum. In order to evaluate the improved algorithm comprehensively, the energy evaluation, SLA violation rate, virtual machine transfer cost and average migration time are compared and analyzed. The results show that the proposed RFFD method can reduce the cost of energy and reduce the cost of virtual machine migration in the case of lower SLA violation rate. Based on the above finding, it is concluded that this new real-time data distribution virtual technology plays an important role in data caching and analysis of cloud computing center.

Key words. Cloud computing, data centers, virtual technology, resource allocation.

1. Introduction

Cloud computing is a product of the integration of distributed computing, parallel computing, grid storage and virtualization, and it has great commercial value. In recent years, with the advent of the networked information age, cloud computing has received widespread attention and rapid development as an emerging business model. It is seen as the third wave of IT following changes in computers and the

[1]The authors acknowledge the National Natural Science Foundation of China (Grant: 51578109), the National Natural Science Foundation of China (Grant: 51121005).

[2]Inner Mongolia Electronic Information Vocational Technical College, 010070, Hohhot, Inner Mongolia, China

internet. However, virtualization is the key technology for cloud computing. Cloud computing data center obtains a lot of advantages, such as flexible management, high resource utilization and strong scalability, making the data center management more convenient and efficient [1].

At the same time, with the rapid development of cloud computing, the surge of energy consumption of basic equipment, such as network equipment and servers, has led to high operating costs, resulting in serious resource and environmental problems. Therefore, at this stage, the urgent problem is to develop new technologies to manage and optimize cloud computing data center resources. Sharkh studied resource allocation problem in cloud computing center under network virtual environment, and designed a new algorithm. Xiao Zhen used real-time migration technology to dynamically manage and integrate virtual machine resources [2]. A resource management method for heuristic cloud computing, data center caching and real-time data allocation based on frame is proposed around the theme of cloud computing data center resource management. Through the evaluation and analysis of cloud computing, data center structure, energy model and related performance indicators, virtual machine management method based on frame is proposed after the improvement of MBFD algorithm. Finally, the simulation is carried out in order to improve the utilization rate of cloud computing data, liberate human resources and compress management costs.

2. Experimental procedure

2.1. Cloud computing and data center for cloud computing

At present, the most authoritative definition of cloud computing comes from American National Institute of Standards and Technology. Cloud computing is a model of computing resources that can be independently controlled. Users can access the resources in the model with a convenient, on-demand way via Internet. These resources exist in a dynamically shared pool of resources and can be obtained and released in a quick and intelligent manner [3]. Cloud computing is not a new technology appeared suddenly, but gradually evolved by combining the traditional computing technology and the network technology. As shown in Fig. 1, compared with traditional computing technology, they are both continuous and distinct. They are the aggregation of advanced computing and service technology.

Cloud computing can be divided into three categories from the service type. The first is to provide services based on virtual infrastructure, that is, infrastructure services (IaaS). The second is to provide software development platform based services, that is, platform services (PaaS). The last one is the service that provides the application software directly to the user, namely software service (SaaS). Because of the diversity of cloud computing services, its architecture is relatively complex. As shown in Fig. 2, from the bottom layer to the top layer, it can be roughly divided into four layers: hardware resource layer, virtualization resource layer, management component layer and service interface layer [4].

Data center (DC) is a complete set of complex facilities that include not only

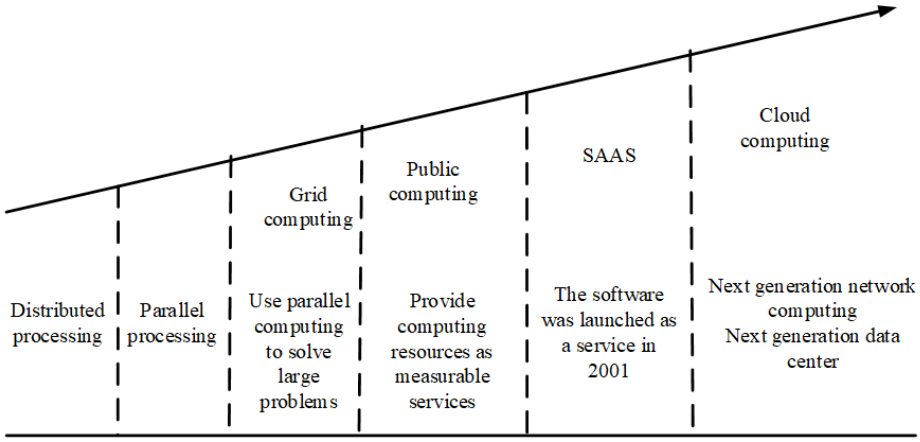


Fig. 1. History of cloud computing

the basic computer system and the physical equipment, but also the monitoring, scheduling and security mechanisms responsible for its normal operation [5]. It is the ultimate bearer of cloud computing services, providing basic support for customer applications such as resources and services. The physical resources such as computing, storage and network in data center are constructed into dynamically adjusted virtual resource pools, so that the basic service units change from physical hosts to virtual machines. The automatic deployment, dynamic expansion and on-demand allocation of cloud computing resources are realized. Users can access cloud computing services on demand and out-of-the-box.

2.2. Overview of virtualization technology

Virtualization technology is the foundation of cloud computing, including network virtualization, storage virtualization, server virtualization, desktop virtualization, application virtualization, presentation virtualization. In general, virtualization is an abstraction layer that separates physical hardware from the operating system to provide greater IT resource utilization and flexibility. Processor, memory, storage and network hardware resources are abstracted into standardized virtual hardware. It is packaged in a hardware-independent virtual machine along with a complete operating environment, including operating systems and applications. The server virtualization principle is depicted in Fig. 3.

By using virtualization software, virtualized data center software abstracts the underlying hardware device and the upper operating system, and manages the resources using appropriate methods. Through virtualization, virtual can create multiple virtual machines on a server. At this point, the virtual machines are isolated from each other, while the applications, operating systems and hardware devices are encapsulated so that the system can be quickly backed up and deployed quickly.

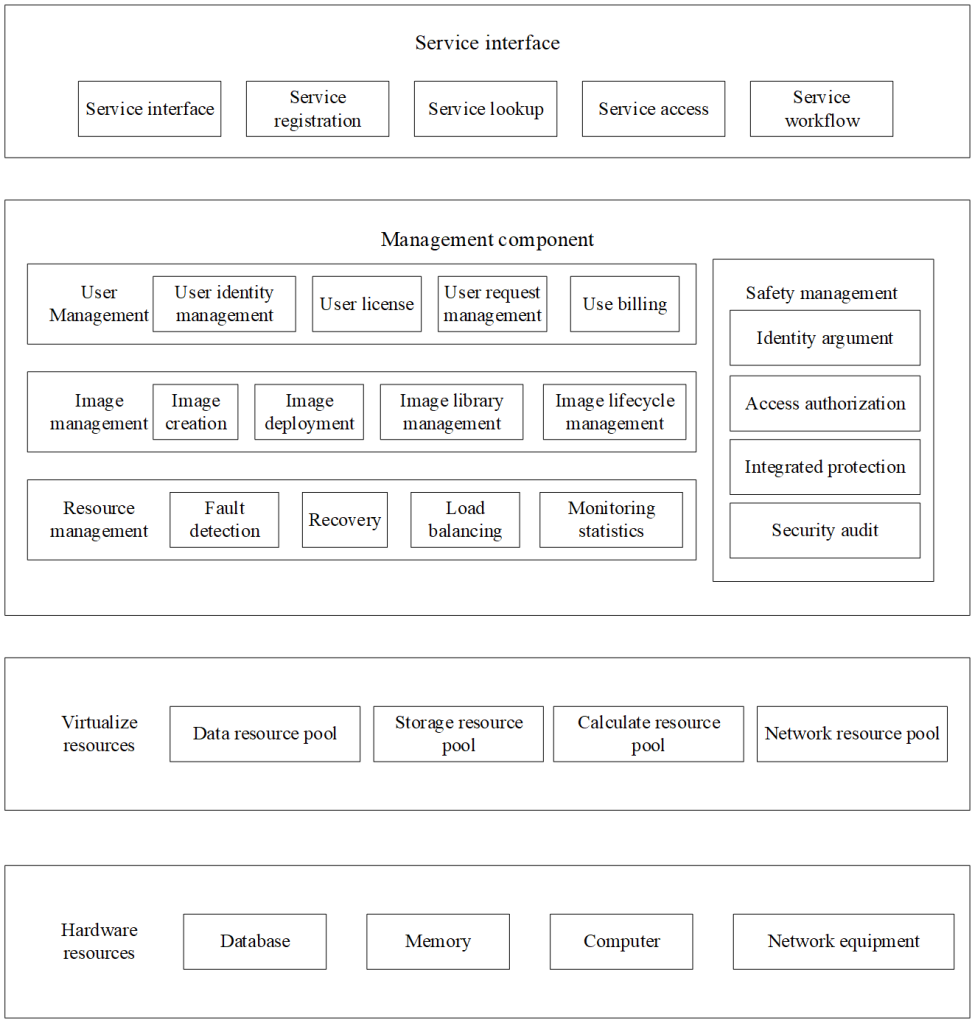


Fig. 2. Cloud computing architecture diagram

2.3. The system model of virtual cloud computing center

Starting from the basic composition of data center, the composition of data center structure is analyzed. It mainly includes host, frame, cooling system and network switch and other physical equipment. At the same time, the energy calculation method of data center and the index of evaluating data center performance are put forward. The energy cost of cloud computing data centers is divided into three parts, namely, host, cooling equipment and network [6]. The energy cost of the host is mainly generated by CPU, memory, hard disk and network interface, in which the energy cost of CPU accounts for most of the total overhead. In order to calculate the energy cost of the host more accurately, the energy overhead of the host is collected

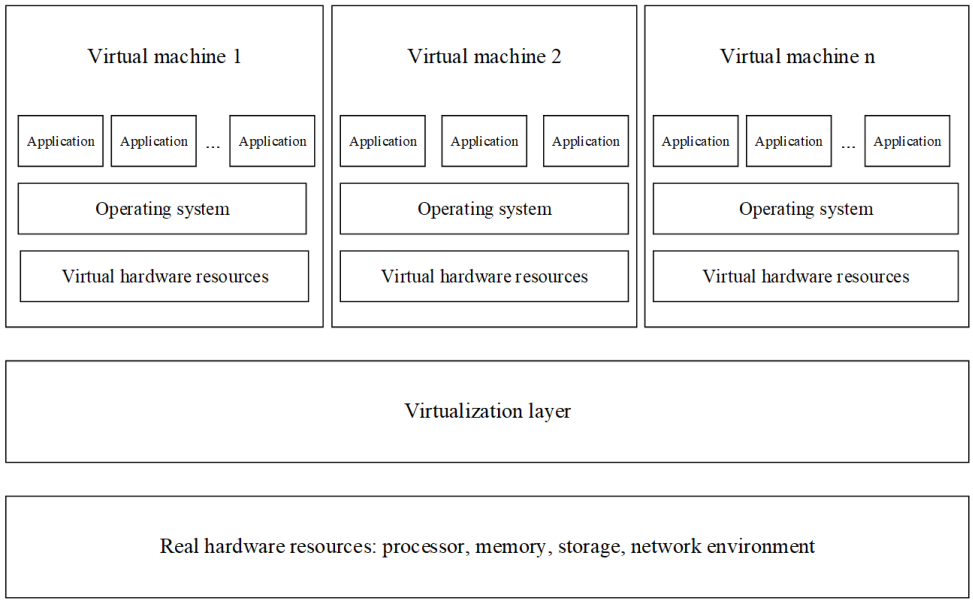


Fig. 3. Server virtualization principle

every 10 % of the host CPU utilization from 0 % to 100 %, with a total of 11 nodes of data. For hosts with CPU utilization falling at each 10 % interval, a specific energy cost can be calculated from the linear relationship

$$P_i = P_{\text{pre}} + (u_i - o_{\text{pre}}) \times (P_{\text{after}} - P_{\text{pre}}) \times 10\,. \tag{1}$$

In formula (1), P_i represents the host energy cost, and u_i is the CPU utilization of the host. P_{pre} represents the energy overhead of the previous node when the CPU utilization is u_i. P_{after} represents the energy overhead of the latter node when the CPU utilization is u_i.

Let P_{TE}, P_{Agr}, and P_{CR} represent the energy overhead of the rack upper-layer switch, aggregation layer switch, and core routing, respectively. Then

$$P_{\text{Net}} = P_{\text{TE}} + P_{\text{Agr}} + P_{\text{CR}}\,. \tag{2}$$

The total energy cost of the rack is $P_{\text{R}} = \sum_{j=1}^{n} P_j$ (n is the total number of running racks), and the total energy cost of the host is $P_{\text{S}} = \sum_{i=1}^{m} P_i$ (m is the total number of running hosts). The total energy cost of the cloud computing data center can be expressed as

$$P_{\text{DC}} = P_{\text{S}} + P_{\text{R}} + P_{\text{Net}}\,. \tag{3}$$

There are four main metrics for quantifying the performance of a virtualized cloud computing data center, namely, energy overhead, SLA breach rate, virtual machine migration cost and average migration time. The cost of energy is calculated as described above. SLA violation rate refers to the ratio of the total time of all virtual

machine SLA violations in the data center to the sum of the time the virtual machines run. Let symbol n represent the number of virtual machines in the data center. $ViolateTime(VM_i)$ is the virtual machine SLA collision time, and $Runtime(VM_i)$ is the virtual machine running time. Then

$$sla\_violate\_rate = \frac{\sum_{i=1}^{n} ViolateTime(VM_i)}{\sum_{i=1}^{n} Runtime(VM_i)}. \tag{4}$$

Virtual machine migration costs is represented by the sum of all virtual machine migration costs. It is related to the specific transfer cost of each virtual machine and the selected destination node. When the destination node and the virtual machine are in the same frame, the transfer cost is the number of migration. When not in the same frame, the migration cost is multiplied by a constant K (the constant value is higher than 1)

$$migrate\_cost = \sum_{i=1}^{n} migrateCost(VM_i). \tag{5}$$

Here, $migrateCost(VM_i)$ is the total cost of the virtual machine being migrated. The average migration time of a virtual machine represents the average time spent in the migration of the virtual machine. Symbol n is the total number of virtual machine migration times.

Let $migrateTime_(i)$ represent the time it takes to migrate. Then

$$average\_migration\_time = \frac{\sum_{i=1}^{n} migrateTime_i}{n}. \tag{6}$$

2.4. Virtual machine management algorithm and improvement

The core problem of virtualization cloud computing data center resource management is the migration of virtual machines and virtual machine placement. Virtual machine migration will virtually increase the overhead of CPU and network bandwidth. Therefore, the minimum migration strategy is adopted to reduce the migration of the virtual machine [7]. The core idea is that the least amount of the virtual machine is used to control the resource utilization of the host within the threshold. The virtual machine placement problem is improved by the best fit decreasing algorithm. First, the descending order is sorted according to the MIPS required by the virtual machine, and then the best placement host is found from the resource utilization higher than the threshold lower value and the host set below the threshold upper limit. If there is no host that meets, it is searched from the collection of hosts whose resource utilization is below the threshold. If there is no host that conforms, a host is turned on and the virtual machine is placed on it. On this basis, virtual machine placement algorithm based on frame is proposed. First, a lower limit is set for rack utilization. When the rack utilization is below the threshold, all virtual machines on the rack are migrated. After that, the host on the rack is turned off, and the rack itself and the network device are used to save energy.

It is assumed that V_j is the set of virtual machines allocated on host j, and $u_{a(v)}$ represents the proportion of CPU allocated to virtual machine v. The purpose of least migration strategy is to find the set $V_{mm} \in V_j$ that satisfies the formula

$$V_{mm} = \begin{cases} \left\{ V \,\middle|\, V \in V_j, u_j - \sum_{v \in V} u_{a(v)} T_u, |V| \to \min \langle u_j \, \mathrm{T_u} \rangle \right. \\ V_j \langle u_j \, \mathrm{T_u} \rangle , \\ \emptyset \langle \text{other} \rangle . \end{cases} \tag{7}$$

In order to evaluate the performance of the resource management method proposed in the virtualized cloud computing data center, the rack components were expanded in the simulation platform. At the same time, five algorithms of MBFD, IMBFD, RFFD, R & D virtual machine placement algorithm, NURBN (non-low load rack optimal node algorithm) and NURMG (non-low load rack minimization interval algorithm) are implemented [8]. Simulation experiments are conducted to verify the performance of different algorithms in terms of energy cost, SLA violation rate, virtual machine transfer cost and average migration time. Meanwhile, a comparative analysis is carried out.

3. Results and discussion

3.1. Total energy cost evaluation

Figure 4 compares the total energy costs of five different virtual machine resource management methods for MBFD, IMBFD, RFFD, NURBN, and NURMG, while Fig. 5 shows the total transfer cost of data center virtual machines. Compared with the MBFD algorithm, the energy utilization rates of IMBFD, RFFD, NURBN and NURMG have been improved by 1.7 %, 8.1 %, 6.3 % and 10.1 % respectively. The RFFD, NURBN and NURMG can achieve a relatively large increase in efficiency because these rack-based virtual machine management methods can reduce the number of racks running, thereby reducing the cooling system and rack energy costs [9]. With the increase of the number of virtual machines, the improved MBFD algorithm is more efficient than virtual machine placement algorithm based on frame, especially the rack based virtual machine placement method, which makes the energy saving more obvious.

3.2. Cost analysis of virtual machine migration

Both NURBN and RFFD have the preference to migrate virtual machines to rack hosts whose frame load is greater than the threshold limit. This reduces the number of racks that load is below the threshold, thus reducing the migration of the virtual machine. As a result, the cost of their virtual machine migration is reduced by 2.7 % and 0.2 %, respectively, compared with MBFD. With the increasing number of virtual machines, the cost of IMBFD, NURMG and NURBN transfer is in a

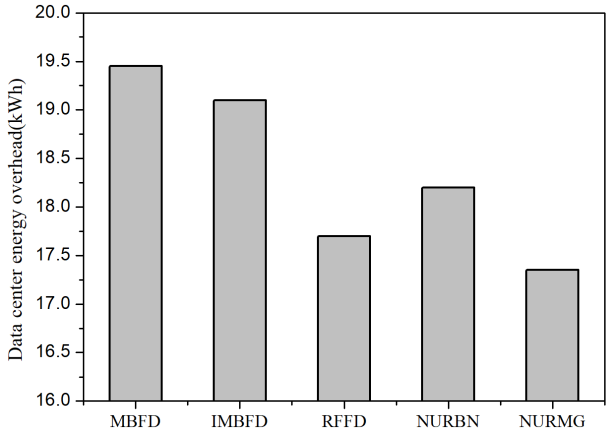


Fig. 4. Comparison of energy cost of data center under different resource management methods

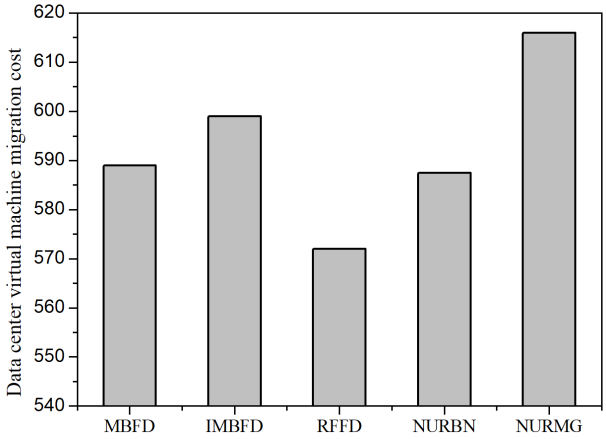


Fig. 5. Comparison of total transfer cost of data center virtual machines under different resource management methods

relatively stable state with the MBFD contrast. On the contrary, RFFD increases as the number of virtual machines increases, and the cost of virtual machine migration decreases further.

3.3. Comparison of service grade contract violation rate

As shown in Fig. 6, NURMG has the highest SLA violation compared to other methods. Because it minimizes the distance between the host's resource load and the set host load upper limit, it is easy to cause SLA violation because of changes in the resource load. NURBN attempts to find the destination host on the rack where all the resources load is greater than the threshold threshold. This can reduce the number of hosts, but also lead to relatively high host resource load. Therefore, com-

pared with RFFD, its SLA violation rate is still much higher [10]. With the increase of the number of virtual machines, the SLA violation rate of resource management method proposed by MBFD has been improved and the range of increase is in a steady state with sufficient resources. Figure 7 compares average migration time of data center virtual machine under different resource management.

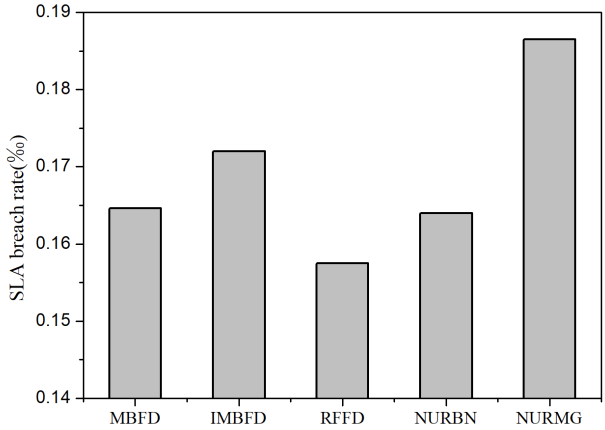


Fig. 6. Comparison of data center SLA violation rates under different resource management methods

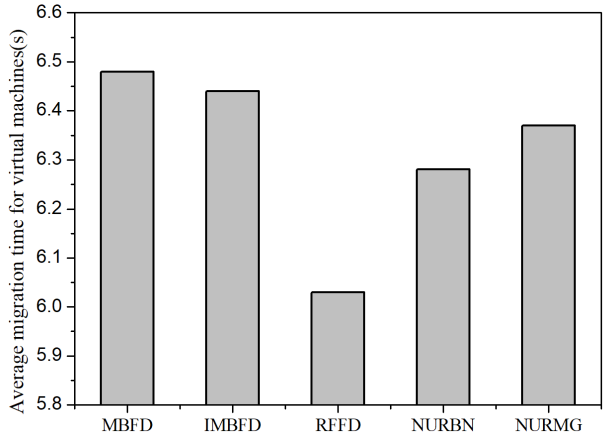


Fig. 7. Comparison of average migration time of data center virtual machine under different resource management

3.4. Calculation of average migration time of virtual machines

The virtual machine placement algorithm based on frame can reduce the average migration time of the virtual machine. Because these algorithms first look for the

destination node on the rack node which is not below the threshold, thus increasing the probability that the virtual machine migrates to the frame. It takes less time for virtual machines to migrate between nodes within the rack. In these methods, RFFD has the least average migration time. With the increase of the number of virtual machines, the average migration time of virtual machines is relatively stable relative to MBFD.

RFFD is the best choice from the comprehensive performance of energy cost, SLA violation rate, virtual machine transfer cost and average migration time. Because the RFFD algorithm preferentially puts the virtual machine on the other physical nodes of the frame. If the other nodes of this frame cannot place the virtual machine, it is arranged in descending order according to the utilization ratio of the rack, and then the first frame for placing the virtual machine can be selected. This not only reduces the number of racks, but also reduces the virtual machine migration costs and average migration time.

4. Conclusion

Aiming at the technology of data center cache and real-time data allocation in cloud computing, a heuristic cloud computing data center resource management method is proposed. The structure and energy calculation model of cloud computing data center are analyzed. The least migration strategy is used to screen the virtual machines, and the number of VMS is guaranteed to be the least. Then, in view of the placement of virtual machines, a virtual machine placement algorithm based on frame is proposed according to the improved heuristic algorithm. By comparing the performance of several algorithms, the results show that compared with the previous methods, the proposed heuristic cloud computing data center resource management method reduces the energy consumption in the case of guaranteeing low SLA collision rate.

Due to the complexity and complexity of cloud computing resources, there are many influencing factors besides the energy cost, SLA breach rate, migration cost and average migration time, which need to be considered in the management of cloud computing data center. Subsequent research can improve the virtual machine migration progress and access to better cloud computing data management performance. At the same time, it can improve the stability of the system in order to develop the virtual technology in the cloud computing data center for further application.

References

[1] A. Beloglazov, J. Abawajy, R. Buyya: *Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing.* Future Generation Computer Systems *28* (2012), No. 5, 755–768.

[2] M. A. Sharkh, M. Jammal, A. Shami, A. Ouda: *Resource allocation in a network-based cloud computing environment: Design challenges.* IEEE Communications Magazine *51* (2013), No. 11, 46–52.

[3] Z. Xiao, W. Song, Q. Chen: *Dynamic resource allocation using virtual machines for*

cloud computing environment. IEEE Transactions on Parallel and Distributed Systems *24* (2013), No. 6, 1107–1117.

[4] S. Mohapatra, S. Mohanty, S. Pattanayak, A. Hota: *Comparison of various platforms in cloud computing.* International Journal of Computer Applications *162* (2174), No. 7, 28–33.

[5] H. T. Dinh, C. Lee, D. Niyato, P. Wang: *A survey of mobile cloud computing: architecture, applications, and approaches.* Wireless Communications and Mobile Computing *13* (2013), No. 18, 1587–1611.

[6] R. Moreno-Vozmediano, R. S. Montero, E. Huedo, I. M. Llorente: *Cross-site virtual network in cloud and fog computing.* IEEE Cloud Computing *4* (2017), No. 2, 46–53.

[7] F. Al-Haidari, M. Sqalli, K. Salah: *Impact of CPU utilization thresholds and scaling size on autoscaling cloud resources.* International Conference on Cloud Computing Technology and Science (CloudCom), 2–5 December 2013, Bristol, UK, IEEE Conference Publications *2* (2013), 256–261.

[8] D. Yang, C. Deng, Z. Zhao: *Dynamic scheduling method of virtual resources based on the prediction model.* International Conference on Collaborative Computing: Networking, Applications and Worksharing, 10–11 November 2016, Beijing, China, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering *316* (2016), 384–396.

[9] W. Tian, Y. Zhao: *A toolkit for modeling and simulation of real-time virtual machine allocation in a cloud data center.* Optimized Cloud Resource Management and Scheduling (2015), 217–243.

[10] S. Hosseinimotlagh, F. Khunjush, R. Samadzadeh: *SEATS: Smart energy-aware task scheduling in real-time cloud computing.* Journal of Supercomputing *71*, (2015), No. 1, 45–66.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 199–208

Damage and failure of concrete based on material mechanics

Qin Wang[1], Chenyu Peng[1], Xiangtian Tong[2]

Abstract. The failure of concrete materials and structures is caused by the damage evolution of meso-scale and macro-scale. To solve this problem, we analyzed the failure process of concrete structure caused by meso-scale crack propagation. The numerical simulation results show that the crack initiation occurred at the initial crack tip and gradually formed the main crack. Also, the results show that the nominal strength decreases with the increase of sample size. The fracture energy increases and the dispersion of the macroscopic mechanical properties decreases. We concluded that the linear elastic fracture mechanics theory can be used to analyze. In addition, the validity of the numerical analysis method and the meso-multi-crack model is verified.

Key words. Concrete, meso-cracks, failure.

1. Introduction

Concrete is a composite material composed of cement mortar matrix, aggregate and the interface between them. It has the characteristics of multiphase, heterogeneity and quasi brittleness. As an important engineering material, concrete is widely used in the construction of roads, bridges, dams, airports, tunnels and slopes and other public infrastructure and housing and other civilian facilities. However, the damage caused by the deterioration of the material properties of concrete structures often leads to immeasurable losses to the lives and property of the people [1].

In the analysis of concrete material and structure design, it is generally used as a uniform solid at the macro scale. However, in the analysis of the material properties, it is necessary to consider the discontinuity between the heterogeneous materials. The discontinuity of the micro scale is initially manifested as the initial damage state of the concrete formed by the meso-crack distribution. After the damage evolution process characterized by meso-crack coalescence (including fusion and crossover), it is shown as the discontinuity of macroscopic scale at last, that is, the generation of macroscopic crack. Therefore, the failure of concrete material is caused by the initiation of meso-cracks in concrete [2]. In addition, the propagation of fine lines,

[1]Xijing University, Xian, China
[2]Scegc No. 5 Construction Engineering Group Company Ltd., Xian, China

coalescence and macroscopic cracks will aggravate the damage of concrete in the process of cross scale evolution.

According to the different levels of failure behavior, the failure of concrete can be divided into three levels such as material, component and structure. The starting point of the effect is the initiation of the micro defects in the material, and the end point is the destruction of all levels. Generally, the failure analysis of concrete material is based on meso-scale, and the failure analysis is based on the macro scale. The failure analysis of components can choose different research scales based on the research objectives. In this way, the method of adopting different research scales for different structural levels meets the needs of engineering practice to a certain extent. However, the characteristics of the cross-scale evolution of concrete damage are ignored. It is difficult to describe the micro crack propagation in concrete, which leads to the deterioration of the performance of the macro components or structures [3]. Due to the failure of evolution, the meso-damage characteristics of concrete damage are restricted to the macroscopic property or may significantly affect the macroscopic properties. However, it is not clear how the micro structure, the material properties and the state of the phase affect the macroscopic properties of concrete under any conditions [4]. Although there is no uniform definition of the scale of the research, the solution of these problems is very important for the design of concrete and concrete structures.

2. Numerical simulation of damage and failure of three-point bending beam

The structure and loading conditions of the three-point bending beam are relatively simple, which is one of the most commonly-used tensile (I Type) fracture specimens. In this section, based on the meso-crack model, the heterogeneity of concrete material is considered. The damage and failure of concrete beams with initial crack and without initial macro crack are simulated. This paper discusses the process of component failure caused by the damage evolution, and presents a multiscale simulation method for the damage evolution process.

2.1. Failure simulation of damage evolution of concrete beams with initial cracks

In order to study the macroscopic fracture phenomenon of concrete members, one or two prefabricated cracks (usually more than 50 mm) should be prepared in the specimens, which can ensure that the fracture begins to expand from the tip of the crack. Bazant and Pfeiffier are selected to study the effect of concrete fracture size on concrete three-point bending beam specimens, and the plane stress problem is simplified [5]. The plane size of the numerical specimen is $800\,\text{mm} \times 200\,\text{mm}$ ($L \times H$). The initial macroscopic crack length in the middle of the beam is $a_\text{o} = 50\,\text{mm}$, the relative crack depth is $a_\text{o}/H = 0.25$. There is concentrated load F in the middle of the span of the beam. The proposed method can successfully simulate the failure process of concrete beam specimens subjected to the failure of different scales. As

shown in Fig. 1, its upper part represents the load and midspan displacement curve ($F - d$). Figure 1, bottom part, represents the load crack tip opening displacement curve (F – CMOD) of the specimen.

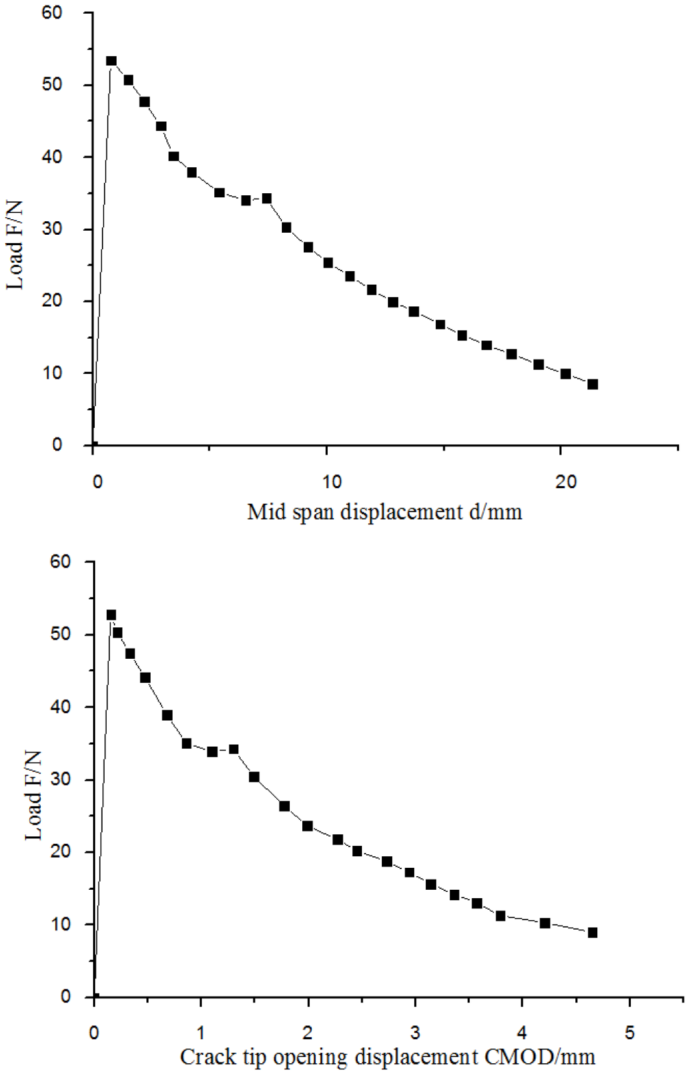


Fig. 1. Load displacement curve and load crack opening displacement curve of concrete beam specimen: upper part–load displacement curve, bottom part–load crack opening displacement curve

In this paper, the linear elastic fracture mechanics is used to judge the crack propagation, and it is assumed that the crack propagation is closely related to the damage of the meso-scopic element. Therefore, it can be seen from the figure that the unit at the beginning of the initial crack tip is damaged at the initial stage of

the stress, At the same time, the $F-d$ curve and the F – CMOD curve keep a good linearity, that is, the initial failure of the concrete samples with initial macro cracks have obvious brittleness. The specimen will soon reach the peak load with the gradual evolution of the damage. At the beginning of the macroscopic crack tip, a meso-unit first appears tensile damage, which marks the beginning of crack propagation [6]. After that, the damage began to accelerate, and the crack initiation occurred at the initial crack tip and gradually formed the main crack. Damage evolution enters the macro level from the meso-level, and the sample enters the unstable stage.

2.2. Failure simulation of damage evolution of concrete beams with initial cracks

Based on the study of the upper section, the characteristics of damage evolution and crack propagation of three-point bending beam with initial macroscopic crack and its influence on the mechanical properties of the specimens were studied. The main crack is obtained by the initial crack propagation. Because the failure process of concrete is random, the position and shape of the main crack are uncertain [7]. As for the practical engineering, the study on the specimens without initial macroscopic crack should be more practical.

In order to investigate the general situation, the numerical analysis of the damage and failure of concrete beam without initial macro crack is carried out by using the concrete meso-crack model. The samples are similar to the upper section in concrete mix proportion, aggregate gradation and meso-mechanical parameters. The same method can be used to simulate the failure process caused by the cross-scale evolution of the specimens [8]. As shown in Fig. 2, it represents the load displacement curve $(F-d)$. Because the concrete sample does not contain the prefabricated crack, there is no crack opening displacement data, that is, so-called F – CMOD curve.

As shown in Fig. 2, the meso-unit at the interface is subject to tensile damage in the early stage of concrete beams, and the meso-scopic crack begins to expand along the interface. The nonlinearity of the F-d curve is not very obvious at this time. An extended interface crack reaches a certain number before peak load and begins to expand into the mortar matrix. The evolution of the crack mainly comes from the damage of the meso-scopic element in the mortar matrix and $F-d$ exhibits a certain nonlinearity. After reaching the peak load, the fusion and intersection of the local micro cracks will occur. The main crack was formed in the specimen, and the evolution of damage evolves from micro level to macro level. The sample began to enter the unstable stage and the bearing capacity decreased.

3. Analysis of the size effect of concrete beams caused by the damage evolution

The size effect of concrete members is that the experimental side value of strength, toughness and other mechanical parameters have a regular change with the increase of specimen size. The mechanical parameters obtained from the concrete test not

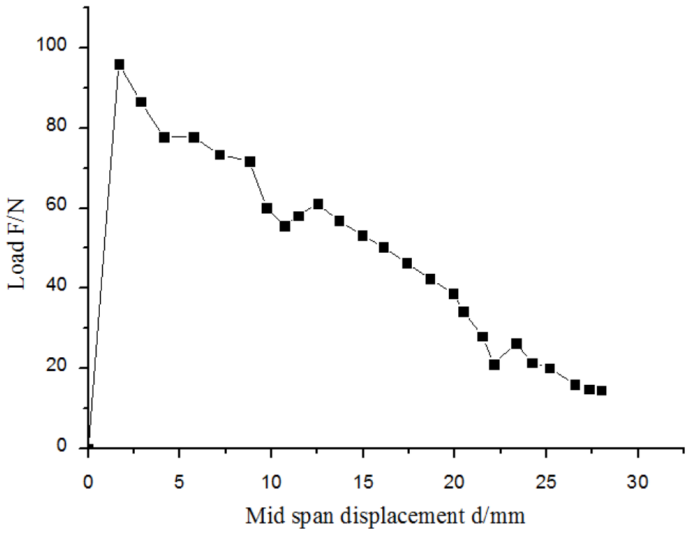


Fig. 2. Load mid-span displacement curve of concrete beam specimen

only depend on the performance of concrete, but also the size of the sample. For a long time, the size effect of concrete has been an obstacle to the design of concrete structures, life prediction and structural failure analysis. There has been a lot of work around this problem, such as Weibull statistical theory, energy release theory, fractal theory and so on. But the problem has not been completely resolved. In particular, the mechanism of concrete size effect has not been clearly understood and described.

3.1. Numerical test of size effect of concrete beam

The height of the concrete beam specimen is H (it is the characteristic size of the structure), and the length is $L = 4H$. The three-point bending load is adopted, and the load is F. In order to study the size effect of concrete materials, 5 kinds of specimens with H of 50 mm, 70 mm, 100 mm, 140 mm, and 200 mm were selected in this paper. The physical model of the concrete specimen of each size is randomly placed by the aggregate, and 7 samples of the same size are generated.

Based on the optimized numerical analysis method, the multi-scale modeling of concrete members is carried out, which can simulate the whole process of concrete beam failure. Figure 3 shows the failure mode of a concrete beam with a size of $H = 50$ mm and $H = 200$ mm. The results of the numerical simulation of each step can be used to obtain the load and displacement curves of the concrete. It includes linear phase, nonlinear strengthening stage and softening stage. As shown in both parts of Fig. 3, a–d is the representative point of the initial state of the sample, the limit of reinforcement, the softening stage and the failure state.

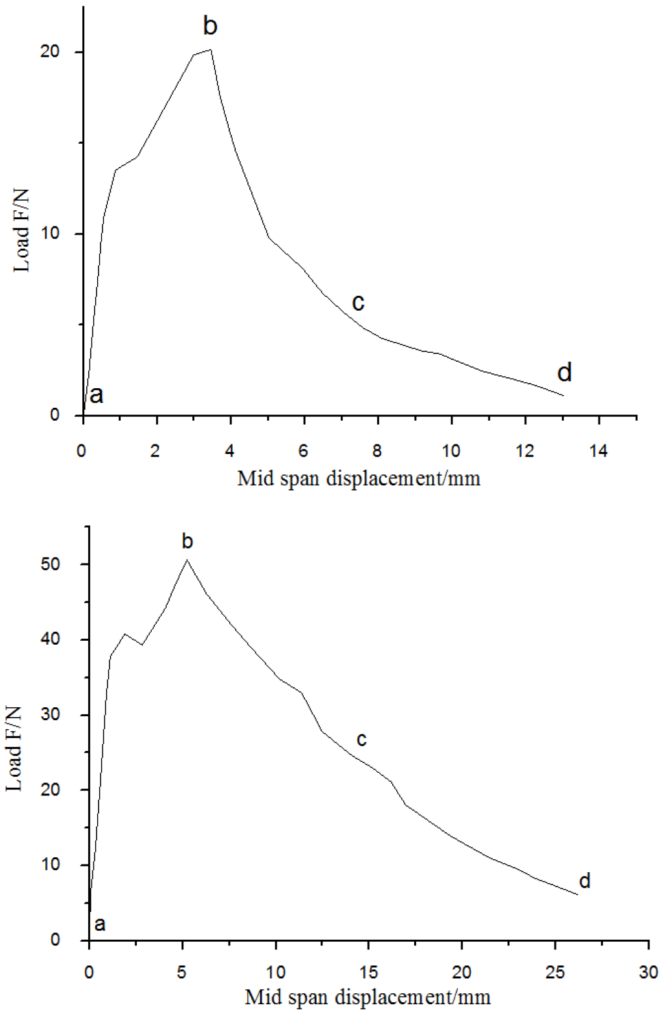


Fig. 3. Load mid span displacement curve of concrete beam: up–$H = 50$ mm, bottom–$H = 200$ mm

3.2. Numerical results analysis

The $F - d$ curves of the 7 samples of the same size were collected and fitted, and the fitting curve was obtained as shown in Fig. 4. It is used to represent the bending failure characteristic curve of the specimen.

In order to analyze the size effect of concrete structures, it is necessary to calculate the mechanical parameters of different sizes, including strength, fracture energy and so on. The maximum stress of concrete beam in this paper is:

$$\sigma_N = 6F/bN\,. \tag{1}$$

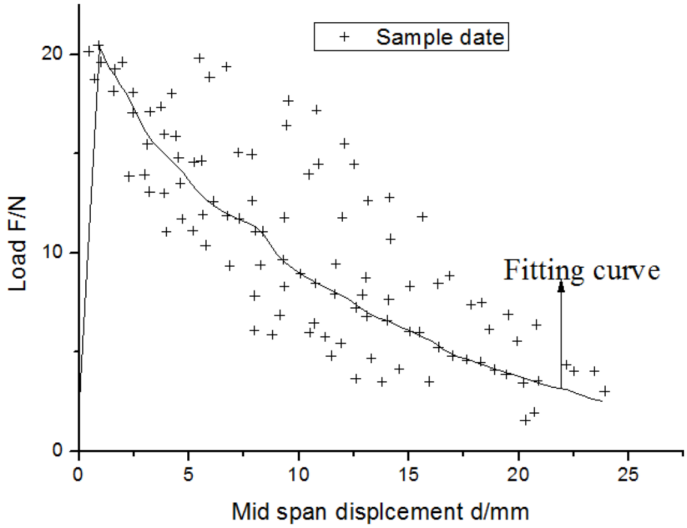


Fig. 4. Tensile fitting curve of 7 concrete beams in $H = 50\,\text{mm}$

In formula (1), b is the thickness of the beam (plane strain state, $b = 1\,\text{mm}$). Let d be the mid span displacement of the beam. The relative deflection of the sample is then

$$W = d/N\,. \tag{2}$$

Based on the load displacement curve in figure 4, the curves of the maximum stress (strength) and relative deflection of concrete beams with different sizes are obtained. As can be seen from the figure, the strength of the concrete beam decreases with the increase of the size, which indicates that the damage can lead to the size effect. In addition, the size effect also affects the post peak curve. When the small structure is destroyed, the relative deflection of the structure is larger. For larger structures, there are abrupt changes and jumps in the peak curve, and the relative deflection is smaller when the final failure occurs. The results of numerical simulation are summarized, and the maximum stress, the standard deviation of mid span stress and the fracture energy of concrete specimens with different sizes are obtained. As can be seen from Table 1, there is a size effect on the mechanical properties of concrete members due to the damage evolution. It shows that the strength of concrete specimens and the dispersion of the stress data in the mid span decrease with the increase of size. In addition, the fracture energy will increase.

Table 1. The size effect of the maximum stress and fracture energy

Sample size (mm)	50	70	100	140	200
Mid span maximum stress (MPa)	2.86	2.58	2.40	2.34	2.27
Mid span stress standard deviation	0.51	0.29	0.31	0.19	0.14
Fracture capacity (kJ/m)	2.41	3.14	3.51	5.22	9.11

3.3. Results

In this paper, the failure of concrete component is taken as the research focus, and the three-point bending beam of concrete is selected as the research object. The numerical analysis method and the multiple crack model are optimized and improved, and failure analysis of components caused by damage evolution is carried out. Therefore, the numerical simulation of three point bending of concrete beam with initial macro crack and without initial macro crack is carried out, while the failure process caused by the damage of the cross-scale evolution was investigated. On the basis of this, a multiscale simulation method is proposed to simulate the damage evolution process. And then the scale effect of concrete beams with different sizes is studied and the size effect and mechanism of the damage evolution are discussed. In addition, based on the fatigue failure mechanism of concrete, the multi-crack model is improved. Numerical simulation of fatigue damage of concrete beam is carried out, which analyses the fatigue damage accumulation caused by the damage evolution and gives the prediction method of fatigue life. The main research work and results are summarized as follows:

(1) The failure process of concrete beam subjected to the failure of the concrete beam with initial macro crack and without initial macroscopic crack is simulated. The results show that the concrete beam with initial macro crack has obvious brittleness. The damage evolution is mainly focused on the initial macroscopic crack tip, and the final main crack is obtained from the initial macroscopic crack propagation.

(2) 5 kinds of concrete beam specimens with different sizes were selected by using the optimized numerical analysis method, and each sample was made of 7 samples. The numerical experiments on the size effect of concrete members are carried out, and the mechanism of the size effect of concrete is discussed from the perspective of the evolution of damage. The results show that the nominal strength decreases with the increase of sample size. The fracture energy increases and the dispersion of the macroscopic mechanical properties decreases.

4. Conclusion

In this paper, the failure mechanism of concrete materials used in engineering is studied, and the characteristics of concrete damage and failure are analyzed. This paper mainly introduces the theory and method of concrete fracture mechanics analysis, the theory and method of concrete damage mechanics analysis. Based on the analysis of the existing methods of meso-numerical analysis of concrete, the meso-crack model is chosen as the research tool in this paper.

The failure process of concrete induced by the meso scale damage evolution is calculated and analyzed by using meso-crack model. Based on the analysis of the meso-characteristics of concrete materials, the concrete material is idealized and induced to some degree. Meanwhile, the meso-crack model of concrete is established and the reasonable model parameters are selected. The numerical simulation technique is used to simulate the failure process of concrete. Based on the verification of the effectiveness of the multi-crack model, the damage evolution process is quantita-

tively analyzed by using the damage mechanics analysis method. The relationship between macro and meso-damage variables of concrete material is discussed. The numerical simulation of uniaxial tensile tests of concrete specimens with different mixing ratios was carried out. The influence of damage on the macro mechanical properties of concrete was studied. It is proved that the linear elastic fracture mechanics theory can be used to analyze. In addition, the validity of the numerical analysis method and the meso-multi-crack model is verified.

References

[1] B. Sun, Z. Li: *Adaptive multi-grid FE simulation on dynamic damage and seismic failure of concrete structures.* Journal of Vibroengineering *18* (2016), No. 1, 276–288.

[2] R. Guo, S. Li, D. Zhou: *The three-stage model based on strain strength distribution for the tensile failure process of rock and concrete materials.* Acta Mechanica Solida Sinica *29* (2016), No. 5, 514–526.

[3] X. Gu, Q. Zhang, D. Huang, Y. Yv: *Wave dispersion analysis and simulation method for concrete SHPB test in peridynamics.* Engineering Fracture Mechanics *160* (2016), 124–137.

[4] C. Ma, W. Z. Chen, J. Y. Sun: *Numerical implementation of spatial elastoplastic damage model of concrete in the framework of isogeometric analysis approach.* Mathematical Problems in Engineering (2016), paper 4273024, 13 pages.

[5] F. Safaei, C. Castorena, Y. R. Kim: *Linking asphalt binder fatigue to asphalt mixture fatigue performance using viscoelastic continuum damage modeling.* Mechanics of Time-Dependent Materials *20* (2016), No. 3, 1–25.

[6] Z. Yu, S. Li, D. Lu, C. Lu, J. Liu: *FFailure mechanism of single-layer steel reticular domes with reinforced concrete substructure subjected to severe earthquakes.* International Journal of Steel Structures *20* (2016), No. 4, 1083–1094.

[7] M. Luković, B. Šavija, E. Schlangen, G. Ye, K. van Breugel: *A 3D lattice modelling study of drying shrinkage damage in concrete repair systems.* Materials (Basel) *9* (2014), No. 7, 575.

[8] C. Qian, Y. Nie, T. Cao: *Sulphate attack-induced damage and micro-mechanical properties of concrete characterized by nano-indentation coupled with X-ray computed tomography.* Structural Concrete *17* (2016), No. 1, 96–104.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 209–218

Pioneer robot motion control based on ZigBee wireless electronic communication technology

Zhili Wang[1], Suxiang Weng[1]

Abstract. In order to explore the multi-intelligent network collaborative control that involves the multi-intelligent control, network communication, computer technology application and so on, American pioneer robot is used as the intelligent body. Through ZigBee, the wireless communication network is set up. On the platform of host computer built by LabVIEW, the consistency of multi-intelligent biped robot network is studied. The research content is shown as follows: the application of discrete consistency control algorithm in multi-intelligent biped robot network collaborative control is achieved. The shortcomings of using theoretical value for collaborative control in the previous collaborative control are overcome. In addition, the application of sensors on the steering gear of robot is realized. The entire experiment process makes use of the actual value of motion angle of robot steering gear for the consistency experiments. Compared with the experiments using the theoretical value, it has high practical significance. Based on the requirements of multi-intelligent interaction mechanism, starting from human-computer interaction, LabVIEW is adopted to set up biped robot cooperative control host computer. What is more, the monitoring of robot collaborative control process is achieved, which well implements the requirements of interaction. The results showed that the robot consistency control evolution step is displayed accurately, which achieves the expected effect. At last, it is concluded that the biped robot display platform well meets the needs of multi-intelligent intelligence, achieving good effect on the control of multi-intelligent robot.

Key words. ZigBee, wireless communication, pioneer robot, cooperative control.

1. Introduction

In recent years, with the development of wireless communication technology, swarm robot technology and embedded system, the organic combination of the three, that is, wireless communication between swarm robots, has become the focus of today's research. With the rapid development of social production technology, the field of robot application is also being expanded. But in terms of the development of current level of robot technology, single robot is very limited in information ac-

[1] Jinlin Medical college, 132001, Jilin, China

quisition, processing, control ability and so on. For complex tasks and changing work environment, it faced difficult problems of communication and coordination, complex wiring and so on. A swarm system composed of many robots, through coordination and cooperation to accomplish the tasks that a single robot cannot or will not be able to accomplish, will be a trend in the development of robots. At present, multi-robot communication usually adopts wired communication. Although the technology is mature and reliable, it has not been widely applied because the cable has limited the range and flexibility of the robot.

ZigBee network technology is a wireless communication protocol based on IEEE 802.15.4. It has the characteristics of low cost, low power consumption and low transmission speed. It also supports many network topology structure, such as star, tree, mesh and so on [1]. Multi-robot collaboration only needs to transfer limited information intermittently and to save energy as much as possible. In order to solve the problem of multi-robot cooperation, the use of ZigBee module is proposed to build the hardware platform for group robot mutual communication, and the ZigBee node is introduced into the hardware system. A mesh network is constructed through the ZigBee module, which can achieve reliable point-to-point communication. At the same time, the network can reach the purpose of simple wiring and convenient collaboration, which extended the work space of group robots and improved the work efficiency.

2. State of the art

ZigBee technology is a new wireless network technology with short distance, low complexity, low power consumption, low data rate, and low cost [2]. It is between the wireless tag technology and Bluetooth technology, mainly used for short distance wireless connection, to complete the function of network and communication between nodes. Compared with other short distance communications, it has great advantages. This article makes use of ZigBee technology to complete the information transmission and networking of the intelligent robot system.

ZigBee is an IEEE802.15.4 standard based, and newly emerging wireless network technology with short range and low rate. It is a technical proposal ranged between Bluetooth technology and wireless tag technology [3], mainly used for short distance wireless connection and wireless data transmission. It has its own radio standard, which coordinates and communicates through thousands of tiny sensors. These sensors require only a small amount of energy to transfer data from one sensor to another by radio waves in relays, so it is highly efficient in communication [4]. Finally, these data can be entered into computers for analysis or to be collected by another wireless technology.

ZigBee can be viewed as a wireless data network platform consisting of up to 65000 wireless digital modules, similar to CDMA networks or GSM networks for mobile communications [5]. Among them, each ZigBee network data transmission module is similar to a base station of a mobile network, which can communicate with each other over the entire network within the entire network range [6]. Moreover, the entire ZigBee network can also be connected with other existing networks. The

transmission distance between each network node can be extended from the standard 75m to hundreds of meters or even thousands of meters.

For simple point to point, and point to multi point communication, packaging structure is relatively simple, mainly composed of synchronous preamble, data, and CRC check-sum several parts. ZigBee is to use the concept of data frame. Each wireless frame includes a large number of wireless packaging, which contains a lot of time, address, command, synchronization and so on information. The real data only occupies a small part, which is the key that ZigBee can realize the network organization management and implementation of reliable transmission [7]. At the same time, ZigBee uses MAC technology and DSSS (Direct Sequence Spread Spectrum) technology to achieve high reliability and large-scale network transmission.

From another point of view, the emergence of MANETs is inevitable. The wired communication mode limits the application scope greatly and reduces the flexibility greatly. What's more, wired communications have poor resistance to damage and it cannot work in some special cases, like earthquakes. As a result, it requires a network that can withstand strong damage, and communication continuity is guaranteed under any circumstances. The army, in particular, has a strong demand for it. The study origin of ad hoc network is similar to that of TCP/IP protocols. In 1970s, Defense Advanced Research Projects Agency DARPA funded the new network architecture model using packet radio network for data communication in the battlefield environment in PRNET [8]. Later, in 1983 and 1994, it funded the SURAN (Survivable Adaptive Network) and the GloMo (Global Mobile Information Systems) project research. Ad hoc, which is called ad hoc network, absorbs the above networking ideas and ensures the rapid networking and super invulnerability of the army in the battlefield environment [9]. Similar to TCP/IP, ad hoc not only has wide application in the military field, but also lays a foundation for civil communication services, which has become the research focus of the industry. At present, there are as many as 70 kinds of routing protocols of ad hoc, which fully illustrates this point. For the network application of the previously described special environment (such as natural disasters, scientific investigation, exploration, battlefield environment and so on), these kinds of networks are powerless to do so. To communicate and implement the dynamic process of cooperative communication in special environment requires the network that has dynamic and rapid deployment, and does not rely on or rarely depends on the existing wired network, which is the mobile Internet [10].

Short duration. The delay sensitive applications are optimized, and the communication delay and the delay activated from the sleep state are very short [11]. ZigBee's response rate is faster, which generally just takes 15 ms transferring from sleep to work. The connection for node into the network only takes 30 ms, which further saves the electricity. By comparison, Bluetooth requires 3∼10 s, and WiFi requires 3 s.

3. Method

3.1. ZigBee wireless communication module

Similar to the multi-agent collaborative consistency, the information exchange between nodes is accomplished through communication. In the researches on motion pioneer robot control, the information exchange between nodes plays a very important role. The ZigBee wireless communication system is adopted in the experiment. As a newly developed wireless communication technology, ZigBee, in the network communication with short distance and low rate requirements, becomes more and more popular, and it is widely used in practical engineering. ZigBee has good network capacity, and a large number of nodes can be extended to realize wireless communication between multiple nodes. Secondly, with license free frequency, low cost, high efficiency, convenient and fast communication system, it can be used repeatedly. The ZigBee wireless communication system adopted in this experiment is the ZigBee wireless communication system development kit produced by Chengdu Wireless Dragon Communication Technology Co.Ltd. It is an economic, efficient, convenient and fast development kit. The system supports the USB high-speed download, and supports IAR integrated development environment, with online download, debug and simulation function, and it provides ZigBee protocol source code. In this experiment, we use ZigBee wireless communication system development kit, which is developed by Wireless Dragon Company and used for communication and location. The system uses the IAR development environment, which can download, compile and debug the communication code, and provide ZigBee protocol code by Wireless Dragon Company, which supports downloading the source program through USB.

The ZigBee stack structure consists of a set of modules called layers. Each layer performs a specific set of services for the above layer: the data entity provides the data transfer service, and the management entity provides all the other services. Each service entity provides an interface for the upper layer through a service access point (SAP), and each SAP supports a variety of service primitives to implement the required functionality. The ZigBee stack is built on the basis of the IEEE802.15.4 standard, and the protocol stack structure is described in Fig. 1. It is based on the standard open system interconnection (OSI) seven layers model, but defines only the layers that involve ZigBee. The IEEE802.15.4-2003 standard defines two lower layers: the physical layer (PHY) and the media access control (MAC) sublayer. The ZigBee alliance builds the network layer (NWK) and the application layer architecture on this basis [12].

3.2. Wireless data transmission of robot

The core part of this paper is to apply the ZigBee wireless communication system in intelligent robot system. First of all, the hardware design scheme of communication platform is designed, which designs and implements the connection circuit. Through the data communication of motor angle value, pioneer robot motion control system is achieved. In addition to positioning, one of the essential role of ZigBee

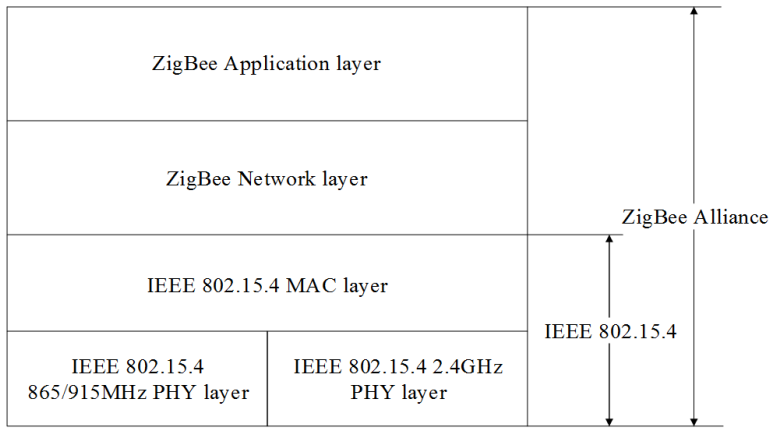


Fig. 1. ZigBee structural system diagram

system is the networking and communications. Firstly, the coordinator establishes the network, and then the coordinator, the router and the terminal node communicate with each other to realize the sending and receiving of the data information. Three basic and typical network topology structures, namely star networks, cascade networks and mesh networks, are introduced previously. In order to better study the network characteristics of ZigBee wireless communication system, we design several different forms of networks according to the network topology in the three forms, to verify if all the nodes run within the network can realize the communication connectivity, reliability and so on.

In order to achieve communication between the robots, a series of networking and communication experiments are done with ZigBee. Two biped robots are used as multi-intelligent individuals, labeled as A robot and B robot. This experiment adopts A and B two robots' right arm for collaborative control experiment. Each robot, as the communication node, is added to ZigBee wireless communication network. Two robots communicate with each other to obtain the state of right arm of another robot, and communicate its own angle value to each other. Then, the next step action is calculated through the consistency control protocol, and ultimately the right consistency control is achieved. And then, through the host computer interface monitoring and processing, information of each robot arm is reacted to the host computer. Because the information receiving of robot is real-time, the monitoring process is real-time and dynamic. The design diagram of the wireless communication network in this paper is shown in Fig. 2.

4. Results and discussion

This experiment is a two-order network topology model experiment. The information receiving in the process of experiments is based on sequential cooperative control. In the process of experiment, A robot arm information is displayed through

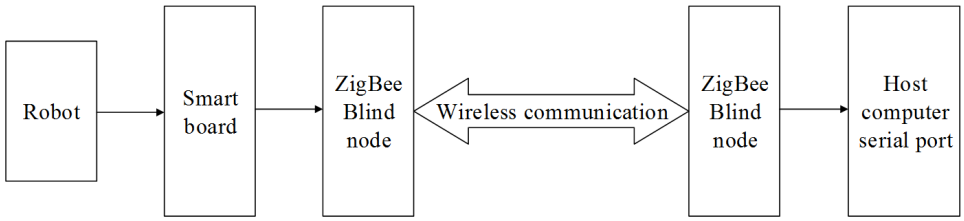


Fig. 2. Wireless communication network block diagram

the control components. The horizontal axis for displaying control is selected as the number of steps of the coordination and control, and the vertical axis is the robot steering angle information. The angle information is adjusted according to the actual size. The display control is shown in Fig. 3.

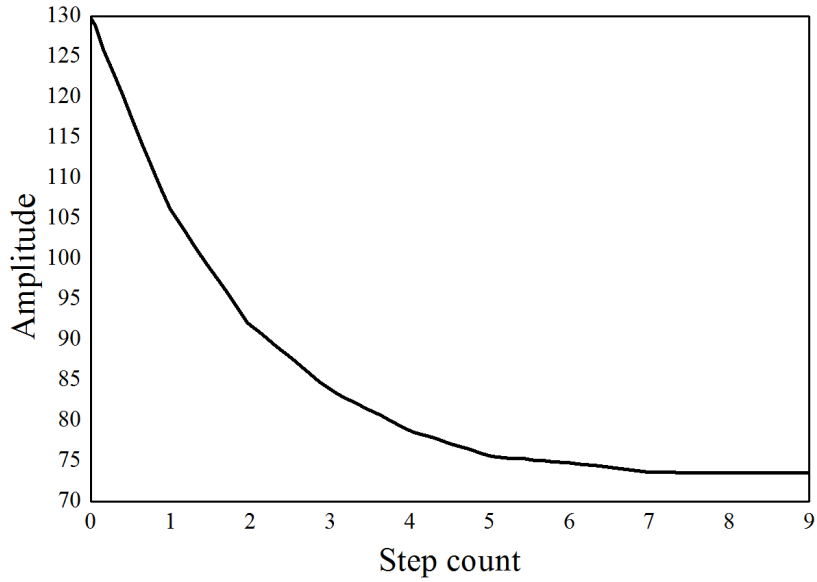


Fig. 3. Robot A right arm angle value

In the experiment, the information of the B robot is displayed by the corresponding display control components, and the angle value of the steering wheel of the B right arm is displayed, as shown in Fig. 4.

In the experiment, through the design of the program, the angle difference between the two servos in the process of consistency cooperation can be displayed, which can reflect the evolution process of cooperative control to some extent, as shown in Fig. 5.

The following analysis is made for the three curve graphs:

In the two robots monitoring experiment, steering angle can reach agreement, but

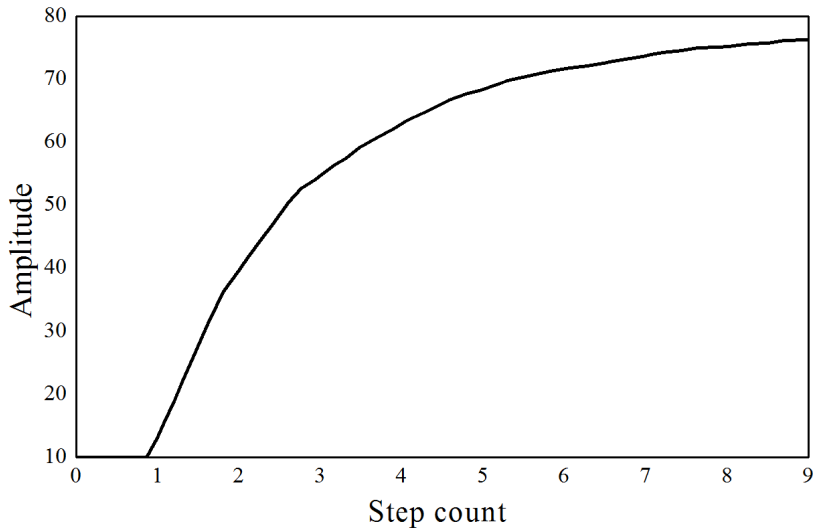


Fig. 4. Robot B right arm angle value

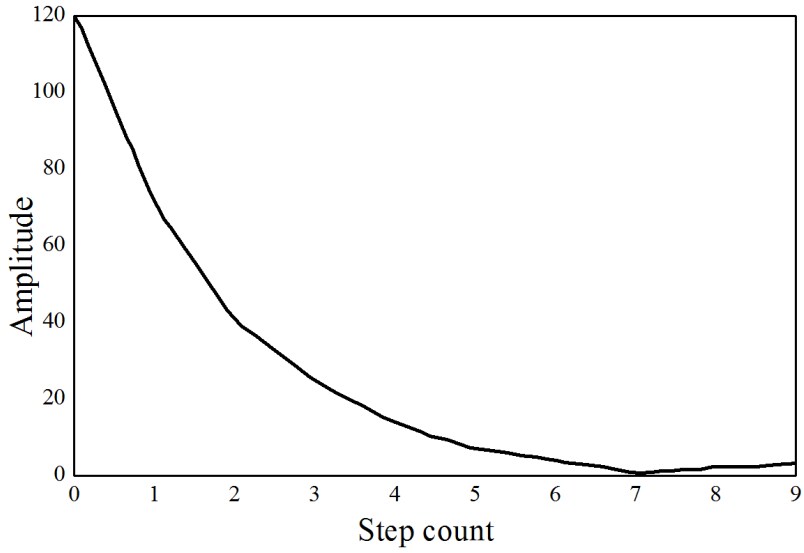


Fig. 5. Robot angle difference value

there is still a gap with the experimental theoretical value, and in the calculation of the theoretical value, two different steering gears can finally make intermediate values. It can be seen from the experimental results that the angle of motion of the steering gear in the initial experiment is relatively large, and then the change value becomes smaller, and finally the consistency is achieved. From the arm monitoring diagram of robots A and B, because the horizontal coordinate is the number of steps

needed by the collaborative evolution, and the time for performing a step basically is certain, so it is able to see the time that the experiment needs. Through the data in the upper computer, we can conclude that the experiment can indeed monitor the actual value of the robot rudder, and it has very good practical significance.

Through the experimental data, several instructions are made for the experiment. Based on the actual value of the angle, the cooperative control experiment is made. The robot steering angle value only makes feedback of integer data, so it is calculated by integer data. But in the calculation of the theoretical value, when the sensor is failed in reading, it will still take the theoretical value. In the selection of control conditions, the previous experiment is based on the theoretical value to be out of the program when the difference of two angles is less than 0.2 degrees. But in the process of experiment, due to the application of actual value, through the instructions, the robot's motion angle and actual movement angle has 2 degrees or so error. In the iterative process, with the accumulation of errors, the cooperative control evolution process is not as before, having relatively fixed consistency angle value, thus the conditions for ending the cycle will change.

5. Conclusion

This paper mainly studies the multi-agent system based on ZigBee wireless communication module, which has far-reaching theoretical and practical significance. It focuses on the ZigBee technology which has a huge advantage in the short distance communication technology. The protocol stack structure and network characteristics are studied. First of all, a wireless communication between nodes under several different network ways is designed and achieved. This is the important foundation of its application in multi-intelligent robot system. In this paper, the software and hardware platform of the wireless communication system is designed and implemented. Under the Z-Stack protocol stack software platform, a communication platform for communication is constructed, which successfully completes the functional requirements of cooperative control in the multi-robot motion process on communication functions. The experimental results basically verify the consistency of the system, and for the application of ZigBee, a new idea is broaden. The work done in this paper can be divided into the following three aspects:

Firstly, the multi-intelligent system development and research content object are studied. The multi-robot system is led out, the communication requirements of the system are analyzed, and the communication scheme of the robot system based on ZigBee wireless network is designed. For the popular ZigBee technology, the full range and multi-angle study is carried out. And it is compared with several other popular short distance communication technologies. In addition, its characteristics advantages are analyzed, and the network technology is especially discussed. Four kinds of networking modes are designed and achieved, and the communication between nodes is completed. An in-depth study is conducted for the ZigBee protocol stack, and the hierarchical structure and data format are explored in detail. For the communication needs, it is developed in the application layer platform. What is more, the hardware platform of wireless communication system is set up, the de-

sign, implementation, debugging and analysis are performed, and the consistency of multi-agent action robot system is verified.

References

[1] Ill-Woo Park, Jung-Yup Kim, Jungho Lee, Jun-Ho Oh: *Mechanical design of the humanoid robot platform, HUBO.* Advanced Robotics *21* (2007), No. 11, 1305 to 1322.

[2] M. Bennewitz, W. Burgard, S. Thrun: *Optimizing schedules for prioritized path planning of multi-robot systems.* Proc. IEEE International Conference on Robotics and Automation (ICRA), 21–26 May 2001, COEX Seoul, Korea, *1*,271–276.

[3] T. Simeon, S. Leroy, J. P. Lauumond: *Path coordination for multiple robots: a resolution complete algorithm.* IEEE Transactions on Robotics and Automation *18* (2002), No. 1, 42–49.

[4] W. Ren, R. W. Beard: *Consensus seeking in multi-agent systems under dynamically changing interaction topologies.* Transactions Automatic Control *50* (2005), No. 5, 655–661.

[5] S. Liu, L. Xie, H. Zhang: *Distributed consensus for multi-agent systems with delays and noises in transmission channels.* Automatica *47* (2011), No. 5, 920–934.

[6] J. Xu, Z. Wu: *A discrete consensus support model for multiple attribute group decision making.* Knowledge-Based Systems *24* (2011), No. 8, 1196–1202.

[7] R. Cavalcante, B. Mulgrew: *Adaptive filter algorithms for accelerated discrete-time consensus.* IEEE Transactions on Signal Processing *50* (2009), No. 3, 1049–1058.

[8] Z. Li, J. J. Lee: *Global synchronization of coupled map lattices.* Proc. IEEE International Conference on Mechatronics and Automation (ICMA), 5–8 August 2007, Harbin, China.

[9] R. Olfati-Saber, R. M. Murray: *Consensus problems in networks of agents with switching topology and time-delays.* IEEE Transactions on Automatic Control *49* (2004), No. 9, 1520–1533.

[10] Z. Li, G. R. Chen: *Global synchronization and asymptotic stability of complex dynamical networks.* IEEE Transactions on Circuits and Systems II *53*, (2006), No. 1, 28–33.

[11] K. Liu, G. Yang, F. Cui, J. Ma, B. Cheng, Y. Wang: *Exploration of innovation in teaching single-chip microcomputer with practice running through the course.* Journal of Education in New Century *1* (2013), No. 1, 9.

[12] H. Wu, S. Cheng, Y. Peng, K. Long, J. Ma: *IEEE 802.11 distributed coordination function (DCF): analysis and enhancement.* Proc. IEEE International Conference on Communications. Conference Proceedings (ICC), 28 April–2 May 2002, New York, NY, USA *1*, 605–609.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 219–228

Optimal path planning of robot based on ant colony algorithm[1]

Jieyu Chen[2]

Abstract. In order to develop a scheme for optimal path calculation of robots, the optimal path planning of robot based on ant colony algorithm is proposed. The path planning of mobile robot is one of the core contents of robot research. It has the characteristics of complexity, constraint and nonlinearity. Ant colony algorithm (ACA) is a bionic optimization algorithm developed in the last ten years. The algorithm has shown its excellent performance and huge development potential in solving many complex problems. On the basis of ant colony algorithm and genetic algorithm (GA), the GA-ACA algorithm and ACA-GA algorithm are proposed and applied to robot path planning. Based on the MATLAB7.5 software development environment, a mobile robot path planning simulation system based on ant colony algorithm is designed. The simulation results verify the effectiveness of the proposed algorithm. After the simulation calculation, it is shown that the GA-ACA and ACA-GA algorithms have better comprehensive performance than ant colony algorithm. Based on the above finding, it is concluded that the optimal path planning of robot based on ant colony algorithm is suitable for the control and tracking of robot during its operation.

Key words. Path planning, ant colony algorithm, genetic algorithm, GA-ACA algorithm, ACA-GA algorithm.

1. Introduction

With the development of computer technology, control theory, artificial intelligence theory and sensor technology, the research of robot has developed to a new stage [1]. Among them, mobile robot, as an important branch, has been paid more and more attention in the field of research both at home and abroad. In recent years, mobile robot technology has played an important role in many fields such as industry, agriculture, aeronautics and space exploration. Because of its broad application prospects, mobile robot technology has become a hot issue at home and abroad. Path planning is the security guarantee for mobile robot to accomplish the task, and it is also an important sign of the intelligent degree of mobile robot. The

[1]The author acknowledges the National Natural Science Foundation of China (Grant: 51578109), the National Natural Science Foundation of China (Grant: 51121005).

[2]Jiaxing Radio & TV University, Jiaxing, Zhejiang, 314001, China

improvement of the path planning algorithm will change the navigation performance of the mobile robot and improve the intelligence level of the mobile robot [2]. At the same time, it will reduce the uncertainty of mobile robot in the course of movement, increase the flexibility of mobile robot movement, and provide a solid foundation for the development of high intelligent mobile robot. The path planning of mobile robot is one of the most important problems in the research of robot control system, and it is also one of the core contents in the field of robot research. The purpose of path planning is to hope that future mobile robots have high-level capabilities such as perception, planning and control. So that it can collect information from the surrounding environment, build a model about the environment, and use this model to plan and execute high-level tasks.

2. Literature review

At home and abroad, the path planning research can be divided into two methods: traditional methods and intelligent methods. Among them, traditional methods include visual graph, free space method, grid method, artificial potential field method and topological method. The intelligent methods include genetic algorithm, fuzzy logic method, neural network method and ant colony algorithm. Ant colony algorithm belongs to a new algorithm. Ant colony algorithm uses distributed parallel computing mechanism, which can effectively find the global optimal solution [3]. It has strong robustness and adaptability, and is easy to combine with other algorithms. However, ant colony algorithm has some drawbacks. Although compared with genetic algorithm, the search time of ant colony algorithm has been obviously shortened, it is still not ideal enough. On the other hand, when the path planning problem is large, the ant colony algorithm tends to stall or sink into the local optimal solution. In order to overcome the shortcomings of the two algorithms and form complementary advantages, genetic algorithm and ant colony algorithm can be combined. Two improved algorithms are proposed based on ant colony algorithm combined with genetic algorithm, which are GA-ACA algorithm and ACA-GA algorithm [4]. And these new algorithms are applied to robot path planning. Based on the MATLAB7.5 software development environment, a mobile robot path planning simulation system based on ant colony algorithm is designed. The simulation results show that the GA-ACA and ACA-GA algorithms have better comprehensive performance than ant colony algorithm.

3. Methods

3.1. Robot path planning based on ant colony algorithm

The establishment of environment model is a very important part of robot path planning. The actual working environment of a robot is a realistic physical space, and the space that the path planning algorithm handles is the abstract space of the environment. Environment modeling is a mapping from physical space to abstract

space. The environment model is built by grid method, and the actual working space of robot is simulated. A grid is used to represent the environment map of the robot's work, and complex computation can be avoided when dealing with the boundaries of obstacles. In the application of grid method, the most important part is the partition of grid granularity. The smaller the grid size is, the more accurate the obstacle will be. But at the same time, it takes up a lot of storage space, and the search scope of the algorithm increases exponentially. Grid size is too large, and the planning path will be very inaccurate. The grid granularity is $1\times$l.

In the grid environment with n of total grid, the steps of ant colony algorithm for robot path planning are as follows:

Step 1: Set initial parameters and initialize ant colonies.

Step 2: An ant is transferred to the next grid.

Step 3: The step 2 is repeated until all dead ants complete the selection of the subsequent grid.

Step 4: Local pheromones are updated.

Step 5: Step 2, step 3, and step 4 are repeated until all ants are moved to the target grid.

Step 6: The shortest path, the length of the shortest path, and the average length of the path taken by all ants in the cycle need to be calculated.

Step 7: The global pheromone is updated.

Step 8: The tabu table is cleared, the number of cycles is $Nc = Nc + 1$, if $Nc = Nc_{\max}$ was transferred to step 2; If $Nc > Nc_{\max}$, jump out the entire loop, output the optimal path and the optimal path length.

3.2. Robot path planning based on genetic algorithm

Genetic algorithm (GA) is a global optimization algorithm based on natural inheritance and natural selection proposed by Professor J. Holland from America in 1975. It has implicit parallelism, good global optimization ability, and strong robustness and flexibility. The main operations of genetic algorithm are selection, crossover and mutation. The core contents are parameter coding, generation of initial population, design of fitness function, design of genetic operators and setting of control parameters. The concrete steps are as follows:

Step 1: The solution space of the problem is encoded.

Step 2: The initial population is randomly generated and the fitness function of all individuals in the population is evaluated.

Step 3: Determining whether the convergence criteria of genetic algorithms are satisfied. If satisfied, output search results; otherwise, continue with the following steps.

Step 4: According to the size of the fitness function, copy operations are performed in a certain way.

Step 5: Cross operation is performed according to cross probability Pc.

Step 6: According to the mutation probability Pm, the mutation operation is performed.

Step 7: Go to step 2.

3.3. Robot path planning based on GA-ACA algorithm

In order to overcome the shortcomings of ant colony algorithm and genetic algorithm, the genetic algorithm and ant colony algorithm are fused to complement its advantages. Firstly, the initial pheromone distribution of the problem is generated by the stochastic search, rapidity and global convergence of the genetic algorithm. Then, the parallelism of the ant colony algorithm, the positive feedback mechanism and the high efficiency of the algorithm are fully utilized to solve the problem. The algorithm is superior to genetic algorithm in solving efficiency, and is superior to ant colony algorithm in time efficiency. A heuristic algorithm for solving both efficiency and time efficiency is proposed. The algorithm, which uses genetic algorithm to generate the initial pheromone distribution, and then uses ant colony algorithm to solve the problem, is called GA-ACA algorithm. The concrete steps are as follows:

Step 1: The grid sequence number in the grid environment model is used as the path encoding, that is, the path individuals are represented by the grid sequence number. In addition, the barrier, the grid number and the repeated grid serial number are not allowed in the sequence. The initial path population is generated randomly, and the fitness function related to path length is chosen.

Step 2: The fitness function of the path individuals in the path population is calculated, and the individual path of crossover and mutation is selected by roulette wheel according to fitness function.

Step 3: Cross probability Pc is used to perform crossover operations on path individuals. The concrete method is to randomly generate a random number $Rand$ between 0 and 1. If $Rand < Pc$, cross operation is performed, otherwise, the operation is not performed [5].

Step 4: The mutation is performed for the path individual according to the mutation probability Pm. The concrete method is to generate a random number $Rand$ between 0 and 1 randomly. If $Rand < Pm$, the mutation operation is performed, otherwise it will not be executed.

Step 5: The step 2 to step 4 is repeated until a set of convergence conditions is reached or the set cycle times are generated, and several groups of optimized path individuals are generated.

Step 6: According to the optimized path generated by step 5, the initial distribution of pheromone is formed, and the initial parameters of ant colony algorithm are set up. All ants are placed in the initial grid of the path planning, and the initial grid is added to the ant tabu list.

Step 7: Each ant chooses the next grid according to the state transition rule and the next grid is added to the ant's tabu list.

Step 8: The step 7 is repeated until the ant constructs a path to update the pheromone locally.

Step 9: Step 7 and step 8 are executed repeatedly until all ants build a good path to update the pheromone globally.

Step 10: The tabu list of all ants is cleared, and step 7 to step 9 are repeated until the desired cycle number is reached or certain termination conditions are met.

Step 11: The optimal path is output.

3.4. Robot path planning based on ACA-GA algorithm

The operation of ACA-GA algorithm is to select two paths of ants randomly after the ant colony completes a cycle. The two paths are crossed according to a certain cross probability. If the crossover produces a path that is superior to the optimal path of the iteration, the pheromone is released on the better path. The crossover operation of genetic algorithm is introduced into ant colony algorithm, which can increase the diversity of solution and accelerate the speed of problem solving. The concrete steps are as follows:

Step 1: The initial parameter is set and the ant colony is initialized. All ants are placed in the initial grid of the path planning, and the initial grid is added to the tabu list of all ants.

Step 2: Each ant chooses the next grid to move according to the state transition rule, and the next grid is added to the tabu list of the ants.

Step 3: The step 2 is repeated until the ant constructs a path to update the pheromone locally.

Step 4: Step 2 and step 3 are executed repeatedly until all ants build a good path.

Step 5: The optimal path in the iteration is selected, and then another path in the iteration is selected randomly to perform cross operation according to the given cross probability.

Step 6: The pheromone is updated globally and the tabu list of all ants is removed. Step 2 to step 5 is repeated until the desired number of cycles is set or certain termination conditions are met [6].

Step 7: The optimal path is output.

4. Results and discussion

In order to verify the ACA-GA algorithm, the genetic algorithm, ant colony algorithm and ACA-GA algorithm are simulated, respectively. Figures 1, 2 and 3 are the optimal path evolution maps searched by genetic algorithm, ant colony algorithm and ACA-GA algorithm, respectively [7].

From the results depicted in Figs. 1, 2, 3 and simulation, it is shown that the genetic algorithm finds the sub-optimal path after 5 iterations and converges to the sub-optimal path after 20 iterations [8]. The ant colony algorithm finds the optimal path after 6 iterations, and converges to the optimal path after 30 iterations. The ACA-GA algorithm finds the optimal path after 3 iterations and converges to the optimal path after 15 iterations. Compared with the three algorithms, the ACA-GA algorithm can converge to the optimal path in the least number of iterations.

In order to verify the GA-ACA algorithm, the genetic algorithm, ant colony algorithm and ACA-GA algorithm are simulated, respectively. Figures 4, 5 and 6 are the optimal path evolution maps searched by genetic algorithm, ant colony algorithm and GA-ACA algorithm respectively [9].

From the results in Figs. 4, 5, 6 and simulation, it is shown that the genetic algorithm finds the sub optimal path after 3 iterations and converges to the suboptimal

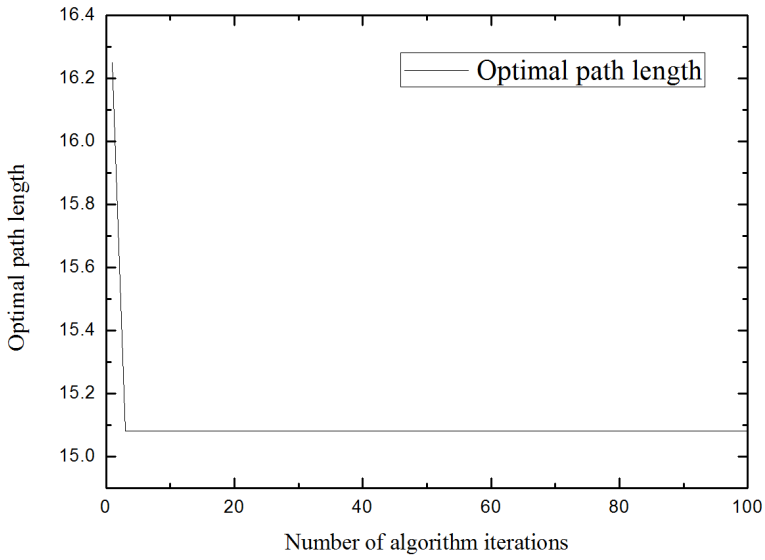


Fig. 1. Optimal path evolution of genetic algorithm

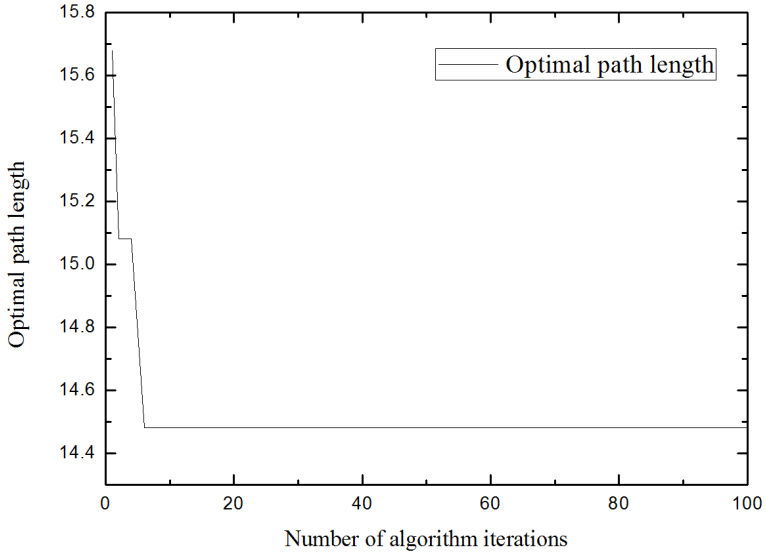


Fig. 2. Optimal path evolution of ant colony algorithm

path after 60 iterations. The ant colony algorithm finds the optimal path after 5 iterations, and converges to the optimal path after 50 iterations. The GA-ACA algorithm finds the optimal path after 3 iterations and converges to the optimal path after 10 iterations. Compared with the three algorithms, the GA-ACA algorithm can converge to the optimal path in the least number of iterations [10].

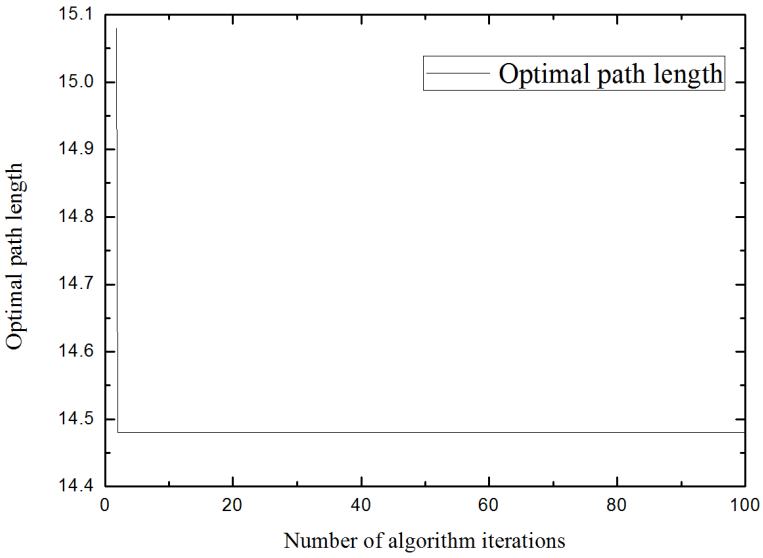


Fig. 3. Optimal path evolution of ACA-GA algorithm

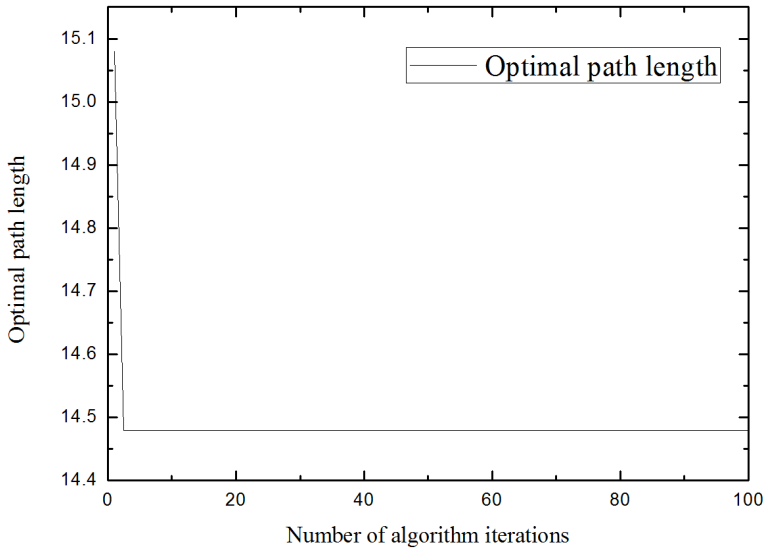


Fig. 4. Optimal path evolution of genetic algorithm

5. Conclusion

On the basis of ant colony algorithm and genetic algorithm, two improved algorithms are proposed, which are GA-ACA algorithm and ACA-GA algorithm, and they are applied to robot path planning.

GA-ACA algorithm uses genetic algorithm to generate the initial pheromone

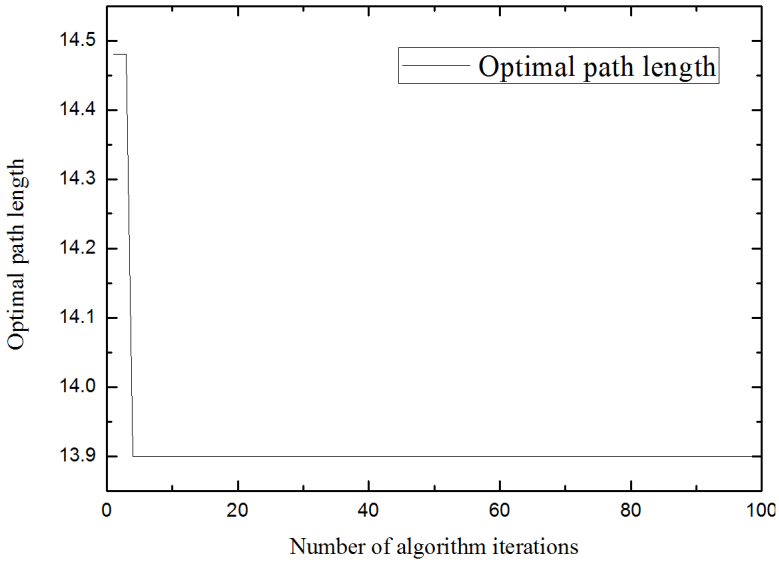


Fig. 5. Optimal path evolution of ant colony algorithm

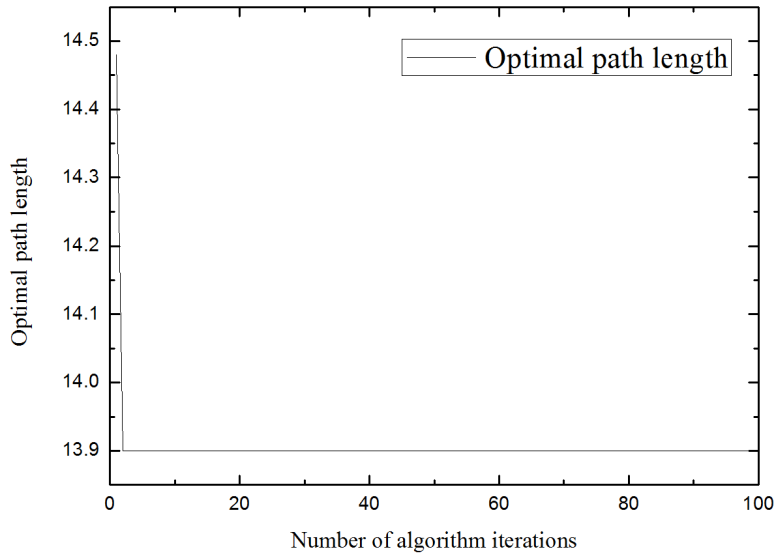


Fig. 6. Optimal path evolution of GA-ACA algorithm

distribution, and then uses ant colony algorithm to solve the problem, that is to say, different algorithms are adopted in different stages. In the initial stage of solving some problems, ant colony algorithm lacks the effective guidance of pheromone, which leads to the same algorithm as greedy algorithm, and the convergence speed is not fast. Based on the GA-ACA algorithm, the ant colony algorithm is effectively

guided by pheromone at the beginning of the solution, and can quickly converge to the optimal solution. Simulation results show that the performance of GA-ACA algorithm is better than ant colony algorithm.

ACA-GA algorithm introduces the crossover idea of genetic algorithm into the ant colony algorithm. After the ant colony completes each cycle, the path of the optimal ant is chosen. Then, another path is chosen randomly. The two paths are crossed, and if the crossover produces a better path than the best path in this cycle, the pheromone on the better path is updated. The ACA-GA algorithm increases the diversity of knowledge. Simulation results show that the performance of ACA-GA algorithm is better than ant colony algorithm.

References

[1] R. Wang, H. Tiang: *Two-dimension path planning method based on improved ant colony algorithm.* Advances in Pure Mathematics *5* (2015), No. 9, 571–578.

[2] Y. Dai, M. Zhao: *Manipulator path-planning avoiding obstacle based on screw theory and ant colony algorithm.* Journal of Computational & Theoretical Nanoscience *13* (2016), No. 1, 922–927.

[3] M. K. Gao, Y. M. Chen, Q. Liu, C. Huang, Z. Y. Li, D. H. Zhang: *Three-dimensional path planning and guidance of leg vascular based on improved ant colony algorithm in augmented reality.* Journal of Medical Systems *39* (2015), No. 11, 133.

[4] J. Tian, M. Gao, G. Ge: *Wireless sensor network node optimal coverage based on improved genetic algorithm and binary ant colony algorithm.* EURASIP Journal on Wireless Communications and Networking (2016), 104.

[5] D. Zhou, L. Wang, Q. Zhang: *Obstacle avoidance planning of space manipulator end-effector based on improved ant colony algorithm.* SpringerPlus *5* (2016), No. 1, 509.

[6] R. Mao, X. Ma: *Research on path planning method of coal mine robot to avoid obstacle in gas distribution area.* Journal of Robotics (2016), No. 12, paper 4212076.

[7] W. Li, M. Sun, C. Cao, M. Sun, H. Shen, J. Zhu: *Ant colony algorithms based on graph knowledge transfer in geometric constraint solving.* Journal of Information and Computational Science *12* (2015), No. 15, 5631–5639.

[8] B. Xu, H. Min: *Solving minimum constraint removal (MCR) problem using a social-force-model-based ant colony algorithm.* Applied Soft Computing *43* (2016), No. C, 553–560.

[9] J. F. Wang, X. Wu, X. Fan: *A two-stage ant colony optimization approach based on a directed graph for process planning.* International Journal of Advanced Manufacturing Technology *80* (2015), Nos. 5–8, 839–850.

[10] K. V. J. Raam, K. Rajkumar: *A novel approach using parallel ant colony optimization algorithm for detecting routing path based on cluster head in wireless sensor network.* Indian Journal of Science & Technology *8*, (2015), No. 16, 452–463.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 229–236

Optimal allocation of structural sensor in civil engineering based on simulated annealing genetic algorithm

Tian Jiale[1]

Abstract. Aiming at the problem of combinatorial optimization of sensor optimal configuration, this paper presents an improved adaptive simulated annealing genetic algorithms based on Modal Assurance Criterion (MAC). Taking the maximum off-diagonal element of the MAC matrix as a target function and the sensor number is not change as a constraint condition, a dualistic coding genetic algorithm is proposed, and after improve the traditional simulation annealing algorithms, use it as an independent operator of GA. Therefore, adaptive crossover and mutation probabilities are adopted to prevent premature convergence. The results of research show that the hybrid algorithm optimizes the number and position of the sensor at the same time, and obtains the optimal configuration of the sensor to meet different accuracy requirements.

Key words. Optimal sensor placement, modal assurance criterion (MAC), dualistic coding, simulated annealing genetic algorithm.

1. Introduction

Structural health monitoring technology [1], using the integrated sensor network structure in real-time monitoring structure of the environmental excitation (human or natural) response signal, and from the structural health status-related information, combined with advanced signal processing methods, extract structural damage parameters, determine whether the structure is damaged and the location and extent of damage. The most important thing of optimal sensor placement is to select the appropriate optimization method [2]. Traditional optimization algorithms mainly include: effective independent method, kinetic energy method, Guyan model reduction method, these methods have their own limitations. In recent years, some new intelligent optimization methods have been developed, mainly simulated annealing algorithm, genetic algorithm and neural network algorithm, these new algorithms to solve complex problems provides a new way of thinking and means. Genetic algorithm (GA) has a good global search capabilities, it is currently the most widely

[1]College of City Construction, Jiangxi Normal University, 330022, China

used method [3]. However, the practice result shows that genetic algorithm has some shortcomings, the most important problem is that it can bring easily premature convergence and local search capability is bad. Another widely used algorithm is the simulated annealing algorithm (SA) [4]. SA algorithm is a heuristic random optimization algorithm, because the algorithm uses Metropolis probability acceptance criteria, it has a strong local search capabilities to avoid falling into the local optimal solution. But the algorithm has the problem of bad global search ability, slow convergence rate and low efficiency. In addition, the existing sensor optimal configuration method, most of them are directly given the number of sensors, how to determine the optimal design to meet the number of sensors is still a stubborn problem.

In this paper, an adaptive genetic algorithm based on simulated annealing is proposed to optimize the number and position of the sensor, in order to optimize the configuration of the sensors with different precision requirements. The basic idea is to combine the simulated annealing algorithm (SA) with strong local search ability and genetic algorithm (GA) to learn from each other and enhance the local search capability of GA [5]. At the same time, in order to avoid premature convergence, adaptive crossover and mutation probabilities are introduced. And the feasibility of the algorithm is verified by a numerical

2. Literature review

Civil engineering is closely related to the life of the country and the people. Along with the rapid development of economy, the civil engineering structures such as bridges, dams, high-rise buildings and ocean platforms have increased dramatically. Their service period is generally several decades or even hundreds of years. At the same time they are affected by long-term environmental erosion, material aging, the long-term effects of load, fatigue and other adverse factors coupling effect, this will result in structural damage and structural damage accumulation and resistance, coupled with earthquakes, storms, floods and other natural disasters, many factors lead to different degrees of structural damage, serious may lead to catastrophic accidents, endangering the state and the people's lives and property, resulting in adverse social impact [6].

In recent years, civil engineering construction in our country is rapidly rising, the changing high-rise buildings, exhibition centers, railways, highways, bridges, ports and other major livelihood projects across the country have sprung up. However, many large civil engineering structures due to large size, complex forces, lack of monitoring and other reasons causes frequent problems [7]. Civil engineering accident not only caused huge economic losses, but also seriously threatened the people's life safety and affected the economic and social development and stability.

If civil engineering structures can be predicted and evaluated before an accident or disaster occurs, it is important to take early measures to prevent and mitigate the losses caused by the accident. With the rapid development of modern science and technology, in order to ensure the safety, applicability and durability of various engineering structures, people should know the running state of the structure in

real time and carry out structural health monitoring. Health monitoring of civil engineering structures is inseparable from the sensor layout, the use of different types of sensors to collect engineering structure of the various data through software analysis and processing to get the relevant health information to identify the location of the injury, so that people take measures to do the appropriate preparation [8]. However, civil engineering structures tend to be large in size, numerous in number of nodes, taking into account the economic and structural operation of the state, the layout of the sensor is not the more the better, people should use a limited number of sensors to obtain information about the structure of the information, which can reflect the whole and local state change of the structure. That is to say, the optimal placement of sensors has become one of the key problems in the structural health monitoring system. This problem determines whether the system can truly and accurately obtain information on the whole structure and the partial operation state.

It is the key problem of the optimal layout of the sensors to arrange as few sensors as possible to obtain as much of the civil engineering structure of health information, and also the important issue of structural health monitoring to be resolved first. To sum up, in the civil engineering structure to carry on the optimal placement of the sensor has great application value, and also has profound social significance.

3. Research contents and methods

Sensor optimization problem is a special kind of knapsack problem—the given sensor configuration in the optimal location, amd the mathematical model is actually a 0–1 planning problem. If the ith gene code is 1, the sensor is arranged on the ith degree of freedom, if the th gene code is 0, the sensor is not arranged at this degree of freedom. Assuming that the number of sensors is m, and the number of candidate points is n, the mathematical model can be represented by the following equation, the key problem is how to express x_i such that the objective function X is maximized. The function is subject to

$$\sum_{i=1}^{n}(x_i - m), \quad x_i = 0 \text{ or } 1, \ i = 1, 2, \cdots, n\,. \tag{1}$$

Simulated annealing algorithm, as one of many stochastic optimization algorithms, simulates the high temperature objects with large internal energy. The simulated process is the gradual annealing process of the particles inside the material, including the annealing at different temperatures point to equilibrium state until the normal temperature of the ground state of the process, when the particles inside the material reaches the ground state, the internal energy is in the minimum value. During the annealing process, the energy of the system is subject to Boltzmann distribution, that is

$$P(f) = \exp\left(-\frac{f}{kT}\right)\,. \tag{2}$$

First of all, according to the above formula to build energy function f, the optimal solution is found by using Metropolis sampling and an annealing process that tends to be orderly. Here, the energy function f refers to the non-diagonal element mean value of the MAC of the mode confidence matrix, that is

$$f = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} MAC_{ij}}{n(n-1)} . \tag{3}$$

In the formula, $\mathrm{MAC}_{ij},\ i \neq j$ is called the non-diagonal element of the modal confidence matrix MAC, n is the selected modal number, and the important goal of the sensor optimization arrangement is to ensure that the value of f is minimized. Considering that the asymptotic convergence of the simulated annealing algorithm needs to be realized with certain conditions, a set of cooling schedule parameters should be set as the termination condition for the iterative calculation using MATLAB programming. The specific parameters are as follows:

Initial temperature t_0: in order to prevent the simulated annealing process from evolving into a local stochastic search process, it is necessary to ensure that the quasi-equilibrium is achieved in the initial stage of operation, a sufficiently large initial temperature t_0 is selected according to the Metropolis criterion. The attenuation function of the control parameter: in general, select $t_k + 1 = \alpha t_k$, where α is called the attenuation coefficient, often take 0.5∼0.99.

Iteration termination criterion: according to the solution obtained by the algorithm in each operation stage, we can see whether the convergence quality of the existing solution has a big improvement, that is to say, the algorithm stops when the solutions of some adjacent Mopkob chains are not significantly improved. It is clear that this criterion is determined by the degree of convergence of the solution at each stage, which ensures that the optimal result of the solution converges to an approximate solution with sufficient accuracy.

The length of the Mopkob chain L_k: L_k selection first needs to determine the attenuation function, and then L_k should ensure that the control parameters corresponding to the various values can be restored to quasi-balanced state.

4. Results and analysis

4.1. Optimum layout of high-rise shear wall structure sensor based on simulated annealing genetic algorithm

In MATLAB software, the maximum value of non-diagonal element of MAC matrix is obtained by changing the number of measuring points, as shown in Fig. 1.

As can be seen from this figure, the MAC non-diagonal element maximum value decreases with the increase of the number of measuring points, at the same time as the number of measuring points increases, MAC non-diagonal element maximum slightly ups and downs, the overall trend is gradually decreasing. It can be seen from the coordinate data in the figure that when the number of measuring points reaches 12, the maximum value of MAC non-diagonal elements has reached

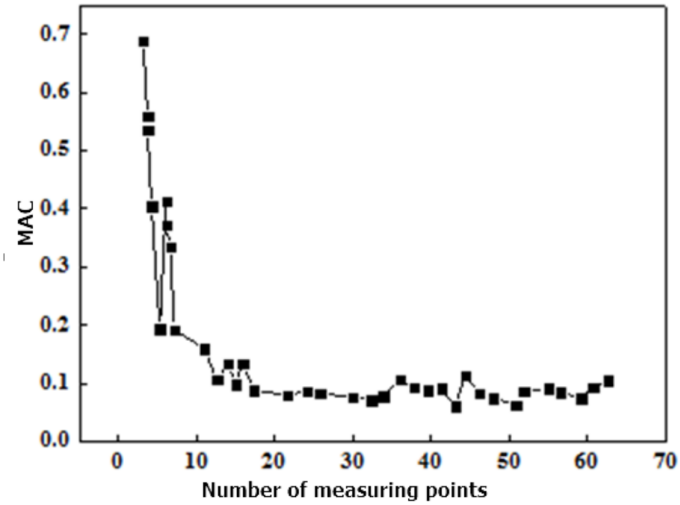


Fig. 1. The relationship between the maximal value of non-diagonal element of MAC matrix and the number of measuring points

0.0943. When the number of measuring points is 17, the MAC non-diagonal element maximum value is 0.08404; when the number of measuring points is 21, the MAC non-diagonal element maximum value is 0.07231, which indicates that the model has good convergence under the simulated annealing genetic algorithm. Taking into account the economic factors, only 12 sensors can be installed to meet the technical requirements, which also once again embodies the significance of the optimal sensor placement. Therefore, the number of sensors to be installed here is set to 12, run the simulated annealing genetic algorithm program to get the optimal scheme of the MAC criterion index 0.1251, the mac index value and the annealing temperature with the iteration number of the relationship shown in Figure 2. The optimization scheme is shown in Table 1.

When the number of measuring points is 8, the MAC non-diagonal element maximum is 0.02033; when the number of measuring points is 10, the MAC non-diagonal element maximum is 0.01469. From Fig. 3, it can be seen that when the number of measuring points is equal to 8, the maximum value of MAC non-diagonal elements tends to be minimum. When the number of measuring points is greater than or equal to 10, with the increase in the number of measuring points MAC non-diagonal maximum value no longer changes. Therefore, considering the economic factors, the number of sensors to be installed in the steel truss bridge model is 8, and the simulated annealing algorithm program can be used to get the iterative curve of the mac criterion index value in each generation. The relationship between the fitness value and iteration number as shown in Fig. 2.

4.2. Optimum layout of steel truss bridge sensors based on simulated annealing genetic algorithm

The optimal layout of simulated annealing genetic algorithm in steel truss bridge is studied here. First obtained MAC non-diagonal element maximum and the number of measuring points of the curve, as shown in Table 2.

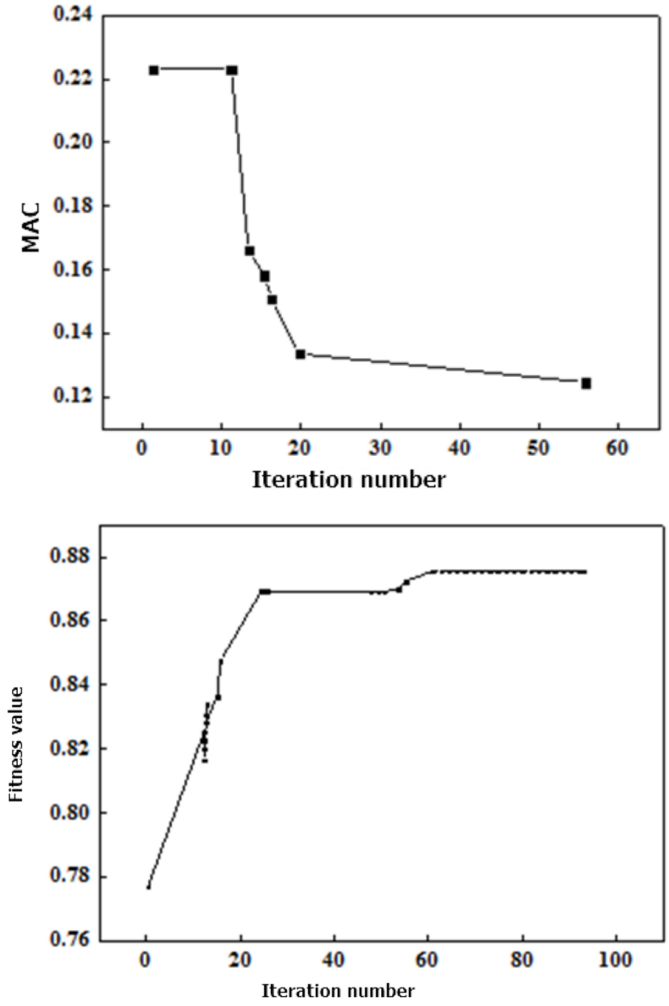


Fig. 2. Configure GA-SA plots for 12 sensors: up–relationship between the maximum value of the MAC non diagonal element and the iteration number, bottom–relationship between fitness value and number of iterations in each generation

Table 1. Optimal placement of 12 sensors

Sensor number	Node number	Direction	Sensor number	Node number	Direction
1	16	X	7	292	X
2	91	Y	8	328	X
3	157	X	9	352	X
4	181	X	10	409	X
5	190	Y	11	415	X
6	286	X	12	457	Y

Table 2. Optimum layout scheme of steel truss bridge with 8 sensors

Sensor number	Node number	Direction	Sensor number	Node number	Direction
1	3	Z	5	17	X
2	7	Z	6	17	Z
3	11	Z	7	21	X
4	12	X	8	22	Z

Optimum layout of multi-story frame sensor based on simulated annealing genetic algorithm: The content of the research process is the same as (2), the maximum value of the non diagonal element of MAC decreases with the increase of the number of points, at the same time, with the increase of the number of measuring points, the maximum value of the MAC non diagonal elements is slightly ups and downs, and the overall trend is gradually decreasing. It can be seen from the coordinate data in the figure that when the number of measuring points reaches 12, the MAC non-diagonal element maximum has reached 0.1769, when the number of measuring points is 17, the MAC non-diagonal element maximum is 0.1692, when the number of measuring points is 21, the maximum value of MAC non-diagonal element is 0.1386, which indicates that the model has good convergence under simulated annealing genetic algorithm. Taking into account the economic factors, only 12 sensors can be installed to meet the technical requirements, which also once again embodies the significance of the optimal sensor placement. Here, the number of sensors to be installed is set to 12, running the simulated annealing genetic algorithm program, get the best scheme of the MAC criterion index of 0.2123.

5. Conclusion

This paper combines the genetic algorithm and simulated annealing algorithm, proposed an efficient global optimization of hybrid genetic algorithm, and applied to the optimal layout of bridge health monitoring sensors. The main conclusions are as follows:

The position of the optimal measuring point increases with the increase of the number of sensors and presents the inheritance, that is, when only 5 sensors are installed, the sensor will also appear when the 12 sensors are installed. It can be seen that the search of hybrid genetic algorithm is based on the contribution of each measuring point to the amount of test information. The important measuring points are selected first, and the required number of measuring points are selected according to the importance order of each measuring point.

The algorithm is based on two-dimensional coding genetic algorithm as the main algorithm, and the simulated annealing algorithm is introduced into it as a separate operator. On this basis, MATLAB optimization interface was designed, and the optimal placement of the sensor is carried out for the high-rise shear wall, the steel truss bridge and the multi-story frame structure, the position of the sensor in the measurement of the first 5 order modes is given. The results show that the method is ideal and the expected results are achieved.

References

[1] T. H. Yi, H. N. Li: *Methodology developments in sensor placement for health monitoring of civil infrastructures.* International Journal of Distributed Sensor Networks *8* (2012), No. 8, paper 612726.

[2] I. Fister, X. S. Yang, J. A. Brest: *A comprehensive review of firefly algorithms.* Swarm and Evolutionary Computation *13* (2013), 34–46.

[3] T. H. Yi, H. N. Li, G. Song, X. D. Zhang: *Optimal sensor placement for health monitoring of high-rise structure using adaptive monkey algorithm.* Structural Control and Health Monitoring *22* (2015), No. 4, 667–681.

[4] K. V. Yuen, S. C. Kuok: *Efficient Bayesian sensor placement algorithm for structural identification: a general approach for multi-type sensory systems.* Earthquake Engineering & Structural Dynamics *44* (2015), 757–774.

[5] A. Khare, S. Rangnekar: *A review of particle swarm optimization and its applications in solar photovoltaic system.* Applied Soft Computing *13* (2013), No. 5, 2997–3006.

[6] M. Mohammadi, M. Nastaran, A. Sahebgharani: *Development, application, and comparison of hybrid meta-heuristics for urban land-use allocation optimization: Tabu search, genetic, GRASP, and simulated annealing algorithms.* Computers, Environment and Urban Systems *60* (2016) 23–36.

[7] U. Can, B. Alatas: *Physics based metaheuristic algorithms for global optimization.* American Journal of Information Science and Computer Engineering *1* (2015), No. 3, 94–106.

[8] M. G. Soto, H. Adeli: *Placement of control devices for passive, semi-active, and active vibration control of structures.* Scientia Iranica *20* (2013), No. 6, 1567–1578.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 237–244

Transformer fault detection based on infrared power image

Nan Yao[1], Xi Wu[2]

Abstract. The purpose of this study is to study the transformer fault detection based on infrared power image. Based on the infrared detection method of transformer, a scheme of on-line fault detection for power transformer is proposed. The transformer fault on-line detection and diagnosis system is designed. The effect of on-line detection of transformer faults is tested. The system focuses on a particular area. Through the fault temperature threshold and the rate of temperature change, it can predict the operation of the equipment. Combining the information in the infrared image database, it can determine the type of transformer failure. The results show that the application of infrared diagnosis technology in transformer fault diagnosis ensures the normal operation of transformer, improves the utilization ratio of transformer, and reduces the economic loss of power system. Therefore, it can be concluded that the system realizes on-line fault detection of infrared image of transformer.

Key words. Infrared detection, power transformer, image processing, fault diagnosis.

1. Introduction

In the process of power transmission and distribution, power transformers are one of the most critical devices in the grid [1]. It is responsible for power transmission and voltage and current conversion and other functions [2]. Its operating state directly affects the integrity of the entire power system operating chain. Therefore, it is very important to reduce and prevent transformer failure [3]. The safe operation of the transformer is not only an economic problem, but also a matter of social public interest [4]. With the expansion of power production scale, the security of power generation is becoming higher and higher [5]. In the live state, the maintenance technique determines the operation of the equipment [6]. The on-line detection technology of transformer can improve the reliability of equipment, reduce the cost of equipment maintenance and improve the economic benefits [7]. The infrared thermal imaging or the infrared thermography is based on the principle of infrared radiation [8]. By measuring the infrared radiation energy of the object surface, the

[1]Jiangsu Electric Power Company Research Institute, Nanjing, 211107, China
[2]State Grid Wuxi Power Supply Company, Wuxi, 214061, China

temperature distribution on the surface of the object is converted into an image of an intuitive thermal image (greyscale or color chart) for subsequent analysis [9]. Because of its fast detection speed, large area and wide range, infrared thermography is more and more widely used in thermal diagnosis of power equipment [10].

Infrared inspection technology is a new state maintenance technology which has been developed in recent years [11]. Through non-contact infrared thermal imaging technology, it can detect the running state of the equipment, so as to keep abreast of the operation of the equipment, find out the cause of the accident and predict the future status of the equipment [12]. At present, infrared detection technology has been widely used. It has become an important means of on-line detection of power equipment [13]. In the detection of power equipment failure, infrared diagnostic technology has the advantages of safe operation, high sensitivity, accurate judgment, visual image, high detection and diagnostic efficiency, and free from electromagnetic interference [14]. It can be calculated and analyzed. It can detect and diagnose a large number of internal and external defects in power transformers, and quickly perform infrared imaging on the device's thermal state. Through the analysis of the infrared image distribution of the transformer, the hidden faults and defects of the transformer in operation are diagnosed. The infrared detecting technology can find the hidden trouble of transformer in time and accurately. It reduces energy losses caused by equipment failures, thereby avoiding certain electrical equipment accidents. In the safe operation of power equipment, fault detection technology has played a significant role [15].

2. State of the art

As far as the international scope is concerned, in the middle of 1960s, the Swedish national electric power company firstly applied infrared thermal imaging technology to the fault diagnosis of power equipment. In 1980s, under the auspices of the American Electric Power Research Institute, American road research and development company developed a new on-line monitoring instrument for internal temperature of high-pressure equipment, that is, fluorescent fiber thermometer. The British and Swiss state power bureau and the former Soviet Union electricity management department have set up standards and procedures for using infrared inspection technology to detect power equipment. In 1993, Detroit Edison and Illinois State Electric Company introduced the latest developments in the detection of electrical equipment in overhead transmission lines, substations and power stations by infrared detection technology. It shows that the infrared detection and diagnosis technology has become an important technical means for monitoring the power equipment, discovering the hidden dangers and preventing the sudden and sudden malignancy. The American Electric Power Research Institute published the RP1289-1 report in June 1994. It points out that optical fiber point thermometer plays an important role in ensuring the operation life of transformer and reducing the fault of power transformer. In China, in 1960s, the Northeast power technology improvement bureau, Shenyang Electric Power Bureau and Changchun Institute of Optics and optics jointly developed the first generation of power equipment detection and diagnosis infrared

thermometer.

In 1988, the North China electric power test research institute undertakes the main research on infrared diagnosis technology of power generation and transformation equipment. In 2008, the national development and Reform Commission promulgated the code for application of infrared diagnostics of live equipment (DL1T664-2008). In October 2009, in the "national Power Grid Corp high voltage skills competition" held in Xi'an, the infrared diagnosis was listed as one of the important events. Dai Wenyuan proposed a theoretical model for establishing the relationship between the size of the defect to be measured and the brightness level in the infrared thermal image. Wang Yanwu proposed a method to simulate the two-dimensional temperature field according to the temperature distribution curve. The defect types within the material are determined by simulating the temperature field. However, according to the surface temperature measurement results, the theoretical model for accurate identification of defect size needs further study. Based on the infrared radiation theory, the infrared radiation model of the inner parts of the shell was established when the internal components were overheated. Based on the infrared imaging temperature measurement of shell surface, the inverse problem of heat conduction is studied by using conjugate gradient method, and the heating temperature and orientation of internal components are identified. According to the characteristics of electrical equipment failure and the characteristics of infrared diagnosis of electric power, Jiang Dingyou studied a method of automatic fault recognition of red heat image of high pressure equipment. Yang Zhengbo, who use infrared imager to obtain the transmission line temperature change image. Based on the analysis of the characteristics of the infrared image of the transmission line, the edge information of the image is obtained by comparing the advantages and disadvantages of each color space and applying HSI color space conversion. With the advantage of preserving the edge information by median filtering, the improved median filter is used to eliminate the interference. The highest temperature region is extracted by the gradient method, which can quickly and accurately diagnose the fault of the transmission line. In China's power industry equipment fault diagnosis, infrared diagnosis has made some progress. However, from the equipment diagnostic engineering, it is still in the initial stage. Therefore, standardization and intelligentization are the future directions of infrared diagnostic technology.

3. Methodology

In fact, through the analysis of the temperature field of the running equipment and the study of the thermal image, the nature of the equipment fault is determined. That is, by revealing the local overheating or abnormality of the surface by means of the transformer, the source of the fault is found. This requires a large number of on-site operation of the transformer for infrared scanning detection. The transformer fault can be distinguished only by mastering the law of thermal image of transformer faults. The infrared image fault detection and diagnosis system mainly uses the infrared image. It converts the image information into the temperature information of the equipment. According to the temperature value and change of parts, the

operation of the equipment is predicted.

The design idea of fault diagnosis system is as follows: the infrared image is enhanced by linear transformation and histogram equalization algorithm. Wavelet packet threshold algorithm is used to denoise infrared image. The Canny operator edge detection method is used to segment the image. The improved H-u invariant moments are used to extract infrared image features, and the nearest neighbor classifier is used for image recognition. With Visual Basic 6.0 as the developing tool, the on-line fault detection and diagnosis system of power transformer is designed. The key of fault diagnosis is to establish the database corresponding to the fault and temperature information. Database-saved data and infrared image data can also provide analytical data for archival analysis. Infrared image database includes the following: Transformer parts of the normal operation of the infrared image and the corresponding fault infrared images, the average temperature of each part of the equipment is running, the operating voltage / current, the maximum temperature of the current part of the transformer, the operating voltage / current and ambient temperature, the type of image (faulty, faulty description, cause of failure) and so on. The processing of infrared images is shown in Fig. 1.

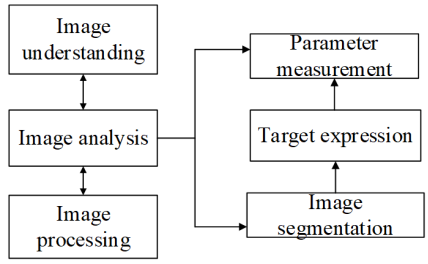


Fig. 1. Activity period of the sensor node and the wireless communication base station

After the normal operation for a period of time, the surface temperature distribution of the transformer will tend to be stable or the rate of change tends to be stable, that is to say, it will enter a relatively stable state. However, from the beginning of the work to the steady state, the surface temperature changes follow certain rules. According to this rule, the temperature variation curve can be plotted. When the monitoring site fails, the surface temperature change curve will change, and the temperature will rise at a certain rate.

Transformer infrared diagnostic steps: According to the image displayed at the abnormal temperature of the device, the infrared image of the abnormal part is acquired. The infrared thermal image is processed and segmented to facilitate the diagnosis of equipment faults. According to the operation principle of the equipment and the operation data of voltage and current, the fault types are judged. According to DL/T664-2008 standard, equipment defects are classified. According to the characteristics of temperature difference and image, the qualitative conclusions of defects are made. Depending on the nature of the defect, the cause of the site fever is found.

4. Result analysis and discussion

As shown in Fig. 2, the maximum temperature of the transformer monitoring site changes with time. It can be seen from the figure that the initial temperature rise rate is the largest and decreases with time. Finally, the state tends to be stable. In the figure, curve 3 represents the temperature change during normal operation of the transformer. It can be observed that the transformer operating temperature finally rises to a steady state. When the temperature is stable, the operating temperature of curve 2 is greater than the threshold value T. It can be seen from the figure that the temperature change rate k of curve 2 after stabilization is almost zero. It proves that the transformer can continue to run, but it is also in general alert. The curve 1 is mutated when the transformer temperature is about to enter a steady state. Temperature change rate $k > 0$ will indicate a serious failure. It will prompt the dialog box, warning the device failure. Therefore, in the transformer fault incubation period, the potential failure can be diagnosed by measuring its temperature change rate.

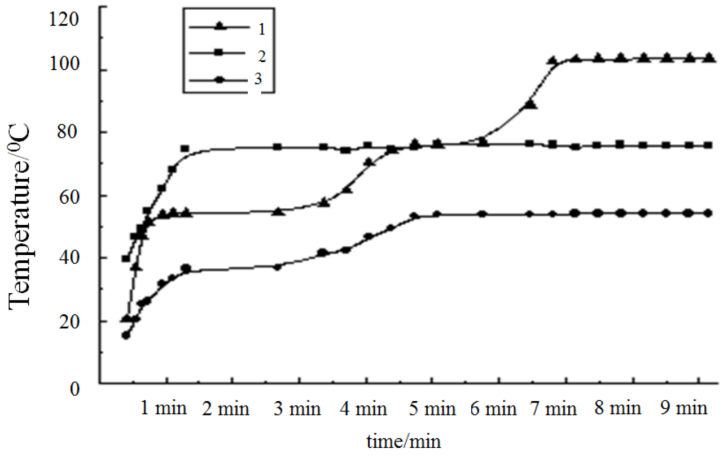


Fig. 2. Change of the maximum temperature of transformer with time

The infrared thermal image of the collected transformer casing was tested. The fault diagnosis of the transformer is tested by the fuzzy temperature difference method. The general failure of the equipment is obtained, and possible causes of such failure are listed. Field inspection showed that the connection of transformer bushing was bad.

5. Conclusion

In this paper, based on infrared image recognition and diagnosis technology, the on-line detection of power transformer is carried out. Through the theoretical analysis and experimental simulation test, the application of infrared diagnosis

technology in transformer fault diagnosis is realized. The utility model ensures the normal operation of the transformer, improves the utilization rate of the transformer and reduces the economic loss of the power system. In the future work, on the basis of a large number of detection experiments, the nature of transformer faults needs to be further studied. Through the infrared thermal image, the equipment failure is described in detail. It reveals the root causes of various failures and improves the ability to diagnose faults.

The following aspects need to be improved and perfected. The improvements in the algorithm are limited. By using classical image processing and segmentation algorithm, the infrared image is processed and segmented. Although a certain effect has been achieved, it is necessary to do more improvements, in order to achieve better display results. Because of the faults of the transformer, the source of the infrared image is limited, and the number of the contrast samples in the database is not enough, which makes the experimental results convincing. In future tests, the fault information in the database should be filled at any time. The problems such as the infrared thermal image characteristics, the internal relationship between the infrared thermal image and the fault type are still needed to be studied deeply and systematically.

References

[1] J. S. N'Cho, I. Fofana, Y. Hadjadj, A. Beroual: *Review of physicochemical-based diagnostic techniques for assessing insulation condition in aged transformers.* Energies *9* (2016), No. 5, paper 367.

[2] W. A. Gaspar, E. J. de Oliveira, L. de Mello Honório, L. C. N. Machado: *Modified successive geometric segmentation method applied to power transformers faults diagnosis.* Journal of Control, Automation and Electrical Systems *26* (2015), No. 2, 159–169.

[3] O. Aljohani, A. Abu-Siada: *Application of digital image processing to detect short circuit turns in power transformers using frequency response analysis.* IEEE Transactions on Industrial Informatics *12* (2016), No. 6, 2062–2073.

[4] G. Rigatos, P. Siano: *Power transformers' condition monitoring using neural modeling and the local statistical approach to fault diagnosis.* International Journal of Electrical Power & Energy Systems *80* (2016), 150–159.

[5] S. Degeratu, P. Rotaru, S. Rizescu, S. Danoiu, N. G. Bizdoaca, L. I. Alboteanu, H. O. Manolea: *Condition monitoring of transformer oil using thermal analysis and other techniques.* Journal of Thermal Analysis and Calorimetry *119* (2015), No. 3, 1679–1692.

[6] A. J. C. Trappey, C. V. Trappey, L. Ma, J. C. M. Chang: *Intelligent engineering asset management system for power transformer maintenance decision supports under various operating conditions.* Computers & Industrial Engineering *84* (2015) 3–11.

[7] K. Bandara, C. Ekanayake, T. K. Saha, P. K. Annamalai: *Understanding the ageing aspects of natural ester based insulation liquid in power transformer.* IEEE Transactions on Dielectrics and Electrical Insulation *23* (2016), No. 1, 246–257.

[8] D. Suresh, S. P. Singh: *Reduced-rating hybrid active power filter comprising zero-sequence transformer and three-phase three-wire active power filter for three-phase four-wire distribution system.* Electric Power Components and Systems *44* (2016), No. 13, 1503–1514.

[9] O. Koreh, K. Torkos, M. Bashir Mahara, J. Boressay, V. Izvekov: *Study of*

water clusters in insulating oils by fourier transform infrared spectroscopy. IEEE Transactions on Dielectrics and Electrical Insulation *5* (1998), No. 6, 896–902.

[10] P. Przybylek: *A new method for indirect measurement of water content in fibrous electro-insulating materials using near-infrared spectroscopy.* IEEE Transactions on Dielectrics and Electrical Insulation *23*, (2016), No. 3, 1798–1804.

[11] J. Jiang, G. M. Ma, H. T. Song, H. Y. Zhou, C. R. Li, H. B. Wang, Y. T. Luo, H. Wu: *Tracing methane dissolved in transformer oil by tunable diode laser absorption spectrum.* IEEE Transactions on Dielectrics and Electrical Insulation *23* (2016), No. 6, 3435–3442.

[12] A. Glowacz: *DC motor fault analysis with the use of acoustic signals, coiflet wavelet transform, and K-Nearest neighbor classifier.* Archives of Acoustics *40* (2015), No. 3, 321–327.

[13] X. Mao, X. Zhou, L. Zhai, Q. Yu: *Dissolved gas-in-oil analysis in transformers based on near-infrared photoacoustic spectroscopy.* International Journal of Thermophysics *36* (2015), Nos. 5–6, 940–946.

[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.

[15] S. S. Kumar, M. W. Iruthayarajan, M. Bakrutheen: *Investigations on the suitability of rice bran oil and corn oil as alternative insulating liquids for transformers.* IEEJ Transactions on Electrical and Electronic Engineering *11* (2016), No. 1, 10–14.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 245–254

Mobile Internet anomaly traffic detection technology research based on improved wavelet neural network

Qingshan Li[1, 2]

Abstract. Mobile Internet anomaly traffic detection has a very important significance for ensuring the effective operation of the network and raising the robustness of service providing ability. Quantum particle swarm optimization algorithm is combined with the wavelet neural network, and the parameters of the neural network are optimized by quantum particle swarm optimization. We build a wavelet neural network model based on quantum particle swarm algorithm. Because quantum particle swarm optimization algorithm is easy to fall into local optimum, and it reduces the diversity of population and the global search ability at the same time, we put forward improved quantum particle swarm optimization algorithm based on adaptive local search scheme. Then the improved scheme is used to optimize parameters of wavelet neural network. The experiment results show that the proposed scheme has higher detection rate of abnormal state and lower misjudgment rate of normal state than tradition quantum particle swarm optimization algorithm.

Key words. Anomaly traffic, wavelet neural network, detection rate.

1. Introduction

Wireless communication, mobile broadband and embedded technology progress, makes the calculation function of smart phones and other mobile devices enhanced. Smart phones become the unity equipment combining telecommunication network and Internet communication, and gradually become information center of personal. Along with the rapid growth in the number of users, its malicious software is more and more, which is not only harmful to mobile device security, but also does bad to user privacy property security. How to protect the safety of mobile devices and network is concerned by academia and industry.

In the smartphone security protection technology, operating system security reinforcement and digital signature technology is not enough to prevent malicious

[1]School of EECS, Peking University, Beijing, 100871, China
[2]MOE Key Lab of Network and Software Security Assurance, Peking University,Beijing, 100871, China

software. Most of the mainstream smartphone security software based on feature matching can detect public malicious software, but it cannot detect the unknown malware, and due to the particularity of mobile network, the characteristic library of the software is difficult to guarantee the timely update. For mobile network security protection, the traditional intrusion detection system is difficult to cope with the new mobile network DoS attack. The anomaly detection technology can detect unknown malicious programs, as well as new DoS attack, which is one of the current research focuses in the field of mobile Internet security protection.

For more complex data sets, different anomaly detection techniques have different difficulties. The anomaly detection technology based on clustering and the adjacent method has low detection performance for high dimensional data set. When the dimension is high, the distance between the normal instance and abnormal instance is difficult to draw. These two kinds of technology need to use suitable distance measurement standard to determine the exception. Anomaly detection technology based on spectrum can reduce the dimensions, but performance depends on the hypothesis that normal and abnormal instance has difference after they are projected onto a low dimension. The anomaly detection technology based on classification needs to train normal and abnormal instance. Anomaly detection based on statistical method depends on the assumption of distribution function. Anomaly detection based on information theory needs sensitivity standard to determine the abnormal event. Statistical and signal-based network traffic recognition for anomaly detection was investigated by Michal [8]. Network anomaly detection by cascading k-Means clustering and C4.5 decision tree algorithm was proposed by Muniyandi [9]. Discriminating DDOS attack from flash crowds by means of flow correlation coefficient was put forward by Yu Shui [10]. A design of history based traffic filtering with probabilistic packet marking Against Dos attacks was put forward by Tadashi Kiuchi [11]. Discriminating DDoS attack traffic from flash crowd through packet arrival patterns was presented by Therasak Thapngam [12]. A kind of real time DDoS detection method using fuzzy estimators was proposed by S. N. Shiaeles. Distributed collaborative DDoS detection method based on traffic classification features was investigated by Z. Xiong. D. Stevanovic investigated detection method of malicious and non-malicious website visitors by means of unsupervised neural network learning. Statistical and signal-based network traffic recognition for anomaly detection was proposed by M. Choras. Here, we investigate mobile Internet traffic anomaly detection based on wavelet neural network aiming at the high-dimensional nonlinear behavior of network traffic on small-time scale and propose improved scheme in view of drawbacks of wavelet neural network. It is organized as follows. In the next section, a kind of mobile Internet anomaly traffic detection method based on improved wavelet neural network is put forward. In section 3, in order to test the performance of mobile Internet anomaly traffic detection method, experiments are done. In the end, some conclusions are given.

2. Mobile Internet anomaly traffic detection based on improved wavelet neural network

Wavelet neural network (WNN) and quantum particle swarm optimization are used in traffic anomaly detection of mobile Internet. Wavelet neural network is trained by quantum particle swarm optimization. The parameters combination of wavelet neural network is taken as a particle in the quantum particle swarm optimization algorithm. The parameter vector with the optimal fitness value is searched, then wavelet neural network trained by quantum particle swarm is used in anomaly traffic detection. The wavelet neural network structure is shown in Fig. 1

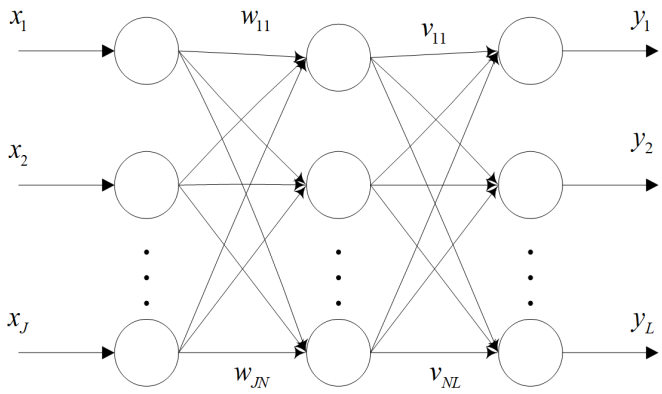


Fig. 1. Adopted wavelet neural network structure

The structure of continuous wavelet neural network is similar to BP neural network, which includes the input layer, hidden layer and output layer. Neuron activation function of hidden layer is wavelet function $\psi(x)$. Training process of wavelet neural network is based on the idea of error back propagation, weight and wavelet parameters are adjusted according to gradient descent direction.

Function $\psi(x) \in L^2(R)$ is the activation function of hidden layer node. There holds

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\psi(w)|^2}{|w|} dw < \infty \,. \tag{1}$$

The output result of node i in the output layer is

$$y_i(t) = \sigma(x_n) = \sigma\left(\sum_{j=1}^{N} v_{ji}\psi_{a_j b_j}(\sum_{k=1}^{J} w_{kj}x_k(t))\right) \,, \quad i = 1, 2, \ldots, L \,, \tag{2}$$

where $\sigma(x_n)$ is the Sigmoid function

$$\sigma(x) = \frac{1}{1 + \mathrm{e}^{-x}} \,. \tag{3}$$

Wavelet neural network uses gradient descent method to adjust the network connection weights, expansion coefficient and shift coefficient based on the error function.

$$E = \frac{1}{2}\sum_{i=1}^{L}\left(y_i(t) - d_i(t)\right)^2 . \tag{4}$$

Here, t represents the current time, x_k represents the input vector of input node k, y_i represents output vector of output node i, w_{kj} is the weight of input node k and hidden layer node j. Quantity v_{ji} represents the weight of hidden layer node and output layer node i, a_j and b_j are expansion coefficient and shift coefficient of the hidden layer node j, E is the error function, and d_i represents the expected output of output node i.

Further

$$\text{net}_j = \sum_{k=1}^{J} w_{kj}x_k(t)\,, \tag{5}$$

$$\psi_{a_jb_j}(\text{net}_j) = \psi\left(\frac{\text{net}_j - b_j}{a_j}\right), \tag{6}$$

and

$$y_i(t) = f\left(\sum_{j=1}^{N} v_{ji}\psi_{a_jb_j}(\text{net}_j)\right). \tag{7}$$

The gradient descent method is used to optimize the wavelet neural network parameters, which is easy to make the algorithm trapped into local optimum and cause oscillation effect. PSO algorithm and QPSO algorithm can overcome the above shortcomings. But quantum particle swarm optimization algorithm and standard PSO is easy to fall into local optimum. Therefore, wavelet neural network parameters are optimized by improved quantum particle swarm optimization algorithm. In the Dth dimension of space, there are m particles. The position of the ith particle is $x_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$, $i = 1, 2, \ldots, m$. The local optimum of the ith particle is $p_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$ and the optimal position of the whole particle swarm is $p_g = (p_{g1}, p_{g2}, \ldots, p_{gD})$, $g \in \{1, 2, \ldots, M\}$. The evolution equation of quantum particle swarm optimization is

$$\text{mbest}(t) = \frac{1}{m}\sum_{i=1}^{m} p_i(t) = \left[\frac{1}{m}\sum_{i=1}^{m} p_{i1}(t), \frac{1}{m}\sum_{i=1}^{m} p_{i2}(t), \ldots, \frac{1}{m}\sum_{i=1}^{m} p_{iD}(t)\right], \tag{8}$$

$$p = (r_1p_{id} + r_2p_{gd})/(r_1 + r_2)\,, \tag{9}$$

$$X_{id}(t+1) = p_{id}(t) \pm \beta\,|\text{mbest}(t) - X_{id}(t)|\ln\frac{1}{u}\,. \tag{10}$$

In formulas (8)–(10), $i = 1, 2, \ldots, m$, $d = 1, 2, \ldots, D$, u, r_1 and r_2 are random number belonging to [0,1]. Symbol t represents the current iteration times, D represents the dimension of particle, M represents the swarm scale, $p_i(t)$ represents the

current best position of the particle, which means the ith particle in the tth iteration. Quantity $p_g(t)$ represents the global optimal position and mbest(t) represents the average best position in the swarm, which means average value of particle in the tth iteration. Symbol $p_{id}(t)$ represents the random point between $p_i(t)$ and $p_g(t)$. Quantity β represents the shrinkage and expansion coefficient. When quantum particle swarm optimization algorithm is applied to the practical problem, there are many kinds of control methods for the parameter β. One simple way is to set β as a fixed value.

Another effective method is based on linear decreasing. We put forward a kind of adaptive mechanism based on error function

$$z = \frac{f_i - f_{\text{gbest}}}{\min(\text{abs}(f_i), \text{abs}(f_{\text{gbest}}))}\,, \tag{11}$$

where f_i represents the best fitness value of the ith particle, $f_i = f(\text{pbest}(i))$. Symbol f_{gbest} represents fitness value of gbest. The error function is used to represent proximity degree between the particle and the global optimal position gbest. For a certain particle, if value of the error function is the smaller, it means that particle is closer to the global optimal point. As a result, the search area of the particle is narrowed. For the point that is away from the global optimal value, the value of β should be smaller, and, the value of β should be bigger for the point near to the global optimal value. This is because it is almost impossible to search the point far away from their current position, so the value of β should be larger. Otherwise, it is impossible to search a new optimal position. On the contrary, for particles away from the global optimal value, the value of the parameter β is set to be small to ensure the convergence of population. In this way, we construct one adaptive function and can work out value of β according to error function of fixed particle.

In order to improve the algorithm global search and local search ability, and prevent algorithm trapped in local optimal solution, we put forward the improvement strategy, namely adaptive local search strategy. This strategy adjusts the size of local search neighborhood area adaptively according to the search state of swarm algorithm. Changing the size of the neighborhood area is implemented by neighborhood function. Each component of current solution is added by a random variable

$$p_{id}^{'} = p_{id} + \varepsilon\,, \tag{12}$$

where $p_{id}^{'}$ is the solution in the neighborhood area of p_{id}, and ε is a random real number belonging to $(-\delta, \delta)$. Symbol δ is determined by iteration times and fitness value of current optimal solution as

$$\delta = \frac{a}{\text{iteration}} \left| \overline{f(p)} - f(p_g) \right|\,, \tag{13}$$

where $f(p_g)$ is the fitness value of current optimal solution p_g, $\overline{f(p)}$ is the average fitness value of all particles in current swarm. iteration is current iteration times and a is a given real number. The process of improved quantum particle swarm optimization algorithm is as follows:

Step 1. Initialize the population size, the number of iteration, the number of swarm, dimension, individual best position and position of current global optimal solution.

Step 2. Update all particles according to evolution equation.

Step 3. Initialize the position of p_i and p_g with the smallest fitness value.

Step 4. Adaptive local search algorithm is used to search particle swarm is carried out and fitness value of each particle is worked out.

Step 5. Update p_i and p_g.

Step 6. If it meets the constraint condition, the optimal solution is outputted. Otherwise, it turns to step 2 to search until the condition is satisfied.

Then the improved quantum particle swarm optimization algorithm is used to optimize parameters of wavelet neural network. Firstly, we determine the maximum iteration times $T_{\max}$, population size M, particle search space dimension D, shrink and expansion coefficient a_j, shift coefficient b_j, node connection weight value w_{kj} and v_{ji}. We define a vector $x_i = \{a_i, b_i, w_i, v_i\}$. Then x_i is trained and optimized as one particle of improved quantum particle swarm optimization algorithm. The particle with optimal value is mapped to wavelet neural network parameters. The training sample is input to train wavelet neural network and square error of each network on the training set is worked out:

$$E = \frac{1}{2U} \sum_{s=1}^{U} \sum_{p=1}^{L} |c_s^p - f_s^p|^2 \ , \tag{14}$$

where c_s^p and f_s^p represent ideal and actual outputs of training sample s on the output node p. Symbol U represents the total number of sample and L represents the number of output node. Thirdly, we determine whether the algorithm meets termination condition. If it does not meet the condition, new particle individuals are generated according to quantum particle swarm optimization algorithm and these new individuals are optimized again. If it meets termination condition, the parameters with the optimal fitness value is taken as the final results.

3. Experiment and analysis

Ns-2 is used as simulation environment, and the data set is KDD CUP 99 data set, which contains 5000000 connection records, including four types of attacks, namely DOS, U2R and R2L and PROBE. Ten percent of the data is taken as wavelet neural network training samples. The training sample contains 22 kind of different attack way with 494021 connection records, of which only 97278 connection records is normal, and each connection record has 41 different attributes. Network node receives data from the mobile Internet, then the first data is standardized. After standardization, these network data is input into wavelet neural network for testing. KDD CUP 99 data set has 41 different properties in each connection record, so the structure of wavelet neural network is 41-60-5. Five output layer nodes respectively correspond to the four types of attacks and a normal type. Particle swarm

optimization, quantum particle swarm optimization and improved quantum particle swarm optimization is used to optimize parameters of wavelet neural network. Through the simulation experiment results, we compare advantages and disadvantages of three kinds of method. The population size is $M = 100$, the maximum iteration times is $T_{\max} = 1000$, the values of w_{kj} and v_{ji} belong to (-1,1). The value ranges of a_j and b_j belong to (1,100). In the quantum particle swarm optimization, $\beta = 0.5 \cdot (T_{\max} - T)/T_{\max} + 0.5$. In the particle swarm optimization algorithm, inertia weight is determined by $w = (w_{\text{init}} - w_{\text{end}})(T_{\max} - t)/T_{\max} + w_{\text{end}}$ and $T_{\max} = 1000$ means the maximum iteration times. Symbol t represents the current iteration times, w_{init} represents the initial inertia weight, w_{end} represents termination inertia weight and the value of w is from 0.4 to 0.9, $c_1 = c_2 = 2$. The average detection rate and misjudgment rate is shown in Table 1. Detection rate of four kinds of network anomaly is shown in Table 2. In the simulation experiment, the network parameters of wavelet neural network are very different respectively after optimized by above three kinds of training algorithm. From Table 1, we can see that detection rate and misjudgment rate of improved quantum particle swarm optimization is better than the other two kinds of algorithm. Conhat improved quantum particle swarm optimization has faster convergence speed and it has better convergence effect under the same number of iteration times. In mobile internet anomaly detection, the improved quantum particle swarm optimization has several advantages, which not only improves detection rate of network traffic abnormal state, but also reduces the misjudgment rate of normal state.

Table 1. Average detection rate and misjudgment rate

algorithm	average detection rate	average misjudgment rate
PSO	90.87	8.79
QPSO	93.17	5.67
Improved QPSO	95.01	4.65

Table 2. Detection rate of four kinds of network anomaly

algorithm	DOS	PROBE	U2R	R2L
PSO	88.77	88.57	94.27	91.87
QPSO	91.43	91.20	96.48	93.57
Improved QPSO	92.63	93.17	97.51	95.09

4. Conclusion

Network anomaly detection, which establishes the normal network traffic behavior model to detect the abnormal behavior of the network, is an important means of intrusion detection. In recent years, with the continued growth of the number of mobile Internet users and the rapid deployment of new network application, threat

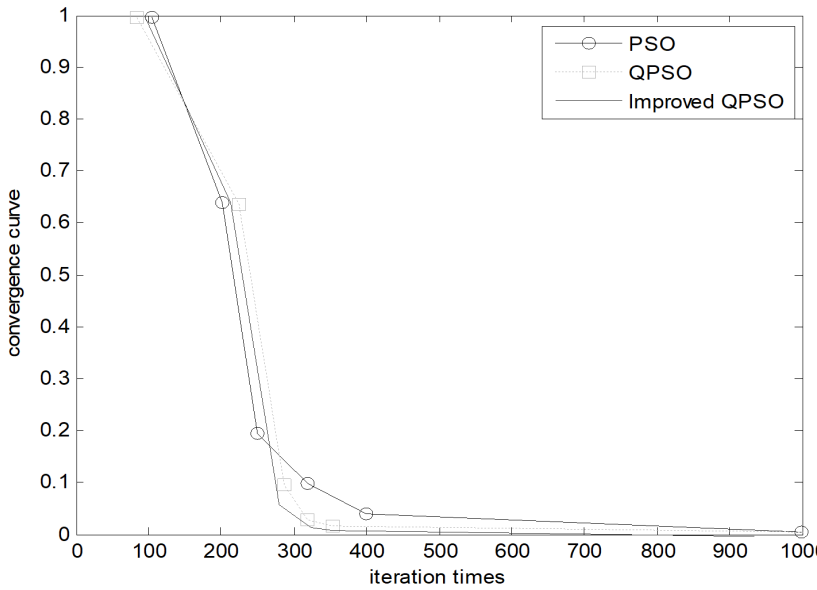


Fig. 2. Convergence curve of three kinds of algorithm

of attack against the network traffic has become increasingly serious. Aiming at the high-dimensional nonlinear behavior of network traffic on small-time scale, a novel wavelet neural network model optimized by improved quantum wavelet neural network is presented. The experiment results show that the proposed scheme is effective. In the condition that mobile node has limited computing power and battery capacity, data dimension reduction methods should be investigated. We should retain the network characteristics as much as possible, minimize the data dimension, decrease the complexity of the algorithm and reduce battery consumption of mobile node.

References

[1] X. Ma, Y. Chen: *DDoS detection method based on chaos analysis of network traffic entropy*. IEEE Communications Letters *18* (2014), No. 1, 114–117.

[2] J. Yang, D. Woolbright: *Correlating TCP/IP packet contexts to detect stepping-stone intrusion*. Computers & Security *30* (2011), Nos. 6–7, 538–546.

[3] X. Wu, Y. Chen: *Validation of chaos hypothesis in NADA and improved DDoS detection algorithm*. IEEE Communications Letters *17* (2013), No. 12, 2396–2399.

[4] Y. Gao, C. Chen, J. Bu, W. Dong, D. He: *ICAD: Indirect correlation based anomaly detection in dynamic WSNs*. IEEE Wireless Communications and Networking Conference, 28–31 March 2011, Cancun, Quintana Roo, Mexico, IEEE Conference Publications (2011), 647–652.

[5] G. Thatte, U. Mitra, J. Heidemann: *Parametric methods for anomaly detection in aggregate traffic*. IEEE/ACM Transactions on Networking *19* (2011), No. 2, 512–525.

[6] J. Yu, H. Kang, D. H. Park: *An in-depth analysis on traffic flooding attacks detection and system using data mining techniques.* Journal of Systems Architecture *59* (2013), No. 10, Part B, 1005–1012.

[7] Y. Ye, T. Li, Q. Jiang, Y. Wang: *CIMDS: Adapting postprocessing techniques of associative classification for malware detection.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) *40* (2010), No. 3, 298–307.

[8] M. Choraś, Ł. Saganowski, R. Renk, W. Holubowicz: *Statistical and signal-based network traffic recognition for anomaly detection.* Expert Systems *29* (2012), No. 3, 232–245.

[9] A. P. Muniyandi, R. Rajeswari, R. Rajaram: *Network anomaly detection by cascading K-Means clustering and C4.5 decision tree algorithm.* Procedia Engineering *30* (2012), 174–182.

[10] S. Yu, W. Zhou, W. Jia, S. Guo, Y. Xiang, F. Tang: *Discriminating DDoS attacks from flash crowds using flow correlation coefficient.* IEEE Transactions on Parallel and Distributed Systems *23*, (2012), No. 6, 1073–1080.

[11] T. Kiuchi, Y. Hori, K. Sakurai: *A design of history based traffic filtering with probabilistic packet marking against DoS attacks.* IEEE/IPSJ International Symposium on Applications and the Internet, 19–23 July 2010, Seoul, South Korea, IEEE Conference Publications, (2010), 261–264.

[12] T. Thapngam, S. Yu, W. Zhou, G. Beliakov: *Discriminating DDoS attack traffic from flash crowd through packet arrival patterns.* IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 10–15 April 2011, Shanghai, China, IEEE Conference Publications, (2011), 952–957.

Received July 12, 2017

Acta Technica *62* No. 2B/2017, 255–264

Dynamic simulation of crawler excavator walking mechanism

Donglin Jiang[1]

Abstract. In order to develop a mechanical simulation experiment, a dynamic simulation of crawler excavator walking mechanism is designed. The performance of walking mechanism of excavator is tested by dynamic simulation. According to the data analysis, it is known that crawler walking mechanism is an important part of mine excavator and a walking device with better adaptability than wheeled walking mechanism. The crawler can travel on a road with poor working conditions, such as deep snow, swamp, mud and sand, while the wheeled devices cannot work in these conditions. Therefore, in combination with the actual working conditions of excavators, two kinds of mines are simulated and studied. The dynamic simulation of the meshing force between the driving wheel and the track plate during the walking process is carried out, and the movement state of the driving wheel in the forward and backward conditions is simulated and analyzed. The experimental results show that the grounding of the crawler device is much smaller than that of the wheel type, and the load impact can be greater than that of the wheeled type. In addition, the working environment of the excavator is very bad, and it is often run in the mines in the sand and mud. Through simulation analysis, it is concluded that the walking mechanism of crawler excavator is more efficient and convenient than the traditional wheeled excavator.

Key words. Excavators, load impact, dynamics.

1. Introduction

China has vast land resources and rich mineral resources, and these abundant resources are the strong material base for building a modern and powerful country. China's mineral resources are mostly suitable for open pit mining. Only 65 % of the iron ore reserves are suitable for open-pit mining. Other large coal mines, non-metallic minerals and non-ferrous metal mines are also suitable for open-pit mining. At present, iron ore open-pit mining has accounted for 88.4 % and non-ferrous metal mine open-pit mining accounted for 47 %. In addition, chemical raw materials open-pit mining accounted for 50 % and construction materials open-pit mining accounted for more than 50 % [1]. Experience from all over the world has proved that the production capacity and the demand for minerals is increased because of the large

[1]College of Mechanical Engineering, Changchun Normal University, Hangchun Jilin, 130032, China

mining equipment used in open pit mining [2]. The surface mining method is 5–10 times more efficient than the underground mining method, and the cost is 1–2 times lower than that of the underground mining method [3]. This method is easy to realize automatic mining. It has the advantages of safe production and high recovery rate. Its rapid and efficient economic results have promoted the rapid development of opencast mining [4]. For a long time, the development of all kinds of crawler excavators in our country, including crawler walking mechanism, have been dominated by the traditional mode [5]. The emergence of a new type of machine usually involves several steps, such as preliminary design, prototype production, industrial test, improvement, finalization and mass production [6]. The design and research model based on physical prototype has the disadvantages of high cost and long cycle, and it cannot be tested repeatedly on the physical prototype [7]. Therefore, our country's mechanical excavator product replacement is slow, it is difficult to adapt to the rapid changing market demand. At the same time, with the development of mining excavator toward large-scale, the structure of the product becomes more and more complex [8]. The manpower, material and financial resources invested in the development of a physical prototype are also increasing [9]. Therefore, the traditional way of product development is used to develop new large-scale mining excavators, the demand for funds is huge, but also difficult to adapt to market demand. Based on this, the dynamic simulation analysis of crawler excavator walking mechanism is studied [10].

2. Overview of crawler walking mechanism

The crawler walking device was invented in 1830s by Dimitri Chagyanskiy [11]. It moves the moving vehicle according to two closed parallel rotating tracks. It can travel on the roadlessness, snow, mud, swamp area where wheeled wheels are not available.

Although crawler walking device has been widely used in engineering, it is still not perfect. Compared with wheel walking device, it has low mechanical efficiency and working reliability as well as the complex mechanism. In addition, due to the complexity of crawler movement, previous studies often separate each link to simplify analysis [12]. Coupled with the uncertainty of ground mechanics, the design work relies too much on empirical data, empirical formulas, and the designer's own design experience.

The working principle of the crawler type walking device is shown in Fig. 1: the whole caterpillar vehicle is supported on the caterpillar trolley; the rear end of the track trolley is a driving wheel 1, the front end is a guide wheel 2, the middle part is a supporting wheel 3. The crawler passes the load to the lower track of the crawler through these wheels. The track 4 is an endless chain with an upper branch supported on the idler pulley 5. When the driving cycle is turned, the tracks engaged with the drive wheel have a tendency to move [13]. However, because the adhesion between the lower track branches and the soil is greater than that of the driving wheel, the guide wheel and the supporting wheel, the track does not move [14]. The caterpillar wheel, guide wheel and supporting wheel roll along the caterpillar track,

and the whole caterpillar vehicle walks forward [15].

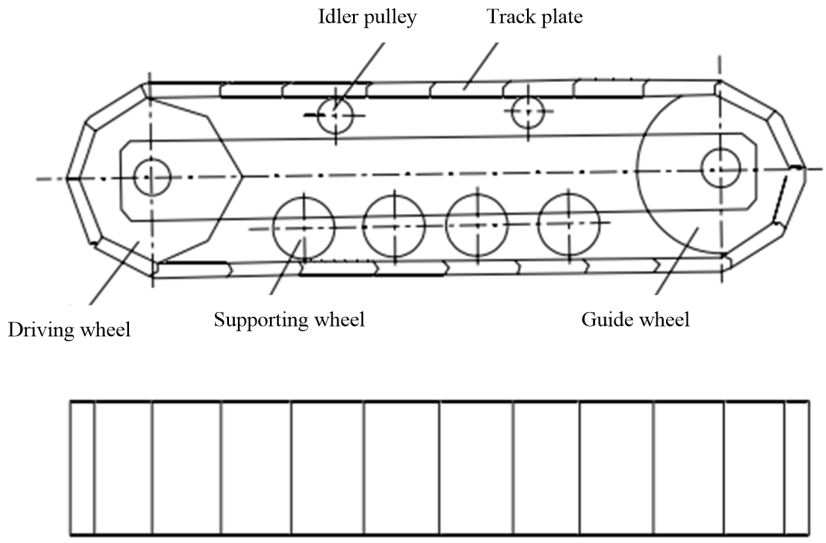


Fig. 1. Sketch of working principle of crawler walking device

3. Method

Before establishing the virtual prototype of crawler walking mechanism, it is necessary to analyze the structure and working principle of the crawler walking mechanism.

The crawler walking mechanism is composed of a driving wheel, guide wheel, caterpillar plate, supporting wheel and plate. The function is to change the torque from the motor to the traction force. Working principle: when the drive wheel turns, it needs to pull out the track below the support wheel. But because the weight of the whole machine makes the track and the ground engage tightly, the soil produces shearing stress, forming the traction force that makes the tracked vehicle advance. This pressure can also be interpreted as the force exerted on the track by the ground, which is transmitted to the whole by the caterpillar through the drive wheel, and the whole machine is moved. The root of the movement of the whole machine comes from the friction of the ground facing the track plate.

The function of the track is to ensure the high pass of the vehicle on various complicated ground, reduce the running resistance and increase the adhesion to the ground. By interaction with the ground, it can generate traction and braking force of the crawler propulsion device. Different soil conditions will affect the interaction of the entire machine and the ground, and ultimately affect the whole machine, such as driving resistance and other driving performance. Therefore, a suitable vehicle ground mechanical model must be established according to the actual soil

conditions. In the prototype model, considering the actual working environment of the excavator, the focus of the simulation analysis, the complexity of the simulation and the ground mechanical model are calculated according to the rigid soil road.

In the ground model, the pressure between the tracked plate and the ground is defined by the impact force, the impact force generated by the interaction force between the track and the ground. In ADAMS, the contact force is a special force acting on the component. The contact force is generated when the two members contact each other and deform. The magnitude of contact force is related to the size of deformation and the speed of deformation. If the two components are separated from each other without contact, the contact force is zero. The impact force) can be expressed by the following formula

$$F = -k\left(q - q_0\right)^n - cq\,. \tag{1}$$

In the formula, F is the ground force, k stands for the stiffness coefficient, $q - q_0$ is the penetration, n is the deformation index, C is the damping coefficient and q is the deformation speed.

The friction between the tracked plate and the ground is obtained by the formula

$$F_{\mathrm{f}} = \mu F\,, \tag{2}$$

where F_{f} is the ground friction and μ represents the friction coefficient between ground and track plate.

In the hard ground model, the ground is assumed to have stiffness and damping.

4. Results and discussion

4.1. Excavator driving analysis

As the initial speed of the machine is zero, and the resistance is certain, a certain drive torque is needed, and the drive torque is estimated according to the resistance and drive wheel parameters: $T_{\mathrm{T}} = 800\,\mathrm{kN\,m} \geq T_{rmf}$.

Since the Z axis is in the opposite direction to the track, the speed curve is negative as it moves forward. It is shown by the curves in Figs. 2 and 3 that the horizontal change of the center of mass of the driving wheel is relatively steady. Some periodic fluctuations of velocity and acceleration are caused by sudden force when engaged with the track plate. As the dynamics model is built, there is a small gap between the ground and the track system. Therefore, the simulation has just started and the tracks will suddenly fall to the ground. It can be seen from the curve that there is a certain impact between the two projects. The damping in the contact force will soon consume this part of the shock, so this part of the shock will not affect the analysis later.

When driving at a constant speed, the drive force is the same as the resistance, and the speed is constant. Therefore, a driving pair is added to the driving wheel to replace the driving torque so as to drive at a uniform speed. Accord-

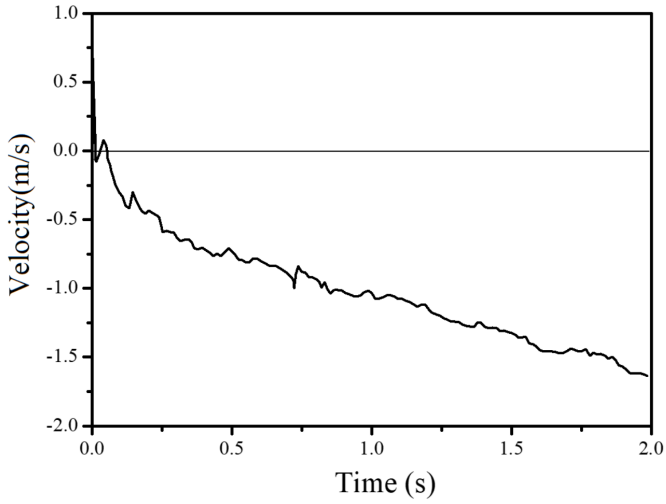


Fig. 2. Horizontal speed of the center of mass of the driving wheel

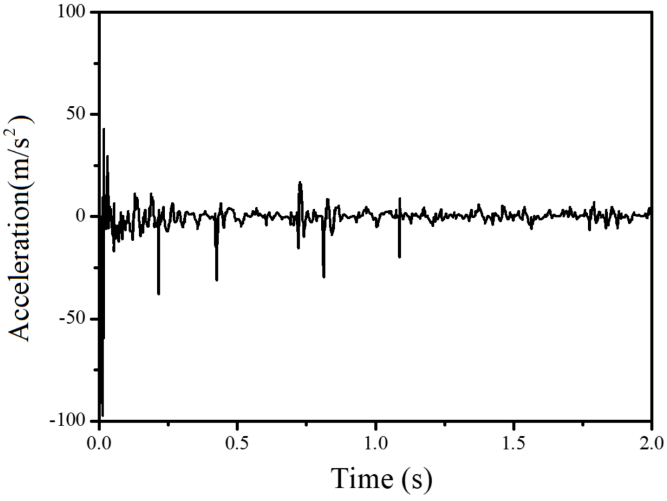


Fig. 3. Acceleration in the horizontal direction of the center of the wheel

ing to the maximum speed of walking, the maximum angular velocity of the driving wheel is calculated: $\omega_{\mathrm{T}} = v/R$, $R = 1.1\,\mathrm{m}$, $v_{\max} = 1.6$ km/h $= 0.4444\,\mathrm{m/s}$, $\omega_{\mathrm{T}} = 0.4444/1.1 = 0.404 = 23.16^{\circ}/\mathrm{s}$.

As can be seen from Figs. 4 and 5, when the driving wheel is applied with a constant rotational speed, the horizontal velocity and the acceleration are not uniform and are periodically changed. The reason is that when each crawler plate is engaged, the whole force is uneven. In the crawler walking mechanism, the periodic variation of speed is inevitable. The smaller the length of the caterpillar track is, the closer it becomes. Compared with the simulation video, the maximum speed and

acceleration in the curve is perpendicular to the mass center of the driving wheel when the crawler plate runs. When the crawler plate is vertical below the driving wheel, the driving wheel and the pedrail plate have a great force in the horizontal direction.

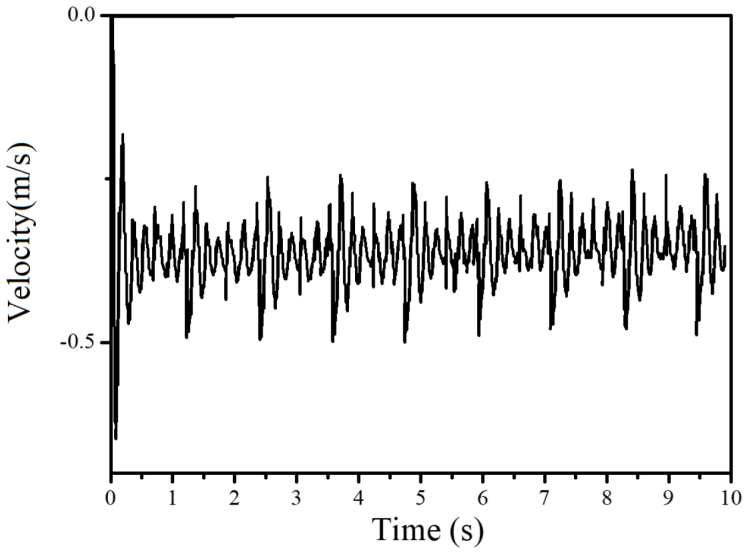


Fig. 4. Speed of the drive wheel in the horizontal direction

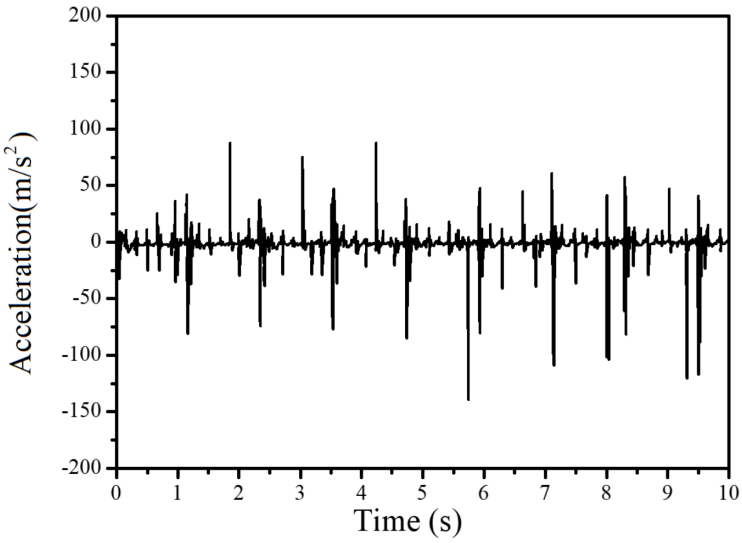


Fig. 5. Acceleration of the drive wheel in the horizontal direction

When the speed of the drive wheel is constant, the horizontal acceleration changes

periodically, but the change is great and the impact is great. In each period of change there are some sudden increases with irregular change. Because in simulation, the contact surface between the driving wheel and the track plate appears the phenomenon of edge contact with the plane, so the force suddenly increases. But in real environments, this absolute edge and plane contact do not exist, so these can be ignored.

4.2. Excavator astern driving

Different from the forward driving, the internal resistance of the whole track system in astern driving will become larger, and the force acting on the track will change. Drive torque $T_{\mathrm{T}} = 980\,\mathrm{kN{\cdot}m}$.

As shown in Figs. 6 and 7, excavators are inevitably retrograde at work. Compared to the forward constant torque, the speed of the astern constant torque varies greatly, and the overall trend is accelerated retreat. Compared with the simulation animation, the peak of the speed curve is the maximum of the tangential force of the driving wheel after the contact between the crawler plate and the driving wheel. The acceleration of a constant torque moving in the horizontal direction fluctuates frequently.

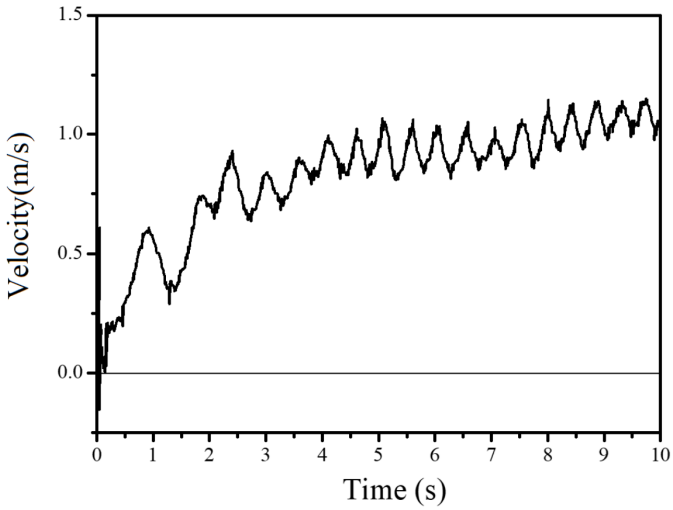


Fig. 6. Speed of the constant torque of the drive wheel in the horizontal direction

When driving back at a uniform speed, a reverse motion pair is needed at the driving wheel. The maximum speed is $v_{\max} = 1.6\,\mathrm{km/h}$, and the maximum angular velocity is $\omega_{\mathrm{T}} = 0.4444/1.1 = 0.404 = 23.16^{\circ}/\mathrm{s}$.

As shown in Figs. 8 and 9, when the driving wheel rotates in the uniform direction, the horizontal speed and the acceleration of the driving wheel center are relatively smooth. When retracting, the engagement of the track shoe with the drive wheel is not affected by external forces such as the ground, and the engagement is more

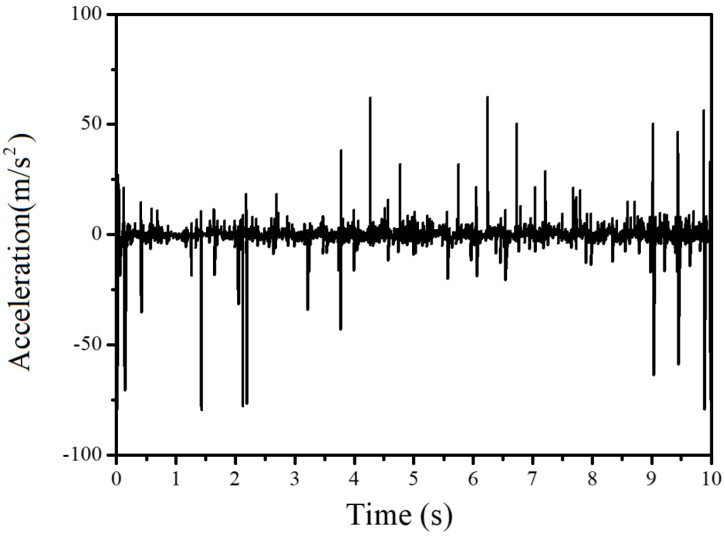


Fig. 7. Acceleration of the constant torque of the drive wheel in the horizontal direction

stable. However, there are ground-facing pressure and horizontal frictional forces when engaging forward. Complex forces lead to greater volatility.

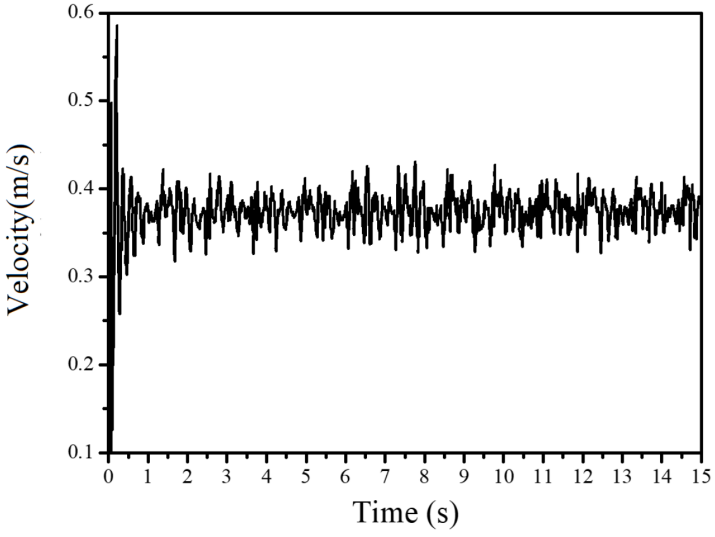


Fig. 8. Speed of the wheel center in horizontal direction

4.2.1. Design of Backstepping controller

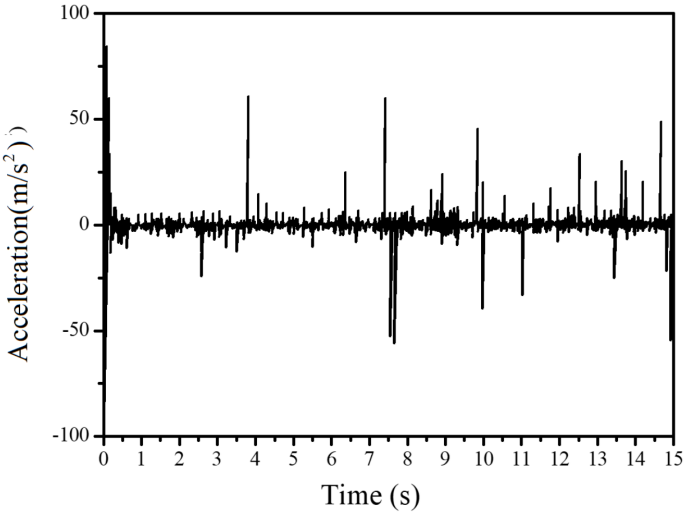


Fig. 9. Acceleration of the wheel center in horizontal direction

5. Conclusion

The dynamic simulation of the crawler type walking device is carried out by using the crawler traveling mechanism of the mining machine and the actual walking condition. Some problems of dynamic modeling of track walking mechanism are solved by self-programming. The dynamic simulation analysis of excavator crawler walking system is carried out, and the two conditions of forward and backward are analyzed, respectively. In each case, the constant speed and constant torque are compared and analyzed. In the forward condition, a continuous four track shoe meshing with the drive wheel is selected. The simulation results show that the four track shoes and the driving wheel are in the horizontal and vertical direction. During the running of the track, the engagement force curve fluctuates periodically due to the periodic engagement of the drive wheel and the track shoe. In addition, the abrupt change in the constant speed exceeds the constant torque force mutation because of the initial conditions of the ADAMS calculation equation. In the retreating condition, the running resistance of the backward traveling crawler device is greater than the running resistance of the forward traveling. Through the simulation of the horizontal movement of the wheel center, the constant speed of the recursive torque is larger than the constant speed, but the speed change is stable and the impact is less compared with constant speed.

References

[1] A. C. de Camargo, M. A. B. Regitano-d'Arce, C. R. Gallo, F. Shahidi: *Gamma-irradiation induced changes in microbiological status, phenolic profile and antioxidant activity of peanut skin.* Journal of Functional Foods *12* (2015), 129–143.

[2] F. Geng, J. H. Saleh: *Challenging the emerging narrative: Critical examination of coalmining safety in China, and recommendations for tackling mining hazards.* Safety Science *75* (2015), 36–48.

[3] E. Cho, A. M. Fiore, D. W. Russell: *Validation of a fashion brand image scale capturing cognitive, sensory, and affective associations: Testing its role in an extended brand equity model.* Psychology & Marketing *32* (2015), No. 1, 28–48.

[4] B. Limnirankul, T. Onprapai, P. Gypmantasiri: *Building local capacities in natural resources management for food security in the Highlands of Northern Thailand.* Agriculture and Agricultural Science Procedia *5* (2015), 30–37.

[5] B. Li, T. Du, B. Yu, J. van der Gucht, F. Zhou: *Caterpillar-inspired design and fabrication of a self-walking actuator with anisotropy, gradient, and instant response.* Small *11* (2015), No. 28, 3494–3501.

[6] D. Grzelczyk, B. Stańczyk, J. Awrejcewicz: *Prototype, control system architecture and controlling of the hexapod legs with nonlinear stick-slip vibrations.* Mechatronics *37* (2016), 63–78.

[7] X. Xin, J. Zhang, A. Zhu, C. Zhang: *Effects of long-term (23 years) mineral fertilizer and compost application on physical properties of fluvo-aquic soil in the North China Plain.* Soil and Tillage Research *156* (2016), 166–172.

[8] M. Y. Liu, L. L. Zhang, J. Li, Y. Li, N. Li, M. Q. Chen: *Characteristics of the cross-sectional vorticity of the natural spawning grounds of schizothorax prenanti and a vague-set similarity model for ecological restoration.* PlosOne *10* (2015), No. 8, e036724.

[9] A. Bermúdez, A. L. Rodríguez, I. Villar: *Extended formulas to compute resultant and contact electromagnetic force and torque from Maxwell stress tensors.* IEEE Transactions on Magnetics *53* (2017), No. 4, paper 7200409.

[10] S. Frimpong, M. Thiruvengadam: *Multibody dynamic stress simulation of rigid-flexible shovel crawler shoes.* Minerals *6*, (2016), No. 3, paper 61.

[11] C. Xiao, C. Zhang: *Dynamic simulation analysis of working device for hydraulic excavator based on ADAMS.* Telkomnika *14* (2016), No. 3A, 194–201.

[12] L. E. Ericsson, J. P. Reding: *Dynamic simulation through analytic extrapolation.* Journal of Spacecraft and Rockets *19* (1982), No. 2, 160–166.

[13] D. E. Hill, J. R. Baumgarten, J. T. Miller: *Dynamic simulation of spin-stabilized spacecraft with sloshing fluid stores.* Journal of Guidance, Control, and Dynamics *11* (1988), No. 6, 597–599.

[14] K. R. Saul, X. Hu, C. M. Goehler, M. E. Vidt, M. Daly, A. Velisar, W. M. Murray: *Benchmarking of dynamic simulation predictions in two software platforms using an upper limb musculoskeletal model.* Computer Methods in Biomechanics & Biomedical Engineering *18* (2015), No. 13, 1445–1458.

[15] D. A. Marshall, L. Burgos-Liz, M. J. IJzerman, W. Crown, W. V. Padula, P. K. Wong, K. S. Pasupathy, M. K. Higashi, N. D. Osgood: *Selecting a dynamic simulation modeling method for health care delivery research-part 2: report of the ISPOR dynamic simulation modeling emerging good practices task force.* Value in Health *18* (2015), No. 2, 147–160.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 265–272

Electric control system of numerical control machine tool based on PLC

Xiaofang Hou[1]

Abstract. In order to explore the advantages and disadvantages of electric control system of numerical control machine, the electric control system of numerical control machine is first of all designed. Firstly, the overall design scheme of electric control system is set up in accordance with the requirements of electric control system. In addition, through the calculation of electric load, the electric parameters of electric components are determined, which is convenient for the choice of types and wire of electric equipment. At the same time, the electric control system of main equipment is designed, focusing on the design of spindle, feed shaft, and PLC module. At last, the automatic tool setting control system is discussed, doing some preliminary work for realizing automatic tool setting in the future. The results showed that the electric control system has high accuracy and reliability. And it can be concluded that the electric control system can be widely applied in microelectronics, computer and so on fields.

Key words. Arduino, wireless sensing, network node.

1. Introduction

The performance of the control system is decided by the advantages and disadvantages of electrical control mode of CNC machine tools. A machine tool is essentially an electromechanical energy conversion device whose function is to convert electrical energy into mechanical energy required for processing. Therefore, the distribution and management of electrical energy is the energy guarantee of machine tools, and it plays a role similar with the heart in machine tools. The electrical control system of machine tools includes machine tool power supply and distribution system, spindle control system, feed shaft control system, cooling oil pump control system and other auxiliary equipment control systems. The introduction of electrical control, especially the electrical switching control, is a major improvement. In this paper, it mainly used to control cooling motors, oil pumps motors, and other auxiliary electrical control systems. It is applied to replace relay - contactor control system, which improves the reliability and flexibility of the system, resulting in a qualitative leap in control performance [1].

[1]Shaanxi Institute of Technology, Hanzhong, Shaanxi, 723001, China

Digital control technology refers to the technology making use of digital information to control. The machine tool which uses digital information to control the movement and machining process of machine tools is called NC (numerical control) machine tool. It is the product of the combination of digital control technology and machine tools. CNC machine is a typical mechatronic product, and it is a set of modern machinery manufacturing device with high efficiency, high precision, high flexibility and high automation, which integrates modern machinery manufacturing technology, automatic control technology, detection technology, and computer information technology [2]. It is the same as other mechatronic products, which is also composed of mechanical body, power source, electronic control unit, detection sensor part and executive machine (servo system)

2. Method

2.1. Overall design scheme of electrical control system for NC machine tools

In the CNC machine tools, the overall program of electrical control is shown in Fig. 1. ARM9 is used as the main control system, and DSP+FPGA acts as the motion controller and the core of servo drive, and PLC is used as the control core of the electric auxiliary system [3].

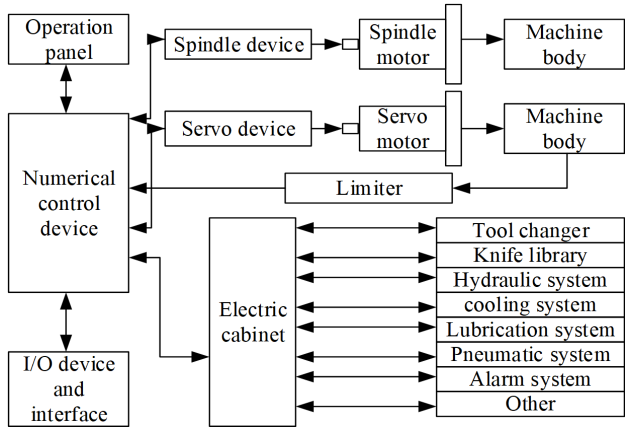


Fig. 1. Scheme of auxiliary electric control system

2.2. Control requirements of electrical control system for NC machine tools

This CNC electrical control system requires to realize the following movements and controls:

Overview of spindle control requirements: to realize the spindle motor start, stop, speed change, and speed control; require the smooth spindle speed, fast acceleration and deceleration response, low speed torque, overload ability and so on.

Overview of feeding shaft control requirements: X, Y, Z axes have each servo motor drive, requiring the shaft to achieve speed and corner control. Precision requirements: repeat positioning accuracy is ± 0.005 mm, and the positioning accuracy is ± 0.001 mm; fast speed requirements: good acceleration and deceleration, relatively smooth three axis, good following performance; requirements for satisfying the requirement of stability with the fastest speed: small vibration, strong anti-interference ability and so on.

Auxiliary control: to realize the control of hydraulic pump, lubricating motor, automatic tool setting, cooling pump motor, and fan motor; system emergency stop and station limit protection; signal control of machine running state.

The main electrical rated load of this CNC machine tool is: voltage: three-phase 380 V, current: 40 A, power: 10 kW, and frequency: 50 Hz. The limit of the electrical parameter is 1.5 times of the rated load. The rated torque of the spindle is 25 Nm, the rated torque of the X, Y and Z axes is 11.5 Nm, and the rated torque of the shaft is 18 Nm.

2.3. Main functions of the motion controller module

The block diagram of the motion controller system is shown in Fig. 2. Its function is: receiving the interpolation and control signal after main controller decoding and compiler information processing, for speed processing and linear and circular interpolation operation. And by comparing the position and speed feedback signal with all kinds of servo systems, it thereby generates the interpolation direction and pulse to the servo system [4].

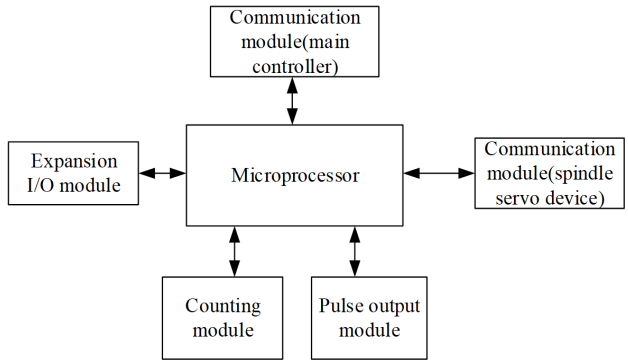


Fig. 2. Block diagram of motion controller system

Microprocessor module: according to the function number and so on information obtained by the main controller decoding, the acceleration and deceleration control and the precise interpolation calculation of the line and arc are performed. Encoder decoding module: decoding is carried out on the received encoder signal and the decoded structure is passed to the counter for counting processing to obtain the position information of the motor. Pulse output module: the interpolation pulse and direction are generated to the servo system according to the result of micro processing interpolation to control the rotation of the motor. Communication module: through

the MAX3490 bus and the main controller for full duplex communication, function number and other control information after the host computer decoding are received, and relative coordinates and absolute coordinates information are sent to the main controller. Extended I/O module: I/O port is reserved to facilitate connection with auxiliary equipment, such as lubrication mechanism, hydraulic system, limit switch and other I/O equipment.

2.4. Functions of PLC in the electric control system

PLC (programmable logic controller) is commonly used for automatic control of equipment, which uses replacing the relay control circuit to realize the control function of CNC machine tools. The significant performance lies in its control function with a high degree of flexibility. In the permission scope, NC axis configuration control signal changes the control signal of auxiliary feed shaft, which can control machine operation panel signal and interface switch control. And only by changing the programming can it make the appropriate configuration efficient immediately, and the same numerical control system can be flexibly applied to CNC machine tools with different configurations [5].

PLC is used in numerical control machine tools, mainly for the auxiliary electrical control of the periphery of machine tools. It is also called programmable machine tool controller, referred to as PMC.

The realization of the control of machine tool auxiliary equipment (including cooling system, lubrication system, lighting system, tool change system, automatic tool setting, control panel and so on) is accomplished. The peripheral circuit (strong circuit) cannot be directly connected with each other, so the peripheral circuit must be converted through the relay loop in the middle [6]. The entry must be converted into a weak signal, and for controlling the peripheral circuit, the output weak signal must be converted into a strong signal.

3. Design of electrical control system for main equipment

3.1. Design of electric control system for spindle motor

Spindle control requirements are: the spindle speed is stable, acceleration and deceleration response is fast, low speed torque is large, and the overload ability is strong and so on [7]. Machine spindle is generally used to provide power for the processing of machine tools. The NC machine model is lokson650 vertical machine tools, so it is spindle drive cutting work-piece rotation. The spindle drive uses the closed loop vector control mode with Yaskawa inverter as the core, and the incremental photoelectric encoder as the feedback as spindle speed meter [8]. The frequency converter and system communication adopt RS485 bus.

3.2. Electric control design of feed shaft

This type of numerical control machine tool is a half line rail machine tool, that is to say, one shaft adopts the line rail, and the other shaft adopts the hard rail. Feed shaft using the line rail has small contact surface with mechanical body, the friction is small, the resistance is small, and the load capacity is relatively light, generally used for light load cutting, machining metal parts, auto parts and other products [9]. While the feed shaft using hard rail has rough surface, the friction is large, the resistance is large, and the load capacity is strong, can be used for processing heavy cutting, which is generally used for processing mold products. The machine adopts half line rail, and its use is a product between the two, mainly used for processing hardware, auto parts, small molds and other products. Therefore, the cutting requirements are divided into: feed shaft requirements and spindle requirements, and we focus on the feed requirements: accuracy requirements: the repeat positioning accuracy is ± 0.005 mm, and the positioning accuracy is ± 0.001 mm [10]. Fast speed requirements: good acceleration and deceleration, relatively smooth three axis linkage, and good following performance. The average processing progress per second is 20 cm/min, and the fastest can reach 1 m/min. Stability requirements: small vibration, anti-interference ability and so on.

3.3. Design of PLC module

Lubrication system of auxiliary electric control system of this system has two inputs and two alarm outputs; cooling system also has two inputs and two alarm outputs; automatic knife has 6 inputs and 6 alarm outputs; lighting and signal lamp system has 1 input and 4 outputs; ATC (Automatic tool cutting) system has 5 inputs and 5 outputs. As a result, in the choice of PLC, taking into account the leave of 1/3 allowance, we use Huichuan's PLC [11]. It has 24 inputs and 15 outputs. The scanning frequency is up to 1000 Hz, which fully meets the requirements.

4. Result and discussion

The automatic tool setting needs to set the cutter parameters, including the center of length tool detection head and the radius tool detection head, the radius length and height between the center of length tool detection head and that of the radius tool detection head. These parameters must be sure to be correct; otherwise, in the tool setting process, it will easily lead to crash tool presetting instrument [12]. After manually changing the cutter, the user shall remove the tool mark corresponding to the cutter number, and avoid the tool setting instrument because of the misoperation of the user (for the re-setting tool, see Fig. 3).

The length manual cutter and the radius manual tool cutting process have the locking axis direction, which is related to the mode of the installation position of the tool setting instrument. The radius too cutting needs to determine the close direction of the tool detection. The step system will be automatically set in processing the standard cutter, and the user does not need to participate in, to avoid misoperation.

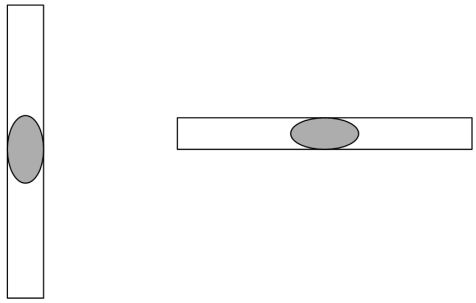


Fig. 3. Mode (1) and mode (2) of cutter installation

4.1. Manual tool setting principle of tool length

(1) Correctly set the parameters of the tool setting tool, the safe position of the cutter, the tool length and the function of the tool setting.

(2) Enter the electronic hand-wheel mode and choose the tool positioning function. The system will be positioned to the security tool in the current position, and then positioned to the center of the length cutter detection head. In the process, the electronic hand-wheel pulse will automatically disappear [13].

(3) Shake the electronic hand-wheel to move the tool head close to the length tool detection head (suggested here). This process can move the non locking shaft.

(4) Press the tool locking function, and then the system will make the electronic hand-wheel gear locked in the file, while telling the locking tool point.

(5) Shake electronic hand-wheel, to make the cutter head slowly contacting the length tool detection head, locking tool parameters (including X, Y, Z and A, four parameters). ARM will store four parameters, and if the standard cutter parameters have been already set. ARM will automatically calculate the cutter tool length compensation value and store to the corresponding cutter number. Finally, A means Auxiliary control, to realize the control of hydraulic pump, lubricating motor, automatic tool setting, cooling pump motor, and fan motor.

(6) Shake the electronic hand-wheel, to make the tool head lift, slowly get away from the length tool detection head, and push out the cutter tool locking function.

4.2. Automatic tool setting control principle of tool length

(1) Correctly set the parameters of the cutter tool setting, the safe position of the cutter, the tool length and the function of the tool setting.

(2) Ensure that the reference knife parameters have been correctly set, the cutter number is correct, and the length is manually or automatically adjusted.

(3) Enter the electronic hand-wheel mode or manual mode and select the length setting function. The system will be positioned to the security tool in the current position, then positioned to the center of the length cutter head, and to the locking value of the cutter tool. In this process, the electronic hand-wheel pulse will automatically disappear, and the manual button mobile function will be screened. If the

system opens knife angle positioning function, it also need to position the spindle.

(4) The system locates the tool head above the locking value Z of the tool point of the cutter number [14].

(5) ARM tells DSP to lock the tool point. The system locates the tool head above the locking value Z of the tool point of the cutter number.

(6) The tool head slowly contacts the length tool detection head, and DSP locks the cutter tool point parameter. ARM will store these four parameters, and automatically calculate the tool length compensation value and store it to the corresponding cutter number.

(7) The cutter head lifts the locking point, and slowly gets away from the length tool head. The tool head is raised to the tool safety point Z, and the tool locking function is introduced.

5. Conclusion

Manufacturing is an important cornerstone of human economic activity, as well as the motive force of human historical development and civilization progress. The level of manufacturing is an important indicator of the degree of industrial development in a country. The manufacturing industry, represented by the traditional electromechanical industry, is undergoing profound changes. Modern science and technology, especially the rapid development of microelectronics, computer, information and other science and technology, and its wide application in all walks of life, have brought profound changes to the machine building industry. In this paper, the electrical control system of NC machine tools is deeply studied and analyzed. An electric control system is designed and installed to satisfy the requirements of high speed, high precision, high reliability and intelligence of NC machine tools. The electrical control system of CNC + PLC + external electric appliance is adopted, and the control system of the main equipment is designed, in which the control of spindle motor, X, Y, Z axes control, and the design of PLC module are mainly described. In addition, at the end of the thesis, the tool setting control system is designed.

In the CNC machine tools, the overall program of electrical control is shown in Fig. 1. ARM9 is used as the main control system, and DSP+FPGA acts as the motion controller and the core of servo drive, and PLC is used as the control core of the electric auxiliary system [3].

References

[1] G. S. Joshi, N. V. Bhujbal, S. M. Kurkute: *Agriculture at a click using PLC & SCADA*. International Journal of Emerging Trends in Science and Technology, Impact Factor: 2.838 *3* (2016), No. 5, 3928–3932.

[2] S. H. Chen, H. Ye, Y. Yang: *Program transmission method of PLC equipments' remote monitoring software based on virtual serial port*. Computer Science and Application *6* (2016), No. 3, 110-118.

[3] D. O. Ikotun, J. A. Ademuyiwa, F. D. Famule: *Comparative analysis of customers'*

queue management of first bank PLC, and Guaranty Trust Bank PLC, Isokun Ilesa, Nigeria. International Journal of Mathematical Sciences and Computing *4* (2016), 1–11.

[4] A. Mathur, M. R. Bhatnagar, B. K. Panigrahi: *PLC performance evaluation with channel gain and additive noise over nonuniform background noise phase*. Transactions on Emerging Telecommunications Technologies *28* (2016), No. 5, e3131.

[5] T. Suzuki: *Propagation characteristics of kHz band PLC signal on MV–LV distribution system*. Electrical Engineering in Japan *199* (2017), No. 2, 47–57.

[6] D. Yoo, H. K. Choi, I. B. Sohn, Y. Kim, S. Kim, W. Kim, J. Kim: *Femtosecond-laser micromachining of a thermal blocking trench for an enhanced PLC variable optical attenuator*. Korean journal of optics and photonics *27* (2016), No. 4, 127–132.

[7] R. Bhinge, J. Park, K. H. Law, D. A. Dornfeld, M. Helu, S. Rachuri: *Toward a generalized energy prediction model for machine tools*. Journal of Manufacturing Science and Engineering *139* (2016), No. 4, paper 041013.

[8] G. Fu, J. Fu, H. Shen, J. Sha, Y. Xu: *Numerical solution of simultaneous equations based geometric error compensation for CNC machine tools with workpiece model reconstruction*. International Journal of Advanced Manufacturing Technology *86* (2016), No. 5, 2265–2278.

[9] X. Zhou, Z. Jiang, B. Song, X. Tang, S. Zheng: *A compensation method for the geometric errors of five-axis machine tools based on the topology relation between axes*. International Journal of Advanced Manufacturing Technology *88* (2017), Nos. 5–8, 1993–2007.

[10] Y. Liu, H. Zhang, X. Wang: *Analysis on influence of perpendicularity error of five axis NC machine tool error modeling accuracy and complexity*. Procedia Engineering *174*, (2017), 557–565.

[11] L. Cai, Z. Zhang, Q. Cheng, Z. Liu, P. Gu, Y. Qi: *An approach to optimize the machining accuracy retainability of multi-axis NC machine tool based on robust design*. Precision Engineering *43* (2016), 370–386.

[12] S. Ding, X. Huang, C. Yu, W. Wang: *Actual inverse kinematics for position-independent and position-dependent geometric error compensation of five-axis machine tools*. International Journal of Machine Tools and Manufacture *111* (2016), 55–62.

[13] J. Um, S. H. Suh, I. Stroud: *STEP-NC machine tool data model and its applications*. International Journal of Computer Integrated Manufacturing *29* (2016), No. 10, 1058–1074.

[14] Q. Cheng, B. Sun, Z. Liu, J. Li, X. Dong, P. Gu: *Key geometric error extraction of machine tool based on extended Fourier amplitude sensitivity test method*. International Journal of Advanced Manufacturing Technology *90* (2017), Nos. 9–12, 3369–3385.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 273–282

High compression ratio static image coding technologies

Xi Chen[1], Yuan Yuan[2], Yani Liu[1], Ruiqiang Cao[1]

Abstract. In order to reduce the block effect in flat areas and retain more image edge details, an adaptive compression algorithm of image block classification based on human visual characteristics in the DCT domain is put forward, which effectively improves the quality of the compressed image. First of all, the core transformation technology of image compression encoding is analyzed, and compression comparison analysis is carried out for three typical static compression standards JPEG, JPEG2000 and JPEG-XR, to analyze the compression effects under the condition of high compression ratio. Secondly, the new idea of image block classification compression based on visual characteristics is put forward, and a new method for the DCT domain image block classification is put forward. By calculating the activity of the image block, the image block is divided into smooth area, edge area and texture area three categories. Finally, the analysis of distortion quantization is made for the unified quantization way, and the fundamental reasons of JPEG blocking effect are discussed. Then, combining with the human visual characteristics, adaptive quantization is realized for smooth area, edge area and texture area. The results showed that the image obtained by adaptive quantization compression has better visual effects and compression quality. In a word, it is concluded that adaptive quantization is beneficial for realizing high compression ratio image encoding.

Key words. Static image, compression coding, visual property, image block classification, adaptive quantization.

1. Introduction

The image is the most direct and vivid expression of the objects, and it is the most important information source for us to get the described object information. At present, the images we obtain are mainly derived from the natural scenery taken by digital cameras or other multimedia devices. The image can be divided into static image and video image according to its content [1]. Image compression has always been a hot topic in the field of image processing. With the constant updating of the

[1]Panzhihua University, Panzhihua, Sichuan, 617000, China
[2]Sichuan College of Architectural Technology, Deyang, Sichuan, 618000, China

times, compression coding technology has become an important subject of image processing, and it is gradually normalized and standardized. At present, image compression technology is very popular, and it is widely used in every field of life.

Image compression is to reduce the amount of image data, which is the typical application of data compression in images [2]. Image compression in essence is a kind of data set [3] which is relatively small in correlation, obtained by transformation and combination of image resource data with strong correlation. The general process of image compression is shown in Fig. 1.

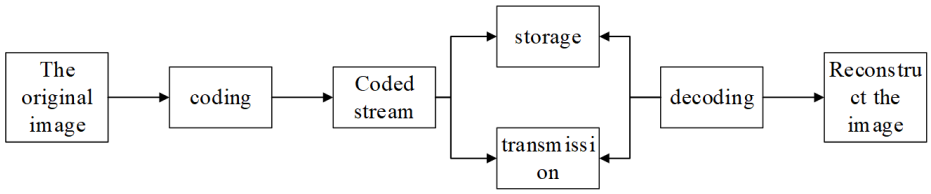


Fig. 1. Image compression process

The purpose of image compression is to minimize the redundancy of the image data. The image is, based on the form of data matrix, to make storage and transmission. As a result, a series of transformations of the data matrix can remove the redundant information, and encode the transformed data according to a certain rules, which can reduce the storage space for information. Digital images have very large amounts of data. A colorful image of 640×480 size, and a 1 G compact disk can holds only about 1000 images. Such a large amount of data undoubtedly brings great difficulties to the storage and transmission of images. There are two ways to solve this problem: to expand data transmission bandwidth; and to reduce the amount of image data. The bandwidth resource is obviously not infinitely expanded. In order to conduct more efficient data storage and transmission, to reduce the amount of image data becomes the best way to solve this problem. In consequence, the encoding algorithm with a higher compression performance becomes the object that the image processing technology pursues. By removing the intra frame and inter frame correlation of images, the image compression is achieved. While the image correlation is only considered in the static image. At the same time, the human visual system has visual characteristics, with different sensitivities to different regions. By using such characteristic of the human eye, in the process of image compression, for regions not sensitive to visual areas, in the case of without affecting the visual effect of image compression, it needs to reduce the encoding precision, so as to achieve the purpose of digital image compression. The quantization process designed in this paper is based on the order of JPEG model, introducing the adaptive quantization factors on the basis of the default quantization table given by the data unit of 8×8. The quantization factors, according to the content characteristic of image block and the sensitivity of human eye brightness, it is adjusted and modified, which improves the quantization accuracy, and improves the visual quality of image in the case of high compression ratio.

Based on the analysis of human visual characteristics, this chapter presents a new idea of compression algorithm combining image block classification, and puts

forward a method of DCT domain image classification. The traditional image features analysis is based on the spatial domain. The classification method is on the basis of DCT domain, to select the DCT AC coefficient energy as a characteristics standard to measure the image block activity, and to make the adaptive weighting about human visual system of AC coefficient energy. Thus, the image blocks are divided into three regions with different visual sensitivities, and good results are achieved.

2. State of the art

The development of image compression encoding technology is mainly divided into three stages [4]: the first generation encoding, mainly to remove data redundancy, the second generation encoding, having better compression effect and encoding rate, and the third generation encoding, combined with human visual characteristics and using multi-resolution coding.

At present, the international static image compression standard are: the JPEG standard [5], which has good compression performance and low algorithm complexity; the JPEG2000 standard [6], which supports both lossless and lossy compression two kinds of working methods, and in the lossless compression condition, it still has a higher compression ratio; JPEGXR standard [7, 8], which supports both lossy and lossless compression, which has higher compression performance than JPEG. In the same bit rate, the image compression quality is two times that of JPEG or equivalent quality only needs half of the volume.

With the development of these image encoding standards, it greatly promoted the development of image communication technology. As a result, the image information and communication technology has obtained large-scale popularization and application. The expansion of image communication technology scale has put forward higher requirements on image encoding technology, thus to further promote the development of encoding standardization work.

As the image compression and encoding technology develops rapidly, a new encoding algorithm emerges in. The encoding algorithm development experienced from simple image compression to complex image compression, and with the increasing standardization of image compression encoding, it promoted the step of the image encoding from theoretical research stage to engineering application stage, and achieved considerable achievements. As a result, the application of image compression has been throughout our life. The image types are different, and the content features are very different. The same coding algorithm is used to compress the images with different contents, and the optimal compression effect cannot always be achieved. In short, each compression algorithm has its advantages and disadvantages. In practical applications, we need to compare the performance of various compression algorithms, and select the most appropriate algorithm.

With the expansion of image application field, the image compression coding has developed rapidly, which has become an important technology in computer field, and has developed into a separate discipline system. But the traditional image compression algorithm, due to not considering human visual characteristics, the

human eye sensitive details are lost, thus affecting the quality of image compression. In consequence, the ability for using the traditional encoding method for compression of image data cannot be improved again. In order to improve the performance of image compression, researchers continue to break through innovation and find new ways and means. As a result, the second generation image encoding method that makes full use of human visual characteristics comes into being. A new encoding method not only considers the spatial frequency characteristics of the image, but also makes full use of the characteristics of human visual, so the image can obtain higher compression ratio and reconstruction quality.

3. Performance comparison of static image compression algorithms

3.1. Evaluation of image compression quality

An accurate and objective evaluation of the quality of compressed images is needed, which is a subject worthy of study in the field of image compression, and a standard to measure the compression performance of compression algorithms. The evaluation is mainly divided into subjective evaluation and objective evaluation [9]. Subjective evaluation is according to the human visual effect as a measurement standard of image compression quality. However, because the observing effect of human eye on images is easily affected by subjective effects, like image types, experimental environment and so on. Therefore, the evaluation result is uncertain, which cannot be directly used to evaluate the quality of image compression. The objective evaluation method is mainly to evaluate the quality of image compression through the calculation of formula. The main aspects are as follows:

Compression ratio, that is the ratio of the original image to the compressed image, which is usually measured by bpp. The difference of compression algorithm is the key factor that affects the compression ratio.

Encoding and decoding time is the coding efficiency.

Reconstruction of image quality. We often use peak signal to noise ratio (PSNR) and mean square error (MSE) as the evaluation criteria to evaluate the quality of image compression. MSE represents the mean square variance between the original image and the compressed image, and the calculation formula is

$$\mathrm{MSE} = \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N2-1} \left(x[n_1, n_2] - \overline{x}[n_1, n_2]\right)^2 . \tag{1}$$

In the above formula, N represents the number of the original images, n represents the number of the compressed images.

3.2. Comparison of compression performance between three algorithms

The evaluation of image compression quality is divided into subjective evaluation and objective evaluation. Subjective evaluation refers to the visual effects of an image given to human. The objective evaluation criteria are measured by the peak signal-to-noise ratio (PSNR) of the image. This paper selected 8 bits gray image Peppers.bmp with resolution of 512×512, to make the JPEG, JPEG2000 and JPEG-XR compression. In addition, the test results are compared, and the compression performance of three kinds of compression methods in high compression ratio is analyzed.

The three compression algorithms are simulated by MATLAB, the PSNR values under different bit rates are calculated, and the compression effects of the three compression methods are obtained, as shown in Table 1:

Table 1. PSNR value under different bit rates for image Peppers.bmp

JPEG		JPEG2000		JPEG XR	
Actual bitrate	PSNR	Actual bitrate	PSNR	Actual bitrate	PSNR
GIVEN BITRATE bpp = 3					
2.97	42.9	3	45.7	2.98	45.01
GIVEN BITRATE bpp = 1					
1.01	37.81	1	42.4	0.98	42.1
GIVEN BITRATE bpp = 0.7					
0.7	35.87	0.7	39.9	0.69	39.89
GIVEN BITRATE bpp = 0.4					
0.41	33.18	0.4	38.3	0.401	38.01
GIVEN BITRATE bpp = 0.2					
0.2	29.78	0.2	36.45	0.2	36.11
GIVEN BITRATE bpp = 0.107					
0.1	21.26	0.1	32.6	0.107	31.9

The compression effect diagrams of three compression methods, JPEG, JPEG2000 and JPEG-XR, can be drawn from Table 1. As shown in Fig. 2, the effect can be visually compared.

3.3. Static image compression combined with human visual characteristics

According to the visual characteristics of human eyes, the human eyes are sensitive to the noise in the image smoothing area, insensitive to the noise in the texture area, and quite sensitive to the contour information of the image. Using the spatial frequency characteristics of human eyes and human visual characteristics, the image is divided into different types of regions, and adaptive quantization is achieved in different regions. This paper presents a new idea of image compression algorithm combined with image block classification. The research shows that the human visual

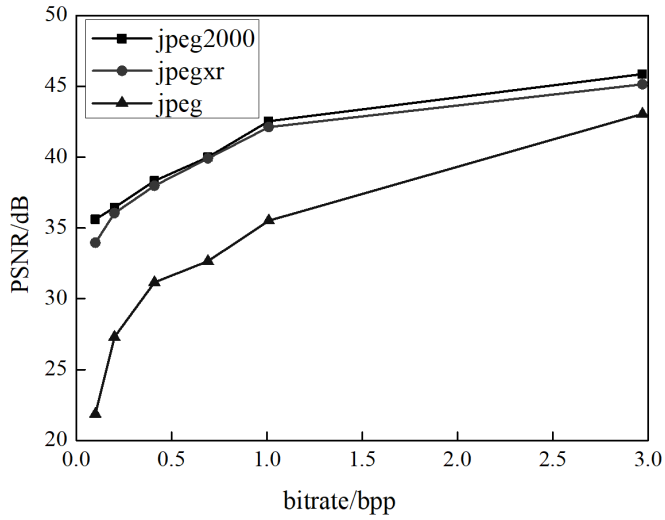


Fig. 2. PSNR value curves of three algorithms

system perceives an image through its regional features instead of pixels. Thus, an image is segmented into a series of regions with different categories, and the importance of information that each region transmits to the human eye is different. Basic idea: firstly, according to the characteristics of image blocks, the test image sub blocks are classified, and then adaptive compression is made for each kind of image blocks. Content features of image block are different, and the compression accuracy is different. In order to make the image compression maintaining high matching property with image features, and to improve the compression performance of compression algorithm, the analysis of image compression characteristics is particularly important. We often use the information entropy, mean or variance as the characteristic values of image. Taking the 8 bits gray scale image "Peppers.bmp" with resolution of 512×512 as an example, the image block activity after considering the visual weight is calculated, as shown in Fig. 3.

The weighted activity is normalized, and the range of activity and the average value are shown in Table 2.

Table 2. Activity intervals and averages

image	min	max	mean
Peppers.bmp	0	1	0.0801
Lena.bmp	0	1	0.1392
Barbara.bmp	0	1	0.2152
Baboon.bmp	0	1	0.2322

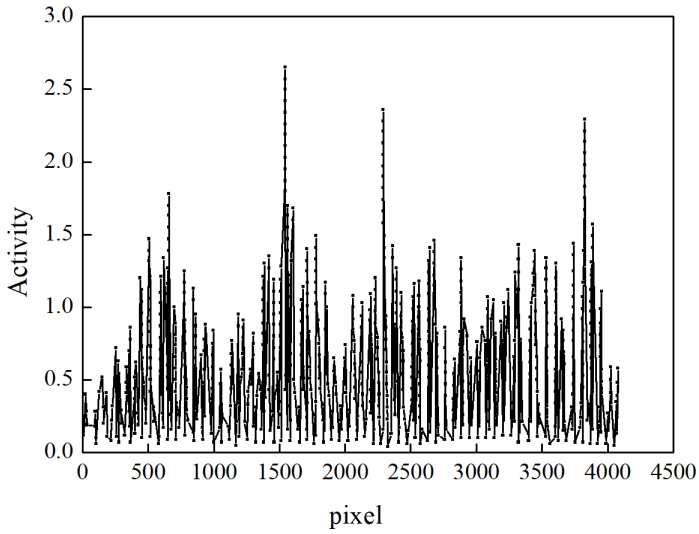


Fig. 3. Coordinate distribution of weighted activity

4. Results analysis and discussion

As can be seen from Table 1 and Fig. 2, the PSNR value of JPEG2000 is slightly higher than that of JPEGXR, and the PSNR values of JPEG2000 and JPEGXR are significantly higher than those of JPEG.

Based on the simulation, calculation and analysis of three kinds of compression algorithms, it can be known that, under the condition of low compression ratio, the compression effects of three kinds of encoding algorithm are very good, almost no distortion. In moderate compression ratio, the compression effect is better, and JPEG exhibited slight blocking effect. When continuing to increase the compression ratio, the block effect is obvious. Under the condition of high compression ratio, JPEG exhibits serious block effect, while JPEG2000 and JPEGXR compression effect can still satisfy the visual requirement. At the same time, it can be seen that, under the condition of the same compression ratio, for JPEG2000 and JPEGXR, there was no significant difference in visual effects. As this algorithm is studied aiming at serious block effect produced under high compression ratio, the images with visual effect under the conditions of high compression ratio are compared. And it can clearly be seen that, the image block effect in edge region and flat region significantly reduced, while the visual effect is improved obviously.

In Fig. 3, the maximum activity value is 2.69, the minimum is 0.016, and the average value is 0.23. It can be seen that image classification results are different under different thresholds, and the threshold range of better visual effect is [2Tv, 4Tv] (Tv represents the average activity). Simulation experiments are made on many images, all satisfying this property.

After normalizing the weighted activity, the range of normalization is [0,1], the corresponding maximum value is 1, the minimum value is 0, and the average value

is 0.08. It can be seen from table 2 that, the content characteristics are different for different images, so the best threshold of activity classification is different. Therefore, the individual image block activity is very large, so in determining the threshold value, it is not feasible to determine the threshold with the maximum for reference, but to take the average value as the reference standard.

5. Conclusion

In the existing image encoding system based on block, in the case of high compression ratio, it has block effect. For the image encoding based on block, the transform and quantization is an independent process. For each block, filters with different parameters were used for filtering. The quantization error generated by block filtering results in that adjacent block boundaries are not continuous, and such discontinuous will have a blocking effect. In the practical application of image coding, the block effect seriously affects the visual quality of images. In order to improve the quality of image reconstruction, new coding algorithms emerge in endlessly, but good compression algorithms often increase the cost of algorithm complexity. The increase of algorithm complexity not only increases the coding time and reduces coding efficiency, but also increases the memory consumption. In the static image compression, bit rate control is very important. Without accurate rate control, it cannot meet the requirement of accurate compression, and it undoubtedly brings great drawbacks to the application of the algorithm. For JPEG standards, the standard only provides a unified quantization table, but cannot set compression ratio. JPEGXR, as a completely new static image compression standard, has the same problem.

Based on the above problems, we, based on the JPEG standard algorithm, design the non uniform quantizer based on content characteristics classification. In that the image in different regions have different sensitivities, the final quantization is adjusted, which significantly improves the image quality at a given bit rate.

It can be seen that the adaptive quantization compression ratio compression standard JPEG has higher PSNR. For different images, the improvement degrees are not the same. The images with high complexity have rich details, so the improvement degree of PSNR value is higher. For the images with rich texture, the improvement degree of PSNR value is smaller. And for the same image, with the increase of compression ratio, the improvement degree gradually declined.

According to the new adaptive quantization method of the human visual characteristics, adaptive quantization is achieved for the smooth area, edge area and texture region. Fist of all, in the selection of a fixed threshold, the image blocks are classified. And then according to the sensitivity characteristics of human eyes to different regions, different correction coefficients are determined, and the final quantization table is determined. Finally, simulation results are made, and the results show that the algorithm has good compression effect.

The main work and innovations of this paper are summarized as follows: The effects of the sensitivity characteristics of human visual on the compression effect are studied. A compression algorithm new idea combining image block classification is proposed. DCT blocks of image are divided into three regions with different

visual sensitivities, the simulation experiments of the algorithm are made, and the compression effect of this algorithm is compared with that of the JPEG algorithm.

The relationship between the local image features and image block classification is discussed. The traditional variance, information entropy and so on image features are abandoned. The sum of total energy of DCT transform AC coefficient matrix is taken as the activity of image sub block, to make classification of image blocks, and to realize adaptive quantization of image blocks after classification according to the eye visual characteristics.

Through a large number of simulation experiments, the adaptive compression method is proved to be effective in improving the quality of image compression from two aspects of subjective evaluation and objective evaluation. It can be used to compress the whole images, and can also be used to compress each sub image after one image is layered (partitioned).

References

[1] C. Liu, L. Yuan, Y. Li: *The design and realization of a image processing system based on DSP and USB*. Proc. 8th World Congress on Intelligent Control and Automation, 7–9 July 2010, Jinan, China, IEEE Conference Publications (2010), 6714–6718.

[2] J. Liu, N. Xu, Z. Mei: *The research on description method of computer operational geographic information in wargame*. Chinese Control and Decision Conference (CCDC), 17–19 June 2009, Guilin, China, IEEE Conference Publications (2009), 4517–4520.

[3] G. K. Wallace: *The JPEG still picture compression standard*. Bibliometrics, Magazine Communications of the ACM *34* (1991), No. 4, 30–44.

[4] J. M. Shapiro: *Embedded image coding using zerotrees of wavelet coefficients*. IEEE Transactions on Signal Processing *41* (1993), No. 12, 3445–3462.

[5] F. Dufaux, G. J. Sullivan, T. Ebrahimi: *The JPEG XR image coding standard [Standards in a Nutshell]*. IEEE Signal Processing Magazine *26* (2009), No. 6, 195–199, 204–204.

[6] S. Dodig, U. Kragh-Hansen, L. Minchiotti, M. Dodig, S. Grgič, B. Benko: *Quality of electrophoresis and isoelectric focusing images compressed using JPEG and JPEG 2000*. Biochemia Medica, Journal of Croatian Society of Medical Biochemistry and Laboratory Medicine *17* (2007), No. 2, 220–230.

[7] R. Rosenbaum, H. Schumann: *Compliant interframe coding for motion-JPEG 2000*. International Symposium on Visual Computing (ISVC), 29 November–1 December 2010, Las Vegas, NV, USA, Springer Nature, Lecture Notes in Computer Science *6455* (2010), 99–108.

[8] T. Bruylants, J. Barbarien, A. Munteanu, P. Schelkens: *Perceptual quality assessment of JPEG, JPEG 2000, and JPEG XR*. SPIE, Proceedings of Optics, Photonics, and Digital Technologies for Multimedia Applications *7723* (2010), paper e77230.

[9] T. Chen, H. Wu, B. Qiu: *Adaptive postfiltering of transform coefficients for the reduction of blocking artifacts*. IEEE Transactions on Circuits and Systems for Video Technology *11* (2001), No. 5, 594–602.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 283–292

Interaction between weak surrounding rock and supporting structure of shallow buried tunnel based on numerical simulation

Feng Xie[1]

Abstract. In order to design a better surrounding rock and supporting structure, the interaction between these of shallow buried tunnel based on numerical simulation is proposed. The stability of surrounding rock and support structure is tested by numerical simulation. During the construction of shallow tunnel, the stress and deformation characteristics of the supporting structure of the surrounding rock are the key to the stability of the tunnel, especially in the weak surrounding rock. The stability evaluation of the surrounding rock support structure is closely related to the design and construction of the tunnel. The selection of reasonable parameters of support structure is the guarantee of tunnel safety construction and operation. Therefore, it is necessary to study the interaction between soft rock and support structure of shallow tunnel. The mechanical and deformation characteristics of soft surrounding rock and supporting structures in shallow tunnels are studied by combining the measured data at the construction site with the finite element numerical simulation. The experimental results show the surrounding rock pressure of tunnel is not uniformly distributed over the supporting structure, and the pressure of surrounding rock in different parts is quite different. Based on the above finding, it is concluded that the difference between measured and calculated values of the stress of supporting structure is larger than that of surrounding rock stress. Therefore, the numerical simulation is suitable for calculating the degree of weakness of shallow separation tunnel.

Key words. Retaining structure, shallow buried tunnel, surrounding rock, finite element method, numerical simulation.

1. Introduction

With the rapid development of traffic in our country, the application of tunnel engineering in highway engineering is more and more (especially in mountain highway). Tunnels often encounter shallow buried strata, loose, weak and fractured surrounding rock belts and other unfavorable geological segments. Shallow buried tunnel has a shallow buried depth, and many surrounding rocks are weathered and

[1]Traffic Committee of Wulong, 408500, Chongqing, China

broken, and the bias is obvious. The stress distribution and deformation condition of surrounding rock and support structure are complex. Especially in the hilly areas where the depth and topography are large, the stress distribution of surrounding rock and the stress and deformation of lining are more complicated. There are many difficulties in the design of the tunnel construction. If the effective measures are not adopted, the accidents such as collapse of surrounding rock, deformation of supporting structure and lining cracking can easily occur. After years of engineering practice, the construction technology of shallow and weak surrounding rock tunnel has been greatly improved, but the design and construction are different in complex environment such as topography, geomorphology, engineering geology and hydrological conditions.

The main problems existing in shallow buried weak surrounding rock tunnel are [1]: In tunnel construction, the integrity of overlying soil in shallow and weak surrounding rock tunnel is seriously damaged. The stress characteristics, bearing mechanism and failure modes are complex. In order to ensure the reliability and safety of tunnel design and construction, it is necessary to study deeply. The construction of shallow weak surrounding rock tunnel is affected by the site. How to adjust the construction technology and support measures in time according to the site conditions will lay the foundation for the quality of the tunnel project, and the economic benefits are remarkable. Proper selection of test section and appropriate test scheme and test method have great influence on the accuracy and authenticity of data collected at the test site. There is a bias phenomenon in shallow soft rock tunnel, and how to select model parameters and mechanical model in numerical analysis is the key technology of numerical analysis.

The above problems are difficult to be solved in the design and construction of shallow and weak surrounding rock tunnel, which needs further study. Combined with Lantau Peak tunnel construction project, based on numerical simulation analysis and field test, the interaction between shallow buried weak surrounding rock and support structure is studied.

2. Literature review

In the late twentieth century, the method of combining engineering practice and finite element analysis was applied to the tunnel construction of weak surrounding rock. In the twenty-first century, a large number of special finite element software was applied in engineering practice, which made the design more scientific and reasonable. Combined with theoretical calculation, construction monitoring and practical experience, the construction design method which is consistent with geological exploration, design and construction is formed.

2.1. Status of foreign research

Through numerical simulation of the interaction mechanism of surrounding rock and support, Ozsan A analyzed the ground support interaction analysis (GSIA), the empirical method based on the rock mass rating (RMR) and the rock quality

index Q classification, and analyzed the reasonable supporting force of the tunnel [2]. Choi analyzed the stability of tunnel with weak surrounding rock by FLAG-3D, and optimized the relevant parameters of support structure [3]. Jeon compared the tunnel with weak interlayer by model test and numerical simulation. The influence of weak interlayer on the stability of tunnel is analyzed. Atkinson and other researchers have carried out the upper and lower limit method of limit analysis for the support reaction of unstable circular tunnel with cohesionless soil layer in the critical state. The calculation results are compared with the existing centrifuge test results, and the stress state of surrounding rock of shallow buried tunnel in cohesionless soil layer is described.

2.2. Status of domestic research

Fu Xinbin adopts the theory of elasticity and the theory of finite element analysis, and studies the mechanical behavior of tunnel construction and the whole stability of tunnel surrounding rock in shallow and weak surrounding rock. It is proposed that the settlement of vault caused by three steps parallel excavation method is less than that of CD method. In shallow buried weak rock, it is beneficial to keep the whole stability of surrounding rock. The reasonable construction method and construction organization of the tunnel that relies on the mountain are studied.

Taking the olecranon tunnel as the research background, the stability criterion of the surrounding rock supported structure is analyzed by Zhu Renjing. The elasto-plastic finite element method (FEM) is used to simulate the dynamic behavior of soft surrounding rock of tunnel. By analyzing the displacement and stress variation of surrounding rock support structure during construction, the stability and safety of surrounding rock support structure during tunnel construction are evaluated.

Liu Qinqin simulated the construction process of weak surrounding rock tunnel with ABAQUS software, and obtained the change law of displacement and stress during the excavation of surrounding rock and initial support structure. The mechanical properties of the contact face between surrounding rock and initial support structure are deeply studied, and a series of conclusions are obtained. For example, it is unreasonable to increase the support effect of surrounding rock by increasing the length of bolt or increasing the thickness of shotcrete layer, so the design parameter of shotcrete bolt support should be taken into consideration [4].

In view of the buried depth of the Foling tunnel and the poor stability of the surrounding rock, Li Chunsheng uses ANSYS analysis software to simulate the soft surrounding rock of tunnel based on three-dimensional numerical. The analysis software includes the reserved core soil method, CD method, CRD method, double side wall guide method. The deformation of surrounding rock, the distribution shape and development law of plastic zone, the stress characteristics of surrounding rock and the variation law of initial support stress are analyzed. The influence mechanism of tunnel construction on the stability of weak surrounding rock is studied.

Xu Daihong uses finite element method to simulate the pre-reinforcement and initial support structure of shallow buried bias tunnel by step method. The surrounding rock stability and construction mechanics characteristic of shallow bias tunnel are

analyzed, and the technical scheme of the reasonable construction of sharp slope of tunnel engineering is proposed.

Yang Feng, Yang Junsheng and Zhao Lianheng have adopted rigid upper bound method and upper bound finite element method to calculate and analyze the stability of surrounding rock of shallow tunnel under different conditions. The calculation chart of stability coefficient of tunnel surrounding rock under undrained condition and the calculation chart of reaction coefficient of tunnel support under drainage condition are obtained, which can be referred to for practical reference [5]. At the same time, in the light of the large deformation and collapse of tunnel surrounding rock and support structure during the excavation of the sub clay layer of Yunyang mountain tunnel, the stability analysis of tunnel surrounding rock is carried out by using the upper bound finite element method of limit analysis.

Based on the construction of the shallow surrounding tunnel in the coastal area, using the finite element analysis software (FLAG), the nonlinear numerical simulation of partial excavation is carried out by Jiao Cang, Zhu Jianglin and Fan Peng. The mechanism of large deformation of tunnel weak rock during excavation is studied.

3. Experimental design and numerical simulation

3.1. Field experiment

In order to study the interaction between soft rock and support structure of shallow tunnel, field test is based on Lantau Peak tunnel [6]. Through the selection of two test sections in the tunnel, the pressure box and the concrete strain gauge are embedded, and the deformation of surrounding rock is monitored, and the data of the test element are collected. Mechanical and deformation characteristics of surrounding rock and support structure in tunnel construction of shallow weak surrounding rock are analyzed to ensure safe construction and operation of tunnel.

Two sections of ZK82+008 and ZK82+311 of Lantau Peak tunnel are selected as test section. The two test sections are in the left line of the tunnel and are V grade surrounding rock. ZK82+008 is at the entrance of the left line of the tunnel, taking the A section. ZK82+311 is at the exit line of the left line of the tunnel, taking the B section. Two the rock mass of test section is more broken. The fracture development of surrounding rock and the basic information of test section are shown in Table 1.

Table 1. Basic information of test section

Test section	Pile No.	Buried depth	Surrounding rock grade	Initial support	Second support
A section	ZK82+008	40 m	V grade surrounding rock	26 cm thick C25 shotcrete ⌀8 steel mesh (spacing: 20×20 double deck) I18 I-beam	45 cm thick C30 reinforced concrete
B section	ZK82+311	15 m			

3.2. Numerical simulation

The buried depth of Lantau Peak tunnel is very shallow. The surrounding rock is broken and mixed with mudstone, and the strength is low [7]. Complicated engineering environment and geological condition bring huge safety hazard to the project construction. Combined with the engineering example of Lantau Peak tunnel, based on numerical simulation and analysis, the interaction between soft surrounding rock and support structure of shallow tunnel is studied, and the stability of tunnel surrounding rock and lining structure is evaluated.

The accuracy of finite element numerical calculation depends mainly on the rationality of the constitutive model and the accuracy of the calculated parameters. The plastic model and the elastic model of tunnel surrounding rock and lining structure are analyzed respectively. The ideal elastoplastic linear Mohr-Coulomb yield criterion is used to analyze the physical and mechanical characteristics of the interaction between the surrounding rock and the lining structure [8].

The tunnel is a slender structure, that is, the cross section of the tunnel is very small relative to the length of the longitudinal direction. It can be assumed that there is no displacement in the longitudinal direction under the action of the surrounding rock, and only lateral displacement occurs. Therefore, On the basis of the actual construction process and supporting measures, the tunnel model is simplified as an elastic-plastic plane strain model during the numerical simulation, and the plane model is extended along the longitudinal length in combination with the specific circumstances of the tunnel.

In view of the characteristics of shallow weak surrounding rock tunnel, based on numerical calculation, the stress distribution characteristics of surrounding rock and support structure, the law of surrounding rock settlement deformation and the law of convergence of characteristic points surrounding rock are analyzed. Detailed analysis is shown in Table 2.

Table 2. Analysis content

Construction method	Analysis content
Bench method	Stress of surrounding rock of tunnel Settlement of tunnel surrounding rock Support structure stress Feature point edge convergence

4. Results and analysis

As it can be seen from Figs. 1 and 2, during this period, the surrounding rock pressure change is relatively large. With the initial injection of concrete strength gradually increased to the design strength, the surrounding rock pressure is basically in a stable state, and increases little with time. Overall, the adjustment of A section surrounding rock pressure eventually tends to be stable and takes about two months.

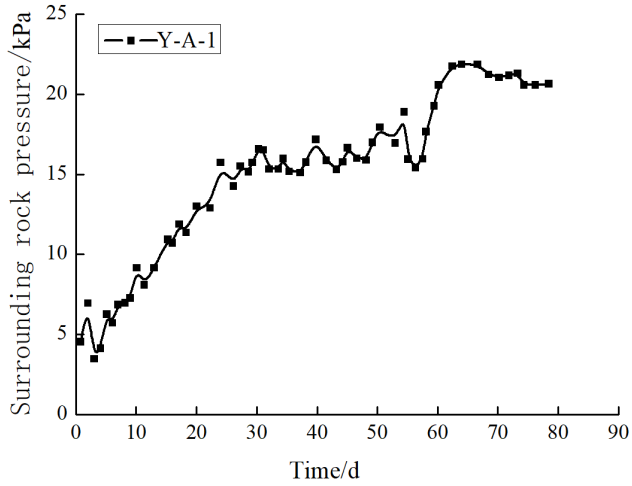


Fig. 1. Variation law of surrounding rock pressure with time at vault of A section

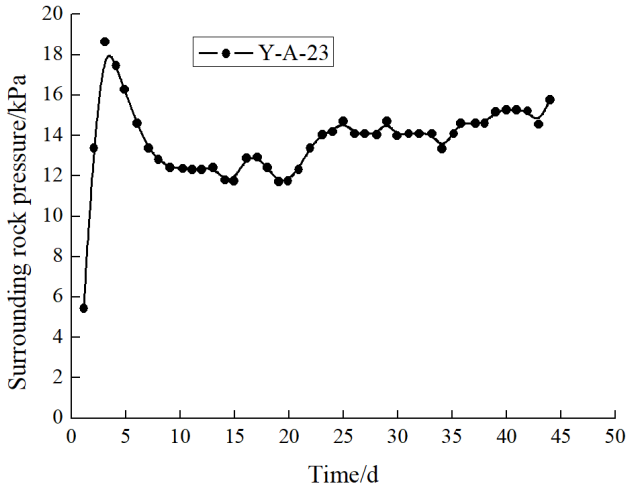


Fig. 2. Variation law of surrounding rock pressure with time under inverted arch of A section

As shown in Figs. 3 and 4, the deformation time of the surrounding rock of the vault is longer than that of the side wall and the inverted arch [9]. The stress of surrounding rock gradually stabilized after 13 days' excavation. However, the stress of surrounding rock gradually stabilized after 15 days' excavation of the side wall and the inverted arch. The adjustment of the surrounding rock pressure of the B section eventually tends to be stable and takes about 2 months.

Through the selection of two typical sections in tunnel, the interaction between soft rock and support structure of shallow tunnel is studied, and the following con-

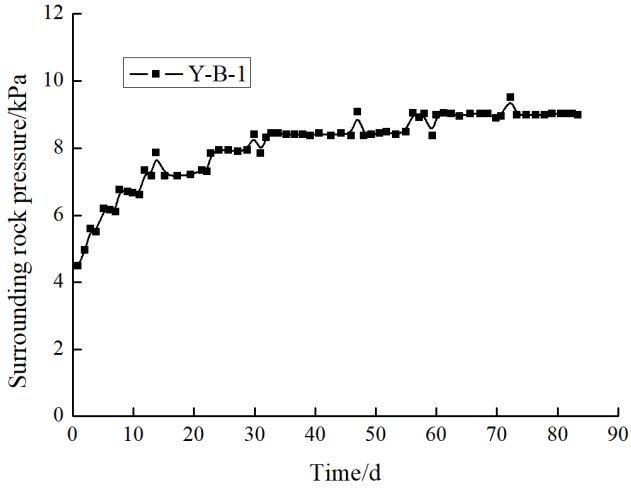


Fig. 3. Variation law of surrounding rock pressure with time at vault of B section

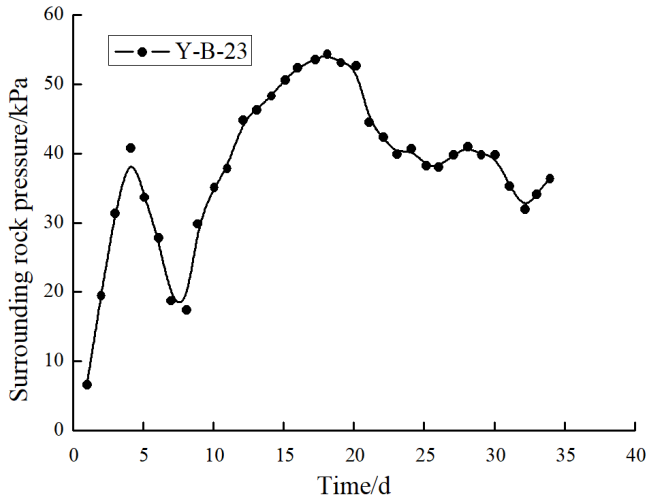


Fig. 4. Variation law of surrounding rock pressure with time under inverted arch of B section

clusions are obtained [10]. According to the monitoring data, the surrounding rock pressure of tunnel is not uniformly distributed over the supporting structure, and the pressure of surrounding rock in different parts is quite different. The pressure concentration of surrounding rock will appear in some areas, and the surrounding rock pressure of invert arch and arch waist is larger. The surrounding rock pressure, the initial lining, the two lining, the two-layer support, the pressure and the concrete stress in the two lining are smaller. It is much smaller than the compres-

sive strength and the tensile strength of the original shotcrete and the two-lined reinforced concrete. It shows that the tunnel structure is safe and stable.

4.1. Numerical calculation results

As shown in Figs. 5 and 6, through numerical simulation, the curves of the stress of the tunnel vault surrounding rock with the overlying rock thickness and the vertical displacement of vault with the overlying rock thickness are obtained.

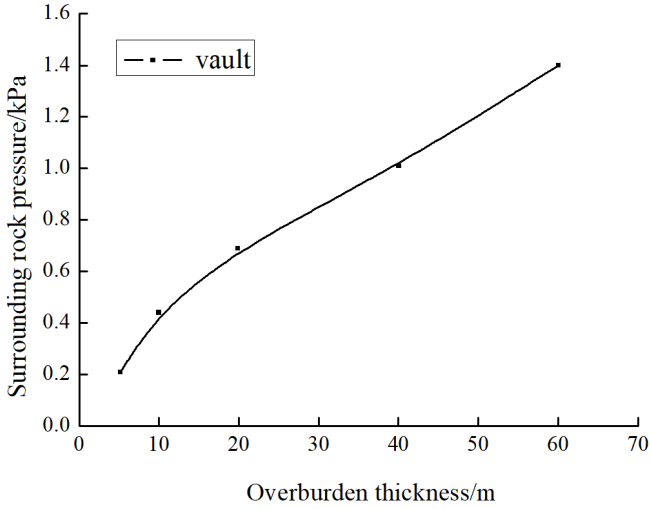


Fig. 5. Curves of stress of surrounding rock of tunnel vault with overlying rock thickness

Through the numerical calculation and analysis of the interaction between the shallow buried weak rock and the support structure, the following conclusions are obtained. When the thickness of the overlying rock on the tunnel is from 5 m to 20 m, the stress of the surrounding rock of the vault increases accordingly, and the stress increases from 0.2032 MPa to 0.6870 MPa. The deformation of the surrounding rock of vault decreases correspondingly, and the settlement of vault decreases from 0.4368 m to 0.0536 m. When the thickness of the overlying rock on the tunnel is from 20 m to 60 m, the deformation of surrounding rock increases, but the amplitude of the increase is small. The dome is increased from 0.0536 m to 0.0677 m, and it is concluded that the thickness of the overlying rock mass is 20 m, which is the limit thickness of arching effect.

5. Conclusion

By comparing and analyzing the results of field test and numerical simulation, it is shown that the difference between measured and calculated values of the stress of supporting structure is larger than that of surrounding rock stress, except for

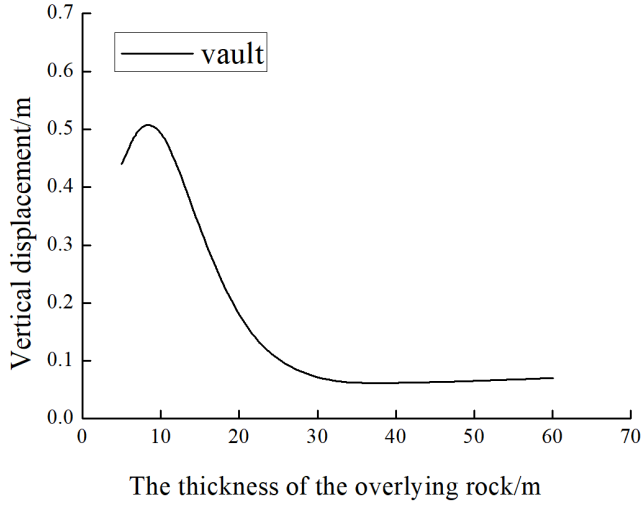


Fig. 6. Vertical displacement of vault with thickness of overlying rock

the abrupt change of 3 points of B section. In numerical calculation, the ideal is to close the surrounding rock with the initial lining, the initial lining and the two lining. But in fact, the tunnel is not smooth. There is no numerical calculation of the effect of the initial lining and the surrounding rock. The first lining and the two lining are actually attached to the waterproof board, and the gap between them leads to the deformation and coordination of the lining under the stress of the surrounding rock. The stress from the surrounding rock is reduced, so the numerical value of the stress of the supporting structure is larger than the measured value. In the B section vault, the measured stress is 0.006 MPa, while the numerical value is 0.025 MPa compressive stress. In the course of pouring concrete, the pressure box is inclined and offset by the impact of the concrete, resulting in the stress distortion of the supporting structure.

The measured data and numerical simulation values are compared from three aspects: deformation of surrounding rock, surrounding rock stress and stress of supporting structure. The two data are different in the same order. Through the finite element analysis, it can basically reflect the actual force and deformation characteristics of the tunnel, and put forward scientific guidance for design and construction.

References

[1] T. Nakai, L. Xu, H. Yamazaki: *3D and 2D model tests and numerical analyses of settlements and earth pressures due to tunnel excavation.* Soils and Foundations *37* (1997), No. 3, 31–42.

[2] A. Özsan, H. Basarir: *Support capacity estimation of a diversion tunnel in weakrock.*

Engineering Geology *68* (2003), Nos. 3–4, 319–331.

[3] S. O. Choi, H. S. Shin: *Stability analysis of a tunnel excavated in a weak rock mass and the optimal supporting system design.* International Journal of Rock Mechanics and Mining Sciences *41* (2004), Supplement No. 1, 876–881.

[4] S. Jeon, J. Kim, Y. Seo, C. Hong: *Effect of a fault and weak plane on the stability of a tunnel in rock-a scaled model test and numerical analysis.* International Journal of Rock Mechanics and Mining Sciences *41* (2004), No. 3, paper 486.

[5] J. H. Atkinson, D. M. Potts: *Stability of a shallow circular tunnel in cohesionless soil.* Géotechnique *27* (1977), No. 2, 203–215.

[6] S. C. Hsu, S. S. Chiang, J. R. Lai: *Failure mechanisms of tunnels in weak rock with interbedded structures.* International Journal of Rock Mechanics and Mining Sciences *41* (2004), Supplement No. 1, 670–675.

[7] G. Gioda, S. Sakurai: *Back analysis procedures for the interpretation of field measurements in geomechanics.* International Journal for Numerical and Analytical Methods in Geomechanics *11* (1987), No. 6, 555–583.

[8] S. Dalgic: *The influence of weak rocks on excavation and support of the Beykoz Tunnel, Turkey.* Engineering geology *58* (2000), No. 2,137–148.

[9] W. Q. Ding, Y. F. Qiao,Y. L. Jin, Q. Z. Zhang: *A novel support system for shallow buried caverns based on the mining method.* Journal of Geophysics and Engineering *13* (2016), No. 1, 123–132.

[10] S. Koneshwaran, D. P. Thambiratnam, C. Gallage: *Blast response and failure analysis of a segmented buried tunnel.* Structural Engineering International *25* (2015), No. 4, 419–431.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 293–302

Matching technique of image extraction and machine learning based on face recognition

Yunfei Li[2,3], Zhaoyang Lu[2], Jing Li[2]

Abstract. In order to design a kind of new technology used for feature extraction, a matching technique of image extraction and machine learning based on face recognition is proposed. As a biometric identification technology, human face recognition involves many fields such as artificial intelligence, pattern recognition, computer vision, image processing and analysis, image coding, computer graphics and so on. Compared with other identification methods, face recognition technology has some advantages, so it has a wide range of applications. It is mainly used in the field of national security, military security and public security, intelligent access control, intelligent video surveillance and other fields. A video based real-time face recognition system is designed in this paper. In this system, we adopt video capture technology. At the same time, the location and size of the human face can be detected by the function provided by the visual library, and the human face image can be extracted. Finally, the device can achieve 5 frames per second of recognition speed, and fully meet the application requirements. The experiment results show that the recognition rate of facial expression is improved after illumination normalization. After the improved histogram equalization method is used, the recognition rate is higher than that of the traditional histogram equalization method. Based on the above finding, it is concluded that the information security technology based on face information recognition will be widely used and benefit mankind.

Key words. Face recognition, feature extraction, video capture technology.

1. Introduction

Face recognition is a technique that uses computer to analyze face images and extract valid information from them so as to identify people's identity [1]. Compared with other mature biometric methods, the face recognition system has the advantages of low cost, no invasion and no human participation. But there are also

[1]The authors acknowledge the National Natural Science Foundation of China (Grant: 51578109), the National Natural Science Foundation of China (Grant: 51121005).

[2]School of Telecommunications Engineering, Xidian University, Xi'an, 710071, China

[3]School of Network Security and Information, Weinan Normal University, Weinan, 714099, China

some disadvantages: for example, it is vulnerable to the environment, hairstyles, age, make-up and jewelry. The human face is represented by a geometric feature vector, and the classifier is designed according to the hierarchical clustering in pattern recognition. This requires that the selected geometric feature vector has a certain unique, and can reflect the difference between different faces. At the same time, it also has a certain flexibility to eliminate time span, light and other effects. Geometric feature vector is a feature vector based on the shape and geometric relationship of human facial organs. Its components usually include the Euclidean distance and curvature angle between two points specified by the human face. Deformable template face has a better method of describing active contour, which makes the facial organ model parameterized. However, there are two problems [2]. The selection of weighting coefficients at various costs is empirical and difficult to generalize. The process of generalization is time-consuming and not suitable for real-time applications. he main work of this paper is to study the basic theory and key techniques of face detection and recognition. The problems of face detection, geometric normalization, feature extraction, feature selection and face recognition in video based color images and gray images are discussed. And the design and implementation of a video based real-time face recognition system is proposed.

2. Literature review

The earliest research of face recognition technology can be traced back to 1950s, and the researchers at that time were mainly involved in the field of social psychology. By 60s, a number of engineering documents had been published. But the research of automatic face recognition in real sense started from Kanade and Kelly in 1970s. The techniques used at that time were basically typical pattern recognition techniques, such as classifying and identifying the distances between important facial feature points. With the development of computer technology, from the 80s to the early 90s, face recognition technology has been greatly developed, and has entered the field of practical application. At this stage, statistical recognition methods based on facial appearance have been greatly developed. Experiments proposed by Eigenfaces and Fisherfaces on large scale face databases have obtained fairly good results. At the same time, the recognition method based on facial features has gradually developed. This method is not sensitive to the changes of light and angle of view, and the location of human faces. It is beneficial to improve the recognition rate, but the feature extraction methods used are not mature enough and reliable. Since the late 90s, a number of commercial face recognition systems have gradually entered the market. In recent years, face recognition has developed rapidly in the world as a computer security technology [3]. At this stage, more research has been focused on video based face recognition. The image is captured by a computer camera, then the position of the human face is detected from the video, and the captured face image is pretreated with light, size and position. Then, the features of human face are extracted and useful features are selected. Finally, the selected features are trained and identified to determine the face category of the feature and identify the identity of the person.

3. Methods

3.1. Overview of face detection

A standard face template is first predefined and parameterized. Then, an input image is provided, respectively, to calculate the facial contours, and the correlation between the eyes, nose, and mouth and the template. If the correlation is within a certain threshold, the input image is judged to be human face. Otherwise, it can be ruled as a non-face. Sakai and other researchers modeled with the eye, nose, mouth, and human face contours and detected the positive faces in photos. Each sub template is defined according to the line segmentation, and the line of the input image is extracted based on the maximum gradient. Then, the extracted line is matched with the sub template, and the relation between the sub image and the contour template is computed to detect the candidate area of the face, and the other sub templates are used to match the candidate region [4].

Face detection is to determine whether the human face exists in the input image. If there is a human face, then the number and the location and size of each face are determined. As a key technology in face information processing, face detection is the first part of face recognition system. Whether the location of a human face can be detected from an image or video stream correctly has a significant impact on the subsequent feature extraction and recognition. At the same time, the research of face detection is of great academic value. Human face is a kind of natural structural object with complex details. The challenge of this kind of target detection problem is that the difference position of human face and camera will lead to difference face image, such as the acquisition of human face image in front, and side and 45-degree angle. There may be glasses, beards and other appendages on the face, including their different shapes, sizes, and colors. At the same time, it is also influenced by people's appearance, facial expression and skin color, and the conditions of photography, such as intensity of light, light distribution and characteristics of cameras. Therefore, the methods to solve these problems need to be found to construct a face detection and tracking system successfully, which will provide important implications for other similar complex pattern detection problems. The method of detecting human faces from an image can be divided into the following categories: skin color based methods, template matching methods, feature based methods, and neural network based methods.

3.2. Traditional histogram equalization algorithm (HE)

The gray histogram of the image represents the number of pixels with each gray level in the gray image, and reflects the frequency of each gray level in the image. It is one of the basic statistical features of the image. Histogram equalization is the use of histogram statistics to modify the histogram. The gray values of each point in the image are changed by some corresponding relation, so that the histogram of the transformed image is relatively uniform and flat with respect to the original image, and the gradation is clear. Thus, the purpose of image enhancement and reducing

the influence of illumination on recognition are achieved. If the variable r represents the gray level of the pixel in the image, the r is normalized. That is, r is a decimal between (0, 1), where $r = 0$ stands for black and $r = 1$ stands for white. For a given image, each pixel obtained in [0, 1] intervals within the gray scale is random [5]. R is a random variable, which can be used to represent the gray distribution of the original image by the probability density function $Pr(r)$. The gray distribution feature of an image can be seen from the distribution of gray level. If the gray values of most pixels in the image are in the region close to 0, the whole image is dim, otherwise, the whole image is brighter. The ideal histogram equalization process is designed to make the histogram of the processed image be flat. That is, each gray level has the same frequency of occurrence, so that each gray level has a uniform probability distribution, so that the image looks clearer. Therefore, it is necessary to find a transform relation $S = T(R)$, so that the new gray histogram after transformation is more straight than the histogram before transformation. The traditional histogram equalization adopts the cumulative probability distribution function of R as the transfer function, that is,

$$S = T_r = \int_0^r Pr(w)\,\mathrm{d}w\,. \tag{1}$$

In formula (1), w is the integral variable, and $\int_0^r Pr(w)\,\mathrm{d}w$ is the cumulative probability distribution function of r. Let r_k is used to represent discrete gray values, and $Pr(r_k)$ is used to represent $Pr(r)$. Then

$$Pr(r_k) = \frac{n_k}{n}, \quad (0 < r_k < 1, \quad k = 0, 1, 2, \ldots, L-1)\,. \tag{2}$$

In the formula, n_k is the number of pixels in this image and n is the total number of pixels in the image. Symbol L represents the gray value of the image, which is divided into L classes. The discrete form of the transformation function is

$$S_k = T_{rk} = \sum_{j=0}^{k} \frac{n_j}{n} = \sum_{j=0}^{k} Pr_{rj}, \quad 0 < r_k < 1, \quad k = 0, 1, 2, \ldots, L-1)\,. \tag{3}$$

It can be proved that when the histogram of the image is uniformly distributed, the amount of information contained is the largest.

3.3. Improved histogram equalization algorithm (NHP)

In formula (3), for a 256×100 with 256-gray level image, if the pixel number of grayscale k in original image is less than 100, then, in the converted image, there will be a combination of the adjacent gray levels, and it is impossible to have its own gray value. This results in that the gray range of the transformed image is difficult to reach the maximum gray range permitted by the image format. At the same time, because of the "gray phagocytosis", it is also easy to cause the loss of image information. Especially when the histogram distribution of original image is very

uneven, such as the human face image acquired in bright or very dark environment, the level of image is worse and the information loss is more serious [6].

Jiang Duan and GuoPing Qiu proposed an improved histogram modification algorithm. Based on the RGB color space of image, this method uses high resolution gray value, which floats to indicate the original image pixel gray value, so as to effectively reduce the influence caused by the annexation of gray. Algorithm principle: the lower form is used as the gray transformation formula of RGB color space, in which Lu is the floating-point number type. In our experiment, Lu takes two points after the decimal point

$$\mathrm{Lu}(x,y) = 0.299 \times R_{(x,y)} + 0.587 \times G_{(x,y)} + 0.144 \times B_{(x,y)}\,. \tag{4}$$

Local histogram equalization (LHE): Histogram equalization is the transformation of the distribution of pixel values across the entire range of images. However, the best results achieved in the global range may not be optimal in local areas. The local histogram equalization adopts the same algorithm as the traditional global histogram equalization. It changes the image area of the algorithm, and decomposes the image into a number of $m \times n$ windows. The region of each window is balanced by histogram, so that the images in each region can get the best results. In the experiment, the original image is divided into M rows and N columns, and the $m \times n$ regions are balanced by histogram respectively.

Adaptive histogram equalization (AHE): The adaptive histogram equalization algorithm uses the histogram of the local correlation region with the pixels as the center. It applies the result of histogram equalization in this region to the pixel point method instead of the histogram of the whole image area to modify the pixels. That is to say, for each pixel of the image, the first thing to do is to find a region (such as the window of the $m \times n$ centered at that point). In this area, the mapping function is calculated according to the histogram equalization algorithm described above, and the new values of the pixels are calculated according to the mapping function. Adaptive histogram equalization needs to compute the histogram of each pixel of the image. Therefore, the amount of computation is considerable. However, since the corresponding regions of adjacent pixel points are related, their correlation can be used to optimize the computational process [7].

4. Results and discussion

In the experiment, the Gabor filter is used to extract the facial features, and the K mean distance classifier is used to recognize the features. PCA and PCA+LDA are used to reduce the features of the two methods. The training and testing samples are derived from the AR face database, and the recognition rate is shown in Fig. 1. 134 people in the AR database are selected, each of whom contains 7 frontal images with no mask on the face [8]. The recognition rates using various feature selection methods are shown in Figs. 2 and 3.

The experimental results show that the recognition rate of facial expression is improved after illumination normalization. After the improved histogram equalization

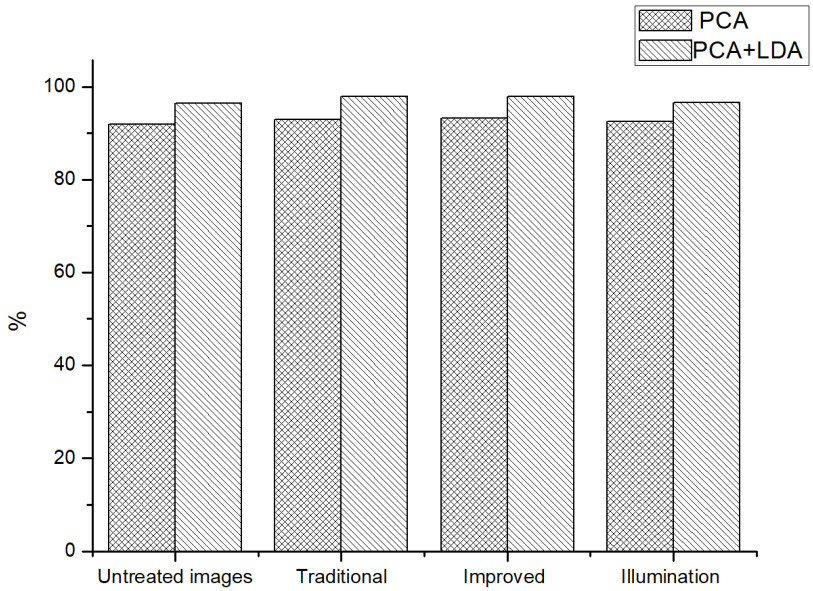


Fig. 1. Recognition rate of different illumination normalization in face recognition system

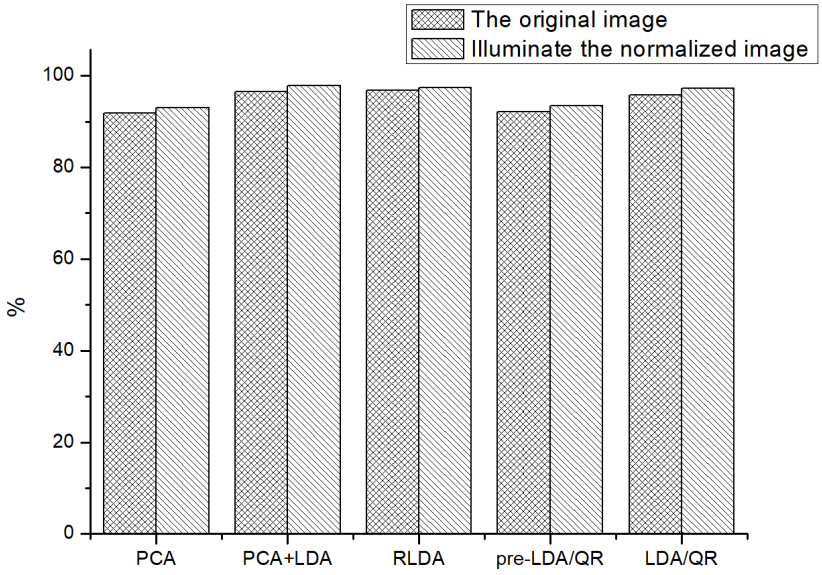


Fig. 2. Comparison of different feature selection with global Gabor

method is used, the recognition rate is higher than that of the traditional histogram equalization method. However, the improved histogram equalization requires that

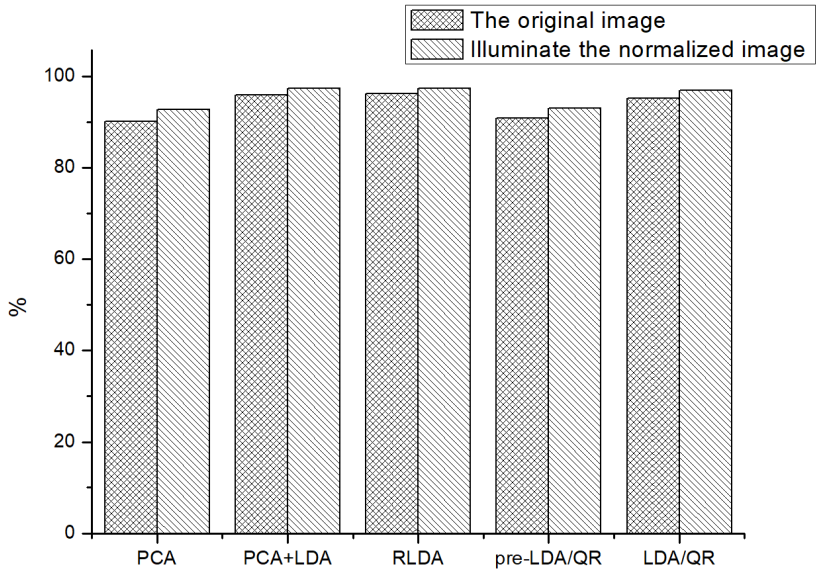


Fig. 3. Comparison of different feature selection with local Gabor

the original image contains RGB color space information, which limits its range of applications. However, the illumination normalization method based on affine transformation illumination model is poor in real-time performance, and the processing speed increases with the complexity of the filter used. At the same time, in the experiment, the Gabor feature is reduced to dimension. This process may remove the feature vector that represents the illumination property of the image. Therefore, the effect of different illumination normalization methods has some limitations. The use of PCA+LDA, RLDA5 and LDA/QR feature selection method can get higher recognition rate [9]. Among them, RLDA needs to spend more computing time, and PCA+LDA is the second, LDA/QR has the lowest computational complexity. Since the number of training samples in our system is relatively small, it can be seen from the above analysis that the computational complexity of PCA+LDA is very different from the computational complexity of LDA/QR. And the recognition rate of PCA+LDA feature selection method is 0.2 to 0.75 percentage points higher than that of LDA/QR. Therefore, PCA+LDA is adopted as dimension reduction feature selection method in this system.

A real-time face recognition system based on video is designed and implemented. In this system, DirectShow video capture technology is adopted. After capturing an image, firstly, the location and size of the human face are detected by the function provided by Intel OpenCV visual library, and the human face image is extracted. Then the face image is pre-processed, including eye location and geometric normalization. Then the feature extraction and feature reduction is combined to extract the low dimensional Gabor features of human face. Finally, the feature is input into the classifier to classify the feature, and the classifier outputs the recognition

result. The classifier used in this system is Euclidean distance classifier. The system diagram is shown in Fig. 4.

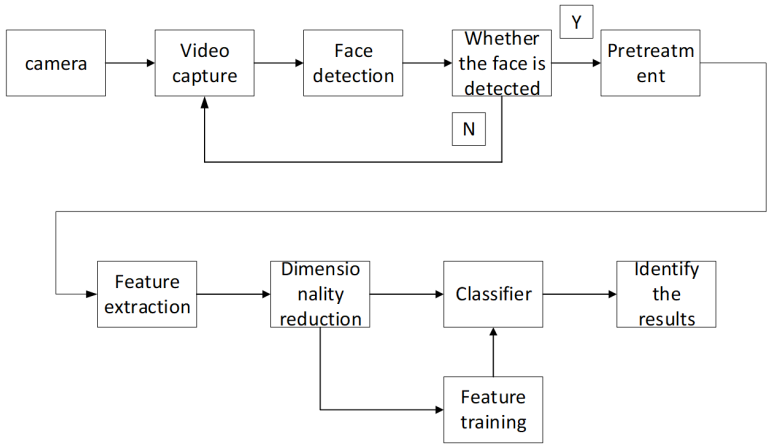


Fig. 4. Framework of real-time face recognition system

In the real-time face recognition demo system based on video, a common USB camera is used, and the resolution of the image is 320×240. In order to improve the speed of face detection, we stipulate that only the face not less than 20×20 is detected. In the face detection part, multiple faces is detected at the same time, but for the convenience of identification, the first face is only recognized. At the same time, in order to improve the accuracy of recognition, multi frame recognition results voting strategy is adopted (as shown in Fig. 5). The recognition results of 5 consecutive face images are classified into the categories with the highest number of recognition results. As the final recognition results, the input feature classification results are output.

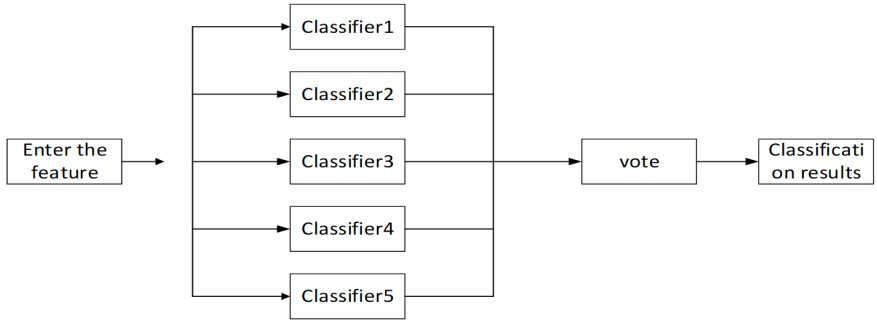


Fig. 5. Strategy of classifier

Microsoft Visual C++NET is selected as the development platform to complete the face image acquisition, detection, preprocessing, feature extraction and selection, identification of the whole process. Because C++ has the characteristics of high code efficiency, convenient memory and powerful function, it can make up for the

inconvenience of large amount of data, complex operation and long time-consuming in image processing. Matlab is selected as a tool for feature selection training, and Matlab's powerful matrix computing function is used for offline computation. The establishment of real-time face recognition system is image acquisition and regular USB camera. The image separation rate is 320×240 (that is, the size of the camera monitor area). The CPU size is P4 3OG. Memory is 512 MB. The software development platform is Microsoft Visual C++ NET. Table 1 presents the average processing speed of a real-time face recognition system. Since image capture and subsequent processing processes are synchronous and interactive, the processing speed given in the previous table is the average processing speed for each process separately. As can be seen from Table 1, the entire process of the system takes about 200 ms time, and has basically met the requirements of real-time systems.

Table 1. Processing speed

Image capture	Face detection	Preprocessing	Feature extraction and selection	Classification recognition
$<$29 ms	$\sim$100 ms	$<$29 ms	50$\sim$ 100ms	$<$10 ms

5. Conclusion

The structure of face recognition system is divided into image capture, face detection, image preprocessing, feature extraction and feature selection, and face recognition. The problems of face detection, geometric normalization and feature selection are discussed. Real time face recognition demo system based on video is designed. It uses a common USB camera, and real-time recognition of the front of the camera face is realized through pretrained templates It has the advantages of fast recognition speed and high accuracy. It can achieve real-time monitoring effect, and can also be used in practical applications.

With the development of micro electronics, computer and network technology, traditional identification methods such as magnetic cards, IC cards and passwords are facing severe tests. The appearance and rapid development of face recognition technology are expected to solve these problems in the near future. Because any person's facial information features are always followed by their own, they do not worry about losing face and being stolen. In addition, as face features are relatively complex and difficult to copy, it has higher reliability and security as a security key. Therefore, with the rapid progress of people's in-depth research for face recognition technology and computer science and technology, in the near future, the information security technology based on face information recognition will be widely used and benefit mankind.

References

[1] J. Cao, Y. Zhao, X. Lai, M. Eng Hock Ong, C. Yin, Z. X. Koh, N. Liu: *Landmark recognition with sparse representation classification and extreme learning machine.* Journal of the Franklin Institute *352* (2015), No. 10, 4528–4545.

[2] J. Lu, G. Wang, P. Moulin: *Localized multifeature metric learning for image-set-based face recognition.* IEEE Transactions on Circuits and Systems for Video Technology *26* (2016), No. 3, 529–540.

[3] Z. Lei, M. Pietikäinen, S. Z. Li: *Learning discriminant face descriptor.* IEEE Transactions on Pattern Analysis and Machine Intelligence *36* (2014), No. 2, 289–302.

[4] J. Lu, V. E. Liong, X. H. Zhou, J. Zhou: *Learning compact binary face descriptor for face recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence *37* (2015), No. 10, 2041–2056.

[5] G. Deepak, J. Lee: *Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines.* Sensors *13* (2013), No. 6, 7714–7734.

[6] J. Tompson, M. Stein, Y. Lecun, K. Perlin: *Real-time continuous pose recovery of human hands using convolutional networks.* Journal ACM Transactions on Graphics (TOG) *33* (2014), No. 5, Article No 169.

[7] R. Girshick, J. Donahue, T. Darrell, J. Malik: *Region-based convolutional networks for accurate object detection and segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence *38* (2016), No. 1, 142–158.

[8] Y. T. Yu, G. H. Lin, I. H. R. Jiang, C. Chiang: *Machine-learning-based hotspot detection using topological classification and critical feature extraction.* IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems *34* (2015), No. 3, 460–470.

[9] S. Puttinaovarat, P. Horkaew: *Urban areas extraction from multi sensor data based on machine learning and data fusion.* Pattern Recognition and Image Analysis *27* (2017), No. 2, 326–337.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 303–314

A study of sustainable design for abandoned coal mines' ecological remediation

Yang Yan-ping[1], Luo Fu-zhou[1]

Abstract. To explore ecological remediation plans for abandoned coal mines in line with the sustainable development perspective, this paper, based on the sustainable development theory, sets Wangshi'ao coal mine as an example and studies application models and ecological elements allocations of abandoned coal mines' ecological remediation. Through investigations and calculations, this paper finds out that in ecologically remediating abandoned coal mines, sustainable designs can facilitate abandoned coal mines evolving into zonal nature ecosystem and accelerate nutrient accumulation of abandoned soil, which can improve abandoned coal mines' ecosystem services value. Research results can provide references to studies on theory and practice of ecological remediation planning and design of abandoned coal mines.

Key words. Abandoned coal mine, ecological remediation, sustainable design, Wangshi'ao coal mine.

1. Introduction

Coal is the largest and most widely distributed conventional energy in the world, accounting for 25 % of the world's primary energy consumption. For a long time, coal mining has been damaging the original ecosystem stability and natural landscape homogeneity, which leads to environmental degradation and lots of abandoned coal mines. Abandoned coal mines refer to lands that are damaged and cannot be used without remediation because of the exploitation of coal resources. After mining for a long time, abandoned coal mines will re-pollute and destroy the surrounding environment through dust, collapse, landslide and erosion, which will further deteriorate the surrounding environment [1].

Environmental hazards and social contradictions caused by coal mining have attracted wide attention. Long before World War II, in the mining areas, the United States and some developed countries had begun ecological remediation after mining. For example, in 1920s, the United States conducted re-planting experiment in Indi-

[1]School of Management, Xi'an University of Architecture & Technology, Xi'an, 710055, China

ana's coal gangues and German afforested the Lignite mine in open pit coal mine etc. China's ecological remediation of abandoned coal mines embarked on a legal track in the late 1980s, beginning an organized land reclamation stage, such as land reclamation in Huaibei coal mine subsidence area in Anhui province. At present, with the development of world economy and coal industry, abandoned coal mines' ecological remediation has become a serious problem, waiting to be solved during social and economic development.

2. Literature review

Recently, domestic and foreign scholars have conducted massive researches on ecological remediation in coal mine areas. Coal mines' ecological remediation refer to comprehensive restore surface damage caused by coal resources mining, and improve ecological system's structure and function in the coal mine area, so that the ecosystem can maintain itself and reach a new ecological balance. Because of long-term coal mining and grazing activities, similarities between remediation areas' soil seed bank and nearby control mountain's surface vegetation are relatively high [2], so using these indigenous species to ecologically remediate abandoned lands in coal mines can reduce costs and avoid ecological risks associated with non-native species [3]. In coal mining areas, there is a correlation between plant species' diversity and soil's physical and chemical properties [4], and different tree species have varying functions in improving rhizosphere soil quality, which provides an important basis for selecting suitable tree species used for ecological remediation in abandoned coal mine areas [5].

Compared with undisturbed soil, after coal mining and soil reclamation, soil's microbial diversity and biomass decrease significantly [6]. To promote soil amelioration and comprehensive land management in abandoned coal mines, there have been massive studies in spatial distribution and variation characteristics of soil heavy metals in abandoned coal mines [7], soil reconstruction, fertility management and microbial flora regulation in reclamation and ecological remediation of abandoned coal mines [8], and soil's physical and chemical properties changes in abandoned coal mines under different remediation modes [9], results from which provide scientific basis for vegetation restoration in abandoned coal mines.

Existing studies are focusing on plant species preference and soil matrix improvement, and the planning and design practice of abandoned coal mines' ecological remediation have not been widely touched upon, which results in demonstration projects' lead and radiation. Therefore, this paper applies the sustainable design concept to abandoned coal mines' ecological remediation, and combines practice procedures in ecologically remediating Wangshi'ao coal mine to the sustainable design model and its application rules in ecologically remediating abandoned coal mines.

3. Research methodology

Sustainable design, based on ecology, builds a sustainable ecological structure suitable for a variety of environments by following the natural patterns of development. In recent years, sustainable design has gradually been applied to agricultural park planning, urban and rural landscape planning and other fields, which proves its certain universality.

Sustainable design, which takes natural aesthetics as the standard, relies on ecosystem's self-succession and is supplemented by artificial remediation methods, is the integration of artificial restoration and natural restoration mode, with the premise of no large-scale engineering remediation for constructing abandoned coal mines' ecosystem. From the perspective of ecological ethics, sustainable design belongs to an alternative theory of abandoned coal mines' ecological remediation, and has a set of sustainable ecological design concepts.

In accordance with labor intensity or frequency of visits, sustainable design, from the inside to the outside, zones abandoned coal mining areas into five parts (as shown is Fig. 1). Selecting diversified species is to increase niche by vertical planting, marginal cultivation and other models; the disposition of biological elements is dominated by perennial plants, considering the functional diversity of each organism and the functional synergy between organisms.

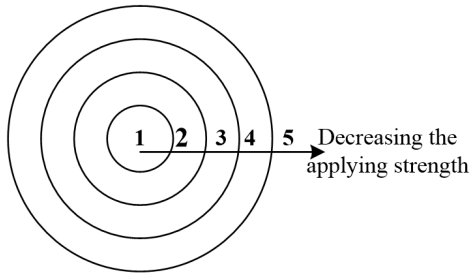


Fig. 1. Sustainable design's zoning plan

Sustainable design emphasizes the use of parks, lawns, residential areas, roads, and other marginal areas as well as not reasonably utilized open lands for growing grain crops with landscape function. It also focuses on organic combination of production, living elements and biological elements.

The value orientation of sustainable design advocates harmonious development of man and nature, with ecological perspective to observe things, which reflects sustainable development concepts, that is people's demands are fundamental and the needs of people for the people, and symbiosis between man and nature. It can promote the transformation and upgrading of traditional culture to ecological culture.

4. The application of sustainable design in abandoned coal mine's ecological remediation

Wangshi'ao coal mine locates in 12.5 kilometers east suburb of Tongchuan city in Shaanxi Province, which is one of the 156 key projects built by the Soviet Union during the first Five-Year-Plan. It started construction in December 1957 and applied into use in November 1960 with an annual production capacity of 120 tons, the largest mechanized shaft in Northwest China at that time. The rise and fall of Wangshi'ao coal mine is a microcosm of the development of China's coal industry.

4.1. The application model of sustainable design

Planning according to different zones. In practice, sustainable design does not concern every zone, and each zone are not divided equally from the inside to the outside area. According to Wangshi'ao coal mine's location, surrounding environment and the intensity of land use, it is divided into four zones from the inside to the outside (as shown in Fig. 2). The first zone is the main area for people's activities, so it requires the most amount of energy and management. From the first zone to the fourth zone, the biological elements gradually decrease so there are different ecological construction goals, which means the time for management and maintenance is gradually reduced. The planning principle for each zone can be seen in Table 1.

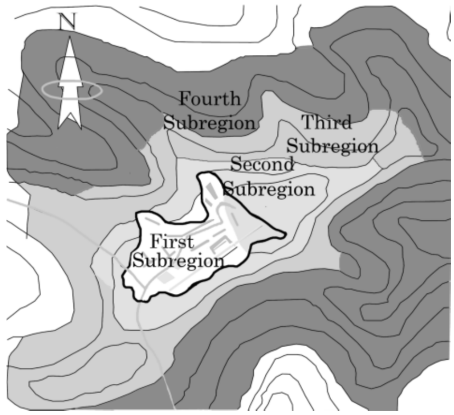


Fig. 2. Zoning diagram of the ecological restoration in Wangshi'ao coal mine

Sector analysis. Sustainable design focuses on the application of natural energy flow patterns, creating a regional microclimate environment by human, as shown in Fig. 3. In Wangshi'ao coal mine, a low water tank is at the southwest side in the first zone, connecting with drains in other zones, which can cool the hot air from the southwest during summer; there is a degradation tank in the shade at the second zone and a firewood storing area at the sun exposure area; a hedge is set at northeast side of the fourth zone to fend off dry and cold monsoon during winter, etc.

Cyclic layout. Sustainable design emphasizes zoning across areas and mobilizing ecological elements as much as possible to participate in material recycling. Wang-

shi'ao coal mine ecosystem makes full use of the symbiotic relationship between animals and plants, forming the industrial chain of material recycling as shown in Fig. 4. In addition, drains are excavated from natural gullies in each zone. During the rainy season, surface water, through the third zone, is injected into catchments in the second or even the first zones, which can be served as breeding grounds for aquatic animals and plants or used for irrigation.

Table 1. Principle of zoning plan of the ecological restoration in Wangshi'ao coal mine

Subregion	Construction objectives	Visiting frequency
1	Industrial heritage tourism area, ecological vegetable garden	Several times per day
2	Food forest, orchard, fowl	Daily
3	Grain crops, Chinese herbal medicines	Once every several days
4	Gullies, forested area, steppe area	Once every several weeks

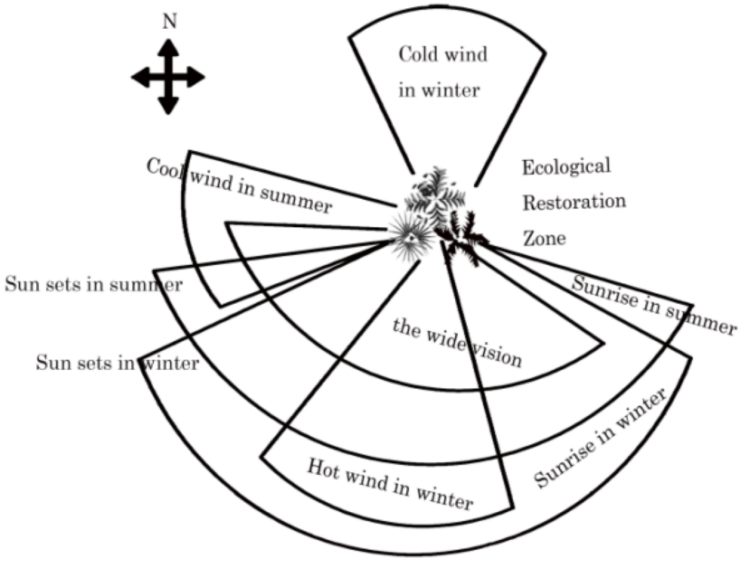


Fig. 3. Sector analysis of sustainable design

4.2. Configuration of biological elements in sustainable design

Creating niche. In the first zone of Wangshi'so coal mine, we plant ornamental fruits and vegetables in marginal areas at sides of the pond, near the building, and along the path, achieving ecosystem stability through plant density and biodiversity.

A mixture of apple, cherry, walnut and other economic forest is planted in the second zone with beans, peppers and potatoes scattered among them and the lowest layer is strawberries, vegetables, mint and other low plants to construct an uneven and spreading plant forest.

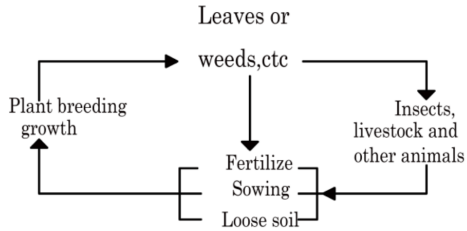


Fig. 4. Material cyclic use layout in sustainable design

Prioritizing local plants. Local perennial plants are preferred in sustainable design, and grain and forage crops such as wheat, rape, alfalfa etc. are planted at the third zone with cumin, pepper, dandelion and other spices scattered among them. In the fourth zone, there are natural secondary forest of Pinus, poplar, oak etc. and wild herbs including bupleurum, forsythia, rhizoma atractylodis, scutellaria etc. In addition, temperate deciduous broad-leaved forest mixes with grasses and shrubs, within which there are small animal groups of partridge, pheasant, deer and sheep etc.

Utilizing Synergistic relations between biological elements. At the first zone of Wangshi'ao coal mine, vines are planted among rose plants where vine roots absorb chemical material released from roses roots to fend off pest white flies; clover and other nitrogen fixing and cold resistant species are planted in the grassland; under fruit trees, there grow vegetables, such as legumes and peppers for nitrogen fixation and pest control and at the edge area there are beekeeping farms and small dairy farms, etc.

4.3. Measurement of ecosystem service value in Wangshi'ao coal mine

To evaluate the application of sustainable design in abandoned coal mines' ecological remediation, it is necessary to estimate the value of ecosystem services in Wangshi'ao coal mine in recent years and understand its trend of change.

The ecosystem types in Wangshi'ao coal mine are mainly forest, farmland and grassland. The ecological function service types are food production, material supply, gas regulation, climatic regulation, water conservation, soil protection and waste treatment and biodiversity as well as entertainment culture [10]. In the ecosystem of Wangshi'ao Coal Mine, the calculation formula of production service value of food produced by per hectare of the farmland ecosystem is [11]

$$V_{\mathrm{e}} = \frac{1}{7}\sum_{j=1}^{n}\frac{A_j P_j Y_j}{T}, \quad j = 1, 2, 3, \cdots, n\,. \tag{1}$$

In formula (1), V_e denotes the service value of food produced by per hectare of the farmland ecosystem, yuan/hm^2, A_j is the planted area of the grain crop j in the farmland ecosystem, P_j is the national average price within the calculation period of the grain crop j, yuan/t, Y_j is the yield of the grain crop j, t/hm^2, T is the total area of the grain crop n, hm^2. 1/7 is the ecological function service value that the farmland ecosystem can offer under natural conditions with human factor intervention fully eliminated. The crops of the farmland ecosystem in Wangshi'ao coal mine are wheat, corn and rapeseeds (Table 2).

Table 2. Crop planting in the farmland ecosystem of Wangshi'ao coal mine

	Name	2014	2015	2016
Planted area (hm^2)	Wheat	26	31	36
	Corn	17	25	25
	Rapeseed	9	13	27
	Total	52	69	88
Yield per unit (t/hm^2)	Wheat	2.15	2.29	2.55
	Corn	3.16	3.50	3.78
	Rapeseed	0.98	1.39	1.35
Unit price (yuan/t)	Wheat	2.507	2.513	2.360
	Corn	2.243	2.241	1.903
	Rapeseed	5.080	3.726	3.910

Note: the data is available from the Propaganda Department of Wangshi'ao Coal Mine, Yearbook of Tongling City and www.cngrain.cn.

According to formula (1), it was calculated that the service values of food produced by per hectare of the farmland ecosystem in Wangshi'ao coal mine ecosystem during 2014–2016 were 839.12 yuan/hm^2, 914.73 yuan/hm^2 and 875 yuan/hm^2.

According to the Unit Area Ecological Service Value Equivalent Scale of Terrestrial Ecosystem in China of the scholars XIE Gaodi et al., the unit area ecological service value equivalent scale of Wangshi'ao coal mine ecosystem was obtained (Table 3). According to formula (2) below, the ecological service value per hectare of the ecosystem types in Wangshi'ao coal mine was calculated (Table 4).

$$V_{mi} = c_{mi} V_e \,. \tag{2}$$

Here, V_{mi} is the value of the type m ecological service provided by per hectare of type i ecosystem, yuan/hm^2, c_{mi} is the coefficient of value equivalent of the type m ecological service provided by per hectare of type i ecosystem versus the functional service of food produced by per hectare of the farmland ecosystem; V_e is the service value of food produced by per hectare of the farmland ecosystem, yuan/hm^2.

According to the area of ecosystems in Wangshi'ao coal mine (Table 5) and based on the annual ecosystem service value in Table 4, the ecological service function values of ecosystems in Wangshi'ao coal mine could be calculated (Table 6).

Table 3. Unit area ecological service value equivalent scale of Wangshi'ao coal mine ecosystem

	Forest	Grassland	Farmland
Food production	0.1	0.3	1
Material supply	2.6	0.05	0.1
Gas regulation	3.5	0.8	0.5
Climatic regulation	2.7	0.9	0.89
Water conservation	3.2	0.8	0.6
Soil protection	3.9	1.95	1.46
Waste treatment	1.31	1.31	1.64
Biodiversity protection	3.26	1.09	0.71
Entertainment culture	1.28	0.04	0.01

Table 4. Ecological service value per-hectare of Wangshi'ao coal mine ecosystem from 2014 to 2016

	Forest	Grassland	Farmland
Food production	0.1	0.3	1
Material supply	2.6	0.05	0.1
Gas regulation	3.5	0.8	0.5
Climatic regulation	2.7	0.9	0.89
Water conservation	3.2	0.8	0.6
Soil protection	3.9	1.95	1.46
Waste treatment	1.31	1.31	1.64
Biodiversity protection	3.26	1.09	0.71
Entertainment culture	1.28	0.04	0.01

5. Discussion

Combined with the ecological remediation of Wangshi'ao coal mine and calculation results of its ecosystem's service function value, it is found that the main roles of sustainable design in abandoned coal mines' ecological remediation are as follows.

5.1. Promoting the remediation of abandoned coal mines' biodiversity to achieve evolution to regional natural ecological systems

Sustainable design establishes a pioneer plant community by simulating natural ecosystems and maximizing the use of resources and topographic conditions in abandoned coal mines. With successive regional species' invasion, abandoned coal mines have more similarities with regional natural ecosystem, along which there are improvements in ecological system's self-maintenance ability, stability, ecological carrying capacity and anti-risk ability.

Table 4. Ecological service value per-hectare of Wangshi'ao coal mine ecosystem 2014 to 2016

	2014			2015			2016		
	Forest	Grass-land	Farm-land	Forest	Grass-land	Farm-land	Forest	Grass-land	Farm-land
Food pro-duc-tion	83.91	251.73	839.12	91.47	274.42	914.73	87.5	262.5	875.00
Material supply	2181.71	41.96	83.91	2378.30	45.74	91.47	2275	43.75	87.50
Gas regula-tion	2936.92	671.30	419.56	3201.56	731.78	457.37	3062.5	700	437.50
Climatic regula-tion	2265.62	755.21	746.82	2469.77	823.26	814.11	2362.5	787.5	778.75
Water conser-vation	2685.18	671.30	503.47	2927.14	731.78	548.84	2800	700	525.00
Soil protec-tion	3272.57	1636.28	1225.12	3567.45	1783.72	1335.51	3412.5	1706.25	1277.50
Waste treat-ment	1099.25	1099.25	1376.16	1198.30	1198.30	1500.16	1146.3	1146.25	1435.00
Biodi-versity	2735.53	914.64	595.78	2982.02	997.06	649.46	2852.5	953.75	621.25
Enter-tain-ment culture	1074.07	33.56	8.39	1170.85	36.59	9.15	1120	35	8.75
Total	18334.76	6075.23	5798.33	19986.86	6622.65	6320.8	19118.8	6335	6046.25

Table 5. Land use area of the ecosystems in Wangshi'ao coal mine from 2014 to 2016, hm^2

	2014	2015	2016
Forest	62	70	84
Grassland	76	89	113
Farmland	57	73	91

5.2. Speeding up the accumulation of soil nutrients in abandoned coal mines and promoting ecological remediation systems integrating into biogeochemical cycles

Each zone of derelict land in the Wangshi'ao coal mine grows clover, soybean, locust tree, sea-buckthorn, oleaster, lespedeza, and other nitrogen fixing plants, while animal manure, leaves, straw and alfalfa are good sources for soil nutrients such as nitrogen, phosphorus, potassium, calcium, iron, magnesium and zinc. Sustainable

design is to promote abandoned coal mines' ecosystem to integrate with biogeochemical cycles by means of nutrient retention and loss, and organic matter synthesis and degradation.

Table 6. Ecological service value of the ecosystems in Wangshi'ao coal mine from 2014 to 2016

	2014	2015	2016
Forest ecosystem	1.14×10^6 (59%)	1.40×10^6 (57%)	1.61×10^6 (56%)
Grassland ecosystem	0.46×10^6 (24%)	0.59×10^6 (24%)	0.72×10^6 (25%)
Farmland ecosystem	0.33×10^6 (17%)	0.46×10^6 (19%)	0.55×10^6 (19%)
Total value of biological services	1.93×10^6 (100%)	2.45×10^6 (100%)	2.88×10^6 (100%)

5.3. Enhancing abandoned coal mines' ecosystem service function value and speeding up the ecological balance construction

The ecosystem service value of Wangshi'ao coal mine shows an increasing trend year by year. Although the annual increment is very low, each ecosystem function service value basically maintains balanced change condition, which means that sustainable design can help maintain a relatively stable change in species composition and proportion of ecosystems, and promote the ecosystem to move towards dynamic equilibrium.

5.4. 5.4 Promoting the coordinated development of ecological, economic and social subsystems in abandoned coal mines

While constructing ecosystem, sustainable design also considers planting agricultural plants with edible and landscape functions, and taps coal mines' social and cultural values to incorporate abandoned coal mines' ecology, economy and social elements into one ecosystem and promote coordinated development among abandoned coal mines' subsystems: ecology, economy and society.

6. Conclusion

To explore ecological remediation plans for abandoned coal mines in line with the sustainable development perspective, this paper, based on the sustainable development theory, analyzes ecological remediation patterns for abandoned coal mines and configuration of ecological elements. This paper finds that sustainable design has four stages in abandoned coal mines' ecological remediation. The first is to establish a diversified pioneer plant community, promoting abandoned coal mines to evolve into regional natural ecological systems. The second is to accelerate soil nutrients' accumulation in abandoned lands by using biological elements. The third is to enhance abandoned coal mines' ecosystem service function value, and promote sustainable development of ecosystems' structure and function. The fourth is to incorporate economic and social elements into ecological remediation and promote

coordinated development among abandoned coal mines' subsystems: ecology, economy and society.

Up to now, there is no universal evaluation system for ecosystem services value [12], and the ecosystem services value per unit area is affected by its biomass and its spatial location [13]. Those uncertainties will lead to calculating deviations of ecosystem's service function, influence environmental economic accounting of abandoned coal mines, and increase difficulties in controlling abandoned coal mines' ecological mechanism and ecological function zoning.

References

[1] H. L. Zhang, L. N. Sun, T. H. Sun, L. F. Chen: *Substrate amelioration and vegetation reconstruction in ecological remediation of abandoned mines: Research advances.* Chinese Journal of Ecology *31* (2012), No. 2, 460–467.

[2] M. Zhang, F. Chen, Y. Wu, Y. Ma, S. Guan, Y. Huang: *Characteristics of the soil seed bank of planted and natural restored draw-down zones in the Three Gorges Reservoir Region.* Ecological Engineering *103* (2017), Part A, 127–133.

[3] S. Fernández, C. Poschenrieder, C. Marcenò, J. R. Gallego, D. Jiménez-Gámez, A. Bueno, E. Afif: *Phytoremediation capability of native plant species living on Pb-Zn and Hg-As mining wastes in the Cantabrian range, north of Spain.* Journal of Geochemical Exploration *174* (2017), 10–20.

[4] H. Lei, Z. Peng, Y. Hu, Y. Zhao: *Vegetation and soil restoration in refuse dumps from open pit coal mines.* Ecological Engineering *94* (2016), 638–646.

[5] S. Mukhopadhyay, S. K. Maiti, R. E. Masto: *Use of reclaimed mine soil index (RMSI) for screening of tree species for reclamation of coal mine degraded land.* Ecological Engineering *57* (2013), 133–142.

[6] P. Dorr de Quadros, K. Zhalnina, A. G. Davis-Richardson, J. C. Drew, F. B. Menezes, F. A. de O. Camargo, E. W. Triplett: *Coal mining practices reduce the microbial biomass, richness and diversity of soil.* Applied Soil Ecology *98* (2016), 195–203.

[7] X. Sun, Y. Li: *The spatial distribution of soil heavy metals and variation characteristics of datong abandoned coal mine area in Huainan City.* Scientia Geographica Sinica *33* (2016), No. 10, 1238–1244.

[8] C. L. Cao, X. S. Yu, B. Geng, X. Liu, J. Ye, C. X. Zhu: *Soil reclamation for abandoned opencast coal mine and a case study on Pingshuo surface coal mine.* Journal of Xi'an University of Science and Technology *33* (2013), No. 1, 51–55.

[9] Y. Guo: *Soil physical and chemical properties and species diversity characteristics during the vegetation recovery process in wasteland of opencast coal mine.* Forestry Science & Technology *37* (2012), No. 5, 51–54.

[10] G. D. Xia, C. X. Lu, Y. F. Leng, D. Zheng, S. C. Li: *Ecological assets valuation of the Tibetan Plateau.* Journal of Natural Resources *18* (2003), No. 2, 189–196.

[11] L. Wang, J. Yang, Y. M. Hu, J. H. Li, Z. L. Jia: *Quantitatively study on landuse ecosystem service value in Guangzhou City.* Bulletin of Water and Soil Conservation *29* (2009), No. 4, 229–234.

[12] G. D. Xie, C. X. Zhang, L. M. Zhang, : *Improvement of the evaluation method for ecosystem service value based on per unit area.* Journal of Natural Resources *30* (2015), No. 8, 1243–1254.

[13] L. Ovsepyan, A. Mostovaya, V. Lopes de Gerenyu, I. Kurganova: *Changes in microbial activity of soils during the natural restoration of abandoned lands in central Russia.* EGU General Assembly, 12–17 April 2015, Vienna, Austria, Geophysical Research Abstracts *17* (2015) ID. No. 592.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 315–328

Prediction of water quality class based on an ELM optimized by a MFOA

Chi Zhang[1]

Abstract. As an extreme learning machine easily runs into local optimization and is slow in convergence, a MFOA-ELM-based water quality evaluation model was proposed so as to enhance the prediction accuracy and applicability of water quality evaluation model. In order to prevent the FOA from running into local optimization, the correction factor was introduced into FOA to put forward a MFOA. High-precision water quality prediction and evaluation can be realized by optimizing the weight and threshold of ELM with MFOA under the condition of optimal parameters. Water quality monitoring data of Chao Lake between 2010 and 2015 were selected as objects of research. The water quality of Chao Lake was predicted and evaluated based on the research on the parameter change trends of different water qualities. Through the comparison of 3 water evaluation and prediction methods, namely, MFOA-ELM, FOA-ELM and PSO-ELM, it is found that the prediction accuracy of MFOA-ELM reaches up to 98.36 %, which is higher than the prediction accuracies of PSO-ELM and FOA-ELM. Thus, it shows that MFOA-ELM is more accurate and applicable in evaluating and predicting water quality, verifying the validity and reliability of MFOA-ELM. Meanwhile, compared with other algorithms, MFOA-ELM is faster in convergence and better in effect.

Key words. FOA, ELM, water quality evaluation, particle swarm optimization, correction factor, convergence rate.

1. Introduction

Water quality evaluation and prediction aim to correctly reflect the quality and pollution of water environment through the water quality evaluation, predict the future development tendency of water environment quality, and provide a basis and method for scientific decision-making of management, protection and governance of water environment. At present, water prediction methods mainly consist of grey prediction method [1], artificial neural network [2] and support vector machine (SVM) [3] etc. Although a satisfactory result can be obtained through water prediction based on the grey theory, this prediction method is low in prediction accuracy and not suitable for long-term prediction [4–5]. To increase the prediction accuracy of

[1]Shandong University of Science and Technology, 266590, China

water quality, reference [6] and reference [7] respectively proposed a water quality prediction model based on weighted array, exponential smoothing and GM (1,1) model combination. Although this method can increase the prediction accuracy to a certain extent, it fails to fundamentally solve the problem of prediction error. In order to enhance the accuracy of water quality evaluation, reference [8] proposed a water quality evaluation model based on grey renewal GM(1,1) model. The experimental result shows that the prediction accuracy of this method is superior to that of a traditional GM(1,1) model. Based on the accurate non-linear mapping capability and generalization ability of BP neural network, reference [9] proposed a water quality prediction model based on BP neural network. The experimental result shows that the method is high in accuracy and generalization ability. As the small sample data of BP neural network is low in prediction accuracy and easily runs into local optimization, reference [10] proposed a water quality prediction model based the improved BP neural network, increasing the prediction accuracy. In consideration of the disadvantages of BP neural network such as local optimization and slow convergence, references [11–12] established a water quality prediction model based on GA-BP neural network by optimizing the weight, threshold and network structure of BP neural network with GA; through the optimization with GA, both its prediction accuracy and the stability in its prediction result can be enhanced. As the single SVM prediction model is low in prediction accuracy and slow in speed, reference extracted the characteristic information of water quality data via the wavelet analysis and proposed a water quality prediction model based on wavelet transformation and SVR. Reference proposed a water quality parameter prediction model based on improved weighted SVM, largely increasing the prediction accuracy. According to small samples and time series data of jump water quality, reference proposed a water quality prediction model based on ELPM data preprocessing and PSO algorithm optimizing the parameters of least square VSM.

FOA (Fruit Fly Optimization Algorithm), which is a swarm intelligence algorithm proposed by simulating the fruit fly's forging behavior, has the advantages such as few control parameters and fast convergence. At present, it is found that there is no article concerning the application of FOA in water evaluation. In consideration of the disadvantages of a traditional water quality evaluation model such as low prediction accuracy and poor adaptability, a ELM optimized by MFOA (Modified Fruit Fly Optimization Algorithm) is proposed here to evaluate water quality. The research result shows that MFOA has obvious advantages in both the optimization effect and computing speed so it has a satisfactory effect.

2. MFOA

2.1. FOA

The flow of FOA is shown below:

Step 1. Initialize the algorithm parameters and set the population size and maximum numbers of iterations to be popsize and iteration respectively; set the initial positions of fruit flies to be X_{begin} and Y_{begin}.

Step 2. Find out the optimizing directions of individual fruit fly and calculate the distance according to the formulas

$$x_i = X_{\text{begin}} + \text{Value} \times \text{rand}\,, \tag{1}$$

$$y_i = Y_{\text{begin}} + \text{Value} \times \text{rand}\,. \tag{2}$$

In the formulas above, x_i and y_i refer to positions of individual fruit flies. Quantity Value refers to the scouting distance of fruit fly.

Step 3. Calculate the distance d_i between the individual fruit fly and the original point and smell concentration s_i of individual fruit fly according to the formulas

$$d_i = \sqrt{x_i^2 + y_i^2}\,, \tag{3}$$

$$s_i = \frac{1}{d_i}\,. \tag{4}$$

Step 4. Calculate the decision function of smell concentration and obtain the smell concentration of current position of individual fruit fly in the form

$$\text{Smell}_i = \text{Function}(s_i)\,. \tag{5}$$

Step 5. Search for the best smell concentration Smell_b and best positions given by x_b and y_b among the fruit fly population.

2.2. MFOA

According to formula (3) and formula (4), it can be seen that the decision value of smell concentration s_i gets very small after calculating the reciprocal in formula (4). Then, if smell concentration s_i is used as the decision function, FOA will run into local optimum, causing the problem of "prematurity".

To prevent FOA algorithm from running into local optimum, a MFOA (Modified Fruit Fly Optimization Algorithm) was proposed by introducing the correction factor β into the basic FOA. Its modified formulas are shown below.

$$d_i = \sqrt{x_i^2 + y_i^2}\,, \tag{6}$$

$$s_{Mi} = \frac{1}{d_i} + \beta\,. \tag{7}$$

In the formulas above, s_{Mi} refers to the smell decision function of MFOA. Now

$$\beta = \begin{cases} g \times d_i, \\ \\ K \times X_{\text{axis}} \text{ or } K \times Y_{\text{axis}}, \end{cases} \tag{8}$$

where g obeys uniform distribution and K refers to a constant.

3. ELM (extreme learning machine)

For N different samples

$$(\boldsymbol{x}_i, \boldsymbol{t}_i), \quad \boldsymbol{x}_i = [x_{i1}, x_{i2}, ..., x_{in}]^{\mathrm{T}} \in R^n \text{ and } \boldsymbol{t}_i = [t_{i1}, t_{i2}, ..., t_{im}]^{\mathrm{T}} \in R^m$$

a unified model of SLFN with the number of nodes in a hidden layer of $\tilde{N}$ and the excitation function of $g(x)$ is shown below:

$$\sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g_i(\boldsymbol{x}_j) = \sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g(\boldsymbol{a}_i \cdot \boldsymbol{x}_j + b_i) = \boldsymbol{t}_j, \quad j = 1, \cdots, N. \tag{9}$$

In the formula above, $\boldsymbol{a}_i = [a_{i1}, a_{i2}, \cdots, a_{in}]^{\mathrm{T}}$ refers to input weight connecting the ith hidden layer node, b_i refers to the bias of ith hidden layer node, $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{im}]^{\mathrm{T}}$ refers to the output weight of ith hidden layer node, and $\boldsymbol{a}_i \cdot \boldsymbol{x}_j$ refers to the inner product of $\boldsymbol{a}_i$ and $\boldsymbol{x}_j$.

Let $E(\mathbf{W})$ refers to the sum of squared error between the expected value and actual value. The problem solved is to find out the optimal weight $\mathbf{W}(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\beta})$ to minimize the cost function $E(\mathbf{W})$. Its mathematical model can be expressed as

$$\begin{aligned} &\underset{\mathbf{W}=(\boldsymbol{a},\boldsymbol{b},\boldsymbol{\beta})}{\operatorname{argmin}} \; E(\mathbf{W}) = \underset{\mathbf{W}=(\mathrm{a,b},\beta)}{\operatorname{argmin}} \; ||\boldsymbol{\varepsilon}||^2, \\ &s.t. \; \textstyle\sum_{i=1}^{\tilde{N}} \boldsymbol{\beta}_i g(\boldsymbol{a}_i \cdot \boldsymbol{x}_j + b_i) - \boldsymbol{t}_j = \boldsymbol{\varepsilon}_j, \quad j = 1, \cdots, N. \end{aligned} \tag{10}$$

$\boldsymbol{\varepsilon}_j = [\varepsilon_{j1}, \varepsilon_{j2}, ..., \varepsilon_{jm}]$ refers to the error of jth sample. The schematic diagram of ELM is depicted in Fig. 1.

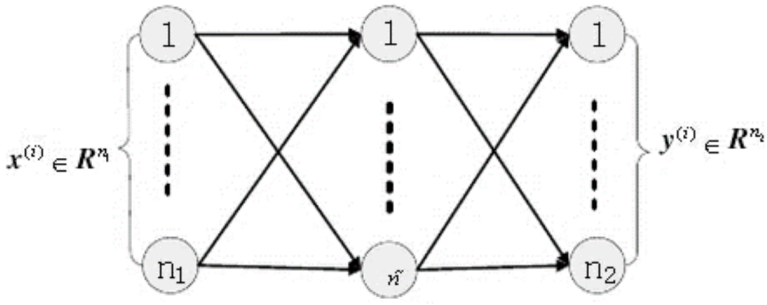


Fig. 1. Schematic diagram of ELM

4. Water quality prediction with the ELM optimized by MFOA

4.1. Water quality evaluation indexes

Water quality evaluation is to calculate and determine the water quality class of water sample through a certain mathematical model based on water quality evaluation criteria and all index values of water sample. As there are abundant indexes of water quality analysis, this paper uses ammonia nitrogen, dissolved oxygen, chemical oxygen demand, permanganate index, total phosphorus and total nitrogen as water evaluation indexes in combination with the quality standards for surface water environment; their corresponding water quality classes are shown in Table 1:

Table 1. Water quality classes and content standards

Type	Type 1	Type 2	Type 3	Type 4	Type 5
Ammonia nitrogen (mg/l) $<$	0.15	0.50	1.0	1.5	2.0
Dissolved oxygen (mg/l) $>$	7.5	6.0	5.0	3.0	2.0
Chemical oxygen demand (mg/l) $<$	15	15	20	30	40
Permanganate index (mg/l) $<$	2.0	4.0	6.0	10	15
Total phosphorus (mg/l) $<$	0.02	0.10	0.20	0.30	0.40
Total nitrogen (mg/l) $<$	0.20	0.50	1.0	1.5	2.0

4.2. Fitness function

On the premise of guarantee the minimum prediction errors in water quality classes, the MFOA is used to optimize the weight and threshold of ELM. Since the weight and threshold among the parameters of ELA need to be optimized, its fitness function is

$$\text{Minimize Fitness}(w_{ij}, b_j) = \sum_{i=1}^{m} (o_i^k - d_i^k)\,. \tag{11}$$

In the formula, d_i^k and o_i^k, respectively, refer to the input and output of ELM; w_{ij} and b_j, respectively, refer to the weight and threshold of ELM.

4.3. Algorithm steps

The algorithm flow of water quality prediction with the ELM optimized by MFOA is shown below.

Step 1. Normalize water quality sample data and establish training samples and test samples.

Step 2. Set the population size and maximum number of iterations of MFOA to be popsize and iteration, respectively.

Step 3. Input the established training samples into the ELM. Calculate the fitness function value of individual fruit fly according to the objective function formula (11).

Search for the positions and optimal values of individual fruit fly and global optimal fruit fly.

Step 4. Update the speed and position of fruit fly.

5. Experimental analysis

5.1. Data source

Water samples from the Chao Lake were collected as objects of water quality evaluation. The sampling water intakes were the Nanfei River lake inlet, Pai River lake inlet and Chao Lake dam entrance; the longitudes and latitudes of sampling points are shown in Table 2. The water quality sampling time of Chao Lake was between 2010 and 2015. The sampling frequency was once every quarter. The changing trend of all indexes of water samples are shown in the Figs. 2, and 3.

Table 2. Longitudes and latitudes of sampling points

No.	Sampling point	Longitude	Latitude
1#	Nanfei River lake inlet	$117°24'40''$	$31°42'15''$
2#	Pai River lake inlet	$117°18'15''$	$31°41'30''$
3#	Chao Lake dam entrance	$117°51'46''$	$31°34'18''$

5.2. Empirical results

In order to verify the validity and reliability of algorithm proposed in this paper, the water quality data of three sampling points in the Chao Lake between 2010 and 2015 were used as objects of research. The parameters of MFOA are set below: the population size is 20, the maximum number of iterations is 100, the maximum number of iterations of ELM is 100, the target error is 0.001, the number of internuncial neurons is 20. The water quality prediction result with MFOA-ELM is shown in Fig. 4. The water quality class evaluation results of MFOA-ELM, FOA-ELM and PSO-ELM are shown in Table 3 and Table 4.

According to the water quality class prediction results of MFOA-ELM, FOA-ELM and PSO-ELM in Table 3 and Table 4, it can be seen that the prediction results of MFOA-ELM and PSO-ELM and better than that of FOA-ELM, and the water quality evaluation result and misjudgment rate of MFOA-ELM are the best, thus verifying the superiority and reliability of MFOA-ELM.

5.3. Comparison of convergence rates of different algorithms

In order to compare the convergence rates of ELMs optimized by MFOA, FOA, PSO and GA [10], these algorithms were randomly operated for 4 times and the comparison of their convergence results are shown in Figs. 5 and 6. Compared with

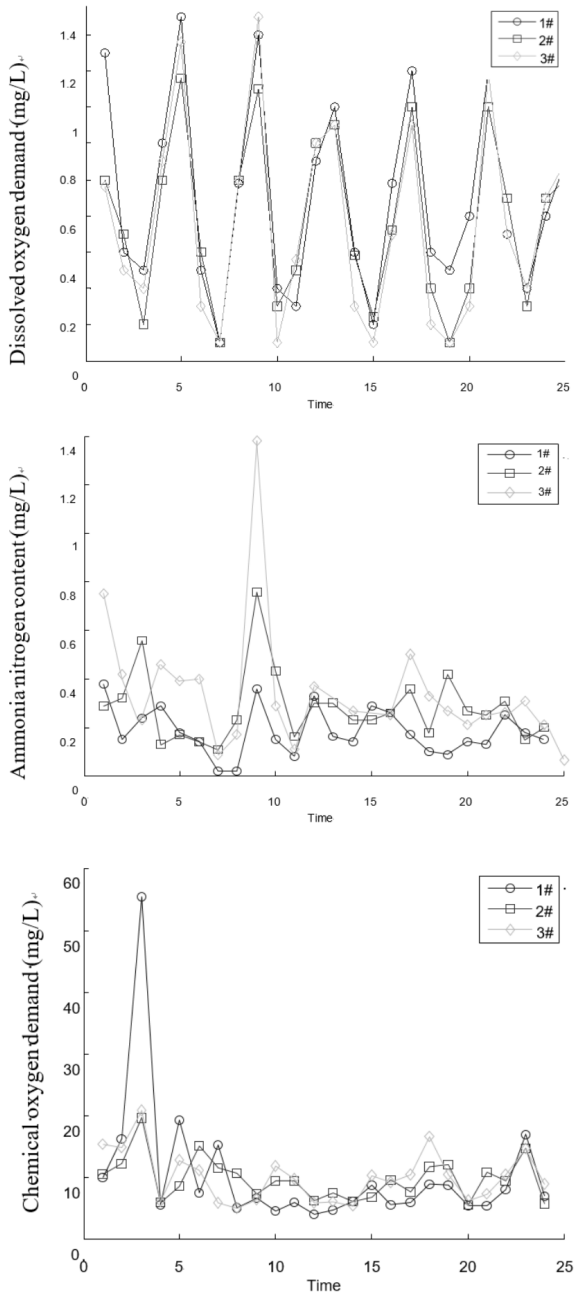


Fig. 2. Dissolved oxygen content, ammonia nitrogen content and chemical oxygen demand

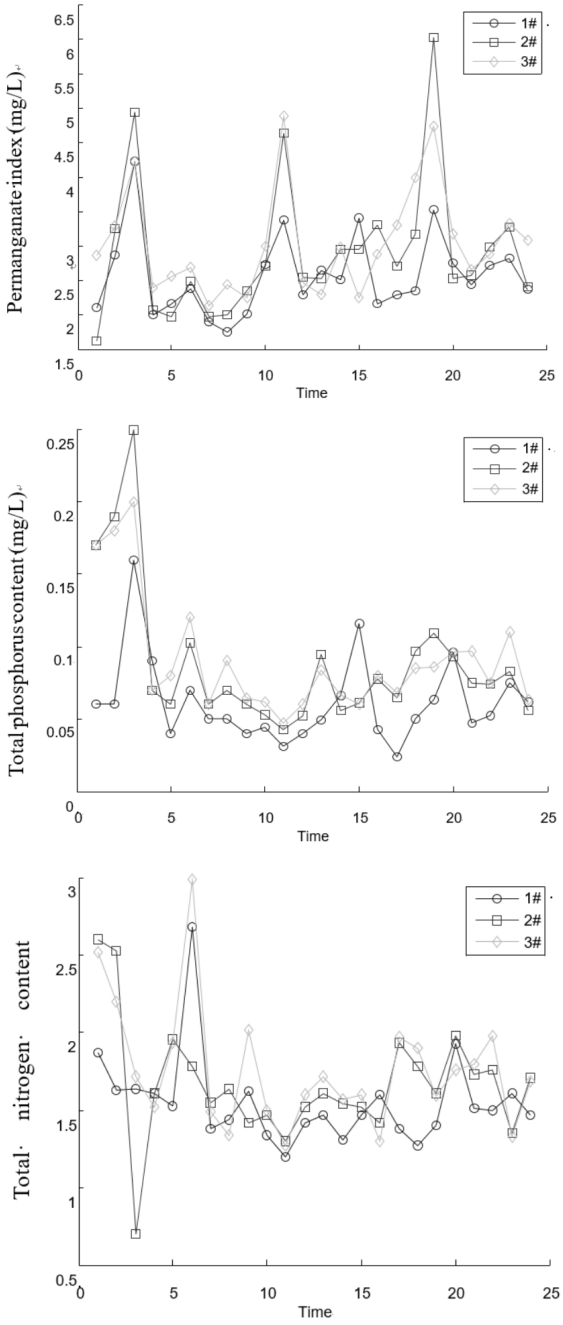


Fig. 3. Permanganate index content, total phosphorus content and total nitrogen content

FOA, PSO and GA, MFOA is faster in convergence.

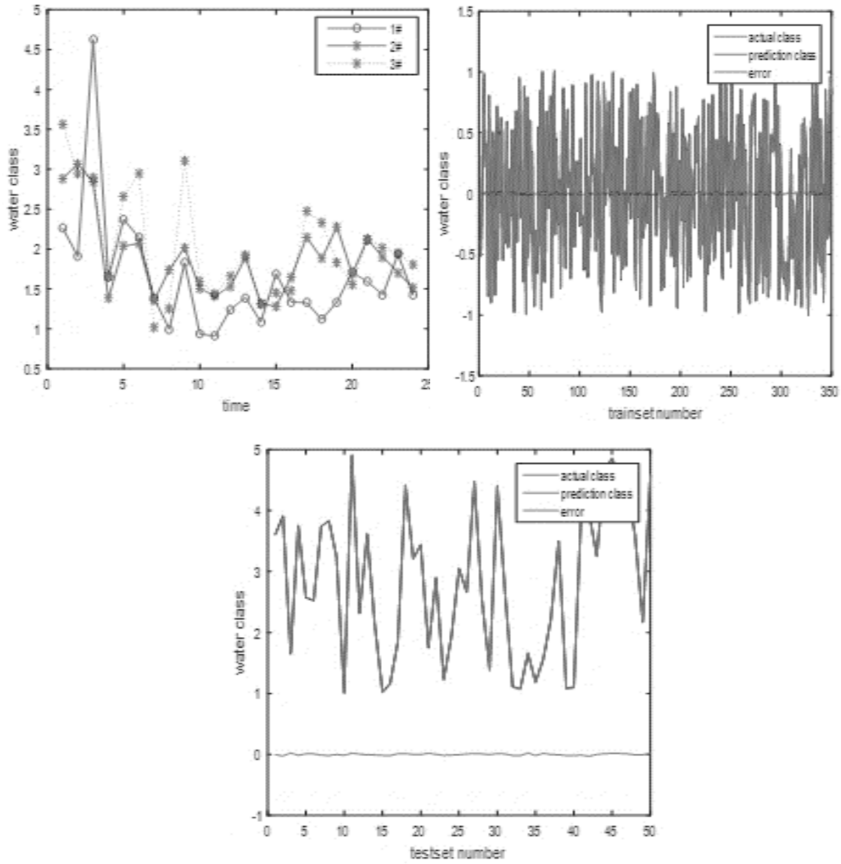


Fig. 4. Water quality prediction result: up left–prediction result, up right–training result, bottom–test result

Table 3. Water quality class prediction results of ELMs optimized by MFOA, FOA and PSO

Method	Time (s)	Grade 1	Grade 2	Grade 3	Grade 4
MFOA_ELM	40.0672	100.00%	99.35%	96.12%	99.56%
FOA_ELM	89.5534	98.54%	90.65%	58.33%	76.74%
PSO_ELM	0.8036	98.67%	96.12%	98.52%	90.43%
PSO					
Method	Time (s)	Class 1	Class 2	Class 3	Class 4
MFOA_ELM	40.0672	100.00%	99.35%	96.12%	99.56%
FOA_ELM	89.5534	98.54%	90.65%	58.33%	76.74%
PSO_ELM	0.8036	98.67%	96.12%	98.52%	90.43%

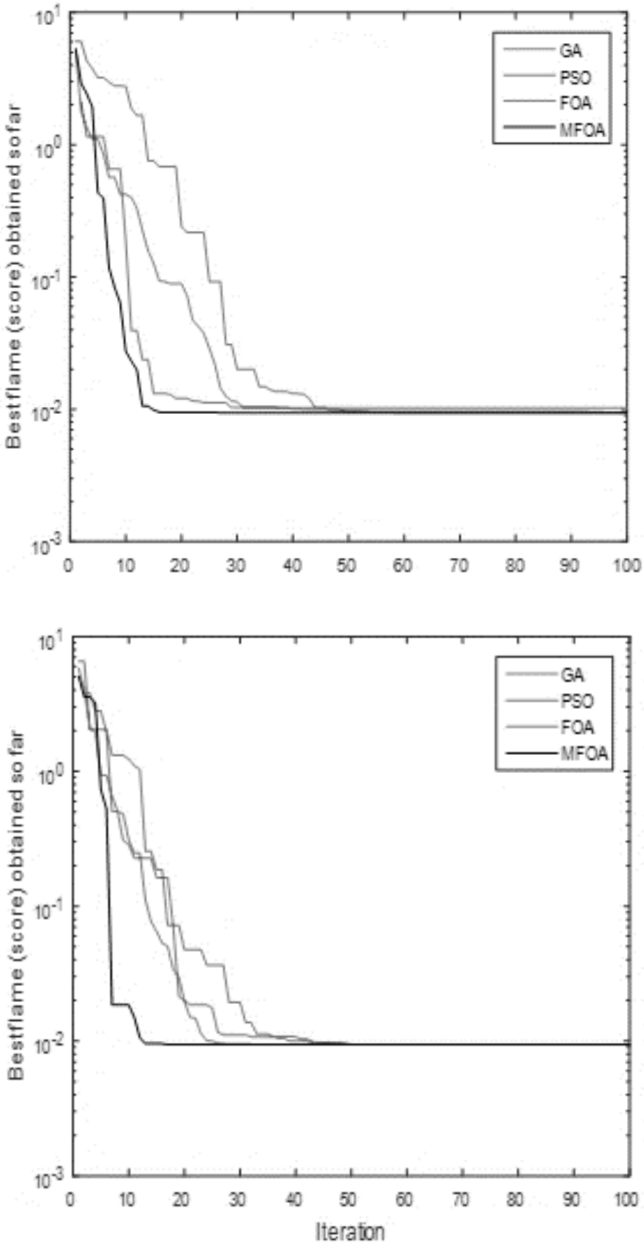


Fig. 5. Comparison of convergence rates: up–first time, bottom–second time

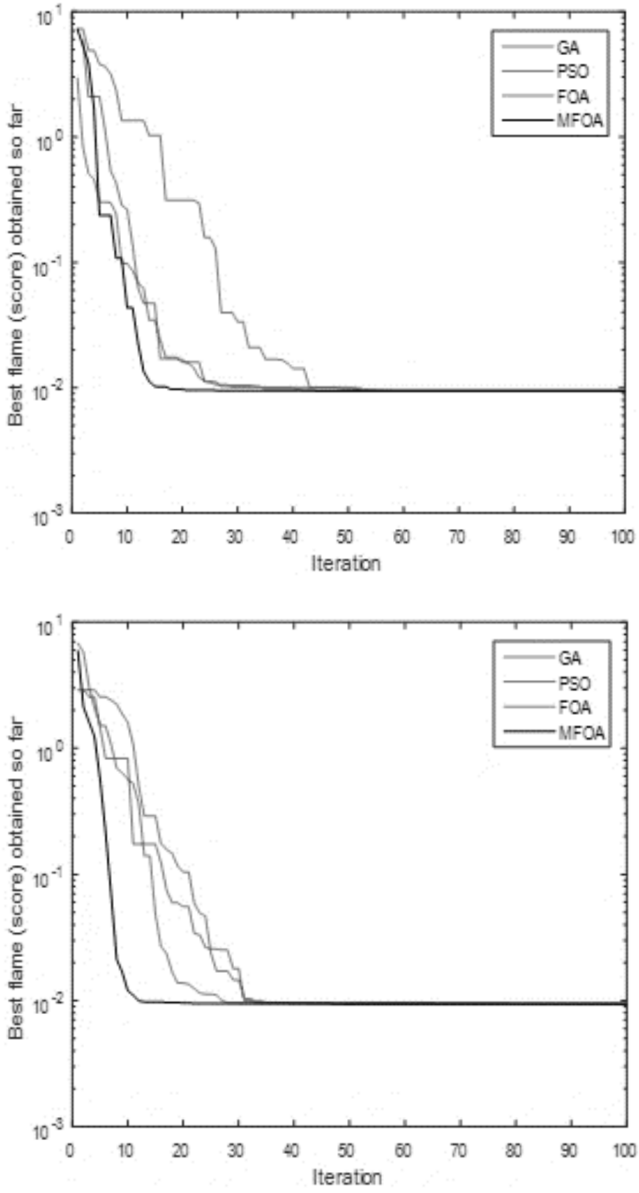


Fig. 6. Comparison of convergence rates: up–third time, bottom–fourth time

6. Conclusion

In consideration of the disadvantages of a traditional water quality prediction and evaluation model such as low prediction accuracy and poor adaptability, an ELM optimized by a MFOA was proposed to establish a water quality evaluation

model. Through the comparison of 3 water quality evaluation and prediction methods, namely, MFOA-ELM, FOA-ELM and PSO-ELM, it is found that the prediction accuracy of MFOA-ELM reaches up to 98.36 %, which is higher than those of PSO-ELM and FOA-ELM. Thus, this indicates that water quality evaluation and prediction with the MFOA-ELM is higher in accuracy and adaptability, thus verifying the validity and reliability of MFOA-ELM. Meanwhile, compared with other algorithms, MFOA-ELM has a higher convergence rate and a better effect.

Table 4. Water quality class prediction results of ELMs optimized by MFOA, FOA and PSO

Method	Class	1	2	3	4
MFOA-ELM	1	100.00%	0.00%	0.00%	0.00%
	2	0.65%	99.35%	0.00%	0.00%
	3	0.00%	3.00%	95.00%	2.00%
	4	0.00%	0.00%	0.65%	99.35%
FOA-ELM	1	98.65%	1.35%	0.00%	0.00%
	3	0.00%	0.67%	24.00%	75.33%
	3	0.00%	3.00%	95.00%	2.00%
	4	0.00%	0.67%	58.00%	41.33%
PSO-ELM	1	98.67%	0.00%	1.33%	0.00%
	2	0.00%	96.00%	3.67%	0.33%
	3	0.00%	2.00%	96.67%	1.33%
	4	0.00%	1.00%	8.33%	90.67%

References

[1] W. Yong, D. Hu: *Deficiency of the smoothness condition and its remedy.* Systems Engineering - Theory & Practice *29* (2009), No. 8, 165–170.

[2] Y. Zhang, J. Wang, A. M. Vorontsov, G. Hou, M. N. Nikanorova, H. Wang: *Using a neural network approach and time series data from an international monitoring station in the Yellow Sea for modeling marine ecosystems.* Environmental Monitoring and Assessment *186* (2014), No. 1, 515–524.

[3] S. Devi, A. K. Jagadev, S. Patnaik: *Learning an artificial neural network using dynamic particle swarm optimization–backpropagation: Empirical evaluation and comparison.* Journal of Information and Communication Convergence Engineering *13* (2015), No. 2, 123–131.

[4] J. G. Yang, S. Y. Weng, H. Zhou, K. F. Cen: *An optimized BP network model using genetic algorithm for predicting the ignition-stability index of pulverized coal.* Journal of Chinese Society of Power Engineering *26* (2006), No. 1, 81–83.

[5] K. P. Wu, S. D. Wang: *Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space.* Pattern Recognition *42* (2009), No. 5, 710–717.

[6] H. Yoon, S. C. Jun, Y. Hyun: *A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer.* Journal of Hydrology *396* (2011), Nos. 1–2, 128–138.

[7] D. Karaboga, B. Akay: *A survey: Algorithms simulating bee swarm intelligence.* Artificial Intelligence Review *31* (2009), Nos. 1–4, 61–85.

[8] I. Primpas, G. Tsirtsis, M. Karydis, G. D. Kokkoris: *Principal component analysis: Development of a multivariate index for assessing eutrophication according to the European water framework directive.* Ecological Indicators *10* (2010), No. 2, 178–183.

[9] F. Birgand, C. Faucheux, G. Gruau, B. Augeard, F. Moatar, P. Bordenave: *Uncertainties in assessing annual nitrate loads and concentration indicators: Part 1. Impact of sampling frequency and load estimation algorithms.* Transactions of the ASABE, American Society of Agricultural and Biological Engineers *53* (2010), No. 2, 437–446.

[10] M. Aminian, F. Aminian: *A modular fault-diagnostic system for analog electronic circuits using neural networks with wavelet transform as a preprocessor.* IEEE Transactions on Instrumentation and Measurement *56* (2007), No. 5, 1546–1554.

[11] J. Cui, Y. Wang: *A novel approach of analog circuit fault diagnosis using support vector machines classifier.* Measurement *44* (2011), No. 1, 281–289.

[12] P. Jantos, D. Grzechca, T. Golonek, J. Rutkowski: *Heuristic methods to test frequencies optimization for analogue circuit diagnosis.* Bulletin of the Polish Academy of Sciences Technical Sciences *56* (2008), No. 1, 29–38.

[13] Y. Jin, G. J. Chen, H. Liu: *Fault diagnosis of analog circuit based on wavelet neural network.* Chinese Journal of Scientific Instrument *28* (2007), No. 9, 1600–1604.

[14] X. S. Gan, W. M. Gao, Z. Dai, W. D. Liu: *Research on WNN soft fault diagnosis for analog circuit based on adaptive UKF algorithm.* Applied Soft Computing *50* (2017), 252–259.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 329–338

Mechanical remote control technology based on artificial intelligence[1]

Bo Cui[2], Xianchuang Fan[2], Yuejiao Niu[2]

Abstract. For solving the uncertainty of data transmission, which reduces the operational performance of control system and results in the system instability, the deterministic control of robot arm network remote control system is studied. By introducing the receiving buffer, the random time-varying delay is converted to fixed delay, and then the controller is designed for the fixed time delay. The basic ideas of the three main methods of deterministic control, namely state predictive control, step-by-step transform control and predictive control are introduced. Through the corresponding state transitions, the relationship between the three methods is discussed, and the theoretical results are proved by the simulation experiment. The results showed that the obtained results are correct. Based on the above findings, it is concluded that the deterministic control is of great help in solving the uncertainty of data transmission in the networks.

Key words. Remote control, deterministic control, fixed time delay.

1. Introduction

With the high integration and rapid development of the information industry, the remote control technology is becoming more and more popular and widespread [1–2]. The remote control is to separate the controller and the controlled object in the physical position, and to adopt certain means of communication to realize the transmission of control information and feedback information [3]. Generally, when the operator is far away from the controlled objects, it will often use the remote control technology, especially in the complex, harsh and dangerous environment. The operator remotely controls the controlled object to complete the required planning and decision-making, which is able to reflect the advantages of remote control technology, such as space exploration, marine development, remote medicine, nuclear industry remote experiments and so on. Of course, in these special areas, remote control will have higher speed, reliability and security requirements. Because of the

[1]The authors acknowledge the National Natural Science Foundation of China (Grant: 51578109), the National Natural Science Foundation of China (Grant: 51121005).

[2]School of Information & Engineering, North China University of Science and Technology, 030051, China

many applications of remote control, there are many kinds of communication means applied, such as space-based network, communication satellite, special optical cable, microwave and so on. The channel capacity, speed, transmission distance and application situations of various media are different, which makes the remote control system realize different functions and service objects [4–5]. In a network based remote control system, the transmission delay in the network is random and time-varying, so the whole system can be transformed into a stochastic system for the processing. Therefore, many scholars adopted the method of stochastic optimal control [6–7], used a linear stochastic system model to describe networked control systems with stochastic delay characteristics, and carried out linear two-order Gauss control. However, the stochastic control method requires that the network delay must obey a certain determined distribution. When the network delay does not obey a certain determined distribution, the stochastic control method is no longer applicable [8–9]. Luck and Ray [10] proposed a deterministic control idea, introduced the receive buffer, and transformed the random time-varying delay into fixed delay. And then in allusion to the fixed time delay, the controller is designed to solve the problem of uncertainty of the network delay. At present, the control methods of deterministic network remote control system mainly follow the three ideas of state predictive control [11], step change control [12] and Smith predictive control [13]. This paper introduces the three methods, and based on this, it focused on the research on state prediction control and staged transformation control based on state feedback, which combined with the infinite time two-order optimal controller. In addition, it proved the essence consistency of the two methods, and discussed the equivalence of the three control methods in the case that the system initial state is zero.

2. Method

2.1. State predictive control

According to Fig. 1, at the moment of kT, the controller receives the system state of $x(k)$, and the control law of $u(k)$ after operation is the moment of $(k+m)T$ when it is sent to the controlled object by the actuator, and at that time, the state of the controlled object has been changed to $x(k+m)$. Therefore, the idea of the state predictive control is that the controller, at the kT moment, uses the state equation of the controlled object to predict the system state of the $(k+m)T$ moment, and to send the control law in view of the state of the $(k+m)T$ moment.

Within the moments $(0, mT)$, because of the role of deterministic delay, the controlled objects cannot receive any control information. Starting from the mT moment, the actuator, with time as the T cycle, provided the control information $u(0)$, $u(1)$ and so on, for the controlled objects according to the order.

The system state $x(k)$ obtained at the moment of kT and the control amount of m systems before the moment of kT is $u(k-1), u(k-2), \cdots, u(k-m)$. And according to

$$\begin{cases} x(k+1) = Ax(k) + Bu(k-m), \\ y(k) = Cx(k) \end{cases}$$

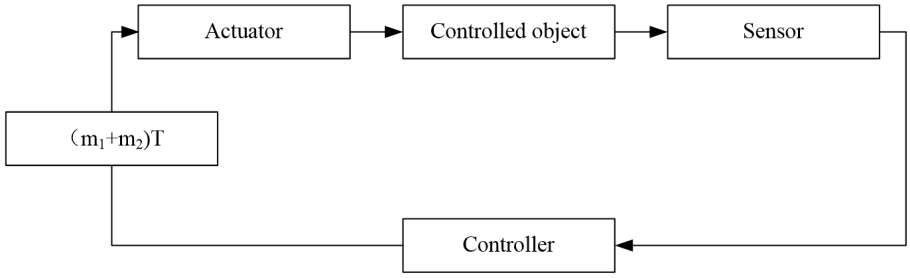


Fig. 1. Equivalent structure diagram of deterministic network remote control system

the system state at the moment of $(k+m)T$ can be predicted as $\overline{x}(k+m)$. Now

$$\begin{aligned}
&\overline{x}(k+m) = A\overline{x}(k+m-1) + Bu(k-1) \\
&= A\left[A\overline{x}(k+m-2) + Bu(k-2)\right] + Bu(k-1) \\
&= A^m x(k) + A^{m-1}Bu(k-m) + A^{m-2}Bu(k-m+1) + ... + Bu(k-1) \\
&= A^m x(k) + \sum_{i=1}^{m} A^{i-1}Bu(k-i).
\end{aligned}$$

And then from the predicted value $\overline{x}(k+m)$ of state, the finite time two-order optimal control law that is the mot matching with the controlled object actually receiving control information at the moment of $(k+m)T$ can be calculated as

$$u(k) = -K\overline{x}(k+m) .$$

At that time, the system state feedback gain matrix K is: $K = \left(R + B^{\mathrm{T}}PB\right)^{-1} \cdot B^{\mathrm{T}}PA$. Here, P refers to the positive definite solution of the algebra Riccati equation $P = A^{\mathrm{T}}PA - A^{\mathrm{T}}PB\left(R + B^{\mathrm{T}}PB\right)^{-1}B^{\mathrm{T}}PA + Q$. The corresponding performance indicator of infinite time is

$$J = \sum_{K=0}^{\infty}\left[x(k)^{\mathrm{T}}Qx(k) + u(k)^{\mathrm{T}}Ru(k)\right] .$$

2.2. Smith predictive control

The Smith predictive control is to the Smith predictor to compensate the pure delay, and transforms the time-delay system into an equivalent no delay system. When the initial state of the system is zero, the shape of the output response curve of the controlled object after compensation and the response characteristics are exactly the same as those without delay, but only a delay in time. It is supposed that the closed-loop system without delay is the state feedback control, and the system structure is shown in Fig. 2. According to the equivalent structure of network remote control system diagram shown in Fig. 1, a delay link is added before the controlled object in Fig. 2. The system structure diagram of using the Smith prediction control for the compensation is shown in Fig. 3, where the dotted line is the designed Smith predictor. If the model of the controlled object in the Smith predictor is very accu-

rate, Fig. 3 can be further simplified and Fig. 4 is obtained. It can be seen that the Smith predictive control is equivalent to passing a time delay value of the control action in the time coordinate.

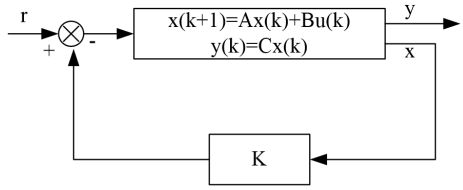


Fig. 2. Structure of closed loop control system without time delay

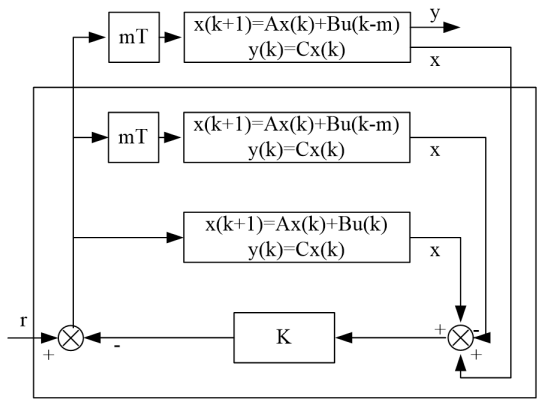


Fig. 3. Structure of network remote control system using Smith predictive control

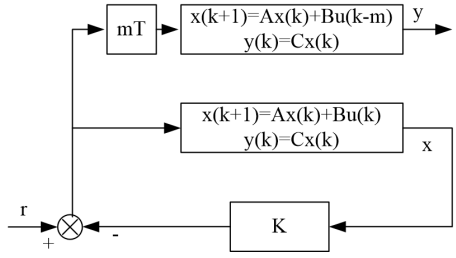


Fig. 4. Equivalent structure diagram of network remote control system using Smith predictive control

2.3. Step change control

The idea of step change control is: first of all, through a system state transformation, the system with delay is transformed into the system without delay. And then, according to the system without delay, the optimal state feedback control law

is designed. Finally, through the inverse transformation of the transformation made, the control law of the system with delay is obtained.

The state transformation with control memory is introduced:

$$z(k) = x(k) + \sum_{i=1}^{m} A^{i-m-1} B u(k-i) .$$

Then

$$z(k+1) = x(k+1) + \sum_{i=1}^{m} A^{i-m-1} B u(k+1-i) .$$

After substituting

$$\left\{ \begin{array}{l} x(k+1) = Ax(k) + Bu(k-m) \\ y(k) = Cx(k) \end{array} \right\} .$$

into the above formula, it can obtained

$$\begin{array}{l} z(k+1) = x(k+1) + \sum_{i=1}^{m} A^{i-m-1} B u(k+1-i) \\ = Ax(k) + Bu(k-m) + \sum_{i=1}^{m} A^{i-m-1} B u(k+1-i) \\ = Ax(k) + A^{-m} B u(k) + \sum_{i=1}^{m+1} A^{i-m-1} B u(k+1-i) \\ = A\left[x(k) + \sum_{i=1}^{m} A^{i-m-1} B u(k-i)\right] A^{-m} B u(k) = Az(k) + A^{-m} B u(k) . \end{array}$$

Let $\overline{B} = A^{-m} B$, then the above formula can be expressed as

$$z(k+1) = Az(k) + \overline{B} u(k) .$$

3. Basic idea of deterministic control

In this paper, we mainly study the cases of single data packet transmission and system single loop, without considering the loss of data packets during network transmission. The schematic diagram of the network remote control system with time delay can be shown in Fig. 5. In the figure, τ_{sc} suggests the transmission delay from the sensor nodes to controller nodes, and τ_{ca} refers to the transmission delay from the controller nodes to the actuator nodes. The operation time τ_{c} of the controller itself is generally incorporated into τ_{ca}. The execution time of the controlled object is negligible compared with the network delay. The influence of many factors on the network results in that the network delay τ_{sc} and τ_{ca} have stochastic time-varying characteristic, and the analysis of the system and the design of the controller are faced with great difficulties. The deterministic control theory is used to solve the delay problem in the network remote control system.

Assuming that the delay has boundaries, and $\tau_{\mathrm{sc}} \leq m_1 T$, $\tau_{\mathrm{ca}} \leq m_2 T$, where m_1 and m_2 are positive integers, and T is the sampling cycle of the system. The sensor, controller, and actuator all choose the working way of time driving, which is the requirements of deterministic control method.

The model for deterministic remote control system is constructed. In the re-

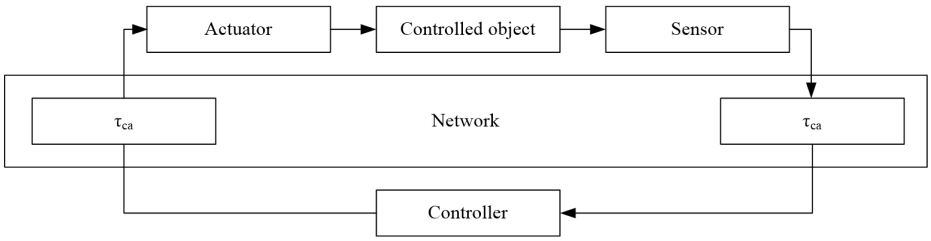


Fig. 5. Structure of network remote control system

ceiving terminal of controller and actuator, a FIFO queue buffer is arranged in the length of m_1T and m_2T, respectively. As a result, the transmission delay between the sensor to the controller is fixed as m_1T, and transmission delay between the controller to the actuator is fixed as m_2T. In this way, the random delay in the network is converted to fixed delay, and the design of the controller is simplified. This approach not only cleverly avoids the uncertainty of network delay, but also ensures the order of receiving information.

Figure 6 shows the closed loop system diagram of this deterministic control. Since the converted time delay is fixed, the controller can be exchanged with the delay section in terms of position, as shown in the upper and bottom parts of Fig. 7. For the controlled systems, the three structures are equivalent.

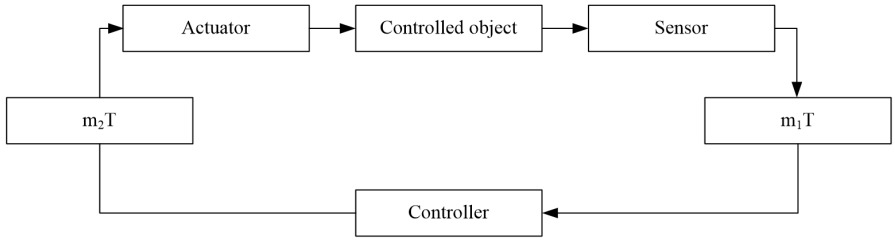


Fig. 6. Structure of deterministic network remote control system

4. Results and discussion

The sampling cycle of the system is $T = 0.05s$, and the discrete state equation of controlled object is:

$$x(k+1) = \begin{bmatrix} 1 & 0.04961 \\ 0 & 0.9843 \end{bmatrix} x(k) + \begin{bmatrix} 0.00014 \\ 0.00571 \end{bmatrix} u(k) ,$$

The infinite time two-order performance indicator function is chosen as

$$J = \sum_{K=0}^{\infty} \left[x(k)^T Q x(k) + u(k)^T R u(k) \right]$$

taking

$$Q = \begin{bmatrix} 50 & 0 \\ 0 & 1 \end{bmatrix},$$

$R = 0.01$, so as to solve the state feedback matrix in the state predictive control $K = [64.1596,\ 32.0961]$ and the state feedback matrix in the step change control $\widetilde{K} = [64.1596,\ 52.3870]$.

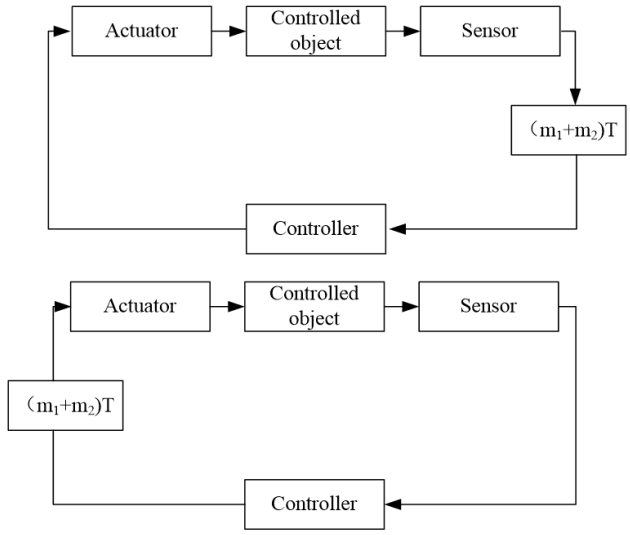


Fig. 7. Equivalent structure diagram of deterministic network remote control system

Assuming that the delays τ_{sc} and τ_{ca} in the network remote control system are randomly changing, and the maximum value is not more than 4 times of T value. Let $m = 8T$, the deterministic control method is adopted [14]. Figure 8 shows the output response curve of the system when the initial state of the controlled object is $[1,\ -2]^{\mathrm{T}}$, and the State Predictive Control and the step change control are adopted, respectively. It can be seen from the simulation results that the two curves coincide completely, and the essential consistency between the state predictive control and the step change control is verified. Figure 9 shows the step response curves of the system under three conditions, namely, the state predictive control, the step change control and the Smith predictive control, when the initial state of the controlled object is zero. The three curves also coincide completely, and the equivalence of the three methods when the initial state of the system is zero is verified. Figure 10 shows the output response curves of the system under three conditions, namely, the state predictive control, the step change control and the Smith predictive control when the initial state of the controlled object is $[1,\ -2]^{\mathrm{T}}$. The results show that, when the initial state of the system is not zero, the control effects of the state predictive control and the step change control are better than the traditional Smith predictive control.

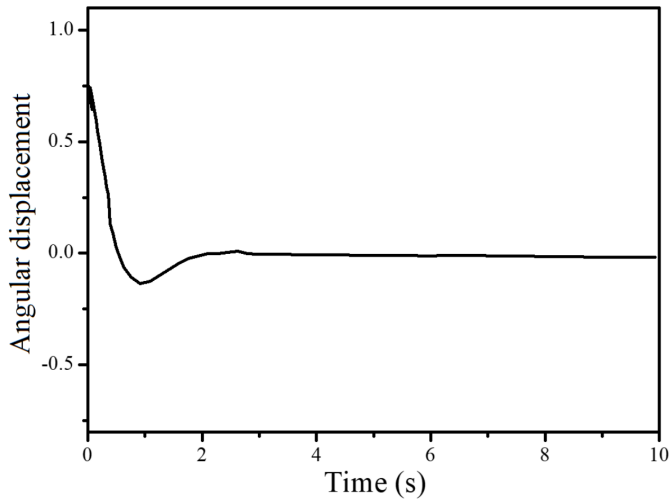


Fig. 8. Output response curves of state predictive control and step change control

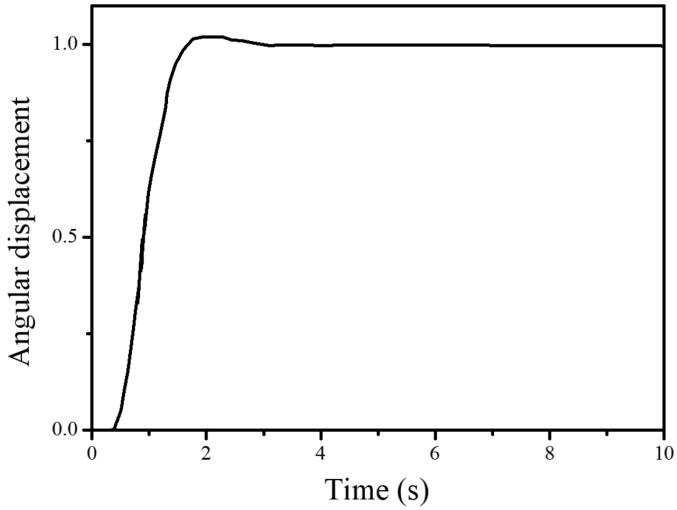


Fig. 9. Output response curves of three control methods when the initial state is zero

5. Conclusion

In this paper, three control methods of deterministic network remote control system are discussed, namely the Smith predictive control, the state predictive control and the step change control. On the basis of research idea of these three methods, through theoretical analysis, the essence consistency of state predictive control and step change control is verified. In addition, the equivalence relation of the two in

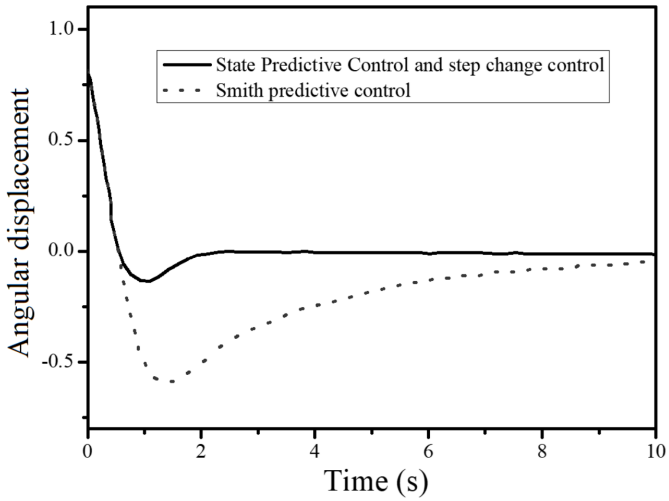


Fig. 10. Output response curves of three control methods when the initial state is not zero

the case of initial state of system of zero with the Smith prediction control. The simulation results verify the correctness of the obtained results.

References

[1] G. Yang, L. Xie, M. Mäntysalo, X. Zhou, Z. Pang, L. D. Xu, S. Kao-Walter, Q. L. Chen, R. Zheng: *A health-iot platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box.* IEEE Transactions on Industrial Informatics *10* (2014), No. 4, 2180–2191.

[2] M. A. Harris, K. P. Patten: *Mobile device security considerations for small- and medium-sized enterprise business mobility.* Information Management & Computer Security *22* (2014), No. 1, 97–114.

[3] D. H. Jenkins, J. F. Rappold, J. F. Badloe, O. Berséus, L. Blackbourne, K. H. Brohi, F. K. Butler, A. P. Cap, M. J. Cohen, R. Davenport, M. de Pasquale, H. Doughty, E. Glassberg, T. Hervig, T. J. Hooper, R. Kozar, M. Maegele, E. E. Moore, A. Murdock, P. M. Ness, S. Pati, T. Rasmussen, A. Sailliol, M. A. Schreiber, G. A. Sunde, L. M. van de Watering, K. R. Ward, R. B. Weiskopf, N. J. White, G. Strandenes, P. C. Spinella: *Trauma hemostasis and oxygenation research position paper on remote damage control resuscitation: Definitions, current practice, and knowledge gaps.* Shock *41* (2014), Suppl. No. 1, 3–12.

[4] U. Bodin, K. Wolosz: *Proportional throughput differentiationwith cognitive load-control on WSN channels.* EURASIP Journal on Wireless Communications and Networking (2015), No. 1, paper 186.

[5] T. N. Le, W. L. Chin, H. H. Chen: *Standardization and security for smart grid communications based on cognitive radio technologies—a comprehensive survey.* IEEE Communications Surveys & Tutorials *19* (2017), No. 1, 423–445.

[6] M. Mehmetoglu, E. Akyol, K. Rose: *Deterministic annealing optimization for Witsenhausen's and related decentralized stochastic control problems.* CoRR abs/1607.02893 (2016), arXiv: 1607.02893v1 [cs.SY] 11 Jul 2016.

[7] R. Caballero-Águila, A. Hermoso-Carazo, J. Linares-Pérez: *Optimal state estimation for networked systems with random parameter matrices, correlated noises and delayed measurements*. International Journal of General Systems *44* (2015), No. 2, 142–154.

[8] Z. Zhang, K. Long, J. Wang, F. Dressler: *On swarm intelligence inspired self-organized networking: its bionic mechanisms, designing principles and optimization approaches*. IEEE Communications Surveys & Tutorials *16* (2014), No. 1, 513–537.

[9] S. Huang, B. Liang, J. Li: *Distributed interference and delay aware design for D2D communication in large wireless networks with adaptive interference estimation*. IEEE Transactions on Wireless Communications *16* (2017), No. 6, 3924–3939.

[10] R. Luck, A. Ray: *Experimental verification of a delay compensation algorithm for integrated communication and control systems*. International Journal of Control *59*, (1994), No. 6, 1357–1372.

[11] M. H. Vafaie, B. M. Dehkordi, P. Moallem, A. Kiyoumarsi: *A new predictive direct torque control method for improving both steady-state and transient-state operations of the PMSM*. IEEE Transactions on Power Electronics *31* (2016), No. 5, 3738–3753.

[12] S. Anand, B. G. Fernandes, J. Guerrero: *Distributed control to ensure proportional load sharing and improve voltage regulation in low-voltage DC microgrids*. IEEE Transactions on Power Electronics *28* (2013), No. 4, 1900–1913.

[13] G. A. Papafotiou, G. D. Demetriades, V. G. Agelidis: *Technology readiness assessment of model predictive control in medium-and high-voltage power electronics*. IEEE Transactions on Industrial Electronics *63* (2016), No. 9, 5807–5815.

[14] S. Vaidyanathan, S. Sampath, A. T. Azar: *Global chaos synchronisation of identical chaotic systems via novel sliding mode control method and its application to Zhu system*. International Journal of Modelling, Identification and Control *23* (2015), No. 1, 92–100.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 339–350

Multi-circuit system observational data fusion method

Honglai Yan[1]

Abstract. At present, the fusion method of multi-circuit system observation data is complicated and the calculation process is large, which is difficult to be extended to the observation application of grid space. This paper presents a high-precision multi-circuit system observation data fusion method. Firstly, a high precision multi-circuit system observational data fusion method is put forward on the basis of the redundancy set estimation (hereinafter referred to as RSE for short), which combines the process of the information fusion with the set operation link of the algorithm itself. Finally, in order to verify the feasibility and effectiveness of the method studied in this paper, the grid observational data fusion simulation experiment is carried out. This method can guarantee the relatively high observation precision, and at the same time, does not significantly increase the amount of calculation of the single-circuit system redundancy set estimation algorithm, therefore, is of relatively high real-time performance.

Key words. Multi-circuit system, active observational data fusion, redundancy set estimation, optimal angle.

1. Introduction

In the current society, the application of the multi-circuit systems is becoming more and more popular, and various multi-circuit systems with high intelligence are playing or about to play an important role in the different occasions. However, with the continuous extension of the fields of the human social activities and the continuous development of the circuit systems research, the single multi-circuit system is faced a number of difficulties in the replacement of human beings to accomplish the missions including large-scale disaster relief, scientific inspection and so on as well as the battlefield environmental monitoring and other military missions, for example: poor reliability, small operating range, and low task accomplishment efficiency, etc. While the relevant researches show that, the multi-circuit system composed of a number of circuit systems can exactly solve these problems through the coordination and cooperation [1, 2]. Therefore, the multi-circuit system is considered to be of broad application prospects.

[1]Xi'an International University, 710077, Xi'an, Shaanxi, China

At present, majority of the passive data fusion methods are based on certain estimation method, and the most widely used method is the traditional estimation method on the basis of the probabilistic statistical theory. Among them, literature [3] made use of the Kalman filter method to deduce the relationship between the covariance matrix and the state variables of the multi-circuit system, and it has proven that the covariance matrix of the minimized target position is equivalent to the maximum eigenvalue of the minimized target covariance matrix, so as to obtain the condition of optimal observation of the multi-circuit system through calculation. In literature [4], the problem of observing ground moving targets through Doppler-based adaptive observer was studied. In this paper, the relationship between the observation results and the noise as well as the observation angle of the single circuit system was discussed. And then the Fisher information matrix and Cramer-Rao boundary principle was adopted, by optimizing the covariance matrix of the observation results, the conditions of optimal observation were obtained, so as to realize the optimization of the position and velocity of the ground targets. In literature [5], the authors assumed that the target states observational by several different circuit systems obeyed the normal distribution, and then applied the theory of distribution multiplication to fuse the multiple observations so as to obtain a more precise observational data fusion with smaller uncertainty, and achieve the purpose of the multi-circuit system coordinated observation to improve the measurement precision. In view of this, a new nonlinear filtering method, namely, the Extended Set-membership Filter (hereinafter referred to as RSE for short), has been proposed and applied to the fusion of observational data. This method only requires the noise distribution to be bounded, which can be satisfied during the actual observation [6]. In literature [7], the authors tried to solve the problem of multi-circuit system observational data fusion by the application of the RSE method.

In this paper, a method of active observational data fusion of the multi-circuit system on the basis of RSE algorithm in the grid space is proposed. This method includes two parts: Firstly, on the basis of the detailed analysis of the work in literature [8], the algorithm is improved, and the observational data fusion algorithm of the multi-circuit system and the RSE estimation algorithm are merged organically to improve the real-timing, and precision of the data fusion algorithm. Then, by the optimization of the observation conditions, a method of coordinated planning of the behavior of the multi-circuit systems by the application of the concept of optimal angle is put forward to optimize the observation results, so that the active observational data fusion of the multi-circuit system can be finally realized. Therefore, in this paper, both the circuit system and the moving target are represented by the following kinematic model on the basis of Newton's law of motion (after discretization), respectively expressed as the following:

$$\begin{cases} p_{i,k+1}^R = p_{i,k}^R + v_{i,k}^R \cdot \Delta T, \\ v_{i,k+1}^R = v_{i,k}^R + a_{i,k}^R \cdot \Delta T, \\ a_{i,k+1}^R = u_{i,k}, \end{cases} \tag{1}$$

$$\Gamma\left(p_{i,k}^{R}, v_{i,k}^{R}, a_{i,k}^{R}\right) \leq 0\,,$$

$$\begin{cases} p_{k+1}^{T} = p_{k}^{T} + v_{k}^{T} \cdot \Delta T + \omega_{1,K}, \\ v_{k+1}^{T} = \omega_{2,k}, \end{cases} \tag{2}$$

$$y_{i,k} = h_i\left(p_k, v_k\right) + \omega_{3,i,K}.$$

Here, the superscripts R and T, respectively, stand for the circuit system and the target system; $p_{i,k} = (x_i, k, y_i, k, z_i, k)^T$ $(i = 1, 2, \ldots n)$ stands for the position of the system at the time k of the i-th circuit system; $v_{i,k}^{R} = (v_{i,x,k}, v_{i,y,k}, v_{i,z,k})^T$ and $a_{i,k}^{R} = (a_{i,x,k}, a_{i,y,k}, a_{i,z,k})^T$ stand for the velocity and the acceleration at the time k respectively; $u_{i,k}$ stands for the control input of the i-th circuit system at the time k; it is assumed in this paper that the velocity variation of the circuit system u_k, that is, the acceleration is the controllable input of the circuit system; the sampling time; the motion constraint inequation of the system of the circuit system; ΔT stands for the sampling time, $\Gamma(\cdot, \cdot, \cdot)$ stands for the motion constraint inequation of the circuit system; p_k^T stands for the position information of the target circuit at the time k, the corresponding v_k^T stands for the velocity information of the target circuit at the time k; $y_{i,k}$ stands for the observation value of the target circuit by the i-th circuit system at the time k. In this paper, it is assumed that the circuit system does not have any priori knowledge of the motion performance of the target circuit, and thus it is impossible to conduct accurate prediction. Therefore, only the second-order motion equation is considered, rather than the third-order equation of motion as shown in the circuit system (1). Symbols $\omega_{j,k}$ $(j = 1, 2)$ stand for the process noise of the moving object; $\omega_{3,i,k}$ stands for the measurement noise when the target circuit is observed by the i-th circuit system. According to the assumption, all the three noise vectors shall satisfy the following conditions:

$$\omega^{\mathrm{T}} Q^{-1} \omega \leq 1\,.$$

where Q is a positive definite symmetric matrix.

In this paper, it is assumed that the observation of all the circuit systems on the target is implemented by the adaptive observer, and the observation equations are

$$r_{i,k} = \left[\left(x_{T,k} - x_{i,k}\right)^2 + \left(y_{T,k} - y_{i,k}\right)^2 + \left(z_{T,k} - z_{i,k}\right)^2\right]^{1/2} + n_{r,i}\,, \tag{3}$$

$$\theta_{i,k} = \tan^{-1} \frac{z_{T,k} - z_{i,k}}{\left[\left(x_{T,k} - x_{i,k}\right)^2 + \left(y_{T,k} - y_{i,k}\right)^2\right]^{1/2}} + n_{\theta,i}\,, \tag{4}$$

$$\alpha_{i,k} = \tan^{-1} \frac{y_{T,k} - y_{i,k}}{x_{T,k} - x_{i,k}} + n_{\alpha,i}\,. \tag{5}$$

In these three equations, $r_{i,k}, \theta_{i,k}, \alpha_{i,k}$ stand for the three observations obtained by the adaptive observer, as shown in Fig. 1. There, $(n_{r,i}, n_{\theta,i}, n_{\alpha,i})^T$ stand for the

measurement of the noise, namely, the $\omega_{3,i,k}$ in equation (2). It should be noted that if the observer is not an adaptive observer, the corresponding observation equation $h(\cdot)$ will be changed accordingly, but this will not affect the application of the observational data fusion method described in this paper. In addition, different circuit systems can also carry different observers, and the corresponding observation equation will be different as well.

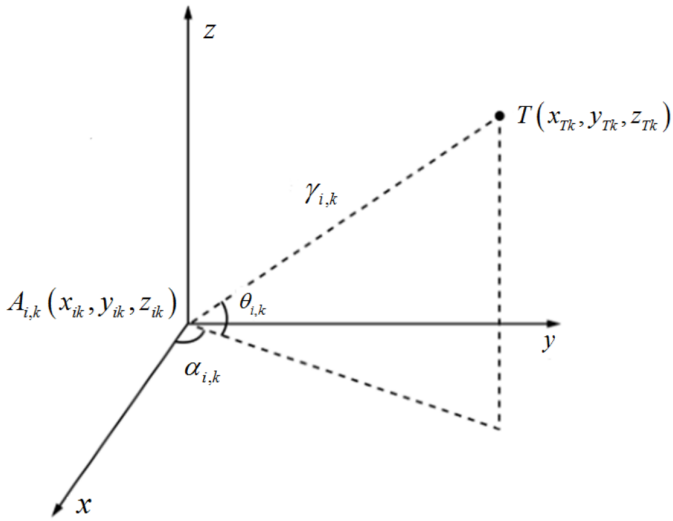


Fig. 1

2. Observational data fusion method on the basis of RSE

This sectionmainly solves how to conduct fusion on the Observational data for the multi-circuit system in the active Observational data fusion, so as to obtain more precise target state information, that is, the passive Observational data fusion problem.

According to the principle of the RSE method, when observing the target with a circuit system, the target should be included in the observational observation set; if another circuit system observes the same target simultaneously, another observation set containing the target shall be obtained, and the rest can be deduced in the same manner, the target should be included in the intersection of the observations set and the predictions set of these circuit systems. Figure 2 shows the diagram of the observational data fusion of two circuit systems. Therefore, by solving the intersection of these sets more accurate results can be obtained.

In this paper, the fusion method of observational data proposed makes use of the characteristics of RSE's own calculation, and integrates this intersection solving process into the estimation method, that is, the process of solving the intersection of the prediction set and the observation set is replaced by solving the intersection of the prediction set and two observation sets. The detailed procedure of the algorithm

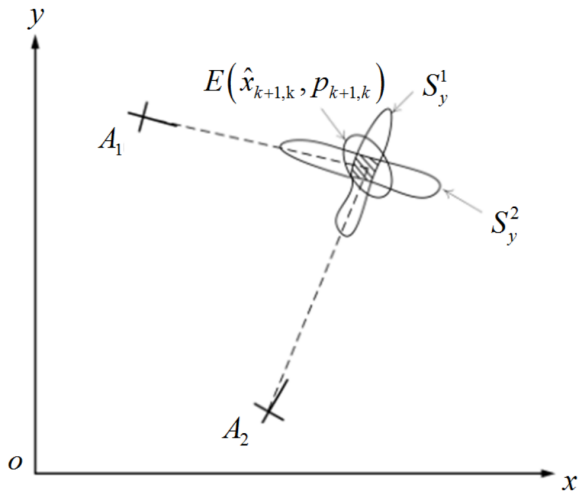


Fig. 2. Basic idea of the observational data fusion method

is shown as the following (the variable with the left superscript sign $i\,(i = 1, 2.., n)$ indicates that the variable is calculated from the observational data of the first circuit system. For the ease of understanding, this paper only sets out the algorithm with one collaborative circuit system, and the multiple collaborative circuit system method can be deduced in the same manner):

Step 1. Initialization

$$\hat{x}_{k,k} = \hat{x}_I, P_{k,k} = P_I\,.$$

Step 2. Single machine forecasting

$$\hat{x}_{k+1,k} = f\left(\hat{x}_{k,k}\right)\,, \tag{6}$$

$$P_{k+1,k} = \frac{A_k p_k A_k^T}{1-\beta_k} + \frac{\hat{Q}_k}{\beta_k}\,. \tag{7}$$

Here,

$$A_k = \frac{\partial f\left(x_k\right)}{\partial x}|_{x_k=\hat{x}_{k+1,k}}\,.$$

Step 3. Collaborative updates

Step 3.1. Fusion of the collaborative circuit system observational data

$${}^2\hat{x}_{k+1,k+1} = \hat{x}_{k+1,k} + {}^2K_k\left({}^2y_{k+1} - h\left(\hat{x}_{k+1,k}\right)\right)\,, \tag{8}$$

$${}^2P_{k+1,k+1} = {}^2\delta_k \times \frac{P_{k+1,k}}{1-{}^2\rho_k} - {}^2\delta_k \times \frac{P_{k+1,k}}{1-{}^2\rho_k} \times C_{k+1}^T\,{}^2W_k^{-1} C_{k+1} \times \frac{P_{k+1,k}}{1-{}^2\rho_k}\,, \tag{9}$$

where

$$C_{k+1}=\frac{\partial h\left(x_k\right)}{\partial x}|_{x_k=\hat{x}_{k+1,k}},\quad {}^2W_k=\frac{C_{k+1}P_{k+1,k}C_{k+1}^T}{1-{}^2\rho_k}+\frac{{}^2\hat{R}_{T,k+1}}{{}^2\rho_k},$$

$${}^2K_k=\frac{P_{k+1,k}}{1-{}^2\rho_k}C_{k+1}^T{}^2W_k^{-1},$$

$${}^2\delta_k=1-\left[{}^2y_{k+1}-h\left(\hat{x}_{k+1,k}\right)\right]^T\times{}^2W_k^{-1}\times\left[{}^2y_{k+1}-h\left(\hat{x}_{k+1,k}\right)\right],$$

$${}^2\hat{R}_{T,k+1}=\frac{{}^2\bar{R}_{T,k+1}}{1-{}^2\beta_R}+\frac{R_{k+1}}{{}^2\beta_R},\quad {}^2\beta_R=\frac{\sqrt{\operatorname{tr}\left(R_{T,k+1}\right)}}{\sqrt{\operatorname{tr}\left({}^2\bar{R}_{T,k+1}\right)}+\sqrt{\operatorname{tr}\left(R_{T,k+1}\right)}},$$

$${}^2\rho_k=\frac{\sqrt{\operatorname{tr}\left({}^2\hat{R}_{T,k+1}\right)}}{\sqrt{\operatorname{tr}\left(C_{k+1}P_{k+1}C_{k+1}^T\right)}+\sqrt{\operatorname{tr}\left({}^2\hat{R}_{T,k+1}\right)}}.$$

Step 3.2. Fusion of the main circuit system observational data

$$\hat{x}_{k+1,k+1}={}^2\hat{x}_{k+1,k+1}+{}^1K_k\left({}^1y_{k+1}-h\left({}^1\hat{x}_{k+1,k+1}\right)\right), \tag{10}$$

$$P_{k+1,k+1}={}^1\partial_k\frac{{}^2P_{k+1,k+1}}{1-{}^1\rho_k}-{}^1\delta_k\frac{{}^2P_{k+1,k+1}}{1-{}^1\rho_k}\times{}^2C_{k+1}^T\times\frac{{}^2P_{k+1,k+1}}{1-{}^1\rho_k}. \tag{11}$$

Here,

$${}^2C_{k+1}=\frac{\partial h\left(x_k\right)}{\partial x}|_{x_k={}^2\hat{x}_{k+1,k+1}},\quad {}^1W_k={}^2C_{k+1}\left(\frac{{}^2P_{k+1,k+1}}{1-{}^1\rho_k}\right){}^2C_{k+1}^T+\frac{{}^1\hat{R}_{T,k+1}}{{}^1\rho_k},$$

$${}^1K_k=\frac{{}^2P_{k+1,k+1}}{1-{}^1\rho_k}\times{}^2C_{k+1}^T\times{}^1W_k^{-1},$$

$${}^1\delta_k=1-\left[{}^1y_{k+1}-h\left({}^2\hat{x}_{k+1,k+1}\right)\right]^T\times{}^1W_k^{-1}\times\left[{}^1y_{k+1}-h\left({}^2\hat{x}_{k+1,k+1}\right)\right],$$

$${}^1\hat{R}_{T,k+1}=\frac{{}^1\hat{R}_{T,k+1}}{1-{}^1\beta_R}+\frac{R_{T,k+1}}{{}^1\beta_R},{}^1\beta_R=\frac{\sqrt{tr\left(R_{T,k+1}\right)}}{\sqrt{tr\left({}^1\bar{R}_{T,k+1}\right)}+\sqrt{tr\left(R_{T,k+1}\right)}},$$

$${}^1\rho_k=\frac{\sqrt{tr\left({}^1\hat{R}_{T,k+1}\right)}}{\sqrt{tr\left({}^2C_{k+1}{}^2P_{k+1,k+1}{}^2C_{k+1}^T\right)}+\sqrt{tr\left({}^1\hat{R}_{T,k+1}\right)}}.$$

From this the location set $E\left(\hat{x}_{k+1,k+1}\right)$, $P_{k+1,k+1}$ of the targets can be estimated.

After the initialization of the main circuit system is completed, the prediction set of the target location is calculated by the application of equations (6) and (7), which are exactly the same as the prediction process of the single machine. It can also be

seen from equations (6) and (7) that the results of the prediction are related only to the state of the target at the previous time, while irrelevant to the state of the circuit system, that is, no matter what the state of the two circuit systems is, the prediction set of the target is the same, so it is only required for the prediction step to complete the main circuit system. Then the main circuit system calculates the corresponding observation set according to the observational data of the collaborative circuit system and its own observational data, and calculates the intersection of the three sets by the application of the RSE update algorithm twice, and then the fusion result of the observational data can be obtained. The aforementioned process to solve the intersection of these three sets of processes can be divided into the following two steps:

1) First solve the intersection of the observation sets S_y^2 of the observation set $E(\hat{x}_{k+1,k})$, $P_{k+1,k}$ and the collaborative circuit system. This step is exactly the same as the RSE of the single machine, therefore, the calculation result $E\left({}^2\hat{x}_{k+1,k+1}\right)$, ${}^2P_{k+1,k+1}$ is an ellipsoidal set that meets the equation (6), denoted as E_{temp};

2) Solve the intersection of E_{temp} and the main circuit system observation set. In this process, E_{temp} is regarded as a prediction set, introduce the set into the RSE method, and make use of the main circuit system to update its observation set so as to obtain the intersection of these two sets, which is the time observational data fusion result.

Under normal circumstances, according to the assumption of RSE, both of the circuit systems of the observational data set contain the real system state points; therefore, the intersection must be non-empty. However, when some extreme conditions are encountered (for example: Some observers fail and result in large errors in the observational data), the observation sets of the two circuit systems may not intersect, that is $E_{\text{temp}} \bigcap S_y^1 = \phi$. In this case, in order to be able enable the algorithm to be carried on recursively, we can directly take the estimated results in 1) as the final observation results, and carry out the next recursive operation.

3. Simulation experiment

The simulation experiment is carried out by Matlab on the PC platform. And the experimental parameters are as the following: Envelope matrix of the process noise: $Q = diag\{0.0001, 0.0001, 0.0001\}$; Envelope matrix of the observation noise: $R = diag\{0.001, 0.001, 0.001\}$; Envelope matrix of the initial target state: $P_0 = diag\{0.1, 0.1, 0.1\}$; Weight value: $\omega_1 = 0.01$, $\omega_2 = 0.0005$, $\omega_3 = 2.7$, $\omega_4 = 1$, $\omega_5 = 1.1$.

Figure 3 shows the trajectory of the active observation process of the two circuits. The top left figure is the aerial view of the grid. The other three figures are the projection of the trajectory of the circuit in plane $x - y$, plane $y - z$ and plane $x - z$. For the ease of viewing, all the ellipsoids in the figure are all magnified by a factor of 15. The middle solid line in the figure indicates the trajectory of the movement of the target, and the other two solid lines represent the trajectory of the two circuit systems, respectively. It can be seen from the figure that, with the progress of observation, the uncertain ellipsoid set of the target state becomes

smaller and smaller, that is, the observation on the target becomes more and more precise. To further validate the method proposed in this paper, we assume that there is a sudden change in the trajectory of the target, but it can be clearly seen from the figure that this mutation does not affect the tracking of the target by the two circuit systems, and that they can still properly plan their own trajectories and perform observation on the target, the ellipsoid set of the target state can converge to a relatively smaller stable value.

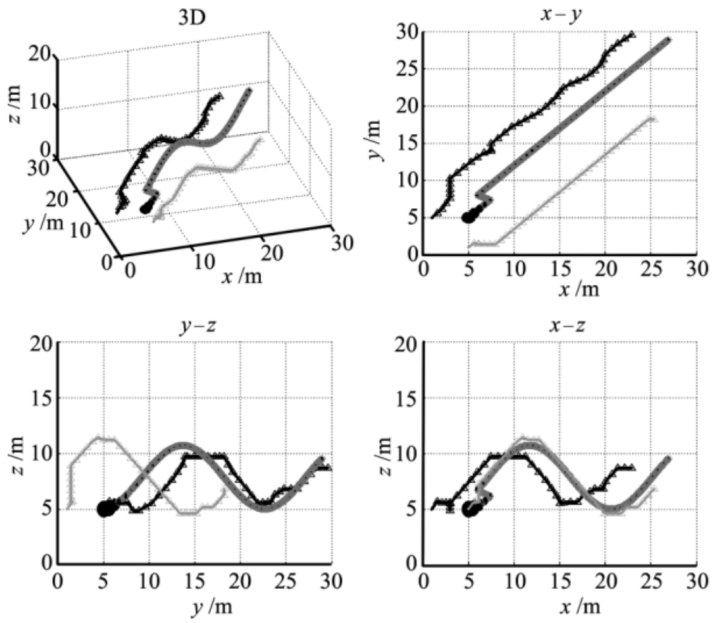


Fig. 3. Trajectory of the active observational data fusion process

The length of three axes of the ellipsoid can reflect the size of the ellipsoid, so the trajectory of the envelope matrix $P_{k,k}$ in equation (8) can be used as a parameter to measure the ellipsoid size. Figure 4 shows the variation trend of the trajectory of the matrix $P_{k,k}$, which is a quantitative description of the variation of the ellipsoid set. The solid line in the figure shows the results obtained with the observational data of only one circuit system, and the dotted line shows the result of data fusion by the adoption of two circuit systems. The results of the observational data fusion are significantly superior to those of the single machine observations (0.0090 vs. 0.0034) from both the convergence rate and the observation precision. In the figure, there is a sudden variation in the vicinity of 5 s, which is due to the sudden variation of the target trajectory. However, after the mutation, the two circuits system can make the observational ellipsoid converge to a relatively smaller value by properly planning the respective trajectories, which indicates that the algorithm proposed in this paper has good stability.

Figure 5 shows the variation of the observational data fusion observational angle

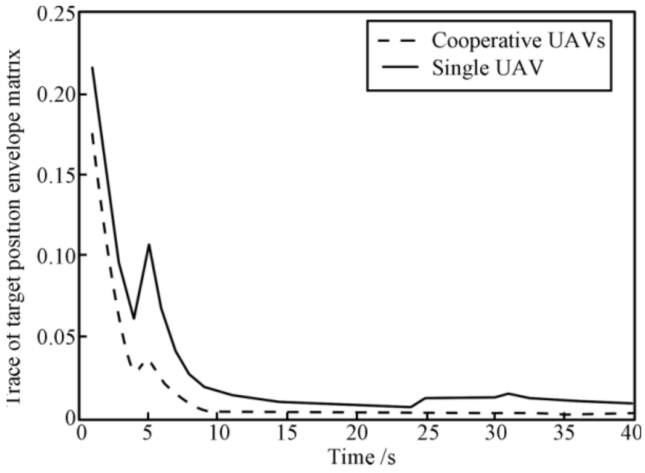


Fig. 4. Variation of the trajectory of target state envelope matrix

during the process of observation, and the observational data fusion angle maintains at 90° or so in the whole observation process. Therefore, the method proposed in this paper can effectively plan the path of the circuit system, and realize the approximation optimization fusion of the observational data on the target.

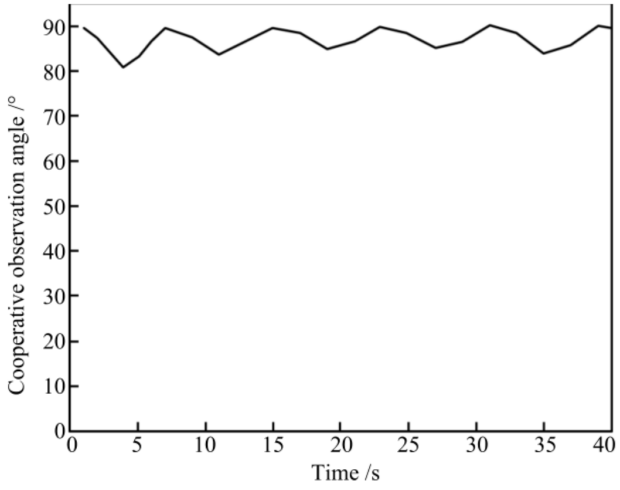


Fig. 5. Variation of the observational data fusion angle

In order to illustrate the fastness of the proposed method in more detail, we compare the observational data fusion method mentioned in Section 2.2 with the method proposed in literature [9]. Table 1 lists the computation time of the two algorithms under the scenario of three circuit system observational data fusion. It can be clearly observed from the table that, the average time of iteration is 0.313 s

at each step of the data fusion method proposed in this paper, which is very close to that of the single-circuit system observation method. From the simple passive observational data fusion method, the data fusion process of the method proposed in this paper does not introduce too many computations. While the method proposed in literature [10] consumes an average of 0.391 s for the iteration each time, which is 1.2 times of the method proposed in this paper. Therefore, it can be seen that the algorithm proposed in this paper is more applicable for the real-time application environment. And the differences between these two algorithms are shown in literature [11].

Table 1. Time comparison of the two observational data fusion methods

Algorithm	Time (s)
Single circuit system observation method	0.311
Three-circuit system observational data fusion (the method proposed in this paper)	0.313
Three-circuit system observational data fusion (the method put forward in literature [7])	0.391

4. Conclusion

In this paper, a RSE-based multi-circuit data active observation fusion method is put forward. This method makes use of the calculation characteristics of RSE and integrates the fusion process of the observation results of the multi-circuit system into the estimation algorithm, so as to realize the real-time optimization observation of the target. The main advantages of this method are summarized as the following:

1) The optimization observation on the target can be realized for the circuit system so as to ensure the precision of the observation.

2) On the basis of the estimation method, the observational data fusion of the two circuits is implemented through the introduction of less calculation to improve the rapidity of the algorithm.

3) In addition, less approximation process is introduced to improve the precision of the observation results. Finally, the simulation results are provided to verify the feasibility and effectiveness of the algorithm.

References

[1] C. L. Fu, D. L. Gong, L. I. Jie: *Multi-sensor consistency data fusion algorithm in observation uncertainty.* Transducer & Microsystem Technologies *32* (2013), No. 7, paper 113.

[2] U. Zengin, A. Dogan: *Real-time target tracking for autonomous uavs in adversarial environments, a gradient search algorithm.* IEEE Transactions on Robotics *23* (2007), No. 2, 294–307.

[3] K. Zhou, S. I. Roumeliotis: *Optimal motion strategies for range-only constrained*

multisensor target tracking. IEEE Transactions on Robotics *24* (2008), No. 5, 1168–1185.

[4] G. Gu, P. R. Chandler, C. J. Schumacher, A. Sparks, M. Pachter: *Optimal cooperative sensing using a team of UAVs*. IEEE Transactions on Aerospace and Electronic Systems *42* (2006), No. 4, 1446–1458.

[5] L. Wang, J. W. Wan, Y. H. Liu, J. X. Shao: *Cooperative localization method for multi-robot based on PF-EKF*. Science in China Series F: Information Sciences *51* (2008), No. 8, 1125–1137.

[6] B. Zhou, J. D. Han: *A UD factorization-based adaptive extended set-membership filter*. Acta Automatica Sinica *34* (2008), No. 2, 150–158.

[7] J. Ousingsawat, M. E. Campbell: *On-line estimation and path planning for multiple vehicles in an uncertain environment*. International Journal of Robust and Nonlinear Control *14* (2004), No. 8, 741–766.

[8] P. Yang, R. A. Freeman, K. M. Lynch: *Multi-agent coordination by decentralized estimation and control*. IEEE Transactions on Automatic Control *53* (2008), No. 11, 2480–2496.

[9] C. D. Pathiranage, K. Watanabe, K. Izumi: *T–S fuzzy model adopted SLAM algorithm with linear programming based data association for mobile robots*. Soft Computing *14* (2010), No. 4, 345–364.

[10] D. Zu, J. D. Han, D. L. Tan: *Transverse vibrations of non-homogeneous rectangular plates with variable thickness*. Acta Automatica Sinica *33* (2007), No. 10, 1036–1042.

[11] E. A. Arkenbout, J. C. F. de Winter, P. Breedveld: *Robust hand motion tracking through data fusion of 5DT data glove and nimble VR kinect camera measurements*. Sensors (Basel) *15* (2015), No. 12, 31644–31671.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 351–356

Design of wood drying control system based on PLC

Donglin Wang[1,2], Songlin Yi[2], Qin Yang[2]

Abstract. The wood drying process control is a very important part, directly related to the quality of timber drying. In this paper a control system based on programmable logic controller (PLC) and the configuration interface are designed to dry wood. The system can realize automatic control of drying through the host computer PC and PLC. First a suitable drying criterion is set and the drying process is conducted on the timber. At the same time, the data of temperature, humidity and wood moisture content etc. in the kiln are collected. After filtering and standardization, the above data are put into the PID table. The proposed algorithm is applied to achieve PID regulator, then the interface outputs signal to control the opening of the corresponding valve to adjust the temperature and humidity in kiln and the wood moisture content of timber. Experimental results show that the control system can achieve the real requirements.

Key words. Wood drying, PLC, configuration, Proportion Integration Differentiation (PID)
.

1. Introduction

The wood drying is one of the important measures which can preserve the good characteristics of wood, enhance materials, and rationally use of wood. The wood drying process control is a very important part, directly related to the quality of timber drying the pros. This paper uses programmable logic controller (PLC), which is widely used in the field of automatic control, to establish the control system for wood drying. Use PID as regulating means to control temperature, humidity, and moisture content and other drying parameters. In accordance with the drying technologies to achieve effective and accurate real-time control purpose of the drying process.

[1]Corresponding author
[2]Beijing Forestry University, Beijing, 100083, China

1.1. The structure of control system

The core control part of the system adopts the combination of lower computer-PLC and upper computer-PC through configuration software [1–3]. PLC has the advantage of working stability, strong anti-interference, programming convenience, and is easy to expand etc. [4]. PLC can form a good human-computer interface through configuration programming which can observe the real-time change of parameters in the drying process and modify the various control parameters and techniques. PLC controls the various implementing agencies in the drying scene and make data processing[5–6]. PC displays and saves all kinds of real-time parameters and alarm messages. It also can set and modify parameters and calculate parameters by intelligent algorithms. System block diagram is shown in Fig. 1.

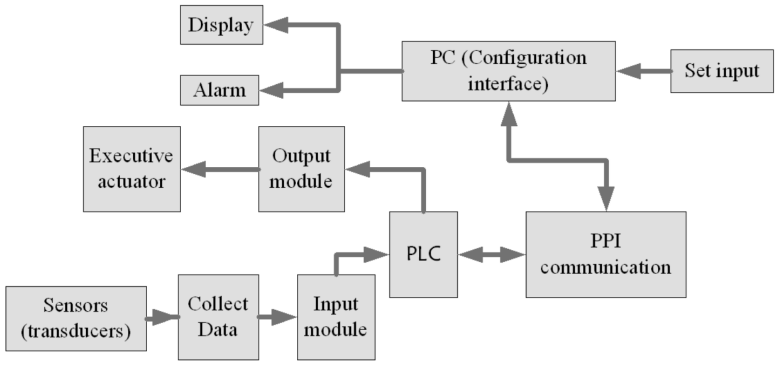


Fig. 1. Structure of system

First of all, select or input a suitable drying criterion bases on timber to be dried and the whole process of drying is according to the criterion to complete. The first work of the system is to collect the real-time data of temperature, humidity, wood moisture content, etc. in the kiln. The system is designed as six-way loop collection program. Eliminate the filter wave in bad data and arithmetic mean value to get the real-time temperature, humidity and wood moisture content. To obtain the given value of temperature and humidity according to the current moisture content and the drying criterion. And after the standardization of data, put it into the PID table which is corresponding to temperature and humidity. The adjusted program is designed to achieve PID regulator. The analog output interface outputs signal to control the opening of the corresponding valve in order to control the temperature and humidity in kiln and adjust the wood moisture content, drying timber in the drying criterion.

1.2. PLC Selection and module expansion

The system uses Siemens S7-200 series PLC with 224 CPU [7]. The total expansion of analog I/O module are five EM235. This is shown in Fig. 2.

All sensors (transducers) have in the system output 0–10 V voltage signal. The

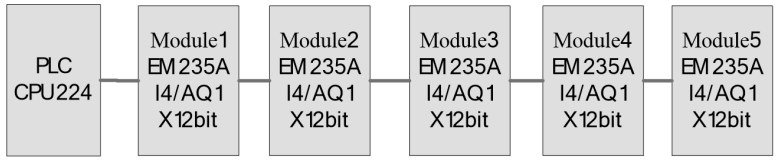


Fig. 2. Hardware module chart

five EM235-analog input and output modules all use voltage wiring, while DIP switch is slid to 0–10V. The system uses multi-point acquisition mode, the setting are as follows: Input analog signal: kiln temperature × 6 Road, kiln humidity × 6 Road, wood moisture content × 6 Road. Output analog signal: heating × 1 Road. humidifying × 1 Road. Input switch variable signal: fan operates and wind direction confirms × 2 Input point, amendment of tree species × 4 Input point. Output switch variable signal: intake × 1 Output point, exhaust × 1 Output point, fan forward and backward × 1 Output point, fan operates × 1 Output point. Heating and humidification control valves are all electric valves, regulating the opening to complete the temperature and humidity regulation in drying kiln. As the analog output is 0∼10 V voltage signal, the VVI41.15-2.5 type electric valve is selected, and the actuator is SQX62. The hardware resource allocation and functions of the system are shown in Table 1.

1.3. Sensor selection and detection circuit

Temperature and humidity are the main parameters of drying kiln. Because the environment is complex and the range of changes in temperature and humidity is large in drying kiln, the sensor should have the characteristic such as temperature drift is small, measuring range is large, damp-resisting, anti-corrosive and so on. The system selects KZC2.HD/9 temperature and humidity integrated sensor (transducers) with a strong aluminum body. It can measure the relative or relative humidity and temperature humidity of air and other non-corrosive gases. It is able to be used at pressures up to 25 bar as well as temperatures up to 160 °C. Its temperature measuring range is –40 °C∼160 °C and humidity measuring range is 0–100 % RH [8].

Transducer outputs are 0–10 V of standard voltage which can be directly connected to the EM235 analog input module.

EAL (Electronic apical locator) is used to measure the real-time wood moisture content and probes are used to measure the moisture content of drying kiln at 6 value. When the wood moisture content is between 28 % to the fiber saturation point, the relationship between the logarithm of specific resistance and the logarithm of moisture content abide by linearity relations. When it is the above fiber saturation point, the moisture content and the logarithm of resistance rate has linear relationship. According to the electrical properties of wood, use the measurement of the resistance value of wood in the wood drying process to online detect changes in the moisture content of wood. The test range of moisture content is of 8 % to 80 %, and the corresponding change range in resistance value is 1400 MΩ ∼ hundreds kΩ.

The input resistance with the high input impedance, high common mode rejection

ratio (CMRR) of the bootstrap amplifier operation expanded in inverse proportion to enlarge, then enter the logarithmic amplifier, converted to the voltage 0∼10 V input PLC. Logarithmic amplifier circuit is shown in Fig. 3.

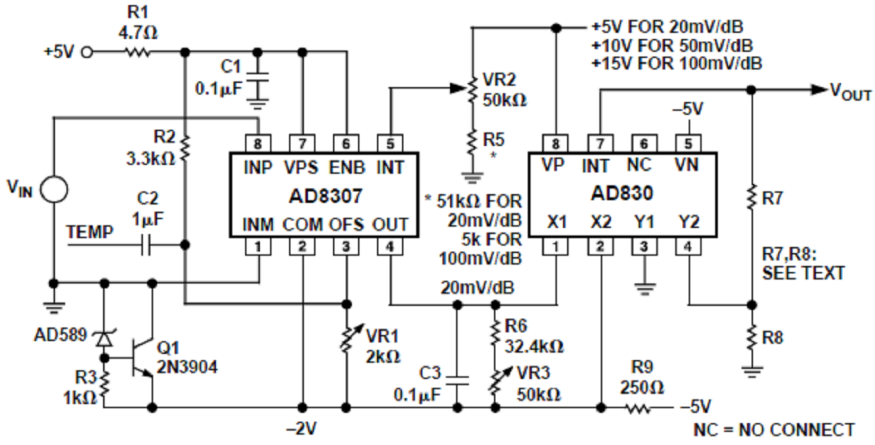


Fig. 3. Logarithmic amplifier circuit

The characteristic of this circuit is to use a logarithmic amplifier as the conversion module of a resistor voltage. As the amplifying circuit requires high input impedance, therefore use the bootstrap circuit to improve the input resistance. Circuit output connects AD8307 logarithmic operational amplifier input Vin. AD8307 can be used in industrial temperature range: from –40 °C up to 85 °C, and can be used for DC input. AD830 is used to eliminate the impact of output resulting from negative power [9–10]. Select 50 mV/dB, R7 = 3 kΩ, R8 = 2 kΩ,the output is VOUT =10-log10 (Rx).

The standard 0–10 V output voltage can be directly connected to the EM235 Analog Input Module.

2. Design of control system software

The design of system software uses modular thinking. System software is made up of function modules that have relatively independent of the function. According to different functions, the modules can be divided into the following sections:

(1) Control of the main program module. The main functions of system's main program are: initialization, data processing, open timer interrupt, transfer alarm program and Completed drying base table changes when PC gives directions.

(2) Initialization module. Connected to the system power, and then automatically start the control system software in the first scan cycle, go to "System Initialization" module. The module was completed for: Memory cleared, Moisture content - the initialization of temperature and humidity table, PID parameter initialization, Set the final moisture content.

(3) The subroutine of Moisture content data pre-processing shown in Fig. 4.

(4) PID adjustment subroutine (timer interrupt module). The main functions of the module: The establishment of temperature, humidity PID table PC chooses whether adjusts temperature, humidity PID.

(5) Output module (heating valve, steam valve output).

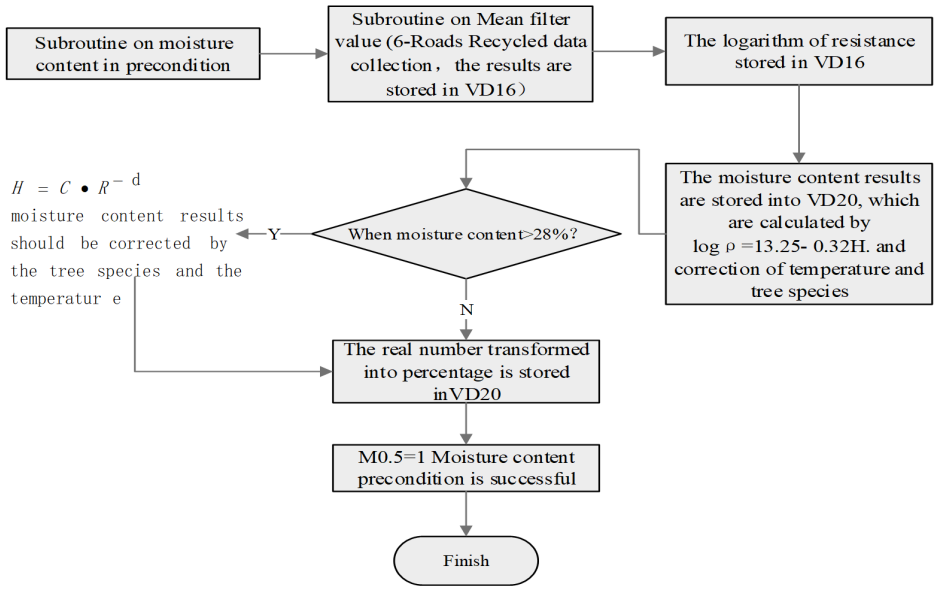


Fig. 4. Subroutine flow chart of Moisture content data pre-processing

3. Conclusion

The automatic test system of wood drying kiln is used in the dying process for the poplar trees. The drying experiment is going on for four consecutive days, wood moisture content decreased from 74.2 % to 8.3 %. In order to verify the accuracy of the system, the wood moisture content is measured by weighing method and system measurement method. Experimental results showed that the error between the moisture content came from multi-parameter test system of the wood drying kiln and the moisture content came from weighing method is less than ±1 % in the final moisture content phase. Results show that the error of the moisture content of the final stage value is less than ±1 %, comparing with that of weighing method. The control system for wood drying process will shorten the drying time, reduce drying energy consumption and improve the drying quality.

References

[1] E. B. Priyanka, C. Maheswari, B. Meenakshipriya: *Parameter monitoring and*

control during petrol transportation using PLC based PID controller. Journal of Applied Research and Technology *14* (2016), No. 2, 125–131.

[2] J. Laurila, M. Havimo, R. Lauhanen: *Compression drying of energy wood*. Fuel Processing Technology *124* (2014), 286–289.

[3] M. Aktaş, S. Şevik, M. B. Özdemir, E. Gönen: *Performance analysis and modeling of a closed-loop heat pump dryer for bay leaves using artificial neural network*. Applied Thermal Engineering *87* (2015), 714–723.

[4] J. Chang, D. Wang, J. Gao: *Intelligent auto-control system of wood drying*. Journal of Beijing Forestry University (2003), No. 2, 72–75.

[5] L. Obregón, L. Quiñones, C. Velázquez: *Model predictive control of a fluidized bed dryer with an inline NIR as moisture sensor*. Control Engineering Practice *21* (2013), No. 4, 509–517.

[6] M.Bederina, M.Gotteicha, B.Belhadj, R. M. Dheily, M. M. Khenfer, M. Queneudec: *Drying shrinkage studies of wood sand concrete – effect of different wood treatments*. Construction and Building Materials *36* (2012) 1066–1075.

[7] M. Simo-Tagne, R. Rémond, Y. Rogaume, A. Zoulalian, B. Bonoma: *Modeling of coupled heat and mass transfer during drying of tropical woods*. International Journal of Thermal Sciences *109* (2016), 299–308.

[8] S. Geving, J. Holme: *Vapour retarders in wood frame walls and their effect on the drying capability*. Frontiers of Architectural Research *2* (2013), No. 1, 42–49.

[9] L. Ge, G. S. Chen: *Control modeling of ash wood drying using process neural networks*. Optik - International Journal for Light and Electron Optics *125* (2014), No. 22, 6770–6774.

[10] Z. Yang, Z. Zhu, F. Zhao: *Simultaneous control of drying temperature and superheat for a closed-loop heat pump dryer*. Applied Thermal Engineering *93*, (2016), 571–579.

Received May 7, 2017

Acta Technica *62* No. 2C/2017, 357–364

Construction project cost forecasting method based on artificial neural network model

Fang Cuilan[1,3], Yi Deyong[2]

Abstract. Traditional method of relying on drawings, based on fixed or physical quantity of the project cost forecast method takes a long time, drag the progress of the project. This paper presents a method of building engineering cost forecasting based on artificial neural network model. This paper firstly proves that artificial neural network can be used to forecast the cost of construction project, and then uses RBF neural network to build the construction cost forecasting model, and uses 23 sets of construction engineering data to verify the analysis. The results show that RBF neural network is accurate, fast and feasible for forecasting construction cost, and its prediction error is less than 7 %.

Key words. Construction engineering, cost forecasting, artificial neural network, RBF neural network, engineering characteristics.

1. Introduction

Construction project cost forecast is a basic link in the feasibility study of construction engineering, and it is also a bidding basis of construction project. Its accuracy and fastness directly affect the investment decision and bid competition ability of the project. The traditional cost forecasting method mainly relies on the drawing to calculate the project cost based on the quota or the physical quantity. This method of forecasting takes a long time and is too dependent on the standard specification and does not take good account of the changing mechanism of the market. This can only cause the project cost forecast and the actual project cost error is more and more, and the market gradually derailment. At present, engineers are trying to find a more efficient and accurate prediction method to make up for the shortcomings of traditional methods. Some scholars have proposed fuzzy mathematical probability theory, gray theory and other methods, but because of the lack of

[1]Chongqing Vocational Institute of Engineering, Chongqing, China, 402260
[2]Chongqing Xintong Engineering Cost Consulting Co. Ltd. Chongqing, China, 400020
[3]Corresponding author

cost of the dynamic and marketability, the lack of predictive results of timeliness, not accurate enough. RBF, BP and other artificial neural network method is a new cost forecasting method, which is through the depth of learning to predict training, which can accurately and quickly predict the cost of construction projects.

Many scholars in domestic and abroad have done a lot of research on the prediction of construction cost of artificial neural network. Yang Jiyue constructs the construction project cost forecasting model based on BP neural network, and the accuracy and fastness are better [1]. Niu Dongxiao compares the current methods of using more construction cost forecasting, and proposes a variable structure neural network prediction model, and proves the feasibility of the method from theory and practice [2]. Huang Kunpeng used the value engineering, gray theory and fuzzy mathematics and other methods to control the cost of construction projects were systematically analyzed and compared the advantages and disadvantages of these three methods [3]. Chen Yusu analyzed the basic theory, network model and valuation method of BP neural network method for estimating construction cost [4]. Cao Li established a risk management system based on neural network method, and carried out the detailed design of the system, including the design of knowledge base, main interface and risk comprehensive evaluation [5]. For resolving the limitations of the two methods for gray prediction model and artificial neural network, Chen Congfa introduced the random oscillatory model and the ant colony optimization algorithm The two methods are combined and verified by examples, which provides a good idea for the study of construction cost forecasting method [6]. Osarna Moselhi introduced the application method of neural network method in construction engineering, and provided theoretical support for forecasting the construction cost by using neural network [7]. J. Z. Wu uses the neural network method to simulate the service life of the large-grid structure of the reservoir, which provides a reference for how to select the parameters when the neural network method is used in the building engineering [8]. M. R. Vigder conducted a study on the prediction and control of building costs in the Ottawa area of Ontario, Canada, and compared a variety of forecasting methods [9]. Anonymous introduced a variety of cost forecasting methods and knowledge, providing a more advanced foreign project management and cost forecasting concept [10]. F. C. Chen uses BP neural network to simulate the adaptive self-regulation engineering, and demonstrates the effectiveness of BP neural network in engineering management [11]. Haykin S system introduces the neural network method, and provides theoretical support for the application of neural network method in engineering cost [12].

2. Basic theory of artificial neural network

2.1. Definition of artificial neural network

Artificial neural network first appeared in 1890 William James's "Psychology", the book proposed the famous MP model (McCulloch-Pitts), which created a neural network research precedent. Artificial neural network is similar to human brain or human neural network, with learning, association, self-adaptation, self-organization

and anti-interference ability, it can solve the general information processing methods cannot or difficult to solve the problem, especially some image thinking, reasoning, Summarize, summarize the problem. Artificial neural network to solve the problem fast and accurate, there is a certain degree of fault tolerance, very suitable for the construction cost forecast. Artificial neural network structure model shown in Fig. 1.

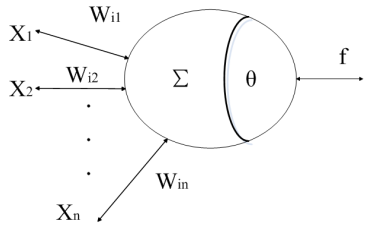


Fig. 1. Artificial neural network structure model

Neural network learning process is also known as training, refers to the environment through the neural network to stimulate the adjustment of the free parameters of the neural network, making the neural network in a new way to make a certain response to the external environment of the process. Neural network learning is divided into two kinds: there are teacher learning (learning with a teacher) and no teacher learning (Learning without a teacher). There is a mentor to learn that there is a mentor in the training process. Although the neural network is unknown to the external environment, the tutor has an understanding of the external environment. The tutor can be represented by a set of samples consisting of input and output. The supervisor signal or the desired response represents the optimal result of the implementation of the neural network, that is, the neural network input is used to adjust the neural network parameters so that the neural network output is as close as possible to the pilot signal or desired response. No tutor to learn without a tutor signal, but through continuous use with the external environment to minimize the performance of the scalar index and completed, the training process did not evaluate the entire training process, but to provide a neural network learning on the quality of the method Measure the scale by which the free parameters of the network are optimized.

Artificial neural network is an advanced means of bionic computing, which has three distinct characteristics:

Nonlinearity. Artificial neurons can be linear or non-linear. If the artificial neural network is made of non-linear artificial neurons, the artificial neural network is nonlinear and can be used to simulate many complex nonlinear problems.

Input-output mapping (Input-Output Mapping), artificial neural network has the ability to learn, one can get through the training input to the mapping between the output method. We can input a lot of training samples, the sample has an input signal and output signal, artificial neural network through this group of training samples to sum up the input to the output of the mapping, or can become a functional relationship, and after several trainings gradually reduce the error until the satisfaction so far, so that the input from the input to the mapping. This method

is used to carry out the basis of construction cost forecast.

Adaptability (Adaptivity). Artificial neural networks can adjust the connection weights to accommodate changes in the initial environment. Artificial neural networks can be easily trained to adapt to changes in the initial environment if they encounter minor changes in the initial environment after training in a particular environment. Said it has a certain degree of self-adaptability.

2.2. Feasibility of artificial neural network used in predicting construction cost

The cost of construction works has a direct relationship with the characteristic factors of construction engineering. This relationship is obviously a non-linear relationship, and its relationship with the most important features of construction is very close, or the weight Larger. Because of this non-linear, complex, multi-input problem, the cost forecast of construction is often not easy to predict accurately. The most prominent feature of artificial neural networks is the ability to solve nonlinear problems and complex problems, such as modeling, time series analysis, pattern recognition and signal processing and control, and so have excellent performance. We can take the characteristic factors of the construction engineering as the input of the artificial neural network, take the cost of the construction project as the output of the neural network, and make the multi-group construction project which has been completed, the clear characteristic factor and the final real cost as the training sample, The Mapping Relationship between Characteristic Factors and Construction Cost. Artificial neural network since the date of birth, many scholars at home and abroad has been in its in-depth research and application, the current application of artificial neural network has been very wide. With the artificial neural network algorithm matures, based on artificial neural network construction cost forecast is completely feasible.

2.3. Two commonly used artificial neural network methods

There are two main methods of artificial neural network: BP (Back Propagation) algorithm and RBF (Radial-Basis Function) algorithm, namely multi-layer sensor network method and multi-variable interpolation radial basis function method.

BP algorithm has the ability of non-linear approximation of two-level perceptron, which makes the network connection more connected, which makes the BP algorithm have extremely powerful computing ability. It is a kind of artificial neural network algorithm which is more popular and widely used.

The RBF algorithm uses the radial basis function as the "base" of the hidden unit to form the hidden layer space. The hidden layer transforms the input vector, transforms the low-dimensional input data into the high-dimensional space, so that the linear indivisibility problem in low-dimensional space can achieve linear separability in high-dimensional space.

There are two main differences between the two methods:

The number of parameters is different. BP algorithm requires global approximation of the nonlinear mapping relation of the samples, while the nonlinear function

of the local exponential decay used by the RBF algorithm is local approximation to the mapping of the nonlinear input and output. If the same precision is achieved, the RBF algorithm needs of the parameters than the BP algorithm much less.

The calculation accuracy is different. RBF network model to predict the results of the project cost is relatively stable, the basic accuracy can be controlled within $\pm 10\,\%$, while the BP network model of the accuracy of the forecast is easy to large ups and downs.

The calculation speed is different. RBF algorithm because there is only one layer of hidden layer, the calculation speed is relatively fast, easy convergence. The BP algorithm often has a slow convergence rate, easy to fall into the local minimum point of the problem.

From the above three reasons, we can see that the performance of RBF network is better than BP network as a whole, so this paper uses RBF neural network to forecast the construction cost.

3. RBF neural networks

3.1. RBF network structure

From the structure point of view, RBF network is a three-tier forward network, the first layer for the input layer, which is composed of signal source nodes. The second layer is the implicit layer (RBF algorithm has only one layer of hidden layer), the number of hidden units is determined by the problem of the fixed, hidden unit of the transformation function is a center of the radial symmetry and attenuation Non - negative nonlinear function. The third layer is the output layer, it responds to the input layer, the structure shown in Fig. 2.

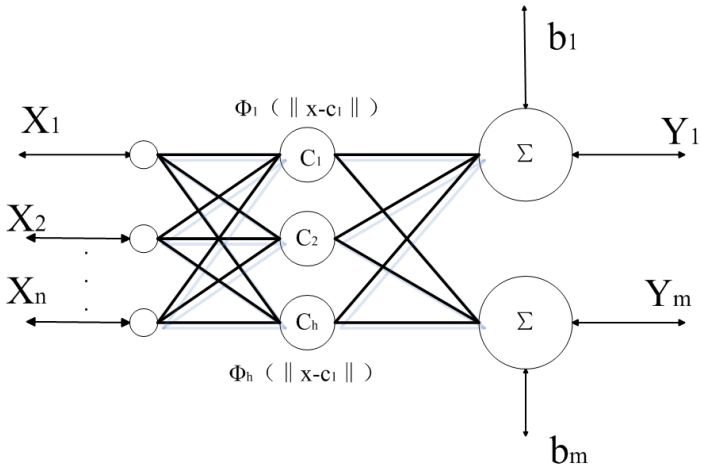


Fig. 2. RBF network structure diagram

3.2. RBF algorithm flow

The algorithm for determining RBF is actually a process of determining the number of hidden nodes in the RBF network, the data center of the radial basis function, and the output weights. After these parameters are determined, the hidden layer to the output layer is a multiple linear system of equations, and then learn the weight can be used to solve the least squares method. Commonly used algorithms include clustering method, gradient training method, orthogonal least squares learning algorithm and so on.

The most classic RBF network algorithm is based on the K-means clustering algorithm RBF network, the method is proposed by Moody and Darken. The main idea is to use the method without instructor to study the clustering analysis to determine the number of hidden nodes in the RBF network, and then find out the distance of the data center to determine the expansion constant of the activation function in the hidden layer, Finally, you can use the least squares method to directly find the weight, that is shown in Fig. 2.

4. Verification analysis of construction cost based on RBF neural network

In this paper, we use MATLAB to write the algorithm program, mainly using the software toolbox radbas (radial basis transfer function) and newrb (radial basis network function), based on the minimum gradient method to prepare the algorithm to simulate the simulation training program.

In this paper, a total of 23 sets of construction project cost examples were collected to verify the algorithmic program. For the selection of engineering features, this paper chooses the eight engineering features which are most influential to project cost, namely, foundation type, main structure type, wall type, layer number, layer height, single floor building area, inner and outer wall practice and door and window type , The above eight engineering features as RBF network training sample values for input to the single cost, the amount of concrete, the amount of steel used in these three factors as the output of the sample. After training, the error map and the forecast result output table are shown in Fig. 3 and Table 1, respectively.

In the training, the error err_goal is set to 0.01, the width of the hidden layer Gaussian function is set to 0.5, the maximum number of iterations is 100, As can be seen from Fig. 3, only after 13 iterations after less than 0.01, the error convergence rate is relatively fast.

In the training, because the data is not a lot, only 23 groups, and RBF network structure is relatively simple, so only after 13 iterations to achieve the expected range of error, and the actual cost and forecast cost error is small, in addition to individual data, The error is less than 7 %, all data are less than 10 %. Due to the limited space, this paper only lists the results of samples 5 and 15, the prediction results are shown in Table 1.

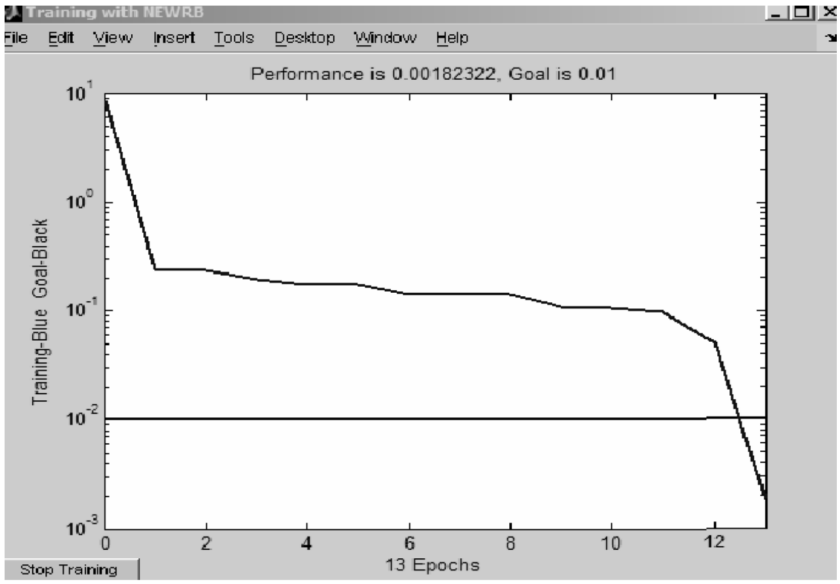


Fig. 3. Error training chart

Table 1. The distribution of the dynamic comfort

Sample No.	No.5			No.15		
Input item	Each cubic cost	Concrete	Steel bars	Each cubic cost	Concrete	Steel bars
Actual value	954.6	1992	588	712.8	1752	1207
Predictive value	977.8	2103	561	676.2	1667	1285
Tolerance	2.43 %	5.54 %	-4.59 %	-5.13 %	-4.82 %	6.46 %

5. Conclusion

In this paper, the construction cost is forecast by using RBF neural network, and the results of 23 sets of construction engineering are verified, and get the following conclusions Artificial neural network model to predict the cost of construction projects is accurate, fast and feasible.

RBF neural network is simple in structure, fast convergence, less required parameters, and less prediction error, and can be applied to construction cost forecast.

Based on the engineering cost forecast of eight engineering features of the 23 sample sizes mentioned in this paper, although the error is less than 10 %, but most still more than 5 %, it is still necessary to further increase the number of samples and refine the engineering characteristics to improve predictive accuracy.

References

[1] J. W. Zhang, D. L. Ye, G. C. Ye, J. Z. Zhou: *Engineering project risk evaluation system construction and research based on BP model.* Applied Mechanics and Materials *367–370* (2013), 2050–2053.

[2] G. H. Kim, J. E. Yoon, S. H. An, H. H. Cho, K. I. Kang: *Neural network model incorporating a genetic algorithm in estimating construction costs.* Building and Environment *39* (2004), No. 11, 1333–1340.

[3] W. Li, S. C. Li, D. Wang: *Risk evaluation of the wind power project investment based on BP neural network.* Advanced Materials Research *108–111* (2010), 256–261.

[4] G. Zhang, B. E. Patuwo, M. Y. Hu: *Forecasting with artificial neural networks: The state of the art.* International Journal of Forecasting *14* (1998), No. 1, 35–62.

[5] H. Adeli: *Neural networks in civil engineering: 1989–2000.* Computer-Aided Civil and Infrastructure Engineering *16* (2001), No. 2, 126–142.

[6] J. S. Shane, K. R. Molenaar, S. Anderson, C. Schexnayder: *Construction project cost escalation factors.* Journal of Management in Engineering *25* (2009), No. 4, 221–229.

[7] H. Golizadeh, A. N. Sadeghifam, H. Aadal, M. Z. A. Majid: *Automated tool for predicting duration of construction activities in tropical countries.* KSCE Journal of Civil Engineering *20* (2016), No. 1, 12–22.

[8] C. M. Wen, S. L. Hung, C. S. Huang, J. C. Jan: *Unsupervised fuzzy neural networks for damage detection of structures.* Structural Control and Health Monitoring *14* (2007), No. 1, 144–161.

[9] A. L. Yver, L. M. Bonnaillie, W. Yee, A. McAloon, P. M. Tomasula: *Fractionation of whey protein isolate with supercritical carbon dioxide-process modeling and cost estimation.* International Journal of Molecular Sciences *13* (2012), No. 1, 240–259.

[10] Y. Xu, F. Elgh, J. A. Erkoyuncu, O. Bankole, Y. Goh, W. M. Cheung: *Cost engineering for manufacturing: Current and future research.* International Journal of Computer Integrated Manufacturing *25*, (2012), Nos. 4–5, 300–314.

[11] S. G. Tzafestas, G. G. Rigatos: *Neural and neurofuzzy FELA adaptive robot control using feedforward and counterpropagation networks.* Journal of Intelligent and Robotic Systems *23* (1998), Nos. 2–4, 291–330.

[12] G. H. Kim, S. H. An, K. I. Kang: *Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning.* Building and Environment *39* (2004), No. 10, 1235–1242.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 365–374

Risk assessment of existing concrete frame structure in geological hazards area with high incidence

Jiawei Sun[1], Hainan Liu[1]

Abstract. To provide a quantitative measure of hazard performance evaluation, we carried out a probabilistic seismic risk analysis and proposed an economic performance evaluation index. The economic performance evaluation index is based on the performance of the hazard engineering. We also proposed the vulnerability index analysis based on risk and robustness analysis. We compared the seismic demand and seismic damage risk of the different index structures of the corresponding prototype structures. The results indicate that the reinforced concrete frame structure can meet the demand of "the great hazard will not make too great changes". The model we proposed has higher accuracy and computational efficiency

Key words. Risk assessment, concrete frame structure, geological hazards.

1. Introduction

At home and abroad, a lot of researches have been done on the study of hazard loss estimation, and the research object is mainly divided into two categories: the study of regional loss estimation and the estimation of single building loss. The so-called regional loss estimation is to take a large number of building loss estimation in a region as the research objects, to make the economic loss estimation. While single building loss estimation study is the accurate estimation of loss for a specific site on the concrete building. As to the research method, the hazard loss estimation can be divided into deterministic method and probabilistic method [1]. And both of them are based on the corresponding risk analysis. The difference is that the deterministic risk analysis is only to consider the given magnitude of given level in a specific site, while the probabilistic seismic hazard analysis is to consider the exceedance probability of all possible ground motion strengths.

Probabilistic seismic risk assessment is the basis of hazard risk decision and safety management. On the basis of the previous studies, this paper will first of all explore

[1]Shaanxi Institute of Geo-Environment Monitoring, Xi'an, 710000, China

the probabilistic seismic risk assessment index. The paper introduces the seismic risk evaluation indexes such as risk probability, risk loss, internal rate of return, net present value, dynamic investment recovery period, vulnerability index and risk based robustness index and so on hazard risk assessment indexes. On this basis, the risk assessment of concrete frame structure in geological hazards area with high incidence.

2. Probabilistic risk analysis of concrete frame structure

2.1. Probabilistic seismic risk analysis function

The probability λ_{LS} of exceeding a certain limit state failure of the structure each year can be calculated by the following formula:

$$\lambda_{\mathrm{LS}} = \int_X F_{\mathrm{R}}(x) \,|\, \mathrm{d}H(x)| \,. \tag{1}$$

In the above formula, $H(x)$ refers to the hazards risk function, indicating the probability that the hazards with a certain strength happened in the design site each year [2]. The hazards risk function $H(x)$ can be expressed by the maximum distribution function of extreme of II type:

$$H\,(x) = P\,[\mathrm{IM} \geq x] = 1 - \exp\left[-\left(\frac{x}{u}\right)^{-k}\right] . \tag{2}$$

In (2), u indicates the hazards scale parameter and k represents the shape parameter.

In the probability frame, $H(x)$ usually uses power exponent for the approximation, suggested as Log-linear relationship

$$\lambda_{\mathrm{IM}}(x) \approx \left(\frac{x}{u}\right)^{-k} \approx k_0 x^{-k} \,. \tag{3}$$

In (3), k_0 and k are the shape parameters, which can be obtained by DBE (Design Based Earthquake) and considering the earthquake motion strength fitting corresponding to the hazards MCE [3].

In 1994, Cornell substituted seismic vulnerability function

$$F_{\mathrm{R}}(x) = P\,[D \geq C\,|\mathrm{IM} = x\,]$$

(IM refers to the intensity measure, $D \geq C$ indicates the structure reaches or exceeds a certain extreme state where D is the demand and C suggests the seismic capacity and $F_{\mathrm{R}}(x)$ is called seismic vulnerability function) into (1), and then he derived and obtained the analytical expression of probabilistic seismic risk

$$\lambda_{LS} = H(m_{\mathrm{R}}) \exp\left[\frac{1}{2} k^2 \beta_{\mathrm{R}}^2\right] . \tag{4}$$

By using the seismic vulnerability function based on displacement, we can get the analytical expression of probabilistic seismic risk only considering the intrinsic uncertainty [4]

$$\lambda_{\mathrm{LS}} = H\left\{\left(\frac{m_C}{\exp\left(\beta_0\right)}\right) 1/\beta_1\right\} \exp\left[\frac{1}{2}k^2\left(\frac{\beta\left|{}^2_{D|\mathrm{IM}} + \beta_C^2\right.}{\beta_1^2}\right)\right]. \tag{5}$$

In (5), if the capability uncertainty in the analytical expression of probabilistic seismic risk is not considered, then we can get:

$$\lambda_D = H\left\{\left(\frac{d}{\exp\left(\beta_0\right)}\right)^{1/\beta_1}\right\} \exp\left[\frac{1}{2}k^2\left(\frac{\beta_{D|\mathrm{IM}}^2}{\beta_1^2}\right)\right]. \tag{6}$$

In the frame of PBEE (Performance Based Earthquake Engineering), (5) and (6) can be understood as the different stages of uncertainty transmission [5]

$$\lambda_{\mathrm{LS}} = \int_{\mathrm{edp}}\int_{\mathrm{im}} G\left(\mathrm{dm}\,|\mathrm{edp}\right) |\,\mathrm{d}G\left(\mathrm{edp}\,|\mathrm{im}\right)\| \,\mathrm{d}\lambda\ (\mathrm{im})|\,, \tag{7}$$

$$\lambda_D = \int_{\mathrm{im}} G\left(\mathrm{edp}\,|\mathrm{im}\right) |\,\mathrm{d}\lambda\left(\mathrm{im}\right)|\,. \tag{8}$$

From (7), it can be seen that λ_{LS} considers the uncertainty of structure capability and it can be seen as the results of hazards risk transmission to structure damage layer, which is called "Probabilistic seismic damage risk". From (8), it is known that λ_D does not consider the uncertainty of structure capability, which is called "Probabilistic seismic demand risk" [6]. And the "Probabilistic seismic demand risk function" shown in (6) is also called "structural seismic hazard function". The seismic risk analysis of structures can be regarded as the extension of the seismic hazard from the site to the structure. If the seismic demand d is taken in a reasonable range, the seismic demand risk is expressed in the form of seismic demand risk curve. If we only pay attention to the seismic demand d^{LS} at a certain level, the seismic demand risk represents the average annual probability of the event $\left\{d > d^{\mathrm{LS}}\right\}$. However, the seismic damage risk characterizes the probability of earthquake damage that occurs in different states each year.

2.2. Probabilistic seismic demand risk analysis

Based on the index prototype structure DBE and MCE Sa (T1,5%) value designed, we can get the structural hazard function parameters, as shown in Table 1. According to the seismic code, the hazards intensity increases for 1 degree, and the hazards action increases twice. As a result, we obtain SaMCE/SaDBE = 2 [7].

The seismic demand risk curve of the prototype structure is obtained based on the above formulas. From the results, it is known that with the increase of the level of structural fortification, the risk of seismic demand has not weakened, but

increased. According to the research in the previous parts, it shows that with the increase of the fortification level, seismic vulnerability of the structure under the same seismic fortification level gradually weakened [8]. That is to say, the improvement of structure fortification enhances the ability of earthquake resistant structure. However, the increase in fortification levels also increases the risk of structural seismic damage. Of course, if there is an earthquake risk in the same site, the structure with a stronger ability to resist ground motion should have a smaller seismic demand risk.

Table 1. The distribution of the dynamic comfort

Structures	Sa_{DBE}/g	Sa_{MCE}/g	k_0
F3-1	0.07	0.13	3.29×10^{-6}
F3-2	0.17	0.33	2.90×10^{-5}
F3-3	0.17	0.33	2.90×10^{-5}
F3-4	0.32	0.64	1.41×10^{-4}
F3-5	0.43	0.86	2.81×10^{-4}
F3-6	0.70	1.40	9.00×10^{-4}
F5-1	0.05	0.10	1.62×10^{-6}
F5-2	0.11	0.21	1.03×10^{-5}
F5-3	0.15	0.31	2.45×10^{-5}
F5-4	0.22	0.44	5.74×10^{-5}
F5-5	0.29	0.59	1.14×10^{-4}
F5-6	0.52	1.03	4.38×10^{-4}
F8-1	0.04	0.08	8.97×10^{-7}
F8-2	0.08	0.16	5.04×10^{-6}
F8-3	0.11	0.23	1.20×10^{-5}
F8-4	0.15	0.30	2.29×10^{-5}
F8-5	0.24	0.48	6.99×10^{-5}
F8-6	0.32	0.64	1.38×10^{-4}
F10-1	0.03	0.07	6.07×10^{-7}
F10-2	0.07	0.13	3.31×10^{-6}
F10-3	0.10	0.19	7.89×10^{-6}
F10-4	0.12	0.25	1.47×10^{-5}
F10-5	0.20	0.40	4.63×10^{-5}

3. Assessment of concrete frame structure risk

3.1. Assessment of concrete frame structure risk

Taking into account the uncertainty of structural capacity, the results of seismic damage risk analysis considering the intrinsic uncertainty of the prototype structure

are obtained according to the formula, as shown in the following table.

Table 2. Probabilistic seismic damage risk values for index architype buildings considering intrinsic uncertainty

Structures	SD(10^{-2})	MD(10^{-3})	ED(10^{-4})
F3-1	1.46	0.27	0.56
F3-2	7.05	0.46	5.55
F3-3	7.05	1.29	5.55
F3-4	1.64	1.37	10.51
F3-5	1.15	2.30	18.51
F3-6	15.50	0.18	49.32
F5-1	0.32	0.53	1.23
F5-2	1.41	1.87	0.62
F5-3	1.16	2.26	1.08
F5-4	1.55	1.31	1.43
F5-5	1.17	0.75	1.69
F5-6	0.49	1.87	3.83
F8-1	0.60	0.16	0.16
F8-2	3.01	0.84	2.72
F8-3	3.02	1.72	1.78
F8-4	2.32	1.53	5.11
F8-5	1.77	2.68	3.08
F8-6	0.93	0.76	7.14
F10-1	0.43	0.27	0.64
F10-2	1.90	1.14	2.86
F10-3	4.24	2.71	6.03
F10-4	6.28	3.58	1.01
F10-5	2.93	4.21	1.45

Taking the structure F5-2 as an example, the structural seismic demand risk curve is compared with the structural damage risk, as shown in the Fig. 1. As can be seen from the graph, the value of the seismic damage falls above the seismic demand risk curve [9]. In order to make a further comparison, this paper compares the seismic demand and seismic damage risk of the different index structures of the corresponding prototype structures. And the comparison results suggest that the risk of hazard demand is less than the risk of earthquake damage due to the uncertainty of capacity.

3.2. Probabilistic seismic risk analysis in the service life of structures

It should be pointed out that the probabilistic seismic risk involved in this paper is expressed in the form of annual average exceedance probability [10]. In the existing anti-seismic standards, it is not clear that the limit of annual average exceedance probability has been used as the criterion for evaluating the seismic risk of structures.

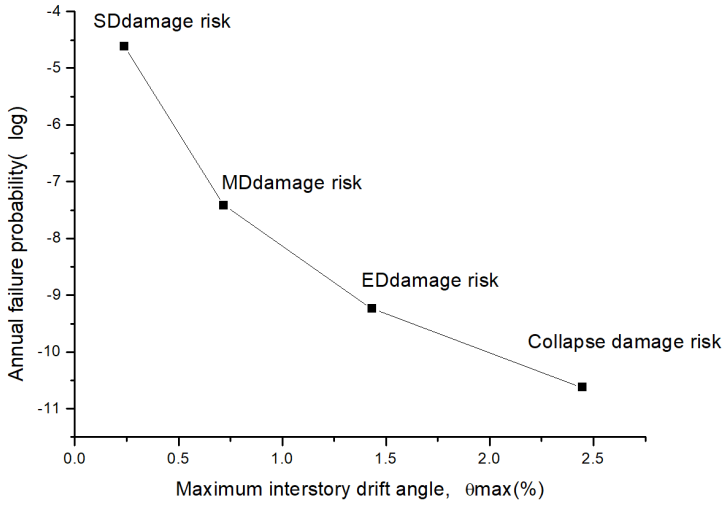


Fig. 1. Comparisons of the seismic demand risk curve and the seismic damage risk values for building F5-2

There are three levels of fortification requirements for structural design in the anti-seismic standards of China: "the small hazard is not serious, the moderate hazard can be repaired, and the great hazard will not make too great changes", in which the occurrence probability of small, moderate, and great hazards within 50 years is 63.2 %, 10 % and 2 % in turn [11]. Because our specification does not consider the influence of structural uncertainty, from the viewpoint of probability, the three level design can be understood as: "the probability for structure to have a slight damage within 50 years does not exceed 63.2 %, the occurrence probability of moderate damage does not exceed 10 %, and the probability of collapse is not more than 2 %" [12]. Based on the above understandings, this paper extends the analysis results of seismic damage risk to 50 years to assess the probability of hazard risk. The failure probability of structure in different damage levels within 50 years can be calculated according to the Poisson hypothesis, shown as follows

$$P_{\mathrm{DM}} = 1 - (1 - \lambda_{\mathrm{DM}})^{50} \ . \tag{9}$$

In the period of usage within 50 years, the probability of the damage of the prototype structure is larger, and the probability of failure of most structures is greater than 63.2 %, which is inconsistent with the requirements of "the small hazard is not serious". With the increase of the structure fortification intensity, the risk of structural damage is not reduced, but increased [13]. This shows that improving the structure fortification and increasing its resistance level is to reach the "balance" at the expense of hazard risk". The probability of the intermediate failure of many index prototype structures is greater than 10 %. While the damage only with F3-6 structure is more than 2 % [14]. Therefore, reinforced concrete frame structure designed according to the standard meets the requirements of "the great hazard

does not make too great changes", which is certain.

3.3. Results

This chapter is based on the analysis of seismic vulnerability function, using the probabilistic risk relationship in the classical form of power function. We derive the analytic function of probabilistic seismic risk, which includes: only considering the intrinsic uncertainty probabilistic seismic demand risk function and the probability of hazard damage risk function. By comparing the seismic demand and seismic damage risk of different limit states, it can be seen that the risk of seismic demand is less than the risk of earthquake damage since the capacity uncertainty is not considered [15].

The goal of probabilistic risk analysis is to evaluate the seismic safety of structures. To achieve this goal, this paper expanded the probability characterized as annual exceeding probability seismic risk as the failure probability with the use period of structures, and assessed the seismic safety of the reinforced concrete frame structure designed according to the evaluation standard in our country. The results show that the structure to improve the fortification and increase the resistance level is a "balance" reached at the expense of seismic risk. At the end of this chapter, according to the results of seismic vulnerability of population structure obtained, we further analyze the seismic risk of concrete frame structure. And it is found that: the reinforced concrete frame structure designed according to the code of our country cannot meet the requirement "the small hazard is not serious, and the moderate hazard can be repaired", but it can meet the demand of "the great hazard will not make too great changes".

4. Conclusion

In this paper, a new generation of performance based earthquake engineering risk framework is taken as the research background, the open hazard engineering simulation software as the research platform, and the reinforced concrete frame structure with large volume as the research object. The paper intends to assess the safety and potential risk of concrete structures designed aaccording to the current specification, with seismic analytic function vulnerability and risk function and efficient simulation technique as the research tools. The main conclusions are as follows:

In the aspect of uncertainty analysis methods, the proposed method has higher accuracy and computational efficiency. The sensitivity parameters presented in this paper can be used to describe the effect of the change of random variables in the standard normal space on the function response.

In the aspect of probabilistic seismic demand analysis, the seismic intensity evaluation system proposed in this paper can be used to evaluate the statistical characteristics of ground motion from a probabilistic point of view.

In the aspect of seismic risk assessment, the risk of hazard demand is less than the risk of hazard damage because of the uncertainty of capacity. The structure increases the resistance ability by improving the fortification is under the premise

of losing hazard risk. From the perspective of hazard risk, the reinforced concrete frame structure designed according to the standards in China cannot strictly meet the design requirement of "the small hazard is not serious, and the moderate hazard can be repaired", but can fully meet the demand of "the great hazard will not make too great changes".

References

[1] R. Zeibak-Shini, R. Sacks, L. Ma, S. Filin: *Towards generation of as-damaged BIM models using laser-scanning and as-built BIM: First estimate of as-damaged locations of reinforced concrete frame members in masonry infill structures.* Advanced Engineering Informatics *30* (2016), No. 3, 312–326.

[2] S. W. Park, S. W. Choi, Y. Kim, B. K. Oh, H. S. Park: *Optimal retrofit design of reinforced concrete frame with infill wall using fiber reinforced plastic materials.* WASET, International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering *10* (2016), No. 5, 572–576.

[3] N. Ganesan, P. V. Indira, S. R. Prasad: *Structural behaviour of steel fibre reinforced concrete wall panels in two-way in-plane action.* Indian Concrete Journal *84* (2010), No. 10, 21–28.

[4] A. Kadid, A. Boumrkik: *Pushover analysis of reinforced concrete frame structures.* Asian Journal of Civil Engineering (Building and Housing) (2008), No. 9, 75–83.

[5] G. A. Anwar, J. Anwar, P. D. Versaillot: *Effect of CFRP retrofitting on seismic vulnerability of reinforced concrete frame structures in Pakistan.* American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS) *26* (2016), No. 1, 145–158.

[6] P. Yan, B. Sun: *Earthquake damage of self-built reinforced concrete frame structures in M_S 8.1 Nepal earthquake.* IOP Conference Series: Earth and Environmental Science, International Conference on Energy Materials and Environment Engineering, 10–12 March 2017, Bangkok, Thailand, IOP Publishing *61* (2017), paper 012093.

[7] D. J. Kakaletsis, C. G. Karayannis: *Experimental investigation of infilled reinforced concrete frames with openings.* ACI Structural Journal *106* (2009), No. 2, 132–141.

[8] M. N. Bugeja, J. M. Bracci, W. P. Moore Jr.: *Seismic behavior of composite RCS frame systems.* Journal of Structural Engineering *126* (2000), No. 4, 429–436.

[9] Z. Lan, Y. Tian, L. Fang, S. Liang, X. Wang: *An experimental study on seismic responses of multifunctional vibration-absorption reinforced concrete megaframe structures.* Earthquake Engineering and Structural Dynamics *33* (2004), No. 1, 1–14.

[10] L. Zhang, H. Zhao, T. Wang, Q. Chen: *Parametric analysis on collapse-resistance performance of reinforced-concrete frame with specially shaped columns under loss of a corner column.* Open Construction and Building Technology Journal *10*, (2016), No. 1, 466–480.

[11] L. Wang, X. Wei, N. N. Feng, Q. Gao: *Research on seismic performance of reinforced concrete frame with unequal span under low cyclic reversed loading.* Open Civil Engineering Journal *10* (2016), No. 1, 373–383.

[12] V. Kilar, S. Petrovčič, D. Koren, S. Šilih: *Cost viability of a base isolation system for the seismic protection of a steel high-rack structure.* International Journal of Steel Structures *13* (2013), No. 2, 253–263.

[13] V. T. Dzyubenko, L. V. Zvorygin: *Characteristics of rock pressure manifestations on reinforced-concrete tubular shields.* Soviet Mining *5* (1969), No. 5, 583–586.

[14] J. W. Hwang, B. K. Oh, Y. Kim, H. S. Park: *CO2 emission and cost optimization of reinforced concrete frame designed by performance based design approach.* WASET, International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering *10* (2016), No. 3, 336–341.

[15] Z. Lu, X. Chen, X. Lu, Z. Yang: *Shaking table test and numerical simulation of an RC frame-core tube structure for earthquake-induced collapse*. Earthquake Engineering and Structural Dynamics *45* (2016), No. 9, 1537–1556.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 375–384

Virtual storage technology and its application in digital library

Dai Chuankun[1,3], Zhang Nan[2]

Abstract. The digital library technology based on virtual storage technology has brought great impetus to higher education and the improvement of students' comprehensive quality. In order to better promote the perfection of the virtual storage technology of digital libraries in China, in this paper, on the basis of understanding the relevant theories, the configuration of the server of the virtual storage technology of the digital library was further analyzed. Then the quantitative analysis was conducted to determine the advantages of digital library development. Finally, the performance of the virtual storage system of the digital library was tested. Test results show that the system has stronger compression capacity, and compared with the traditional library, it has the advantages of large amount of resources and high sharing ability, and thus can promote the further improvement of the mode of higher education in China. The study provides theoretical support for the establishment of digital libraries in universities and colleges, and provides reference for the development of higher education in China.

Key words. Virtual storage technology, digitization, library..

1. Introduction

With the development of the times, the world economy level is improving rapidly, and people's demand for knowledge is gradually increasing too, especially in the current era of rapid development of various industries, the role of knowledge is extremely important. As for the important link of knowledge popularization, the establishment of the library has brought positive effects to the popularization of knowledge. As knowledge continues to improve, more libraries are beginning to build up in the world, and the collection of books in each library is also increasing. Many colleges and universities have begun to establish different ways of libraries, and which has brought a certain degree of positive impact to the students' learning. The mass establishment of libraries is essential to the continuous improvement of the overall

[1]Dongfang College, Shandong University of Finance and Economics, Taian, Shandong, 271000, China
[2]Shandong Sport University, Jinan, Shandong, 250102, China
[3]Corresponding author

knowledge level of the world, and each industry needs more comprehensive talents with the development of the times. In this trend, many colleges and universities begin to pay attention to the promotion of students' comprehensive abilities. Relying solely on classroom education may not improve students' overall level of achievement better. In the spare life, students should have more learning resources, which can further provide a supporting role for the promotion of self-level. And nowadays, the demand for the size and the amount of paper resources has also gradually increased in traditional libraries. Under this trend, the concept of digital library has been put forward and further applied to the construction of actual data resource information, so as to provide certain scientific supports and references for the reserve of world's knowledge.

2. State of the art

With the development of the times, the education industry has become an important guarantee for economic development. It can train the related personnel and further provide a certain amount of information resources for the promotion of the comprehensive-level talents [1]. In many schools, the establishment of libraries has gradually become one of the evaluation indicators of the comprehensive level of colleges and universities. As an important center for documentation and information in the construction of universities, libraries provide certain services for the promotion of students' academic levels. However, with the development of education industries, the increasing number of colleges and universities and the increasing enrollments, the demand for library construction and collection is also increasing [2]. This indirectly leads to the tension of some land resources. In order to better meet the needs of the volume of books, many colleges and universities have begun to increase the purchase rate of books, which has brought a great impact on the demand for paper resources. Under this trend, the concept of digital library has been mentioned. Because of the better use of computer technology in the current era, the establishment of libraries relies more on certain electronic information platforms to complete the transmission and supply of relevant data and information. At the same time, as a new additional form of traditional library, digital library has more data resources and is more shared, and it will not cause environmental damage and waste of resources [3]. As a resourceful expression form under the new era, the library in this model is more individualized and more extensive, and it can provide more resources for the demanders more efficiently and quickly. Therefore, it has gradually been paid more attention by scholars and researchers, and its related theories have constantly improved and summarized.

3. Methodology

With the development of the times, the world has gradually taken the economy as one of the important subjects of the development of the times. The rapid development of the economic level has brought certain impetuses and positive influences

on the development of various industries in the world. Especially the comprehensive level of the industries is rising, at the same time, the requirements of industries are also rising, and the needs for continuous screening of related human resources provide certain internal impetuses to the development of the industries [4]. The improvement of the overall and comprehensive level of the world industry also requires more high-quality talents to enter the process of enterprise development, which makes the training of talents gradually become an important direction of the development of the times. Since the reform and opening policy of China, the economic level has been greatly improved and developed. Meanwhile, the development of all sectors of our country has further promoted the development of education [5]. China's higher education has gradually reached the world advanced level, and the investment in the education industry has gradually increased, so the development of education has become one of the most important subjects in the development of the times. The development of education is of great significance to the promotion of China's comprehensive strength. More and more talents have been trained for all sectors of our country, and the process of personnel training in China has become more perfect and scientific with the continuous improvement of educational resources, and the quality of the employed population has also been continuously improved [6]. However, it is not enough to rely on the knowledge transfer process of the classroom teachers to realize the comprehensive development of the students, and students in the spare time also need to further understand other knowledge through relevant resources. And in the process of learning, if there are relatively difficult problems, students should not only communicate with the teacher, but also need to do their own work to refer to the relevant point of view. Only in this way, the theoretical support and scientific basis can be further provided for the improvement of students' comprehensive level and ability. In this trend, in the establishment of colleges and universities, China has also invested a large degree of funding for the establishment of various university libraries (see Fig. 1). The increasing number of libraries and the increasing volume of books make it possible for students to have more resource information, so that they can learn more knowledge, better enrich themselves and improve their comprehensive qualities [7]. However, with the emergence of more colleges and universities, the increasing number of students and the increasing desire for knowledge, all these problems directly or indirectly bring pressures to the traditional libraries. At the early days of the construction of a lot of libraries, the construction area may not meet the current number of people. According to statistics, with the increasing number of readers in our country, many libraries in many areas or colleges have been unable to meet the actual number of books, which has caused that some people have study problems. What's more, the books in the libraries are mostly paper books, and people may cause damage to books after a long time of reading, which causes certain limitations to the dissemination of relevant knowledge; and the use of a large number of books has a certain impact on paper resources, because more and more trees are cut down and used to make paper, thus affecting the sustainable development of the ecological environment [8]. Therefore, in conclusion, our current traditional library model may no longer be able to better meet the desire for better solutions, and the application of a large number of books may also lead to

an increase in the use of paper, the size of the library and so on. Under this trend, our country urgently needs a new library construction mode and applies it in the current education industry, so the introduction and development of related concepts of digital library provide a certain way for solving this uncoordinated phenomenon in our country [9]. Therefore, in this paper, the construction of digital libraries and the application of related virtual storage technologies were paid much attention to, so as to provide some scientific supports for the improvement of related theories and techniques in China.

Fig. 1. Development of university libraries

First of all, the circulation of library books from 2010 to 2014 in some colleges and universities in Nanjing (A, B, C, D and E) was analyzed in this study. The analysis results are shown in Fig. 2.

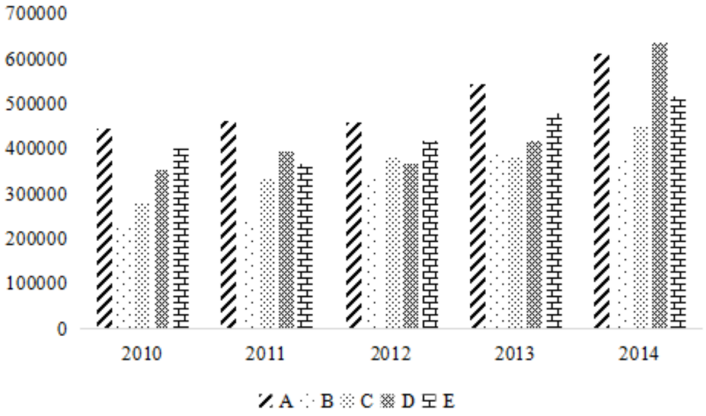


Fig. 2. Overview of data circulations in five colleges and universities in Nanjing

The data circulation of the five university libraries in Nanjing was summarized in this paper. It is further found that with the development of the time, books in

libraries show a trend of increasing year by year, and it shows that China's traditional libraries are facing greater pressures. Therefore, it is necessary to set up a digital library to alleviate the related pressures.

Then, through the analysis of relevant data, the virtualization technologies of academic libraries in our country and the related theories and concepts were determined for analysis, and the related configurations of the hardware facilities used in the digital library were summarized.

On the basis of a clearer theory, the pressure capacity of the virtual storage technology of the digital library of the university was further analyzed in the paper through the Benchmark Factory Software, so as to provide the data reference for the later development of the virtual storage technology [10].

Finally, the advantages of virtual storage technology for digital libraries in China were further analyzed in this study by means of questionnaires, so as to further provide theoretical basis for the further development of digital libraries in universities. Information about the investigator is shown in Table 1.

Table 1. Overview of personnel information used in this questionnaire

Investigator	Number of people	Average age	Proportion of men and women
College students	20	24.71	3:2
University teachers	20	35.11	1:1
Librarians	10	37.24	1:1

4. Result analysis and discussion

With the development of the times, China's education industry has been greatly improved. In this trend, China's traditional libraries are also facing great pressures, and China needs large land areas to achieve the construction of the library. And with the increasing demand, our lands and paper resources have been under certain pressures. Since the introduction of the theory and technology of digital libraries, our country has begun to use this technology extensively, and it has further promoted the pressure relief of traditional libraries. As one of the main forms of combination of traditional libraries and information technologies, digital library is the current trend of development, which combines the advantages of information technology, resource, information and sharing with the traditional library, so it brings a great deal of practical significance for the mass dissemination of book resources. As for the concept of digital library, it is not only to display some traditional paper data resources in digital forms to those who need it, but also to provide a certain type of services for all demanders as a carrier [11]. For the concept of digital libraries (Fig. 3), many researchers believe that the new library model should serve as a service system which integrates all of the more dispersed data resources.

This system can not only display all the knowledge and theory to readers and provide more information for readers as the traditional library model, but also pro-

Fig. 3. Development of digital libraries

vide the relevant theoretical research process for readers, so as that the readers can further understand the relevant theory of the origin and development of the whole link from the source. Therefore, in general, the digital library not only has the storage link of the traditional library for the relevant data information, but also includes the sharing and processing of all the data information [12]. As an important way to share and transfer data resources in the current era, digital library has mainly undergone the emergence of related theories and techniques, the continuous improvement and development of theories and technologies, and the application of relevant theories. Among them, the emergence of the related theories and technologies is mainly in the premises that the demand for data resources is increasing in the current era, and the establishment of a large number of libraries has created some damage to the space and other resources of people's living environment, people begin to put forward the combination of information technology and the traditional library construction mode which develops rapidly in the current era, so as to achieve more sharing of resources. The development of data information technology has positive influences on the theory and technology of the digital library. More data resources are displayed on a certain platforms because of the development of data and information technology, and the use of more innovative information technologies can effectively achieve the sharing and transmission of data information, thereby reducing the waste of related resources. Therefore, many scholars have combined this kind of digital technology with the construction idea of data library, so as to lay the foundation for the development of other researches [13]. On the basis of relevant theories, many scholars have begun to consider how to store large amounts of data resources. Only by further improving the storage efficiency can some technical supports be provided for the further improvement and development of digital libraries. Nowadays, the rapid development of computer technology has brought some impetuses to the development of information storage technologies in digital libraries, especially with the development of the virtual technology, more and more data storage theories have been combined with the virtual storage technology, so that the new virtual storage technology has been developed, so as to provide influences for related resources of

the subsequent library [14]. Many researches have already considered that virtual storage technology is one of the most important parts of digital library developments. Only by doing better analysis and researches of this link can some theoretical and technical supports be provided for the improvement of digital library [15]. In this study, based on the further reading and summarizing of the relevant data, the host configuration of the main digital virtual storage technology used in a university in our country was summarized. The results are shown in Table 2.

Table 2. Virtual storage configurations of digital libraries in a university in China

Device name	Brand	Model	Parameter	Quantity (unit)
The server	DELL	PowerEdge R710	2U chassis CPU: one Xeon 5606 memory 16G, 3X146G SAS hard disk-RAID5, double redundancy power supply, HBA slot. CD driver.	3
Optical fiber storage (HBA)	DELL			3
Storage /IP/SAN	DELL	MD3200i	24T, 24*1TSATA	2
UPS	Emerson	NXR-40KVA	2 hours	1
Cabinet	Totems	AD.6937	37U	1
KVM	Tanto	XL-1708	8 rims, with display and keyboard operation	1

Then, the Benchmark Factory software was used to test the running pressure of the virtual storage system of a digital library in a university. The results are shown in Table 3. The results show that the system has experienced all the test items and there are no crashes and snaps. Therefore, it can be considered that the virtual storage system has a strong ability to resist compression.

Table 3. Test result analysis of related indicators of Benchmark Factory software

User load	TPS	Average response time (s)	Total executions	Total rows	Total errors
10	0.564	0.569	98	541	0
50	2.669	0.098	483	2638	0
100	5.295	0.159	946	5183	0

Through the questionnaire, the advantages of the virtual storage technology for digital libraries in our country were analyzed in this paper. The results are shown

in Fig. 4. The results show that compared with the traditional libraries, the digital library based on virtual storage technology has more resources and more sharing. Therefore, the digital library based on the virtual storage technology is more suitable for the development of higher education in our country. And through the questionnaire survey, it can be found that more people believe that our country should apply the related theories and techniques of digital library to the actual construction of university libraries, and the results are shown in Fig. 5.

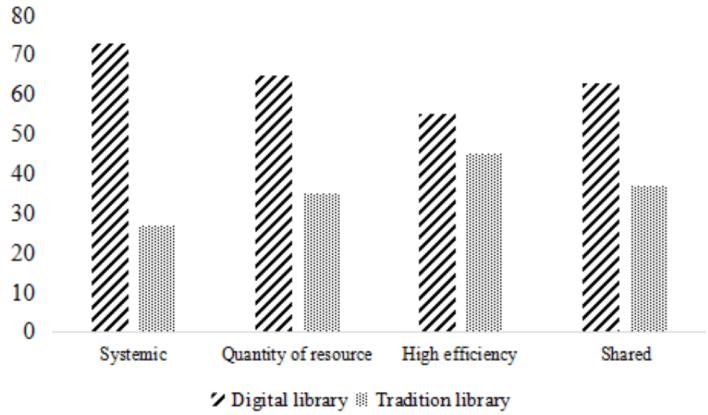


Fig. 4. Comparative analysis of digital libraries and traditional libraries

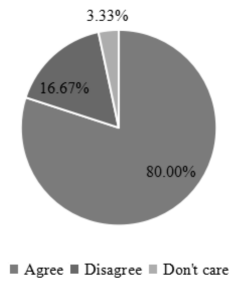


Fig. 5. Researches on the acceptance of digital libraries based on the virtual storage technology in Chinese universities

5. Conclusion

With the development of the times, the economic level of the present era has been greatly improved. In this trend, all industries in the world have been developed to a great extent. And the development of various industries has also begun to have more demands for high-quality talents, which has made the education industry in colleges and universities obtain great development. For the development of higher education,

it is very important for libraries to study and perfect the relevant knowledge for students. Relevant theories were clarified in this study by reading relevant data, so that the relevant theories were cleared; and then the configuration of the server of the virtual storage technology was analyzed. Through the questionnaire, the digital library has higher advantages, which is very necessary for the development of education in China. The research results show that this kind of technology is very necessary for the development of higher education in our country. The purpose of the research is to provide theoretical basis and reference for the development of the digital library technology in China. This research still has some shortcomings for the discussion and comparison of virtual storage devices, but it can still be used as a reference for related researches, and can provide some data supports for the development of the digital library.

References

[1] D. X. Lu, Y. X. Niu, L. Q. Zou: *Analysis of virtual storage technology and its application in the library.* Advanced Materials Research *756–759* (2013), 1289–1294.

[2] X. W. Kang, Y. J. Yang, X. Du: *Application of virtual storage technology in disaster tolerant system.* Computer Engineering (2009), No. 21, 36–38,41.

[3] X. Y. Zhou, Z. H. Hu, J. K. Shun: *Security architecture of digital library research in general knowledge environment.* Applied Mechanics and Materials *198–199* (2012), 601–605.

[4] T. Kirchhoff, W. Schweibenz, J. Sieglerschmidt: *Archives, libraries, museums and the spell of ubiquitous knowledge.* Archival Science *8* (2008), No. 4, 251–266.

[5] W. Yang: *Embedded discipline-oriented service: The new idea on transformation and development of research University Libraries.* Information and Documentation Services *33* (2012), No. 2, 88–92.

[6] F. W. J. Voges: *Sustainable development and the socially embedded firm. An inquiry into the nature, causes and transformation of structural unsustainability in contemporary liberal capitalisms.* Dissertation, University of Stellenbosch, South Africa (1999).

[7] J. Xie, L. Sun: *Macaudata.com: A shared virtual library and digitization project.* Journal of Interlibrary Loan, Document Delivery and Electronic Reserve *23* (2013), Nos. 4–5, 179–190.

[8] E. Henthorne: *Digitization and the creation of virtual libraries: The princeton university image card catalog–reaping the benefits of imaging.* Information Technology and Libraries *14* (1995), No. 1, 38–40.

[9] M. Sullivan, M. N. Ochoa: *Digital library of the caribbean: A user-centric model for technology development in collaborative digitization projects.* OCLC Systems & Services, Emerald Group Publishing Limited *25* (2009), No. 4, 249–262.

[10] F. Bettio, A. J. Villanueva, E. Merella, F. Marton, E. Gobbetti, R. Pintus: *Mont'e scan: Effective shape and color digitization of cluttered 3D artworks.* ACM Journal on Computing and Cultural Heritage *8* (2014), No. 1, 4–22.

[11] J. Feng, N. X. Zhao: *A new role of Chinese academic librarians-The development of embedded patent information services at Nanjing Technology University Library, China.* The Journal of Academic Librarianship *41* (2015), No. 3, 292–300.

[12] L. Zhang: *Discussion on subject knowledge service pattern of Agricultural University Libraries.* Journal of Computer & Communications *4* (2016), No. 1, 33–38.

[13] Q. H. Weng , D. S. Lu: *A sub-pixel analysis of urbanization effect on land surface temperature and its interplay with impervious surface and vegetation coverage in Indianapolis, United States.* International Journal of Applied Earth Observation and Geoinformation *10* (2008), No. 1, 68–83.

[14] Y. Xiong, S. Huang, F. Chen, H. Fe, C. Wang, C. Zhu: *The impacts of rapid urbanization on the thermal environment: A remote sensing study of Guangzhou, South China*. Remote Sensing *4* (2012), No. 7, 2033–2056.

[15] B. Yang, J. Li, Q. Zhang: *G language based design of virtual experiment platform for communication with measurement and control*. Procedia Engineering *29* (2012), 1549–1553.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 385–394

Key technologies of distributed file system for big data analysis

Wei Wan[1]

Abstract. With the continuous development of Internet technology, people have higher requirements for the ability of using computer to analyze and process data. In order to further clarify the way of effectively optimizing the distributed file system, the research status of the distributed file system at home and abroad was introduced first. The construction and application of the algorithm model were expounded. Finally, the data obtained from the system test was analyzed and the conclusions were drawn. The results show that the performance of distributed file system based on big data analysis is high and the time consuming is short. The results of the study were summarized by the author, and the relevant suggestions for optimizing the key technologies were proposed, so as to provide a strong basis for the effective application of distributed file system technology.

Key words. Big data technology, key technology.

1. Introduction

In recent years, the rapid development of big data technology has promoted the development of computer storage technology to a certain extent, and has led to the development of China's economy and society. The distributed file system is a data management system with better data storage and analysis ability. The stable operation of the distributed file system is inseparable from the application of big data technology. With the advent of the Internet era, people's demand for data diversity and structured applications arises at the historic moment. At the same time, the need for data storage technology is becoming more and more urgent. The distributed file system built by big data technology meets the demand of data storage management to a great extent. At present, more and more people use big data technology to manage large amount of files, and apply distributed system to store data and information. Distributed file system has been widely recognized by the majority of users, but with the rapid increase in the number of data storage, the bearing capacity of the distributed file system has become a concern. Therefore,

[1]School of Electronic Commerce, Jiangxi Institute of Economic Adminstrator, NanChang, Jiangxi, 330088, China

the study of distributed file system technology for big data analysis is of great significance.

Based on this, the key technologies of distributed file system for big data analysis were analyzed and studied by constructing algorithm model and system test method in this paper.

2. State of the art

The research on big data technology and distributed file system was originated in foreign countries. At present, foreign academia has made some achievements in the research of distributed systems. Our scholars' research on big data technology and distributed file system mainly focuses on the development of big data technology and the application of distributed file system.

The above researcher are the introduction and discussion of the development of distributed file system, although these studies have discussed the distributed file system in detail, the existing research results are mostly theoretical research of distributed file system, which lacks the research of the influence of big data technology on distributed file system. Therefore, in view of the shortcomings of the existing research, in this paper, the algorithm model and system testing methods were proposed, and the key technologies of distributed file system for big data analysis were analyzed and studied. In addition, in the third part, the specific content of the research object and algorithm model were expounded; in the fourth part, the algorithm model and system test data and results of data analysis were concluded; and finally, in the last part, the relevant conclusions were summarized.

3. Methodology

In this paper, the key technologies of distributed file system for big data analysis were mainly analyzed and studied by constructing Paxos algorithm model and system testing method. The distributed file system is a typical distributed system, and the architecture of the distributed system represents or determines the service capability of the system [1]. The emergence of large data technology effectively optimizes the design of distributed file system, and the system's difficult architecture design brings its huge service capability. The distributed file system refers to the physical storage resources managed by the file system, not necessarily directly connected to the local nodes, but connected to the nodes through the computer network [2]. Table 1 shows the distributed file system related issues and the main content of the main solution technology. The design of a distributed file system is based on the client server model. A typical network may include multiple servers for multi-user access. Specifically, a variety of distributed file systems can adapt to a certain application environment, and play a superior performance, which can well meet the requirements of the computing system on the storage system in the storage technology development at all stages. Song has believed that the design and implementation of distributed file system are based on a certain storage structure,

which is divided according to the organization of storage media. Design and implementation of a distributed file system is a storage structure based on the storage medium according to the division of the organization, storage structure of the file system is divided into virtual storage structure and storage structure of object [3].

Table 1. Issues and techniques related to distributed file systems

Question	Relevant techniques adopted
Data equilibrium distribution	Improved consistent hashing algorithm
Data conflict processing	Vector clock
Temporary troubleshooting	Data return mechanism
Permanent fault recovery	Hash tree applications
Error monitoring	Error investigation

The Paxos algorithm is a consistency algorithm based on message passing. This algorithm is considered to be the most efficient of similar algorithms. In order to describe the Paxos algorithm, a virtual Greek city called Paxos is created. The island enacts laws in accordance with the political model of parliamentary democracy, but no one is willing to put all his time and energy on it. So neither the speaker nor the speaker or the clerk passing the note can promise to appear when someone else needs it, nor can promise to approve the resolution or pass the message. It is assumed that there is no Byzantine General problem, that is, although it is possible that a message is passed twice, but this will never be the wrong message. As long as waiting for enough time, the message will be passed to. Members correspond to each node, and the laws enacted correspond to the state of the system. Each node needs to enter a consistent state, and the consistency requirement corresponds to only one version of the legal [4]. Parliament corresponds to a distributed system, parliamentarians correspond to each node, and the laws enacted correspond to the state of the system. Each node needs to enter a consistent state, and the consistency requirement corresponds to only one version of the legal provision. The uncertainty of the MP and the attendant corresponds to the unreliability of the node and message passing channel [5] is

$$\lambda = \frac{1}{\mathrm{e}}\,. \tag{1}$$

The construction of Paxos algorithm model mainly includes metadata replication consistency and algorithm optimization and fault tolerance analysis and so on. The test of consistency of metadata replication should first define the failure rate according to formula (1), the time for data replication tasks should be set, and finally the reliability of the data should be calculated and estimated based on formula (2). Data reliability is the primary measure of metadata consistency testing. Consistency of metadata replication is the basis for the construction of the algorithmic model.

$$R(t) = \mathrm{e}^{-\lambda t}\,. \tag{2}$$

Other election of Axos algorithm with distributed computing as well as in the message passing algorithm based on two-phase commit process, due to the large amount of communication, the maintenance of data consistency process cost is too large, so that the big data platform using the consistency model, reducing the reliability and availability of the upgrade the space [6]. For example, due to its large amount of traffic, the two-stage commit process makes the process of maintaining data consistency too costly. As a result, the large data platform adopts a simple consistency model, which reduces the space for upgrading reliability and availability. Therefore, the optimization goal of the Paxos algorithm model is to reduce the traffic of the algorithm and improve the performance. The Paxos algorithm is optimized as follows:

Firstly, the core idea of the algorithm model is the basis of the operation, and the parameters of the distributed system are assumed and the research nodes of the server are found. Equation (3) is applied to the calculation and analysis of the damaged nodes of the distributed system. The calculation and analysis of the damaged nodes is mainly to clarify the basic data of the distributed system. At the same time, the analysis result of the damaged node can be used as a reference for the distributed system to reduce the communication loss, which is beneficial to the distributed system to improve its operation efficiency. At the same time, the operation of damaged nodes is the basic premise of system performance testing

$$R(t) = p\{T \geq t\} = 1 - \prod_{t=1}^{m} [1 - R_i(t)] . \tag{3}$$

Secondly, formula (3) is applied to analyze the frequency of fault occurrence and the probability of the specific situation of the algorithm model, and the results obtained can be used as a reference for the analysis of the running efficiency of the distributed file system. The problem of the distributed file system is caused by the number of nodes sending and receiving and the wrong path. And the failure problem is the key point of algorithm model optimization and improvement [7]. The fault problem is the key content of the optimization and improvement of the algorithm model. At the same time, the delay in the sending and receiving of node information is also a measure of the main indicators of the distributed file system. Formula (4) is used to calculate the delay time of the node information of the distributed file system, and determine the operating efficiency of the distributed file system through the obtained delay time. If the delay time is too long, then the distributed file system is less efficient. Once this happens, the algorithm model can be optimized and improved by modifying the algorithm model and repeating the calculation of the delay time [8].

$$R(t) = \mathrm{e}^{-\lambda t} \sum_{k=0}^{m-1} \frac{(\lambda t)^k}{k!} \tag{4}$$

Thirdly, the fault tolerance analysis of the algorithm model is mainly based on the formula (5). The fault-tolerant analysis of the algorithm model is based on the optimization of the algorithm model. The main process includes the following

aspects: Firstly, the metadata is numbered and the data information is sent. Then, according to the situation of the delivery and feedback of the data information, the transmission and reception of data information is counted and the specific fault tolerance is obtained. The calculation of fault tolerance is mainly to measure the accuracy of the application of the algorithm model, so as to ensure the effectiveness of the applied algorithm model. In the process of calculating the fault tolerance of the algorithm model, the algorithm parameters may appear in the case of infinite loop competition, because any one of the larger parameter values will always repeat the same motion with the larger number. Therefore, the fault tolerance analysis of the algorithm model must be based on the reasonable setting of the algorithm parameters.

$$R(t) = \sum_{j=k}^{m} \left(\begin{array}{c} m \\ j \end{array} \right) \mathrm{e}^{-\mathrm{j}\lambda t} \left(1 - \mathrm{e}^{\lambda t}\right)^{m-j} . \tag{5}$$

Many large data analysis algorithms will eventually boil down to basic machine learning and data mining algorithms [9]. However, for dealing with large-scale data sets, many existing serialized machine learning and data mining algorithms are difficult to complete model training and data processing within an acceptable time frame. The Paxos algorithm proposed in this paper takes full advantage of big data to analyze the advantages of application algorithms, and the model constructed by this algorithm largely reflects the ability of big data analysis technology to deal with massive complex data. The design goal of big data distributed file system for large distributed data storage system design, and provide a general scalable distributed cache scheduling framework, through the framework of the big data used different data access patterns and caching strategies and upper integration, and provides a set of efficient scheduling strategy covering many different data access mode for the user, thus speeding up the application of the upper big data read and write data access performance [10]. Through this framework, different data access patterns and cache policies and large data applications on the upper layer are integrated, which can provide users with a set of cache scheduling policies that cover a wide range of different data access patterns, thus speeding up the read and write access performance of big data applications in the upper layer. Figure 1 shows the overall structure of the big data distributed file system. The distributed file system cache scheduling framework has very good versatility and independence, its architecture, access mode and cache strategy can be applied to any distributed storage system.

The operation and algorithm testing of the Paxos algorithm model requires an asynchronous reliable communication system. The transmitted data may be lost, delayed or repeated, but will not be tampered with. In this paper, the ordinary computer was chosen as the server, and the corresponding node was selected to analyze and study the key technologies of the distributed file system. In the course of the experiment, the data of the node was tested and the delay and loss were tested to ensure the credibility of the system test, and then the consistency of the communication node was checked. After the experiment, the number of nodes was controlled through the error mechanism, so as to effectively shorten the communication time and improve the communication efficiency. The algorithm model was

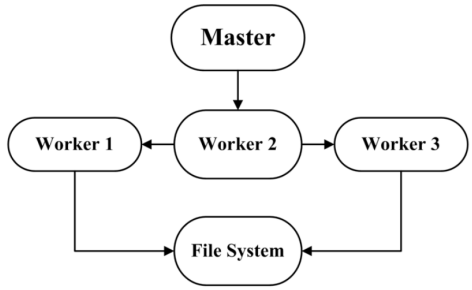


Fig. 1. Framework for big data distributed systems

applied to analyze the distributed file system of big data analysis, so as to effectively utilize the computer resources and improve the system availability on the basis of comprehensive understanding of node loss.

4. Result analysis and discussion

Different distributed file systems were set up in the computer A and B inside, and simulation experiments and tests were carried on for reading and writing large numbers of files. Computer A was set up the distributed file system with big data analysis, computer B was set up in general distributed file system. It was assumed that each node configuration of the simulation experiment was memory 4 G and hard disk 500 G. The simulation results obtained by running the file system are as follows:

Table 2. Comparison and analysis of system operation time

Number of read and write files	Time consuming of A system	Time consuming of B system
1	375	1190
3	1239	3465
5	2004	5987
7	2689	7982
9	3578	9861
11	4538	12445

The time-consuming situation of the operation of the distributed file system directly determined the operating efficiency of the system. As shown in Table 2, the time-consuming index obtained by the A-computer system test proposed in this paper was significantly lower than the time-consuming index obtained by the B-computer system test. In other words, the time-consuming index of big data distributed systems was only about one-third of the time consumed by other general distributed file systems. Thus, the operation time of big data analysis distributed

file system was smaller, which reflected the good operation efficiency of big data analysis distributed file system to a certain extent.

Based on the above analysis of the operational advantages of the big data analysis distributed system, the performance test of the system operation of the computer A which had the big data analysis distributed file system was carried on. The results obtained by performing performance tests on different system operations under different numbers of servers are shown in Table 3. As can be seen from the above data, the performance of the new, access to information, delete, and rename the directory to create the metadata system operation were increased with the increase of the server. Among them, the highest increase in performance was the directory creation and renaming of two system operations, which increased by 159.32 % and 156.45 %, respectively. The performance gains of other system operations were slightly lower than those of the two system operations. Overall, the increasing number of servers has greatly improved the performance of metadata.

Table 3. Metadata performance improvement test analysis

Operate	Newly built	Get information	Delete	Directory creation	Rename
1	159.32	156.45	–3.92	–9.23	–7.37
3	112.12	147.25	5.13	6.33	5.16
5	103.56	136.92	6.51	8.30	8.29
7	100.73	116.56	7.7	9.52	9.05

Distributed file system is mainly used to store files. Therefore, the system needs to meet the needs of users to read and write files. Fig. 2 shows the changes in the reading (up) and writing (bottom) performance indexes of distributed file system before and after the application of big data technology.

As shown in Fig. 2, A1 and B1 is the application of big data technology optimization computer A and B system to read and write performance curve, A2 and B2 after the application of big data technology to computer A and B system to read and write performance curve, System is a typical distributed file system data read and write performance curve. From Fig. 2 it can be seen in the data, before the application of big data technology, the B1 performance index was less than the A1 performance index, the number of visible, reads the metadata server computer system B is significantly lower than that of A computer system, and the value of A2 and B2 two performance curves are almost equal, visible, of B computer system to optimize the application of big data technology with the computer A system to read and write approximately the same performance level, distributed file system of the computer built-in read metadata server capability has been significantly improved. Thus, the application of big data technology greatly promotes the performance of metadata reading and writing.

Distributed file systems need to complete different commands and tasks. The performance changes of the main system operations under different commands and tasks are shown in Table 4 and Fig. 3. The performance improvement ratios of three operations of deleting, obtaining information and renaming under different metadata processing tasks were 2.17 %, 4.55 % and 1.09 % respectively. The dynamic

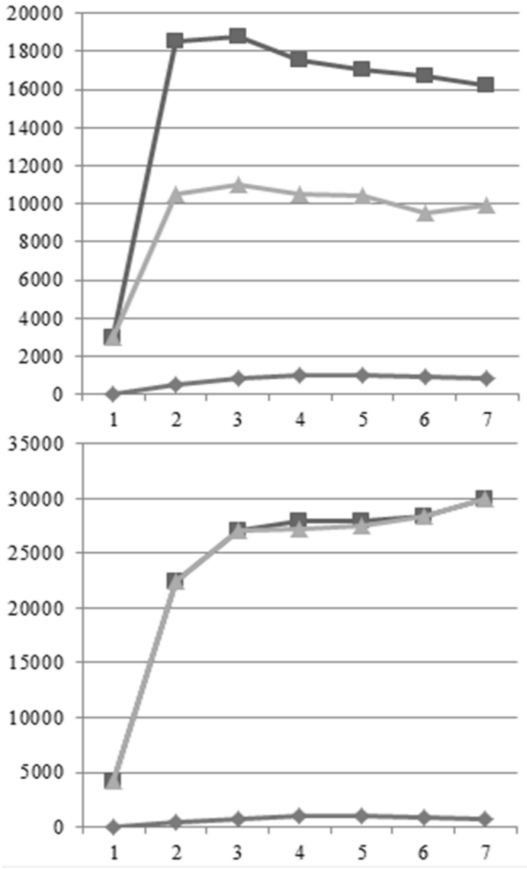


Fig. 2. Comprehensive test analysis of metadata read and write performance

trend of other performance was not stable, but overall, most of the performance index was slightly improved, the performance index declined, and the decrease in each operation was smaller, which was not more than 7 %. In summary, the main operational performance index of the distributed file system optimized by big data technology was higher, the overall performance of the distributed file system with big data analysis was better. Therefore, it was believed that on the basis of ensuring the normal operation of metadata processing, the application of big data technology can further enhance the scalability and usability of distributed file system. However, big data technology plays a less active role in other random operations of distributed systems. The improvement of mixed performance of distributed systems for big data analysis can be carried out in two aspects: enhancing system metadata throughput and reducing metadata consumption, thereby improving the overall performance of distributed file systems and promoting the continuous development of distributed file systems for big data analysis.

5. Conclusion

In order to find out the way to optimize the distributed file system by using the big data technology, the method of Paxos algorithm model was constructed, the metadata performance of distributed file system with large data analysis was systematically tested, the optimization of distributed file system was studied. Finally, the main conclusions were obtained as follows: the time-consuming of each operation of the distributed file system for big data analysis proposed in this paper is relatively short; at the same time, the overall performance level of the distributed file system optimized with big data technology is high.

Table 4. Analysis of metadata operations under different tasks

Task	Newly built	Get information	Delete	Directory creation	Rename	Build page	Fault feedback
Count	0.99	0.99	0.99	0.49	57.14	32.51	0.99
Text search	1.70	2.11	1.69	0.85	50.00	32.20	1.70
Random input	2.17	2.17	2.17	1.09	0.00	7.61	1.09
Sort	2.21	1.77	1.77	0.88	14.60	15.93	1.33

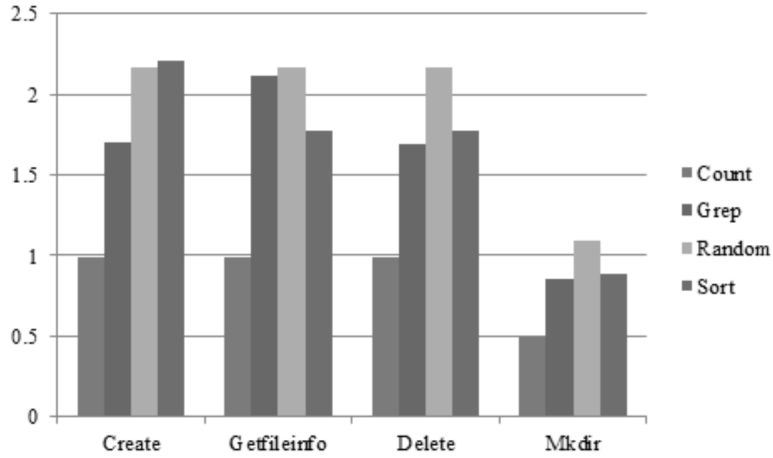


Fig. 3. Metadata performance comparison of different tasks

In summary, the Paxos algorithm model proposed in this paper is simple and reliable, and the application of the model is helpful to further analyze the effectiveness of the distributed file system. However, the research on big data analysis of distributed file system through the application of Paxos algorithm model lacks an analysis of the system user's visit. It can be seen that the algorithm model has some

shortcomings in the application of comprehensiveness. Therefore, in the future research, it is possible to add the operation and system test of user access performance index in the process of model analysis, so as to provide a more favorable basis for the application feasibility of Paxos algorithm model.

References

[1] R. C. Ionescu: *A scalable system for primal-dual optimization.* Journal of Computer Research and Development arXiv:1507.01456v1, 2015.

[2] J. Wang, X. Zhang, J. Zhang, J. Yin, D. Han, R. Wang, D. Huang: *Deister: A light-weight autonomous block management in data-intensive file systems using deterministic declustering distribution.* Journal of Parallel and Distributed Computing *108* (2017), 3–13.

[3] J. Song, C. Guo, Z. Wang: *HaoLap: A Hadoop based OLAP system for big data.* Journal of Systems & Software *102* (2015), 167–181.

[4] E. A. Murray: *Method of securing files under the semi-trusted user threat model using symmetric keys and per-block key encryption.* US Patent US 9363247 B2, 2016

[5] J. Li, C. Pan, M. Lu: *A seismic data processing system based on fast distributed file system.* International Journal of Computers & Technology *14* (2015), No. 5, 5779–5788.

[6] M. Alam, S. Sethi: *Covert Channel Detection framework for cloud using distributed machine learning.* CoRR - Computing Research Repository - arXiv (2015), abs/1504.03539.

[7] Y. Cao, D. Sun: *Migrating large-scale air traffic modeling to the cloud.* Journal of Aerospace Information Systems *12* (2015), No. 2, 257–266.

[8] Z. Wang, Q. Ding, F. Gao, D. Shen, G. Yu: *iHDFS: A distributed file system supporting incremental computing.* IFIP Advances in Information & Communication Technology *503* (2015), 151–158.

[9] S. Saranya, M. Sarumathi, B. Swathi: *Dynamic preclusion of encroachment in Hadoop distributed file system.* Procedia Computer Science *50* (2015), 531–536.

[10] M. Mirakhorli, J. Cleland-Huang: *Detecting, tracing, and monitoring architectural tactics in code.* IEEE Transactions on Software Engineering *42* (2016), No. 3, 205–220.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 395–404

Construction and application of multi-dimensional management framework system of libraries in colleges and universities[1]

Meiling Xie[2], Yuanli Wang[2], Qian Zhao[2], Yan Zhang[2]

Abstract. Colleges and universities must face the work of book management, and the existing methods can't meet the demand. In particular, students can't find library information online. Based on this, in this paper, the management system of multi-dimensional framework of university library was constructed. In the study, the principles and roles of the multi-dimensional management framework were analyzed; the J2EE+SQL Server technology combination was utilized to implement the library management system; the SSH framework was introduced. Library management system consists of system management, library management, reader management, book lending management, and loan notification functions. The final system test results show that the system constructed in this paper can effectively improve the work efficiency of the staff in practical applications, so it has remarkable effects.

Key words. Library, SSH framework, system construction.

1. Introduction

2. Modeling of underwater robot kinematics

The subject comes from an institution of higher learning. After decades of development, the school library now has tens of thousands of books. However, due to too many varieties, the collection of each book is very limited, and the quantity is kept about 5 and 6, and therefore, it is necessary to improve the borrowing speed of books. If each book is kept by a user for too long, it will make it impossible for other users to borrow it for a long time. The above model is a common defect in

[1]This work was support by the science and technology project of Baoding (No. 14ZR081)
[2]Agricultural University of Hebei, Baoding, Hebei, 071000, China

the existing library system. For this reason, the system needs to introduce loan notification function to remind users to return books in time. In addition, the existing system does not support the online lending function, which remains to be addressed in the new system. The new library management system can effectively improve the work efficiency and the running efficiency of books. In a word, the existing book management model has many shortcomings, such as lower security, efficiency and poor ease of use, so it is unable to meet the management needs of library staffs. Therefore, the introduction of new means of library management has become an inevitable trend of development.

2.1. State of the art

It was not until the end of last century that libraries in China began to change to modernization. At first, many large libraries introduced information technology to replace the traditional model. Early library management systems used the C/S model mostly. The function of the system is to manage the book information, so the user can't find relevant information, and they can only go to the library to deal with related information [1]. In the countries where computer technology is widely used, most of the computer technology is used in the economic business, and a small part of it is applied in the business of science and technology [2]. Some foreign countries tried to introduce information technology as early as 1954. After half a century of development, the library has now formed a complete management system, and our country is trying to narrow the gap with foreign countries. Therefore, domestic related personnel need to make efforts to improve and popularize library informatization in China. Compared with foreign countries, there is still a big gap in the field of information technology research in China [3]. In 1998, when visiting the Beijing library, Premier Li Lanqing said, "in the future, information technology will become an important part of the library". Various resources in the library will be digitized to realize online inquiry and management. Therefore, digitalization is the future trend of library development [4]. In the digital library, various business involved in the books management must be realized, including borrowing, returning and so on, so as to improve the efficiency of the business [5]. At present, many large universities in China have already realized the informatization construction of library management. In China, the more famous Tsinghua University, Peking University and Zhejiang University have already realized the online processing of library management services. However, in many basic colleges and universities, the construction process of information technology is relatively slow, and there are even phenomena of using manual mode to deal with book business [6].

3. Methodology

The main task of the library management system designed in this paper is to manage all aspects of library information, including library management. Management of the above information must be conducted online. The physical structure of the library management system is shown in Fig. 1.

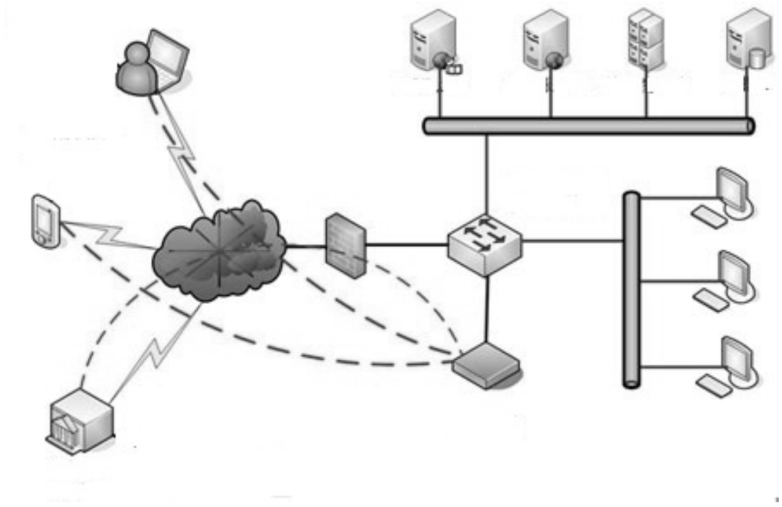


Fig. 1. Physical structure of the library management system

In order to make the final library management system meet the application requirements of the library, the system should follow some principles in the design process, including the following aspects:

The practical principle: whether the system is matched with the library management business is the first thing to be considered, otherwise, the system does not have practicability and will not be applied [7].

Usability principle: practicability is the first rule of library management system. On the practical basis, the ease of use should be taken into account. If the operation complexity of the library management system is too high, then it will affect the work efficiency, which is not advisable. At the same time, the system with higher availability can reduce training costs, and users can master the use of library management system in the shortest possible time [8]. Ease of use can be considered in terms of operating hints and operating documents.

Advanced and mature principles: when building a library management system, advanced technology architecture should be used, so as to ensure that the system can meet the demand for a period of time. Otherwise, it should be rebuilt, which will increase the cost of library management information costs, so it is not desirable.

The principles of stability: staff, students, teachers and administrators and so on often need to access the library management system to deal with library management business. Therefore, if the system is not stable enough and often inaccessible, then the business will be interrupted and the library staff's work efficiency will be affected, which is contrary to the purpose of promoting the information construction of library management [9].

The working principles of the library management system are shown in Fig. 2.

From the above picture, the library management system consists of three layers, and each layer undertakes different tasks, as follows.

Interface layer: the interface layer is the window of the library management

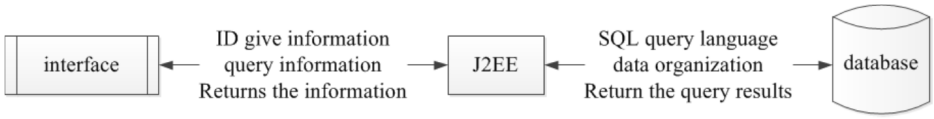


Fig. 2. Architecture diagram

system, and all the business requests need to accept the request through the interface layer, and feedback the result of the request processing to the user. In order to build the system interface independently, the Struts framework is introduced in the system [10].

Business layer: the business layer is the core of the library management system, and all the business logic is responsible for it. At the same time, it is also responsible for interaction with the database.

Data layer: the data layer stores the library data and borrowing data in the books management system. In addition, it also provides JDBC technology for data interaction.

Library management system needs to include the following basic functions:

Library management module: library information in libraries is dynamic, and it will often buy books. To this end, it needs to provide such functions as adding, deleting, modifying, importing and exporting books information [11].

Reader management module: the reader management module is mainly used to manage the readers' related information, including the management of readers' types, the management of readers' information and the management of readers' rules.

Book lending management module: the book lending management module mainly manages the borrowing and returning of books, including books borrowing, returning, and overdue charging and other modules [12].

Loan notification module: the loan notification module is mainly used to remind users to return books in time, including SMS notification function and mail notification function.

System management module: this module maintains system security mainly from the point of view of code. It consists of login, logout, user management, data backup, data recovery and other functions [13].

The functional architecture diagram of the system is shown in Table 1.

Table 1. System function framework

Library management module	Add, delete, modify, import, export
System management module	Login, logout, user management, data backup, data recovery
Book lending module	Borrow management, return management, overdue return alarm
Notification module	SMS notification
Reader management module	Reader information management, reader type management and reader rule management

System management is the first barrier for the security of library management system, and it mainly guarantees the stable operation of the system from the code level. Data backup and data recovery function can improve the security of the system data; login function can ensure the legitimacy of system operation; construction authority distribution system for the library management system can clear the user management function; the write off function prevents login information stored in the browser from being stolen [14].

The library management module can be adapted to the dynamic changes of library books, and it includes the addition, deletion, modification, import and export of library information. The book adding function can add books information to the system one by one. This function is inefficient when there are books in bulk entry. Therefore, the introduction of books import function can realize the batch increase of books information. When there is an error in the library information, it can be deleted or modified from the system database. Library staff can export books and information from the system to the local for inspection.

Readers are an important part of the library management system, and they main use library lending function that is composed of reader type management, reader information management, reader rule management. The reader type management is mainly used for adding, deleting and modifying the types of readers, such as teachers users and students users; the reader information management is mainly used to add, delete, modify and query the relevant information of readers; readers rule management is mainly used for the management of various penalty rules. As can be seen, each function is used to add, delete, and modify information [15].

The book lending management module mainly manages the borrowing and returning of books, including books borrowing, returning, and overdue charging and other functions. Book borrowing mainly includes two parts, such as loan application, loan application processing and so on, so as to support the user's online loan application processing. When returning, the time of borrowing books will be cleared. The overdue fee is used to record the fine after the violation of the reader, so as to carry on the statistical analysis.

The loan notification module is mainly used to remind users to return books in time, including SMS notification function and mail notification function.

4. Result analysis and discussion

The test is the last link to deploy the library management system to the library, so this link plays an important role in the whole library management system construction process. Without strict system testing, if the system is deployed to the library for practical application, it may have a serious impact on the work of the entire library. If functions are not implemented or performance is not up to date, a large amount of cost will be spent on maintenance after the system is deployed. Most importantly, it does not meet the business processing requirements of library management. As a result, rigorous testing is required in two areas: performance testing and functional testing.

First of all, with the login function of the system and the reader type add function

as an example, the library management system is tested by the use case to determine whether the system function meets the functional requirements.

Table 2, for example, is a test case for the login function test.

Table 2. Test case for login function test

Input equivalence class	Test case specification	Test data	Expected result	Reasons for selection
Username and password	User name is empty	null, user	The user name cannot be empty	The username is null and does not meet the validation requirements
	The password is empty	user, null	The password is not empty	The password is null and does not meet the requirements of the verification
	Enter the correct user name, the wrong password	user, 1234567	The login failed. Check that your user name or password is correct	User name and password must match
	Enter the correct password, the wrong user name	1234567, user	The login failed. Check that your user name or password is correct	User name and password must match
	Enter the correct user name and password	user, user	The login failed. Check that your user name or password is correct	Both can login and need to assign permissions based on the user type

Login function is the first barrier of library management system. Any user who wants to operate the relevant function must be authenticated by the login function. At the same time, the user type corresponding to the user will be found and assigned permissions based on the type. In this way, the clear permission distribution system of the library management system can be guaranteed. The login function of the system requires the user to enter the user name and password.

Table 3 is the test case for added functions of reader type.

In the process of adding a reader type, the type of information to be submitted includes the name of the reader type and the description of the reader type, and so on. If the reader type information can be successfully added to the library management system, the following conditions must be satisfied. Firstly, the type name cannot be repeated; secondly, the type name cannot be null.

Table 3. Test case for added functions of reader type

Input equivalence class	Test case specification	Test data	Expected result	Reasons for selection
The name of The name of the reader type and the description of the reader type	Reader type names are empty; others satisfy system formatting requirements	null, ss	Prompts "add failed" and red flag for the file reader type name	The reader type name is required, not null
	The reader type repeats the name; others satisfy the system formatting requirements	test, test	Prompts "add failed" and prompts the reader for type names not to be repeated	The reader type name cannot be repeated
	The reader type name is correct; the type description is null	test1, null	Prompts readers to add success type	The reader type description can be null
	The reader type name is correct; the type description is not null	test1, test	Prompts readers to add success type	There is no formatting requirement for reader type descriptions

Table 4 is the test case for SMS notification function.

The lending notification function mainly uses the short note and the mail way to notify the borrowing book person to return the book in time. Key testing items include SMS server configuration, SMS content, and so on. The points for testing include server configuration, receive numbers, and SMS content. The receive number should be numeric and the content cannot be null.

Performance testing is a very important part of the testing process. If the system performance is not up to the set target, the library management system cannot be deployed to the library network environment. Otherwise, it will bring losses to the library. When testing, the basic indexes such as response time, concurrent user number and memory share should be paid more attention to. If the actual test results meet the set goals, then the system meets the standards.

In testing, it is impossible to organize 100 people to run concurrent tests on the library management system. Thus, the professional test tool Loadrunner was introduced to simulate users and perform concurrent operations. During the test, 10 people were logged in every 2 seconds. In this process, attention should be paid to the response time and memory footprint of the login function. When the number of concurrent users reached 100, each index of the library management system was viewed and recorded. Specific test indicators were shown in Table 5.

Table 4. Test case for SMS notification function

Input equivalence class	Test case specification	Test data	Expected result	Reasons for selection
The configuration of the SMS server and the number of the receiver and the content of the SMS	The SMS server is not configured; others meet the system formatting requirements	False 13366666666 return books	Send "failed"	SMS server must be configured
	The SMS server is configured. The phone number is incorrect	True 1336A666666 return books	Send "failed"	The phone number must be correct
	The content is empty; the other items are correct	True 13368666666 Null	Send "failed"	Send content cannot be empty
	All items meet the format requirements	True 13366666666 return books	Send "successful"	Send successfully

Table 5. Results of system performance test

Concurrent users	Response time (seconds)	Application server ICPU utilization (%)
10	0.23	1
20	0.25	1
30	0.32	1
40	0.45	3
50	0.55	7
60	0.60	10
70	1	15
80	1.5	18
90	2	20
100	3	25

According to the test results in Table 3, the response speed of the system did not increase with the number of concurrent users and there was no occurrence of a significant change. When the number of concurrent users reached the system specifications, the response rate was still less than 5 seconds, but the performance remained normal. From this, it can be known that the existing performance of the

library management system can meet the requirements of the library and can be deployed into the actual environment of the library.

5. Conclusion

The rapid development of information technology has promoted the application of information technology. All fields are promoting informatization construction, so as to improve work efficiency and competitiveness. In order to solve the shortcomings of the existing library management system, the university library management system based on multi-dimensional management framework was studied in this paper. Based on the analysis of the multidimensional management framework, the SSH framework was introduced. According to different functional requirements, the library management systems were classified, implemented and tested. Through this study, the following conclusions were obtained: on the basis of in-depth investigation of the requirements of various types of users of the system, functional requirements of library management system were identified, consisting of five parts, such as system management, library management, reader management, book borrowing management and loan notification. On the basis of the determination of functional requirements, the performance requirements of the system remained to be determined; and in the design phase, the following tasks: data access model design, security design and technical framework design should be focused on. The most important point is to determine the implementation details of each module function in sequence diagrams. Of course, there are still some shortcomings in this paper, thus needing further improvements in future studies. For example, with the updating of library facilities and equipment, in the future, books will be pasted with radio frequency cards to gradually help the realization of self-borrowing and self-returning functions of books.

References

[1] S. G. Zhang, H. H. An, Y. W. Zhang: *Improving and perfecting project contract management by synergy management mechanism.* Applied Mechanics and Materials *315–317* (2014), 3732–3735.

[2] X. Zhao, J. Xie, W. J. Zhang: *The impact of information sharing and ordering coordination on supply chain performance.* Supply Chain Management: An International Journal *7* (2002), No. 1, 24–40.

[3] J. Li, F. Jin: *Two-retailer competitive simulation model in supply chain.* International Journal of Advancements in Computing Technology *4* (2012), No. 19, 635–643.

[4] Q. He, L. Luo, Y. Li, Y. Lu, J. Wang: *Developing a measurement model for degree of management synergy of mega-projects.* Advances in Information Sciences & Service Sciences *5* (2013), No. 9, 651–659.

[5] S. Chatterjee: *Types of synergy and economic value: The impact of acquisitions on merging and rival firms.* Strategic Management Journal *7* (1986), No. 2, 119-139.

[6] T. Mo, Z. Wang, X. Xu, X. Wang: *A virtualization-based service system development method.* Journal of Service Science and Management *2* (2009), No. 1, 1–9.

[7] E. Dheur, J. Ferguson, R. Martens, A. Petrilli, B. Smale: *Modernizing cor-*

porate MIS: from information system modelling to implementation. Proc. European ORACLE Users Group Conference, 22–27 March 1992, Cannes, France, CERN Accelerating science (1992).

[8] J. Stansfield: *Prevalence of stuttering and cluttering in adults with mental handicaps.* Journal of Intellectual Disability Research *34* (1990), No. 4, 287–307.

[9] Y.van Zaalen-Op't Hof, F. Wijnend, P. H. de Jonckere: *Differential diagnostic characteristics between cluttering and stuttering—part one.* Journal of Fluency Disorders *34* (2009), No. 3, 137–154.

[10] A. Rauber, D. Merkl, M. Dittenbach: *The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data.* IEEE Transactions on Neural Networks *13*, (2002), No. 6, 1331–1341.

[11] Y. Li, K. C. Ng, D. J. Murray-Smith, G. Gray, K. C. Sharman: *Genetic algorithm automated approach to the design of sliding mode control systems.* International Journal of Control *63* (1996), No. 4, 721–739.

[12] J. L. Qiu, L. Zhou: *Through the prism of the internet cafemanaging access in an ecology of games.* China Information *19* (2005), No. 2, 261–297.

[13] H. Y. Zhang, S. P. Li, M. J. Sun: *Dissipative structure theory and the management system of University Library electronic reading room.* Microcomputer Information *26* (20101), No. 9, 49–51.

[14] D. S. Sanders, S. Read-Brown, D. C. Tu, W. E. Lambert, D. Choi, B. M. Almario, T. R. Yackel, A. S. Brown, M. F. Chiang: *Impact of an electronic health record operating room management system in ophthalmology on documentation time, surgical volume, and staffing.* JAMA Ophthalmol *132* (2014), No. 5, 586–592.

[15] S. Read-Brown, D. S. Sanders, A. S. Brown, T. R. Yackel, D. Choi, D. C. Tu, M. F. Chiang: *Time-motion analysis of clinical nursing documentation during implementation of an electronic operating room management system for ophthalmic surgery.* Proc. American Medical Informatics Association (AMIA), Annual Symposium, 16–20 November 2013, Washington DC, AMIA Proceedings (2013), 1195–1204.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 405–414

Trajectory operation and coordinated control of MATLAB manipulator in automobile manufacturing industry

Yang Gang[1]

Abstract. The trajectory operation and coordinated control of the MATLAB manipulator have a positive impact on the automation development of the automobile manufacturing industry. In order to improve China's current automobile manufacturing MATLAB manipulator application technology in the industry, the screw tightening of the engine cylinder head in the automobile manufacturing industry was taken as an example in this paper. The MATLAB manipulator was designed, and the trajectory operation and coordination control precision of the engine cylinder head mechanical arm were analyzed. Results showed that the trajectory motion and coordination control precision of the robot arm were closely related to the actual demand. Moreover, the precision grade of the mechanical arm used in this study could meet the actual needs of the automobile manufacturing industry. The conclusions provide theoretical basis and reference for the development and improvement of the mechanical arm of other production links in the automobile manufacturing industry.

Key words. MATLAB manipulator, trajectory operation, coordinated control, automobile manufacturing industry..

1. Introduction

With the rapid development of the world economy, great progress has been made in all sectors of the world. With the continuous improvement of the industrial structure and related theories, the comprehensive level of various industries has been greatly improved. Nowadays, with the development of the new era, a variety of more advanced technologies begin to appear. Especially with the development of computer technology, many automated industrial patterns have been studied and applied to the actual production and life. The development of these new technologies has brought convenience and development for people's production and life, and made people's production and life more colorful. In the era of the development environment, the automobile industry is a relatively new industry, which has grad-

[1]Chongqing Jiaotong University, Engineering Training Center, Chonggong, 400074, China

ually become one of the important traffic tools in people's life and production, and played an important role in people's life. With the continuous progress of science and technology, some of the more innovative technologies are gradually applied to the actual automotive industry processing, which have brought important driving forces and positive impact for the further improvement of the automotive industry and the promotion of comprehensive strength. As one of the most important product of the rapid development of science and technology, the development of MATLAB manipulator technology has played a complementary and auxiliary role for the development of automobile manufacturing technology in the present era, and it has a certain impact on the improvement of manufacturing efficiency in the process of automobile manufacturing. In this paper, the main influences of the trajectory operation and coordinated control of MATLAB manipulator on automobile manufacturing technology are analyzed, so as to provide a theoretical basis and reference for the development of automobile manufacturing industry in our country.

2. State of the art

With the development of the times, the automobile has gradually become one of the most important means of transportation in people's daily life. With the increasing demand for automobiles, the automobile manufacturing industry has gradually become one of the important pillar industries in the world economic development [1]. In recent years, the inevitability of the development of automobile manufacturing industry makes many scholars consider the application of more innovative concepts and techniques to the actual production of automobiles and to form a more innovative model of development, so as to bring certain impetus to the development of the automobile manufacturing industry [2]. Transmission is one of the most important parts of the automobile manufacturing industry. Many researchers have begun to recognize the inevitability of the development of this manufacturing process. With the development of various innovative technologies, especially the computer technology, it has a positive impact on the upgrading of automation technology [3]. As a new type of automobile manufacturing automation technology, MATLAB manipulator has a certain impetus for the promotion of manufacturing efficiency and human resource saving in the process of automobile manufacturing. Many enterprises have already introduced the new technology into actual automobile manufacturing industry, and combined it with the actual production, so that the car industry can obtain greater development and progress [4].

2.1. Methodology

With China's entry into the new century, the dynamics and influence of reform and opening to the outside world have led to a great deal of progress and development in many industries in China. Especially in the manufacturing industry, taking the automobile industry as an example, China has become an important import and export country of automotive products in the world. With the introduction of new theories and technologies, the industry model and concept of automobile man-

ufacturing industry have become more mature and perfect, which gradually become an important pillar industry in China. With the rapid development of automobile manufacturing industry in China, a number of new technologies have been developed and applied to the actual production of automobile manufacturing. The manipulator technology based on MATLAB has gradually attracted the attention of some automobile manufacturing industries in our country. In China's automotive industry, the application of MATLAB manipulator technology has gradually increased, and related theoretical research has made great progress [5]. However, with the rapid development of MATLAB manipulator technology in our country, the final application results are not satisfactory because of the relative weakness in some parts of the automobile manufacturing industry. For example, MATLAN manipulator technology is still relatively deficient in the manufacture and research and development of some auto parts in China's automobile manufacturing industry, and the relative technical level is relatively low. A series of factors have resulted in poor production capacity for some of the more sophisticated parts of the automotive industry in China, and many cars have to be imported from other countries. It makes China's automobile manufacturing and production subject to other countries, and the whole industry has the relatively high cost of production, which may cause China's automobile manufacturing industry experiencing a bottleneck period. Some statistics believed that in 2015, for China's automobile manufacturing industry, the local auto manufacturers only accounted for 5 % in the sales volume of ten foreign auto manufacturers or enterprises, and the sales of related automotive products were shown in Fig. 1.

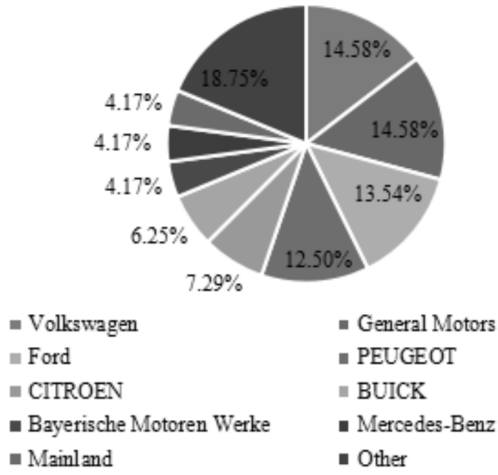


Fig. 1. Sales and marketing of China's auto manufacturing industry in 2015

In this paper, on the basis of a clear awareness of the status quo of China's automobile manufacturing industry, it is necessary to analyze and study the shortcomings of China's auto manufacturing industry, so as to optimize the limitations of its development. In addition, it can reduce the gap of China's automobile manufacturing industry and other foreign advanced enterprises, improve the competition ability

of China's automobile manufacturing industry, and make China get a foothold in the monopoly of the automobile manufacturing industry in developed countries in today's world [6]. This study will study and analyze the key technology and theory of MATLAB manipulator in China's automobile manufacturing industry, and then discuss the core part of it. On this basis, the main performance of MATLAB manipulator in automobile manufacturing industry is studied [7]. Then, this research will discuss the operation trajectory and coordination control system of MATLAB manipulator in the automobile manufacturing industry of our country, so as to confirm the research route of this research. The research directions are as follows.

(1) In this study, the current situation of automobile manufacturing industry in our country is analyzed by reading relevant data. The defects and shortcomings of MATLAB manipulator application in China's automobile manufacturing industry are summarized, and the key links and technologies that need to be solved are discussed.

(2) On the basis of a clear understanding of the related technologies and theories of robotic manipulators, the manipulator model is constructed by using the related mathematical models of forward and backward kinematics and dynamics used in the research of manipulator [8]. Through the analysis and discussion of the relevant data, we further optimize the trajectory operation system and coordinated control system of the MATLAB manipulator, so as to complete the following related experiments. The construction process of the relevant mathematical model theory is as follows.

The azimuth description model of each link of robot arm is

$$\overset{A}{B}R = \left[{}^{A}X_{B}\, {}^{A}Y_{B}\, {}^{A}Z_{B}\right] = \begin{bmatrix} n_x & n_y & n_z \\ o_x & o_y & o_z \\ a_x & a_y & a_z \end{bmatrix}, \tag{1}$$

where $\overset{A}{B}R$ is the rotation matrix of each component of the manipulator, and n, o, a are the projection components of different links in different directions [9], as shown in Fig. 2.

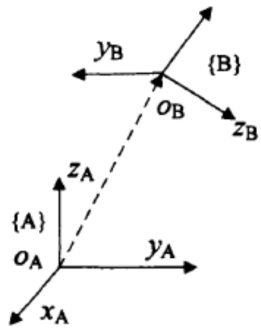


Fig. 2. Description of azimuth of different links in different directions

After setting up the parameters of the manipulator, the mathematical model is constructed based on the theory of forward kinematics. The model is constructed

as follows.

$$ {}^{i}_{i-1}T = \mathrm{Rot}(X_{i-1}, \alpha_{i-1})\mathrm{Trans}(X_{i-1}, \alpha_{i-1})\mathrm{Trans}(X_{i-1}(Z_i, d_i)\mathrm{Rot}(Z_i, \theta_i)\,, \qquad (2) $$

where ${}^{i}_{i-1}T$ is the transformation matrix of the manipulator, X_i and Z_i belong to different links of the manipulator, respectively, α_i and θ_i represent the rotation angle of the robotic arm during the operation, and d_i represents the translational distance in the running of the manipulator.

On the basis of understanding the inverse kinematics and related parameters, in order to obtain a higher manipulator trajectory scheme, the related mathematical model of forward kinematics is further introduced [10]. The model design is

$$ {}^{i}_{i+1}T = {}^{1}_{2}T^{2}_{3}T^{3}_{4}T^{4}_{5}T^{5}_{6}T \cdots {}^{i}_{i+1}T\,. \qquad (3) $$

(3) Finally, this study takes a car manufacturing industry in China as the research object. In view of the lack of automobile manufacturing in the operation process of the automobile manufacturing enterprise, more and more single robot arms are used only in the production process. It leads to low production efficiency and it cannot better meet the actual demand for automobile production. For this series of uncoordinated problems, further research is made on the trajectory operation and coordination control of the multi-robot manipulator, so as to provide a solution for the development of the entire automotive industry.

3. Results analysis and discussion

In order to improve the comprehensive strength of China's automobile manufacturing industry, many researchers in China have begun to apply the MATLAB mechanical arm technology to the actual automobile production (see Fig. 3).

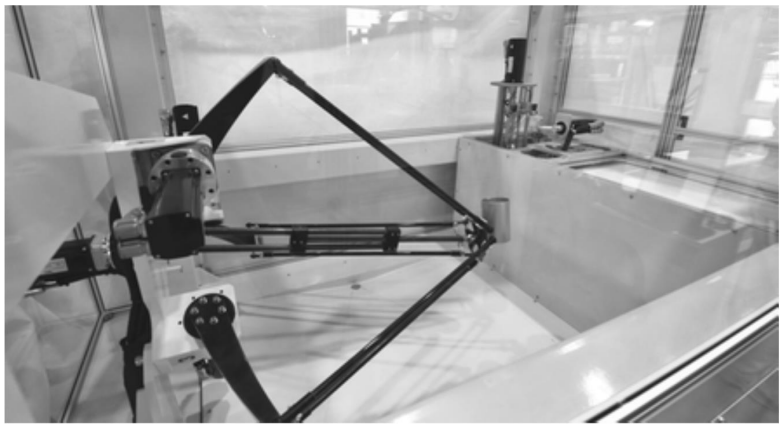


Fig. 3. Development and application of MALTAB manipulator

However, due to the relatively low level of research on this theory in our country, the key technology of MATLAB manipulator is still insufficient and defective [11].

Based on the investigation of related data and automobile manufacturing industry, the development trend and research center of mechanical arm technology in our country are summarized as follows [12]:

(1) In order to better combine the mechanical arm with the car in the manufacturing process and form a modular operation mechanism, it is necessary to improve the material of the manipulator so as to improve the overall practicability of the manipulator and improve the bearing capacity of the manipulator.

(2) With the continuous development of automotive technology, the demand for some more sophisticated parts has begun to increase. This requires the design of manipulator in the related links may need to optimize the processing, so that the manipulator has a certain degree of flexibility, and achieve the same hand or beyond the actual operation ability.

(3) It can improve the accuracy and stability of the manipulator in the automotive manufacturing industry, so that the final production of related automotive components can reach the world advanced level.

(4) On the basis of the rapid development of computer technology and sensors, the organic combination of robot arm and human being is further realized, so that the manipulator can better recognize the needs of the entire industry, provide some planning for subsequent development, and improve the efficiency of the final automobile industry.

This research mainly focuses on the current status of the important parts-engine in the automobile manufacturing industry. As an important part of the automobile, the engine has a very important influence on the operation of the automobile. Only with a higher performance of the engine can make the car have higher running speed, so that the level of China's automobile industry has been greater development and progress [13]. In the traditional automobile manufacturing process, the fastening of the engine bolts is done manually. However, the efficiency of manual operation is relatively low, and the process of human operation may be affected by some human factors, it may have a negative impact on the final quality of cars and ultimately weaken China's automobile influence [14]. Therefore, the main operation process of the robot arm object is to fix the bolt of the cylinder head of the engine [15]. The main object of the study was the four-cylinder engine in the automobile production. The 3D model and size are shown in Fig. 4. Before using the manipulator for automatic assembly, the engine enters a fixed mounting position. In addition, the bolt part of the cylinder head is manually installed with screws, and the final installation task is finished by the automatic rotation of the manipulator.

In the study, the installation requirements of the cylinder head installation of the automotive engine types involved in other references are summarized. The results are shown in Table 1, which can be used for the index design of the follow-up manipulator, so as to further improve the final applicability of the manipulator involved.

Table 1. Assembly requirements for some engine cylinder heads

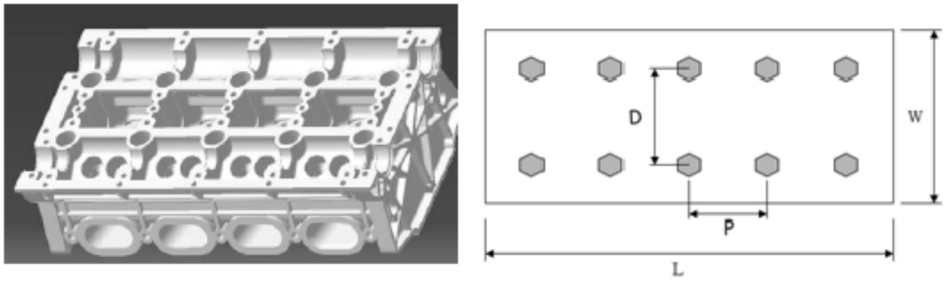


Fig. 4. 3D appearance and dimensions of a four-cylinder engine for study

Model	Cummins 6BT	6114	EQ6102	6108ZQ
Bore	102 mm	114 mm	102 mm	108 mm
Bolt longitudinal spacing	85 mm	101 mm	85 mm	95 mm
Target torque	165∼180 Nm	210∼230 Nm	150∼165 Nm	175∼190 Nm
Tightening process	120 Nm + 90°	142 Nm + 100°	90 Nm + 80°	110 Nm + 90°

On the basis of summarizing and describing the installation process of other automotive engine types and cylinder heads, the parameters of the manipulator are set by summarizing the related data, so as to lay a foundation for the follow-up manipulator trajectory operation and coordination control system optimization. By using the formula mentioned above, the parameters of the relevant manipulator links used to study the screw rotation of the automotive engine cylinder head are calculated, as shown in Table 2.

Table 2. Link $D-H$ parameter setting of the manipulator

n	a_{n-1} (mm)	α_{n-1} (°)	d_n(mm)	θ_n (°)	θ_n range (°)
1	0	0	0	θ_1	–170∼170
2	170	–90	0	θ_2	–85∼120
3	560	0	0	θ_3	–170∼85
4	155	–90	365	θ_4	–180∼+180
5	0	90	0	θ_5	-135∼135
6	0	–90	0	θ_6	–360∼360

The design of manipulator in the tightening process of automobile engine cylinder head is studied. Firstly, the force of the cylinder head before and after the bolt tightening is analyzed. In order to make the overall stress of the cylinder head of the engine more balanced, not cause the shape change of the cylinder head, and cause material damage during tightening of cylinder head bolt, two feasible tightening sequences are designed for the bolt tightening sequence of the manipulator, as shown

in Fig. 5. The method of bolt tightening used in this study is the torque-angle method.

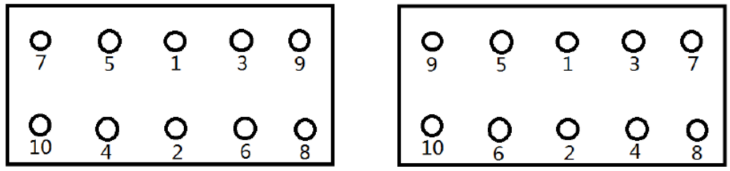


Fig. 5. Tightening sequence of engine cylinder head bolts

The parameters of the mechanical arm are set up, and the screwing order and tightening method of the engine cylinder head bolt are clearly defined. In this study, through the calculation of the relevant mechanical model, the model of the cylinder head bolt screwing machine of the automobile engine is designed, as shown in Fig. 6.

Fig. 6. Manipulator model of cylinder head bolt screwing of automobile engine

Finally, the application and analysis of the trajectory and coordination control accuracy of the manipulator designed are carried out. The results are shown in Table 3. The results show that the trajectory motion and coordination control precision of the manipulator are closely related to the actual demand. The different precision grades of this kind of mechanical arm can meet the actual demand in the automobile manufacturing industry.

Table 3. Precision operation results of trajectory motion and coordinated control of manipulator in different demands

Accuracy class		Lead accuracy of mechanical table operation (μm)									
		C0		C1		C2		C3		C4	
ELTP (mm)		RE	VV	RE	VV	RE	VV	RE	VV	RE	VV
>100	≤200	3.5	3	4.5	5	7	7	10	8	20	18
>200	≤315	4	4	6	5	8	7	12	8	23	18
>315	≤400	5	4	7	5	9	7	13	10	25	20
>400	≤500	6	4	8	5	10	7	15	10	27	20
>500	≤630	6	4	9	6	11	8	16	12	30	23
>630	≤800	7	5	10	7	13	9	18	13	35	25
>800	≤1000	8	6	11	8	15	10	21	15	40	27
>1000	≤1200	9	6	13	9	18	11	24	16	46	30

Note: ELTP–Effective length of thread part, RE–Running error, VV–Variable value

4. Conclusion

Nowadays, the demands for real life goods have increased the scale of production in manufacturing industry. As an important role in people's production and life, the development of the manufacturing industry has been increasing year by year. There is no doubt that the development of MATLAB manipulator technology will push the automobile industry. However, due to the lack of research and development of the manipulator technology in our country, there are still some phenomena of imperfect theory and immature development in the actual automobile manufacturing and production. In order to perfect and research the related theory, the related theories of trajectory operation and coordinated control of MATLAB manipulator were summarized in this paper. On the basis of understanding the theories, the tightening process of engine cylinder head in automobile production process was studied as an actual case. In the research, the parameters of the trajectory, operation and coordinated control of the MATLAB robot arm used in the engine cylinder head were set up, and the model was constructed. Finally, through the precision analysis of the mechanical arm model, it was found that this kind of robot arm could meet the needs of actual production. In this research, the manipulator of the engine cylinder head was simply designed with few levels and some defects. However, this study could still be used as a reference for the development of other manipulator.

References

[1] S. Arimoto, M. Sekimoto, S. Kawamura: *Task-space iterative learning for redundant robotic systems: Existence of a task-space control and convergence of learning.* SICE Journal of Control, Measurement, and System Integration *1* (2008), No. 4, 312–319.

[2] B. Bidikli, E. Tatlicioglu, E. Zergeroglu, A. Bayrak: *An asymptotically stable robust controller formulation for a class of MIMO nonlinear systems with uncertain dynamics.* International Journal of Systems Science *47* (2016), No. 12, 2913–2924.

[3] C. T. Freeman, E. Rogers, A. M. Hughes: *Iterative learning control in health care: Electrical stimulation and robotic-assisted upper-limb stroke rehabilitation.* IEEE Control Systems *32* (2012), No. 1, 18–43.

[4] D. D. Nam, Y. Yamashina, T. Namerikawa: *Bilateral teleoperation of multiple cooperative robots with time-varying delay.* IEEE International Conference on Control Applications, 8–10 September 2010, Yokohama, Japan, IEEE Conference Publications (2010), 2053–2058.

[5] H. B. Kang, J. H. Wang: *Adaptive robust control of 5 DOF upper-limb exoskeleton robot.* International Journal of Control, Automation and Systems *13* (2015), No. 3, 733–741.

[6] R. Marino, P. Tomei, C. M. Verrelli: *Learning control for nonlinear systems in output feedback form.* Systems & Control Letters *61* (2012), No. 12, 1242–1247.

[7] P. Mitra, G. Niemeyer: *Model-mediated telemanipulation.* International Journal of Robotics Research *27* (2008), No. 2, 253–262.

[8] V. Stepanyan, A. Kurdila: *Asymptotic tracking of uncertain systems with continuous control using adaptive bounding.* IEEE Transactions on Neural Networks *20* (2009), No. 8, 1320–1329.

[9] B. Xian, D. M. Dawson, M. S. de Queiroz, J. Chen: *A continuous asymptotic tracking control strategy for uncertain nonlinear systems.* IEEE Transactions on Automatic Control *49* (2004), No. 7, 1206–1211.

[10] Y. Zhang, B. Xian: *Continuous nonlinear asymptotic tracking control of an air-breathing hypersonic vehicle with flexible structural dynamics and external disturbances.* Nonlinear Dynamicse *83*, (2016), Nos. 1–2, 867–891.

[11] C. Bergeles, G. Y. Yang: *From passive tool holders to microsurgeons: Safer, smaller, smarter surgical robots.* IEEE Transactions on Biomedical Engineering *61* (2014), No. 5, 1565–1576.

[12] H. Rafii-Tari, C. J. Payne, G. Z. Yang: *Current and emerging robot-assisted endovascular catheterization technologies: A review.* Annals of Biomedical Engineering *42* (2014), No. 4, 697–715.

[13] R. Cappato, H. Calkins, S. A. Chen: *Prevalence and causes of fatal outcome in catheter ablation of atrial fibrillation.* Journal of the American College of Cardiology *53* (2009), No. 19, 1798–1803.

[14] E. Wissner, F. Ouyang, K. H. Kuck: *Examining the causes of ablation failure in the Wolff-Parkinson-White syndrome.* EP Europace *12* (2010), No. 6, 772–773.

[15] F. Sacher, M. Wright, U. B. Tedrow, M. D. O'Neill, P. Jais, M. Hocini, R. Macdonald, D. W. Davies, P. Kanagaratnam, N. Derval, L. Epstein, N. S. Peters, W. G. Stevenson, M. Haissaguerre: *Wolff-Parkinson-White ablation after a prior failure: A 7-year multicentre experience.* EP Europace *12* (2010), No. 6, 835–841.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 415–424

Network computing principle and application analysis based on distributed peer-to-peer

Lei Liu[1]

Abstract. The development of network computing principle based on distributed peer to peer is very important for the improvement of computer operation function. In order to improve the relative theory and technology of computer industry in our country, the principle of network computer based on distributed peer to peer was recognized in this paper. CPF technology was used as an example of this research, and the network computing algorithm was applied to the actual network operation process. The results show that the distributed peer-to-peer network computing principle has a great positive impact on the efficiency and accuracy of the computing process. The conclusions provide a theoretical support for the development of network computing technology in China, as well as s significant reference for the progress of China's overall economic industry.

Key words. Distributed, reciprocity, network computing principle.

1. Introduction

With the rapid development of the world economy, a lot of innovative sciences and technologies have been gradually developed in people's productions and lives, and have played a certain role in promoting the development of various industries with the combination of related enterprises and fields, and have had a positive impact. Among them, the development and application of computer network technology is one of the important technologies for the rapid development of science and technology in the current era. This research is mainly aimed at the computation principle and application of network computer technology, the purpose is to provide a theoretical basis for the development and progress of China's computer industry and provide a technical support for the improvement of the comprehensive economic level of our country.

[1]Department of Computer Science and Engineering, The City College of Jilin Jianzhu University, Changchun, Jilin, 130000, China

2. State of the art

Computer technology has more data information, so that a lot of industries can further analyze the status and trend of development through the analysis on the relevant data information [1]. Because of the support of a large amount of data and information, the industry can rely on the analysis of basic data in the process of development to sum up the pros and cons of each link of the development of the industry, so that it is of great significance and influence to the development plan and the target formulation of the industry. And the application of computer network technology in real life breaks the restriction of time and space better, which makes the exchange of information among various sectors increase, and provides more communication ways for the interactive development of enterprises [2]. Furthermore, it also can promote enterprises to communicate with related technologies and theories, and promote the development and improvement of related technologies too, so as to rectify their shortcomings, further enhance the overall strength and progress of enterprises and provide certain data supports and scientific basis [3].

Through the use of the computer network industry, there have been many major changes in the mode of development happening quietly in many industries, which are very important for the development of the industry itself and the promotion of the country's comprehensive strength [4]. With the rapid development of computer network technology, more and more enterprises and areas begin to apply this technology into the operation process of traditional industries, at the same time, the characteristic of data sharing of the computer network information technology promotes the development of the industry [5]. However, while each industry makes great progresses, it also requires further improvement and promotion of computer network technologies and theories, especially for the emergence of new innovation areas with more advanced development levels nowadays. The demand for computer network technology has gradually increased, which requires that the computer hardware facilities should be more perfect. The computer's speed and data transmission are more efficient, and then it has more accurate computing methods, thus promoting its own development [6]. In this trend, the upgrading of the computer network principle and theoretical improvement has gradually become one of the important subjects in the research of computer technology in the world. Only by a better investment in the subject of human and material resources, can we make the development of computer network technology continue to analyze its shortcomings. And it can also improve the related operation principle, so as to realize the win-win situation of computer network technology and other industries [7]. The application of computer network technology in China has gradually matured, and the development of computer technology has also provided a certain technical support for the progress of various industries in China.

2.1. Methodology

With the continuous improvement of China's comprehensive economic level, all the industries in our country have been developed to a certain extent, and the

introduction of computer network technology has greatly promoted the progress of various industries in China (Fig. 1).

Fig. 1. The development of computer technology

Various industries in our country have begun to combine the theory and technology of the computer with the traditional way of development of the industry gradually, which has brought more positive influences to the further improvement of the comprehensive level of the industry [8]. As the basic information and data in China's industry continues to rise, the amount of data continues to expand too, so that the analysis and processing of a large number of data resources have become an important research content of the industry. This requires computer network technology to support large data systems. And because of the increasing interaction among different industries, the importance of the exchange of basic information that is related to the industry has become a major way for all sectors of the industry to develop together. In the development of these industries, they use network computer technology to share their basic information by remote transmissions, so that more data resources can be excavated in many ways, which further realizes the efficient use of data resources and reduces the waste of relevant resources and information [9]. The development of computer network technology also further breaks down the restriction of time and space between different industries and regions, which has a certain impact on the lower cost of the development of the industry [10]. Through reading and analyzing the related data, it can be seen that the development of network computer technology has a great influence on the promotion of the comprehensive level of various industries in China. And our country has begun to gradually put more manpower and material resources into the development of computer network technology, and has made some considerable achievements. The researches of computer network technology in our country are more in the development of computer hardware, and the research and analysis of relative principles are still few, so that the network computing principle of the network computer cannot be further analyzed more fully, which indirectly affects the practical application of the related enterprises in the network computer technology to a certain extent. It may cause the lack of knowledge about the relevant theories of the enterprise, which cannot combine the theories and technologies with the development needs of the enterprise,

so it is extremely detrimental to the development of the enterprise. In this uncoordinated phenomenon, researchers in China have begun to gradually strengthen the importance of network computing principles, and the relevant theories have been applied to the actual production activities [11]. Therefore, the purpose of this research is to analyze the network computing principle and its application status based on distributed peer-to-peer. The purpose of this study is to provide a scientific basis for the continuous improvement and enrichment of the computer network theory in China, and provide technical supports for the comprehensive improvement of the overall economic level of our country.

In this study, the relevant data was firstly read and summarized, then the network computing principles based on distributed peer-to-peer and the related concepts were summarized [12]. Then, based on the related algorithms, the related characteristics of distributed peer-to-peer computing principles were analyzed (the DCF protocol under distributed peer-to-peer technology was the object of this study). The model equations used are as follows:

$$backoffTime = Random * \mathrm{a}SotTime\,, \tag{1}$$

$$BackoffTime = [0, CW]slotTime(CW_{\min} < CW < CW_{\max})\,, \tag{2}$$

$$AIFS[AC] = AIFSN[AC] * aSlotTime + aSIFSTime\,. \tag{3}$$

Here, $BackoffTime$ represents a distributed peer-to-peer model operation waiting time, $SlotTime$ is the protocol frame interval, $CW_{\min}$ represents the minimum contention window, $CW_{\max}$ represents the maximum contention window, AIFS is arbitration inter frame spacing, AIFSN is arbitration inter frame coefficient, and $SIFSTime$ represents the minimum time interval between frames.

Based on the full analysis and understanding of the relevant operating mechanism, the advantages of distributed peer-to-peer network computing principles were further analyzed.

2.2. Result analysis and discussion

With the rapid development of the world economy and the continuous progress of science and technology, our production and life have gradually entered a new era of information technology and industrialization with high developments. In this new era of development, the world's various industries are constantly receiving the impact of the development of information technologies. In this context, the rapid development of network computer technology has provided positive effects on the progress of various industries, such as the world's culture, economy, military, medicine, etc. [13]. With the rapid development of this technology, many industries are undergoing unprecedented changes and progress (Fig. 2). Nowadays, people's thinking and living habits are changing under the trend of constant accumulations of data and information, and the development of network computer technology is driving the change and progress of today's society with a more powerful means. Many scholars

believe that network computer technology, as the new and contemporary science and technology, plays a huge positive role in the collection and processing of data information. With the increasing emphasis on some basic information in various industries and fields, the combination of network technology and computer technology has gradually become the trend of the development of the industry. Especially the development of the large amount of data of network computer technology has promoted the progress of people's society and the development of the industry to a higher level. Under the influence of computer network technology, the cognition and achievement of various trades and fields are gradually enriched [14].

Fig. 2. Development of computer and other industries

With the rapid development of the world economy and the continuous optimization of the industry, the computer network technology has more massive databases, and its basic information collection process is more perfect, and it can make further collection and processing on the relevant data, so as to achieve different client sharing. As a result, it is constantly used in many fields and industries, such as in some commercial, financial, medical, health and aerospace industries. The development of these industries can access relatively large basic data information. The development of the industry and the improvement of the speed and efficiency require all sectors to introduce the network computer technology and rely on the basic performance of the technology, so as to realize the sustainable development of the industry. With the development of the times, the development of computer network technology has become an important research topic in the development of the contemporary era. The research of computer network technology has gradually turned from the research of its hardware facilities to the analysis and discussion of the related program calculation principles [15]. Especially with the development of network computer technology in various industries and fields, the influence has gradually increased. Now, many researchers have begun to gradually analyze the principles of network computer technology. On the basis of many limitations and restrictions of the existing mechanism, the shortcomings of the original mechanism are improved, so as to design an operational theory with the relatively perfect systematical analysis. This provides a certain technical support for the development of computer network

technology and a scientific basis for the application and combination of computer technology in other industries. A lot of computer network algorithm operation theories have been gradually mentioned and applied, and they have brought the cross-age significance for the quantifiable dimension of human life. The principle of distributed computer network is a new network computing model that is developed under the trend of expanding the Internet scale and increasing the bandwidth. The computer system of each terminal user is connected with each other, and each node in the connection process arbitrarily connects the plurality of lines without any rules. When a terminal line is damaged and blocked, it can rely on other lines to complete the use of related functions, so that the computing method can be more accurate and efficient to complete the sharing of information resources. The distributed technology of network computing came into being around 1980. The premise of the design is to let more users share the resources of some networks, and make the information run on multiple computers simultaneously. Several of the main techniques designed are summarized in Table 1.

Table 1. A brief overview of distributed technologies

	Distributed technologies	Brief overviews	Technological advantages
1	Grid computing	The large computational problem is broken up into many smaller parts, and then they are executed by a number of computers	It not only solves large-scale computing problems, but also supports cross-domain computing, and it takes full advantages of the large amount of idle resources of heterogeneous computers.
2	Peer-to-peer technology	It rises as the antithesis of the C/S model	There are no central nodes in the model, and the node resources are equal. Each node can serve as a resource for the server, and it also can also serve as a client to apply resources to other nodes.
3	Middleware technology	A service program between the operating system and application software that can provide the running and development environment for the upper application software	It can provide resource sharing for different distributed technologies. As an independent and reusable system software, middleware can calculate resource management and communication in the distributed system.
4	Message passing	A communication mechanism adopted by a distributed system that uses *send* and *receive* primitives to send and receive messages respectively	Large scale and high performance and portability. The technology can be used for parallel computer writing.
5	Web Service	Component programming	Whether it runs on the same operating system and whether it is programming the same language or not do not affect the communication between applications.

The main technical forms are divided into 5 major categories, and each of these distributed technologies has its own advantages and characteristics that are related to the technology. Only by analyzing and applying the characteristics of all kinds of distributed technologies, can it provide a certain theoretical support for the perfection of the relevant theories and the integration of traditional industries.

Then, the DCF protocol under the distributed peer-to-peer technology was taken as the object of this study, and the related characteristics were analyzed. First of all, the mechanism of DCF was explored, and the mechanism was analyzed, as shown in Fig. 3, and the operation mechanism of the graph was analyzed. The result shows that the operation model deals with the related data. Thus, in the process of data transmission, the priority degree and importance of data are further analyzed. It arranges the channel for the more important data information, thus ensuring that the priority data information can be further arranged and processed, and it also reduces the data transmission conflict in the transmission process due to the insufficient channel, so as to effectively ensure the efficiency and accuracy of data transmissions. Therefore, this kind of distributed peer-to-peer network computing technology can effectively realize the transmission and sharing of different types of data information, which has certain advantages.

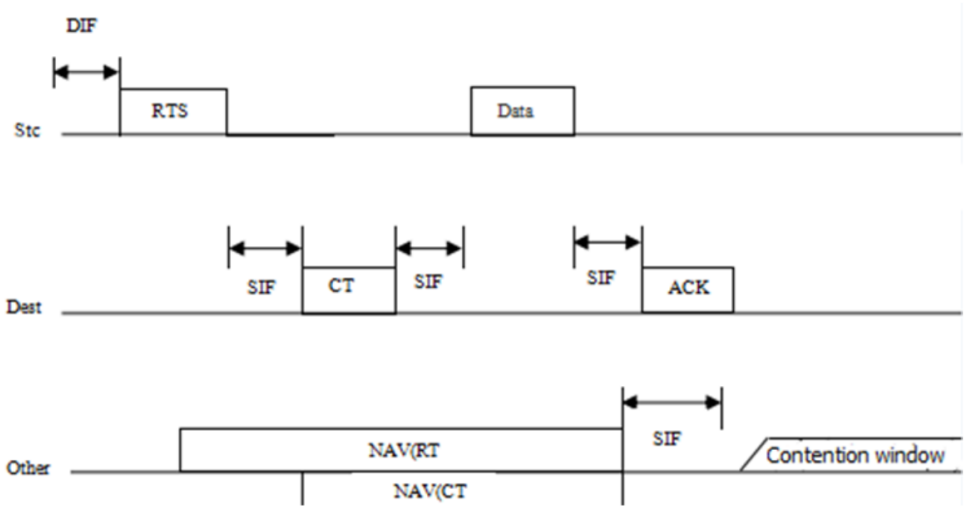


Fig. 3. Distributed peer-to-peer computing principles and process diagrams

On the basis of full understanding of the concepts and related theories of PCF's distributed peer-to-peer network computing technology, the relation between the priority level of data information and the access category of data information was analyzed. The results are shown in Table 2. On this basis, the relevant operational index data of PCF's distributed peer-to-peer network computing technology was further generalized. The results are shown in Table 3.

Through the summarization and analysis of the related distributed peer-to-peer network computing technology principles, and based on the cognition of theories and concepts of PCF's distributed peer-to-peer network computing technology, the

related data was analyzed, and then the advantages of the application of the distributed peer-to-peer network computing principle were confirmed. The analysis results are shown in Figure 4.

Table 2. A summary of the relation between data information priority levels and access categories of data information

Priority	UP	AC	Designation
Lowest to highest	0	AC-BK	Best Effort
	1	AC-BK	Background
	2	AC-BK	Background
	3	AC-BK	Best Effort
	4	AC-VI	Video
	5	AC-VI	Video
	6	AC-VO	Voice
	7	AC-VO	Voice

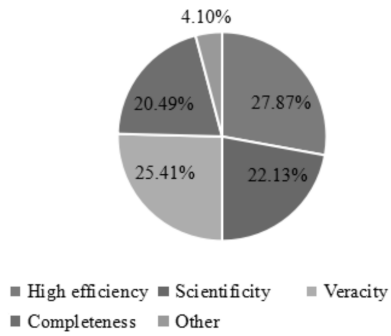


Fig. 4. Analysis of the advantages of distributed peer-to-peer computing

3. Conclusion

With the development of science and technology, the rapid development of modern technologies has brought great convenience for our productions and lives, so that people can contact the outside world more easily in today's era. Network technology, as one of the important technologies in the development of science and technology in the present era, has brought great impetus to the change of human lives. In particular, a large number of data reserves have been better transferred and shared, which has further realized the maximum utilization of data resources. There is no doubt that such a process can promote the production and life of the present era.

Table 3. Related operational index data for PCF's distributed peer-to-peer network computing technology

AC	CWmin	CWmax	AIFSN	TXOP		
				For PHYs defined in Clause 15 and Clause 18	For PHYs defined in Clause 17 and Clause 19	Other PHYS
AC-BK	aCWmin	aCWmax	7	0	0	0
AC-BE	aCWmin	aCWmax	3	0	0	0
AC-VI	(aCWmin+1)/2–1	aCWmin	2	6.016 ms	3.008 ms	0
AC-VO	(aCWmin+1)/4–1	(aCWmin+1)/2–1	2	3.246 ms	1.504 ms	0

With the development of the times, the demand for network technology continues to increase, and now the computer technology is beginning to gradually develop toward a more perfect theory. The development of distributed peer-to-peer network computing technology is one of the main manifestations of the rapid development of computer technologies. However, the relative theory of this technology is still lacking in our country. Therefore, in this study, the related theories and concepts of the technology were summarized and further illustrated by examples. The purpose of this research is to provide a theoretical basis for the development of distributed peer-to-peer computing technology in China, and provide a reference for the progress of the overall scientific and technological level in China. Because the related theory of equality is relatively complex, the author's theoretical level is limited, the result has certain limitation.

References

[1] K. W. Kwong, D. H. K. Tsang: *A congestion-aware search protocol for heterogeneous peer-to-peer networks.* Journal of Supercomputing *36* (2006), No. 3, 265–282.

[2] S. Sezer, S. Scott-Hayward, P. K. Chouhan, B. Fraser, D. Lake, J. Finnegan, N. Viljoen, M. Miller, N. Rao: *Are we ready for SDN? Implementation challenges for software-defined networks.* IEEE Communications Magazine *51* (2013), No. 7, 36–43.

[3] H. Li, P. Li, S. Guo, A. Nayak: *Byzantine-resilient secure software-defined networks with multiple controllers in cloud.* IEEE Transactions on Cloud Computing *2* (2014), No. 2, 436–447.

[4] I. Martinez-Yelmo, A. Bikfalvi, R. Cuevas, C. Guerrero, J. Garcia: *H-P2PSIP: Interconnection of P2PSIP domains for global multimedia services based on a hierarchical DHT overlay network.* Computer Networks *53* (2009), No. 4, 556–568.

[5] M. H. Aghdam, N. Ghasem-Aghaee, M. E. Basiri: *Text feature selection using ant colony optimization.* Expert Systems with Applications *36* (2009), No. 3, Part: 2, 6843–6853.

[6] W. Zhao, C. E. Davis: *Swarm intelligence based wavelet coefficient feature selection for mass spectral classification: An application to proteomics data.* Analytica Chimica Acta *651* (2009), No. 1, 15–23.

[7] Y. Marinakis, M. Marinaki, M. Doumpos, C. Zopounidis: *Ant colony and particle swarm optimization for financial classification problems.* Expert Systems with Applications *36* (2009), No. 7, 10604–10611.

[8] J. A. Lazzús, M. Rivera, I. Salfate, G. Pulgar-Villarroel, P. Rojas: *Application of particle swarm+ant colony optimization to calculate the interaction parameters on phase equilibria.* Journal of Engineering Thermophysics *25* (2016), No. 2, 216–226.

[9] N. Karaboga: *Transverse vibrations of orthotropic non-uniform rectangular plate with continuously varying density.* Journal of the Franklin Institute *346* (2009), No. 4, 328–348.

[10] G. Zhu, S. Kwong: *Gbest-guided artificial bee colony algorithm for numerical function optimization.* Applied Mathematics and Computation *217* (2010), No. 7, 3166–3173.

[11] M. S. Kiran, M. Gündüz: *The analysis of peculiar control parameters of artificial bee colony algorithm on the numerical optimization problems.* Journal of Computer & Communications *2* (2014), No. 4, 127–136.

[12] F. Kang, J. Li, H. Li: *Artificial bee colony algorithm and pattern search hybridized for global optimization.* Applied Soft Computing *13* (2013), No. 4, 1781–1791.

[13] S. C. Satapathy, A. Naik, K. Parvathi: *Weighted teaching-learning-based optimization for global function optimization.* Applied Mathematics *4* (2013), No. 3, 429–439.

[14] P. Rocca, G. Oliveri, A. Massa: *Differential evolution as applied to electromagnetics.* IEEE Antennas and Propagation Magazine *53* (2011), No. 1, 38–49.

[15] A. Qing: *Comment on "Differential evolution as applied to electromagnetics".* IEEE Antennas and Propagation Magazine *53* (2011), No. 4, 169–171.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 425–434

Research of an improved variable step size and forgetting echo cancellation algorithm[1]

Li Ang[2,3], Zheng Baoyu[3], Li Lei[3]

Abstract. In communication systems, the received signal may be interfered by various additive noises. They occur in many places of communication network and reduce the communication quality. Based on the adaptive filter theory, according to the ITU-T G.165 standard, combined with the advantages of forgetting factor and variable step algorithm, a new adaptive echo cancellation algorithm, which is the improved variable step size LMS forgetting factor algorithm, was verified. Through the MATLAB analysis, it was proved that it has a better performance in echo cancellation.

Key words. Adaptive filter, LMS, variable step, forgetting factor, MATLAB.

1. Introduction

In a communication system, the signal at the receiving end is usually interfered by various additive noises. As a kind of communication noise, echo appears in many parts of the communication network [1], [2]. It will affect the speech dialogue naturalness and clarity [3], and sometimes produces shrill voice [4], reduces the signal-to-noise ratio [5], and even interferes with the normal work of the system communication [6]. Therefore, a signal processing technique called echo cancellation (EC) has been developed and flourished [5], [6].

The basic principle of adaptive echo cancellation is to use an adaptive filter to identify and simulate the echo path, generate a copy of the echo, and then subtract the copy from the received signal, and then get the desired signal [7], [8]. Figure 1 shows the principle of adaptive echo cancellation system.

In Fig. 1, after the remote signal $x(n)$ sent by the caller A passes through the

[1]This work is sponsored by NSFC (6167011861, 61070234), and Jiangsu graduate student training innovation project 2016.

[2]Nanjing University of Science and Technology ZiJin College, Nanjing Jiangsu, 210046, China
[3]Nanjing University of Posts and Telecommunications, Nanjing Jiangsu, 210003, China

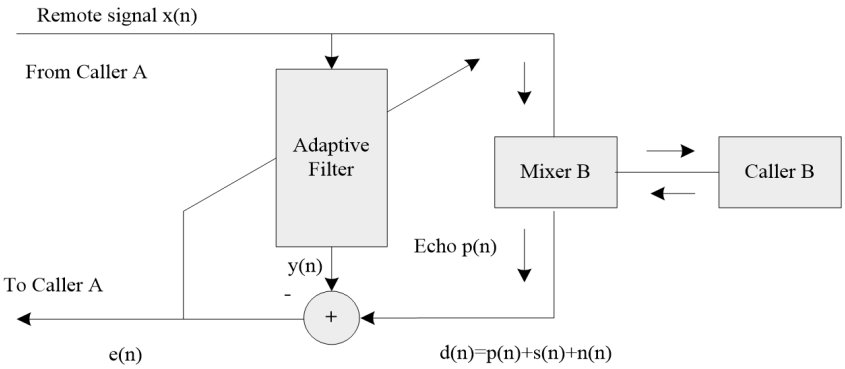


Fig. 1. Principle of adaptive echo cancellation system

adaptive filter, the false echo $y(n)$ is duplicated, and then will generate echo signal $p(n)$ after mixer B. When $p(n)$ meets the signal $s(n)$ (generated by caller B) and noise $n(n)$ (in order to simplify the analysis, sometimes it is accepted that $n = 0$), they will be superimposed to the desired signal $d(n)$, as is shown in formula (1). The signal generated by the elimination of false echo is shown in formula (2).

$$d(n) = p(n) + s(n) + n(n)\,, \tag{1}$$

$$e(n) = d(n) - y(n) = p(n) + s(n) + n(n) - y(n)\,. \tag{2}$$

In practice, when the unit impulse response of the adaptive filter can better simulate the transfer function of the echo path, it is considered that the amplitude of $y(n)$ is equal to that of $x(n)$, and the phase is opposite, so the echo $p(n)$ is basically offset.

2. 2. Analysis of the advantages of forgetting method and variable step method

The advantage of the method is to enhance the anti noise performance of the system, so that the adaptive filter on the basis of the gradient direction can maintain the correction of the direction of convergence [9], and can also restrain the tendency of convergence direction deviating from the gradient direction [10]. To some extent, it improves the drawback of LMS algorithm which has a wide dispersion and slow convergence, so the convergence speed and robustness are increased [11].

The advantage of variable step size algorithm lies in the initial stage of adaptive process [9]. At this point, the error $e(n)$ is large, but the convergence factor $\mu(n)$ is variable and large too, so the convergence rate is accelerated. When the error is gradually reduced, $\mu(n)$ is gradually reduced too, and finally a small steady-state error can be obtained [11].

3. New algorithm based on forgetting and variable step size method

Combining the advantages of the above two methods, a new minimum mean square error (LMS) algorithm is proposed.

Algorithm inputs include the weighted coefficient filter vector $\boldsymbol{w}(n)$, gradient estimation $\boldsymbol{\gamma}(n)$, variable step length convergence factor $\mu(n)$, adaptive filter input vector $\boldsymbol{x}(n)$, and expected output is $d(n)$.

Algorithm outputs include that adaptive filter output $y(n)$, altered variable step size convergence factor $\mu(n+1)$, and altered adaptive filter weight vector $\boldsymbol{w}(n+1)$.

The new algorithm steps are as follows:

1) Adaptive filtering.

$$y(n) = \boldsymbol{w}^{\mathrm{T}}(n)\boldsymbol{x}(n)\,. \tag{3}$$

2) Error estimation.

$$e(n) = d(n) - y(n)\,. \tag{4}$$

3) Gradient vector estimation of forgetting factor.

$$\boldsymbol{\gamma}(n) = \rho\boldsymbol{\gamma}(n-1) + e(n)\boldsymbol{x}(n) \tag{5}$$

4) Parameter estimation of variable step size.

$$p(n) = \beta p(n-1) + (1-\beta)e(n)e(n-1) \tag{6}$$

5) Updating of variable step size convergence factor.

$$\mu(n+1) = \alpha\mu(n) + \xi p(n) \tag{7}$$

(6) Renewal of weight coefficient vector

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \mu(n)\boldsymbol{\gamma}(n) \tag{8}$$

When $\rho = 0$, amnesia disappears, and the algorithm will degenerate into the basic variable step size LMS algorithm.

4. Simulation and performance evaluation

4.1. Simulation settings

The frequency of the remote input signal is 300 Hz, and the amplitude is 1 according to the formula

$$x(n) = \sin(600\pi n)\,. \tag{9}$$

The near-end sine input signal frequency is 1000 Hz and the amplitude is 1/6 according to the formula

$$s(n) = \frac{1}{6}\sin(2000\pi n)\,. \tag{10}$$

The echo signal frequency is 300 Hz, and the amplitude is 0.8, see the formula

$$p(n) = 0.8\sin(600\pi n)\,. \tag{11}$$

The expected signal is

$$d(n) = p(n) + s(n) = \frac{1}{6}\sin(2000\pi n) + 0.8\sin(600\pi n)\,. \tag{12}$$

Assuming the Gauss white noise is $N(\frac{10}{p+1}, 0.001)$, the order of the adaptive filter is 10 ($p = 9$), and variable step sizes are $\alpha = 0.97$, $\xi = 6 \times 10^{-3}$, and $\beta = 0.99$. The lower bound of the boundary value is $\mu_{\min} = 10^{-5}$, and the upper bound is $\mu_{\max} = 5 \times 10^{-3}$.

4.2. Simulation results analysis

The simulation results of variable step size LMS algorithm with forgetting factor are as follows:

Figure 2 displays that the convergence of variable step size LMS algorithm without forgetting factor spent 0.54 s.

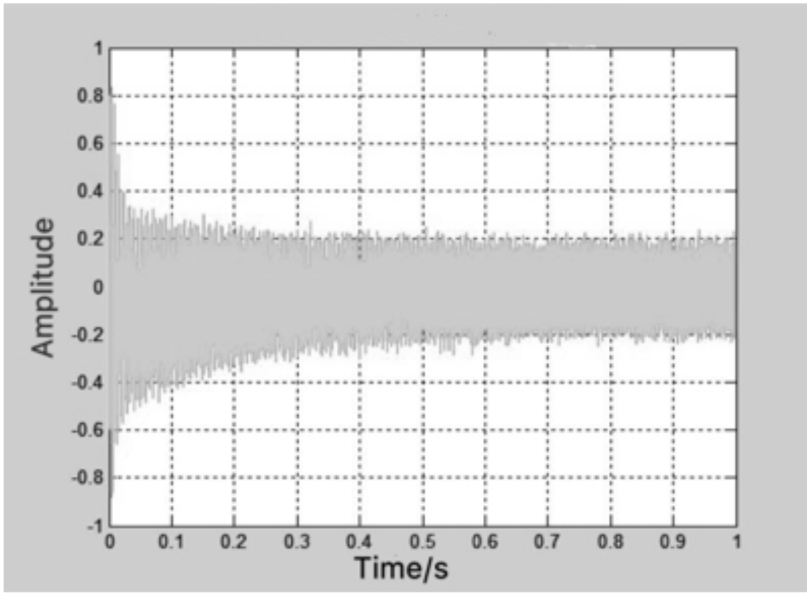


Fig. 2. Convergent results of variable step size LMS algorithm

Figure 3 shows that the convergence time of the variable step size LMS algorithm with forgetting factor 0.89 is 0.26 s, and the convergence rate is obviously faster than that of the variable step size LMS algorithm without forgetting factor. This is easier to observe in Figs. 4 and 5.

Figures 6 and 7 show that the variable convergence factor of the variable step LMS

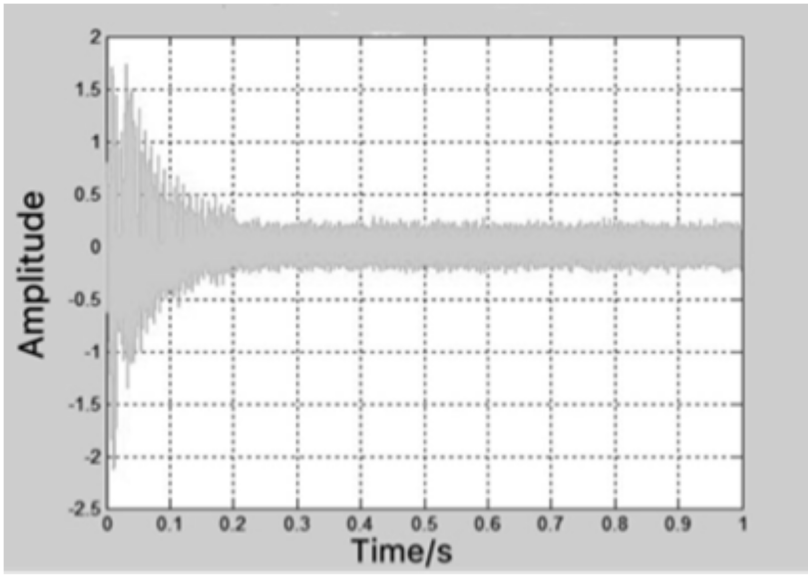


Fig. 3. Convergence time of the variable step size LMS algorithm with forgetting factor 0.89

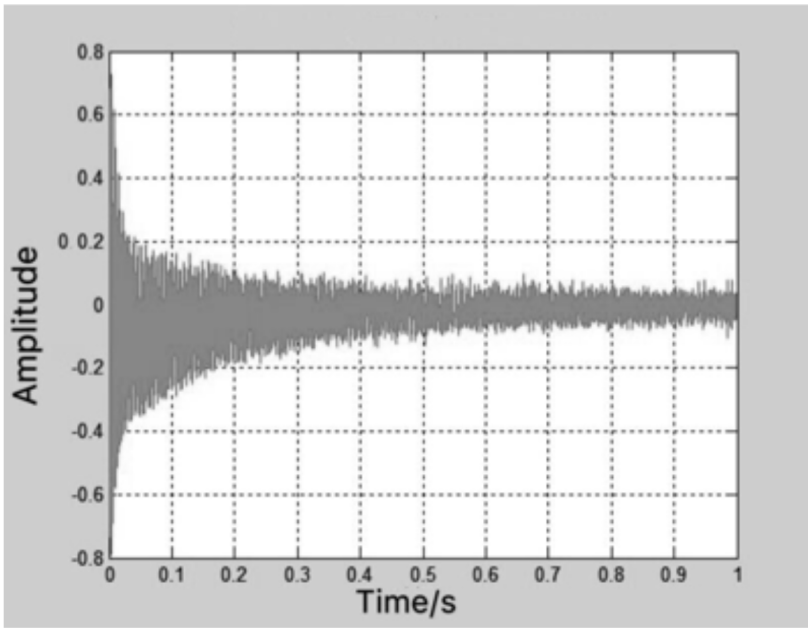


Fig. 4. Convergent process of variable step size LMS algorithm

algorithm with forgetting factor 0.89 changes from 0.005 to 0.002 and stabilized, and the maximum step size is not more than the upper value 0.005, so its steady-state

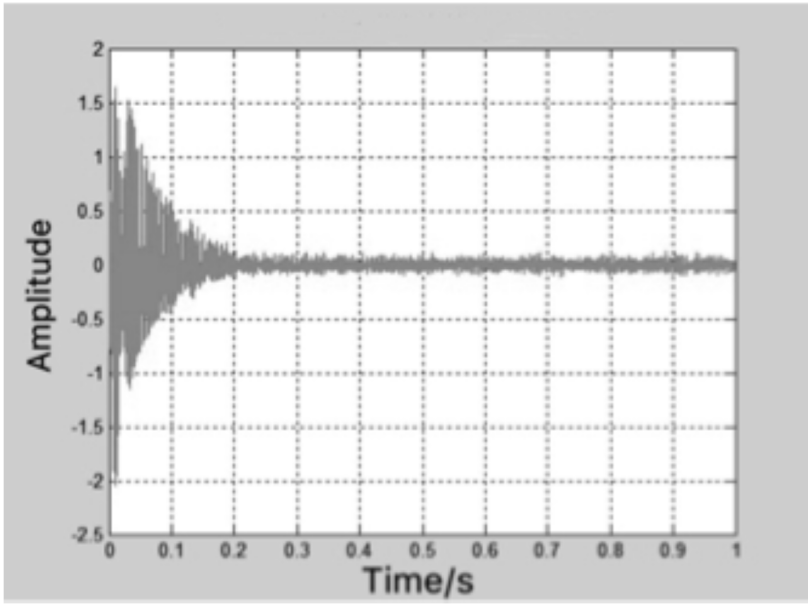


Fig. 5. Convergence process of the variable step size LMS algorithm with forgetting factor 0.89

error is smaller, and convergence speed of the variable step convergence factor $\mu(n)$ is faster than that without forgetting factor algorithm.

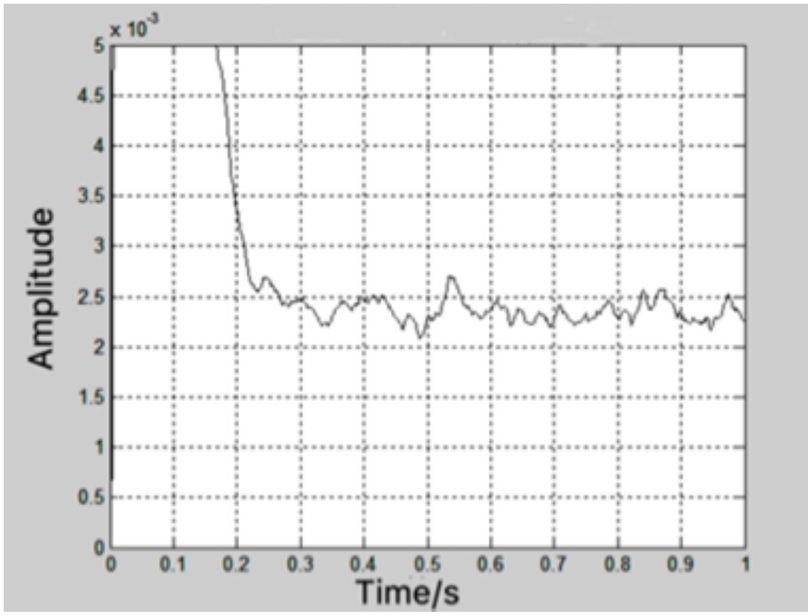


Fig. 6. Step change in the variable step size LMS algorithm

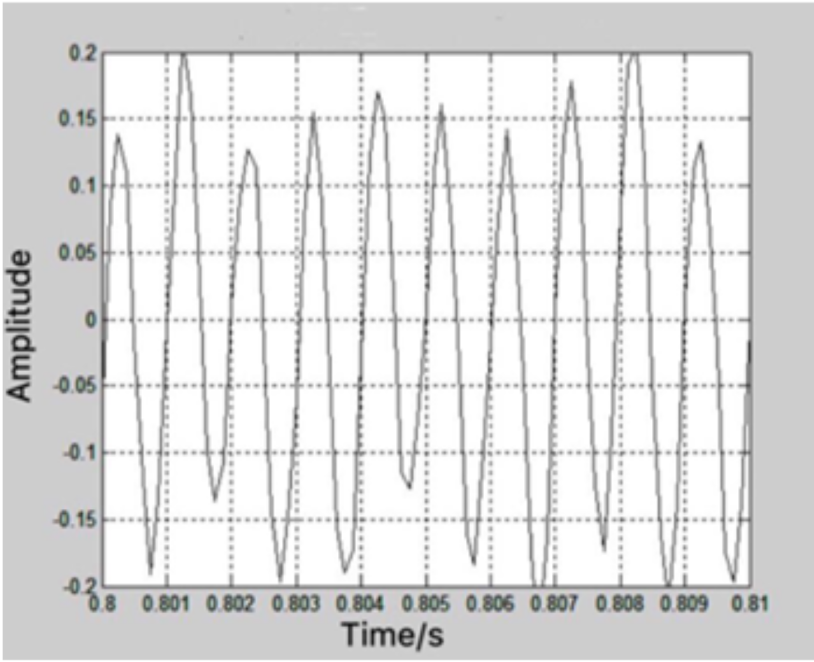


Fig. 7. Step change in the variable step size LMS algorithm with forgetting factor 0.89

Figures 8 and 9 shows that the echo cancellation effect of the forgetting factor is close to the signal from the near end, and the anti-noise ability is better than that of the variable step size algorithm without forgetting factor.

Figure 10 shows that the performance of the algorithm is worse than that of the median forgetting factor ρ_{m}, and leads to wave attenuation. Figure 11 shows that the algorithm becomes divergent when the radius of convergence is exceeded. After many experiments, the convergence radius of forgetting factor was determined, its value $\rho_{\mathrm{c}} = 0.931$.

Compared with the variable step size LMS algorithm without introducing forgetting factor, the convergence speed and noise immunity (robustness) of the variable step size LMS algorithm with forgetting factor are greatly improved.

5. Conclusion and future work

In this paper, a new adaptive echo cancellation algorithm, the improved variable step size LMS forgetting algorithm, is proposed, which is based on the two ideas of forgetting and variable step size. It can not only keep the convergence speed, the steady state error and the low algorithm complexity of the variable step size algorithm, but also absorb the anti-noise performance and fast convergence speed of the forgetting algorithm. So it is proved that this is an effective method.

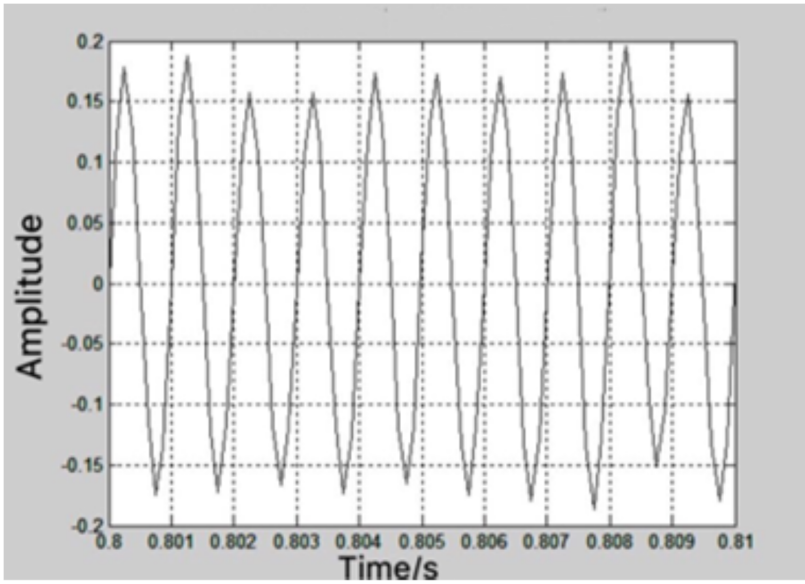


Fig. 8. Echo cancellation results of variable step size LMS algorithm

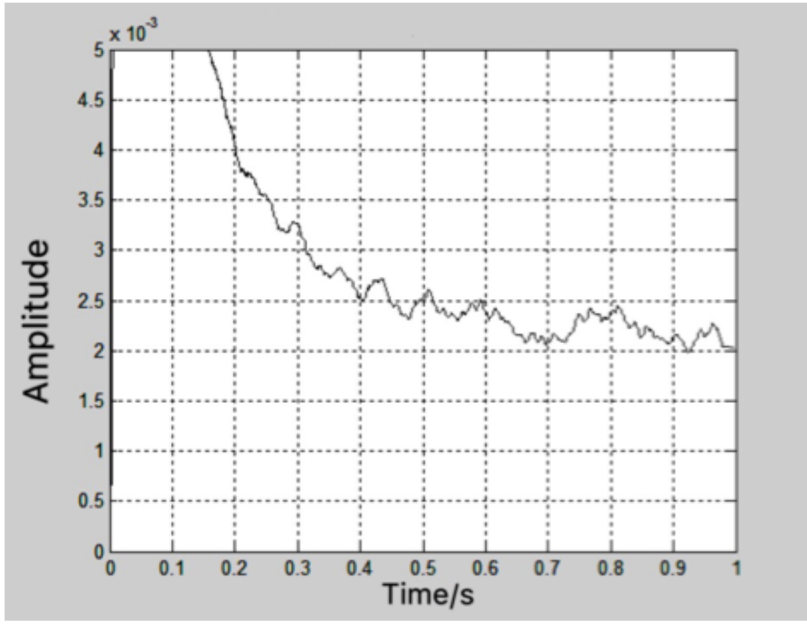


Fig. 9. Echo cancellation results of variable step size LMS algorithm with forgetting factor 0.89

References

[1] A. Kar, M. N. S. Swamy: *Tap-length optimization of adaptive filters used in stereo-*

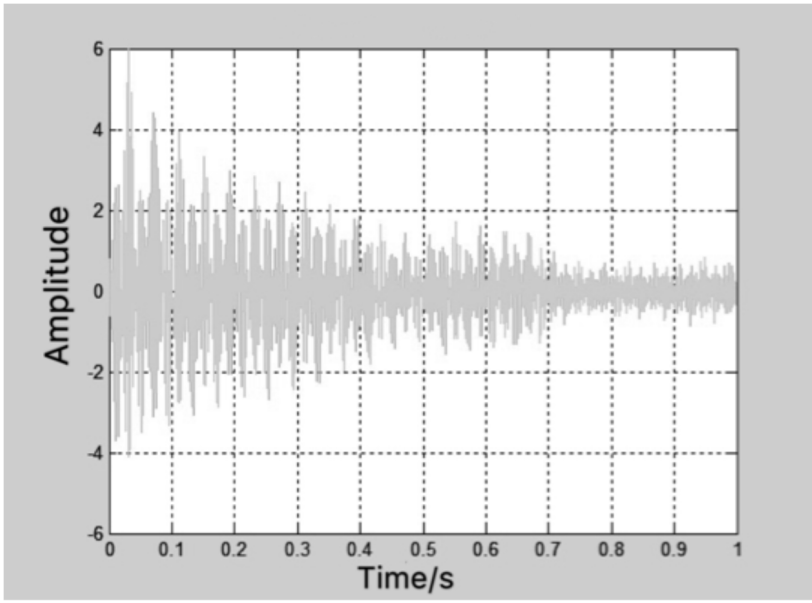


Fig. 10. Wave of forgetting factor 0.9285

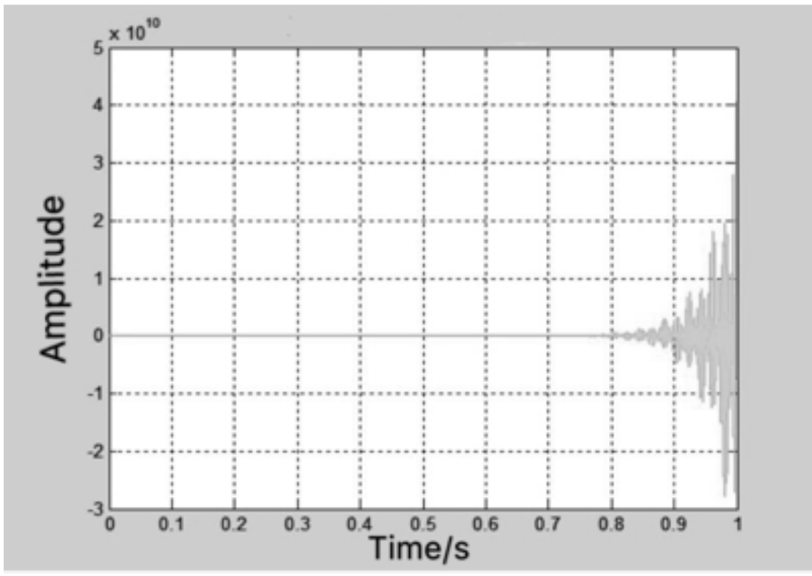


Fig. 11. Divergence of forgetting factor 0.94

phonic acoustic echo cancellation. Signal Processing *131* (2017), 422–433.

[2] D. Acharjee, N. Hasan, S. Maity: *Finite difference time domain simulation of active cancellation of radar echoes*. Procedia Computer Science *92* (2016), 233–236.

[3] M. Hamidia, A. Amrouche: *Improved variable step-size NLMS adaptive filtering algorithm for acoustic echo cancellation*. Digital Signal Processing *49* (2016), 44–55.

[4] C. Paleologu, J. Benesty, S. Ciochină: *Widely linear general Kalman filter for stereophonic acoustic echo cancellation.* Signal Processing *94* (2014), 570–575.

[5] A. Jain, S. Goel, K. Nathwani, R. M. Hegde: *Robust acoustic echo cancellation using Kalman filter in double talk scenario.* Speech Communication *70* (2015), 65–75.

[6] Y. Huang, S. Yao: *The application of adaptive filter in echo cancellation.* Instrument Technology *1* (2009), No. 8, 26–28.

[7] L. S. Gay: *An efficient, fast converging adaptive filter for network echo cancellation.* Proc. IEEE Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers (Cat. No. 98CH36284), 1–4 November 1998, Pacific Grove, CA, USA, IEEE Conference Publications *1* (1998), 394–398.

[8] L. J. Sun, B. Dai, S. Y. Zhang: *A new algorithm of echo cancellation in mobile applications.* International Conference on Information and Management Engineering (ICCIC), Innovative Computing and Information, 17–18 September 2011, Wuhan, China, Proceedings Springer Science & Business Media *232* (2011), 443–449.

[9] B. Ma, H. Dong, Y. S. Zhu: *An improved subband adaptive filter for acoustic echo cancellation application.* Procedia Engineering *15* (2011), 2244–2249.

[10] S. Cecchi, L. Romoli, P. Peretti, F. Piazza: *Low-complexity implementation of a real-time decorrelation algorithm for stereophonic acoustic echo cancellation.* Signal Processing *92* (2012), No. 11, 2668–2675.

[11] C. Stanciu, J. Benesty, C. Paleologu, T. Gänsler, S. Ciochină: *A widely linear model for stereophonic acoustic echo cancellation.* Signal Processing *93* (2013), No. 2, 511–516.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 435–444

Application of workshop production control based on double tray management

Chengwu Zheng[1]

Abstract. The purpose of the paper is to study the workshop management control based on double pallet management. Pipe production is an important part of shipbuilding. By using the improved fuzzy clustering analysis method, the pipe family is constructed. The design of the processing tray is perfect. First of all, the production management flow of tube workshop with double pallets is introduced. Then, the design of the double pallet pipe fittings is carried out. Finally, the fuzzy clustering method makes intelligent division of data, which avoids manual division. The result shows that this method can improve the production efficiency, shorten the processing period and improve the production level of the pipeline workshop. Therefore, it can be concluded that this method can meet the demand of ship manufacturing.

Key words. Double tray, fuzzy cluster analysis, tube shop.

1. Introduction

Pipe production is an important part of ship manufacture. The processing capacity occupies a large proportion in the overall workload of ship construction. The production schedule of the pipe shop will directly affect the production cycle of shipbuilding [1–2]. At present, the research on pipe workshop is becoming more and more popular [3]. Some researchers have studied the classification and fabrication of outfitting pieces in modern shipbuilding modes. Throughout the shipbuilding process, the promotion of group technology and pallet management applications should be taken seriously [4]. Wu Di et al. used the "COM-object" method to extract the underlying data resources of TRIBON, and developed the ship outfitting pallet data management system based on TRIBON database [5]. Through the ship outfitting pallet data mining method based on multi Agent technology, Chen Ning et al. extracted the basic information from the TRIBOB, in order to facilitate the comparison and processing of relevant data [6]. Aiming at the self-made outfitting and outsourcing outfitting in shipbuilding, Deng Shuo designed an algorithm for optimizing the

[1]Yangtze Normal University, Chongqing, 408100, China

pallet allocation scheme. In addition, for the pallet distribution problem, he also proposed an optimization algorithm [7]. Xu Hanchuan and others proposed a tray optimization set management model. Aiming at the problem of multi-route selection in the process of fitting outfitting, an optimization strategy is proposed [8]. It can be seen that many scholars have done a lot of comparative research on workshop operation control management. However, for the characteristics of multi-species small batch of pipe workshop, the research on the workshop production control system is rare. How to make full use of group technology in the modern shipbuilding model for the construction of pipe family? How to give full play to the advantages of tray management model? It is worth further analysis and research to solve the problems of low production efficiency and serious tardiness in the pipe workshop.

2. Materials and methods

2.1. Double tray management mode

The double tray management mode of the ship pipe is the production management mode that combined with the inner processing pallet and the field installation tray [9–10]. Through the use of group technology, the inner processing pallet is the basic unit to achieve the pipe group processing. Outside the field tray as the basic unit, it implements the pipe unit outfitting and segmented outfitting. The double tray management mode of the pipe shop is shown in Fig. 1.

According to the modern shipbuilding mode, the field installation tray divides the ship into different areas, systems and stages. The intermediate product is oriented. The pipe is carried out with unit outfitting and segmented outfitting. Typically, in accordance with the location of the pipe, the order of installation or function, it determines the same area of the pipe combination [11–12]. The corresponding pallet table should include a list of each pipe unit combination, pipe installation drawings and so on. This form has been adopted by the vast majority of shipyards. According to the pipe characteristics and processing technology, the inner processing pallet is the product of making full use of group technology [13]. It is the basic unit of the pipe workshop processing, and is the "field installation tray" sorting set after the pipe family. The corresponding tray table includes the list of the pipe family, the tube and so on [14]. The processing pallet plan is mainly used to determine the order of processing of each pipe family, as well as the start and end time [15].

In the double tray management mode, the cross-regional installation trains with close delivery times are combined into the tray group. According to the similarity of pipe material, shape, working procedure and process, and considering the production equilibrium principle and other comprehensive factors, sorting and clustering are carried out to form the processing tray of the pipe workshop [16–17]. The processing tray is the turnover box of pipe workshop, and it is the basic unit of pipe workshop production management. The allocation of personnel, materials and accessories supporting, the settlement of the work hour quota, the control of the production process is all around it to operate. The workshop in the pipe manufacturing is completed. According to the assembly rules of the pipe, the clusters are sorted

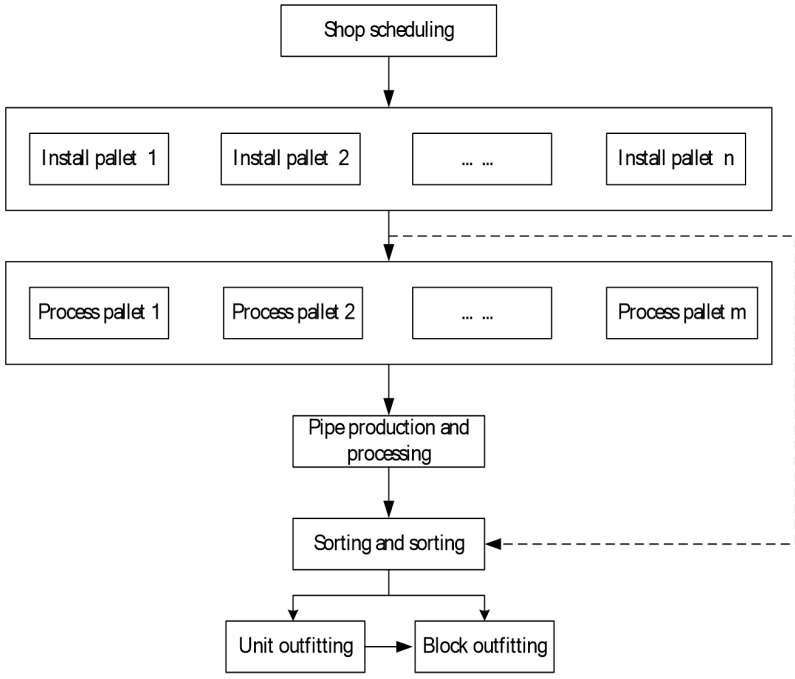


Fig. 1. Management mode of double tray in pipe workshop

again. It is reclassified as a mounting tray, so that subsequent outfitting operations can be carried out efficiently.

The core of the double pallet management mode is the manufacture of the pipe fittings. It takes the processing tray as the basic unit. The unit assembly of the pipe is assembled with the pallet as the basic unit. Its key lies in the processing tray and the installation tray determination, as well as the tray plan accuracy.

2.2. Production management flow of tube workshop controlled by double pallets

This paper will describe the production process of ship pipe workshop based on double pallet management. It provides a further description of the links and logical relationships among key links in the system. The production flow of the pipe workshop controlled by double pallets is shown in Fig. 2.

In order to control the production of ship pipe workshop based on double pallet management, it is necessary to change the production organization form of pipe workshop. In view of the phenomenon that the workshop is received by pallet before the pipe shop is installed, the production organization form of workshop is changed into a double pallet management model with pallet and tray. In the double tray management mode, the cross-regional installation trains with close delivery times are combined into the tray group. According to the similarity of pipe material, size, shape, working procedure and process, the fuzzy clustering algorithm based on part

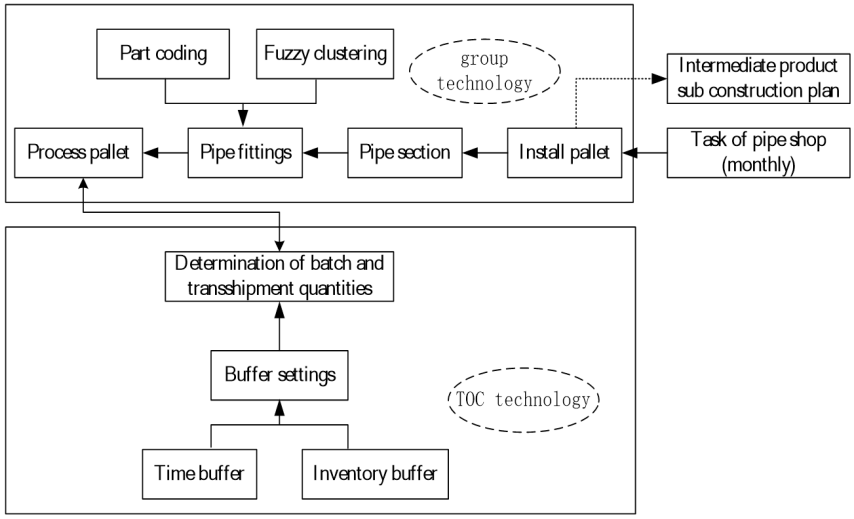


Fig. 2. Production management flow of tube workshop controlled by double pallets

coding is used to sort and cluster and generate pipe fittings. Then, according to batch management, the capacity of the processing tray is determined. According to the regulations, each pipe family is split into a processing tray in the pipe workshop. The processing tray is used as the turnover box of the workshop in the workshop. It is the basic unit of the workshop production management. The determination of processing pallet is the basis of the production control system of ship pipe workshop based on double pallet management.

In the production control of the pipe workshop, this paper mainly uses the TOC (Theory of Constriction) idea and adopts the DBR operation mechanism. According to the principle of DBR mechanism, the primary task is to identify bottleneck resources, which is an important prerequisite to ensure the smooth implementation of DBR mechanism. The identification of the bottleneck can be determined by the capacity of the equipment, the load situation and the number of products in question. Secondly, the production schedule of bottleneck resources is planned, and the processing pallet is the basic dispatching unit. It not only makes full use of group technology, but also facilitates workshop production control. Through the rational formulation of the tray sorting on the bottleneck resources, the operation plan is formulated, thus determining the "drum" of DBR, so as to maximize the role of the bottleneck. Finally, the performance of the whole system is improved. Then, through buffer and batch management, reasonable buffer is set in reasonable place, and the bottleneck resource is maintained high utilization rate without the influence of upstream production fluctuation. In the absence of "hunger" status, reasonable processing batch and transfer batch are set. Thus, the inventory can be kept at a reasonable level while increasing production efficiency.

3. Results

3.1. Fuzzy clustering analysis of parts coding

Set all the parts that need to be classified as: $X = \{x_1, x_2, ..., x_n\}$. Among them, the x_i represents the i-th part in the collection of all parts that need to be classified.

If each part x_i has the attribute m (code bits), then each part's encoding can be represented as $X_i = \{x_{i1}, x_{i2}, ..., x_{ij}, ..., x_{im}\}$ $(i = 1, 2, ..., n;\ j = 1, 2, ..., m;\ m \geq 2)$. Among them, x_{ij} represents the attribute value (code value) of the jth bit attribute object (code bit) of the first i part.

A code bit represents an attribute object of a part. The code value of this bit indicates the information about the property object of the part. The part information is a collection of symbols used to describe a property object (such as a material, shape, etc.) associated with a part. The parts family is divided by using the part properties. First, for part grouping purposes (product design or product manufacturing), part attribute objects are selected. This article is mainly for the parts of the manufacturing process. The attributes mainly include material, size, shape, process characteristics and so on.

To this end, the part group to be classified can be represented by the part coding matrix R_0. Each row of the matrix R_0 represents the attribute value of each attribute of the same part. Each column represents the attribute values of the different parts under the same attribute object.

$$R_0 = [x_{ij}]_{n\times n} = \begin{bmatrix} x_{11} & x_{12} & ... & x_{1m} \\ x_{21} & x_{22} & ... & x_{2m} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & ... & x_{nm}. \end{bmatrix}$$

Fuzzy clustering analysis is usually carried out using differential transformation for data standardization. The worst conversion formula is as follows:

$$x_{ij} = \frac{x_{\mathrm{ij}} - x_{\min}}{x_{\max} - x_{\min}}$$

. In the formula, x_{ij} represents the code value of the jth code bit of the part x_i. Symbols $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of the code values on the same code bit for different parts. Data is normalized to compress the data to the [0,1] interval to reduce the effect of the data dimension on the comparison.

For the similarity matrix construction of part coding, $[r_{ij}]_{n\times n}$ is used to represent the degree of similarity between the ith part and the jth part. The closer the r_{ij} is to 1, the higher the similarity between the two parts. r_{ij} is closer to 0, indicating that the similarity is lower. There are many ways to express r_{ij}. According to the standard Hamming distance method, r_{ij} can be expressed as

$$r_{ij} = 1 - c\sum_{k=1}^{m} |x_{ik} - x_{jk}| .$$

Among them, $\frac{1}{c} = \max\{\sum_{k=1}^{m} |x_{ik} - x_{jk}|, \forall i \neq j\}$, so as to ensure $r_{ij} \in [0\ 1]$.

The fuzzy similarity matrix R has reflexivity and symmetry. However, it usually does not have transitivity, and must be transformed into fuzzy equivalent matrix. Transitive closure is often used. The calculation process is as follows:

$$R \bullet R = R^2$$
$$R^2 \bullet R^2 = R^4$$
$$R^4 \bullet R^4 = R^8$$
$$\ldots\ldots$$

When $R^k \bullet R^k = R^{2k} = R^k$ appears, it means that R is transitive. Therefore, the fuzzy similarity matrix $R = R^k = R^{2k}$ is obtained.

The appropriate threshold (also called confidence level) $\lambda \in [0\ 1]$ is selected. The fuzzy equivalent matrix R' is cut. As the threshold changes, the result will change. With the reduction of the threshold, the number of packets will decrease, and the number of parts in the same group will increase, while the similarity between the components will decrease. It can be seen that expanding the number of parts in each part family is at the expense of reducing the similarity of the parts within the family. The threshold must be chosen and adjusted according to the actual requirements, and the real purpose of the group will be achieved by dynamically grouping the parts.

3.2. Design of fittings for double pallets

At present, the operation of the pipe plant organization is based on the installation of the basic unit for the tray. Therefore, the code is installed for the pipe, and it is unique. Figure 3 is the current pipe workshop common parts coding structure. It is mainly composed of five parts: Ship Engineering number, regional subsection number, pallet number, piping system number and pipe fittings number.

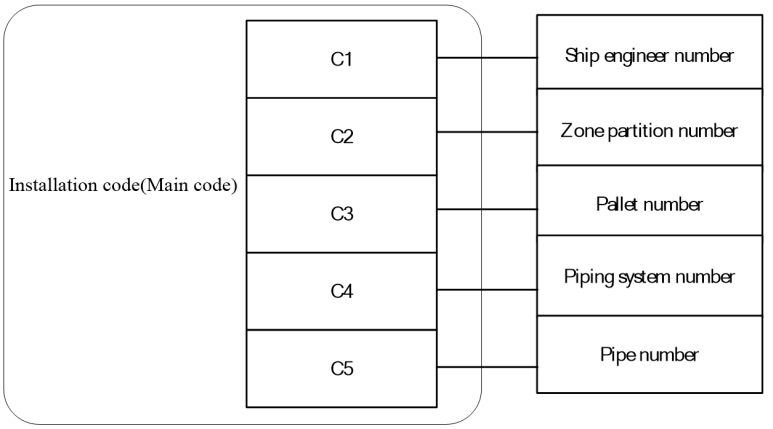


Fig. 3. Tube-oriented structure for the installation

According to the pipe processing status, and considering the processing charac-

teristics of pipe fittings, the attributes of the pipe can be divided into two grades by AHP, as shown in Fig. 4. The application of group technology in pipe workshop is mainly to improve the utilization ratio of material and reduce the switching time of bending process. It should avoid piping backflow due to process differences. It takes full advantage of the professionalism of employees in the process of processing and welding of similar pipe fittings. Therefore, pipe material, pipe specifications, elbow conditions, assembly processing, assembly type and connection type are selected as attribute objects. The effect of group technology on surface treatment processes is not considered here because some pipe shops do not have this process.

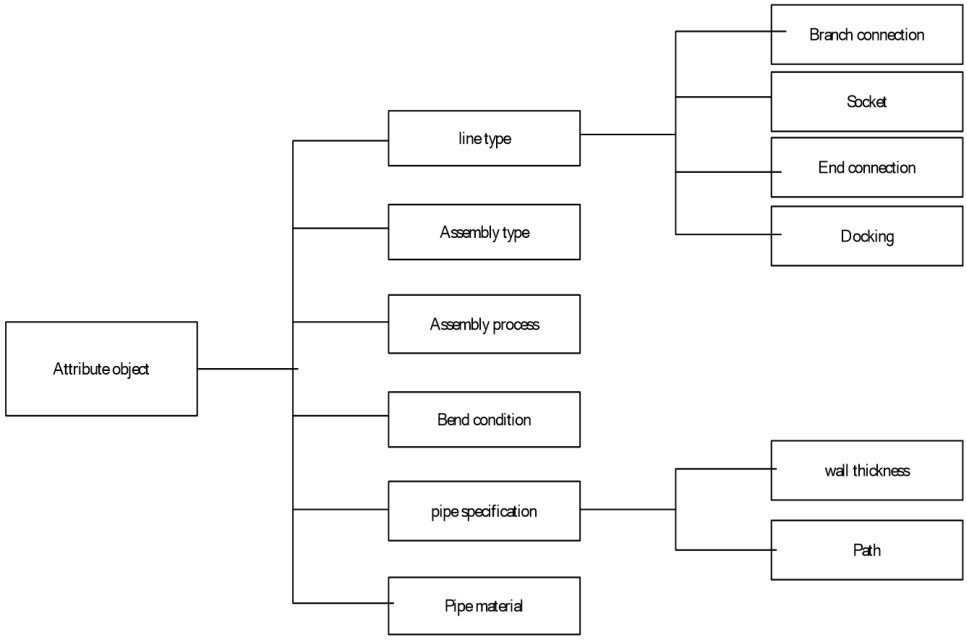


Fig. 4. Design of pipe attribute object for manufacturing

The processing coding structure of the pipe can be determined according to the coding attribute object of the pipe. Combined with double tray management ideas, the pipe can be installed as the main code. The processing codes are used as auxiliary codes, and the combination of the two codes form a new pipe coding structure. Among them, as the main code of the installation code is unique, which is mainly to facilitate the installation of fittings. However, as the auxiliary code, the processing coding is not unique. It is mainly applied to the formation of processing trays, easy to pipe processing, in order to achieve group technology in the pipe workshop applications. When building a pipe and processing pallet for a machining process, it only requires fuzzy clustering analysis of the machining code. The tube coding structure based on double tray management is shown in Fig. 5.

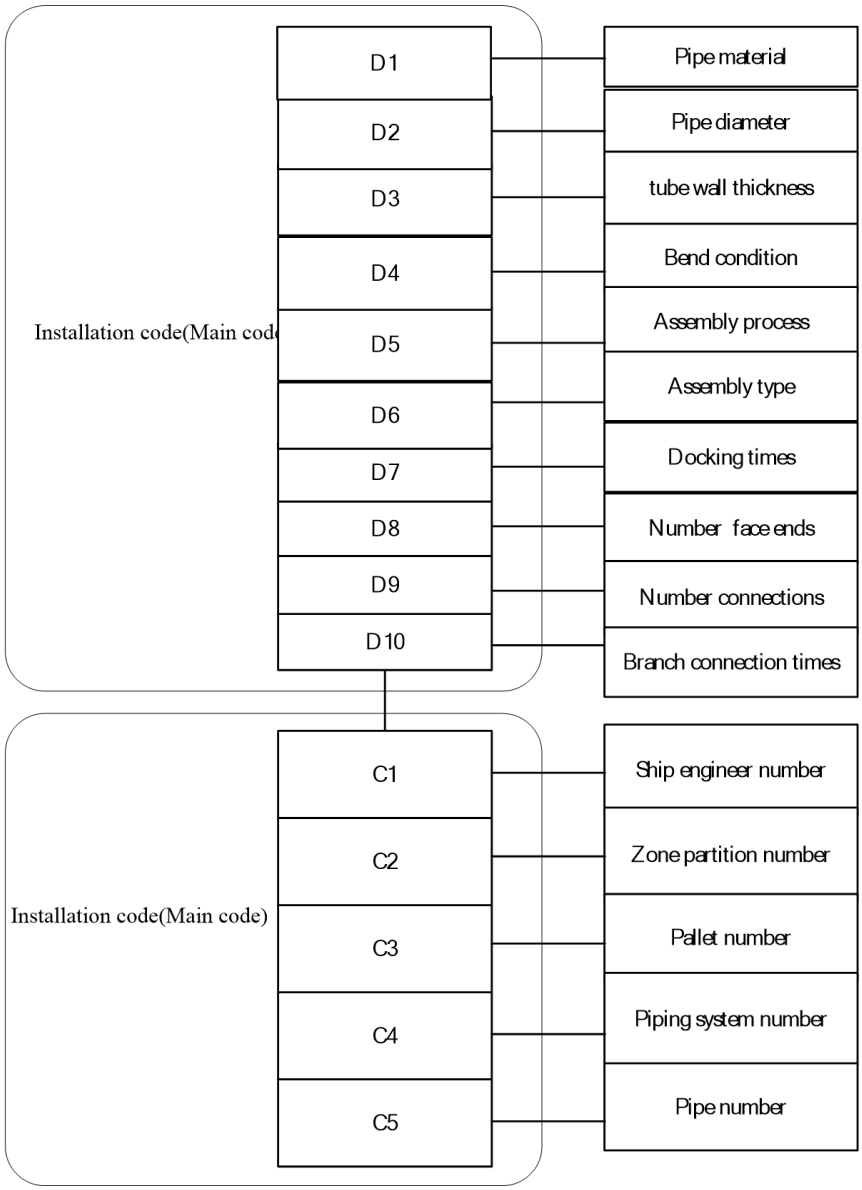


Fig. 5. Pipe coding structure based on double pallet management

4. Conclusion

The shipbuilding industry of our country has been developing rapidly for several decades. Shipbuilding output continues to grow, and the market share in the world continues to expand. In 2013, China's shipbuilding capacity has been ranked first in

the world. At the same time, the international competition of shipping enterprises is more and more fierce. In order to meet the market demand, ship enterprises have been carrying out technical and management reforms. The vessel production is an important part of ship manufacture. Its processing capacity occupies a large proportion in the overall workload of shipbuilding. The production schedule of the pipe shop will directly affect the production cycle of shipbuilding. At present, the tray management of the vessel pipe workshop is not perfect. The production efficiency is low and the processing cycle is long. These problems affect the overall progress and economic benefits of ship enterprises. In this paper, a double pallet management model for pipe workshop is introduced. A pipe coding system based on double pallet management is designed. By using improved fuzzy clustering analysis based on coding, the pipe family is constructed. Finally, the design of processing trays has been improved.

References

[1] C. Pizarro, I. Esteban-Díez, S. Rodríguez-Tecedor, J. M. González-Sáiz: *A sensory approach for the monitoring of accelerated red wine aging processes using multiblock methods.* Food Quality and Preference *28* (2013), No. 2, 519–530.

[2] E. Bedalli, E. Mançellari, O. Asilkan: *A heterogeneous cluster ensemble model for improving the stability of fuzzy cluster analysis.* Procedia Computer Science *102* (2016), 129–136.

[3] M. H. Mello, J. Gosling, M. M. Naim, J. O. Strandhagen, P. O. Brett: *Improving coordination in an engineer-to-order supply chain using a soft systems approach.* Production Planning & Control *28* (2017), No. 2, 89–107.

[4] J. Molka-Danielsen, P. Engelseth, B. T. N. Le: *Vendor-managed inventory as data interchange strategy in the networked collaboration of a Vietnam ship parts supplier and its customers.* Information Technology for Development (2017), 1–21.

[5] T. K. Chapple, A. C. Gleiss, O. J. D. Jewell, M. Wikelski, B. A. Block: *Tracking sharks without teeth: a non-invasive rigid tag attachment for large predatory sharks.* Animal Biotelemetry (2015), No. 3, paper 14.

[6] C. Zambra, J. Romero, L. Pino, A. Saavedra, J. Sanchez: *Concentration of cranberry juice by osmotic distillation process.* Journal of Food Engineering *144* (2015), 58–65.

[7] R. Bharath, V. V. Srinivas: *Delineation of homogeneous hydrometeorological regions using wavelet-based global fuzzy cluster analysis.* International Journal of Climatology *35* (2015), No. 15, 4707–4727.

[8] S. Kaviani, A. M. Hassanli, M. Homayounfar: *Optimal crop water allocation based on constraint-state method and nonnormal stochastic variable.* Water Resources Management *29* (2015), No. 4, 1003–1018.

[9] E. Lopez-Valeiras, M. B. Gonzalez-Sanchez, J. Gomez-Conde: *The effects of the interactive use of management control systems on process and organizational innovation.* Review of Managerial Science *10* (2016), No. 3, 487–510.

[10] P. Mishra, V. Kumar, K. P. S. Rana: *A fractional order fuzzy PID controller for binary distillation column control.* Expert Systems with Applications *42* (2015), No. 22, 8533–8549.

[11] A. R. Gholami, M. Shahbazian: *Soft sensor design based on fuzzy C-Means and RFN-SVR for a stripper column.* Journal of Natural Gas Science and Engineering *25* (2015), 23–29.

[12] A. Pollice, S. Arima, G. J. Lasinio, A. Basset, I. Rosati: *Bayesian analysis of three indices for lagoons ecological status evaluation.* Stochastic Environmental Research and Risk Assessment *29* (2015) No. 2, 477–485.

[13] S. C. Swain, S. Panda, S. Mahapatra: *A multi-criteria optimization technique for SSSC based power oscillation damping controller design.* Ain Shams Engineering Journal *7* (2016) No. 2, 553–565.

[14] S. Donadi, B. K. Eriksson, K. A. Lettmann, D. Hodapp, J. O. Wolff, H. Hillebrand: *The body-size structure of macrobenthos changes predictably along gradients of hydrodynamic stress and organic enrichment.* Marine Biology *162* (2015) No. 3, 675–685.

[15] K. Almohammadi, H. Hagras, D. Alghazzawi, G. Aldabbagh: *Users-centric adaptive learning system based on interval type-2 fuzzy logic for massively crowded E-learning platforms.* Journal of Artificial Intelligence and Soft Computing Research *6* (2016), No. 2, 81–101.

[16] J. Mallick: *Geospatial-based soil variability and hydrological zones of Abha semi-arid mountainous watershed, Saudi Arabia.* Arabian Journal of Geosciences *9* (2016), No. 4, paper 281.

[17] M. E. Hammad, M. I. Dessouky, O. Zahran, H. Kasban, S. M. S. Elaraby, F. E. A. El-Samie.: *Efficient signal processing techniques for distillation column malfunctions identification.* Journal of Nondestructive Evaluation *34* (2016), No. 4, paper 33.

Received July 12, 2017

Acta Technica *62* No. 2A/2017, 445–454

A priority based dynamic bandwidth scheduling in SDN networks[1]

Zun Wang[2]

Abstract. In order to solve the problems of effective storage and the large scale data processing, the data centers running large-scale distributed computing are being constructed all over the world. Hadoop, a distributed computing framework that serves as the core infrastructure of data centers, is proposed. For reducing the time cost brought about by the job implementation in the data migration process, the job scheduling algorithm that plays an important role in the Hadoop assigns the task to task required node data for the implementation as much as possible, so as to shorten the operation response time and improve the performance of the cluster. Through the use of SDN (Software Defined Network) to control the network flexibility, the network control is conducted for the data transfer, so as to avoid the influence of task response time caused by network load change. The experimental results showed that the BS-IDS algorithm can adapt to the dynamic load change of data center effectively, and it has better performance in job response time compared with the traditional delay scheduling algorithms.

Key words. Software defined network, scheduling algorithm, Hadoop.

1. Introduction

In recent years, the Internet has brought great changes to people's daily life, and a large number of applications are being generated every day in the Internet, which has led to the rapid development of Internet [1]. As the most widely used cloud computing and large data processing platform, Hadoop has obtained the world's attention. But in many ways, it still exist the space to improve and enhance the performance, and one crucial issue is the job scheduling [2]. As one of the core technologies of Hadoop, Hadoop job scheduling algorithm is mainly responsible for decision-making for scheduling which job and a suitable task in the job to free computing nodes so that the scheduling is reasonable and the cluster is optimal. However, for the continuously improved job scheduling algorithm, when faced with increasing data and a variety of application scenarios, there are still some shortcomings. As one

[1]The author acknowledges the National Natural Science Foundation of China (Grant No. 51578109) and the National Natural Science Foundation of China (Grant No. 51121005).

[2]Beijing University of Posts and Telecommunications, 100876, China

of the most widely used Hadoop scheduling algorithms, the basis for FIFO (first in first out) to choose the job is based on the time sequence of operations arrived. Its advantages are simple and easy to implement, and the disadvantage is that it treats all the jobs the same, and ignores the urgency of the operation, which may result in that small operations are in the waiting state for a long time that they cannot be scheduled for implementation. In the data center, because the network bandwidth is a scarce resource, under the premise of ensuring the fairness of the operation, in order to improve the probability of task implementation on the local node, researchers at the University of California at Berkeley, based on Fair Scheduler, studied the corresponding delay scheduling algorithm. However, the algorithm only uses a static value as the waiting time threshold for the team's first operation, so it cannot effectively adapt to the dynamic load changes and network changes of the cluster. On this basis, a dynamic waiting time threshold based on network bandwidth and data center load variation is set as the waiting time of the team's first operation [3]. Nevertheless, these algorithms do not take into account the completion time of first team job from a global perspective, ignoring the effects of network load change on the data transfer. As a result, the performance of the algorithms greatly reduced to a certain extent. Therefore, it has important significance to make a detailed analysis of various scheduling algorithms at present, to sum up the corresponding advantages and disadvantages, and through a combination of new technology, and to correct the corresponding defects, essential for reducing the job response time of Hadoop systems and improving the Hadoop cluster performance and so on.

2. Improved delay scheduling algorithm based on SDN

2.1. BS-IDS algorithm

The jobs that users submitted are arranged into multiple job queues, and each queue has a team first job. It is assumed that the selected jobs for scheduling have the remaining m tasks that are not performed, representing $T_1, T_2, ..., T_m$, respectively, then the size of the data block that the task corresponds to is Δ. The data center has n nodes, and they are denoted as $Node_1, Node_2, ..., Node_n$, respectively. The SDN controller periodically transfers the data link bandwidth of the data center to the Master nodes of the Hadoop system. At some points, a free node $Node_{\text{free}}$ requests a task assignment to the master node, so as to ensure the data locality in the task implementation and reduce the response time of the task. The strategy that the BS-IDS scheduling algorithm uses is: if the free node is not the local node of the team first work, the waiting time threshold $t_{\text{threshold}}$ in the team first job is calculated by using the waiting time threshold algorithm, and compared with the waited time t_{wait} respectively. The quantities $t_{\text{threshold}}$ and t_{wait} represent the calculated team's first job waiting time threshold and the waited time, respectively. If the waiting scheduling time of the team's first job does not exceed the waiting time threshold for the team first job, other jobs are scheduled firstly. The real-time available bandwidths of each uncompleted task required data blocks storage node to the free node in the team's first job is achieved by SDN, which are $B_1, B_2, ..., B_m$,

respectively [4]. Symbol B_1 represents the real-time bandwidth available to the free node $Node_{\text{free}}$ from the node stored by the data blocks required by the task T_1. Assuming that each teak in the team first job is scheduled to the current free node work $Node_{\text{free}}$, the data block transfer time overheads caused by the implementation of the task of non local node are $\frac{\Delta}{B_1}, ..., \frac{\Delta}{B_m}$, respectively. From these data block transfer time overhead, the task T_i that a minimum time overhead corresponds to is chosen, and it immediately schedules the task T_i. The calculation process is as follows:

$$T_i = \min\left(\frac{\Delta}{B_j}\right),\ 1 \le j \le m\,. \tag{1}$$

In this case, since the team's first task is scheduled to execute on a non local node, it is necessary to move the data block from the node where the task needs data storage to the computing node. As we all know, network resource is a scarce resource of data center. In order to avoid the effects of network load change on the data blocks transfer, we use SDN bandwidth control capability, to design a bandwidth allocation mechanism based on time section. And it assigns the bandwidth for the data blocks transfer, so as to ensure the efficient allocation of tasks, and improve the performance of BS-IDS scheduling algorithm.

2.2. Bandwidth allocation mechanism based on time section

The delay scheduling algorithm is based on the real-time available bandwidth between the current time nodes to calculate the waiting time threshold. And the network bandwidth resources of data center are the scarce resources, so between the various applications, they will compete for the network bandwidth. In order to ensure the performance of the delay scheduling algorithm, we introduce a time based bandwidth allocation scheme. The main idea is as follows: at the present time $T = t_0$, when there is a free node $Node_{\text{free}}$ requesting for task allocation, if the waited time of the team first job exceeds the waiting time threshold, the task T_i that a minimum data block migration time overhead corresponds to is selected to schedule to the node $Node_{\text{free}}$ for the implementation.

When a task T_i is assigned to a free node $Node_{\text{free}}$ for the implementation, it is assumed that the link Link_i is scheduled by the task T_i for implementation on a nonocal node, the data block that the task T_i corresponds to moves the path has passed over [5]. The Hadoop scheduler calls the interface of the SDN controller, leaving the link Link_i reserved for the task T_i in the time period of $\left(t_0, t_0 + \frac{\Delta}{B_i}\right)$, and the reserved bandwidth size is B_i.

When the data required by the task T_i has completed the migration from the storage node to the free node $Node_{\text{free}}$, the OpenFlow controller will terminate the occupying of the link Link_i.

With the help of bandwidth allocation mechanism based on time section, we can maximize the use of network bandwidth capability, to reduce the impacts of network load change on data transfer, which efficiently makes task scheduling and improves the performance of BS-IDS algorithm.

3. Method

As shown in Fig. 1, the BS-IDS scheduler is mainly divided into task scheduling module, job initialization module, job queue management module, team first job information pool module and waiting time threshold calculation module. Among them, the BS-IDS scheduler inherits and implements the team first modules interfaces of the Hadoop system. In addition, the BS-IDS scheduler designed the first operation information pool and the waiting time threshold calculation module based on the BS-IDS scheduling algorithm.

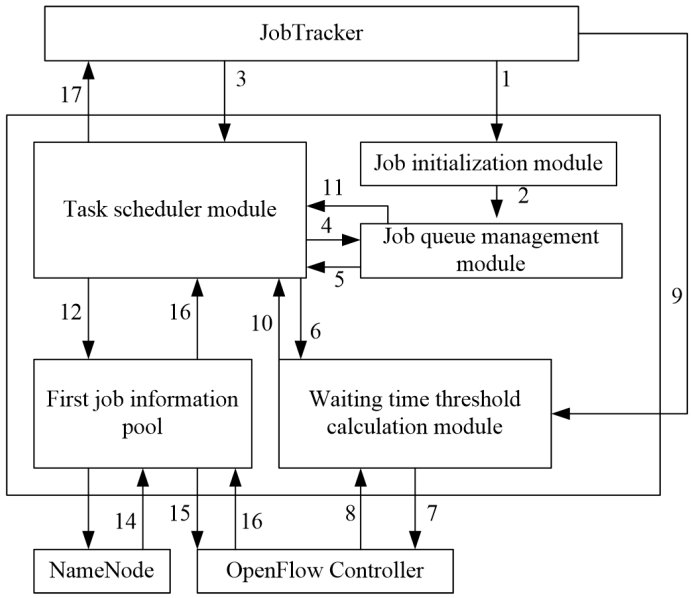


Fig. 1. BS-IDS scheduler module and scheduling flow chart

BS-IDS scheduler flow is shown in Figure 1, and its major process can be divided into the following steps:

Submission and initialization of the job, as shown in step 1 in the figure, when the user submits the job from the client, the BS-IDS scheduler first of all uses the JobTracker.initJob function to initialize it. As shown in step 2 in Fig. 1, a Map task is created for each data block based on the partitioning of the data blocks of the job file. The queue job management module adds successful initialization jobs to the appropriate queue and waits for the implementation.

Request and processing of tasks, as shown in step 3, 4, and 5 in the figure, when the data center has free nodes requesting the task allocation to JobTracker, JobTracker will call the job management and scheduling module to select a job to be scheduled. Secondly, as shown in step 6, 7, 8, 9, and 10, the task scheduling module will give a request to task scheduling module in the BS-IDS scheduler and the waiting time threshold calculation module. And then the calculated team's first job waiting time threshold is given feedback to the task scheduling module. The

task scheduling module team's first job is compared with the waited time and the waiting time threshold. As shown in step 11, if the waited time is less than the waiting time threshold, it first of all schedules other operations; otherwise, perform the following operations. Since the waited time of team's first job exceeds the waiting time threshold, a task in the team's first job needs to be allocated to the current free node. At this point, as shown in steps 12, 13, 14, and 15 in the figure, the task scheduler module sends the task assignment request to the team's job information pool [6]. The team's first job information pool, from the NameNode and OpenFlow controller in the HDFS system, obtained the data blocks backup information of team's first job and network load information of data center. And through calculation, the task is assigned to free nodes to perform, and the task that the data block transfer time overhead between the required block data storage node and current free node corresponds to.

As shown in step 16, the first team job information pool will deliver the task information for the task scheduling module, and allocate the task to the free node. As shown in step 17, the data block that the task corresponds to would be moved to free nodes from the storage nodes. And the MAC address of storage node and free node of the task required block and the transfer time section is transferred to the OpenFlow controller. Then the OpenFlow controller will reserve the data transfer link in the migration period for the task. As shown in step 16 in Fig. 1, the task scheduler module assigns the task to the current free node for the implementation.

4. Results and discussion

The experimental environment is built based on the Hadoop framework of SDN, and its effectiveness is verified. The network environment it builds is shown in Fig. 2, mainly including 6 nodes, two OpenFlow switches, and a router. The 4 nodes are taken as Slave, which is responsible for storing data and performing tasks, and the other two nodes are Master and OpenFlow controllers, respectively.

Relevant allocation parameters of Hadoop system are shown in Table 1.

Table 1. Hadoop system parameters allocation table

Link bandwidth	100 Mbps–1000 Mbps
Data block size	64 M
Data block backup	2

In order to analyze the influence of different scale data on BS-IDS scheduling algorithm, this experiment, according to the number of tasks, divided the jobs into five groups, as shown in Table 2. This experiment used the Wordcount and Sort system of Hadoop system as the test operations. The Hadoop default scheduler FIFO, the delaying scheduler, and the BS-IDS scheduler are tested, respectively, and the test results are analyzed.

Local probability of job data:

As shown in Fig. 3, in the 100 Mbps network environment, the network transmis-

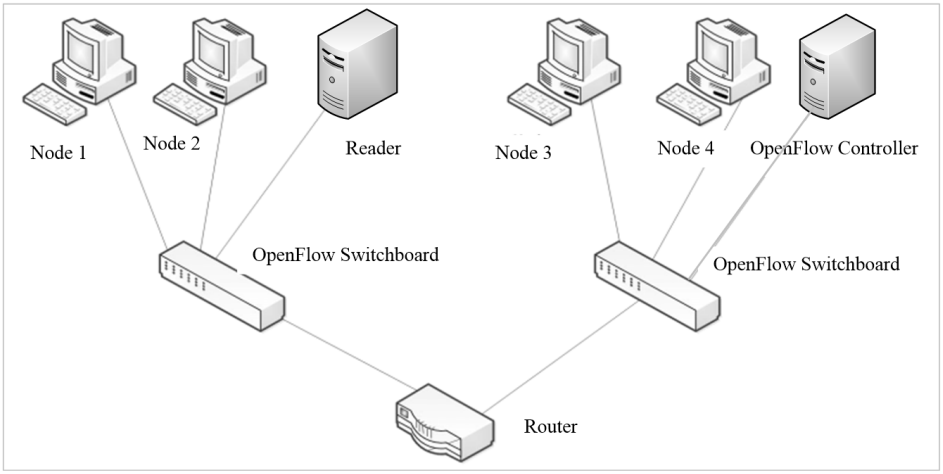


Fig. 2. Hadoop cluster topology based on SDN

sion overhead is great. And the BS-IDS scheduler adjusts the waiting time threshold so that the task is scheduled to run on the local node as far as possible. The rate of data localization is 100 %. In such networks, if the delay scheduling algorithm is adopted, because its setting is a static waiting time threshold, a small amount of data needs to be transmitted through the network. The data localization ratio of group 1, group 2, group 3 and group 4 is 95 %, 99 %, 94 %, and 93.5 %, respectively. For FIFO without adopting delay scheduling, part of the data needs to be transmitted over the network, increasing the network overhead of the data center.

Table 2. Job parameter configuration table

No.	The number of task	The number of job	The size of job
1	1	40	64 MB
2	4	10	256 MB
3	10	4	640 MB
4	20	2	1280 MB

As shown in Fig. 3, in the 1000 Mbps environment, data transmission in the network takes less time. Using the BS-IDS scheduler shortens the waiting time threshold and avoids invalid waiting time overhead. The proportion of the local data is relatively low. The delay scheduling algorithm is set to a fixed waiting time threshold, and there is only a small amount of data reaching the node implementing tasks through the data center network, maintained a higher proportion of local data. Because the BS-IDS scheduler waiting time threshold varies with data center load, when a job is submitted, in the team's first operations, there are a lot of tasks to be implemented [7]. The reaching intensity of the data center is great, thus it needs to set a larger waiting time threshold to schedule the team's first task on the local node, and then it can guarantee a certain number of tasks performed in the local

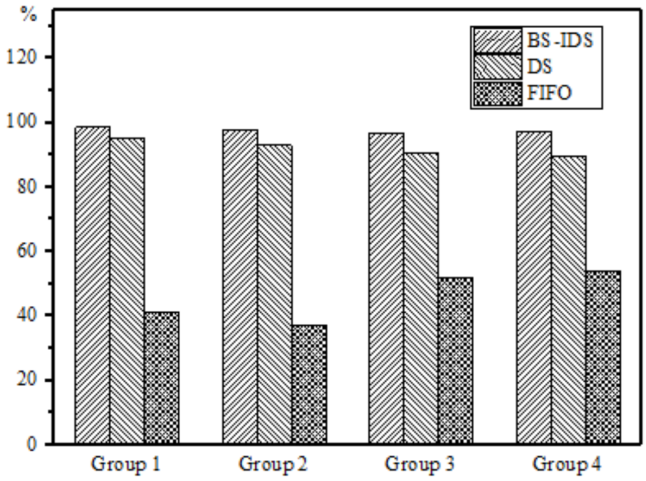


Fig. 3. Comparison of data local probability in the 100 Mbps environment

node. With the implementation of the task, the waiting time threshold is set to be close to 0, which is higher than FIFO.

Job response time comparison:

As shown in Fig. 4, in the 100 Mbps environment, the BS-IDS scheduler schedules the tasks to the nodes that the data is needed for the task, reducing the transmission time overhead of the data in the network. Moreover, during the data transmission time, the SDN bandwidth management and control ability is used to assign the bandwidth for the task and ensure the efficient allocation of tasks, which further reduces the response time of the job. Operation scale is increasing continuously, while the probability of local data by using delay scheduling is falling, which leads to the migration time overhead of data in the network increases, and a corresponding increase appears in job response time. As a result, the performance gap by BS-IDS is growing. While FIFO did not consider local implementation of tasks, it led to a large amount of data network transmission, and its performance was about 40 % of BS-IDS. Taking group 4 as an example, the completion time of the BS-IDS algorithm is reduced by 21 s compared to that of the delayed scheduling, and compared to FIFO, it is reduced by 150 s [8].

As shown in Fig. 4, in the 1000 Mbps network environment, the waiting time threshold of the BS-IDS scheduling algorithm is smaller because of the small network transmission overhead. At this point, the BS-IDS scheduling algorithm is slightly better than the FIFO, and the delay scheduling algorithm results in an extended response time due to the excessive waiting time overhead. Taking group 4 as an example, the job completion time of BS-IDS scheduling algorithm is 10 s less than FIFO, and it is reduced 130 s than that of delay scheduling.

BS-IDS scheduling algorithm's team's first job waiting time threshold variation rule:

In order to compare the effect of network bandwidth on the waiting time thresh-

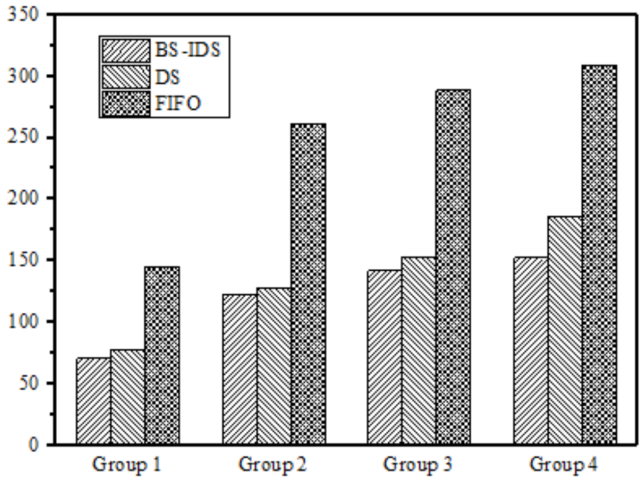


Fig. 4. Comparison of job response time in the 100 Mbps environment

old, this test compares and analyzes the operations with the number of group 1 under the 100 Mbps and 1000 Mbps networks. As shown in Fig. 5, it can be seen that, in the 1000 Mbps environment, the waiting time threshold of BS-IDS scheduling algorithm is smaller than the waiting time threshold of BS-IDS scheduling algorithm under the 100 Mbps network environment. This is because, in high bandwidth environments, BS-IDS scheduling algorithm will assign more tasks to non local nodes, to avoid wasting too much waiting time.

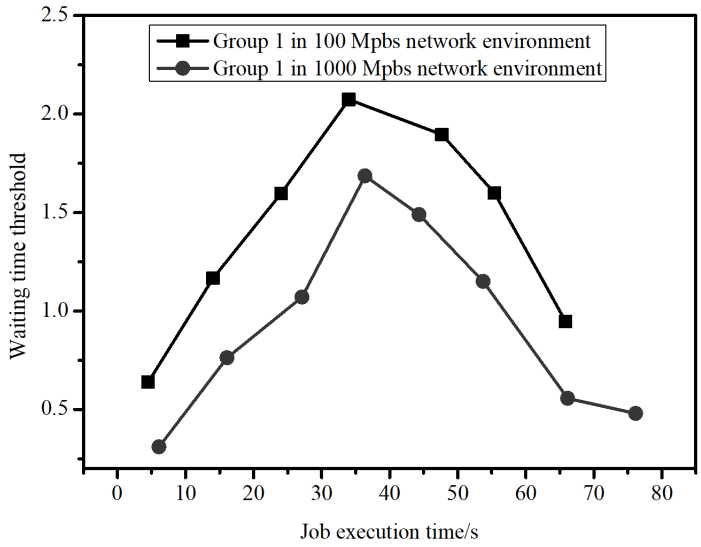


Fig. 5. Variation of waiting time threshold under different network environments

5. Conclusion

With the arrival of the big data era, the data centers operating large-scale distributed computing are built around the world. In the data center, nodes communicate with each other over the network. The limited bandwidth resources between nodes become the key factors that affect the performance of data centers. Using the bandwidth control capability provided by SDN, this paper proposes a scheme to improve the performance of large data processing by using link bandwidth. In the actual test environment, the Hadoop default scheduling algorithm FIFO, delaying scheduling algorithm, and BS-IDS scheduling algorithm are compared, and the efficiency of BS-IDS scheduling algorithm is proved.

References

[1] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami: *Internet of Things (IoT): A vision, architectural elements, and future directions.* Future Generation Computer Systems *29* (2013), No. 7, 1645–1660.

[2] C. W. Tsai, W. C. Huang, M. H. Chiang, M. C. Chiang,C. S. Yang: *A hyper-heuristic scheduling algorithm for cloud.* IEEE Transactions on Cloud Computing *2* (2014), No. 2, 236–250.

[3] D. Trentesaux, C. Pach, A. Bekrar, Y. Sallez, T. Berger, T. Bonte, P. Leitão, J. Barbosa: *Benchmarking flexible job-shop scheduling and control systems.* Control Engineering Practice *21* (2013), No. 9, 1204–1225.

[4] H. Duan, C. Chen, G. Min, Y. Wu: *Energy-aware scheduling of virtual machines in heterogeneous cloud computing systems.* Future Generation Computer Systems *74* (2017), 142–150.

[5] W. Wang, K. Zhu, L. Ying, J. Tan, L. Zhang: *Maptask scheduling in mapreduce with data locality: Throughput and heavy-traffic optimality.* IEEE/ACM Transactions on Networking *24* (2016), No. 1, 190–203.

[6] K. Kaur, T. Dhand, N. Kumar, S. Zeadally: *Container-as-a-service at the edge: Trade-off between energy efficiency and service availability at fog nano data centers.* IEEE Wireless Communications *24* (2017), No. 3, 48–56.

[7] T. Vondra, J. Šedivý: *Cloud autoscaling simulation based on queueing network model.* Simulation Modelling Practice and Theory *70* (2017), 83–100.

[8] M. Kalra, S. Singh: *A review of metaheuristic scheduling techniques in cloud computing.* Egyptian Informatics Journal *16* (2015), No. 3, 275–295.

Received May 7, 2017

Acta Technica *62* No. 2A/2017, 455–464

Q-ary LDPC codes based on RA structure for optimization of optical fibre communication system

Li Yingying[1], Wang Sunan[1,2], Wang Yongxue[1]

Abstract. In order to design a better optical fiber communication system, a multivariate LDPC encoding system based on RA structure is proposed. As a kind of efficient error correcting code, low density parity check code (LDPC) has good performance. It overcomes the shortcomings of other error correcting codes and is the nearest channel coding found from the Shannon limit. Since it was rediscovered, the LDPC code has received a lot of attention. It is another hotspot in the field of channel coding after Turbo code, and has become the preferred coding scheme for the fourth-generation mobile communications. The main research work is to apply the multiple LDPC codes to the optical fiber communication system on the basis of the four-bit interleaved LDPC coded modulation (BI-LDPC-CM) scheme of binary LDPC codes. So that it can meet the growing demand for high capacity transmission of fiber links. In the aspect of hardware implementation, the encoding and decoding of quaternion LDPC code is realized by FPGA. The experimental results show that the construction method of the RA code structure can improve the calculation of the product of the information bits of all lines and the non-zero elements in the check matrix.

Key words. RA structure, q-ary LDPC code, FPGA, Max-log-BP decoding algorithm.

1. Introduction

Low Density Parity Check Code (LDPC) code is a kind of efficient error-correcting code with good performance, which overcomes the shortcomings of other error-correcting codes. It is the channel code found at present and the closest to Shannon's limit [1]. Since the rediscovered LDPC codes have been widely concerned, it is another hotspot in the field of Turbo code channel coding, which becomes the preferred coding scheme of the fourth generation mobile communication. At present, the binary LDPC code has been widely used in a variety of digital communication systems. However, with the same parameters, the Tanner graphs of the q-ary LDPC code are sparser than that of the binary LDPC code, and its girth is larger than the binary

[1]Shenzhen Polytechnic, 518000, China
[2]Corresponding author

LDPC code. This feature can facilitate the optimization design of LDPC codes and reduce the effects of the short-loop and the stop-set on the convergence of the decoding [2].

Therefore, the q-ary LDPC codes can design good codes with a lower error leveling and better error correction ability to meet the high-speed transmission needs of the communication system. We focus on the design of the q-ary LDPC codes based on RA structure and the FPGA hardware implementation based on Max-log-BP decoding algorithm. From the view of performance, speed and resource consumption of codec, it can meet the general communication requirements. The application of this hardware implementation in the underwater acoustic communication network can realize the high-speed and efficient transmission of information. Combined with radio technology, the omnidirectional and three-dimensional communication can be established in the ocean, which can help people to observe and develop the ocean.

2. Literature review

At present the binary LDPC code has been widely used in the optical fibre communication system [3]. However, with the same parameters, the Tanner graphs of the q-ary LDPC code are sparser than that of the binary LDPC code, and its girth is larger than the binary LDPC code. This feature can facilitate the optimization design of LDPC codes and reduce the influence of the short-loop and stop-set on the convergence of the decoding [4]. Therefore, the q-ary LDPC codes can design good codes with a lower error leveling and better error correction ability to meet the high-speed transmission needs of the communication system.

In his paper in 1962, Dr. Gallager proposed the idea of LDPC coding, which is a kind of linear block code, and its special feature is that its check matrix is very sparse [5]. According to the idea of Gallager, when the LDPC code word is constructed by using the check matrix, the generated code word can satisfy the constraint condition of linear, and each symbol contained in the generated code word satisfies certain constraints. In the iterative decoding theory and some other theory aspects, Gallager proposed LDPC code, which far surpasses the Turbo code. But at the time, many researchers were busy with BCH code, Reed-Solomon codes and concatenated codes, and under the impact of the speed of the encoding storage and decoding computation, except for a few researchers, LDPC codes were not able to attract people's enough attention, and it was even forgotten. Until 1996, MacKay and Neal and others found that when they used the optimal decoder to decode, the LDPC proposed by Gallager LDPC was a good code with a lower linear decoding complexity [6], and more and more researchers will pay attention to the LDPC code.

Tanner extended the Gallager construction, introduced the binary graph model [7] into the LDPC code, and proposed the Tanner graph. Tanner graph can show the connection between the code word symbols and its constraints. When the constraints are all binary checksum, you can get Gallager code. In addition to this research, Tanner improved the algorithm idea of Gallager, the decoders used in the process of the implementation of the iterative decoding are independent each other. In terms of LDPC code constraint, the check constraint is improved as the linear constraints,

which makes the LDPC code be widely used. Therefore, Tanner has great contribution to the LDPC code. Wiberg re-studied on the basis of the Tanner graphs, and proposed the "minimum sum" and "sum-product" algorithms. Richardson and Luby studied the threshold effect of LDPC code. If the block length is infinitely close to the infinite value, and the noise interference is very small, below the threshold, then the system can make the bit error rate arbitrarily small. When the noise level is greater than this threshold, the bit error probability is greater than a positive constant. Luby used the threshold to analyse, and firstly studied the irregular structure of LDPC code, the weight of each row and each column of the check matrix for is not a constant, and the "Density Evolution" iterative algorithm is used to build irregular LDPC code [8].

Domestic research is also very rich, and the main contributions are as follows. Hu Xiaoyu (School of Computer Science and Technology, Peking University, Beijing 100871, China) proposed a new PEG algorithm that can maximize the local loops of variable nodes [9]. Yu Kou and Shu Lin propose an algorithm of point and line based on Euclidean space and projection geometry on finite fields to design LDPC codes, which have achieved good performance. In addition, in the hardware implementation, the most is the research based on binary LDPC code research, but the q-ary LDPC codes involved little. Davey and MacKay promoted the binary binary LDPC code to the q-ary LDPC [10]. In a multidimensional domain, the message of q-ary LDPC code is not a single one-bit symbol, and it consists of multiple components [11]. For the non-binary LDPC code, its coding complexity is increased, the calibration information is also complex, but the decoding process of non-binary LDPC code is relatively simple, and the decoding algorithm is different from the binary LDPC code decoding method, and we need to pay attention on it [12].

Now, people mainly explore deeply the LDPC code from the following two levels. First, in order to improve performance, they usually use large code length binary LDPC code to approach Shannon limit; Second, it do not need to use long code length on the study of q-ary LDPC code, and mainly concentrated in the small code length [13]. Regardless of binary or q-ary LDPC codes, most researches focus on the construction of parity check matrix, exploration of new coding and decoding algorithms, and analysis of the performance of codes, so as to lay a solid theoretical basis for the use of LDPC codes in time production. This paper is the study of q-ary LDPC codes.

2.1. Design of FPGA encoder based on RA structure

The code word vector of LDPC code consists of information bits and check bits. The encoding process of the LDPC fast encoding algorithm based on RA structure is very simple, and its implementation is mainly divided into two parts. First, you need to calculate the product of the non-zero elements in the check matrix H and the information bits, and use the idea of iteration to compute the parity bit, and then combine the information bits with the parity bit, which forms the desired code word. In the specific implementation, the form of the look-up table is used to store the column address of the non-zero element of parity matrix H in the ROM, which

can simplify the calculation of parity bit. The entire encoder consists of control module, input module, coding module and output module.

Under the control of the clock and reset signal and combined with the counter, the control module provides the control signal to several other input, coding and output modules through the conversion between the states, including several important control signals: the read write signal of the ROM and RAM and the enable signal of each module. The block diagram of the control module is shown in Fig. 1.

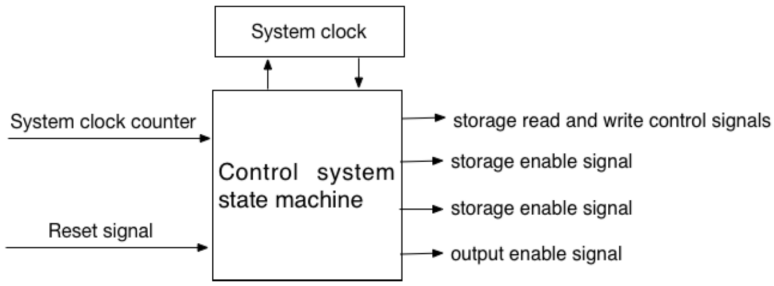


Fig. 1. Design diagram of control system

The role of the input module is that use the ping-pong operation to complete the buffer on the input data, and then extract the corresponding address with the address output by ROM look-up table, and output the prepared data that calculates the parity bit to the encoding module. As a result of the construction method of the RA code structure, the calculation of the product of the information bits of all lines and the non-zero elements in the check matrix becomes the key of this method, that is, calculate $r_i = \sum_{j=1}^{k} h_{ij} \cdot m_j,\ i = 1, 2, \cdots, 486$. The advantage of putting the computation of r_i in the input module is that it reduces the computational complexity of the parity bits in the subsequent encoding module.

Therefore, the calculation of value r_i becomes a critical step in the input module. The specific implementation plan is as follows: The RAM group address storing the storage information element m_j is calculated, and the address is stored in the read-only one-port ROM. Then, the non-zero element value in the H matrix corresponding to the calculation r_i is stored in the other lookup table ROM. Finally, in the cooperation of the clock, the information element $m_1, m_2, ..., m_k$ and the corresponding $m_1, m_2, ..., m_k$ are taken out in turn, and the product of $m_1, m_2, ..., m_k$ and the corresponding $h_1, h_2, ..., h_k$ are calculated. The result is accumulated, and the cumulative result is r_i.

The main function of the coding module is the calculation of parity bit. The intermediate variable r_i output by the input module is transformed to the encoding module to calculate the relevant parity bit p_i and the temporary information bit m_i, which is sent to the output module after the merger. The calculation of the parity bit main use the recursive idea. The specific realization is completed by the multiply-add unit, counter, delay register, and multi-select data selector. The calculation of each specific parity bit is done by the multiply-add unit. The counter is used to generate the chip select signal of the data selector, and the data selector selects the specified parity bit to control the output of the data. The delay register

can generate the delay signal p_{i-j} to achieve the delay function of each parity bit. The specific block diagram of the coding module is shown in Fig. 2.

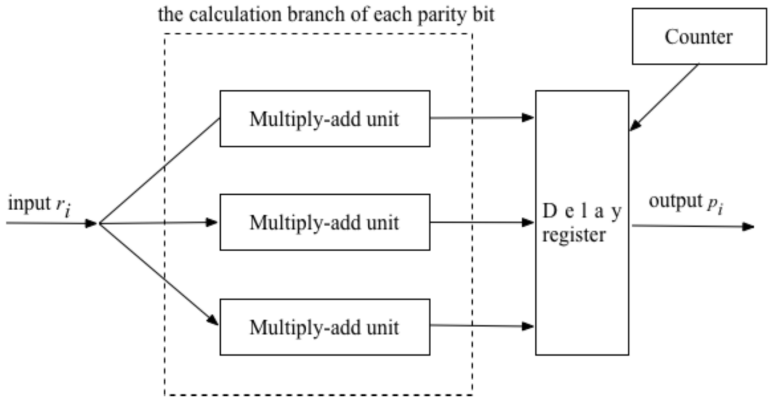


Fig. 2. Structure Block diagram of the coding module

The main function of the output module is to synthesize the final code word, and buffer and output the code word. The module combines the parity bit information obtained by the coding module with the information bit of the buffer and then separates the parity bit from the information bit through a substitution table consisting of a read only single-port ROM whose address is the corresponding parity bit and information bit, and finally double-buffered ping-pong technology is used to read out to the received parity bit and information bits, and then get the code word output [14].

When the encoder is implemented with EP3C55F780C8 model FPGA, the logic unit consumed is 80 %, and the register unit occupies 64 % of the total resource. At this point, the combined clock frequency is 46.34 MHz. Since the quaternion equation is used, each codeword requires two bits, so the data rate is $46.34 \times 2 = 92.68$ MHz. Compared with binary LDPC codes, the hardware resources consumed by FPGAs are relatively large when FPGA is used to implement LDPC codes. In addition, the coding complexity and coding delay are increased.

3. Design of FPGA Decoder Based on Max-log-BP Algorithm

One of the difficulties in the design of LDPC code decoders is that the transmission of information can be effectively carried out on the basis of iteration, so that the decoder can be divided into three kinds of decoding structures: full serial, full parallel and partly parallel.

Full serial architecture: The full serial structure decoder mainly consists of check node units (CNU), variable node units (VNU), control units and intermediate information RAM. The operation of the serial decoder is: The control unit controls the CNU and VNU to alternately complete the operation between the variable node and the check node. The function of each unit is: The CNU updates the check node

information, and VNU updates the variable node information. The intermediate information RAM is used to store the intermediate results of the CNU and VNU iterative operations. The control unit can coordinate the working sequence of CNU and VNU and control the iterative decoding process.

When using the full serial decoding structure, because CNU and VNU are working alternately, it is possible to avoid the access violation of CNU and VNU to intermediate information RAM by designing a multiplexing module to arrange the access reasonably. Because of the alternating work of CNU and VNU, this causes the calculation of the variable nodes to be waited when calculating the check node. In this way, the full-serial decoder can avoid the occurrence of access conflict and consumption of hardware resources at least, but the processing speed is slow. The full-serial decoder is suitable for hardware resource requirements, but the speed of data processing is not demanding.

Full parallel structure: The form of the full parallel structure is very similar to the bipartite graph, containing m CNUs and n VNUs. In the decoder of fully parallel architecture, both the check node and the variable node have their respective independent computing units. In the iteration, it does not have to wait for the calculation results of other units. In the previous clock cycle, all the check nodes can be updated, and the last clock cycle can complete the update of all variable nodes. Therefore, the intermediate information generated by the iterative computation is not required to be stored like a full serial structure decoder. This enables the decoder of the structure to have high speed of operation and can basically perform real-time decoding. However, the number of computing units in the full parallel architecture decoder is too large, and the resource consumption is very large. The critical path length is increased when routing is adopted, which makes the utilization of the chip very low and is not suitable for long format data format.

Partially parallel structure: Logical resources are important in the design of FPGA. Most designs are expected to be implemented with minimal logical resources in order to reduce costs. The full serial structure of the decoder occupies less resources, because it uses the idea of logical multiplexing. Through the repeated use of CNU and VNU, the consumption of resources is reduced. In combination with the structure of full parallel decoding and the idea of multiplexing, a partially parallel structure is used to decode multivariate LDPC codes. The design of FPGA pays attention to the idea of module reuse, and this part parallel design can fully reflect this idea. Structurally, CNU and VNU between partial parallel decoders are cached through a RAM storage array of dual ports. Compared with the decoder with full parallel structure, the number of CNU and VNU is different, and the decoder of partial parallel structure introduces the problem of parallelism.

For example, it is assumed that the row number of the check matrix H is M, and the column number is N. The ranks of the parallel degrees, respectively, are p and q, and they are divisors of m and n. The decoder has Q VNU arithmetic units and P CNU arithmetic units. Each VNU arithmetic unit is responsible for updating the calculated N/q column variable nodes, and each CNU arithmetic unit is responsible for updating the calculated M/p line check nodes. Therefore, it takes N/q times to complete the full update of the variable node. That is, in N/q clock cycles, the N

column variables are updated in chronological order. Furthermore,Q variable nodes can be calculated in each clock cycle, and it takes N/q clock cycles to complete the update calculation of all variable nodes. The updating process of the test node is similar, and there is no more description here. If the resource consumption situation of FPGA is concerned, when the decoder uses full parallel mode, the logical units occupied by the check node and the variable node are C_1 and C_2, respectively. Then, when the structure of the decoder is changed to partially parallel, the entire decoding process requires the resources to be $C = (C_1/p) + (C_2/q)$. At the same time, the consumption of hardware resources is reduced by changing the structure of the decoder. Of course, the decoding time of this partial parallel structure is longer. Therefore, in the selection of structures, we should give full consideration to the balance of speed and area, in order to meet the requirements of the design system.

The overall design of the decoder block diagram is shown in Fig. 3. The modular structure runs through the entire decoder design process. It can be seen from the block diagram that the decoder mainly is composed of the input buffer module, CNU module, VNU module, external information storage module, output buffer module, decode decision output module and the control unit module. With the cooperation of the clock, the work of the whole decoder is controlled by the control unit module. First, the initial information enters into the input buffer unit after quantization, and completes the serial conversion, and then the decoder enters working condition. Next, the check node and the variable node are updated, and the CNU and VNU exchange information through the external information storage module. Finally, after the output buffer module completes the parallel-serial conversion, the data enters the decoding decision module to judge the obtained code word C satisfies C·HT=0, or reach the maximum decoding iterations, so as to get the final output code word sequence.

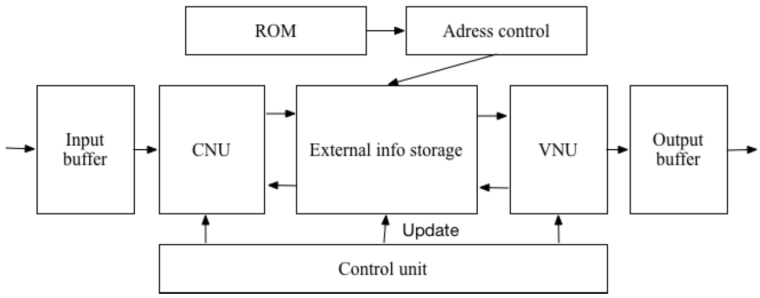


Fig. 3. Overall design diagram of decoder

In the LOG operation domain, each variable point of the q-ary LDPC code has $q-1$ initial information, and the binary LDPC code has only one. Since each variable point of the quaternary LDPC code has 3 channel initial information $L(v) = L(v_1, v_2, v_3)$, the operation amount of the decoder becomes larger and occupies more logic resources. However, in the whole decoding process, the information vector of each variable point is also involved in the operation, so $L(v) = L(v_1, v_2, v_3)$ can be regarded as a whole to store and transfer, and $L(v) = L(v_1, v_2, v_3)$ can be separated

only in the operation of the check node unit and the variable node unit. CNU module achieves the relevant Field word operation [15]. The Field word calculation formula is not introduced too much here.

The load balancing of reducing ends with different inclination is shown in Fig. 4. In this figure, Lx_1 ,Ly_1, Lx_2, Ly_2, Lx_3 and Ly_3 are the input vectors of two variable points for each Field word operation. CNU design part must have two functions. One of the functions is to complete the Field word operation, and the other determines that the output results meet the calibration equation. Only when the CNU module calculates that the result of each checkpoint is 0, it indicates that the calculated code word is correct, otherwise the code word is incorrect, iterating again until the correct code word is reached or the maximum iterations is reached.

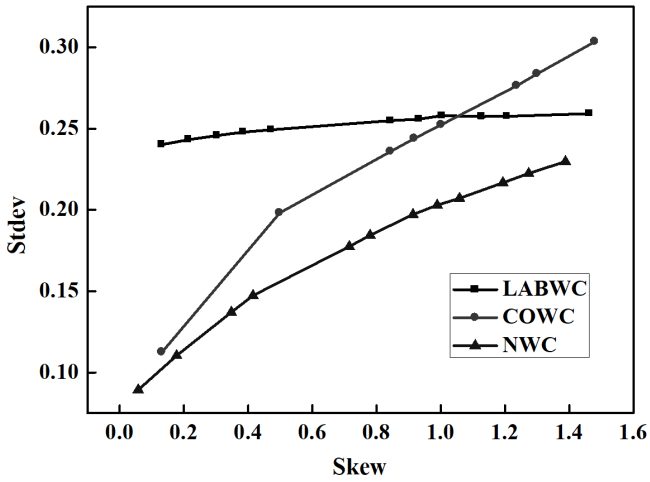


Fig. 4. Load balancing of reducing ends with different inclination

In the design of the VNU module, the calculation of the formula is mainly completed. Symbol $L(qmn)$ denotes the external information that the variable node passes to the check node with respect to the value of the variable node; pn denotes the initialization iteration information; $L(rm'n)$ denotes that the check node passes to the variable node with respect to the log-likelihood information required the check equation when the value of the variable node Isai. Symbol $L_\rho(m, n, l)$ denotes the degree distribution function of the check node. Taking the quaternary LDPC code as an example, the function of the VNU module is to compute the information vector, and the information vector is divided into three sub information vectors. Symbol $L(v)$ denotes the channel initial information of each variable. Symbols $L(A)$, $L(B)$, $L(C)$ and $L(D)$ are the information of four variable nodes respectively.

4. Conclusion

The decoding complexity of LDPC codes is lower and the performance is closer to the Shannon limit. In contrast to the binary LDPC code, the multivariate LDPC

code is more robust in error correction and burst error resistance. It is more suitable for higher order debugging system and efficient data transmission. The multiple LDPC codes are introduced into the optical fiber communication system, and the design of the coder and decoder of the multiple LDPC codes is completed. The main main conclusions of this study are as follows:

First, the encoding method of multivariate LDPC code based on RA structure is introduced in detail. The decoding performance of the maximum BP algorithm in the logarithmic domain, EMS and the modified EMS algorithm in AWGN channel are compared. Finally, a fast decoding method based on FFT-BP and combining four-dimensional modulation is selected. Multi LDPC codes are used in optical fiber communication systems to achieve high coding gain, spectral efficiency and low power consumption, thus satisfying the requirements of high speed and even large capacity transmission of optical fiber.

Second, the FPGA chip used in there is the EP3C55F780C8 of the Cyclone III series device. Using VerilogHDL hardware description language, and combining Quartus II, Matlab and Modelsim development software, the design of encoding and decoding of multiple LDPC codes is realized. Encoding uses an algorithm based on the RA structure, and the decoding uses an algorithm based on the Max-log-BP. The detailed block diagram of each module is given in the design, and the function of each module is explained. The simulation waveform of Modelsim is given. Whether from the resource occupation or the decoding rate, the designed codec of the multiple LDPC codes basically meet the requirements of communication.

Of course, there are still many deficiencies in the design, especially the resource and performance problems in the hardware implementation and the efficiency of the system. The next step of research can be carried out in two aspects: In the hardware implementation phase, the resources of the multi-LDPC code are much more than the resources occupied by the binary LDPC code. Therefore, the problem that needs to be solved urgently is how to enhance the performance and rate of multiple LDPC codes under the condition of limited resources. This needs to improve the existing encoding and decoding algorithms and to maximize the module reuse at the time of design in order to better solve the constraints of resources, performance and speed Although the multivariate LDPC code is introduced into the optical fiber communication system, it is verified only in the theoretical simulation, and has not been built into a specific communication system. The whole optical fiber communication needs to be improved so as to meet the growing demand for high capacity optical fiber transmission.

References

[1] J. Ao, J. Liang, C. Ma, G. Cao, C. Li, Y. Shen: *Optimization of LDPC codes for pin-based OOK FSO communication systems.* IEEE Photonics Technology Letters *29* (2017), No. 9, 727–730.

[2] M. Alibakhshi-Kenari, M. Naser-Moghadasi, R. A. Sadeghzadeh, B. S. Virdee, E. Limiti: *Traveling-wave antenna based on metamaterial transmission line structure for use in multiple wireless communication applications.* AEU - International

Journal of Electronics and Communications *70* (2016), No. 12, 1645–1650.

[3] L. B. Shen, X. M. Zhao, X. Y. Zhu, L. Wang, Q. J. Liu, B. Z. Wei: *A design for a remote condition monitoring system for an optical fibre smart structure based on advanced reduced instruction set computing (RISC) machines (ARM) and general packet radio service (GPRS).* Lasers in Engineering *30* (2015), Nos. 1–2, 15–29.

[4] C. Gong, S. Li, Q. Gao, Z. Xu: *Power and rate optimization for visible light communication system with lighting constraints.* IEEE Transactions on Signal Processing *63* (2015), No. 16, 4245–4256.

[5] W. G. Ma, Y. Cao, W. Wei, X. H. Hei, J. F. Ma: *Energy-Efficient collaborative communication for optimization cluster heads selection based on genetic algorithms in wireless sensor networks.* International Journal of Distributed Sensor Networks (2015), Article No. 103.

[6] H. Wang, H. Zhang, F. Yang, S. Song, Y. Chang, C. Bei, K. Yang: *Parametric optimization of regenerative organic Rankine cycle system for diesel engine based on particle swarm optimization.* Energies *8* (2015), No. 9, 9751–9776.

[7] C. Abellán, W. Amaya, D. Mitrani, V. Pruneri, M. W. Mitchell: *Generation of fresh and pure random numbers for loophole-free bell tests.* Physical Review Letters *115* (2015), No. 25, paper 250403.

[8] D. J. Saunders, J. H. Munns, T. F. Champion, C. Qiu, K. T. Kaczmarek, E. Poem, P. M. Ledingham, I. A. Walmsley, J. Nunn: *Cavity-enhanced room-temperature broadband raman memory.* Physical Review Letters *116* (2016), No. 9, paper 090501.

[9] A. M. Queirós, P. Taylor, A. Cowles, A. Reynolds, S. Widdicombe, H. Stahl: *Optical assessment of impact and recovery of sedimentary pH profiles in ocean acidification and carbon capture and storage research.* International Journal of Greenhouse Gas Control *38* (2015), 110–120.

[10] G. David, K. Esat, S. Hartweg, J. Cremer, E. Chasovskikh, R. Signorell: *Stability of aerosol droplets in Bessel beam optical traps under constant and pulsed external forces.* Journal of Chemical Physics *142*, (2015), No. 15, paper 154506.

[11] T. Feng, J. J. McGuirk: *Measurements in the annular shear layer of high subsonic and under-expanded round jets.* Experiments in Fluids *57* (2016), Article No. 7, paper 25.

[12] A. Figarol, J. Pourchez, D. Boudard, V. Forest, S. Berhanu, J. M. Tulliani, J. P. Lecompte, M. Cottier, D. Bernache-Assollant, P. Grosseau: *Thermal annealing of carbon nanotubes reveals a toxicological impact of the structural defects.* Journal of Nanoparticle Research *17* (2015), No. 4, 1–14.

[13] W. Cai, P. Wang, Y. Li, Y. Zhang, B. Vucetic: *Deployment optimization of uniform linear antenna arrays for a two-path millimeter wave communication system.* IEEE Communications Letters *19* (2015), No. 4, 669–672.

[14] R. Gangula, D. Gesbert, D. Gündüz: *Optimization of energy harvesting miso communication system with feedback.* IEEE Journal on Selected Areas in Communications *33* (2015), No. 3, 396–406.

[15] X. Liu, M. Qiu, X. Wang, W. Liu, K. Cai: *Energy efficiency optimization for communication of air-based information network with guaranteed timing constraints.* Journal of Signal Processing Systems *86* (2017), Nos. 2–3, 299–312.

Received May 7, 2017

Acta Technica *62* No. 2A/2017, 465–474

Quantitative study of correlation strength of mechanical parts based on weighted complex network model

Liu Hang[1]

Abstract. In order to better calculate the correlation strength of mechanical parts, a new calculation method based on weight complex network model is designed. A complex product system module partition method based on the weighted complex network community discovery technology is proposed. In order to construct a complex product system, the component weight complex network model is used to map the parts into network nodes. The relation and intensity between components are mapped into weighted edges and edges of complex networks. The quantitative method of relation strength between function and structure is studied. The experimental results show that a number of module partitioning schemes are obtained to form a scheme set through the different values of network modularity. Based on the above finding, it is concluded that the improved GN algorithm can be used to realize the complex structure of complex product system and the community structure of the complex network.

Key words. Weight, complex network model, intensity, association.

1. Introduction

With the progress of science and technology, the evolution of technology related to subsystems has been accelerated in complex product systems. Emerging customer demand continues to emerge, and current customer demand continues to change. Objectively, the efficiency of R & D of complex product systems needs to be improved continuously to cope with the complex and changeable market environment [1]. The research of existing complex product system mainly focuses on the concepts, characteristics, interpretation, innovation mechanism exploration and research and development, operation management and other factors related to the new research and development of complex product systems [2]. It lacks the research on the basis of several variants. Design knowledge reuse and redesign are developed to improve R & D efficiency [3]. Modular design is widely used in the field of mass

[1]School of Management Engineering, Zhengzhou University, of Aeronautics, Zhengzhou, Henan, 450001, China

customization. It is an effective way to improve the efficiency of product development through the standardization of product structure and design variables [4]. Compared with traditional mass customization products, complex product system has high complexity and obvious heterogeneity, but the traditional modular design theory is difficult to deal with these characteristics effectively [5]. There are many reasons for the heterogeneity of complex product systems [6]. First of all, changes in customer demand require complex product systems to constantly add new features or to improve existing functions, which can easily lead to functional heterogeneity [7]. Secondly, due to the adoption of integrated product development in complex product systems, there is a lack of effective communication between product research and development projects [8]. At present, the flow of core research and development personnel is very large, and it is easy to produce the same functions and principles with different physical structure, forming structural heterogeneity [9]. To solve the above problems, matrix analysis method is used to realize the homogenization reduction of the function and structure heterogeneity of the existing complex product system. A weighted complex network model is introduced to study the module partitioning problem of complex product systems [10].

2. Introduction of relevant concepts

2.1. Complex network concepts

A complex network can be represented as a graph $G = (V, E)$ consisting of a set of nodes, V, and an edge set E. Among them, the element in formula $V = (v_1, v_2, ..., v_n)$ represents the node in the network, and the elements in formula$E = (e_1, e_2, ..., e_m)$ represent edges in a network. The graphical representation of a complex network is shown in Fig. 1.

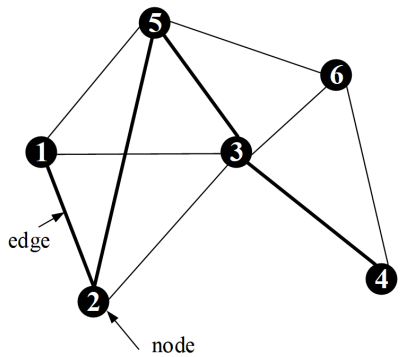


Fig. 1. Graphical representation of complex networks

If it is not specified, the weight of the complex network edge is generally 1. This network can be expressed in binary matrix A, which is called adjacency matrix. The

matrix A is made up of elements A_{ij}, and A_{ij} follows the following rule:

$$A_{ij} = \begin{cases} 1, & \text{if the node } i \text{ is connected to the node } j\,, \\ 0, & \text{other}\,. \end{cases} \tag{1}$$

Then, the adjacency matrix shown in Fig. 1 can be represented as

$$A = \begin{bmatrix} - & 1 & 1 & 0 & 1 & 0 \\ 1 & - & 1 & 0 & 1 & 0 \\ 1 & 1 & - & 1 & 1 & 1 \\ 0 & 0 & 1 & - & 0 & 1 \\ 1 & 1 & 1 & 0 & - & 1 \\ 0 & 0 & 1 & 1 & 1 & - \end{bmatrix}. \tag{2}$$

The adjacency matrix of complex networks requires attention to the following two points: First, the autocorrelation between the nodes of a complex network is not considered, so the diagonal elements of A do not exist. Second, if it is not otherwise specified, A is a symmetric matrix and complex networks are undirected networks. When the topology of a complex network is established, a series of parameters used to characterize its properties can be calculated. Several complex network parameters related to the research are introduced below [11].

In a complex network, the degree of node (degree) can be expressed as k_i, which refers to the number of edges connected to the node i, and is a positive integer [12]. For a complex network with n nodes, the degree of the node can be calculated by summing the elements of the adjacency matrix

$$k_i = \sum_{j=1}^{n} A_{ij}\,. \tag{3}$$

In the formula, A_{ij} represents the elements of the adjacency matrix A.

In a complex network, the number of nodes (between two nodes) is the shortest path through the node: in next text this will be denoted as betweenness. Assuming that the shortest path between the nodes s and t is through the node i, then $n_{st}^i = 1$, and vice versa, if $n_{st}^i = 0$, then the betweenness of the node i can be expressed as

$$b_i = \sum_{s,t}^{n} n_{st}^i\,. \tag{4}$$

A community in a complex network (community) can be defined as a set of closely linked nodes. It has a close connection between the nodes within the community, and the connection between the communities is sparse [13]. Figure 2 shows a complex network instance with three communities.

Modularity is a quantitative index to measure the degree of structure of a complex network community [14]. Through this indicator, a good community structure can be obtained. The most widely used modular computing method is proposed by

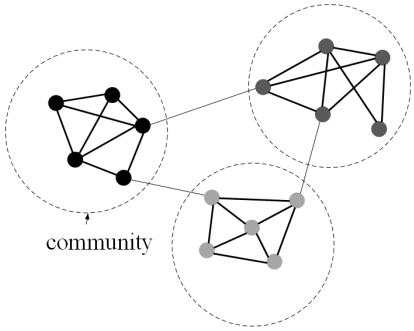


Fig. 2. Structure diagram of complex network group theory

Newman:

$$\lambda = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta \left(C_i, C_j \right) . \tag{5}$$

In the formula, m stands for the total number of edges in the network. If $s = t$, then the function $\delta(st)$ is valued at 1, and vice versa is 0. Symbols C_i and C_j represent the communities in which nodes i and j are located, respectively. The value of a lambda is a natural number between 0 and 1. Different modularity can be used to obtain different community structures.

2.2. Weighted complex network concepts

The weighted complex network can also be represented by an adjacency matrix, which is similar to the ordinary complex network. Figure 3 shows an example of a weighted complex network.

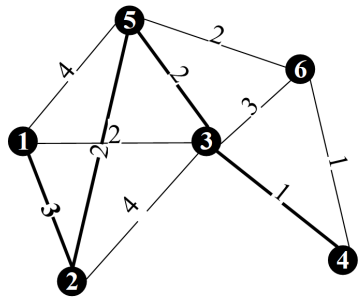


Fig. 3. Weighted complex network example

Its adjacency matrix is

$$A = \begin{bmatrix} - & 3 & 2 & 0 & 4 & 0 \\ 3 & - & 4 & 0 & 2 & 0 \\ 2 & 4 & - & 1 & 2 & 3 \\ 0 & 0 & 1 & - & 0 & 1 \\ 4 & 2 & 2 & 0 & - & 2 \\ 0 & 0 & 3 & 1 & 2 & - \end{bmatrix} . \tag{6}$$

The length of a path can be expressed as

$$L = \sum_{i,j}^{p} A_{ij}, \ i \neq j \, . \tag{7}$$

In the formula, A_{ij} represents the element of the adjacency matrix A, and P represents the number of nodes on the path. Obviously, the length of path $1-2-5-3-4$ in Fig. 3 is 15. The node betweenness, community structure and modularity in weighted complex network are similar to those of the complex networks.

3. Methods

In order to find a more efficient and appropriate way to solve the problem of complex network communities, Girvan and Newman designed the GN algorithm. By identifying the edges between the communities in the network and removing them one by one, the community structure of the complex network is finally obtained through this iterative process. The flow of the GN algorithm is shown in Fig. 4.

The betweenness of all edges in a complex network is calculated by using breadth first rules.

One of the largest edges (possibly more than one) is removed. If the shortest path between two nodes is more than one, the same weight is assigned to them, and then one of them is removed at random.

The betweenness of all remaining edges is calculated.

Step (2) and (3) are repeated until all edges are removed to obtain a community structure of a complex network. If there is a particular degree of modularity constraint, the algorithm ends when the iteration satisfies the module degree constraint.

4. Results and discussion

4.1. Standard analysis of module division

A functional modeling method is used by researchers in the field of product design. For example, as shown in Table 1, the lower the functional level of the two components realizes, the closer the relationship should be. On the basis of the above research, the quantitative evaluation of functional correlation between components

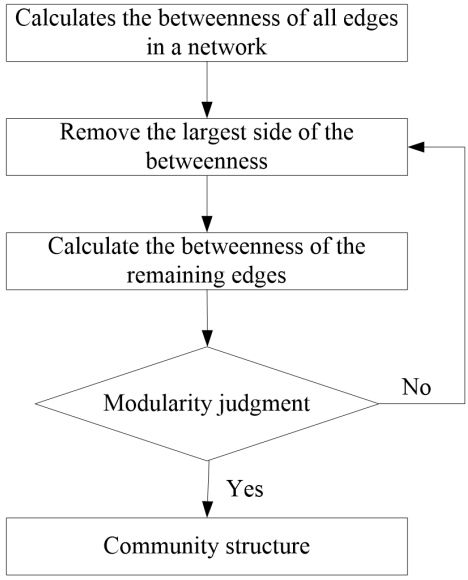


Fig. 4. Algorithm flow

is realized.

Taking the parts of the crawler crane track (Fig. 5) as an example, the above quantization standard is analyzed by example, and the result is shown in Table 2. Taking the gray background cell as an example, "2" means that the drive wheel and the track rack weld together to achieve the same main function. That is to say, the supporting track provides power and support for the whole machine. "10" means that the drive wheel and reducer work together to accomplish a sub function, that is to say, to provide walking power for the whole vehicle. "6" means that the guide wheel and the bracket respectively perform the function of walking, guiding and tensioning, and the guiding function is the subsidiary function of the tensioning function.

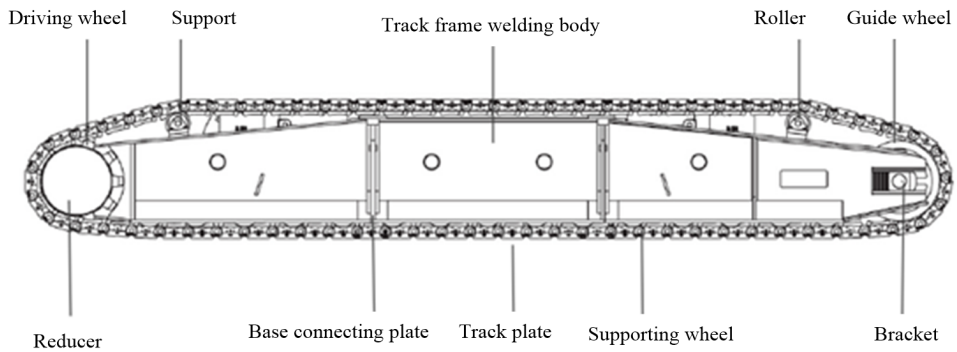
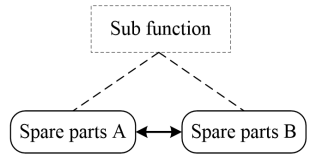
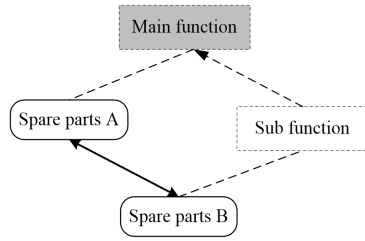
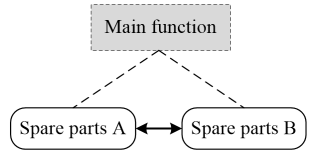
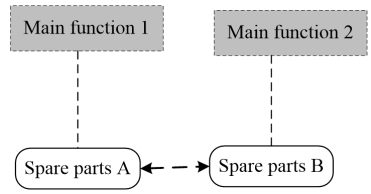


Fig. 5. Sketch map of crawler frame

Table 1. Relative strength of parts function

Relative strength	Score	Description	Graphic
Strong	10	Two components realize the same sub function together	
Common	6	One component is used to implement the main function, and the other is used to implement the sub function under the main function	
Weak	2	Two components are used to achieve the same main function	
Nothing	0	Two components are used to implement different main functions or sub functions	

5. Conclusion

The weighted complex network theory is introduced into the problem division of complex product system modules. Meanwhile, the problem of module partitioning is abstracted into the community discovery problem of weighted complex networks. Firstly, the relation and intensity between components are mapped to edge and edge weights of weighted complex networks. Secondly, because of the complexity of complex product systems, the core elements of product design are considered. That is to say, the function and the structure relation between components are used as the basis of the module division, and the quantitative method of the correlation

strength is put forward. Thirdly, the improved GN algorithm can efficiently realize the component weight of complex product system and find out the community structure of complex network. At the same time, this algorithm avoids the fact that the traditional clustering algorithm and heuristic algorithm cannot obtain the optimal solution because of the large number of parts. Finally, through the different values of network modularity, several module partitioning schemes can be obtained to form a scheme set, which provides a data base for the subsequent module partitioning scheme evaluation.

Table 2. Correlation strength evaluation data of function of crawler frame parts

No.	1	2	3	4	5	6	7	8	9	10
1		2	2	2	10	0	2	2	2	2
2			2	2	2	0	2	2	2	2
3				2	2	0	2	2	2	6
4					2	0	2	2	6	2
5						0	0	0	0	6
6							0	0	0	0
7								2	2	2
8									0	0
9										0
10										

References

[1] [T. S. Kim, S. H. Jeong, D. H. Chang, S. R. In, M. Park, B. K. Jung, Lee KW1, S. J. Wang, Y. S. Bae, H. T. Park, J. S. Kim, W. Cho, D. J. Choi: *Modification to the accelerator of the NBI-1B ion source for improving the injection efficiency*. Review of Scientific Instruments *87* (2016), No. 2, 351–356.

[2] M. D. Andujar-Montoya, V. Gilart-Iglesias, A. Montoyo, D. Marcos-Jorquera: *A construction management framework for mass customisation in traditional construction*. Sustainability *7* (2015), No. 5, 5182–5210.

[3] S. Hong, H. Yang, T. Zhao, X. Ma: *Epidemic spreading model of complex dynamical network with the heterogeneity of nodes*. International Journal of Systems Science *47* (2016), No. 11, 2745–2752.

[4] M. Cucuringu, P. Rombach, S. H. Lee, M. A. Porter: *Detection of core-periphery structure in networks using spectral methods and geodesic paths*. European Journal of Applied Mathematics *27* (2016), No. 6, 846–887.

[5] Z. Gao, S. Li, Q. Shang, Y. Jiao, X. Zhou, C. Fu, H. Xu, D. Shi, K. Chen: *Complex networks approach for analyzing the correlation of traditional chinese medicine syndrome evolvement and cardiovascular events in patients with stable coronary heart disease*. Evidence-Based Complementary and Alternative Medicine (2015), Article No. 824850.

[6] X. Hao, H. An, L. Zhang, H. Li, G. Wei: *Sentiment diffusion of public opinions about hot events: Based on complex network*. Plos One *10* (2015), No. 10, e0140027.

[7] A. Imam, F. Anifowose, A. K. Azad: *Residual strength of corroded reinforced concrete beams using an adaptive model based on ANN*. International Journal of Concrete Structures and Materials *9* (2015), No. 2, 159–172.

[8] R. Wang, J. Feng: *Grid-based correlation localization method in mixed line-of-sight/non-line-of-sight environments.* KSII Transactions on Internet and Information Systems *9* (2015), No. 1, 87–107.

[9] I. M. Schmutte: *Job referral networks and the determination of earnings in local labor markets.* Journal of Labor Economics *33* (2015), No. 1, 1–32.

[10] E. T. Mohamad, D. J. Armaghani, E. Momeni, S. V. A. N. K. Abad: *Prediction of the unconfined compressive strength of soft rocks: A PSO-based ANN approach.* Bulletin of Engineering Geology and the Environment *74* (2015), No. 3, 745–757.

[11] X. D. Quintana, M. Arim, A. Badosa, J. M. Blanco, D. Boix, S. Brucet, J. Compte, J. J. Egozcue, E. de Eyto, U. Gaedke, S. Gascón, L. Gil de Solá, K. Irvine, E. Jeppesen, T. L. Lauridsen, R. López-Flores, T. Mehner, S. Romo, M. Søndergaard: *Predation and competition effects on the size diversity of aquatic communities.* Aquatic Sciences *77* (2015), No. 1, 45–57.

[12] U. Ryb, A. Matmon, Y. Erel, I. Haviv, A. Katz, A. Starinsky, A. Angert: *Controls on denudation rates in tectonically stable mediterranean carbonate terrain.* Geological Society of America Bulletin *126* (2017), Nos. 3–4, 2080–2084.

[13] J. S. Kim, K. S. Lee, W. J. Cho, H. J. Choi, G. C. Cho: *A comparative evaluation of stress–strain and acoustic emission methods for quantitative damage assessments of brittle rock.* Rock Mechanics and Rock Engineering *48* (2015), No. 2, 495–508.

[14] K. L. Aughenbaugh, T. Williamson, M. C. G. Juenger: *Critical evaluation of strength prediction methods for alkali-activated fly ash.* Materials and Structures *48* (2015), No. 3, 607–620.

Received May 7, 2017

Acta Technica *62* No. 2A/2017, 475–484

The key technology of fast stitching of remote sensing images of small unmanned aerial vehicles

Kai Ma[1,2], Qingqing Huang[1], Tiebo Sun[1], Tingting Sui[1], Jinhao Liu[1,3]

Abstract. The purpose is to study the key technology of fast stitching of small unmanned aerial vehicle remote sensing image. Based on the characteristics of small UAV images, the rapid image stitching software is designed. It has the capability of on-site processing. At present, the light and small UAV platform cannot get accurate images when shooting outside the elements. Therefore, when the collinearity equation is used as the UAV image geometric correction model, the calculation of the exterior orientation elements of the image needs to be taken into account. For a large number of ineffective image matching relationships, the Euclidean space distance relationship of camera GPS information is used to quickly construct latent image matching pairs, so as to speed up post-image matching. The results show that the proposed method can improve the overall operational performance. Therefore, it can be concluded that this method can solve the problem of UAV image stitching.

Key words. UAV images, Ransac, quick stitching.

1. Introduction

Because of its convenience and flexibility, UAVs are attracting more and more attention from all walks of life [1]. UAV can efficiently carry out research and practical work. In the remote sensing and mapping industry, UAVs can display their characteristics of maneuvering and fast. At present, UAV flight platforms are used to carry sensors, including card cameras, micro single cameras, measuring cameras, small LIDAR and hyperspectral cameras [2]. However, compared to the rapid development of sensor hardware and platform, post-processing software efficiency and processing capacity is obviously insufficient, especially on-site processing capacity [3]. Image mosaic is a method that combines many images together to form a spa-

[1]Beijing Forestry University, Beijing, China, 100083, China
[2]Xinjiang Agricultural University, Xinjiang, China, 830052, China
[3]Corresponding author

tially consistent image. This image can provide more information in time and space. The core of the entire mosaic algorithm is the registration algorithm between multiple images. In the automatic image processing process of optical images, a key link is the geometric correction of the image without artificial participation. It is directly related to the process of pre-processing and splicing images between the geometric accuracy.

In order to meet the requirements of geometric accuracy of actual unmanned image, it is very important to study geometric correction based on strict imaging model. At present, in the field of photogrammetry, splicing is generally done with high-performance workstations or servers [4]. Day is the processing unit time. Its processing capacity cannot meet the needs of the rapid response characteristics of the business needs, such as disaster relief, site mapping guidance. There are many problems. The degree of automation and processing efficiency are low, and it relies on more measurement conditions. As computer performance is improved, multi-core CPU and image parallel capability GPU are included. At the same time, some new algorithms are proposed in the field of computer vision. It provides the possibility to solve the above problems in UAV image mosaic. Therefore, on the basis of the above background, the small UAV image is designed for the overall scheme of the low electric control system, and the process of image mosaic processing is discussed. The core link is concerned.

2. State of the art

2.1. The basic principle

On the basis of the Lindeberg signal spatial scale theory, the extraction of Sift feature points is done under the establishment of multi-scale space [5]. First of all, the Gaussian pyramid image is constructed by using a series of Gaussian functions of scale factors. Then, the Gaussian difference pyramid image is obtained by subtracting the adjacent two Gaussian images. The local extreme point is detected as a candidate extremum in the Gaussian difference pyramid image and the corresponding scale is recorded. A three-dimensional quadratic function is used to fit the feature points to precisely determine the position and scale of the feature points. At the same time, the low contrast and unstable edge feature points are eliminated to enhance the stability of the feature points. Finally, the main direction of the feature points is determined by using the pixel gradient information in the neighborhood of the feature points. According to the main direction, the feature point neighborhood is rotated to reconstruct the 128-dimensional eigenvector, so as to ensure the invariance of the rotation of the feature. By using the Sift algorithm, reliable matching features can be extracted from images of different viewpoints in the same scene. These local feature descriptors have the following characteristics. It has invariance to target scale and rotation. There is a strong robustness to the difference of the shooting angle of the target, the affine transformation, the illumination changes and so on [6]. It has the characteristic of locality. To some extent, it can reduce the influence of Sift feature descriptors such as occlusion and complex background. The

uniqueness is high, and the amount of information is rich. It is suitable for accurate matching in large feature descriptor database. It can perform precise feature position localization, which can reach sub-pixel levels. In addition, it also has the characteristics of multi quantity, which can fully describe the image information [7].

2.2. GPU-Sift feature extraction

GPU (Graphic, Processing, Unit) is an arithmetic unit that different from CPU. Especially for graphics processing, it has its unique advantages. Today, CPU has entered the era of multi-computing nuclear. However, compared to GPU calculations, the number of cores is still very limited. For example, the Tesla C2050 of NVidia has 448 CUDA cores. Sinha [8] discusses all aspects of Sift feature extraction, which is distributed throughout the CPU-GPU computing framework. The basic process is depicted in Fig. 1.

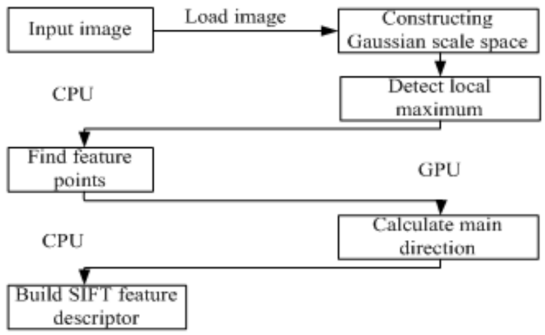


Fig. 1. GPU-Sift basic flow chart (GPU and CPU hybrid implementation)

Considering that all aspects of the Sift feature point extraction are not applicable to the parallel architecture of the GPU, the following steps are used to parallelize the GPU:

First, Gauss images and Gauss difference images are generated.

Second, feature point detection, which includes the extreme value of the scale space detection and feature points of the sub-pixel level positioning.

Third, calculate the main point of the feature point.

By using images of different sizes, the computing time of the main links in GPU-Sift is calculated. The computing platform is NVidia 8800GTX, 768 GB GPU memory, Win XP, Intel 3 GHz CPU.

3. Methodology

3.1. GPU-Sift feature matching

After determining the potential image matching pair, it is concerned that the Sift feature points between the two images are matched. Since the Sift feature matches the main operation time by retrieving the Eurasian distance from the KD tree to the

nearest two character descriptors, this step only completes the parallel processing of the GPU. CPU is computed to construct the KD tree. The experimental results show that the GPU-Sift matching algorithm can improve the speed of image feature extraction. In the case of ensuring the image matching rate is stable, the operation speed is increased by about 3 times. By using GPU, computing efficiency can be improved to a certain extent. However, it makes a certain request for GPU memory. So, when the image feature matches, the CPU version of the matching algorithm is first considered. Unless the time performance of this session becomes the bottleneck of the entire process, the GPU version matching algorithm is considered.

3.2. Sift feature matching based on Ransac

Random Sampling Consistency Ransac (Random Sample Consensus) is an iterative algorithm for estimating the parameters of a mathematical model. The main idea of Ransac is to solve the parameters of the mathematical model that most samples (feature points) can satisfy by sampling and validating strategies. In the iteration, the minimum number of samples and the parameters of the model are needed to collect from the data set at a time. Then, the number of samples that match the parameters of the model is counted in the data set. The most sample parameters are considered to be the final model parameter values. The sample point of the model is called the Inliers, and the sample point that does not conform to the model is Outliers.

For Sift feature matching, the misaligned pair is Outliers. The algorithm flow is as follows:

First, select a pair of match pairs (5 points) randomly from the initial match point set.

Second, using the selected matching point to calculate the essential matrix model parameters between the two images.

Third, by using the distance threshold, the interior point that matches the model is calculated. The ratio of the number of points to the total number of points is counted, and the maximum number of iterations is updated accordingly.

Fourth, repeat first and second steps until the maximum number of iterations is reached, or enough of the Inliers points are found.

Fifth, through the above process, the essential matrix between the two images is determined. Based on this essential matrix, the matching pair of all Sampson errors is greater than the threshold, leaving the remaining Inliers pair. Eventually, the purification of the match point is completed.

3.3. The design of PLC module

The auxiliary electrical control system's lubrication system has two inputs, which are two alarm outputs. The cooling system also has two inputs for both alarm outputs. The automatic tool has 6 inputs and 6 alarm outputs. The lighting and semaphore system has an input and four outputs. The tool change system has 5 inputs and 5 outputs. Therefore, in the choice of PLC, taking into account the

remaining 1/3 of the margin, Huichuan PLC was used. It has 24 inputs and 15 outputs. The scanning frequency is 1000 Hz, which fully meets the requirements.

4. Result analysis and discussion

4.1. Camera model

The camera's basic imaging model is often referred to as a small aperture imaging model. It is given by the three-dimensional space to the center of the plane projection transformation. The spatial point C is the projection center (the site), and the distance from the main point p is f (focal length). The projection of the spatial point C on the image plane is the intersection of the ray and the image plane at the point C and the point X. The results are shown in Figs. 2 and 3.

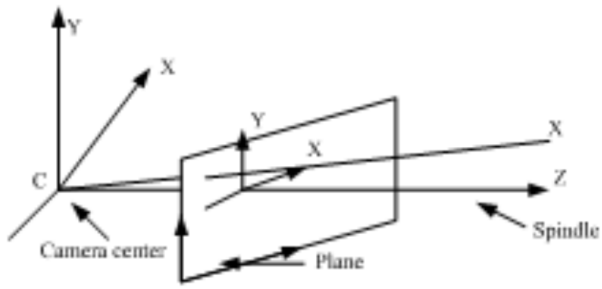


Fig. 2. Camera small hole imaging model 1

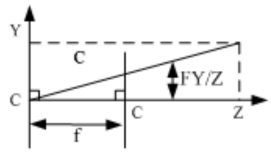


Fig. 3. Camera pinhole imaging model 2

To better describe the projective relation from algebra, three coordinate systems are established, which are camera space, Cartesian coordinate system, image plane coordinate system and pixel plane coordinate system. In practical applications, the upper left corner of the image (the upper left corner of the pixel in the upper left corner) is the pixel plane coordinate origin (0,0), and the horizontal line to the right and the vertical line down is the direct direction of the U axis and the V axis. On the image plane, the image plane coordinate system is established in the positive direction of the x-axis and the y-axis along the u-axis and v-axis positive directions of the pixel coordinate system, respectively, as the coordinate origin of the image plane coordinate system. In the space, the camera projection center C is the coordinate

origin of the camera space Cartesian coordinate system, and the direction of the spindle pointing direction coincides with the positive direction of the Z_{c} axis. With the main axis as Z_{c}, the camera space Cartesian coordinate system (the right-hand space Cartesian coordinate system) is established in the positive direction of X_{c} and Y_{c} in the positive and negative directions of the plane coordinate system. According to the principle of triangular similarity, combined with the above three coordinate system, the basic imaging model of the camera is described as follows:

$$u = x + u_0,\ v = y + v_0\,, \tag{1}$$

$$x = \frac{f x_{\mathrm{c}}}{z_{\mathrm{c}}},\ y = \frac{f y_{\mathrm{c}}}{z_{\mathrm{c}}}\,. \tag{2}$$

The corresponding matrix of the above formula is

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} X_{\mathrm{c}} \\ Y_{\mathrm{c}} \\ Z_{\mathrm{c}} \\ 1 \end{bmatrix} \tag{3}$$

In the above equations, (u, v), (x, y), (X_c, Y_c, Z_c) correspond to coordinates in the pixel plane coordinate system, image plane coordinate system and camera space Cartesian coordinate system. (u_0, v_0) is the coordinate value of the coordinate system as the main point p in the pixel plane coordinate system. f is the camera focus value. According to the triangular similarity principle, the above formula is obtained. It is also expressed in the case of adding a nonzero scale factor. Therefore, (X_c, Y_c, Z_c) in the above formula can be different from other variables.

4.2. Two-view geometry and three-view geometry

The two-view geometry is used to describe the polar geometric relationships between the two images, which express the inherent projective properties between the two images. It does not involve any geometric structure of the photographic scene. It is only related to the camera's inner and outer azimuth elements. The matching relationship between the two-view geometric description is shown in Fig. 4.

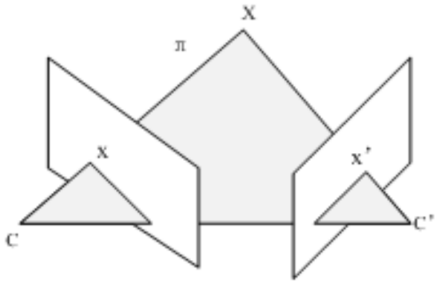


Fig. 4. Point matching relationships under two-view geometric descriptions

The three-view geometry is used to describe the polar geometries between three images, which express the inherent radio graphic properties of the three images. As with the two geometries, the three-dimensional geometry does not involve any geometric structure of the photographic scene. It is only related to the camera's inner and outer azimuth elements. The point matching relationship under the three-pole geometric description is shown in Fig. 5.

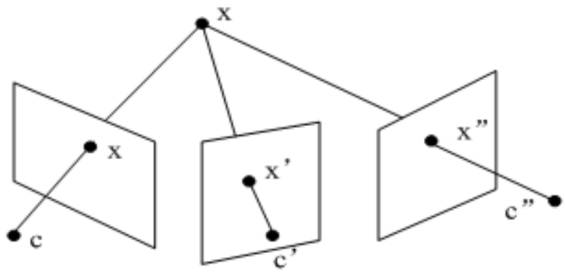


Fig. 5. Point matching relationship under three-view geometric descriptions

Two geometric geometries describe the geometric relationships between the two mating points using the basic matrix. However, when a point on one of the images is given, the specific position on the other image cannot be determined by the basic matrix. It can only be determined on a straight line through the pole. When an object point has three image points corresponding to it, the triplet tensor is used to describe it. When the object points are at known points, the corresponding image point position on the third image can be uniquely determined. Therefore, based on this strong constraint, we can remove some of the misplaced points in the two images, increasing the robustness of the existing model.

In the case of known azimuthal elements and camera lens distortion, there are many ways to solve the rotation matrix and the translation vector transformation between the two-view model and the three-view model. For example, through the relationship between points and points, triplet tensor is established. Then, the rotation matrix and the translation vector in the triplet are decomposed. According to the relative rotation and translation relationship between the two images, the rotation matrix and the translation vector between the triads are directly determined by matrix transformation. As with the two-dimensional geometric description, the amount of the three-dimensional geometry is also different from the overall scale factor.

4.3. Multi-view geometry

In the camera's inner azimuth elements and lens distortion, the problem is to use multiple two-dimensional images in Euclidean space to reconstruct the image corresponding to the three-dimensional point of the scene. Through the geometric characteristics of two views and three views, the rotation matrix, translation vector and projective depth of all three tuples can be computed from the image set by using the matching point pair relation between images. Although the relationship

between the three tuples and the three tuple is uncertain, it effectively removes the external points and provides reasonable input for further multiview computation.

In photogrammetry, the nonlinear beam method is used to compute the exterior orientation elements of the camera and the coordinates of the points in the scene. Although this method relies heavily on reasonable initial values, it can obtain stable solutions in the case of good initial values. According to this, the larger part of the residual point can be removed so that the reliability of the subsequent treatment can be improved. At present, in computer vision, there are two main strategies for providing reasonable initial values for nonlinear beam method adjustment: one is incremental beam adjustment, and the other is global beam adjustment.

5. Conclusion

For low-level light and small UAV images, the stability of the platform is poor when acquiring images. It is susceptible to external weather conditions. The three rotation angles of the image are relatively large, and geometric deformation is particularly sensitive to ground elevation changes. When the size of the survey area image is particularly large, the geometric correction error generated by the approximate polynomial correction model is amplified. It is difficult to meet the requirement of image geometric correction precision when the image is spliced. At present, the light and small UAV platform cannot get accurate images when shooting outside the elements. Therefore, when using the colinearity equation as the geometric correction model of the UAV image, it is necessary to consider how to calculate the external azimuth elements of the image. On the basis of previous studies, aiming at the characteristics of remote sensing images of low - level light and small unmanned aerial vehicles (UAVs), the key technologies of UAV images are researched and analyzed. Combined with three parameters of Ransac's influence performance and iteration times, the parameters of Sift characteristic matching relation for high - resolution UAV images are given. It provides technical support for its efficient and fully automated process. Finally, the rapid processing software for remote sensing images of low altitude light and small UAVs is implemented.

References

[1] Y. Sun, K. Hara: *Fortran code of the projected shell model: Feasible shell model calculations for heavy nuclei*. Computer Physics Communications *104* (1997), Nos. 1 to 3, 245–258.

[2] Z. Zhou, M. Yan, S. Chen, Y. Lan, X. Luo: *Image registration and stitching algorithm of rice low-altitude remote sensing based on Harris corner self-adaptive detection*. Transactions of the Chinese Society of Agricultural Engineering *31* (2015), No. 14, 186–193.

[3] W. Yu, Y. Tao, A. X. Donghai: *Fast automatic stitching for images of unmanned aerial vehicle in emergency response*. Journal of Computer-Aided Design and Computer Graphics *25* (2013), No. 3, 410–416.

[4] E. Faleiro, U. Kuhl, R. A. Molina, L. Muñoz, A. Relaño, J. Retamosa: *Power*

spectrum analysis of experimental Sinai quantum billiards. Physics Letters A *358* (2006), No. 4, 251–255.

[5] Z. Li, Z. Chen, L. Wang, J. Liu, Q. Zhou: *Area extraction of maize lodging based on remote sensing by small unmanned aerial vehicle*. Transactions of the Chinese Society of Agricultural Engineering *30* (2014), No. 19, 207–213.

[6] I. Nitze, B. Barrett, F. Cawkwell: *Temporal optimisation of image acquisition for land cover classification with random forest and MODIS time-series*. International Journal of Applied Earth Observation and Geoinformation *34* (2015) 136–146.

[7] M. Z. Raqab: *Optimal prediction-intervals for the exponential distribution, based on generalized order statistics*. IEEE Transactions on Reliability *50* (2001), No. 1, 112 to 115.

[8] S. N. Sinha, J. M. Frahm, M. Pollefeys: *Feature tracking and matching in video using programmable graphics hardware*. Machine Vision & Applications *22* (2011), No. 1, 207 to 217.

Received May 7, 2017

Acta Technica *62* No. 2A/2017, 485–494

Harmonic measurement of power system based on artificial neural unit network

Zhaojun Wang[1], Jiao Zhen[2], Hongxia Guo[1], Zhelong Wang[1], Fucun Li[1,3], Lijun Liu[1]

Abstract. The purpose is to study the harmonic measurement method of power system based on artificial neural network. Based on the previous harmonic measurement method, the harmonic measurement method based on neural network is studied. The application of adaptive neural network in harmonic measurement is explored. WATLAB simulation software is used to simulate and validate the algorithm. For adaptive neural networks, a single neuron-based network architecture is used. The results show that the method is of higher adaptability and accuracy for the unknown power network with harmonic source. However, for changing harmonic sources, it has poor real-time performance. Therefore, it can be concluded that the proposed method is only suitable for the power systems with unknown harmonic sources.

Key words. Neural network, harmonic measurement, simulation verification.

1. Introduction

With the development of power electronics technology, the wide application of power electronic devices has brought serious harmonic pollution to the power system [1]. In the transportation, metallurgy, chemical and other industrial transportation fields, all kinds of power electronic equipment are widely used, which makes the harmonic problems in the power grid become increasingly serious. Many power electronic devices with low power factor bring extra burden to the grid, which affects the quality of the power supply. Harmonic pollution of power electronic devices has become a major obstacle to the development of power electronic technology. The problem of harmonic measurement and control is imminent [2]. Therefore, the research of harmonic measurement is an important work in power system analysis

[1]State Grid Shandong Electric Power Research Institute, Jinan, Shandong, 250002, China
[2]State Grid Shandong Electric Wucheng Power Supply Company, Dezhou, Shandong, 253300, China
[3]Corresponding author

and control. It is an important prerequisite for relay protection, fault location and fault types. It is a meaningful work to study the method of harmonic measurement. At present, in the power system, the harmonic measurement methods are Fourier transform, instantaneous reactive power, neural network and wavelet transform. By collecting data, the previous harmonic measurement method is analyzed and summarized. Aiming at the shortcomings of these methods, a harmonic measurement method based on neural network is proposed. Then, MATLAB simulation software is used to simulate and analyze different neural network algorithms [3].

2. State of the art

The neural networks have been studied for thirty years. As early as 1957, F. Rosenblatt proposed and designed the famous perceptron concept for the first time. It consists of threshold neurons, which attempt to simulate the perception and learning functions of animals and humans [4]. In 1991, P. Saratchandran proposed a neural network method based on BP algorithm. However, this method requires a large number of samples to train learning weights. The speed of convergence and the various situations of different constraints are not discussed. It just gives the general situation of the algorithm. The improved algorithm has been pursued by many scientists [5]. In 1989, Moody and Darken proposed the neural network algorithm of RBF, which brought new life to the research and application of neural networks. In addition to the breakthroughs and advances in neuroscience research, the main reason is the need for the development of computer science and artificial intelligence [6]. From the point of view of artificial neural networks, a unified and complete theoretical system has not yet been formed. The formation, construction, design and performance evaluation of various network models and algorithms are only specific analysis of specific problems, which mainly rely on the results of computer simulation. It cannot give a strict and scientific general law and method. On the basis of analyzing the basic principle of harmonic measurement device, an adaptive neural network is used to measure the harmonics, and a network and algorithm based on adaptive harmonic measurement are constructed [7]. Then, the validity of the network is verified by simulation software. The radial basis function neural network is used to measure the harmonics. Through the determination of the algorithm, an example is selected for simulation analysis [8].

3. Methodology

3.1. Neural network model

The basic unit of the neural network is neurons. It is widely interconnected by a large number of neurons [9]. Artificial neural network is the simplification and simulation of biological neurons, which is the basic processing unit of neural network. It is a nonlinear component with multi-input and single output. Its input-output

relation can be described as

$$I_j = \sum_{n}^{j=1} w_j x_j - \theta_j \,, \tag{1}$$

$$y_i = f(I_i) \,. \tag{2}$$

Here, $x_j,\ j = 1, 2, \ldots, n$ is an input signal from other cells. Symbol θ_j denotes the bias (threshold) of the neuron unit, and w_j represents the connection weight from cell j to cell i. Symbol n denotes the number of input signals, y_i is the neuron output, and t is the time. Function F is a transfer function, which is sometimes called an incentive function. It often employs 0 and 1, two-valued functions, or S type functions. The transfer function can be a linear function, but it is usually a nonlinear function like a step function or a S-shaped curve. The commonly used neuron nonlinear functions are as follows:

Threshold type function: when y_i takes values 0 or 1, $f(x)$ is a step function.

$$f(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{3}$$

The S-curve is usually a monotonically differentiable function of continuous values within (0,1) or (-1,1). It is commonly used index or tangent and other types of S-shaped curve to represent, such as

$$f(x) = \frac{1}{1 + \exp(-\beta x)},\ \beta > 0 \tag{4}$$

or

$$f(x) = \tanh(x) \,. \tag{5}$$

In the neural network composed of RBF (Radial Basis Function), the structure of neurons can be described by Gaussian function

$$y_i = \exp\left[-\frac{1}{2\sigma_i^2} \sum_j (w_i - w_{jt})^2 \right] . \tag{6}$$

Here, σ_i^2 is a standardized parameter.

The neural network is the nerve structure, which consists of many neurons interconnected together. The neural network model can be obtained by mathematically modeling the relationship between neurons. The application of neural network in harmonic measurement mainly involves the following aspects. The problems are the input method, the construction of the network structure, the determination of the input layer and the determination of the hidden layer [10]. These are the problems that must be considered in the application of neural network theory to harmonic measurement. It will directly affect the speed of convergence. That is to say, it will affect the real-time measurement of harmonics and the accuracy of harmonic mea-

surement. Theoretically, to make the harmonic measurement circuit has a certain accuracy, the measurement should be aimed at different targets [11]. according to the actual waveform distortion of collecting training samples, the harmonic measuring circuit can suitable for measuring a (or several) generated by the nonlinear load harmonic distortion wave component. However, the distortion waveforms in the power system are different from each other, and the harmonic components are different. It is impossible to collect all types of distortion waveforms. Therefore, the training samples should be extracted theoretically [12]. At the same time, in selecting the learning algorithm of the network, the method of higher efficiency and accuracy should be considered, in order to meet the real-time and high precision requirements of harmonic measurement. The construction of neural network has a direct influence on the speed of convergence. Therefore, on the basis of convergence, the network structure should be chosen as simple as possible. By using this method, the real-time requirement of harmonic measurement can be satisfied [13].

3.2. The adaptive noise canceling technology

The adaptive noise cancellation method is a kind of signal detection method proposed by Widrow in the United States. It can separate a signal from additive noise. The principle shown in Fig. 1.

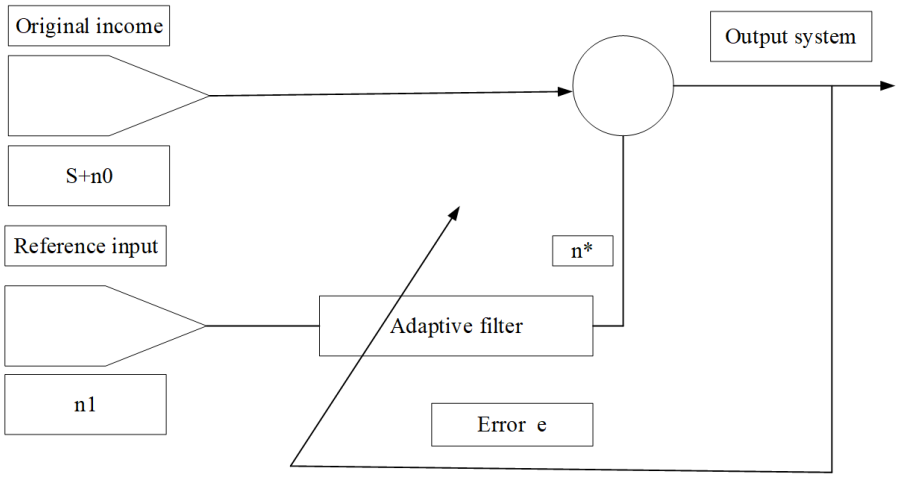


Fig. 1. Adaptive cancellation scheme

The detection system has two inputs. They are the original input $s + n_0$ and the reference input n_1. The symbols s and n_0 are irrelevant, and s and n_1 are not related. However, n_1 and n_0 are related to noise interference. The function of the reference channel is to detect interference. The output n_1 is adjusted by the adaptive filter, so that it is closest to the main channel n_0 in the sense of minimum mean square error. Here, the system output y is also used as the error signal e to adjust the parameters of the adaptive filter. This method requires little or no need for a priori knowledge of signals and noise. The n_0 can be estimated by an adaptive filter,

so that the output of the system is s.

3.3. The harmonic measurement principle based on adaptive filtering

Taking single-phase circuit as an example, this paper expounds the principle of applying adaptive filter to harmonic measurement, and it is also suitable for three-phase circuit. Assuming that the single-phase voltage is $U_{\mathrm{N}}(t) = U_{\mathrm{N}} \sin \omega t$, the non-linear load of the non-sinusoidal current can be expressed as Fourier series in the form

$$I_1(t) = I_1 \sin(\omega t + \phi_1) + \sum_{n=2}^{\infty} I_n \sin(\omega t + \phi_n) = i_i(t) + \sum_{n=2}^{\infty} i_n(t)\,. \tag{7}$$

$$I_n(t) = I_n \cos \phi_n \sin n\omega t + I_n \sin \phi_n \cos n\omega t = i_{ns}(t) + i_{nc}(t),\ n > 1\,. \tag{8}$$

In the above formulas, i_1 and i_n are the fundamental active current and fundamental reactive current, respectively. Symbols i_{ns} and i_{nc} denote the sine and cosine components of the nth harmonic, respectively. The adaptive noise cancellation method is used for harmonic measurements. Quantity i_1 is used as the original input. If $i = i_1 + i_2 + ... + i_n$ is taken as the "noise interference" current, then the other higher harmonics of the total current i need to detect the "signal". Quantity i is not related to i_n. Taking the fundamental sine and cosine signal $\sin \omega t \cos \omega t$ and their $2, 3, ..., n$ multiples of the frequency harmonic as a reference input, they are related to the sine and cosine components corresponding to the "noise interference" current i, but are not related to other higher harmonic currents i_h. Thus, the components of the "noise interference" current I and the approximation values in the sense of the minimum mean square error of the "signal" i_h can be obtained by the multi-channel adaptive filter [14].

3.4. Realization of adaptive filter with neurons

The individual neurons have a certain degree of adaptive and self-learning ability, and the circuit structure is simple. Therefore, a single neuron circuit can be used to implement an adaptive filter to construct an adaptive neural network for harmonic measurement, so that the output can be approximated by the "mean" error to the "signal" to be detected [15]. The sine and cosine signals $\sin \omega t \cos \omega t$ and their $2, 3, ..., n$ multiples are taken as the input of a single neuron, so the input vector $x(t)$ of the neuron is

$$X(t) = [\sin \omega t \cos \omega t, \sin 2\omega t \cos 2\omega t, ..., \sin n\omega t \cos n\omega t]\ . \tag{9}$$

The net input $S(t)$ of the neuron is

$$S(t) = \sum_{i=1}^{n} w_i x_i(t) - \theta_i\,. \tag{10}$$

In the formula, w_i is the connection weight and θ_i is the neuron threshold. The output of the neuron $y(t)$ is

$$y(t) = f\left[s(t)\right] . \tag{11}$$

In the formula, $f(x)$ is the activation function. From (8) and (9), it can be seen that the fundamental and harmonic components are linear combinations of their sine and cosine signals. Thus, the activation function $f(x)$ takes the linear function $f(x) = x$. In addition, since the power system transmission bus does not contain the DC component, the neuron threshold value θ_i is zero. The current sampling value is taken as the target value of the neural network, that is, the difference between the current sampling value $iL(t)$ and the output value $y(r)$ of the neuron is taken as an error $iL(t) - y(t)$. Function $e(t)$ is used to constantly adjust the input weights of neurons. According to the learning rules of neurons, each weight can get the optimal value, so that its mean square error is the smallest. Let the output $y(t)$ of the neuron approach the "noise interference" current i, the error approximates the "signal" i_h, thus, the function of the adaptive filter for harmonic measurements is realized [16].

4. Result analysis and discussion

The algorithm of power harmonic detection based on adaptive neural network is simulated by Matlab software. In the simulation study, the neural network is used to measure the fundamental and 3, 5, 7, 11 harmonic components [17]. Therefore, the reference inputs are $\sin\omega t$, $\cos\omega t$, $\sin 3\omega t$, $\cos 3\omega t$,.....,$\sin 11\omega t$, $\cos 11\omega t$, 10 input vectors, respectively. The LMS algorithm is used to study the network online [18]. The simulation flow chart is shown in Fig. 2.

From the simulation results, it can be seen that the network has good real-time performance [19]. Under the simulation conditions in this paper, the upper target value can be tracked in about two cycles. At the same time, in order to analyze the measurement accuracy of the circuit, Fig. 3 lists the fundamental frequencies at the end of 1 cycle and 2 cycles at the end of the harmonic value of the measurement error [20].

Figure 4 lists the fundamental frequency at the end of the two cycles and three cycles (extended 100 times) at the end of the harmonic phase angle of the measurement error.

As can be seen from the Fig. 4, at the end of the 2 cycles, the accuracy of each harmonic amplitude measured by the circuit is higher, and they are not less than 1 %. The phase angle measurement accuracy is slightly lower, the maximum is 12.27 %. However, the accuracy of the phase angle at the end of the third cycle is also high, and it reaches 0.005 %. Therefore, it can be said, that it exhibits high measurement accuracy simultaneously.

The harmonic measurement circuit is constructed by using adaptive neural network. The circuit detection method for measuring arbitrary harmonics is given, and the network with 3, 5, 7 and 11 harmonics is simulated by MATLAB. The simulation results show that the circuit has good real-time performance, measurement accuracy and adaptability when using adaptive neural network for harmonic

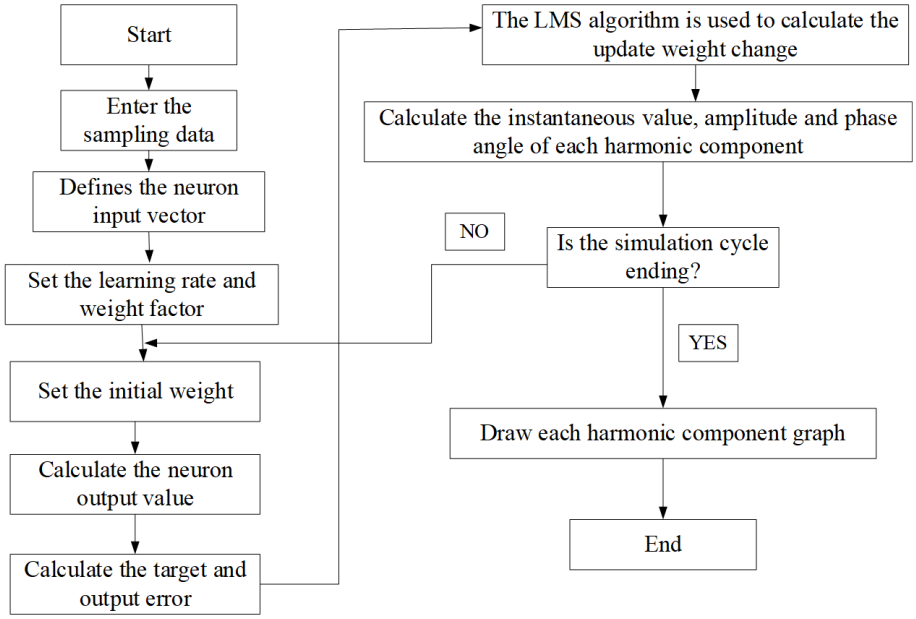


Fig. 2. Simulation flow figure

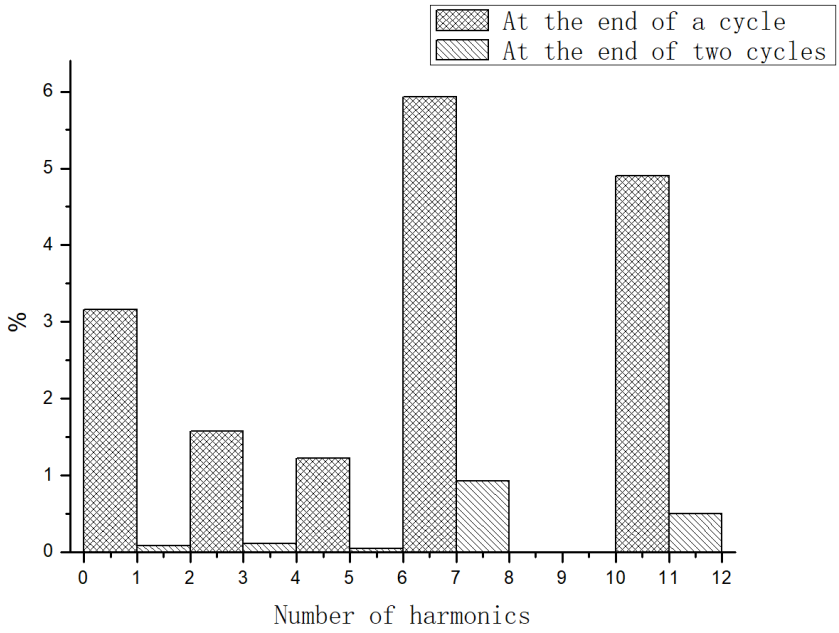


Fig. 3. Measurement error of the amplitude of each harmonic wave at the end of 1 cycle and at the end of two cycles

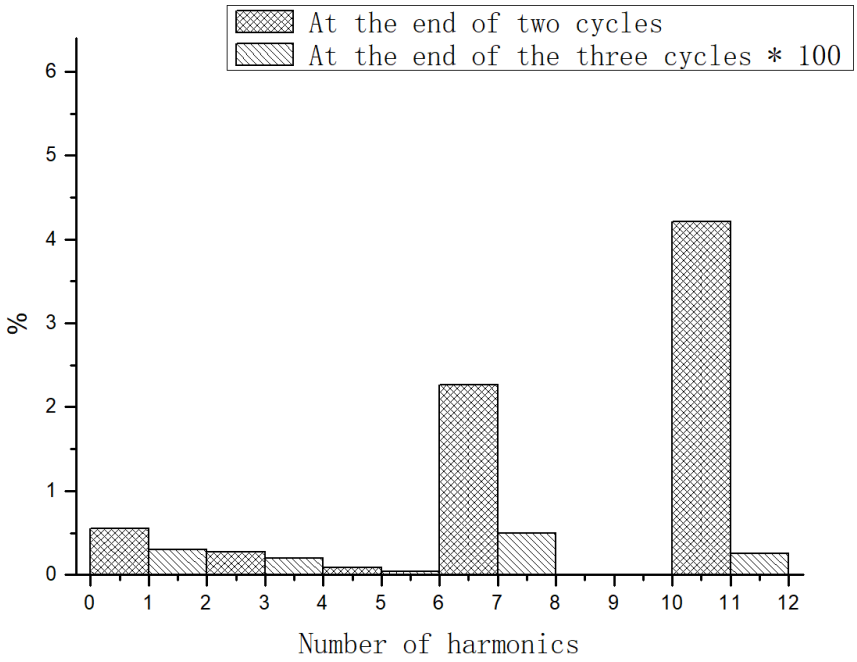


Fig. 4. Two cycles and three cycles (extended 100 times) at the end of the harmonic phase angle of the measurement error

measurement. Adaptive neural network based on single neuron structure is a good method for power network with unknown harmonic source. It is not only simple in structure, but also has certain adaptive ability. For power systems with large load variation, the real-time performance of this method needs further improvement. At last, the simulation test of random distortion current shows that the network has the longest tracking time, and it reaches about 0.08 s.

5. Conclusion

The adaptive neural network based on a single neuron has the characteristics of simple structure and strong network adaptability. For the measured power network, it does not require pre-training. The results show that the proposed method is not only simple in circuit structure, but also has better real-time performance and higher accuracy for power network with unknown harmonic source. However, there is a real-time difference in the large harmonic source, which requires a tracking time of 0.06 to 0.08 s. Therefore, the method is suitable for harmonic sources in unknown power systems. In order to better meet the real-time and measurement accuracy of harmonic measurement, a method based on multilayer feedforward neural network for harmonic measurement is further studied. In the future, the harmonic measurement algorithms should be developed from simple function analysis methods to complex

numerical analysis and signal processing. At the same time, these algorithms should be intelligent. Therefore, it is urgent to establish a general power theory and apply the new theory to harmonic measurement. In this way, harmonic measurement can achieve a breakthrough in real-time and accuracy.

References

[1] C. B. Khadse, M. A. Chaudhari, V. B. Borghate: *Conjugate gradient back-propagation based artificial neural network for real time power quality assessment.* International Journal of Electrical Power & Energy Systems *82* (2016), 197–206.

[2] M. Qasim, K. Parag, V. Khadkikar: *Artificial-neural-network-based phase-locking scheme for active power filters.* IEEE Transactions on Industrial Electronics *61* (2014), No. 8, 3857–3866.

[3] D. Borkowski, A. Wetula, A. Bień: *Contactless measurement of substation bus-bars voltages and waveforms reconstruction using electric field sensors and artificial neural networks.* IEEE Transactions on Smart Grid *6* (2015), No. 3, 1560–1569.

[4] M. Qasim, V. Khadkikar: *Application of artificial neural networks for shunt active power filter control.* IEEE Transactions on Industrial Informatics *10* (2014), No. 3, 1765–1774.

[5] L. Saribulut, A. Teke, M. Tümay: *Artificial neural network-based discrete-fuzzy logic controlled active power filter.* IET Power Electronics *7* (2014), No. 6, 1536–1546.

[6] C. I. Chen, Y. C. Chen: *Comparative study of harmonic and interharmonic estimation methods for stationary and time-varying signals.* IEEE Transactions on Industrial Electronics *61* (2014), No. 1, 397–404.

[7] A. Guellal, C. Larbes, D. Bendib, L. Hassaine, A. Malek: *FPGA based on-line artificial neural network selective harmonic elimination PWM technique.* International Journal of Electrical Power & Energy Systems *68* (2015), 33–43.

[8] H. Rong, J. Lv, C. Peng, L. Zou, Z. Ma, Y. Chen, Y. Zhu: *Dynamic regulation of the weights of request based on the Kalman filter and an artificial neural network.* IEEE Sensors Journal *16* (2016), No. 23, 8597–8607.

[9] F. Bai, X. Wang, Y. Liu, X. Liu, Y. Xiang, Y. Liu: *Measurement-based frequency dynamic response estimation using geometric template matching and recurrent artificial neural network.* CSEE Journal of Power and Energy Systems *2* (2016), No. 3, 10–18.

[10] D. Fan, L. Wei, M. Cao: *Extraction of target region in lung immunohistochemical image based on artificial neural network.* Multimedia Tools and Applications *75* (2016), No. 19, 12227–12244.

[11] E. Deniz, O. Aydogmus, Z. Aydogmus: *Implementation of ANN-based selective harmonic elimination PWM using hybrid genetic algorithm-based optimization.* Measurement *85* (2016), 32–42.

[12] J. C. Yin, N. N. Wang, J. Q. Hu: *A hybrid real-time tidal prediction mechanism based on harmonic method and variable structure neural network.* Engineering Applications of Artificial Intelligence *41* (2015), 223–231.

[13] V. J. Vijayalakshmi, C. S. Ravichandran, A. Amudha, A. M. Ibrahim: *Dual phase analysis based linear regression trained neural network for selected harmonic elimination in a multilevel inverter.* International Journal of Applied Engineering Research *10* (2015), No. 3, 6651–6658.

[14] S. Raza, H. Mokhlis, H. Arof, J. A. Laghari, L. Wang: *Application of signal processing techniques for islanding detection of distributed generation in distribution network: A review.* Energy Conversion and Management *96* (2015), 613–624.

[15] M. Li, X. Huang, H. Liu, B. Liu, Y. Wu, L. Wang: *Solubility prediction of supercritical carbon dioxide in 10 polymers using radial basis function artificial neural*

network based on chaotic self-adaptive particle swarm optimization and K-harmonic means. RSC Advances *5* (2015), No. 56, 45520–45527.

[16] B. S. D. Nugraha, F. Sugarizka: *Non-contact measurement of blood glucose based on artificial neural network*. International Journal of Computer Applications *76* (2013), No. 13, 33–36.

[17] G. Tuna, D. Savran, R. Das, S. Basarici, Y. Kilicaslan: *Evaluation of the effects of measurement interval on artificial neural network-based prediction for wireless water quality monitoring network*. International Journal of Computer Networks and Applications *4* (2017), No. 3, 77–83.

[18] F. Bai, X. Wang, Y. Liu, X. Liu, Y. Xiang, Y. Liu: *Measurement-based frequency dynamic response estimation using geometric template matching and recurrent artificial neural network*. CSEE Journal of Power and Energy Systems *2* (2016), No. 3, 10–18.

[19] X. Su, Y. Wu, H. Pei, J. Gao, X. Lan: *The prediction of coke yield of FCC unit using different artificial neural network models*. China Petroleum Processing & Petrochemical Technology *18* (2016), No. 3, 102–109.

[20] M. Lo Schiavo, B. Prinari, J. A. Gronski, A. V. Serio: *An artificial neural network approach for modeling the ward atmosphere in a medical unit*. Mathematics and Computers in Simulation *116* (2015), No. C, 44–58.

Received May 7, 2017

Acta Technica *62* No. 2A/2017, 495–508

Design and research of wireless vibration signal detecting system for SCM and Bluetooth transmission

Ze-rui Song[1], Heng Zhang[1]

Abstract. Sensor signal is transmitted by cable or optional fiber in the traditional vibration detection. But on an occasion with different wiring arrangement, wireless method should be taken to transmit the signal. Bluetooth technology is selected as a plan for vibration data transmission, and wireless vibration signal detecting system is also designed that is based on SCM and Bluetooth transmission. In hardware design, STC microcontroller is adopted as micro-control center for signal acquisition. Vibration signal is first picked up by sensor, then enlarged and filtered the circuit to transmit to A/D converter. Finally A/D converter is controlled by vibration signal with SCM to finish the acquisition of vibration signal. Digitized signal is made a wireless launch through Bluetooth protocol communication. At the receiving end, Bluetooth transmits the signal that is accepted by the USB interface circuit to the PC computer; in the software design, C programming language is first used to finish the design that is made by SCM to signal acquisition and transmission procedure. And LabVIEW development platform is chosen to research and develop human-computer interaction interface, which can achieve the control of SCM and real-time display receiving data, and the make an analysis about it; at last, the reliability and stability of the wireless vibration signal detecting system is verified by building test experimental platform. With the vibration frequency of vibrostand increased, the relative error of vibration frequency gained from wireless vibration detecting system is also increased, but the relative error is less than 5 %. Thus, the designed wireless vibration detecting system, with higher reliability and stability, basically satisfies the used demand.

Key words. SCM, Bluetooth wireless communication technology, vibration signal, Lab VIEW.

1. Introduction

Generally speaking, the traditional vibration detection method is that the sensor placed at test points and the exported signal of sensor is transmitted to data processing center through relevant cable or optical fiber, then the gained data is analyzed and processed to achieve the detection to the vibration [1–3]. Although

[1]Shandong University of Science and Technology Chian, 266590, China

the traditional vibration detection system is more mature with a wide application range, some inadequacies of this method are still found in the functional application. For example, in the case of the vibration detection in some areas such as rotation components, explosion, nuclear power plant and so on, the vibration signal gained by sensor is hard to be transferred by cable [4, 5].

Besides, what traditional medical facility does during vibration measurement of physiological parameters is, that it transmits the physiological signal to medical facility by cable [6]. This wired method not only makes people nervous and uncomfortable, but also limits people's activities. When people are not in a natural state, human health condition cannot be truly reflected by the gained physiological parameters [7]. For the wired method transmits signal, wireless method is adopted to transmit signal can not only leave out the waste-time-and-energy cable arrangement, but also can resolve the problem that the cable cannot be arranged in the spot, and can move flexibly and conveniently [8]. A wireless vibration sensing system with small size, low cost and high practicability is expected to be developed

2. Literature review

According to the length of wireless transmission distance, it can be classified into two kinds of transmissions: short distance and long distance [9]. At present, there mainly have several popular short distance wireless transmission technologies: infrared, Zig Bee, WLAN, Bluetooth, UWB, FR and other wireless transmission technologies [10, 11]. Infrared transmittal rate is rather higher, but its transmission distance is short (1–3 m) and its angle is limited [12]. Zig Bee not only has many advantages such as low power dissipation and cost, but it also can make a wireless transmission with bi-directional data [13]. Modules needed by WLAN are expensive, which strictly limits its application range. While transmission rate of Bluetooth technology is higher than Bee and its cost is lower than WLAN, so Bluetooth is selected as a plan for vibration data wireless transmission [14, 15].

Wireless detection technology can make up the deficiencies of traditional wired detection technology, especially on some occasions like inconveniently using cable or not using cable transmission, which remarkably show wireless transmission's advantages. According to abroad and domestic researches, wireless technology is mainly applied to industrial control and environmental monitoring, and most of the monitoring objectives are temperature, humidity and illumination intensity in the environment and other slowly changing signals. While for the vibration wireless detection is mainly in low-frequency vibration of bridges, buildings, ocean platforms and other large-scale architectures. With more and more concern about health, wireless technology has a wide prospect in medical treatment and health. Therefore, a wireless vibration detection system based on Bluetooth transmission is designed.

3. Research methods

3.1. Hardware design

Bluetooth wireless communication technology is adopted and a plan is put forward based on Bluetooth transmission for wireless vibration detection system. The whole structure is shown in the following Fig. 1.

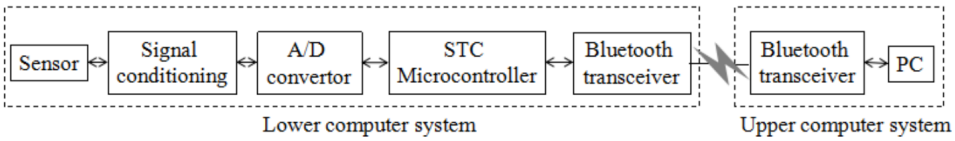


Fig. 1. Whole block diagram of system hardware

From Fig. 1, this wireless vibration detection system consists of upper computer system and lower computer system. The upper computer system hardware is mainly composed of PC computer and Bluetooth. HC05 Bluetooth module is adopted in this research. PC computer is linked with Bluetooth by USB serial interface circuit. And based on the UI of wireless vibration detection system that is designed by Lab VIEW platform in the upper computer system, according to the UI, real-time display and vibration data's storage of vibration signal not only can be achieved, but also the Fourier transform of vibration signal can be analyzed and also can achieve the control of the lower computer system.

The lower computer system mainly consists of SCM, signal processing circuit and Bluetooth and others. So acquisition and wireless dispatch of vibration signal can be achieved.

3.1.1. Design of vibration signal acquisition circuit. Signal processing circuit can enlarge the small signal exported by sensor and have smoothing function. STC89C52 SCM is adopted as microcontroller. According to the design of its peripheral circuit and the software design, A/D convertor and data acquisition can be achieved to control. And then the collected vibration data is transmitted to Bluetooth by UART. Finally, vibration signal is made a wireless launch by Bluetooth.

SCM peripheral circuit generally includes four parts: SCM chip, clock circuit, reset circuit and power source, which means the smallest peripheral configuration can make SCM work normally. Power source adopts 5V direct-current main. Circuit diagram is shown in the following Fig. 2.

Most of detection systems' microcontroller only can receive and process digital signals. While the commonly used output signals of sensor is generally the analog signals. So before inputting these analog signals to the microcontroller, it should be converted into digital signals by analog-digital conversion chip. This system adopts ADC0804 chip. Because analog-digital conversion chip cannot finish the work by itself, it should cooperate with microcontroller to use. Its schematic circuit diagram is shown in the following Fig. 3.

The signal outputted by piezoelectric acceleration sensor is generally too weak.

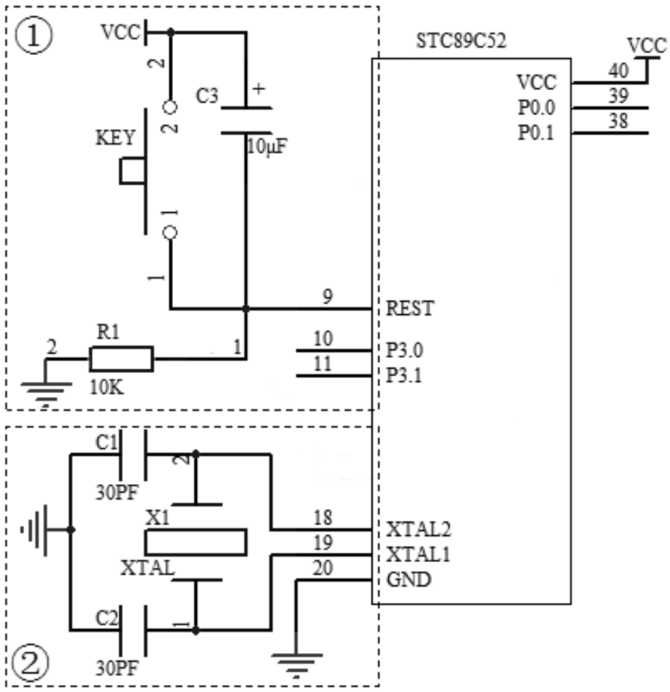


Fig. 2. SCM schematic diagram (automatic and manual reset circuit in imaginary line region (1), SCM clock circuit in imaginary line region (2); P3.0 is the RXD serial input interface, P3.1 is the TXD serial output interface)

In order to improve the whole measuring system's accuracy and measurement accuracy, the outputted vibration signal of sensor needs to be amplified and processed. Vibration signal's conditions mainly contain two parts: smoothing and amplification. The design of smoothing circuit can improve the SNR in the whole measuring system. Because the designed wireless vibration detection system is for low-frequency vibration's detection in this research, no-source LPF circuit is designed, which is shown in Fig. 4.

The output signal of piezoelectric acceleration sensor has two measuring circuits: the first one is voltage amplifier circuit, while another one is electric charge amplifying circuit. The output voltage of this electric charge amplifying circuit is only proportional to its inputted quantity of electric charge, so its largest advantage is make the piezoelectric acceleration sensor's sensitivity not influenced by the circumscribed cable. Its operating circuit is shown in Fig. 5.

In Fig. 5, Q is the output electric charge quantity of piezoelectric acceleration sensor, (-A) stands for reverse input end of operational amplifier, and R_f is the feedback resistance. In this circuit, its positive end is connected to ground, and the output voltage is

$$U_\mathrm{o} \approx U = -\frac{Q}{C_\mathrm{f}}\,. \tag{1}$$

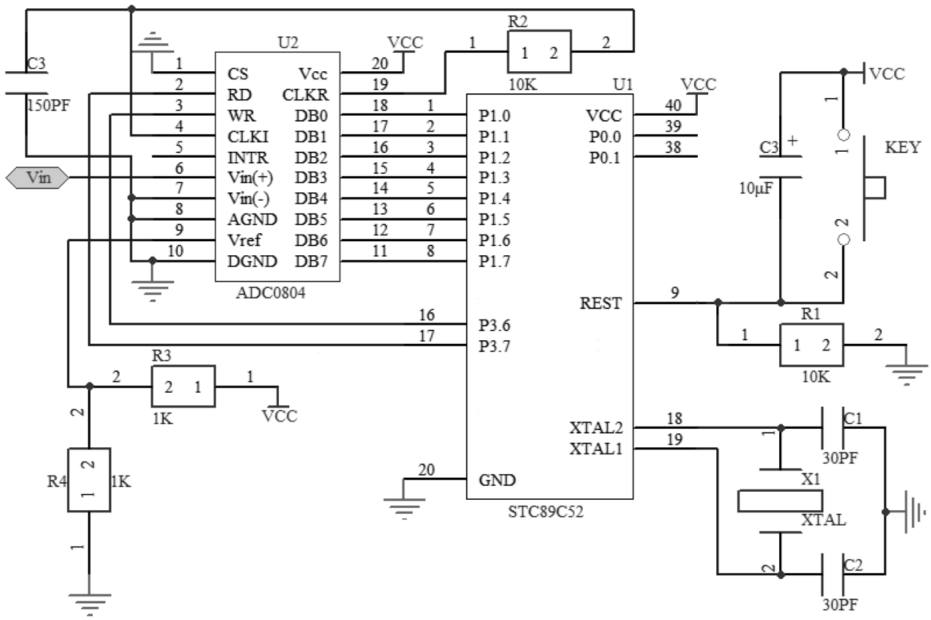


Fig. 3. Working schematic circuit diagram of ADC chip

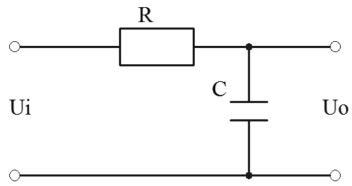


Fig. 4. No-source LPF circuit

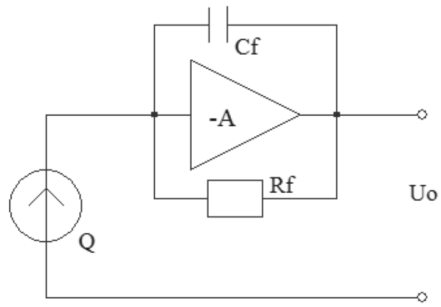


Fig. 5. Electric charge amplifying circuit

In this formula, U is the both-end voltages of feedback capacitance C_f, U_o is the output voltage of electric charge amplifying circuit. High frequency upper limit of

electric charge amplifying circuit generally can reach 180 kHz, while its low frequency lower limit f_{L} depends on the feedback capacitance and feedback resistance:

$$f_{\mathrm{L}} = \frac{1}{2\pi R_{\mathrm{f}} C_{\mathrm{f}}}\,. \tag{2}$$

From formula (2), low frequency lower limit of electric charge amplifying circuit is inversely proportional to time constant $R_{\mathrm{f}}C_{\mathrm{f}}$, and f_{L} can reach $10^{-1} \sim 10^{-4}$ Hz (quasi state).

3.1.2. Bluetooth module design. All the functional parts of Bluetooth can be made as a combinational design in the Bluetooth technology's development process. Because of limited development time, Bluetooth module usually contains an integrated RF transceiver, baseband controller, link supervisor and HCI to a chip.

According to operating mode, Bluetooth module can be classified into two parts: master Bluetooth and slave Bluetooth. Master Bluetooth can search for surrounding slave Bluetooth facilities and request an initiation, but it cannot be searched by other mater Bluetooth. Slave Bluetooth only can be searched and connected. And it can be matched with master Bluetooth, Bluetooth adapters or computers, mobiles and other electronic equipment with Bluetooth function. After matched successfully, serial port communication is on operation. HC05 Bluetooth module is adopted, that is a principal and subordinate Bluetooth module. This module can achieve Bluetooth communication and transmission function only needs little peripheral circuit. Peripheral schematic circuit diagram of HC05 Bluetooth module is shown in Fig. 6.

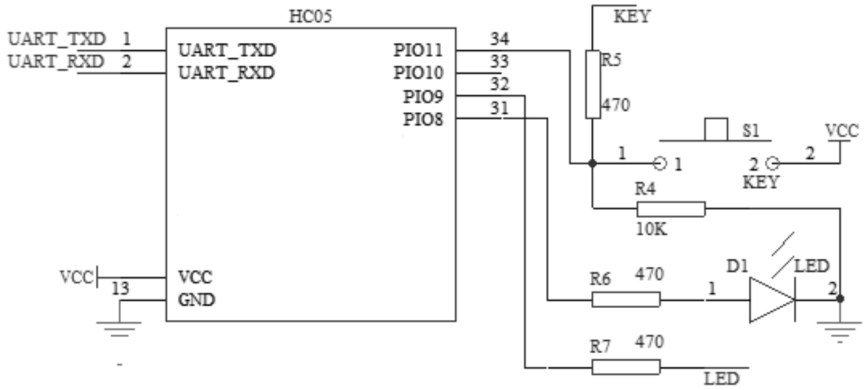


Fig. 6. Peripheral schematic circuit diagram of Bluetooth module

From Fig. 6, SCM's four pins, RXD, TXD, VCC and GND should be linked accordingly with Bluetooth module's four pins, UART_TXD, UART_RXD, VCC and GND. Only doing this, the Bluetooth module can work normally and complete the data's wireless transmission. Besides, the Bluetooth needs testing and other parameters need setting, so the design of Bluetooth wireless transmission can be finished.

4. Software design

Software programming is classified into two parts: low computer programming and upper computer programming.

4.1. Low computer programming design

The whole program compilation and object code generation of low computer are finished in the Keil *upmu*Vision3 development environment. And low computer programming is a software design for SCM. MD is used in the program development process. And first, subprograms of all the modules are designed, and then main program is used to adjust all the module subprograms. Subprogram included delay, suspend, A/D converter, serial port communication and other programs. The design of signal acquisition program is represented in the following.

Vibration signal's acquisition process is achieved by signal acquisition by means of SCM controlling the A/D converter chip. In the acquisition process, A/D is first started, and then waits for whether A/D has been converted. At last, SCM reads the data that A/D converted. In this programming, sequential working drawing of ADC0804 analog-digital conversion chip needs knowing and this chip's normal working can be achieved according to the program adapted by sequential working drawing. Figure 7 is the ADC0804 chip's working sequence chart.

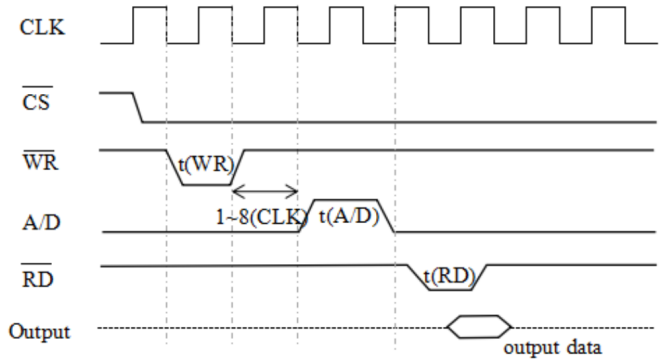


Fig. 7. ADC0804 chip's working sequence chart

According to the above sequence chart, signal acquisition programs are as follows:

CS = 0; WR = 1; _nop_(); WR = 0; _nop_(); WR = 1; P1 = Oxff ; RD = 1; _nop_(); RD = 0; _nop_(); TEMP = P1:

4.2. Upper computer programming design

Program adapted by Lab VIEW,Lab VIEW, is chosen and used for interface development, that includes front panel and program chart. Front panel is also called UI, which is mainly applied to control and display. In the front panel's programming environment, there are two kinds of controls: input control and display control. Input control can set parameters and manipulate data. Display control can display

measured data in real time.

As UI of wireless vibration detection system, upper computer program has many functions in the whole process of wireless vibration detection. Its functional block diagram is shown in Fig. 8.

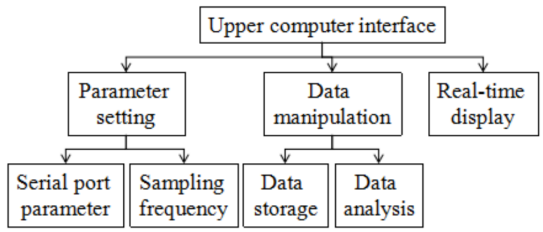


Fig. 8. Functional block diagram of upper computer interface

According to Fig. 8, upper computer programming design is dividing the whole upper computer into different modules and then makes an adaption to those different modules. Finally, the adaption work of the whole program. Figure 9 is the program chart of the whole upper computer programming design.

Low computer can be controlled by manipulating the upper computer's UI. At the same time, using this interface not only can achieve storage and real-time display of RXD, but also can display frequency spectrum curve after making an FFT analysis about vibration signal. It is convenient for researchers to intuitively know the vibration information.

5. Experiment and result analysis

5.1. Vibration test platform building

A vibration test platform is built to verify whether the wireless vibration detection system can normally perform its functions. The structure diagram of the whole test platform is shown in Fig. 10.

In this test experiment, first, vibrostand's vibration signal is picked up through piezoelectric acceleration sensor and converted as electric signal. Then electric signal outputted by sensor is enlarged and smoothed. After it is converted as digital signal by A/D convertor and stored in SCM to fulfill the acquisition of vibration signal. Through UART, SCM transmits the vibration signal form slave Bluetooth module to master Bluetooth module and then completes wireless launch of vibration signal.

Finally, master Bluetooth module is controlled by upper computer UI in the PC computer, which can achieve vibration signal reception wirelessly launched by low computer and store the received vibration signal and display it on the UI. In the meanwhile, vibration signal van be analyzed with Fourier transformation by using upper computer UI, and then display its frequency spectrum curve on the UI.

After the vibration test platform finished, it can test the designed wireless vibration detection system's function. vibrostand's vibration can be tested by using the existing piezoelectric acceleration sensor in the laboratory and the researched

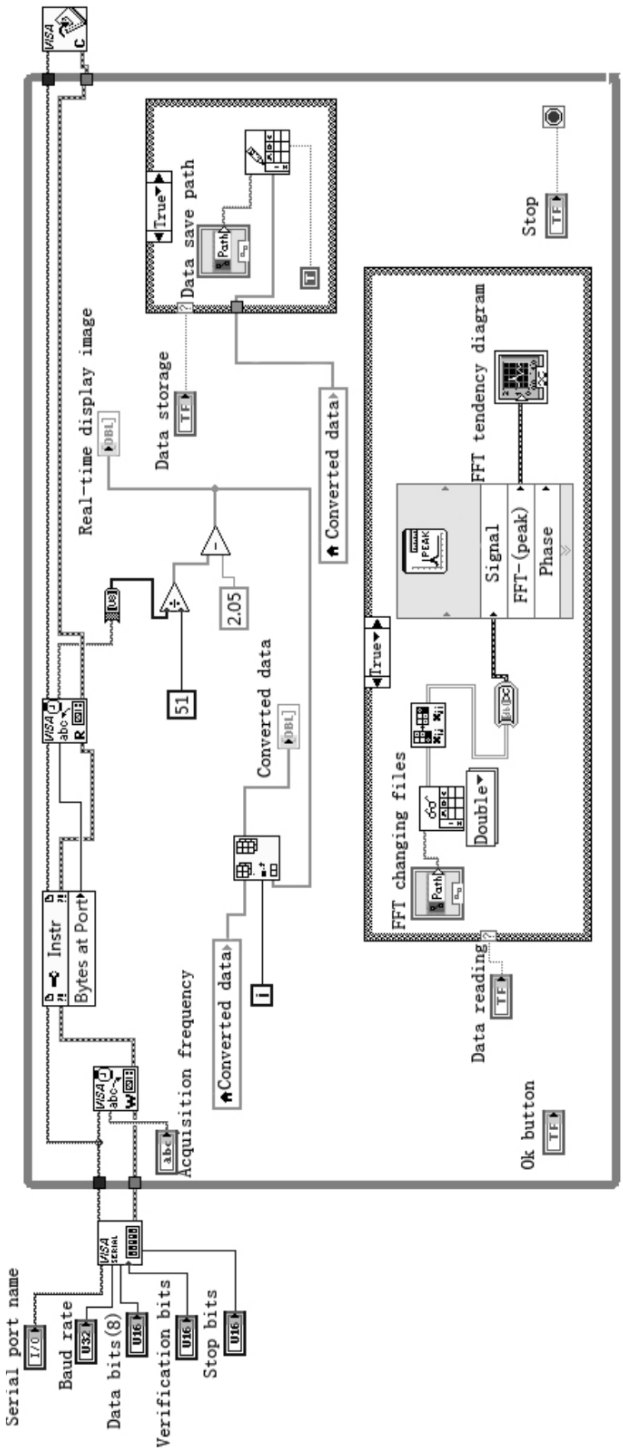


Fig. 9. Whole block diagram of upper computer program

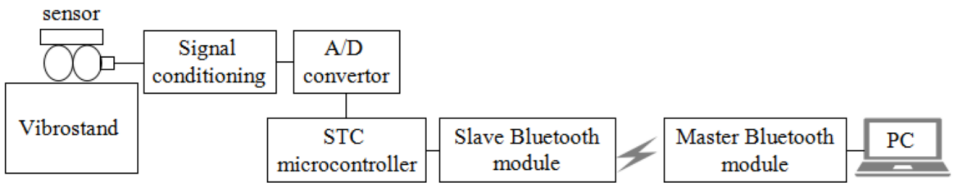


Fig. 10. Block diagram of vibration test platform

wireless vibration detection system, which can verify the reliability and feasibility of the system.

6. Experimental result and analysis

Figure 11 is the real-time display image where 10 Hz sinusoidal vibration detection is derived from the system's upper computer interface to the vibrostand.

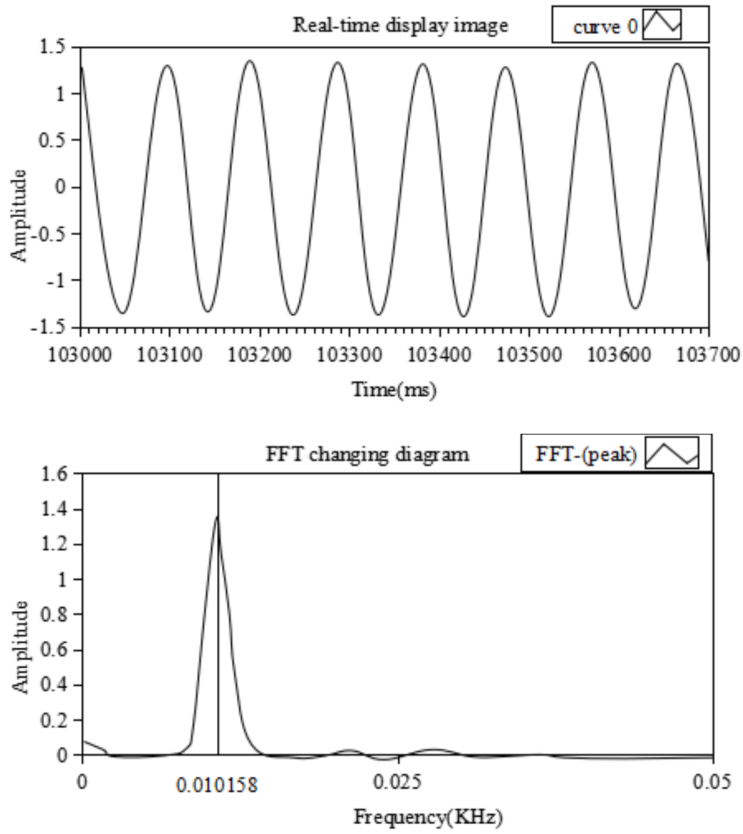


Fig. 11. 10 Hz testing result of vibration signal: up–time domain diagram, bottom–spectrogram

On the upper computer interface, vibration's real-time condition can be observed. Its upper part is the time domain waveform of vibration signal. Its bottom part is the spectrogram of vibration signal being quickly transferred to FFT. According to this spectrogram, vibration frequency can be intuitively watched. From FFT changing diagram, the wireless vibration detection result is: vibration signal frequency $f = 10.158$ Hz. Compared with true value, its relative error is 1.580 %.

In order to fully test the feasibility and reliability of the designed wireless vibration detection system, sinusoidal vibrations of 50 Hz and 100 Hz vibration frequency are respectively tested in the following diagrams. And their vibration detection results are shown in the figure Fig. 12 and 13.

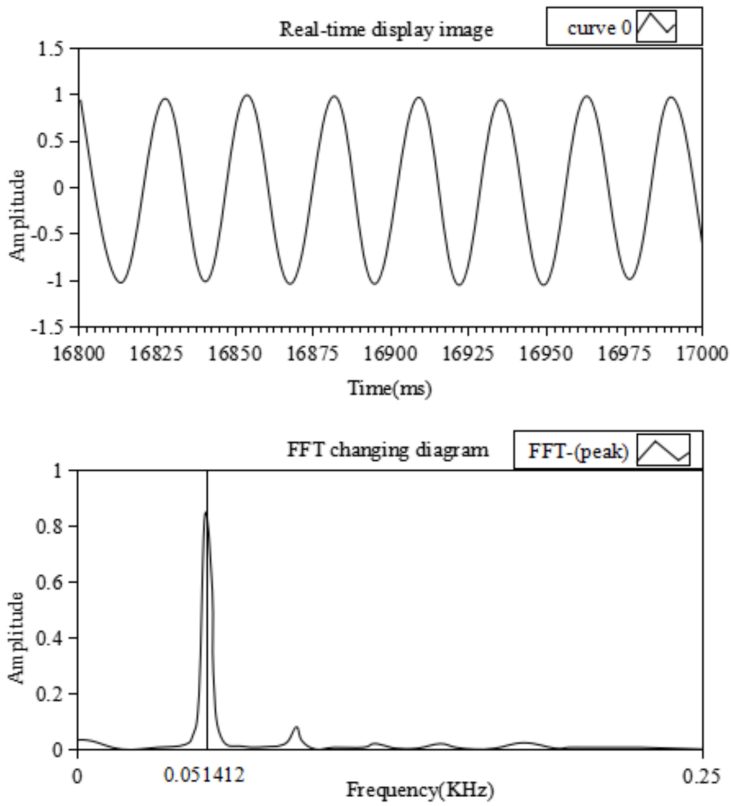


Fig. 12. 50 Hz vibration signal detection result: up–time domain diagram, bottom–spectrogram

As the system has a test to vibrostand's sinusoidal vibration signal, its oscillogram is basically in accordance with the sinusoidal waveform in the time domain diagram. In the spectrogram, it is clearly seen that frequency corresponded by the highest peak point of frequency spectrum curve is the vibration signal frequency. To sum up the testing results, feasibility and reliability of wireless vibration detection system are verified.

From Table 1, with the vibrostand's vibration frequency increased, relative error

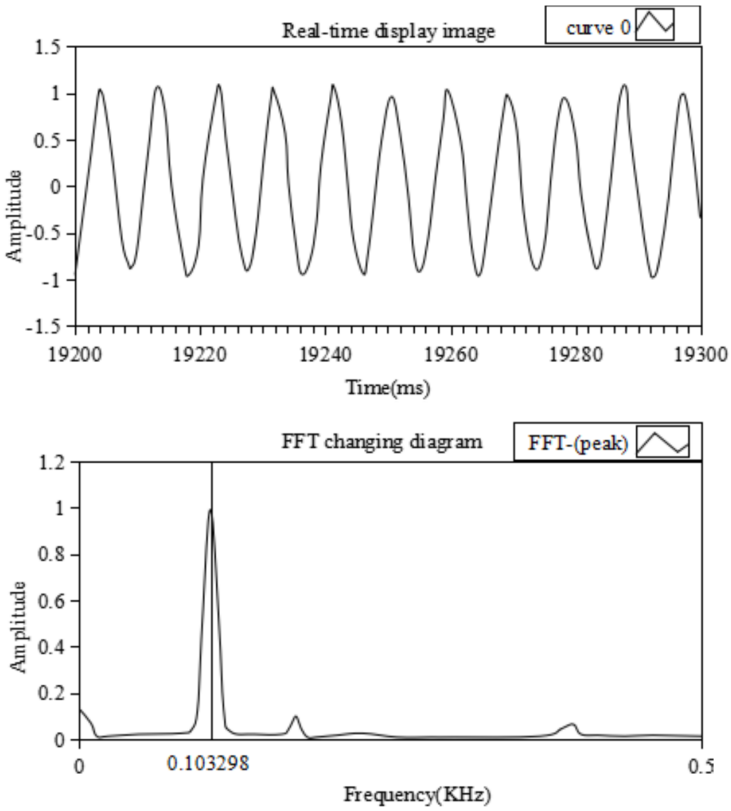


Fig. 13. 100 Hz vibration signal detection result: up–time domain diagram, bottom–spectrogram

of vibration frequency gained from the wireless vibration detection system is also increased accordingly. According to sampling theorem, with vibration signal frequency increased, the measured relative error is getting larger in sampling frequency's condition. From Table 1, if vibration signal frequency is below 100 Hz, the relative error of the measured vibration frequency is no more than 5 %.

Table 1. Relative error of wireless vibration system's measurement result

Vibration signal frequency (Hz)	Measured value (Hz)	Relative error (%)
10	10.158	1.580
50	51.412	2.824
100	103.298	3.298

And the error should include the highest peak position error chosen from frequency spectrum curve. Because the highest peak positions are chosen differently, the corresponding frequency values are various. If the system's sampling frequency is improved, the relative error can be reduced far less than 1 %. However, the present

systematic function has basically met the functional demand, especially in human health.

7. Conclusion

Main content of the research is as follows: In the hardware design, STC single chip is taken as microcontroller center of signal acquisition and ADC0804 chip is selected as ADC. Vibration signal is first picked up by sensor, and then enlarged and smoothing circuit to transmit to A/D converter. At last, the acquisition of vibration signal is completed by SCM controlling the A/D converter. LabVIEW development platform is selected to design upper computer's UI whose design is mainly concentrated on three functions to program: real-time display, vibration data storage and analysis. In order to research the feasibility and reliability of the designed wireless vibration signal detection system, a vibration test platform is built to make a wireless vibration test to the several frequency sinusoidal vibrations. From the testing result, this system has a good performance whether in reliability or instability, which provides a technical reference for wireless vibration detection.

References

[1] E. T. Esfahani, S. Wang, V. Sundararajan: *Multisensor wireless system for eccentricity and bearing fault detection in induction motors.* IEEE/ASME Transactions on Mechatronics *19* (2014), No. 3, 818–826.

[2] G. Cazzulani, S. Cinquemani, L. Comolli, A. Gardella, F. Resta: *Vibration control of smart structures using an array of Fiber Bragg Grating sensors.* Mechatronics *24* (2014), No. 4, 345–353.

[3] W. Wang, O. A. Jianu: *A smart sensing unit for vibration measurement and monitoring.* IEEE/ASME Transactions on Mechatronics Acoustics *15* (2010), No. 1, 70–78.

[4] S. Korkua, H. Jain, W. J. Lee, C. Kwan: *Wireless health monitoring system for vibration detection of induction motors.* IEEE Industrial and Commercial Power Systems Technical Conference (I&CPS), 9–13 May 2010, Tallahassee, FL, USA, IEEE Conference Publications (2010), 1–6.

[5] Y. Q. Ni, B. Li, K. H. Lam, K. H. Law: *In-construction vibration monitoring of a super-tall structure using a long-range wireless sensing system.* Smart Structures and Systems *7* (2011), No. 2, 83–102.

[6] M. Hong: *A real-time pulse wave detection device basing on signal envelope.* IEEE International Conference on BioMedical Engineering and Informatics, 16–18 October 2012, Chongqing, China, IEEE Conference Publications (2012), 812–816.

[7] G. Q. Xi, F. M. Tian, J. H. Wu, Y. J. Wang: *Design of tissue-engineered myocardial contractile force detection system based on VB and MCU.* International Conference on Electronics, Communications and Control (ICECC), 16–18 October 2012, Zhoushan, Zhejiang, China, Publishing IEEE Computer Society (2012), 1952–1955.

[8] J. Lemke: *Remote vibration monitoring system using wireless internet data transfer.* SPIE–The International Society for Optical Engineering, Proceedings of the SPIE *3995* (2000), 436–445, Code 2000SPIE.3995..436L.

[9] H. S. Yoon, S. Y. LeeY, J. T. Kim, J. H. Yi: *Field implementation of wireless vibration sensing system for monitoring of harbor caisson breakwaters.* International Journal of Distributed Sensor Networks *8* (2012), No. 12, paper 597546.

[10] A. S. Weddell, G. V. Merrett, S. Barrow, B. M. Al-Hashimi: *Vibration-powered sensing system for engine condition monitoring*. IET Conference on Wireless Sensor Systems (WSS), 18–19 June 2012, London, UK, IET Conference Publications (2012), 1–5.

[11] K. Loh, J. Lynch, Y. Wang, K. H. Law, M. Fraser, A. Elgamal: *Validation of a wireless traffic vibration monitoring system for the Voigt Bridge*. World Forum on Smart Materials and Smart Structures Technology (SMSST), 22–27 May 2007, China, CRC Press (2008), Chapter No. 26.

[12] Y. Lei, W. A. Shen, Y. Song, J. P. Lynch: *Application of wireless monitoring system for the ambient vibration study of the Wuyuan steel arch bridge*. World Forum on Smart Materials and Smart Structures Technology (SMSST), 22–27 May 2007, China, CRC Press (2008), Chapter No. 54.

[13] K. D. Nguyen, J. T. Kim, Y. H. Park: *Long-term vibration monitoring of cable-stayed bridge using wireless sensor network*. International Journal of Distributed Sensor Networks *9* (2013), No. 11, paper 804516.

[14] G. Feltrin, K. E. Jalsan, K. Flouri: *Vibration monitoring of a footbridge with a wireless sensor network*. Journal of Vibration and Control *19* (2013), No. 15, 2285–2300.

[15] A. S. Ramirez, R. Loendersloot, J. M. J. Becker, T. Tinga: *Design framework for vibration monitoring systems for helicopter rotor blade monitoring using wireless sensor networks*. International Workshop on Structural Health Monitoring - Stanford University, CA, USA, 10–120 September 2013, DEStech Publications *114* (2013), Nos. 3–4, 1039–1046.

Received May 7, 2017

Acta Technica *62* No. 2A/2017, 509–518

Scheduling algorithm for real-time tasks of embedded systems based on dynamic voltage regulation

Chang Weigong[1]

Abstract. In order to test the effectiveness of task energy saving scheduling algorithm, an embedded task energy saving scheduling algorithm based on dynamic voltage regulation is designed. Dynamic voltage regulation technology can effectively reduce energy consumption, so it has been widely used. Now many processors have dynamic voltage regulation capability. With the increasing demand for high performance and energy saving, multi-core processors have become the mainstream of the embedded system market. For the multi-core system for energy-efficient scheduling algorithms, task allocation, sequence arrangement and voltage arrangement have a significant impact on energy efficiency. The purpose is to study how to adjust the order of tasks that the voltage conversion of energy minimum, and study how to design the corresponding dynamic voltage regulation algorithm to solve the problem of multi-core system with fixed voltage conversion time or negligible voltage conversion time. The experimental results show that in the initial scheduling length, the method 2 uses the retimed directed acyclic graph as the object of energy-saving scheduling. Based on the above finding, it is concluded that the proposed method achieves good energy saving effect under the condition of meeting the task setting time limit.

Key words. Dynamic voltage regulation, embedded, energy-saving scheduling, algorithm.

1. Introduction

With the rapid growth in the number of mobile devices and embedded device platforms, energy consumption is gaining more and more attention. High performance is currently a growing requirement, and high performance often comes at the expense of huge energy consumption. Chip area continues to shrink, also makes the energy consumption problem becomes more prominent. Many embedded systems are energy-constrained. Therefore, energy minimization is an important issue in embedded system design [1].

Dynamic voltage regulation technology can effectively reduce energy consumption, so it has been widely used. In order to improve energy efficiency, many current

[1]Tianjin University of Science & Technology, 300072, China

processors support dynamic voltage regulation and provide a variety of supply voltage files [2]. The emergence of a variety of supply voltage processors in the market, making it a reality in the processor layer for energy management. In order to meet the needs of high-performance and reduce energy consumption, multi-core processors have become the mainstream of the embedded system market. In a multi-core environment, since each processing core must share the same clock, blindly using the existing dynamic voltage methods applied to single-core processors will result in more unnecessary power consumption. It can be seen that the problem of multi-core processor is more complex than multiprocessor. Therefore, it is necessary to take multi-core processor as the research object, study how to reduce energy consumption under the conditions of meet the time constraints.

2. Literature review

Weiser first proposed a variety of dynamic voltage scheduling algorithm [3], but the shortcomings of the study are non-real-time tasks. However, his research has enlightened the research of dynamic voltage scheduling algorithm, and then various dynamic voltage scheduling algorithms have been proposed, which has solved the energy optimization problem of different types of tasks running on different types of systems.

For multi-processor/multi-core processor on the dynamic voltage regulation of energy-efficient scheduling, reference [4] proposed an improved greedy algorithm based on dynamic voltage regulation, this algorithm takes into account the concern that unrelated tasks reduce energy consumption on multiprocessors with time constraints, obtains the widespread attention. However, these references do not address the issue of using dynamic voltage techniques to save energy in multi-core systems.

Reference [5] proposes a two-stage approach to energy optimization for heterogeneous distributed systems under time constraints. First of all, the start time of each task as late as possible with a random number, the value will be sorted in ascending order will be given priority to the task, and then use the table scheduling algorithm to achieve the task of distribution, apply the method of relaxation budget to distribute the relaxation. The system is heterogeneous distributed system.

3. Methods

Definition: On a multi-core processor, for a scheduling generated by a retimed directed acyclic graph, When the last computation task in a processing kernel has no descendant node in the original corresponding acyclic graph, or a computing node having a descendant node but representing a descendant node is assigned to the same processing core as the last executed computing task, the scheduling length of the processing core is the completion time of the last computing task executed on it [6]. When the last computation task in a processing kernel has the descendant node in the node corresponding to the original acyclic graph, And the computational tasks in the computational tasks on behalf of their descendant nodes are assigned

to the different processing cores, and the last executed computational task on the processing core is allocated to the different processing cores, the scheduling length of the processing core is the maximum completion time of the communication task generated by the last computing task executed on the processing core [7].

Theorem: In the case where the voltage conversion time is a fixed constant or can be ignored, when each task is executed in a frequency/voltage mode, the tasks on each processor is executed in descending order of voltage or in ascending order, resulting in minimal voltage transfer energy.

Assuming that the tasks on a certain processor core have $n+1$ frequency/voltage mode, the voltage of the ascending order is V_1, V_{1+a_1}, V_{1+a_2}, ..., $V_{1+a_{k-1}}$, V_{1+a_k}, ..., $V_{1+a_{m-1}}$, V_{1+a_m}, ..., V_{1+a_n}, where, $0 \leq a_1 \leq a_2 \leq \ldots \leq a_{k-1} \leq a_k \ldots \leq a_{m-1} \leq a_m \ldots \leq a_n$. The voltage conversion energy of the $n+1$ frequency/voltage mode is

$$A + C_{\mathrm{r}} \cdot [(a_{k-1} - a_k)^2 + (a_k - a_{k+1})^2 + \cdots + (a_{m-1} - a_m)^2 + \\ + (a_m - a_{m+1})^2 + \cdots + (a_{n-1} - a_n)^2], \tag{1}$$

where A is the voltage conversion energy of the first k frequency/voltage modes and C_{r} is the current ratio. Exchange of the position of V_{1+a_k} and V_{1+a_m} will provide

$$A + C_{\mathrm{r}} \cdot [(a_{k-1} - a_m)^2 + (a_m - a_{k+1})^2 + \cdots + (a_{m-1} - a_k)^2 + \\ + (a_k - a_{m+1})^2 + \cdots + (a_{n-1} - a_n)^2]. \tag{2}$$

Subtraction of formula (3) from formula (2) provides

$$2C_{\mathrm{r}} \cdot (a_k - a_m) \cdot [(a_{k-1} + a_{k+1}) - (a_{m-1} + a_{m+1})] \geq 0. \tag{3}$$

The same result can be obtained by swapping the positions of any two frequency/voltage modes, so the tasks on the processing core are performed in the order of ascending order of the voltage levels to minimize the conversion energy [8]. Similarly, the tasks of the processor cores are performed in the descending order of the voltage profiles produced the minimum switching energy. It is considered that the transition time from the operating state to the idle state of the applied processing core may be a fixed constant and cannot be ignored, in this case, when the difference between the given time limit and the completion time of the last computation task on a certain processing core is smaller than the transition time from the operating state to the idle state, the remaining time is run in the frequency/voltage mode of the last computing task running on the processing core, therefore, in order to reduce the energy consumption, the tasks on the respective processing cores are executed in descending order of the voltage level for a processor whose conversion time from the operating state to the idle state is a fixed constant and cannot be ignored. While the transition time from the running state to the idle state is a negligible processor, the tasks on each processing core can be performed in descending or ascending order

of the voltage file [9].

After the above theoretical analysis, an algorithm is proposed to solve the energy-saving scheduling problem of multi-core processors with task-dependent applications. The energy-saving scheduling algorithm takes the isomorphic multi-core system with dynamic voltage regulation capability as the research object, and realizes the compaction energy and the conversion energy. The work flow diagram of the algorithm is shown in Fig. 1, and the implementation steps are shown in Table 1.

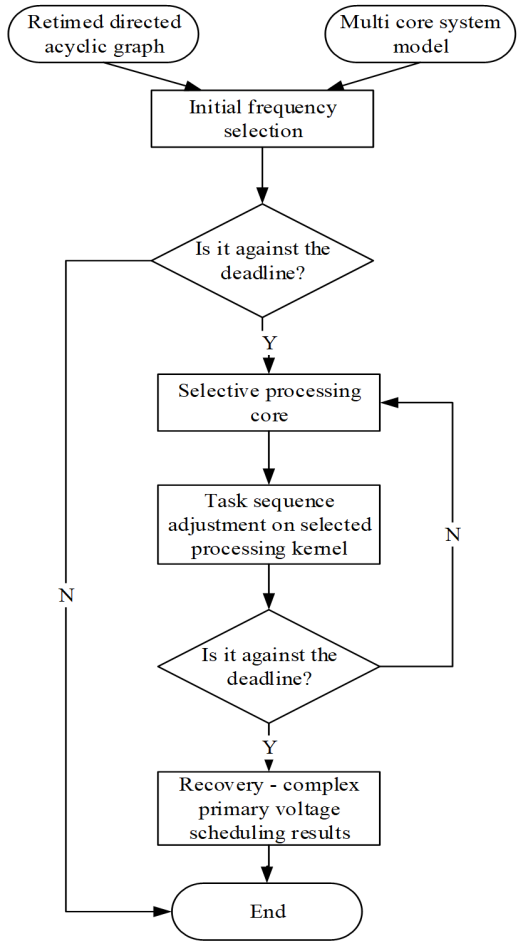


Fig. 1. Design flow diagram of energy - saving scheduling algorithms for dynamic voltage regulation multicore systems

In order to verify the energy-saving effect of the proposed energy-saving scheduling method, the proposed method (method 2) is compared with the method (method 1) proposed in the literature [10]. In the simulation experiment, five randomly generated directed acyclic graphs are used, named task set 1–task set 5, respectively. The number of calculation tasks varies between 27 and 68, while the number of

communication edges varies between 45 and 94. The number of calculation task cycles satisfies a uniform distribution between 10 and 40, specific task characteristics shown in Table 2.

Table 1. Implementation steps of energy-saving scheduling for multi-core system with dynamic voltage regulation

Stage	Step
Initial scheduling	Step 1: The calculation tasks are arranged in descending order of cycles, the mapping of the tasks to the processing kernel is carried out according to the principle of priority assignment of the shortest processing kernel (for the calculation task).
	Step 2: For each processing core, the calculation task with the minimum communication time is set as the computation task last executed by the processing core while the order of the other calculation tasks is kept unchanged.
	Step 3: The computational tasks on each processing core are arranged at the lowest frequency (except for the frequency corresponding to the idle state) of the processing core. If the generated scheduling length is less than or equal to the given time limit, the scheduling is accepted and returned, otherwise go to step 4.
	Step 4: The computational tasks on each processing core are arranged at the highest frequency of the processing core.
Re-scheduling	Step 5: In the processing core with the shortest scheduling length, select a computing task T, When its frequency is reduced by a frequency file and in accordance with its changes in the voltage file in descending order, the computational task on the processing core minimizes the total energy consumption, the computation task T is extended in execution time, that is, the frequency profile of the calculation task is reduced one rating level.
	Step 6: If the new schedule length of the processing core is less than or equal to a given time limit, then accept the change, the calculation task T is executed at the changed frequency profile, the calculation task on the processing core is executed in the order of decreasing voltage range, updating the schedule, go to step 5; otherwise go to step 7.
	Step 7: The calculation task T is executed at the frequency before the conversion, the calculation tasks on the processing core are executed in the order before the change, that is, accept the last change, and return.

For the two computational tasks that have dependencies in the original acyclic graphs, if they are assigned to different processing cores, the number of communication cycles between them is from 1 to 4 randomly selected values. For the two computational tasks that have dependencies in the original acyclic graphs. If they are assigned to the same processing core, the number of communication cycles between them is zero. In the studied system, the frequency on the bus is fixed, set

to 208 MHz. The generated task set was simulated on a multi-core system with 2 and 4 cores respectively. The processor using the Intel PXA270 power model, it can operate in 7 frequency/voltage modes. For Intel PXA270, its mode transition time is negligible. Calculate the voltage conversion energy, the initial time limit is the initial scheduling length of the two nuclear processing system in the method 1, then take the 10 % of the schedule length as the step size, set the time limit, and take the average value from obtained energy consumption.

Table 2. Task set feature table

Task Set Name	Number of nodes	Number of edges	Number of cycles
Task Set 1	27	45	681
Task Set 2	35	54	880
Task Set 3	46	67	1233
Task Set 4	57	74	1520
Task Set 5	68	94	1808

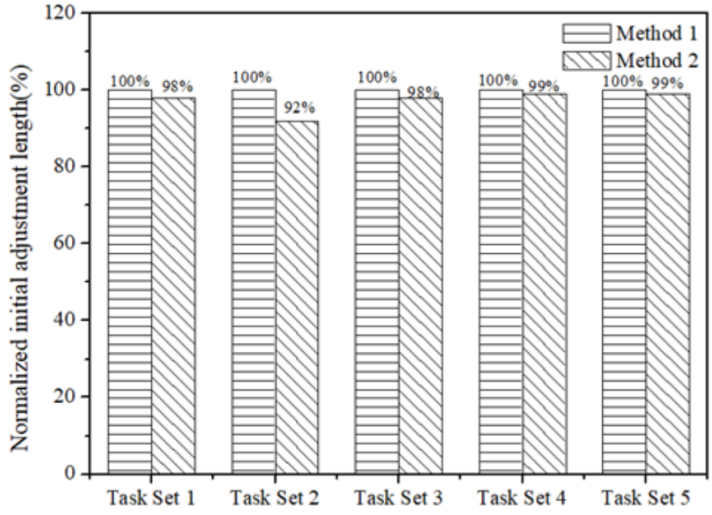
4. Results

The simulation results of method 1 and method 2 are compared respectively from the initial scheduling length and the average energy consumption. Figure 2 shows the initial scheduling length comparison results of method 1 and method 2. Figure 3 shows the results of the average energy consumption comparison of method 1 and method 2.

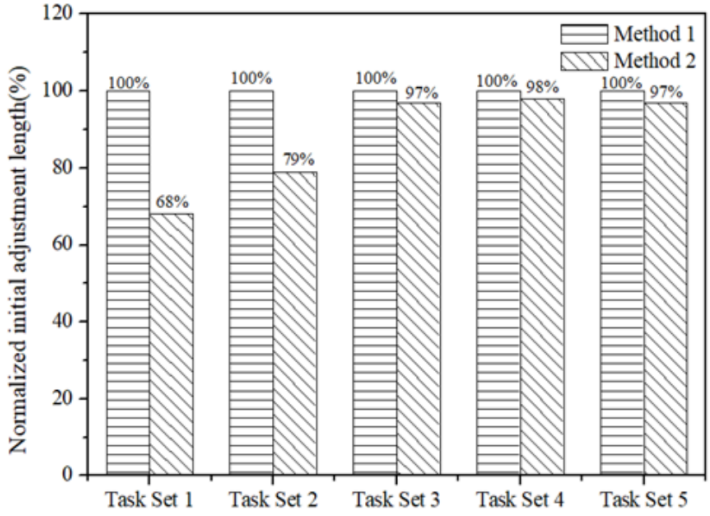
5. Discussion

The results show that, in the initial scheduling length, the method 2 uses the retimed directed acyclic graph as the object of energy-saving scheduling, the tasks in the retimed directed acyclic graphs do not have data dependency in iteration, and improve the parallelism of tasks. Thereby reducing the initial scheduling length. While the method 1 uses a directed acyclic graph, in the directed acyclic graph, the data dependency of the task weakens the parallel ability of the task, resulting in the initial scheduling length of the task is often larger than the initial scheduling length generated by the retimed directed acyclic graph, at the beginning of the average energy consumption, using the retimed directed acyclic graph as the energy-saving scheduling object, so that make the dependencies task in the original directed acyclic graph no longer exist data dependency within iteration, while there is only data dependency between iterations, improve the parallel ability of the task and can produce more slack. When the retimed directed acyclic graph is used as an energy-saving scheduling object, the slack generated on one processing core can be assigned to each task on the processing core, thus providing more frequency/voltage mode adjustment opportunities for the task. Also, since the task execution order is determined in the frequency/voltage mode selection process according to the voltage

level of the task on each processing core, it is possible to simultaneously reduce the dynamic energy consumption and the voltage conversion energy consumption.



(a) Comparison of Initial Scheduling Lengths on Two Processing Cores



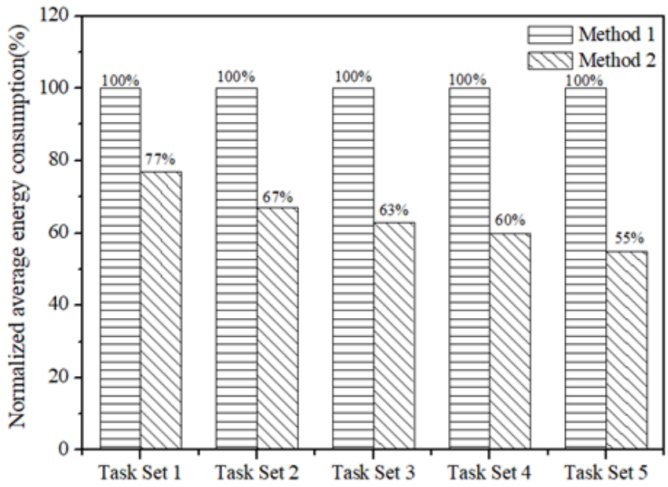
(b) Comparison of Initial Scheduling Lengths over Four Processing Cores

Fig. 2. Comparison of initial scheduling lengths of two methods

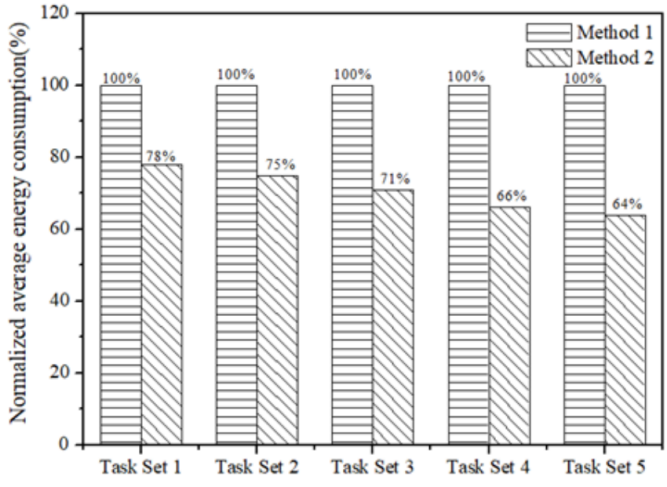
Method 1 due to the influence of data dependency in the iteration, the slack between tasks can only be used to reduce the frequency of the adjacent task and re-

duce the frequency/voltage mode adjustment. Data dependency within the iteration limits the ability to take advantage of multi-core, and therefore leads to a reduction in energy-saving opportunities.

From the simulation results in Fig. 2 upper and bottom parts (a and b), it can be seen that the initial scheduling length of method 2 is smaller than the initial scheduling length of method 1.



(a) Comparison of average energy consumption on two processing cores



(b) Comparison of average energy consumption over four processing cores

Fig. 3. Comparison of average energy consumption of two methods

From the simulation results in Fig. 3 upper and bottom parts (a and b), it can be seen that the energy saving effect of the method 2 is better than that of the method 1.

6. Conclusion

The related research work of energy saving algorithm using dynamic voltage regulation technology is introduced. Taking the multi-core processor as the research object, in order to solve the problem that the voltage conversion time of some processors is fixed or negligible, an energy-saving scheduling method is proposed to reduce energy consumption and voltage conversion energy consumption. The method can constrain the frequency/voltage regulation mode and can use the timed task graph to schedule the object. It can also perform frequency/voltage mode selection and task sequence adjustment. Dynamic voltage regulation technology can effectively reduce energy consumption, so it has been widely used. In order to improve energy efficiency, many current processors support dynamic voltage regulation and provide a variety of supply voltage files. In order to meet the needs of high-performance and reduce energy consumption, multi-core processors have become the mainstream of the embedded system market. In a multi-core environment, since each processing core must share the same clock, blindly using the existing dynamic voltage methods applied to single-core processors will result in more unnecessary power consumption. It can be seen that the problem of multi-core processor is more complex than multiprocessor. Therefore, it is necessary to take multi-core processor as the research object, study how to reduce energy consumption under the conditions of meet the time constraints. The simulation results show that the proposed method achieves good energy saving effect under the condition of meeting the task setting time limit.

References

[1] J. Wang, W. Bao, X. Zhu, L. T. Yang, Y. Xiang: *Festal: Fault-tolerant elastic scheduling algorithm for real-time tasks in virtualized clouds.* IEEE Transactions on Computers *64* (2015), No. 9, 2545–2558.

[2] Y. W. Zhang, R. F. Guo: *Low-power scheduling algorithms for sporadic task with shared resources in hard real-time systems.* Computer Journal *58* (2015), No. 7, 1585–1597.

[3] Y. W. Zhang, R. F. Guo: *Low-power scheduling algorithm for mixed task in real-time system.* Journal of Jilin University (Engineering and Technology Edition) *45* (2015), No. 1, 261–266.

[4] S. Rajpradhan, S. Sharma, D. Konar, K. Sharma: *A comparative study on dynamic scheduling of real-time tasks in multiprocessor system using genetic algorithms.* International Journal of Computer Applications *120* (2015), No. 20, 975–8887.

[5] J. H. Anderson, J. P. Erickson, C. UmaMaheswari, B. N. Casses: *Optimal semi-partitioned scheduling in soft real-time systems.* Journal of Signal Processing Systems *84* (2016), No. 1, 3–23.

[6] Z. Peng, G. Wang: *An optimal energy-saving real-time task-scheduling algorithm for mobile terminals.* International Journal of Distributed Sensor Networks *13* (2017), No. 5, 1–12.

[7] M. L. C. Prabhake, K. Manivannan: *Survey of hard real time task scheduling algorithm on multicore processor*. Asian Journal of Research in Social Sciences and Humanities *6* (2016), No. 12, 135–153.

[8] H. Alhussian, N. Zakaria, F. A. Hussin, H. T. Bahbouh: *USG, an un-fair, semi greedy real-time multiprocessor scheduling algorithm*. REACTION (Conferences and Workshops), Vancouver, Canada (2013).

[9] S. Senobary, M. Naghibzadeh: *Semi-partitioned scheduling for fixed-priority real-time tasks based on intelligent rate monotonic algorithm*. International Journal of Grid and Utility Computing *6*, (2015), Nos. 3–4, 184–191.

[10] H. Alhussian, N. Zakaria, A. Patel: *An unfair semi-greedy real-time multiprocessor scheduling algorithm*. Computers and Electrical Engineering *50* (2016), 143–165.

Received May 7, 2017

Acta Technica *62* No. 2A/2017, 519–528

Low energy building design with integrated GUD system

Lei Huang[1]

Abstract. The aim of the study is to study the low-energy building designed with the integrated GUD system. By using the method of comparative study, its characteristics are analyzed. In the simulation research, the computer model is tested by the measured data. Based on the method of CFD simulation, the integration design of GUD system in architecture is optimized. The influence of different air supply modes, different materials and structural forms on the thermal environment of the building was studied. The design flow of integrated GUD system for low energy building is described. The results show that, by incorporating passive and active energy saving technologies, a balance between creating good thermal comfort environment and saving building energy consumption can be achieved. Therefore, it can be concluded that the concept of integrated design provides a new idea for the design of low-power buildings. The integrated GUD system has laid a foundation for the popularization and application of low energy building design.

Key words. Low-energy building, GUD system, integrated design, CFD simulation.

1. Introduction

PMV-PPD means predicted mean vote-predicted percentage dissatisfied. The PMV-PPD index takes into account six factors, takes into account six factors, such as intensity of human activities, thermal resistance, air temperature, average radiant temperature, air flow velocity and air humidity [1]. PMV values vary from –3 to 3. According to the human thermal sensation, there are seven grades. It corresponds to the PPD index that predicts the percentage of unsatisfied people [2]. According to the calculation method of PMV-PPD, the relationship between air temperature and mean radiation temperature and PMV value can be obtained. AT represents the air temperature and RT represents the average radiation temperature.

GUD means giving-up density. GUD system is a comprehensive indoor thermal environment adjustment technology [3]. It mainly includes three subsystems, which are underground sub-cold and heat source system, underfloor air distribution system and dynamic ventilation wall. The three subsystems of the GUD system exist side by side and play a part together. They form a complete set of building thermal

[1]Zhengzhou Chenggong University of Finance and Economics, Henan, 451200, China

environment regulating system [4]. In view of the specific climatic characteristics of the local area, the building energy consumption is reduced from the aspects of building materials, building usage and architectural [5]. At the same time, with various professional engineers, a variety of building energy efficient facilities as well as the facade of the facilities have been integrated into the architectural design. The relationship between the various parts of the system is integrated as a whole, and in the real sense, environmental friendly low-energy buildings are designed [6].

Based on this, a low energy architecture design method for integrated GUD system is proposed. Through the combination of underground cold and heat sources, floor air supply and ventilation wall technology, the function of building complete indoor thermal environment is realized. Through the combination of active energy saving design and passive energy saving design, the goal of building low energy consumption is completed. In the study, by means of field measurement, computer simulation and design analysis, the low energy building design model of integrated GUD system is explored from a practical point of view. It lays a certain foundation for its popularization and application, and puts forward a new train of thought for the design of low energy consumption building. From the practical point of view, the low energy building design method of integrated GUD system is discussed, which provides a reference and new ideas for architects and researchers concerned with building energy efficiency. It promotes the development of low energy buildings and green buildings.

2. State of the art

2.1. Current status of foreign research

Western countries have been involved in building energy efficiency and ecological architecture theory earlier. In the 60s of last century, Paul Soleri, an American architect in Italy, put forward the idea of ecological architecture (Arology). Later, in Victor Ogoya's "Design with Climate: Bioclimatic Approach to Architectural Regionalism", and Ian McHagg's "Design With Nature", the concept of eco-building was explored. The theoretical system of ecological architecture was gradually established. With the outbreak of the oil crisis in 1973, Western countries began to pay attention to the study of building energy efficiency. The solar energy, wind energy, geothermal energy and other renewable energy technologies have risen, and new technologies such as energy saving, envelope structure, energy-saving technology and equipment have also been developed rapidly. In the next forty years, in the aspects of the architectural design and construction, new building insulation materials, the development and application of building energy conservation regulations and implementation, building energy-saving product certification and management, developed countries have done a lot of work [7]. The content of building energy conservation has also been constantly updated. From concept to practice, green building is gradually improved.

There are many researches on the relationship between architectural design and climate in foreign countries. For example, Victor Olgyay summed up the achieve-

ments of architectural design and climate and geo-relationship studies before the 1960s in the book "Design with Climate: Bioclimatic Approach to Architectural Regionalism". He put forward the design theory of "bio-climate localism". B. Givoni has improved the biological climate of Ogoya in the book "Human Climate and Architecture". Starting from thermal comfort, climatic conditions are examined and analyzed, and then possible design strategies are determined. In "The Architecture without Architects: A Short Introduction to Non-pedigreed Architecture", Bernard Rodolfski reveals the relationship between regional climate and urban form and architectural style with a large number of facts. In "Sustainable Architecture and Urbanism", Dominique Gauzin introduces the thermal environment and energy considerations in architecture and urban design, ranging from technology, climate, policy and so on, and analyzes a number of examples. In the "Sun, wind & light: Architectural design strategies", from the aspects of planning, architecture and detail construction, G. Z. Brown analyzes the utilization of sunlight, wind and light in buildings, and enumerates a lot of related design strategies. In "Sustainable Urban Design", by enumerating important architects' practical projects and design experience, Randall Thomas explores how to achieve architectural unity and energy conservation through design. In his doctoral thesis, Katie Meng Cacace has developed a set of expert system software for instructing architects to apply different energy-saving technologies during the conceptual design phase of the building. In "Thermal Comfort: analysis and applications in environmental engineering", P. O. Fanger explores the criteria for human thermal comfort and its associated influencing factors. He presented a formula for calculating the thermal comfort of the human body. In addition, in later (2002) papers, he also proposed a PMV expansion model for non-air-conditioned environments in warm climate zones [8].

2.2. Current status of domestic research

Building energy efficiency in our country started from the early 80s of last century, and started late. It has gone through three stages [9]. The first stage mainly aims at the reduction of energy consumption of winter heating in the northern area. In 1986 the Ministry of Construction promulgated the "civil construction of energy-efficient design standards (heating residential buildings)" JGJ2686 (now, it has been abolished). It requires a reduction of 30 % in the general design energy consumption level from 1980 to 1981 [10]. At this stage, the passive solar house was developed in the northern area. In the second stage, the government promulgated the "civil construction of energy-efficient design standards (heating residential buildings)" JGJ2695. From 1996 onwards, on the basis of meeting the requirements of the first stage, 30 % was saved (that is, the total energy savings is 50 %). In the third stage, from 2005 onwards, on the basis of meeting the requirements of the second phase, 30 % was saved (that is, the total energy savings is 65 %). At the same time, in 2001, the Ministry of Construction promulgated the "hot summer and cold winter residential building energy efficiency design standards" JGJ134-2001. In 2003, the Ministry of Construction promulgated the "hot summer and warm winter residential building energy efficiency design standards" JGJ75-2003. It also marks

the focus of our energy-saving work from the northern heating area to the southern region. In recent years, with the rapid development of China's overall economic level, the relevant departments have paid attention to building energy efficiency. Various new materials, new technology research and development and application are gradually accelerated, and ecological architecture, green building and other architectural concepts are gradually accepted by designers. "The green building evaluation standards" GB/T50378-2006 and "green Olympic building evaluation system" and other evaluation standards put forward higher requirements for building energy saving and green ecological design. The practice of low energy buildings was also implemented simultaneously. The demonstration projects of renewable energy construction applications have played a good exemplary role in promoting the popularization of new energy-saving technologies. For example, in the super low energy consumption demonstration building of Tsinghua University, it has applied nearly 100 energy-saving and green technologies, and tried to use renewable energy to reduce the impact on the external environment. It provides users with a healthy and comfortable working life space. In addition, there are many cases, such as the Shanghai eco-office demonstration building and the best practice area in Shanghai, World Expo. These buildings integrate advanced energy-saving technology systems both at home and abroad, and meet the requirements of indoor thermal environment, while achieving lower energy consumption.

With the development of low energy consumption construction practice, the domestic scholars have gradually studied the building energy efficiency and green building. Based on the situation of our country and the specific climate characteristics, the research has carried on the theory discussion of the building energy conservation, and proposed the related design strategy, which has the strong pertinence. There are many related studies. For example, in the aspects of building energy conservation and building thermal environment theory, the building energy saving research center of Tsinghua University has edited the Research Report on the annual development of building energy efficiency in China, which analyzes the current situation of China's building energy consumption. It puts forward the potential, goals, problems and main tasks of building energy efficiency in our country, and analyzes the common energy-saving technologies in our country. The scope of application and actual energy saving efficiency are reviewed. At the same time, a large amount of energy consumption data is given for each energy dissipation path in the process of building operation. It has important reference value for mastering the development trend of building energy conservation in China and understanding the merits of building energy saving technology. In his doctoral thesis "building energy consumption gene theory research", Long Enshen built the genetic theory of building energy consumption. By using the theory, the individuality and commonness of energy-saving building are analyzed, which provides a scientific method for grasping the objective law of building energy consumption. In his doctoral thesis "architectural energy conservation theory analysis and application research", according to the climatic characteristics of north China and the current situation of building energy saving, Yu Wenhong has carried on the theory and application research of building energy saving. Based on the theory of heat transfer, the theoretical analysis and

application research of composite wall, architectural glass, construction equipment and building energy saving are carried out. In the study of climate adaptability of human thermal comfort, Mao Yan selected twelve cities with typical climatic characteristics and important economic geographical location in China. It established a climate adaptability model for human thermal comfort in different climatic zones, and proposed a passive climate design strategy to ensure indoor thermal comfort. In the study of "building climate analysis and design strategy", Yang Liu divides our country into nine climatic design zones, and systematically studies the methods of architectural climate design analysis and climate design strategy. In the context of architectural environment control, Song Dexuan expounded the concept of architectural environment control. He has put forward a series of effective design methods and technical means to improve the quality and quality of building environment by means of architectural design.

3. Methodology

3.1. Demonstration building measurement

The experimental research mainly takes the demonstration building, Huazhong University of Science and Technology School of architecture and urban planning as the research platform. After the demonstration project was completed in June 2009, the actual research work began. Because the design of the system is in the stage of experimental exploration, the system has been modified several times on the basis of experiments, and the performance of the system has been improved. Through a series of experimental tests and comparison, the purpose of demonstration building is to study the impact of GUD system operation on the indoor thermal environment of demonstration buildings. And the parameters and control strategies associated with the operation of GUD system are obtained. By understanding the advantages and disadvantages of the system and the characteristics of energy consumption, the experience has been obtained, which lays the foundation for further research on integrated design of GUD systems in low energy buildings. The actual research on the demonstration building is mainly aimed at the operation conditions of the winter and summer systems, the indoor and outdoor thermal environment of the building, the inner and outer surface of the envelope and the temperature. During the transition season, we should pay attention to the change of indoor and outdoor thermal environment, and adjust the building thermal environment with passive means.

The climatic characteristics of the base: according to the data measured by the energy and environment monitoring station, Fig. 1 shows the changes of monthly average temperature and relative humidity in 2015 and 2016. Figure 2 shows the outdoor temperatures in December 2015 and January 2016 for two months. Figure 3 shows the power consumption of the GUD system in 2016 July.

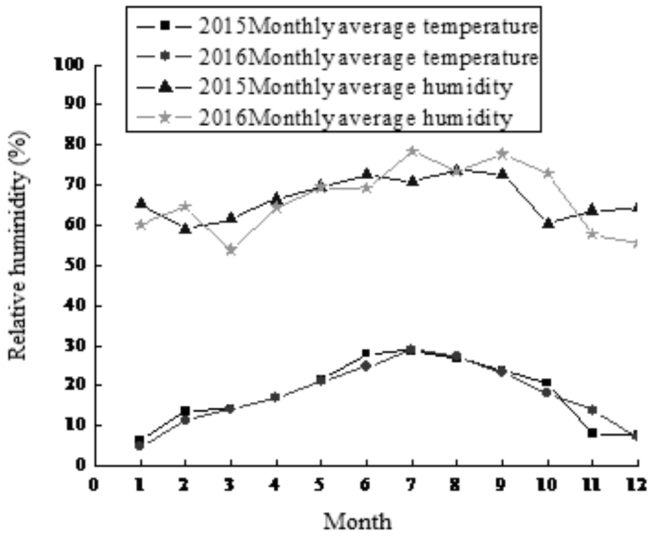


Fig. 1. Comparison of monthly average temperature and relative humidity between 2015 and 2016

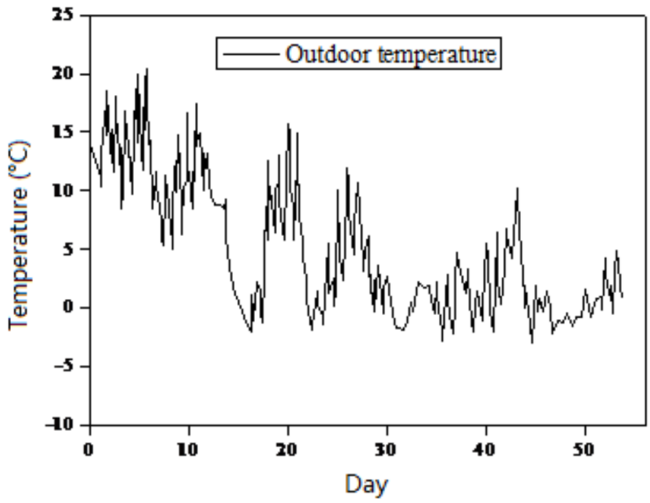


Fig. 2. Outdoor air temperature chart for two months in December 2015 and January 2016

3.2. The CFD simulation of GUD system

Through the actual measurement of the demonstration building, the characteristics and operation rules of GUD system are understood. The integrated design of GUD system in architecture is studied by computer simulation. Taking the office

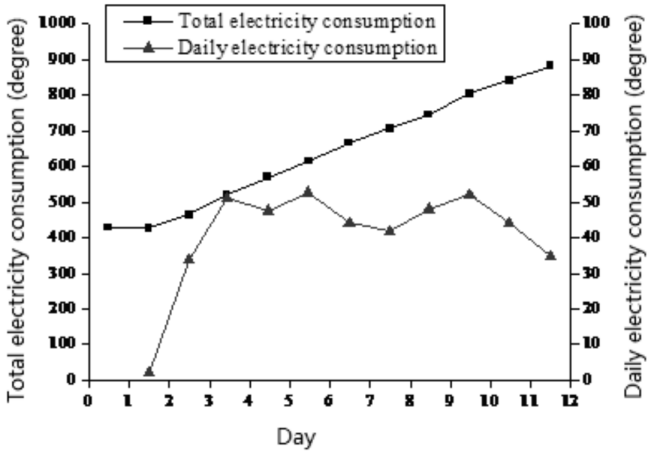


Fig. 3. GUD system power consumption in 2016 July

buildings in hot summer and cold winter area as an example, the low energy consumption architecture design method of integrated GUD system is discussed. The design method of each part of the system combined with the building, as well as the design points of each phase of the architectural design process are summarized.

The integrated design of GUD system in architecture is studied by computer simulation software, and the strategy of improving the thermal environment of building interior is discussed. CFD software Air Pak is used to simulate the thermal environment and wind environment of the building. The influence of different air supply modes, different wall materials and structural forms on indoor thermal environment is studied. The integration design of the system in the building is optimized. The actual data are used to test the computer model in order to ensure the correctness of the simulation. It is necessary to study the change of the thermal environment inside the building after the operation of the GUD system, so the unsteady state simulation (transient simulation) of the model building is carried out. In the Air Pak, the unsteady simulation calculation is performed. Indoor thermal environment changes were calculated within 10 hours after the system was switched on. The interval is set to 10 minutes, and the contrast between simulated and measured results is shown in Fig. 4.

4. Result analysis and discussion

4.1. Analysis of measured results

Based on the research platform of the base, the low energy building thermal environment of integrated GUD system was measured. It can be seen from Fig. 2 that the climate has a bi-directional characteristic. The months with the lowest average temperature are January and December, and the months with the highest average

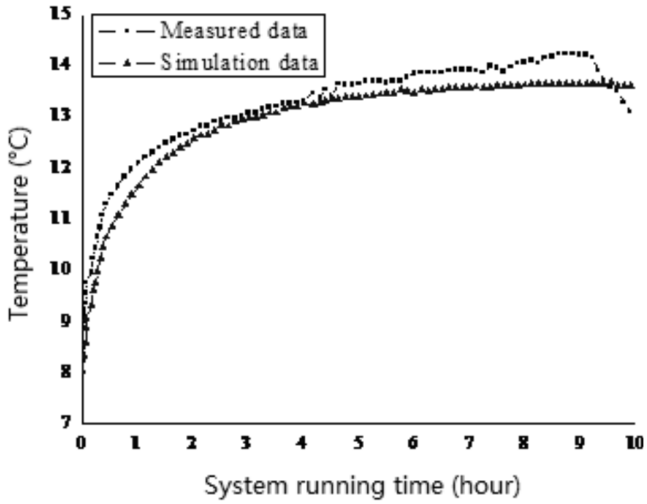


Fig. 4. Unsteady simulation test

temperature are July and August. The relative humidity from June to September is high, and the relative humidity from November to February is low. As can be seen from Fig. 3, the base area experienced several cold wave attacks in December and January. The minimum temperature appeared on December 2015 14–17, and on January 2016, 1–10 and 17–23 days, and the average temperature was about 0 degrees celsius. By testing the operation of the GUD system in the demonstration building under extreme weather conditions, we can see that the stability of the indoor thermal environment is better when the GUD system is turned on.

As can be seen from Fig. 4, when the GUD system is switched on, the average daily power consumption is about 50 degrees. By calculation, the underground cold energy transferred to the room can be obtained. Furthermore, the energy efficiency ratio of the system is 9.56 in winter and the energy efficiency ratio is 5.240 in summer. Thus, the GUD system can create a good thermal comfort environment and save energy balance between buildings.

4.2. Analysis of simulation results

The effect of different tectonic forms of wall on the thermal environment of the building is simulated in different ways of winter and summer. The indoor thermal environment of the integrated GUD system is simulated and analyzed, and the distribution law of indoor thermal environment is further studied. By using the measured data, the CTD simulation of the model building is tested and the relevant settings are modified, so as to reduce the error of simulation and ensure the correctness of the simulation. The simulation shows that it is in good agreement with the measured ones in the first five hours. In the latter part of the simulation, the temperature changes at each point is stable, while the measured temperature continues to rise. This is due to the indoor human activities and the solar radiation factors. On the

whole, however, the simulated and measured temperature trends are basically consistent. The unsteady simulated data in winter and summer are in good agreement with the measured data, which shows that the unsteady state simulation can better simulate the indoor thermal environment change after the equipment is turned on.

5. Conclusion

The utility model meets the architectural function goal. The demonstration building is not strict with the indoor thermal environment. After the GUD system is switched on, the thermal environment inside the building can be improved rapidly. With the operation of the system, it gradually stabilized. Through the use of computer CPD simulation method, the building indoor thermal environment is simulated. From the simulation of the distribution of indoor thermal environment, it can be seen that the distribution of indoor thermal environment is not uniform when the GUD system is opened. There is a certain degree of thermal stratification. The workplace air blowing methods are used to improve the environment of the work area, thereby reducing the overall environmental comfort requirements. By increasing the amount of indoor air supply, the thermal environment in the building can be improved. However, excessive wind speed will affect the indoor temperature stratification

References

[1] P. Raman, S. Mande ,V. V. N. Kishore: *A passive solar system for thermal comfort conditioning of buildings in composite climates.* Solar Energy *70* (2001), No. 4, 319–329.

[2] J. Miriel, L. Serres, A. Trombe: *Radiant ceiling panel heating–cooling systems: experimental and simulated study of the performances, thermal comfort and energy consumptions.* Applied Thermal Engineerings *22* (2002), No. 16, 1861–1873.

[3] P. O. Fanger, J. Toftum: *Extension of the PMV model to non-air-conditioned buildings in warm climates.* Energy and Buildings *34* (2002), No. 6, 533–536.

[4] Z. Ghiabaklou: *Thermal comfort prediction for a new passive cooling system.* Building and Environment, *38* (2003), No. 7, 883–891.

[5] V. Serra, F. Zanghirella, M. Perino: *Experimental evaluation of a climate façade: Energy efficiency and thermal comfort performance.* Energy and Buildings *42* (2010), No. 1, 50–62.

[6] R. C. Richman, K. D. Pressnail: *A more sustainable curtain wall system: Analytical modeling of the solar dynamic buffer zone (SDBZ) curtain wall.* Building and Environment *44* (2009), No. 1, 1–10.

[7] T. Tsoutsos, E. Aloumpia, Z. Gkouskosa, M. Karagiorgas: *Design of a solar absorption cooling system in a greek hospital.* Energy and Buildings *42* (2010), No. 2, 265–272.

[8] E. Gratia, I. Bruyère, A. de Herde: *How to use natural ventilation to cool narrow office buildings.* Building and Environment *39* (2004), No. 10, 1157–1170.

[9] S. Niu, W. Pan, Y. Zhao: *A BIM-GIS integrated web-based visualization system for low energy building design.* Procedia Engineering *121* (2015), 2184–2192.

[10] G. Zapata-Lancaster, Ch. Tweed: *Tools for low-energy building design: an ex-*

ploratory study of the design process in action. Architectural Engineering & Design Management *12* (2016), No. 4, 279–295.

Received May 7, 2017