Acta Technica *62* (2017), No. 6B, 995–1004

Comparison method of multi-source data fusion based on fuzzy evidence theory

Yuliang Jiao[1], Lian'e Li[2], Yi Xiao[1]

Abstract. To improve the accuracy of comparative analysis of Chinese and foreign educational system, the Paper proposes a method of comparative analysis of Chinese and foreign educational system under multi criteria and multi criteria. First, by proposing multi criteria and multi criteria algorithm, the decision model of target recognition is set for given mutually exclusive and independent sets of atomic hypotheses, and then the processing analysis of data is achieved by making use of fuzzy evidence theory data fusion algorithm and prototypical data pre-processing algorithm; next, based on the Wuhan Pharmaceutical Industry, the income of educational system is analyzed with proposed algorithm, which reveals the importance of industry educational training on industry development; and last, the opinions and suggestions are provided for development of educational system in China.

Key words. Multi criteria and multi criteria Chinese and foreign education Industry education.

1. Introduction

As early as 1983, when comrade Deng Xiaoping wrote an inscription for Beijing Jingshan School, he has proposed the strategic thought "Education shall be oriented towards modernization, world and future", and in recent 20 years, under the guidance of Deng Xiaoping education thought, the educational business in China has achieved unprecedented success, which is a recognized fact, however, we shall see that, for "three orientations" of education in China, there is still a big gap in educational system, especially in the educational administration system, compared with some developed countries, so that more attention shall be paid to this. The Paper attempts to compare the advantages and disadvantages of Chinese and foreign edu-

[1]Department of foreign languages, Chongqing Three Gorges University, 404000, China
[2]School of postgraduate, Kunming University of Science and Technology, 650000, China

cational administration system, refer and absorb foreign advantages and experiences of educational administration system especially the developed country to make up for the deficiencies of us, thus making educational administration system in China face the modernization, world and future. The educational administration system means organization, management and lead of organs of state power on education work in accordance with a certain policy to achieve specific objective, of which the main contents include educational administration organization, educational administration legislation, educational planning, educational expenditure and educational supervision system.

Modern educational administration system emerges with the establishment of modernization public educational system, which is the product of social development to a certain historical stage, of which the development is affected by the development level of social productive forces in one hand, and in the other hand, also affected by the determination of social politics and economic system. On account of different national conditions and different social economic development level, there is difference in educational administration system for various countries in the world. The educational administration system in China also has an independent and complete system. Since the third Plenary Session of the 11th CPC Central Committee of the party, great progress has been made in educational enterprise in China, which can be compared with that of some developed countries. Today, China implements the socialist market economy system, equipped with the conditions and foundation to compare with the educational administration system of developed country and each side can be in line with international practice. Given this, the Paper horizontally analyzes the educational administration system in China by selecting some developed countries such as American, Japan, Britain, Germany, France, the former Soviet Union, etc. as reference for comparison.

The Paper establishes a method of comparative analysis of Chinese and foreign educational system under multi criteria and multi criteria and analyzes the difference and common characters in order to correctly understand the problems and defined the development direction in the future. The Paper analyzes and introduces the effectiveness of algorithm based on case verification in the end.

2. Multi criteria and multi criteria algorithm

2.1. Decision model of target recognition

Given mutual exclusive and independent set of atomic hypothesis Θ, and set r possible recognized target types in common, with domain of discourse $U = \{1, 2, \cdots, r\}$ for the ordered set composed of corresponding sequence number of target types, then $\Theta = \{\theta_1, \theta_2, \cdots, \theta_r\}$, where Θ represents frames of discernment and θ_j means the atomic hypothesis of frames of discernment. And set n information source to participate in data fusion of target recognition, with $S = \{1, 2, \cdots, n\}$ for the ordered set composed of corresponding sequence number of n information source, where n information source means n evidence bodies.

Because the sources of knowledge available to each evidence body are different in target recognition, each evidence body provides the information for target recognition at its own particular level and this information has a certain fuzziness. Therefore, the class of each evidence body for recognized target can be expressed by fuzzy set in domain of discourse U and the target recognition decision report of the ith evidence body is C_i recorded by $\forall i \in S$, that is:

$$c_i = \sum_{j \in U} \frac{m_{ij}}{j}, i \in S\,, \tag{1}$$

where, $m_{ij} = \mu c_i(j)$ and $0 \leq \eta_{ij} \leq 1$, namely $\sum_{j=l}^{r} m_{ij} = 1$, $0 \leq \sum_{j=l}^{n} m_{ij} \leq n$, means that the judgment recognized target of the ith evidence body belongs to degree of membership of the j class. m_{ij} means the supporting degree of atomic hypothesis set θ_j in the frames of discernment, but all m_{ik}, $k \neq j$ means the support to complementary set of θ_j or opposition to θ_j, and here m_{ij} is regarded as the supporting degree of simple support function of focal element θ_j, and m_{ik}, $k \neq j$is regarded as supporting degree of simple support function of focal element $\neg\theta_j$. A separable support function for focal element $\neg\theta_j$ can be obtained by merging m_{ik}, $k \neq j$, of which the supporting degree is $1 - \prod_{k \neq j} (1 - m_{ik})$. Therefore, the target recognition result in the form of fuzzy set can be expressed by the following form, namely the fuzzy basic probability distribution function is defined as follows:

$$M_j^i(\theta_j) = \mu c_i(j) = m_{ij}\,, \tag{2}$$

$$M_j^i(\Theta) = 1 - m_{ij}\,, \tag{3}$$

$$M_{\neg j}^i(\neg\theta_j) = 1 - \prod_{k \neq j} (1 - m_{ik})\,, \tag{4}$$

$$M_{\neg j}^i\,(\Theta) = 1 - M_{\neg j}^i(\neg\theta_j) = \prod_{k \neq j}\,(1 - m_{ik})\,, \tag{5}$$

where, $\forall i \in S$, $j, k \in U$. Therefore, for the fuzzy belief function $Bel_i(\theta_j) = M_j^i(\theta_j) \oplus M_{\neg j}^i(\neg\theta_j)$ of θ_j, that is:

$$Bel_i(\theta_i) = \frac{m_{ij} \prod_{k \neq j} (1 - m_{ik})}{1 - m_{ij} \left[1 - \prod_{k \neq j} (1 - m_{ik})\right]}\,, \tag{6}$$

where, $\forall i \in S$, $j, k \in U$. So that the possibility distribution M_K^i of target recognition of the ith evidence body is:

$$M_K^i = \sum_{j \in U} \frac{Bel_i(\theta)}{j}, i \in S\,. \tag{7}$$

In the formula, K is the variable of value in U, which represents that the possibility of K value is the supporting degree of atomic hypothesis θ_j in the frame of discernment after the merger of conflict evidence in the same evidence body. This possibility distribution means the method of target recognition decision result and is a kind of soft-decision in nature, that is, the extended membership degree conducts fuzzy classification between $[0, 1]$, of which the advantage is to obtain whole decision-making on a larger scale compared with general hard-decision when data of target recognition is integrated.

2.2. Fuzzy evidence theory data fusion algorithm

The data fusion at decision level adopts two methods in common, including evidence theory and Bayesian theory, which is a theory that combines multiple evidence to make decisions. Compared with probabilistic decision theory (such as Bayesian theory), it not only can handle the uncertainty caused by inaccurate knowledge, but also can deal with uncertainty caused by unclear. It can satisfy the axiom system which is weaker than the probability theory, and the evidence theory becomes the probability theory when the probability value is known. Probability theory as it were the special circumstances of evidence theory. In the decision fusion of target recognition, because environment and decision condition of each evidence body are different, thus affected by various factors, and for the purpose of fusing the target identification results of each evidence body, that is the general performance of recognized target shall be described reasonably as far as possible to achieve the purpose of correct recognition for target. Because the decision results of target recognition is made by all evidence bodies in the form of possibility distribution, thus, the results of target recognition for n evidence bodies after fusing is also the possibility distribution of value in domain of discourse$U$, denoted as M_K, namely:

$$M_K = \sum_{j \in U} \frac{M_K(j)}{j} = \sum_{j \in U} \frac{Bel(\theta_j)}{j}\,, \tag{8}$$

where, $Bel(\theta_j) = M_K(j)$ means the possibility degree of K value j after results of target recognition for n evidence bodies are fused. Obviously, $Bel(\theta_j)$ is obtained after fusing $Bel_1(\theta_j), Bel_2(\theta_j), \cdots, Bel_n(\theta_j)$, so the key to calculate $Bel(\theta_j)$ is the fusion of $Bel_1(\theta_j), Bel_2(\theta_j), \cdots, Bel_n(\theta_j)$. Here the results of data fusion of target recognition are calculated by fuzzy evidence theory method based on the combination of fuzzy theory and evidence theory. Given evidence body $E = \{(X_1, m_1, \eta_1), \cdots, (X_g, m_g, \eta_g)\}$ composed of three items, respectively for classification, basic probability distribution function BPA and their fractional function, in which $\eta_j = f(X_j, 0)$. The fractional function of some classification in evidence E is defined as:

$$f(X_j, 0) = S(X_j)/S, j = 1, 2, \cdots, g, \tag{9}$$

where, $S(X_j)$ is the area of some classification in evidence E and S means the area of all classification in such evidence body. In fact, this kind of fractional function is similar to a prior probability. In the evidence body discussed in the paper, because the same evidence body is the image of the same resolution ratio actually, so the area ratio of some classification in all classification is the pixel ration of that in fact. In the fusion process of evidence body, the use of this fractional function is to give a certain weight for belief function generated by orthogonal combination rule in evidence theory in effect. If given two evidences:

$$E_i = |\{X_{i1}, m_{i1}\ \eta_{i1}), \cdots, (X_{ig}, m_{ig}, \eta_{ig})\}|\ , \tag{10}$$

$$E_k = \{(X_{k1}, m_{k1}, \varphi_{k1}), \cdots, (X_{kh}, m_{kh}, \varphi_{kh})\}\ , \tag{11}$$

where,

$$\begin{cases} \eta_{ij} = f\left(X_{ij}, 0\right), j = 1, 2, \cdots, g \\ \varphi_{kl} = f\left(X_{kl}, 0\right), l = 1, 2, \cdots, h \end{cases} \tag{12}$$

then the constitution of fused new evidence body E is:

$$E = E_i * E_k = \{(X_{ij} \cup X_{kl}), Bel(\theta_{jl}), \xi_{jl}\}\ , \tag{13}$$

fractional function of some classification in the fusion process of multi evidence bodies is:

$$\xi_{jl} = f(X_{ij} \cup X_{kl}, 0) = \frac{(S(X_{ij}) + S(X_{kl}))}{2S}\ . \tag{14}$$

Set α as corresponding coefficient function after fusion of some classification in evidence body, defined as follows:

$$\alpha[(X_{ij}, \eta_{ij}), (X_{kl}, \varphi_{kl})] = \frac{f(X_{ij} \cup X_{kl}, 0)}{(\eta_{ij}\varphi_{kl})}\ ,\ , \tag{15}$$

then the total fractional function after the merger of two evidence is:

$$\alpha(E_i, E_k) = \sum_{j,l} m_{ij} m_{kl} \alpha[(X_{ij}, \eta_{ij}), (X_{kl}, \varphi_{kl})] \tag{16}$$

$$Bel(\theta_{jl}) = \sum_{j,l} m_{ij} m_{kl} \alpha\left[(X_{ij}, \eta_{ij}), (X_{kl}, \varphi_{kl})\right]. \tag{17}$$

In the fusion process of all n evidence bodies, because each evidence body is independent, so the sequence can not be met for the fusion of n evidence bodies, and any two evidence bodies can be merged to obtain fused value of belief function, fusing with other evidence bodies gradually, and last the fusion belief function $Bel(\theta_j)$ of atomic hypothesis set in the frame of discernment can be obtained. However, after calculating $Bel(\theta_j)$ to $\forall j \in U$, the fuzzy basic probability distribution functions in all evidence bodies are determined in accordance with the following decision rule:

$$M_K = \begin{cases} 1, & if\ Bel(\theta_j) > 0.5 \\ 0.5, & if\ Bel(\theta_j) \le 0.5 \end{cases} \tag{18}$$

2.3. Data pre-processing based on prototype

To determine the membership degree of each model to be classified, given a reference set χ, that is (x_i, μ_i) in $F(\chi)$ is served as training set of (χ, U) space. Each model in training set has fuzzy degree which belongs to domain of discourse U, and the output of fuzzy degree of each model in such training set is served as focal element set of such model, namely $F(\mu_i) = \{\mu_{i1}, \mu_{i2}, \cdots, \mu_{ir}\}$, and for a new model x to be classified, the focal element set of fuzzy basic probability distribution function m_i is $F(m_i) = F(\mu_i)$. Because the increase of traditional DS evidence with the atomic hypothesis focal element in frame of discernment, the amount of calculation shows the exponential growth tendency. To reduce the increase of amount of calculation effectively, in the classification process of each information source that is itself of evidence body, the training set shall be classified preliminarily at first to obtain the prototype of model to be classified, and the corresponding fuzzy basic probability distribution function is obtained from model to be classified in each evidence body through the information provided by the prototype. Assuming that the two sets of I input and J output prototypes respectively are:

$$C = \{c_i \in X, i \in \{1, 2, \cdots, I\}\}\ , \tag{19}$$

$$E = \{e_i \in U, j \in \{1, 2, \cdots, J\}\}\ . \tag{20}$$

The prototype set can be obtained by fuzzy C mean value and fuzzy Kohonen neural network and other clustering method. Assuming that (x_k, μ_k) is sourced from the sample of M independent random variable (X_k, U_k) of probability space $((X, U), A, P_{XU})$, and the probability function $P_{X|U}$ $(U \in E_j\, |X \in C_i)$ is expressed by fractional function $\eta_j = f(X_j, 0)$. If $\pi_{ij} = P_{XU}(X, U)$, M_{ij} belongs to the number of random variable of (x_k, μ_k) couple of $C_i \times E_j$, by defining $M_{ij} \sim B(m, \pi_{ij})$, the strong law of large numbers proves the estimation of fraction function in theory:

$$\begin{aligned} f(X_j, 0) = P_{U|X}\ (E_j\, |C_i) &= \frac{P_{XU}(E_j\, |C_i)}{P_X(C_i)} \\ &= \frac{m_{ij}}{\sum_j m_{ij}} = \frac{S(X_j)}{S(X)}\ . \end{aligned} \tag{21}$$

The output prototype e_j of set will be expressed by the membership degree which belongs to domain of discourse U. For each model to be classified, in accordance with a certain similarity as standard, each c_i in I input prototype provides a output for model to be classified, in this way, the focal element set $|F(m)| \leq 2^J$ for model to be classified will greatly reduce the number of focal elements with such prototype method, thus reducing the amount of calculation in evidence merger effectively.

3. Experimental analysis

3.1. Experimental results

Next, the present condition of vocational education for Wuhan pharmaceutical enterprises will be illustrated by the survey data in 2016 of 22 Wuhan backbone pharmaceutical enterprises (hereinafter referred to as backbone enterprise) under the government of Wuhan Food and Drug Administration.

①Total number of staff. There are 6600 employees in the backbone enterprises, including 3540 technical workers, accounting for 53.6 of the total number of workers, which shows that Wuhan pharmaceutical enterprise is the "production and management type" enterprises.

②Technical grade. In 3540 technical workers, senior workers account for 7.6% of total technical worker, intermediate workers account for 18.5% and primary workers account for 27.6%. Senior technician accounts for 1.3% of total technical worker and technician accounts for 2%. Therefore, high-skilled talent accounts for 11% of total technical workers.

Table 1. Data for total industrial output, total worker and educational level of workers of wuhan pharmaceutical industries from 1992-2006

Year	Total	Workers	Junior	High	Mid	College	Unicer
1992	10.38	10760	3644	4504	2467	113	32
1993	11.37	14700	4904	6080	3434	238	44
1994	13.69	11582	3443	4933	2816	332	58
1995	16.05	12018	3384	5088	2904	562	80
1996	24.17	12200	3406	5138	2954	602	100
1997	36.75	12481	2447	5198	3054	652	130
1998	39.44	14494	3847	5798	3702	952	195
1999	46.55	14550	3854	5807	3724	964	201
2000	53.72	13450	3294	5569	3370	1009	208
2001	49.76	15020	3479	5884	4035	1359	263
2002	63.33	15590	3552	6057	4159	1504	318
2003	60.96	12594	2032	5302	3734	1238	288
2004	63.77	12600	2034	5303	3736	1239	288
2005	54.96	10315	1119	4573	3381	1039	203
2006	69.94	10783	1152	4653	3546	1161	271

According to theory and construction model of multi criteria and multi criteria algorithm, the fusion analysis is conducted for information association between total industrial output value from 2002-2016 and total workers and average years of education of workers with the data of total industrial output value from 2002-2016 and total workers and average years of education of workers of Wuhan pharmaceutical industry. The original data are shown in Table 1.

After operating the above multi-source data fusion procedure, results are shown in Table 2.

Table 2. Results of multi- source data fusion

Residual				
Min	1Q	Median	3Q	Max
-1814	-710.9	-262.7	694.1	3199
Minimum value	25%	50%	75%	Maximum value

Correlation coefficient			
Value	Std.Error	t.value	Pr(>|t|)
-48040.2911	11098.3017	-4.3286	0.0010
0.3052	0.2159	1.4137	0.1829
4123.8507	886.523	4.5628	0.0004
Estimation value of coefficient	Standard deviation	T value	P value

The Table 2 shows that the coefficient of nu is 0.3052, but the P value is 0.1829, not obvious. The coefficient of PJ is 4123.8507, which means to keep the workforce unchanged, with an average increase of 1 years in education, the total industrial output value of pharmaceutical industry will increase by 41.2385 billion yuan, again showing that "the average educational level of technical workers has a very significant impact on the development of the Wuhan pharmaceutical industry".

3.2. Comprehensive analysis

According to above empirical analysis data, the main problems of Wuhan industry vocational education are analyzed comprehensively from enterprise, school, government and industry in the form of "fishbone diagram" as shown in Fig.2.

4. Countermeasures and suggestions

(1) Increase the input in industry vocational education. First, increase support for vocational education in the industry. We shall gradually increase the input of public finance to industry vocational education, in accordance with the principle of "do thing with money", and implement the new financial allocation mechanism of which the main indexes are the number of student at school and mean costs of student training; and we shall arrange special fund to focus on the support for key universities and colleges, key specialty and construction of training base, teacher training and poverty alleviation program, etc.

(2) Create good external conditions for rapid development of industry vocational education. First, strengthen the lead and comprehensive management on industry vocational education. Practically put the development of vocational education in an important position in the overall situation and coordinate development planning,

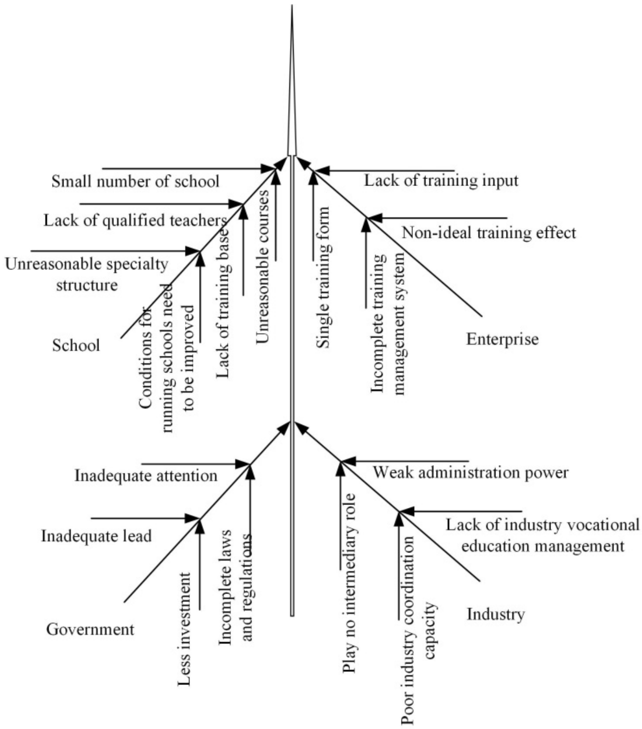

Fig. 1. Problems of educational system in China

resource allocation, condition guarantee and policy measures to provide a powerful public service and good development environment for vocational education.

(3) Quicken specialty construction and course reform. First, implement the specialty construction in counterparts of middle and higher vocational schools in accordance with demand of industry development to face high-skilled talents at different level of industry training on demand. Second, deepen course reform for vocational education, strengthen construction of top-quality courses and textbook, and conduct course lecture combined with enterprise at the same time to achieve match of theory and reality. Strive to establish modularization course system which takes ability as the standard, exercise as the main line and specialized courses as the main body. Third, strengthen information construction of vocational education to promote the application of modern vocational education technology in teaching.

Acknowledgement

The college ESP course curriculum model and construction based on the demand analysis under Grant No. JG60807.

References

[1] Krishnamurthy S, Tzoneva R: (2011) *Comparative analyses of Min-Max and Max-Max price penalty factor approaches for multi criteria power system dispatch problem with valve point effect loading using Lagrange's method*[C]// International Conference on Recent Advancements in Electrical, Electronics and Control Engineering. IEEE, 2011:36-43.

[2] Munkhtsetseg N, Garmaa D, Uyanga S: (2014) *Multi-criteria Comparative Evaluation of the E-Learning Systems: A Case Study*[C]// International Conference on Ubi-Media Computing and Workshops. IEEE, 2014:190-195.

[3] Viswanadhan K G: (2009) *Quality indicators of engineering education programmes: a multi-criteria analysis from India*[J]. International Journal of Industrial & Systems Engineering, 4:270-282(13).

[4] Ivlev I, Jablonsky J, Kneppo P: (2016) *Multiple-criteria comparative analysis of magnetic resonance imaging systems*[J]. International Journal of Medical Engineering & Informatics, 8(2):124–141.

[5] Zhang Z, Sherman R, Yang Z, et al.: (2013) *Integrating a participatory process with a GIS-based multi-criteria decision analysis for protected area zoning in China*[J]. Journal for Nature Conservation, 21(4):225-240.

[6] Chinese D, Nardin G, Saro O: (2011) *Multi-criteria analysis for the selection of space heating systems in an industrial building*[J]. Energy, 36(1):556-565.

[7] Wang J J, Zhang C F, Jing Y Y: (2010) *Multi-criteria analysis of combined cooling, heating and power systems in different climate zones in China*[J]. Applied Energy, 87(4):1247-1259.

[8] Chang C L, Lin Y T, Chiueh P T: (2014) *Single Criterion and Multiple Criteria Analysis: A Comparison of Water Quality Monitoring Designs for a River System*[J]. Water Resources Management, 28(3):645-655.

[9] Jayant A, Giri V, Ojha N: (2015) *Arvind Jayant, V.Giri and Nidhi Ojha(2015) "Comparative Analysis of Multi-Criteria Decision Making Methodologies & Implementation: A Case Study of Warehouse Location Selection"*[C]// The Proceedings of International Conference on Advanced and Agile Manufacturing Systems.

[10] Hajoui O, Dehbi R, Talea M, et al.: (2015) *An advanced comparative study of the most promising NoSQL and NewSQL databases with a multi-criteria analysis method*[J]. Journal of Theoretical & Applied Information Technology.

[11] Li Z C, Huang H J: (2004) *A Comparison of System and User Performance under Single-Criterion and Multi-Criteria Route Guidance Systems*[C]// Eighth International Conference on Applications of Advanced Technologies in Transportation Engineering. 2004:276-280.

[12] Jou Y T, Yang K H, Liao M L, et al.: (2016) *Multi-criteria failure mode effects and criticality analysis method: A comparative case study on aircraft braking system*[J]. International Journal of Reliability & Safety, 10(1):1.

[13] Ulukan Z, Kahraman C, Cebeci U: (2003) *Multi-criteria supplier selection using fuzzy AHP*[J]. Logistics Information Management, 16(6):382-394.

Received May 7, 2017