International Journal of u- and e- Service, Science and Technology
Vol.9, No. 6 (2016), pp.345-354
http://dx.doi.org/10.14257/ijunesst.2016.9.6.31

China Coal Industry International Competitiveness Research Based on Unascertained Clustering

Xiang Chen [1, a], Yang Liu [2, b], Yuxia Liang [3, c] and Xin Zhao [4, d]

[1]*School of architecture, Hebei University of Engineering, Hebei Handan 056038, China*
[2]*The personnel department, Civil aviation university of China, Tianjin 300300, China*
[3]*School of architectural engineering, Shijiazhuang vocational college of science and technology, Shijiazhuang 050000, China*
[4]*China Xinxing Baoxin Construction Corporation, Beijing 100079, China*
[a]*caisuan@126.com*, [b]*461515572@qq.com*, [c]*2274520312@qq.com*, [d]*1113605@163.com.*

Abstract

The unascertained clustering is a new clustering method, which combines unascertained theory and clustering theory to construct the unascertained measure, and uses the unascertained measure as set membership to indicate the membership relation between the samples with the different classes. It overcomes the disadvantage of means clustering algorithm, that a sample definitely belongs to a class, which made greater progress than -means clustering. There are complex non-linear relationship between the coal industry competitiveness and various factors. The article established the evaluation influencing factors system of coal industry international competitiveness. 6 unascertained clustering method to cluster competitiveness. It found out each class center, and gave the membership degree of the samples belong to each class, which better resolved the problem of classifying the coal industry international competitiveness.

Keywords*: unascertained clustering, classified character weight, the coal industry, competition ability*

1. Introduction

From the current actual development situation of China coal industry, its productivity is not commensurate with its principal status, and its development faces many problems. China as the world’s second largest coal reserves country, and the first major coal producer, in today’s highly competitive international environment, how to broad participate in international cooperation and competition, optimize the resources allocation, promote the industry sound development and competitiveness, is an important problem in China coal industry[1]. The key to development of China coal industry is to strive to narrow the gap on efficiency, effectiveness, safety and environmental protections with the international counterparts, which have China coal industry not only has the domestic competition, but also improve the international competitive advantage. Therefore, to explore China’s coal industry competitiveness impact factors, evaluation model, and the countermeasures to improve the competitiveness of China’s coal industry, has become a strong reality and urgency of the major theoretical and practical issue in China’s energy field, that in the face of globalization of world economy opportunities and challenges.

Study on coal industry international competitiveness in China is still relatively scarce and long-term stay in the enterprise microscopic fields. Medium research on

ISSN: 2005-4246 IJUNESST

industry competitiveness mostly learns from foreign coal industry competitiveness theory, the research also remain in the simple analysis. Empirical research is rarely few. It is difficult to adapt to the development of china’s coal industry. Although He Jin-xiang conducted quantitative study, but it been processed with index single and inconsiderable. Researchers in domestic are unanimous in the international competitiveness of China's coal industry; some insist that China have certain competition ability in coal industry, while others think China is lack of international competence [2]. Given this, based on the analysis on affecting factors of coal industry international competitiveness, cluster analysis on coal industry competitiveness by means of unascertained clusters method been used in the paper. Through the research, the center of the clusters and the subjection degrees of all the samples are found, which effectively realized quantitative study on coal industry international competitiveness Study on coal industry international competitiveness in China is still relatively scarce and long-term stay in the enterprise microscopic fields. Medium research on industry competitiveness mostly learns from foreign coal industry competitiveness theory, the research also remain in the simple analysis. Empirical research is rarely few. It is difficult to adapt to the development of china’s coal industry. Although He Jin-xiang conducted quantitative study, but it been processed with index single and inconsiderable. Researchers in domestic are unanimous in the international competitiveness of China's coal industry; some insist that China have certain competition ability in coal industry, while others think China is lack of international competence [2]. Given this, based on the analysis on affecting factors of coal industry international competitiveness, cluster analysis on coal industry competitiveness by means of unascertained clusters method been used in the paper. Through the research, the center of the clusters and the subjection degrees of all the samples are found, which effectively realized quantitative study on coal industry international competitiveness.

2. Unascertained Clustering Algorithm

The kind of unsupervised clustering based on unascertained set in the paper is distinct from the existing cluster method in that two points: the characteristics of each cluster plays different roles according to unascertained clustering and the difference is expressed by classification weight; The unascertained clusters methods conjoins with unascertained theory and clusters theory to establish unascertained measure as a muster to denote the subjection relations between the samples and the classifications.

International competition ability of coal industry has complex nonlinear relation with various factors such as industrial environment, industrial structure, and industrial behavior. The paper analyses influencing factors and collect statistic dataset of 11 major coal production countries. This paper use the method of unascertained clustering to cluster competition ability, find out the center of each category, and to make sample belong to subordinate degree of each category. It resolves the coal industry competition ability cluster problem well.

2.1. Unascertained Means Clustering

Assume there are d features that infected the sample classification. The observed value of sample x_i about characteristic j is x_{ij}. Every dimension $\{x_{1j}, x_{2j}, \cdots, x_{Nj}\}(j = 1,2,\cdots,d)$ was standardized. In this way the sample x_i can be expressed as a point in the d dimensional feature space:

$$x_i = (x_{i1}, x_{i2}, \cdots, x_{id})'(i = 1,2,\cdots,N)$$

If N samples in the d dimensional feature space were divided into C categories, $\Gamma_k(k = 1,2,\cdots,C)$ represented the k th category. m_k is the class center vector in the Γ_k category : $m_k = (m_{k1}, m_{k2}, \cdots, m_{kd})^T$, $(k = 1,2,\cdots,C)$. Obviously, this is a certainty classification.

However, when the m_k was used as a representative approximately indicate the samples in Γ_k category, and the distance between the sample x to class center m_k was used as the approximate measure between x and Γ_k , which in fact, has already turn above certainty category into uncertainty category. Usually, this uncertainty is more close to actual classification. However, if we want to make this kind of uncertainty useful, we must be able to reasonably determine each sample's membership degree to every category. If $\mu_{\Gamma_k}(x)$ was used to represent the x membership degree belong to Γ_k category, as the existence of uncertainty, we can't know the true value of $\mu_{\Gamma_k}(x)$. As the value of $\mu_{\Gamma_k}(x)$ is relative, how to more reasonably determine the value of $\mu_{\Gamma_k}(x)$ is a basic research content of uncertainty classification.

2.2. Instructiveness Acknowledge Acquisition in Means Clustering

N samples in d feature dimensions space were divided into C categories, the class center of $\Gamma_k(k = 1 \sim C)$ is m_k :

$$m_k = (m_{k1}, m_{k2}, \cdots, m_{kd})^T, (k = 1 \sim C) \tag{1}$$

Make

$$\overline{m} = \frac{1}{C}\sum_{k=1}^{C} m_k = (\overline{m}_1, \overline{m}_2, \cdots, \overline{m}_d) \tag{2}$$

Then $\overline{m}$ is the centroid in the particles which composed by C category centers.

$$\sigma_j^2 = \frac{\alpha_i}{C}\sum_{k=1}^{C}(m_{kj} - \overline{m}_j)^2,\ j = (1 \sim d) \tag{3}$$

We proposed that the value of variance σ_j^2 reflect the value concentration and dispersion degree of C class centers $m_1, m_2, \cdots, m_C$ on j th component, in which α_i was adjust constant, which simplest case let $\alpha_i = 1$.

① Suppose $\sigma_j^2 = 0$, then the j th component of C class centers $m_1, m_2, \cdots, m_C$ were the same. Now, if looking from the j th component axis of the d dimension features space, we can find C class centers coincide into one point. In other words, the feature j is malfunction on dividing N samples into different C categories. Removed the feature j, classified N samples on $d-1$ dimension features space, if the classification result is the same, we call the contribution on the classification of the feature j is zero. Put off those features which contributions value on division are zero will not impact the classification.

②The more the value of σ_j^2, the greater the discrete degree of various class center to $\overline{m}$ and the more open the C class each center, which is to say the more contribution of the feature j to divided C classifications.

$$w_j = \sigma_j^2 \Big/ \sum_{k=1}^{d} \sigma_k^2 \tag{4}$$

Obviously, $0 \le w_j \le 1$, $\sum_{k=1}^{d} w_j = 1$, we call w_j as the classification weight of feature j about given classification.

The description of classification weight w_j is: the proportion of feature j 's contribution value to differentiate various class center in all the d kinds of classification.

When $\sigma_j^2 = 0$, if we use "the distance" between sample point x and class center m_k indicate the similarity degree between x and Γ_k, this distance is not the simple distance, but contains some sort of classification information. To be specific, when $\sigma_j^2 = 0$, the j component should not appear in the formula about calculating the corresponding the distance. Therefore, we should use the distance between sample points

$$y_i = (y_{i1}, y_{id}, \cdots, y_{id}) \tag{5}$$

to class center $m_k = (m_{k1}, m_{k2}, \cdots, m_{kd})$ as the similarity measurement between y_i and Γ_k. This distance is a kind of weight distance:

$$\|y_i - m_k\|^2 = \sum_{l=1}^{d} w_l \cdot (y_{il} - m_{kl})^2 \tag{6}$$

2.3. Basic Uncertainty Membership Degree

The greater the weighted Euclidean distance between sample y_i to class center Γ_k, the less the membership degree of y_i belongs to Γ_k category. If we use $\mu_{\Gamma_k}(y_i)$ indicate the membership degree of y_i belong to Γ_k category, although we could not actually know the true value of $\mu_{\Gamma_k}(y_i)$, we know that $\mu_{\Gamma_k}(y_i)$ decreases as the $\|y_i - m_k\|$ increase. So we only need know the memberships relative value. So make

$$\mu_{\Gamma_k}(y_i) = \frac{1}{\|y_i - m_k\| + \varepsilon} \Big/ \sum_{l=1}^{C} \frac{1}{\|y_i - m_l\| + \varepsilon} \tag{7}$$

In which, ε is a control constant which was used to adjust the too big impact to membership degree when the $\|y_i - m_k\|$ is too small. We call the certainty membership of (7) formula as the basic uncertainty membership degree, which referred as basic membership degree.

It's easy to vivificates that the membership degree defined by formula (7) is the uncertainty membership degree. Its rationality is obvious, that formula (7) is as the uncertainty classification membership function corresponding to the certainty classification. When the basic membership degree of every sample belong to various classification is obtained, it will create an uncertain classification about the U universe of discourse.

However, the uncertain classification determined by uncertain classification cannot guarantee the consistency with original certain classification on the classification. On the other word, the certain classification that restored from uncertain classification according to the maximum membership principle, was not equivalent to the original certain classification. This is determined by the specific structure of the classification data. As shown in Figure 1:

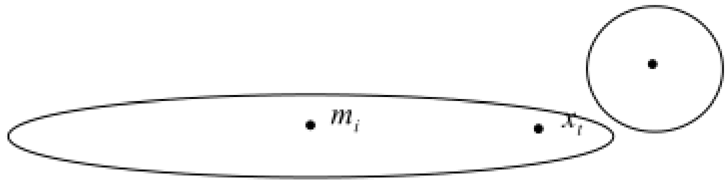

Figure 1. x_t **Shoule Be in Class** Γ_i**, But** x_t **is Closer with** m_j

Therefore, we can adjust the basic membership degree when it's necessary to supervise the clustering. We can use adjusted membership degree replace the basic membership degree of wrong classified sample, to ensure the consistent of the obtained uncertain classification with the given certain classification. Yet, for unsupervised clustering, as there is no known class sample can be used. The class center given by Formula (1) is only the initial class center. In this way, we always use the basic membership degree determined by Formula (7) to create the uncertain classification on U, which can approximate replace the original classification.

2.4. Uncertain Means Clustering Algorithm

Classify N d -dimension feature space sample points into C categories; try to get these C type centers

(1) Data preprocessing

①Data standardization

Let

Then, each component of $Y_{\mathbf{j}} = (y_{\mathbf{1j}}, y_{\mathbf{2j}}, \ldots, y_{\mathbf{mj}})^{\mathbf{T}}$ has a value between 0 and 1.

② Initializing category

$$y_{ij} = (x_{ij} - \min_{1 \le i \le N}\{x_{ij}\}) \big/ (\max_{1 \le i \le N}\{x_{ij}\} - \min_{1 \le i \le N}\{x_{ij}\}) \tag{8}$$

$$Sum(i) = \sum_{j=1}^{d} y_{ij} \tag{9}$$

$$MA = \max_{i} Sum(i) \tag{10}$$

$$MI = \min_{i} Sum(i) \tag{11}$$

$$J = (C-1)(Sum(i) - MI)/(MA - MI) \tag{12}$$

Suppose k is the nearest positive number with $1+J$, put sample y_i into the k th category, so that the N samples will be divided into C different categories. That is, given an initial classification, and calculate the various cluster centers $m_k^{(0)}(k=1\sim C)$ .According to the initial classification, it can calculate various types membership.

(2) Calculate by the following steps:

Step1. From C initial cluster centers $m_1^{(0)},m_2^{(0)},\cdots,m_C^{(0)}$; Calculate

$$\overline{m}^{(0)}=(m_1^{(0)},m_2^{(0)},\cdots,m_C^{(0)}) \tag{13}$$

Step2. Calculate

$$\sigma_j^{2\,(0)}=\frac{\alpha_i}{C}\sum_{k=1}^{C}(m_{kj}^{(0)}-\overline{m}_j^{(0)})^2\,,\ (1\le j\le d\,) \tag{14}$$

Step3. Calculate

$$w_j^{(0)}=\sigma_j^{2\,(0)}\Big/\sum_{l=1}^{d}\sigma_l^{2\,(0)} \tag{15}$$

Step4. Calculate

$$\left\|y_i-m_k^{(0)}\right\|^2=\sum_{l=1}^{d}w_l^{(0)}(y_{il}-m_{kl}^{(0)})^2 \tag{16}$$

Step5. Calculate

$$\mu_{\Gamma_k}^{(0)}(y_i)=\frac{1}{\left\|y_i-m_k^{(0)}\right\|^2+\varepsilon}\Bigg/\sum_{l=1}^{C}\frac{1}{\left\|y_i-m_l^{(0)}\right\|^2+\varepsilon} \tag{17}$$

In which, $\varepsilon=0.01$

$\mu_{\Gamma_k}^{(0)}(y_i)$ got from formula (17) is the initial classification's corresponding basic unascertained membership. Known the N samples basic unascertained membership, it can calculate these N samples should have certainty cluster centers. As follows:

In d dimension feature space, take sample points y_i about Γ_k type's basic membership $\mu_{\Gamma_k}^{(0)}(y_i)$ as point mass endow to point y_i . In this way, N particles constitute the centroids of particle group $\{(y_1,\mu_{\Gamma_k}(y_1)),(y_2,\mu_{\Gamma_k}(y_2)),\cdots,(y_N,\mu_{\Gamma_k}(y_N))\}$ can be determined by physical method. That is particle's corresponding membership vector is:

$$m_k^{(1)}=\sum_{i=1}^{N}\mu_{\Gamma_i}^{(0)}(y_i)\cdot y_i\Bigg/\sum_{i=1}^{N}\mu_{\Gamma_i}^{(0)}(y_i) \tag{18}$$

Thus, C cluster centers: $m_1^{(1)},m_2^{(1)},\cdots,m_C^{(1)}$ are obtained after the first iteration

Step6. Replace $m_i^{(0)}$ with $m_i^{(1)}(i=1,2,\cdots C)$, return to Step 1.

Step7. After 7 iterations, when $\max\left\|m_i^{(t)}-m_i^{(t-1)}\right\|<\delta$, stop the iteration. Output C cluster centers $m_1^{(t)},m_2^{(t)},\cdots,m_C^{(t)}$, in which $m_k^{(t)}$ is the cluster center of type Γ_k

After the above steps, N d-dimension feature space sample points, completed the C-means clustering, given the center vectors of C class-centers.

3. Constructing of Evaluation System for the International Competitiveness of Coal Industry

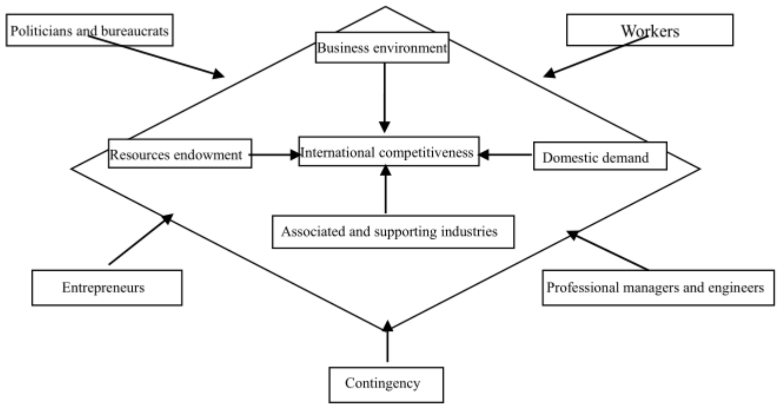

Figure 2. The Model of Nine Factors to Determine Competitiveness of Enterprises

Table 1. International Competition Appraisal Index System of Coal Industry

Total index	First-grade indexes	Second-grade indexes
		state policy
	Business environment	law environment
		prices
		industrial production indicators
	Associated and supporting industries	Energy utilization efficiency
		total energy consumption
	Domestic demand	coal consumption
		import
International competition appraisal index system of coal industry		export
	entrepreneurs	education investment
		storage volume
	Rescores endowment	production volume
		heat contained in coal
		Recoverable coal
	profession manager and engineers	numbers of scientists and engineers
	Politicians and bureaucrats	Management system
	workers	death rate

Diamond model theory provides a relatively complete theoretical framework for analyzing industrial competitiveness, which is universally recognized by academics as the basic theory [3]. In order to ascertain the competitiveness of coal-mining industry in terms of its nature, main factors influencing it and appraising model, concrete research in needed to conduct. Dong-sung cho argues that porters “diamond model” is suitable to explain the industrial competitiveness of developed countries but not desirable for less developed countries and developing countries, modifications are needed to improve “diamond model’, based on which a nine-factor model is established as shown in Table 1.This thesis applies Dong-sung cho analytical model to build comprehensive appraising system composed of eight one first degree indicators and second degree group indicators. With the above-mentioned model, comprehensive appraisals are made to analyze the international competitiveness of coal-mining industry, which is shown in the Figure 2[4].

Learning from analyzing model established by Dong-Sung Cho, the paper formulates an index system including 8 first class indexed and 17 second class indexes to evaluate coal industry international competitiveness in China. The index system is shown in Table 1.

4. International Coal Industry Competitiveness Appraisal Based on Unascertained Clustering

Table 2. Original Data Table

Countries	U.S.A	Canada	Brazil	German	Poland	Russia	Ukrine	Australia	China	India	South Afica
state policy	80	80	75	80	70	70	65	85	70	60	65
law environment	85	80	70	75	60	50	45	75	45	40	55
price(raw coal/ton)	87.38		89.76	121.5	121.4	92	92.1	90.4	95	88.3	81
industrial production index	134.8	120	110.8	121.4	104.2	108.3	103.1	118.8	126	108.1	105
energy utilization efficiency(Geary-Khamis dollar / kg standard oil)	4.30	3.53	6.85	6.11	4.30	1.76	1.68	4.81	4.57	4.91	3.94
energy consumption(million tons of oil equivalent)	2269.3	330.3	266.9	306.4	102.8	685.6	126.4	123.3	2613.2	559.1	126.3
coal consumption(million tons of oil equivalent)	501.9	21.8	13.9	77.6	59.8	90.9	42.4	49.8	1839.4	295.6	92.9
import(10thousand tons)		629.3	6000	3400	310		1040	0	18800	8410	565
export(10 thousand tons)	10700	2700	0	105	1750	10500	105	10800	1466	0	6240
public education expenditure proportion in GDP (%)	6.2	3.9	4.51	4	3.1	3.9	3.4	4.5	2.41	2.9	5.6
storage(1 million tons)	237295	6582	4559	40899	5709	157010	33873	76400	114500	60500	30156
production(100 million tons)	10.04	0.67	0.02	1.89	1.39	3.34	0.62	4.24	34.71	5.86	2.54
heat content of coal	20830	20915	14287	10321	15873	18514	19406	8863	19728	17715	21302
recoverable coal	267312	7251	11148	7428	15432	173074	37647	22	126215	101903	53738
number of scientists and engineers per one million people	4103	3309	168	2873	1460	3394	2121	3320	459	158	
management system	85	80	75	70	65	65	60	75	55	50	65
mortality rate per million tons	0.03			1.19	0.26	0.3	2.84	0.15	3	0.25	0.08

The statistical data of coal production from past 20 years has shown that U.S. India, Russia, South Africa and the rest of 8 countries take up 90% of the total coal production. The paper utilizes the original data from 11 countries economic and social statistics yearbook and export rating result as shown in Table 2.

Data source: international statistics yearbook 2010,bp energy data, international statistics yearbook 2011 and 2012,world economic statistics yearbook, the World Energy Projection System2011, World Energy Projection System Model Documentation, World Resources 2010, world series energy statistics report.

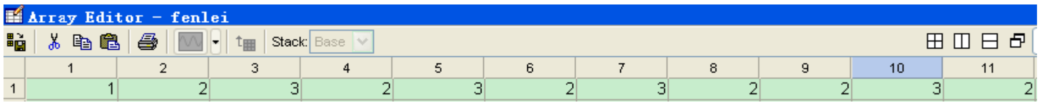

Figure 3. Categorized Cluster Results

The 11 countries are classified into three degrees, which are strong, moderate and weak. Initially the backward indicators (import, death ratios) are turned to forward indicators, and then primary classification of the standard data is obtained, and the membership vector (u_{i1}, u_{i2}, u_{i3}), $i = 1, 2, \cdots, n$ is made. Under the environment of Matlab, according to the steps of unascertained clustering, taking$\varepsilon \leq 0.001$, after 153 iterations, the discrepancy is 0.0099. The clustering result is shown in table 2[5], the first cluster contains Russia, India, Poland and Brazil, the second contains China, Canada, German, Ukraine, Australia and South Africa, the last cluster only contains U.S.A.

The result shows that U.S. is most competitive in its coal-mining industry, China degree of membership to moderate cluster is 0.45942 while the same to strong cluster is 0.34989, thus China can be regarded relatively competitive in the moderate cluster, Russia, India, Poland and Brazil are all in the weaker competitive position, which is commensurate to the present situation of international coal-mining competitiveness.

Table 3. Membership Degree of Sample to Each Class

	U.S.A	Canada	Brazil	German	Poland	Russia	Ukrine	Australia	China	India	South Afica
strong	0.6263	0.1988	0.1176	0.0974	0.1976	0.2769	0.1042	0.2250	0.3499	0.1613	0.2331
moderate	0.2364	0.5021	0.4328	0.5518	0.4014	0.2607	0.5499	0.4374	0.4594	0.3180	0.2333
weak	0.1372	0.3002	0.4496	0.3508	0.4010	0.4624	0.3459	0.3376	0.2007	0.5207	0.5336

5. Conclusions

Adopting unascertained clustering method to overcome C -means clustering algorithm's shortage, each classification result in the paper shows the membership to each class, rather than simply regard a sample with membership 1 belongs to a class. Meanwhile, the method of unascertained clustering recognizes that samples in d-dimension can be divided into C class in that different kinds of samples have different observation value. Their features have different degrees of contributions for distinguishing sample clusters, which is recognized by unascertained clustering, but not by fuzzy C-means clustering. Obviously, the method is more science and actual than C-means clustering and fuzzy C-means clustering [6-8]. It's a new exploration to classification field.

The unascertained cluster theory not only provides reliable classification result but also find the center of the clusters and the membership degrees of all the samples. The method more accurately expressed each sample's membership to each class, in line with the actual situation, which provides the coal mine safety investment dynamic evaluation a feasible method. This is recognized and realized by ascertained clustering method by means of quantitative description, which cannot be done with C-means cluster algorithm.

Acknowledgement

This study is financially supported by philosophy and social science foundation of Hebei (HB12GL065), social science association project of Handan (2012120), construction science and technology research project of Hebei (2012120) and science &technology research and development project of Handan (1323120082-6). Projects supported by National Social Science Foundation (14bjy022)

Reference

[1] L. Shanshan and Z. Min, “Assessment of China’s coal competitiveness based on Porter’s Diamond , Model. Resources& Industries, vol. 2, **(2007)**, pp. 37-40.
[2] H. Jinxiang, “Current international competitiveness of China’s cola industry”, Coal Engineering, vol. 4, **(2003)**, pp. 9-13.
[3] M. E. Porter, “The Competitive Advantage of Nations”, Beijing. Citic Press Group, **(2002)**.
[4] C.Weiguo and Z. Shubin, “Survey of the international competitiveness of industry in domestic”, Teaching and Research, vol. 4, **(2002)**, pp. 57-67.
[5] C. Xiang, Z.Xin and F. Zhongbo, “Evaluation on Science and technology investment power of coal enterprise base on unascertained cluster”, Coal Engineering, vol. 6, **(2012)**, pp. 95-98.
[6] J.-S. Wang and C. S. G. Lee, “Self-Adaptive Neural-Fuzzy Inference System for Classification Applications”, IEEE Trans on Fuzzy systems, Shanghai, China, vo. 10, no. 6, **(2012)** December.
[7] F. Canastra and A. Verrian, “A Novel Kernel Method for Clustering”, IEEE TRANSCATIONS ON PATTERN AND MACHINE INTELLIGENCE, Nanjing, China, vol. 27, no. 5, **(2005)** May.
[8] X. GAO, “Fuzzy Clustering Analysis & Application”, Xidian University Press, **(2004)**, pp. 53-60.

Authors

Xiang Chen, Chen Xiang, female, associate professor, postdoctoral, serving in Architecture Institution, Hebei University of Engineering. Mainly engaged in management science and engineering, architectural design and other aspects research and teaching. Presided over the national social science fund 1, Hebei Province issues 4, Province ministry and agency found 1, Handan city issues 5. Has published more than 20 academic papers at home and abroad, including EI and ISTP search more than 10.

