Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(6):2535-2540

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Research on the evaluation method of regional innovation capability

Yongli Zhang and Shujuan Yuan

Qinggong College, Hebei United University, Tangshan, Hebei, China

---

ABSTRACT

China will develop into an innovative country in 2020. It has become an important topic that study on evaluation method of innovation ability. But there are many human factors in most evaluation methods. This would diminish the validity of a systematic review. The unsupervised clustering algorithm, combining fuzzy C means clustering and support vector machine are used in this paper. The impact of human factors will be overcome. Through the 2013 yearbook data analysis of the experimental results, this method is achieved very good results in evaluation of regional innovation capacity.

Key words: Innovation ability, Fuzzy C means clustering analysis, Support vector machine

---

INTRODUCTION

China has plans to be an innovative country in 2020. In order to scientific and technological innovation capability development process inspection, monitoring and evaluation of the city, it is an important research direction of evaluation index system and evaluation method of the effective. We will enhance the ability of science and technology [1-5]; it cannot do without the city country (region) of the innovation ability of contribution. The comprehensive strength of a city is the gross domestic product (GDP) and the ability of innovation of science and technology together [6-9]. At present, Statistical accounting method of GDP is specific and clear. It is complex to determine the technology innovation ability. There are many factors to influence and reaction innovation ability [10-12]. There is a non-linear relation ship between each factor and uncertainty and ambiguity. These are barrier to evaluation on the innovation capability.

About evaluation of the innovation ability of science and technology, To determine the weight of indexes are evaluated by using AHP and entropy method in reference [1], Factor analysis and fuzzy C means clustering method (FCM) evaluate the ability of innovation of science and technology in reference [2]. In addition there are references using neural network method etc [1-4].The human factor is added to the evaluation on these methods. This is what they have in common. For example, evaluation matrix is determined through expert scoring method in fuzzy comprehensive evaluation but this matrix directly affects the evaluation results. In order to overcome the impact of human factors, the city was clustered using FCM clustering algorithm then the support vector classification was trained by the clustering results. Method for analysis of technological innovation ability is obtained.

FCM CLUSTERING ALGORITHM

Clustering is a method of sample classification. It calculated the distance between samples to determine whether the designated as a class. The calculated distance is compared with a threshold value. There are different classification methods and the classification results according to different distance and different thresholds. The classification process is not affected by the human element. Do not need to known classes of samples as learning tutor.

FCM clustering algorithm is proposed by Dunn in 1974, the algorithm is extended and applied by Bezdek. The

method uses iterative method. It gets the membership parameters and the clustering center by optimizing the objective function. [5-7].

Fuzzy C- partition

A data set $X=\{x_1,x_2,\cdots,x_n\}$ be divided into c classes

$c<n$ and $c,n\in Z$, if $D=(d_{ik})_{c\times n}$ is fuzzy matrix, and $d_{ik}\in[0,1]$; $\sum_{i=1}^{c}d_{ik}=1$; $\sum_{k=1}^{n}d_{ik}\in(0,n)$ D is called fuzzy C- partition.

Criterion Function

Definition: If sample set $X=\{x_1,x_2,\cdots,x_n\}$, $V=\{v_1,v_2,\cdots,v_c\}$, $D=(d_{ij})_{c\times n}$ is fuzzy matrix, Let:

$$J(D,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}(d_{ik})^{p}\left\|v_i-x_k\right\|^2 \tag{1}$$

$J(D,V)$ is called according to the fuzzy partition . $D=(d_{ij})_{c\times n}$ is the criterion function clustering. This definition is given by Bezdek.

Where $p\in R$ and $p>1$, $\left\|v_i-x_k\right\|^2=d(v_i,x_k)$ is vector measure. We usually use the Euclidean distance.

Procedure

Step one The sample set $X=\{x_1,x_2,\cdots,x_n\}$, $c\in R$, $p>1$, $\varepsilon>0$ (The maximum error).

Step two The fuzzy partition matrix $D^{(0)}=(d_{ik}{}^{(0)})_{c\times n}$, d_{ik} is any real number

Step three Calculating $v_i^{(l)}$

$$v_i^{(l)}=\frac{\sum_{k=1}^{n}\left(d_{ik}^{(l)}\right)^p x_k}{\sum_{k=1}^{n}\left(d_{ik}^{(l)}\right)^p} \qquad l=1,2,\cdots;i=1,2,\cdots,c \tag{2}$$

Step four To $k=1,2,\cdots,n$ recalculate $D^{(l)}$ receive $D^{(l+1)}$.

Step four-one if $j(1\le j\le c)$, get $x_k=v_j^{(l)}$, Then:

$$d_{ik}^{(l+1)}=\begin{cases}1, & i=j\\ 0, & i\neq j\end{cases} \tag{3}$$

Step four-two if all of i, get $x_k\neq v_j^{(l)}$, Then:

$$d_{ik}^{l}=\frac{1}{\sum_{t=1}^{c}\left(\frac{\left\|v_i^{(l)}-x_k\right\|}{\left\|v_t^{(l)}-x_k\right\|}\right)^{\frac{2}{p-1}}} \tag{4}$$

Step five Calculating $J(D^{(l)},V^{(l)})$ and $J(D^{(l+1)},V^{(l+1)})$.

If $J(D^{(l)},V^{(l)}) \_ J(D^{(l+1)},V^{(l+1)}) < \varepsilon$ then over. Get $D^{(l+1)}$ and $V^{(l+1)}$ For the optimal clustering center, Otherwise $l = l+1$.Go back to step 3 to recount.

SUPPORT VECTOR MACHINE CLASSIFICATION ALGORITHMS

Support Vector Machine Origin

Vapnik first proposed the support vector machine (SVM) algorithm. It is a kind of intelligent learning algorithm, can be applied to the problem of small sample learning, can solve classification and curve fitting. By means of the principle of structural risk minimization, it can improve the generalization ability of the learning machine. At present, application of support vector machine is very extensive, study more, such as hand written digit recognition, face recognition, fault detection, network security monitoring, and nonlinear function approximation fields mature. SVM is a kind of supervised classification machine. It is applied to the evaluation of scientific and technological innovation ability; it will combine with the clustering analysis algorithm.

support vector classifier

Suppose $\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$ is a sample set, where $x \in R,\ y \in \{1,-1\}$ let:

$$\begin{aligned} w\cdot x_i + b &\geq 1 \qquad y_i = 1 \\ w\cdot x_i + b &\leq -1 \qquad y_i = -1 \end{aligned} \tag{5}$$

w is a classification hyper plane vector method, $\frac{|b|}{\|w\|}$ is the distance form origin to the hyper plane.

(5) can be expressed as the merger: $y_i(w\cdot x_i + b) \geq 1$.the hyper plane $H: w\cdot x_i + b = 0$.origin to the hyper plane $H_1: w\cdot x_i + b = 1$ distance is $\frac{|b-1|}{\|w\|}$, origin to the hyper plane $H_2: w\cdot x_i + b = -1$ distance is $\frac{|b+1|}{\|w\|}$.In the hyper plane H_1 or H_2 points are called support vector. Class interval is $\frac{2}{\|w\|}$, it is distance between two hyper plane H_1 and H_2. The between class distance is greater, the better the classification capabilities of machine. The whole learning process can be considered as the maximum $\frac{2}{\|w\|}$, which will minimize $\frac{\|w\|^2}{2}$.

SVM expression will be:

$$\begin{aligned} &\min \Phi(w,b) = \frac{1}{2}\|w\|^2 = \frac{1}{2}(w\cdot w) \\ &s.t. \quad y_i\left[(w\cdot x_i) - b\right] \geq 1,\ i = 1,2,\cdots,n \end{aligned} \tag{6}$$

Where n is the number of training samples.

The problem described above is called the original problem. The problem can be solved used two times planning. Usually can use the Lagrange method, the original problem is converted into a dual problem [9, 10, 11]:

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j (x_i x_j) - \sum_{i=1}^{l}\alpha_i \tag{7}$$

$$s.t. \quad \sum_{i=1}^{l}\alpha_i y_i = 0 \quad \alpha_i \geq 0 \quad i = 1,2,\cdots,l$$

To solve the above expression, get $\alpha^* = \left(\alpha_1^*, \cdots, \alpha_l^*\right)^T$, then $w^* = \sum_{i=1}^{l} y_i \alpha^* x_i$, let α^* >0 component. For example α_j^* corresponding sample $\left(x_j, y_j\right)$ (This is a sample of support vector), Then the expression $b^* = y_j - \sum_{i=1}^{l} y_i \alpha_i^* \left(x_i \cdot x_j\right)$, count w and b, Ultimately determine the optimal hyper plane $\left(w^* \cdot x\right) + b^* = 0$, Get the decision function $f(x) = \text{sgn}(w^* \cdot x + b^*)$ or $f(x) = \text{sgn}\left(\sum_{i=1}^{l} \alpha_i^* y_i (x \cdot x_i) + b^*\right)$.

PROCESS OF TECHNOLOGY INNOVATION CAPABILITY EVALUATION ALGORITHM

Step one :Selection of experimental data, The data are normalized, standardized. Avoid data difference is too large, affecting the speed of the algorithm.

Step two:The samples for FCM analysis get the clustering results into two categories.

Step three:Please randomly selected a most samples is the training sample of SVM, The rest of the small sample is the testing sample. When the training SVM classifier is completed, we will test using to testing samples. If the recognition rate is not high, adjusting the parameters of the SVM or function. Then classification recognition effect is good.

Step four :The weight vector will obtain from the SVM classifier trained .Then we count data using to the weight vector. The city's innovation capability ranking will be get.

EXPERIMENTAL RESEARCHES AND ANALYSIS

This paper comes from the rankings in 2013 Chinese city innovation ability of randomly selected 20 as the experimental data. Index system is selected from<Monitoring index system of regional innovation capability>.It is promulgated by the Ministry of science and technology of the People's Republic of China. The index system includes the innovation environment, innovation resources, innovation, and innovation output and innovation effect of 5 first level indexes and 53 second level indexes monitoring index system of regional innovation capability. In order to facilitate the experimental and data collection, 30 indexes second level were selected from 53 second level indicators.

These data come from the People's Republic of China National Bureau of Statistics web site Chinese statistical yearbook 2013[12], Chinese statistical yearbook on science and technology of 2013[13] and the national database. There are 30 cities, each city has 32 second level index. The original values of 32 second level indexes were normalized.

Following formula:

$$\boldsymbol{y_{ij}} = \frac{\boldsymbol{x_{ij}} - \min \boldsymbol{x_{ij}}}{\max \boldsymbol{x_{ij}} - \min \boldsymbol{x_{ij}}} \tag{8}$$

Where $\boldsymbol{i} = 1,\cdots,30; \quad \boldsymbol{j} = 1,\cdots,32$.

The use of Matlab2012b software for FCM analysis, remember to type A and type B. The results shown in the table 1 below.

Table 1: The results of clustering

A	Wuxi Jiaxing Kunming Zibo Zhoushan Yangzhou Haikou Baotou Zhengzhou WuHu Jinhua
B	Yingkou Wenzhou Lanzhou Yinchuan Qinghongdao Baoding Xuzhou Wuhai Langfang Handan Jining Tangshan Yichun Anqing Liaoning Chaozhou Nanping Benxi Meizhou

Table 2: Innovation index

city	original index	New index	D-value
Wuxi	92.8944	93.0086	0.1142
Jiaxing	83.1787	82.9544	-0.2243
Kunming	78.6563	78.6808	0.0245
Zibo	76.5074	76.4618	-0.0456
Zhoushan	75.2805	75.3163	0.0358
Yangzhou	75.153	75.1403	-0.0127
Haikou	73.673	73.7451	0.0721
Zhengzhou	73.765	73.7409	-0.0241
Baotou	73.0694	73.0036	-0.0658
Wuhu	72.8355	72.8478	0.0123
Jinhua	71.4094	71.4218	0.0124
Yingkou	70.6731	70.6428	-0.0303
Wenzhou	69.9011	69.9369	0.0358
Lanzhou	68.2218	68.2094	-0.0124
Yinchuan	67.9246	68.0493	0.1247
Qinhuangdao	67.5523	67.6012	0.0489
Baoding	66.5678	66.6799	0.1121
Wuhai	66.7045	66.6318	-0.0727
Langfang	65.5091	65.462	-0.0471
Handan	64.9126	64.9371	0.0245
Jining	64.1939	64.1494	-0.0445
Liaoyang	63.7432	63.8056	0.0624
Yichun	63.6837	63.7051	0.0214
Xuzhou	63.2698	63.2586	-0.0112
Chaozhou	62.8768	62.9292	0.0524
Tangshan	62.461	62.4751	0.0141
Nanping	62.4758	62.3884	-0.0874
Anqing	62.3597	62.371	0.0113
Benxi	61.9939	61.9782	-0.0157
Meizhou	61.5912	61.6033	0.0121

24 cities were randomly selected as training data, the remaining 6 cities as the testing sample. The data is trained by SVM. Matlab2012b software and LS-SVM LAB1.50 tool are used. Kernel function is RBF.

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\| \boldsymbol{x} - \boldsymbol{x}' \|^2 / \sigma^2) \tag{9}$$

The better classification effect of SVM classification is obtained by testing samples. Then get the weight matrix:

$\boldsymbol{w}$ = {−1.6127, 3.1027,−1.1201,0.2478,
1.359,8.3606,7.7146,−1.857,11.2457,
0.1147,−2.4576,−1.2347,0.1254,
−0.1485,14.2576,7.4892,15.4678,
4.5799,28.9235,0.1475,−1.2547,
0.1245,5.7982,−2.3568,1.2457,0.1246,
0.3658,12.5477,5.6988,4.3658,4.5681,
2.3568}

In the decision function weight $|\boldsymbol{w}_i|$ is large, the corresponding index more important. So choose the relatively large $|\boldsymbol{w}_i|$ as a new index system of simplified. Select the top 20 in $|\boldsymbol{w}_i|$ and count the index of innovation. The comprehensive evaluation results were compared on 2013 Chinese city innovation capability list data.

Through comparative research, combined use of FCM and SVM classification algorithm, their differences seem relatively minor for the results obtained and the state announced the results. The result is satisfactory.

CONCLUSION

The FCM analysis and SVM classification technique are applied to comprehensive evaluation of regional innovation ability. By data experiments show that, this method is feasible. If the proper selection of kernel function and its parameters, it can achieve good effect. This method has good generalization ability. It has a good application prospect in innovation capability evaluation.

Acknowledgment
This work is supported by Tangshan City science and technology research and development program (No.2013CX-1), and Research on science and technology to guide the project of higher education of Hebei Province(No.Z2011328).

REFERENCES

[1] Yan Wang. *Statistics and Management*, **2013**, (08).
[2] Xiaoyun Du. *Shandong University of Finance and Economics*, **2013**.
[3] Xiangwei Kong. *Beijing Jiaotong University*, **2009**.
[4] Olivier Chapelle, Vladimir Vapnik, Olivier Bousquet, Sayan Mukherjee. *Machine Learning*, **2002**, (1-3).
[5] Yong Zhang, Zhong-xian Chi, Xiao-dan Liu, Xian g-hai Wang. *Applied Intelligence*, **2007**, (1).
[6] Qi Wu, Rob Law. *Expert Systems With Applications*, **2011**, (10).
[7] Tzong-Huei Lin. *Neurocomputing*, **2009**, (16).
[8] Yaolong Kang, Lilu Feng. *Journal of Shanxi Datong University(Natural Science Edition)*, **2013**, (04).
[9] Zhang B.; Zhang S.; Lu G.. *Journal of Chemical and Pharmaceutical Research*, **2013**, 5(9), 256-262.
[10] Zhang B.; *International Journal of Applied Mathematics and Statistics*, **2013**, 44(14), 422-430.
[11] Zhang B.; Yue H.. *International Journal of Applied Mathematics and Statistics*, **2013**, 40(10), 469-476.
[12] Zhang B.; Feng Y.. *International Journal of Applied Mathematics and Statistics*, **2013**, 40(10), 136-143.