

The Evaluation and Application Research about Regional Innovation Capability Based on Rough Set and BP Neural Network

Wenjuan Wang
College of Information
Management, Chengdu
University of Technology,
Chengdu, China
wangwj@cdut.edu.cn

Bin Xie
College of Information
Management, Chengdu
University of Technology,
Chengdu, China
xieaibin1987@163.com

Yusheng Li
College of Information
Management, Chengdu
University of Technology,
Chengdu, China
liys@cdut.edu.cn (Corresponding author)

Kejia Pan
School of Mathematical
Sciences, Fudan University,
Shanghai, China
kjpan@yahoo.cn

Abstract—Because of the shortage of the traditional methods for evaluating regional innovation capability, it presents a new method based on rough set and BP neural network in this paper. Firstly, index system for evaluating about regional innovation capability is set up, and then rough set is used to simplify the attributes of the index system, it will ease the burden of training and learning of BP network and reduce the amount of computing. The simplified attributes are input into the network which has been designed. Through the training and testing of lots of studied samples, the error of the model is limited in a preconcert range. Thus, an integrative network model has been established. At the end, it proves the accuracy and feasibility of the method with an example.

KeyWords—*Rough Set; BP Neural Network; Regional Innovation Capability*

I. INTRODUCTION

In the opinion of Cooke (1998), innovation is the commercialization of knowledge, it is the process of utilizing knowledge, and innovation can be defined as utilizing new knowledge successfully. According to this perspective, Regional innovation capability is interpreted as the capability of regional to use new knowledge [1]. As the empirical study of Annalee. Saxenian at 1994 on Silicon Valley and Route 128 indicate that, the success of Silicon Valley contributes to the timeless innovation power, and lack of this innovation system is the reason of Route 128's debility [2]. It can come to the conclusion that regional innovation capability lead to regional's further competitive capability; it's also the most important factor to explain the difference of economic prosperity between the regions.

At present, the research on evaluation methods for regional innovation capability is more and more abundance in the academe; the usual methods include vulnerable targets in multiples of law, AHP method, fuzzy comprehensive analysis, principal component analysis, osculating value method, gray clustering method, efficacy coefficient method and so on [3]. There exist many shortages of the traditional evaluation methods such as subjectivity, mistiness and lack theory basic on the weight of indexes.

This paper proposes a method bases on rough set and BP neural network. It can avoid the subjective factors of artificial empowering. This method uses the rough theory to reduce the knowledge and reduce the dimension of input vector under the premise of not losing information, and then take the data of simplified attributes as the input of BP neural network, at last, take the trained neural network into practice.

II. KNOWLEDGE REDUCTION ALGORITHM BASED ON ROUGH SET

Rough set theory (RST), proposed by Z. Pawlak at the 20th century, is a mathematical method to deal with uncertainty, incomplete or inconsistent data [4,5]. Its main idea is to use known knowledge to depict the uncertain or uncertain knowledge approximately. Under the premise of not losing information, reducing knowledge and attribute values, analysis the independencies between attributes to obtain the minimum decision-making rules [6]. The key step is to simplify the attributes.

A. Undistinguished Relationship

Undistinguished relationship, named equivalent relationship, is a core concept in the rough set theory. Only if all the attributes' value of two samples' is equal, they belong to the same equivalence class [7].

The usual method to resolve the undistinguished relationship $Ind(X)$ is to compare each two samples among the whole samples, those which are equal in each attribute belong to the same equivalence class, the solve method of undistinguished relationship refer to algorithm 1 of literature [8].

B. The Importance of Attributes

Definition [9], suppose $X \subseteq A$ is a attribute subset, $x \in X$ is a attribute, tick the importance of X is $Sig_X(x)$,

$$\text{and } Sig_X(x) = 1 - \frac{|Ind(X \cup \{x\})|}{|Ind(X)|}$$

$$\text{If } U/Ind(X) = U/X = \{X_1, X_2, \dots, X_n\}, |Ind(X)| = \sum_{i=1}^n |X_i|^2$$

$Sig_X(x)$ indicate the importance of X .

C. Core Attributes and Reduce Attributes

Core attributes and reduce attributes are two basic concepts in the knowledge reduction. Core is the foundation of all reduction calculates and the collection of attributes which cannot be eliminated [10].

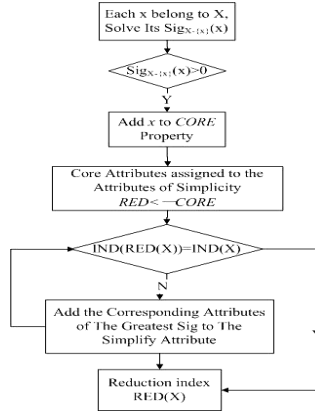


Figure 1 Knowledge reduction algorithm

Definition, suppose $X \subseteq U$, if X is independent and $Ind(X) = Ind(U)$, X is a reduction of U . The all necessary relationship of U compose the core of U , tick it as $Core(U)$. It can be proved that $Core(U) = \cap Red(U)$, $Red(U)$ present all the reduction of U .

It can be easily realize the knowledge reduction algorithm based on importance with proper programming tools such as Matlab. Figure 1 draws the flow chart of the algorithm.

III. BP NEURAL NETWORK

Artificial neural network(ANN) is a complex non-linear conversion network system, composed by a large number of simple basic components-neurons which are connect with each other, it deal with information with the method in parallel by simulating the human brain [11].

A. The Basic Theory of BP Neural Network

BP Network (BP) is a one-way transmission of multi-layer neural network. BP network is the core part of the forward network, embodies the essence of the neural network and the best content [11]. BP network is a network with three or more than three layers; it contains input layer, hidden layer and output layer. The neurons connected with each other between different layers but not in the same layer.

B. Learning Rule of BP Neural Network [11]

The thresholds of neurons are delivered layer by layer from input layer to output layer in the way of middle layer, and the respond will be got in the output layer. And then, in order to reduce the error of target output and actual output, feedback from output layer to output layer so as to amend the value of the right connections. This method, named "error

back propagation algorithm" is BP algorithm. With the amendment of reverse error, the accuracy of response for input mode is increasing.

IV. CONSTRUCT THE INDEXES OF REGIONAL INNOVATION CAPABILITY

Evaluation index is a group of attributes based on the evaluation target. In the recent year, domestic and foreign scholars and relevant agencies have done many researches on the system of regional innovation capability [12]. EU evaluates the innovation capability of its member states on four factors, human resources, the creation of new knowledge, the application and proliferation of knowledge and market. The evaluation system of regions is similar to the countries'. As a result of different culture background, the comparability of regional innovation capability is higher than the countries' [13]. Since reform and opening-up, china's regional innovation system has experienced the evolution process of creation and hatching, growth and continuous improvement; it shows the unique characteristics of the different stages [14]. This article bases on the index system of the National Bureau of Statistics research group on development and strategy of science and technology of china's research results, and builds the evaluation index system as following.

TABLE 1 Regional innovation capability evaluation index

Regional Innovation Capability Evaluation Index System	Innovation Environment A	Innovation Infrastructure A_1 Market Demand A_2 Labor Quality A_3 Financial Environment A_4 Levels of Entrepreneurship A_5
	Knowledge Creation B	Research and development Input B_1 Patent B_2 Research Papers B_3 The Efficiency of Input-Output B_4
	Knowledge Acquisition C	Technical Cooperation C_1 Technology Transfer C_2 Foreign Direct Investment C_3
	Enterprise Innovation Capability D	Research Investment D_1 Design Capability D_2 Manufacturing Capability D_3 Output Value of New Products D_4
	The Economic Benefits of Innovation E	Macroeconomic E_1 Industrial Structure E_2 The International Competitiveness E_3 People's Income Level E_4 Employment E_5

V. THE CASE STUDY BASED ON ROUGH SET AND BP NEURAL NETWORK

According to the evaluation index system of Table 1, collect relevant data of china's each province or municipalities. Select 29 samples of them randomly as the training samples of BP network, 2 samples for testing samples to verify the validity and accuracy of the network. After training the network successful, it can be applied to a new evaluation. Set attribute $A = C \cup D$, A is attribute set, C is condition attribute, D is decision attribute, and $C = \{X_1, X_2, \dots, X_{26}\}$, $D = \{X_{27}\}$, $U = \{Y_1, Y_2, \dots, Y_{28}\}$ is the 29 training samples.

A Knowledge Reduction

According to the algorithm in the part C of II, programming to realize the process of knowledge reduction.

(1) Normalized data, divide the attribute into two kinds, the bigger the better and the smaller the better, and then normalized them separately.

$$\text{The bigger the better } F_j = \frac{(x_j - x_{j\min})}{x_{j\max} - x_{j\min}}$$

$$\text{The smaller the better } F_j = \frac{(x_{j\max} - x_j)}{x_{j\max} - x_{j\min}}$$

F_j is the standardized value of target x_j , $x_{j\min}$ and $x_{j\max}$ are the max and min value of the j st attribute.

(2) Discrete data, according to the need of knowledge reduction, discrete the data.

(3) Use the knowledge reduction algorithm based on the importance to reduce the decretive data. In the computer, the importance is abstracted for function $\text{Sig}(X, Y)$, it represents the importance of attribute, undistinguished relationship is abstracted for function $\text{Ind}(X)$, it represents the classification under X attribute, judge the undistinguished relationship is abstracted for function $\text{IsEqual}(X, Y)$, it represents the classification under X and Y is equal.

According to the process of knowledge reduction, 9 attributes are got among the 21 attributes as the simplified attributes. In Table 2, simplified attributes and its corresponding normalized data are given.

B Empirical Study

According to the result of the knowledge reduction, take the simplified attributes as the input of BP neural network. Construct three-layer neural network, input layer (eight neurons), hidden layer (sixteen neurons) and output layer (one neuron). Utilize the tool box of neural network to train the samples.

The test result of the test samples indicates that the output of network has high accuracy and credibility. Figure 3 shows that it comes to the training requirements through iterating and training 6000 times. Figure 4 shows that the output error of test samples (sample 30 and sample 31) are

both under 8%, it indicate that the training of BP neural network is successful.

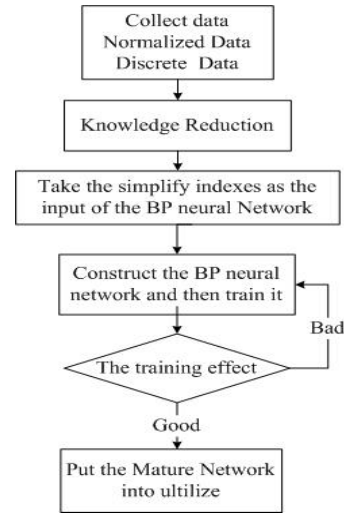


Figure 2 Evaluation process

TABLE 2 Simple attribute table and normalized values

Series Attribute	1	2	3		29	30	31
A_1	1.00	0.70	0.45		0.37	0.32	0.36
A_2	0.62	0.59	0.26		0.19	0.19	0.70
B_4	0.85	0.36	0.32		0.23	0.40	0.93
C_1	0.53	0.44	0.52		0.51	0.71	1.00
C_2	0.27	0.38	0.64	...	0.34	0.58	1.00
D_1	0.61	0.68	0.19		0.31	0.51	0.72
D_3	0.70	0.28	0.49		0.14	0.64	0.73
D_4	0.59	0.31	0.25		0.21	0.42	0.61
E_2	0.76	1.00	0.09		0.06	0.17	0.45
Evaluate Result	0.95	0.61	0.27		0.12	0.23	0.39

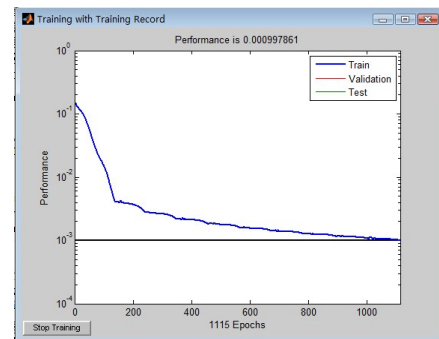


Figure 3 BP neural network performance

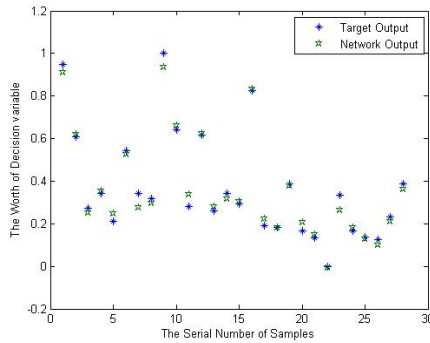


Figure 4 Target output and training output error

Take the samples of Table 3 into the BP neural network, the evaluation results will be got as follows 23.80, 29.72 and 48.59. It can be easily got the conclusion that region C has the best innovation capability, region A is the second and region B is the worst.

TABLE 3 The property value of pending evaluation samples

Attribute Samples	A_1	A_2	B_4	C_1	C_2
Region A	0.37	0.19	0.23	0.51	0.34
Region B	0.32	0.19	0.40	0.71	0.58
Region C	0.36	0.70	0.93	1.00	1.00
Attribute Samples	D_1	D_3	D_4	E_2	
Region A	0.31	0.14	0.21	0.06	
Region B	0.51	0.64	0.42	0.17	
Region C	0.72	0.73	0.61	0.45	

VI. CONCLUSION

Research on regional innovation system is still a new field of study; Scholars have not yet unified argument on its borders, the regional innovation system in the role of the system as well as through the research of regional innovation system to the formation of general policy recommendations. For china, regional innovation research more urgent than in other countries [13]. In this paper, the evaluation method based on rough set and BP neural network, its innovation lies in.

(1) Simplify the attributes with Rough Set theory, delete those unnecessary attributes, it can be reduce the burden of training and learning.

(2) BP neural network avoid the subjective elements and human factors of usual methods such as AHP and fuzzy comprehensive evaluation method. BP neural network will get the evaluation result with integrated tool kit, it don't need determine the weight by man-made, so the results are more effectively and objectively.

(3) BP network evaluation method is simple and the function of its tool kit is strong.

BP network evaluation method has its shortcomings, such as the quality of training samples will affect the stability of the training result, network layers and the neurons of hidden layer also will affect the learning capability and efficiency. Nevertheless, the fruitful results of BP neural network cannot be denied. Mat lab neural network toolbox of BP network model will be an effective way to evaluate the regional innovation capability.

ACKNOWLEDGMENT

This work is supported by Natural Science Foundation of Educational Office of Sichuan Province (No. 07ZA010).

REFERENCES

- [1] Cooke.P and Heirdenreich. Medited Regional Innovation System: the role of governance in a globalized world. UCL Press Limited,1998.
- [2] Annalee Saxenlan. Regional Advantage: Culture Competition in Silicon Valley and Route 128, Boston:Harvard University Press,1994,pp.18-40.
- [3] Yong Yang, Yanli Lin, Junwen Feng. The Measure Model for Innovation capability of R&D Enterprise. JiangSu ShangLun, 2005, vol.8, pp.153-154.
- [4] Z.Pawlak. Rough sets. International Journal of Computer and Information Science, 1982,11(5), pp. 341-356.
- [5] Qingkui Cao, Xiangyang Ren, Chen Liu, Libo Liu. The Analysis and study of the ability of technical innovation of enterprises based on rough set and unascertained measure appraisalment model.Systems Engineering-Theory&Practice. 2006, vol.4,pp.67-72.
- [6] Yong Zhu, Tao Wu. Ecnomic early-earning model based on rough sets and constructive learning neural networks. Journal of HeFei University of Technology,2007,7(30),pp.836-843.
- [7] Shaohui Liu, Qiujian Sheng, Bin Wu, Zhongzhi Shi, Fei Hu, Research on Efficient Algorithms for Rough Set Methods. Chinese Journal of Computers,2003,26(5),pp.524-529.
- [8] Wenxiu Zhang. Rough Set Theory and Method. Beijing: Science Press, 2001.
- [9] Kaiquan Shi, Yuquan Cui. S-Rough Set and Decision. Beijing: Science Press, 2006.
- [10] Shan Liu. A calculate core method on attribute reduction. Computer Engineering and Applications, 2004, vol. 17,pp.100-101.
- [11] Zhexue Ge, Ziqiang Shun. The theory of neural network and matlab application. Beijing: Publishing house of electronics industry, 2008.
- [12] Xiaobin Zeng, Yequn Bao. The Research on Evaluation Indexes of Regional Innovation Capability. Decision and Information, 2008, vol.4,pp.149-150.
- [13] Research Group on Development and Strategy of Science and Technology of China, Annual Report of Regional Innovation Capability of China 2004~2005, Beijing: China Intellectual Property Press, 2005.
- [14] Huanxiang Wang, Fei Sun, Evolutionary Analysis of Chinese Regional Innovation System during the 30 Years of Reform and Opening-up, Science & Technology Progress and Policy, 2008,25(9),pp.36-39.