

Evaluation of Logistics Reputation Based on Principal Component Analysis and BP Neural Network

Ruijun Zhang, Kui Fu
Wuhan University of Technology
Hubei, Wuhan 430081

Abstract—In order to reduce the risk of the buyers' purchase and avoid the problems of future returns and replacement after purchase, we need to sort out the credit of the seller of the electricity supplier to implement different purchase risk control strategies. Based on the principal component analysis and BP neural network, this paper selects the evaluation results of e-commerce buyers, combines with questionnaire surveys and expert analysis to establish a more referential index system so that the credit evaluation model of e-commerce sellers is established. The case study shows that the index system and the mathematical model can accurately judge the seller's credit rating, and have a strong guiding significance for buyers and enterprises to choose partners.

Keywords—BP neural network; E-commerce sellers; Principal component analysis; credit rating

I. INTRODUCTION

The rapid development of e-commerce has led to an exponential growth in the delivery business. At present, the country has more than 8000 enterprises holding express business license. At the same time, express logistics credit varies greatly: Express service quality is not optimistic overall, the loss of courier occurs frequently and the problem of cargo retention and damage still exists. Only in 2016, the state postal service received no fewer than 200 thousand complaints from Express Service [1]. The quality of express service not only damages the rights and interests of consumers, but also leads to the loss of customer resources, which seriously affects the economic benefits of enterprises. Therefore, judging and improving the service process and service quality of express delivery is an inevitable process for express industry to achieve good and rapid development, and also the key to the healthy development of express industry

However, there are few researches on the logistics process of e-commerce. We should establish a scientific and effective credit evaluation model of e-commerce logistics, which can identify and warn the potential risks of the logistics process as soon as possible. At present, Wang Zhuolun establishes the evaluation index system from the aspects of financial attributes and customer attributes, uses zero tempering to eliminate differences in indicators and uses KMV model for logistics enterprises to carry out credit risk warning to provide an important basis for guard against credit risk [2]; Guo Zhanglin

begins with credit risk in Supply Chain Finance, prevents and controls the credit risk so that enterprises suffer less risk [3]; Wu Baode analyzed the credit rating level of SMEs from three levels of macroeconomic indicators, industry indicators and enterprise indicators [4]; Luo and Yong apply fuzzy synthetic method and analytic hierarchy process to study an example to evaluate the customer credit of logistics enterprise [5]. These studies are from the macro level analysis of the entire process of logistics. Although certain conclusions can be reached, the accuracy of the results is difficult to guarantee because the analysis is not specific enough. Therefore, this article has carried on the massive identification and the screening to the massive electricity supplier seller. We select a representative seller of e-commerce to capture the purchase record online. And then we use the principal component analysis to process classification. With the help of software such as SPSS, EXCEL and MATLAB, we separate the main indicators that represent the vast majority of data to Build the logistics credit evaluation model of e-commerce sellers based on BP neural network.

II. ESTABLISH INDEX SYSTEM

A. Selection of index system

Establishing the relevant index of seller's credit evaluation is the precondition to evaluate the seller's credit. Comprehensive consideration can fully reveal the influence factors of all aspects of enterprise credit, and comprehensively reflect the index system of enterprise credit status. Set up what indicators and how to set up are related to the direction of enterprise development. They also concern the scientific nature of the evaluation results and the accuracy and practicability of quasi Selection of index system. More importantly, the enterprise credit construction process can be successfully implemented. Therefore, designing index evaluation index system is the premise and foundation to evaluate logistics enterprise credit correctly. The specific design principles of the evaluation index system are as follows:

1) *Scientific principle.* The index system of evaluation must be established on the basis of scientific analysis and investigation. Content should be able to reflect the systematicness and comprehensiveness of the research, and should stress the key point

Biography: Ruijun Zhang, male, Xiantao, Hubei, Wuhan University of Technology, Professor, Master Tutor; Kui Fu, male, Xiangyang, Hubei, Master degree in logistics engineering, Wuhan University of Technology.

2) *Objectivity principle.* The establishment of evaluation indicators should be able to reflect the objective facts as far as possible and reduce the impact of subjective factors as much as possible.

3) *Principle of comparability.* The selection of each evaluation index should meet the actual situation and is easy to implement and master in practice.

4) *Principle of comparability.* The design of evaluation indicators should be able to make a certain comparability among different enterprises and reflect the advantages and disadvantages of them.

The whole process of the electricity supplier logistics is to provide better service for the logistics demanders. So, evaluation system needs to reflect the service characteristics of the electricity supplier logistics, which are commodity quality, service level and time demand. Packaging is a core factor. We select more important evaluation indicators comprehensively through online questionnaire survey and expert review results.

According to the above principles, we finally select the 11 indicators that can best reflect the logistics credit status of the electricity supplier (Figure 1) to establish the evaluation system of logistics credit indicators.

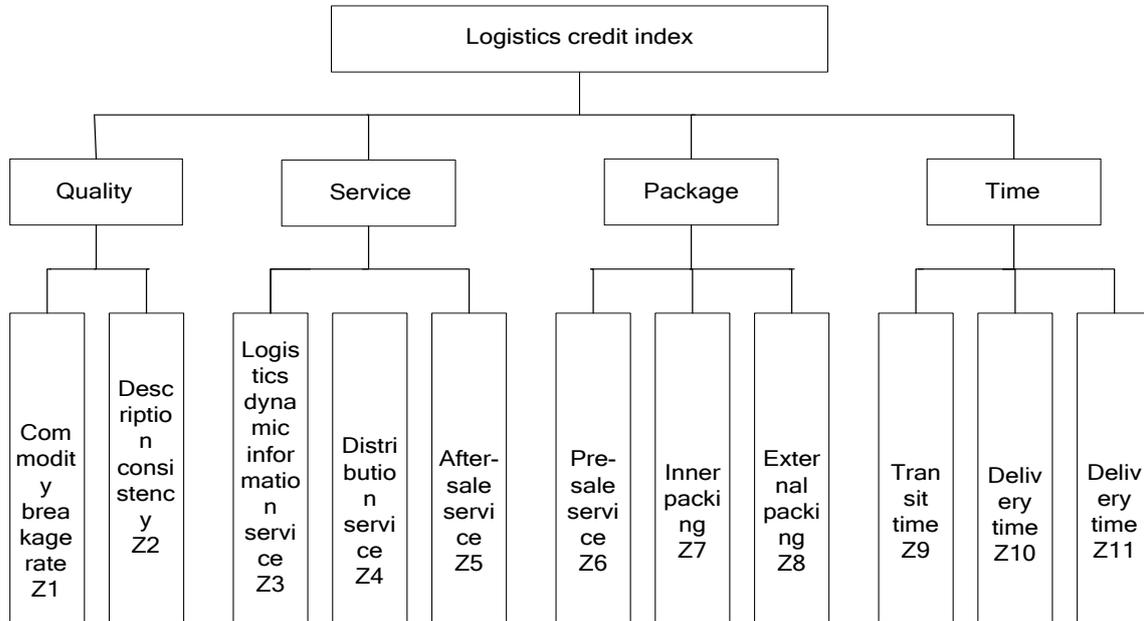


Fig. 1. Logistics reputation index

B. Quantification of index system

Select indicators through the form of questionnaires and quantify the relevant indicators in the form of star rating. As shown in Table 1:

TABLE I. LOGISTICS REPUTATION RELATED INDICATORS

First level index	Two level index	Index scoring quantization
Time	Delivery time	***** Very timely,* Rather late
	Delivery time	***** Very fast,* Rather late
	Transit time	***** Very fast, * Rather late
Package	External packing	***** Beautifully packaged,
	Inner packing	* Rough packing ***** Beautifully packaged * Rough packing

Table I, cont.		
Service	Pre-sales service After-sale service Distribution service Logistics dynamic information service	***** The service is excellent, * The service is terrible ***** The service is excellent, * The service is terrible ***** The service is excellent, * The service is terrible ***** Very accurate, * Very inaccurate
Quality	Description consistency Commodity breakage rate	***** Higher degree of coincidence, * Very inconsistent ***** No breakage, * Badly damaged

Through the analysis of the evaluation of each electricity supplier enterprise, we select a more comprehensive and more in line with the model data, then use the mean substitution method to fill missing data.

III. ESTABLISH A MODEL BASED ON PRINCIPAL COMPONENT ANALYSIS (PCA) AND NEURAL NETWORK (ANN)

A. Introduction of BP neural network

BP neural network (Back Propagation NN) is one of the most widely used artificial neural networks BP. It consists of an input

layer, a number of hidden layers (usually a layer) and an output layer. Each layer has a number of nodes, each of which represents a corresponding neuron. The lower layer and the upper layer are connected by the weight matrix and the layer is the whole connection between the nodes. A typical case of the BP neural network is a three layer network framework with a hidden layer. As shown in Figure 2.

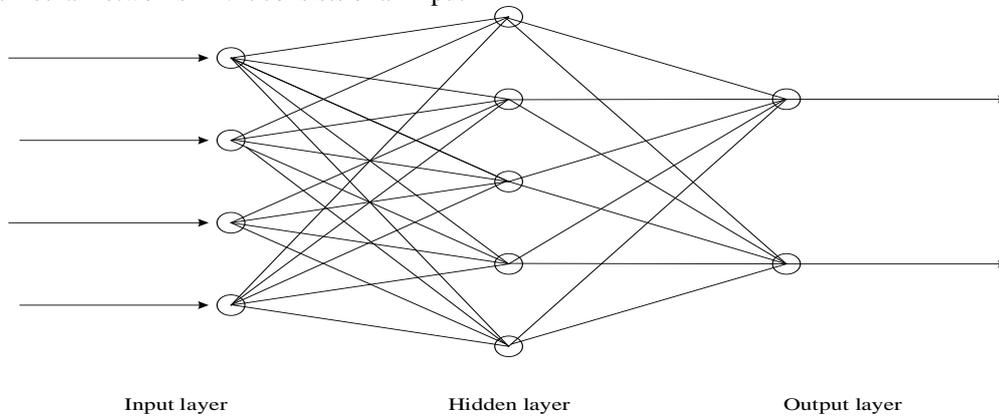


Fig. 2. Structure Model of BP neural network

B. Composition of neural networks

The fitting accuracy of the neural network is positively related to the network level and each layer of the nodes. Although increasing the network layer can improve the fitting accuracy, it can further reduce the error and improve the accuracy, but it will also increase the training time of the network and complicate the network [6]. In fact, we can improve the accuracy by increasing the number of neurons so that it is easier to observe and adjust the training results. According to Kolmogorow theory, under certain conditions, for any given epsilon >0, there exists a three layer neural network, which can approximate any continuous function or exact data by the precision of the mean square variance. Therefore, a three layer BP neural network structure with hidden layer is adopted in this paper.

The number of hidden layer nodes in BP neural network has no theoretical basis to determine its number. If the number is too small, the network will solve the problem with too little information; but too much will increase the training time and also increase the error and reduce generalization ability. Therefore, we use empirical formula to determine the number of nodes in the hidden layer.

Suppose that the BP neural network has N1 input neurons, X output neurons, and Y hidden neurons:

$$y = \sqrt{n_1 + x} + a \tag{1}$$

A is an integer from 1 to 10. Select different a values for testing and get the best results. In order to determine the optimal network structure, this paper will be based on the length of time

and the number of vectors of the actual output and the expected output of the network was trained to determine the size and variance of hidden layer nodes to make the network structure more suitable

C. Determination of output and output of neural networks based on principal component analysis

Principal component analysis (PCA) is also called principal component analysis (PCA), which combines the original information into a group of independent synthetic indexes (i.e. principal components) and can reflect the original main information. The idea is to use the idea of dimension reduction, in which each principal component can reflect most of the information of the original variable, and the information contained in is not duplicated. After the data is standardized, the principal component analysis of the standardized data is done to determine how the output of the neural network output. For neural networks, the number of input nodes will be analyzed by means of principal component analysis to obtain the principal component of variance contribution rate above 85% and replace the 11 related indexes

For the output layer, a single layer is used, and a more accurate value is obtained by using the function approximation method of the network. The output layer can also be multilayer, consisting of multiple neuron layers. In this paper, we choose a neural network with single output layer, then the unique output value of the neural network is a comprehensive evaluation of the credit level of the logistics.

D. Acquisition of data samples

Through network crawler technology, the related indexes of Tmall and Taobao are crawled through the network, and the relevant index data are screened out. Finally we chose 8 more in line with the seller of electricity supplier as a research sample, of which 2 were used as test data, and the remaining 6 were used as training sample data. Principal component analysis was carried out.

The data is pretreated, and the results of the sample data processing are shown in Table 2. Expect "1" show good reputation and recommend buying, and "0" means that the credibility is of the general, do not recommend the purchase.

TABLE II. INDEX DATA

Serial number	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Expect
1	0.7701	0.5405	0.3013	0.6821	0.5293	0.6691	0.5679	0.5467	0.4352	0.4494	0.5124	1
2	0.3744	0.4702	0.423	0.6130	0.3936	0.4106	0.4642	0.3034	0.4402	0.3418	0.3970	0
3	0.4604	0.5384	0.4312	0.5686	0.5194	0.4994	0.5594	0.4688	0.4552	0.3804	0.4800	1
4	0.4854	0.5926	0.4744	0.6456	0.5734	0.4952	0.5972	0.4866	0.5020	0.4912	0.4994	1
5	0.4234	0.4166	0.3785	0.5203	0.4461	0.4768	0.5423	0.4435	0.4501	0.4730	0.5501	0
6	0.4001	0.4976	0.4423	0.5102	0.4868	0.4762	0.5107	0.4021	0.4425	0.3986	0.5003	1

According to the data normalization method, the data processing result of data preprocessing is shown in Table 3:

TABLE III. NEURAL NETWORK DETECTS SAMPLE DATA

Serial number	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11
1	0.3298	0.3351	0.412	0.3076	0.5120	0.3187	0.5523	0.3234	0.3910	0.4102	0.4521
2	0.4988	0.5266	0.4107	0.4505	0.5020	0.3679	0.4461	0.4266	0.4833	0.3107	0.3586

The principal component analysis of input data was carried out by means of SPSS software. According to the principle that the accumulative contribution rate is above 85%, the principal

component is selected so that the eigenvalues of the correlation coefficient matrix and the contribution rate of the variables are obtained, as shown in table 4.

TABLE IV. TOTAL VARIANCE EXPLAINED

Ingredients	Initial eigenvalue			Extract sum of squares load		
	Total	Variance %	Accumulate %	Total	Variance %	Accumulate %
1	5.300	48.179	48.179	5.300	48.179	48.179
2	2.354	21.404	69.583	2.354	21.404	69.583
3	1.807	16.428	86.011	1.807	16.428	86.011
4	.962	8.749	94.760			

5	.337	3.068	97.827			
6	.185	1.679	99.506			
7	.054	.494	100.000			
8	.000	.000	100.000			
9	.000	.000	100.000			
10	.000	.000	100.000			
11	.000	.000	100.000			

Extraction method: principal component analysis.

From the table, 3 principal component variables were extracted from principal component analysis, and the response rate of information was 86.011%, which can replace the original

data to reflect most of the original data. The corresponding three characteristic vectors λ_1 , λ_2 , λ_3 are respectively

$$\lambda_1 = (0.804, 0.660, -0.355, 0.733, 0.592, 0.912, 0.684, 0.940, 0.401, 0.669, 0.642)^T$$

$$\lambda_2 = (0.174, 0.720, 0.298, 0.377, -0.014, -0.016, -0.543, 0.110, 0.734, -0.575, -0.632)^T$$

$$\lambda_3 = (-0.485, 0.152, 0.847, -0.172, 0.483, -0.325, 0.409, -0.034, 0.407, 0.324, 0.156)^T$$

E. Training neural network

This paper uses the three layer BP neural network to model. We use the Sigmoid function as the excitation function

$$f(kx) = \frac{1}{1 + e^{-kx}} \tag{4}$$

$$f(x) = f(x)(1 - f(x)) \tag{5}$$

The corresponding error is :

$$\xi_k = y_k(1 - y_k)(T_k - y_k) \tag{6}$$

The system uses BP algorithm to train the network, and the average error of the control neuron node is not more than 0.0001.

$$\left(\sum_p \sum_{k=1}^n (y_k^p - T_k)^2 \right) / (n \cdot \sum p) \leq \varepsilon, \varepsilon = 0.0001 \tag{7}$$

To simplify the calculation of error back propagation, we simplify the calculation of reciprocal of transfer functions. The calculation of the modified weights of the error propagation back to the network uses the following

$$E_i^{(k)} = O_i^{(k)}(1 - O_i^{(i)}) \sum_{j=1}^{n^{(k+1)}} w_{ij}^{(l)} E_j^{(k+1)}$$

Formula: (8)

$$\Delta w_{ij}^{(k)} = \eta E_j^{(k+1)} O_i^{(k)} \tag{9}$$

$$w_{ij}^{(k)} = w_{ij}^{(k)} - \Delta w_{ij}^{(k)} = w_{ij}^{(k)} - \eta E_j^{(k+1)} O_i^{(k)} \tag{10}$$

The K said the case number, η indicated learning rate, ε said error, E said expectations, $O_i^{(k)}$ said output node i in layer K, $w_{ij}^{(k)}$ said node i the k layer of the k+1 layer node j connection weights. The linear relationship between 3 principal components Y1, Y2, Y3 and each index is obtained from λ_1 , λ_2 , λ_3 :

$$Y_1 = 0.804Z_1 + 0.660Z_2 - 0.335Z_3 + 0.773Z_4 + 0.592Z_5 + 0.912Z_6 + 0.684Z_7 + 0.940Z_8 + 0.401Z_9 + 0.669Z_{10} + 0.642Z_{11} \tag{11}$$

$$Y_2 = 0.174Z_1 + 0.720Z_2 + 0.298Z_3 + 0.377Z_4 - 0.014Z_5 - 0.016Z_6 - 0.543Z_7 + 0.110Z_8 + 0.734Z_9 - 0.575Z_{10} - 0.632Z_{11} \tag{12}$$

$$Y_3 = -0.485Z_1 + 0.152Z_2 + 0.847Z_3 - 0.172Z_4 + 0.483Z_5 - 0.325Z_6 + 0.409Z_7 - 0.034Z_8 + 0.407Z_9 + 0.324Z_{10} + 0.156Z_{11} \tag{13}$$

According to the formula (11) - (13) the first three principal components of each set of sample data are shown in table 5. Test sample data are shown in table 6.

TABLE V. THE FIRST THREE PRINCIPAL COMPONENTS OF SAMPLE DATA

Serial number Serial number	Y ₁	Y ₂	Y ₃	Credit situation
1	3.9990	0.3421	0.5007	1
2	2.7805	0.4067	0.7304	0
3	3.3204	0.3563	0.8104	1
4	3.5934	0.3808	0.9301	1
5	3.1989	0.1348	0.7782	0
6	3.0818	0.2850	0.8309	1

TABLE VI. DETECTION DATA PRINCIPAL COMPONENT

Serial number Serial number	Y ₁	Y ₂	Y ₃
1	2.5622	0.0271	0.9079
2	2.9035	0.5002	0.7524

The learning rate is set to 0.2. In the MATLAB, three layers of neural networks are established. According to formula (1), after several attempts, we determine the number of nodes in the hidden layer 5, which is the best. So far, the structure of neural network is established in this paper is: input layer has 3 neurons, 5 neurons in hidden layer, the output layer is 1 neurons. After 1000 times of training, the overall error meets the requirements. Test data detection results are shown in Table 7.

TABLE VII. NEURAL NETWORK MODEL TRAINING SAMPLE OUTPUT

Project number	1	2
Network output value	0.9942	0.0003
Actual credit	1	0

IV. CONCLUSION

Through the test results, we can see that the logistics reputation evaluation model based on principal component analysis and BP neural network constructed in this paper is effective. Compared with the method without principal component analysis, the convergence is faster and the conclusion is more accurate. At the same time, with the help of MATLAB and SPSS software, the results can be obtained quickly. It can be seen that this method can solve the related problems well. However, because the data samples of this paper are few and the principal component analysis method cannot display the characteristics of large amount of data, the accuracy of network training is also affected. In the future work, we need

to collect more sample data to improve the training accuracy of the network so that the evaluation results can be more accurate.

ACKNOWLEDGEMENT

This research is supported by High School Education Research Program, under Grant 2016239 from Hubei Province, by Natural Science Foundation Under Grant 2014CFB804 from Hubei Province, China.

REFERENCES

- [1] Tsai C, Hu Y, Lu Y. Customer segmentation issues and strategies for an automobile dealership with two clustering techniques [J]. *Expert Systems*, 2015, 32 (1): 65-76.
- [2] Wang Zhuolun, Wang Xingfen. Credit evaluation index system and risk warning of online retail logistics enterprises [J]. *Fujian computer*, 2017, 33 (4)
- [3] Guo Zhanglin, Cui Cui. Study on credit risk based on supply chain finance [J]. *Journal of value engineering*, 2017, 36 (8): 77-79.
- [4] Wu Baode. Construction of credit evaluation system for SMEs based on logistics finance [J]. *logistics technology*, 2017 (3)
- [5] Luo Y, Chen Z. Client Credit Evaluation Index System of Logistics Enterprises Based on Fuzzy Comprehensive and Analytic Hierarchy Process [M]. *Information Computing and Applications*. 2013: 323-332.
- [5] Luo Y, Chen Z. Client Credit Evaluation Index System of Logistics Enterprises Based on Fuzzy Comprehensive and Analytic Hierarchy Process [M]. *Information Computing and Applications*. 2013:323-332.
- [6] Geng Yong, Ju Songdong, Chen Yana. Logistics demand analysis and forecasting based on BP neural network [J]. *Logistics Technology*, 2007, 26 (7): 35-37.