Taobao Shop Credit Evaluation Model Based on BP Neural Network

Dan Bai[1, a],Min Zuo[2, b]

[1]Institute of economic management, Dalian university, Dalian, Liaoning, China

[2]Institute of economic management, Dalian university, Dalian, Liaoning, China

[a]baidan@dlu.edu.cn

[b]072013zuomin@163.com

Keywords: Taobao Shop Credit evaluation model.BP neural network.

Abstract. This paper establishes the Taobao shop credit evaluation model in terms of shop information, goods, and services. Using 242 shop's information from Taobao, after two-step clustering method in SPSS and using BP neural network toolbox, the evaluation index as input, clustering results as output, built the BP neural network of 230 data samples for training, getting the BP network which has a faster convergence speed and higher accuracy rate. On this basis, using 12 shops for the simulation of the endingmodel to predict the credit of shops objectively and eventuallyestablishesthe Taobao shop credit evaluation model.

1. Introduction

By influencing China's economic and social development and changing people's lifestyles,the Internet had becoming a key industry.According to the China Internet Network Information Center recently released its34th China Internet Development Report, as of June 2014, China's online shopping users reached 332 million, compared with June 2013, an increase of 60.6 million, an increase of 22.37%[1]. Although the expansion of the scale of online shopping to stimulate the development of network economy, but also triggered a credit crisis. August 2014, China electricity supplier Research Center released the 2014 *Chinese e-commerce user experience monitoring reports and complaints,*which in China's top ten complaint online shopping site, Taobao ranked first treasure[2].The data show it, the first half of 2014 submitted by online, telephone, mail, public micro-channel number, instant messaging and other forms, platform users throughout the country received a total of 50,180 complaints from e-commerce, an increase of 21.32%. More and more new questions about the integrity of the electricity supplier and service put a higher demand, to improve confidence in e-commerce users, must first let users know under what circumstances is a trusted e-commerce, which involves Taobao shop credit problems.

The traditional method of credit evaluation model is based on statistical methods: commonly include discriminant analysis, Logistic regression method, and Bayesian risk identification method and soon.Often exhibit between Taobao shop credit evaluation and credit rating non-linear characteristics, so the above methods cannot solve this non-linear problem, and the neural network is considered to be more suitable for nonlinear characterization indexes.From the first time in 1990 the Odom introduce neural network to credit evaluation, the neural network method to get attention of researchers and practitioners; Piramuthu using neural network technology to study the Belgian corporate credit, the results show that the neural network has a high accuracy rate[3]; David use neural network technology to establish a commercial bank credit evaluation model, classification accuracy was 87.74%[4]; Yueh-MinHuang such as the use of neural network method and the minimum amount of data to evaluate corporate credit research results show that the neural network approach can be effectively evaluated results[5]; MinXilin, Liu Guohua, who uses BP neural network to establish a predictive model based on trusted e-commerce web user behavior[6]; Lu Red Branch, Cao Jie, Wu Yinghuiusing BP neural network method for commercial bank personal credit evaluation's study[7]. Therefore, this article will use BP neural network model for taobao shop credit evaluation's study.

2. BP Neural Network Theory

BP neural network is relatively simple: Fig.1 shows a typical three-layer BP neural network structure. $X = (x_1, x_2, ..., x_n)^T$ is the input vector, $Y = (y_1, y_2, ..., y_m)^T$ is output vector, n indicates the number of dimensions of the input vector, m is the dimension of the output vector, to form a full interconnection between the layers of neurons, each of the layers within the element does not any contact[8]. The output of the error back-propagation is through some form of anti-pass through the hidden layer to the input layer, layer by layer, and the error is apportioned to all layers of neurons, resulting in an error signal layers of cells, this error signal is corrected Process each neuron weights and thresholds basis weights and thresholds to constantly adjust, that is, BP neural network training process. This process continues until the output error is reduced to an acceptable or to a preset number of times until the study.
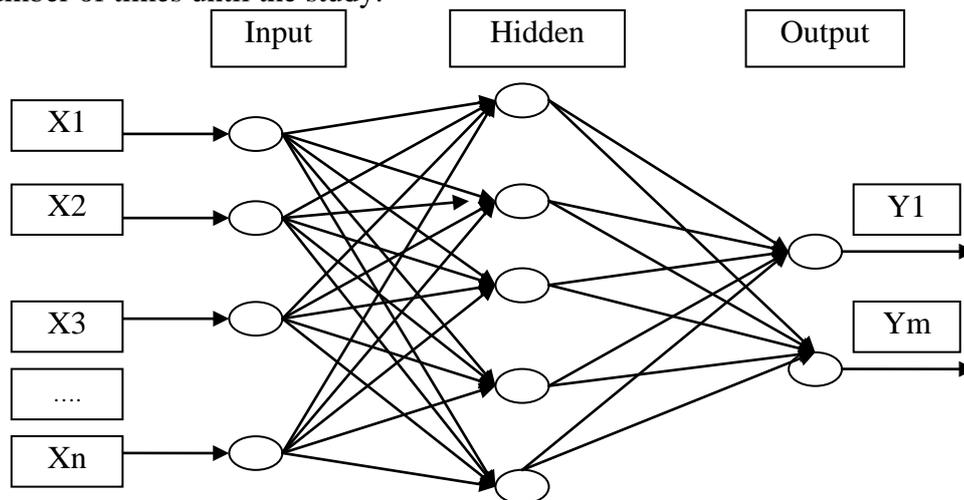

Fig.1 BPNeural Network of Three layers

3. Credit Evaluation Modelof Taobao Shop under BP Neural Network

3.1 Taobao shop credit evaluation index system. Factors related to e-commerce based on credit, combined with the content of consumer complaints, and by reading a lot of literature and the existing evaluation methods and model systems of Taobao, as well as previous studies of the evaluation system of the problems comprehensive research study, this paper established Taobao shop credit evaluation index system which shown in Table 1, the information from the shop, the goods and the services of the shops.

Table 1 The Taobao Shop Credit Evaluation Index System

Item	First level index	Second level index
Taobao shop credit evaluation index system	Informationof the shop	Caution money
		The time of setting up shop
		Punishment
		Favorable rate
		Refund rate
		Top favorites
	Goods of shop	Consistent with the description of the picture
		Number of successful trading
		The cumulative amount of the transaction
	Services of shop	Attitude
		Delivery speed

3.2 Taobao shop credit evaluation model. First, the data selection, this paper based on Taobao shop credit evaluation index system established by the above second level index, gather relevant data from Taobao. In order to ensure the accuracy of the model, the paper selects only "clothing" industry-based data store as a target, choose the shop for sales Lord's clothing, has collected 256 data . Due to some other causes some data is missing, after screening, and ultimately produce 242 valid data, as shown in Table 2 (due to more data, list only part of the data).

Table 2   The Taobao Shop Credit Evaluation Index Date（part）

Number	Credit value	Credit rating	Caution money	The time of setting up shop	The cumulative amount of the transaction	Top favorites	Description of the picture	Attitude	Delivery speed	Favorable rate （%）	Refund rate	Punishment
12	4682572	3J	500	5	238	711057	4.6	4.7	4.7	98.01	1.54	0
13	1482994	2J	1000	7	210	319660	4.7	4.7	4.7	98.63	4.14	0
53	860237	1J	1000	6	5482.6	15662	4.7	4.7	4.6	98.18	8.29	0
14	683832	1J	500	8	1691	328299	4.7	4.8	4.8	99.14	1.98	0
78	318054	5y	51000	9	15960	513	4.8	4.8	4.8	99.97	7.00	0
206	290138	5y	824	8	272	182	4.7	4.8	4.7	99.10	10.51	0
11	254085	5y	1000	2.5	95040	21789	4.6	4.7	4.6	96.94	1.69	0
104	232665	5y	55000	9	2430	531	4.9	4.9	4.9	99.99	4.40	0
136	116170	4y	1000	3	680	72	4.6	4.6	4.6	98.12	3.61	0
10	112547	4y	1000	3	746704	29859	4.5	4.6	4.6	94.97	6.89	0
192	69474	3y	50000	10	41613	480	4.9	4.9	4.9	99.99	10.60	0
92	69056	3y	5000	6	16898	2017	4.8	4.8	4.7	99.99	9.78	0
159	68534	3y	7500	5	300	342	4.8	4.8	4.8	98.97	2.30	1
21	67243	3y	500	9	0	30	4.8	4.8	4.8	99.25	1.34	0
191	63711	3y	15000	3	0	180	4.6	4.7	4.7	99.30	2.30	0

stake：J—golden delicious； Y—silver crown； Z--diamond； X--hearts

Secondly, the preprocessing of data, due to the BP neural network learning is typically supervised learning, the self-learning process is an iterative process, so the credit evaluation model established the need for learning objects. Taobao stores credit rating is currently divided into four major categories of 20 grades, but too many of these buyers are not too sensitive credit rating, credit rating in this article will be divided into five new levels, and using SPSS13.0 software, through a two-step clustering method to determine the number of clusters of the data. The first three columns in Table 2 as the reference index, the refund rate and the penalties as the cost index, the other as the benefit index. When during data processing, these two indicators value will be from positive to negative, and then get through the software analysis results shown in Table 3 and Table 4.

Table 3 The Cluster Distribution

Cluster Distribution		N	% of Combined	% of Total
Cluster	1	6	2.5%	2.5%
	2	87	36.0%	36.0%
	3	22	9.1%	9.1%
	4	104	43.0%	43.0%
	5	23	9.5%	9.5%
	Combined	242	100.0%	100.0%
Total		242		100.0%

Table 4 The Date of Automatic Clustering Export（part）

Credit value	The time of setting up shop	Cumulative amount of the transaction	Number of successful trading	Top favorites	Description of the picture	Attitude	Delivery speed	Favorable rate %	Refund rate	Punishment	TSC 4 82 2
500	5	238	20	711057	4.6	4.7	4.7	98.01	1.54	0	1
1000	7	210	20	319660	4.7	4.7	4.7	98.63	4.14	0	1
1000	6	5483	99	15662	4.7	4.7	4.6	98.18	8.29	0	4
500	8	1691	19	328299	4.7	4.8	4.8	99.14	1.98	0	1
51000	9	15960	105	5513	4.8	4.8	4.8	99.97	7	0	3
825	8	272	4	182	4.7	4.8	4.7	99.1	10.5	0	4
1000	2.5	95040	9698	21789	4.6	4.7	4.6	96.94	1.69	0	1
55000	9	2430	9	531	4.9	4.9	4.9	99.99	4.4	0	3
1000	3	680	8	72	4.6	4.6	4.6	98.12	3.61	0	5
1000	3	746704	13334	29859	4.5	4.6	4.6	94.97	6.89	0	1
50000	10	41613	13	480	4.9	4.9	4.9	99.99	10.6	0	3
5000	6	16898	71	2017	4.8	4.8	4.7	99.99	9.78	0	4
7500	5	300	4	342	4.8	4.8	4.8	98.97	2.3	1	3
500	9	0	0	30	4.8	4.8	4.8	99.25	1.34	0	4
15000	3	0	0	180	4.6	4.7	4.7	99.3	2.3	0	4

3.3 Establishment of credit evaluation model based on BP neural network . To determine the BP neural network model, needing to grasp the following aspects: First, to determine the level of neural network, the second is to determine the number of neurons in each layer, the third is to determine the parameters of BP network, included hidden layer and transfer function of the output layer selection, selection training function, the initial data is normalized, and determining a maximum target error training step.

we use the BP network, by increasing the number of nodes in the hidden layer, to determine the best BP network structure. In this paper, the sample data in the 230 data as training data, 12 data as a test data for simulation model training: the error target goal to 0.015, the training step is set to 100 000, other parameters are default values for the toolbox.

For the determination of the number of layers of neurons, paper selected 11 secondary credit evaluation index as input, the input layer neuron number n = 11; the number of output neurons is determined by the credit rating model results, due to the "shop credit rating "as the target output, with the expression of the corresponding figures were credit rating: 1 for high credit rating; 2 indicates a high credit rating; 3 for the general credit rating; 4 represents a lower level, 5 represents low-grade, because the desired output value is a single digital form, so the network output layer neuron number m = 1. Hidden layer neurons of choice are a very complex issue, and often need to be determined empirically and multiply. Commonly used empirical formula:(1)$\sum_{i=0}^{n} C_l^i > k$,(2)$l = \sqrt{m+n} + \partial$,(3) $l = \sqrt{mn}$,(4)$l = \log_2 n$,(5) $l = 2n + 1$,where, k is the

number of samples, l is the number of hidden layer neurons, to [1, 10] constants. This paper selection formula (2) and (5) to select relatively broad range of the number of hidden layer neurons, can determine the scope of the initial number of neurons in the hidden layer is from 5 to 23, designed by Matlab7.1 taking the hidden layer nodes from 5-23, the output of the error have all the data there are negative, the paper will have to do the absolute value of the output of ABS handle, for the absolute value of error data of different nodes in different averaging, through hundreds of training, changing the weights and thresholds, to give optimum error values of different nodes, can be seen from the right side of Table 5, when m = 10, the average error is the smallest, to determine the optimal number of hidden layer neurons 10.

Choose different training functions, the convergence rate of BP network is different, and the training error also have some impact, the paper set the maximum step training epoch = 100000, respectively usetrainglm, traingdx, trainda, traingdm and traingd training functions frequently such as BP network training, the training results as shown in Table 5 on the left side, so as to determine the training function network trainlmfunctions. Fragment, in the ending model, the hidden layer transfer function istangsig function; the transfer function fortrainlm function; the output layer function ispurelinfunctions.

Table5 The Network Training Error

Training function	The average training error	Step number	Hidden layer nodes	The average training error	Hidden layer nodes	The average training error
trainlm	0.01068	9	5	0.05404	14	0.12831
traingdx	0.44715	20	6	0.06997	15	0.11008
traingda	0.15368	51	7	0.10444	16	0.04844
traingdm	0.14702	38138	8	0.15224	17	0.14214
traingd	0.1423	100000	9	0.12603	18	0.16152
			10	0.01213	19	0.05771
			11	0.09147	20	0.13011
			12	0.14529	21	0.12001
			13	0.07731	22	0.09769

Table 6 TheSimulation Test Output

Number	1	2	3	4	5	6
Desired output	4	4	4	4	2	2
Simulation output	3.998225	3.99607	3.99368	4.006956	2.735248	1.996985
Error value	0.001775	0.00393	0.00632	-0.00696	-0.73525	0.003015
Number	7	8	9	10	11	12
Desired output	2	5	5	3	3	1
Simulation output	2.003612	4.999299	4.994338	2.998056	2.999855	1.059038
Error value	-0.00361	0.000701	0.005662	0.001944	0.000145	-0.05904

Finally, test the model to detect the accuracy. The simulation experiments used in MATLAB, such as the results in Table 6, with12 samples tested, only sample 12 and sample 5 does not reach the set goal = 0.015, the remaining 10 samples were achieved model accuracy reached 83.33%, indicating a high training accuracy of the model, a strong classification ability.

4. Conclusion

In this paper, using BP neural network, designed and built Taobao shop credit evaluation model. The BP neural network algorithm used in the model, is a nonlinear method, with no significant subjective component and human factors, simply enter the processed data to the network, through BP neural network toolbox can be obtained the evaluation results, to avoid the problem of determining the weights, to achieve automation and intelligent evaluation.The process is simple of the seller's credit rating with a more accurate assessment of the credit rating for Taobao shop which is expected to develop an evaluation system, the trained neural network model is embedded in the

system, enter the index value, the system will automatically give a ranking. At the same time, the model can also be used to evaluate some of the Taobao shopactivities, even for a compared analyzedfor the C2C e-commerce in abroad, and thus giving a better understanding of China's e-commerce.

References

[1]*http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201407/t20140721_47437.html.*

[2] *http://www.100ec.cn/detail--6192810.html.*

[3]Piramuthus. Financial credit-risk evaluation with neural and networksystem [J].European Journal of Operational Research(1999).

[4]DavidWebb.Doescreditrationing imply in sufficient lending?[J].Journal of Public Economics.(2000).

[5] Yueh-MinHuang,Chun-Minhung, HewijinchrIstinejiau. Evaluation of neural networks and data Ming methods on a credit assessment task for class imbalance problem[J]. Nonlinear Analysis: Real World Applications(2006).

[6]Min Xilin, Liu Guohua. Use BP neural network to develop network application [J].Computer Applications , (8):163 - 164(2008).

[7] Wu Yinghui. Personal credit risk assessment analyse of BP neural network [J].Technology Start-ups,(7):98-100(2012).

[8]FeiSike Technology R&D Centre.The design of assist BP neural network in Matlab6.5 [M].Electronics IndustryPublishers (2007).