ISSN: 2688-9323

Research on Enterprise Credit Rating based on BP Neural Network Model

Qi Tang*

School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, 233030, China

*tangqi2319@163.com

Abstract

In the credit market, due to the asymmetric information between commercial banks and enterprises, to maximize their own interests, banks need to evaluate their credit risk based on the strength and reputation of small and medium-sized enterprises, and make credit decisions such as whether to lend or not, loan line, interest rate, and term. In this paper, a certain number of enterprises without credit records are taken as the training samples of the BP neural network, and the credit rating evaluation model based on BP neural network is established. The decision-making of enterprises still meets the constraints, and the credit rating of enterprises is obtained, which can be used as a reference for the credit rating evaluation of small, medium, and micro-enterprises.

Keywords

Credit Strategy; Risk Assessment System; Credit Rating; BP Neural Network.

1. Introduction

With the promulgation of a series of national policies to support entrepreneurship, the number of MSMEs has exploded in recent years and has rapidly driven the development of local banks' lending businesses. In the capital market, due to the existence of asymmetric information between commercial banks and enterprises, banks are the disadvantaged and incompletely informed party of information, and the relatively small size of MSMEs and lack of collateral assets, therefore, in actual operation, banks, in order to ensure the maximization of their own interests and the stability of customer sources, usually need to rely on credit policies, information of transaction notes of enterprises and the influence of upstream and downstream enterprises to make an assessment of MSMEs' In order to maximize their own interests and ensure a stable source of customers, banks usually need to assess the strength and creditworthiness of MSMEs based on credit policies, information on their trading notes and the influence of upstream and downstream enterprises, and then make a prediction on their credit risks. Determining credit strategies such as whether to lend, loan amount, interest rate and term based on credit risk and other factors is the core essence of the bank's credit system construction.

After finding the factors affecting the credit rating of enterprises, the values of the influencing factors are used as the training samples of the BP neural network, to establish the BP neural network rating model, test the accuracy of the model, and get the credit rating of enterprises through the test, and make a risk assessment and decision optimization for enterprises.

2. Acquisition of Data and Assumptions

The data in this paper are obtained from the invoice data related to 123 enterprises with credit records and 302 enterprises without credit records provided by question C of the 2020 National

ISSN: 2688-9323

Student Mathematical Modeling Competition. The credit risk of MSMEs is evaluated by analyzing their strength and creditworthiness to determine whether to lend and credit strategies such as loan amount and interest rate setting.

To facilitate the analysis of the problem, the following assumptions are made on the data used in this paper: (1) Assume that the bank's total return to the credit enterprise is always greater than the benefit obtained from safe investment (e.g., purchase of treasury bonds, etc.), i.e., under the condition that the enterprise's default risk is considered, the bank will consider lending to the enterprise only when the bank's expected return is non-negative; (2) Assume that the expected return generated by the enterprise under the credit contract designed by the bank is not smaller than the expected return generated by the enterprise under its own designed credit contract; (3) it is assumed that MSMEs apply for credit loans from only one bank; (4) it is assumed that the model is built without considering the influence of other unexpected factors; (5) it is assumed that there is no misrepresentation of risk information by enterprises and the relevant data are true and reliable.

3. Reputation Rating Evaluation Model based on BP Neural Network

3.1. Theoretical Preparation

The indicator of creditworthiness rating is manually rated by the bank internally according to the actual situation of the enterprise. According to this feature, for enterprises without credit records, certain prediction methods can be used to predict the future default situation and credit rating of the enterprise based on the existing account information. Considering that the credit rating is not a binary variable and the multi categorical logistic regression focuses on the study of influencing factors rather than prediction, the BP neural network method is used to predict the credit rating of an enterprise, where the neuron model in the BP neural network is shown in Figure 1.

In this model, we use BP neural network for relevant prediction. The BP network is able to learn and store a large number of input-output pattern mapping relationships without revealing the mathematical equations describing such mapping relationships beforehand. Its learning rule is to use the fastest descent method to continuously adjust the weights and thresholds of the network by backpropagation to minimize the sum of squared errors of the network. the topology of the BP neural network model consists of an input layer (input), a hidden layer (layer), and an output layer (output).

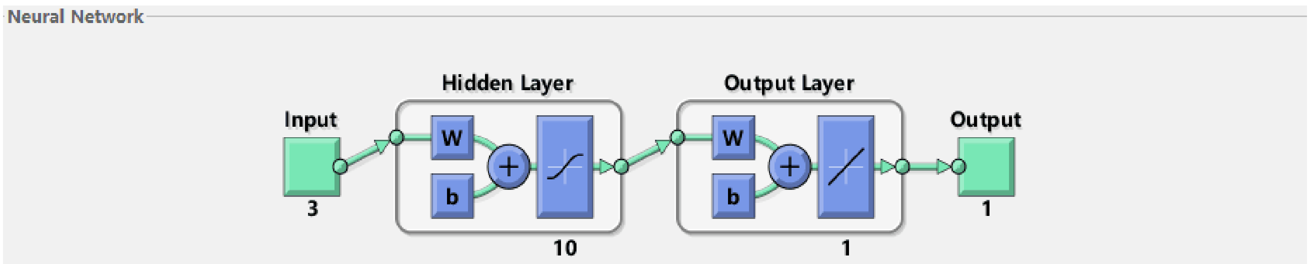

Figure 1. Schematic diagram of the neural network prediction model

3.2. Model Preparation

The structure of the BP neural network-based enterprise reputation level evaluation model is designed as follows.

3.2.1. Determination of the Number of Model Layers

The model selects a 3-layer BP neural network for training, which is easier to achieve than increasing the hidden layers of the model because it relies on increasing the number of nodes in the hidden layers to improve the model accuracy and reduce the model error.

3.2.2. Determination of the Number of Nodes in the Input Layer of the Model

Firstly, the quantitative indexes identified are considered as the input and the quantitative credit risk score is used as the output, and the neural network method is used for training, resulting in the relationship between the test set and the optimal solution as shown in Figure 2.

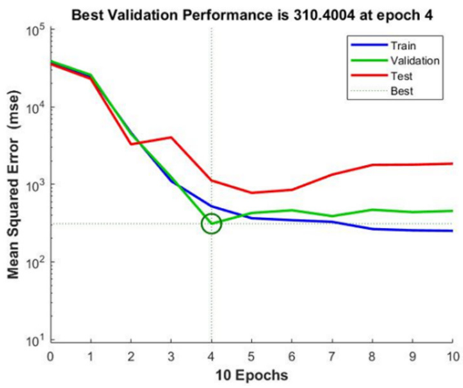

Figure 2. Neural network prediction validation graph

From the figure, it can be seen that if all the first-level indicators in the first question are selected as the training set, the results of the test set obtained differ significantly from the optimal solution, and the model accuracy is reduced, for which it is not accurate.

3.2.3. Determination of the Number of Nodes in the Output Layer of the Model

Referring to the credit rating index, each evaluation index in the first part is the network input information, and the credit rating is set as the network target output. In this paper, the individual credit risk is divided into four grades, A, B, C, and D. Therefore, the number of nodes in the output layer of the model k=4.

3.2.4. Determine the Number of Nodes in the Hidden Layer of the Model

The number of nodes of the hidden layer is generally selected by empirical method or trial-and-error method, with the following reference formula:

$$l < n - 1 \tag{1}$$

$$l < \sqrt{n+k} + i \tag{2}$$

where l represents the number of nodes in the hidden layer, n represents the number of nodes in the input layer, k represents the number of nodes in the output layer, and I represent an arbitrary constant between 0 and 9. Each research index except the 4 reputation rating data is used as the input vector, and the number of units in the input layer of the network is 10.

The other parameters in the neural network are kept constant, and the number of neurons in the hidden layer is adjusted for repeated controlled trials, and the optimal number of neurons in the hidden layer is determined by comparing the error between the output value and the expected value. After the network is trained, the mean square error of the BP neural network reaches the minimum value of 0.0091 when the number of neuron nodes in the hidden layer is

7, and the best approximation of the function is achieved at this time. Therefore, the number of neuron nodes in the hidden layer is chosen to be 7.

Table 1. Neural network prediction training error table

Neuron number	5	6	7	8	9	10	11
Res	0.0281	0.0193	0.0091	0.0234	0.0364	0.0431	0.0483

According to the above analysis, the structure of the proposed BP neural network is: the number of neurons in the input layer is 10; the number of neurons in the hidden layer is 7, and the number of neuron nodes in the output layer is 1.

3.3. Model Solving

The backpropagation algorithm is used to train the weights and deviations of the network iteratively so that the output vector is as close as possible to the desired vector. The training is completed when the sum of squared errors in the output layer of the network is less than the specified error, and the weights and deviations of the network are saved.

Initialization, randomly given each connection power as $[w]$and$[v]$, and threshold value as θ_t and r_t ;

From the given input_output pattern for calculating the output of each cell of the hidden layer, output layer.

$$b_j = f(w_{ij}a_i - \theta_j)c_t = f(v_{jt}b_j - r_t) \tag{3}$$

In the above equation: b_j is the actual output of the jth neuron in the hidden layer;c_t is the actual output of the tth neuron in the output layer;w_{ij} is the connection power from the input layer to the hidden layer; v_{ij} is the connection power from the hidden layer to the output layer.

$$d_{tk} = (y_{tk} - c_t)c_t(1 - c_t)e_{jk} = [d_t v_{jt}]b_j(1 - b_j) \tag{4}$$

The next input pattern pair is selected to return to step 2 for repeated training until the network sets output error to meet the requirements.

According to the above steps, the sample of known complete 123 enterprises is divided into two parts, and the sample data of 90 of them are used as the training set to train the neural network, and the sample data of the remaining 33 enterprises are used as the validation set to fit the validation, and the training results are shown in Figure 4. The training model is provided for the neural network and tested separately, and the specific process is expressed as follows: the total enterprise revenue n, progress factor m, transaction preference F, transaction law L, and invalid invoice proportion Q are input into the first-stage neural network, and the output is the default case V. Then the default case V and invalid invoice proportion Q are input into the second-stage neural network, and the reputation rating R is output after the two-stage neural network training.

ISSN: 2688-9323

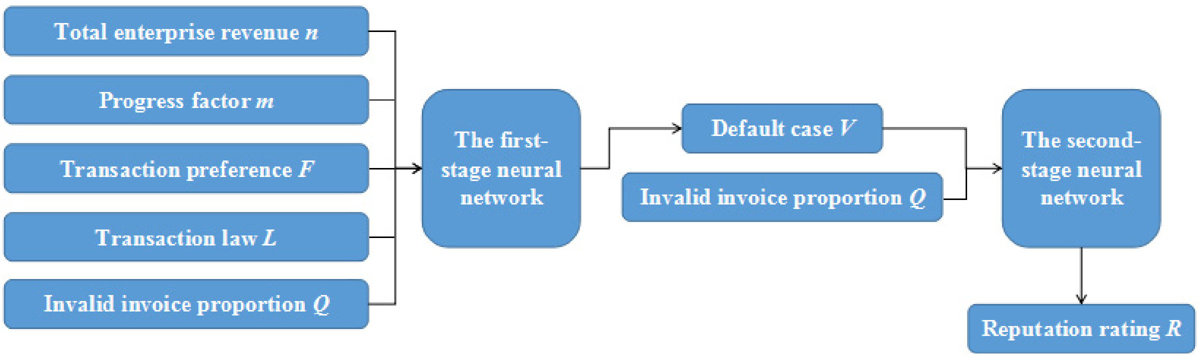

Figure 3. Neural network flow chart

The trained BP neural network can only output the normalized data, in order to get the real data values, we must also denormalize the output data. The denormalization process can use the information in the normalization process, through the function "mapminmax" to achieve. The details are as follows.

BPoutput=mapminmax ('reverse',an, outputps);

where BPoutput is the data after inverse normalization, an is the neural network predicted output, and output is the original output dataset information.

The performance of the neural network at the end of training is plotted as follows.

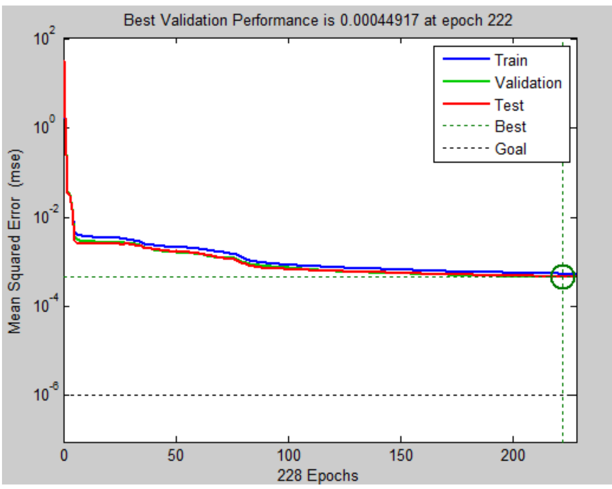

Figure 4. BP neural network training performance graph

At this point of training, the iterative process reaches the minimum mean square error, MSE=0.00044917, and training is completed.

After quantifying the credit registration, the network is trained using the existing data of 123 enterprises, and the data of 302 enterprises are descended for credit rating prediction after the training is completed, and the prediction results are as follows (due to the large amount of data, only data of 10 enterprises are shown here).

ISSN: 2688-9323

Table 2. Table of credit rating forecast results

Enterprise	Output Results	Credit Rating	Enterprise	Output Results	Credit Rating
E124	181.1061	A	E129	51.4694	C
E125	177.6945	A	E130	65.1373	B
E126	45.3612	C	E131	168.3685	A
E127	62.2184	B	E132	42.2855	C
E128	18.0842	D	E133	16.6390	D

The predicted output does not differ much from the desired output and basically achieves our required results.

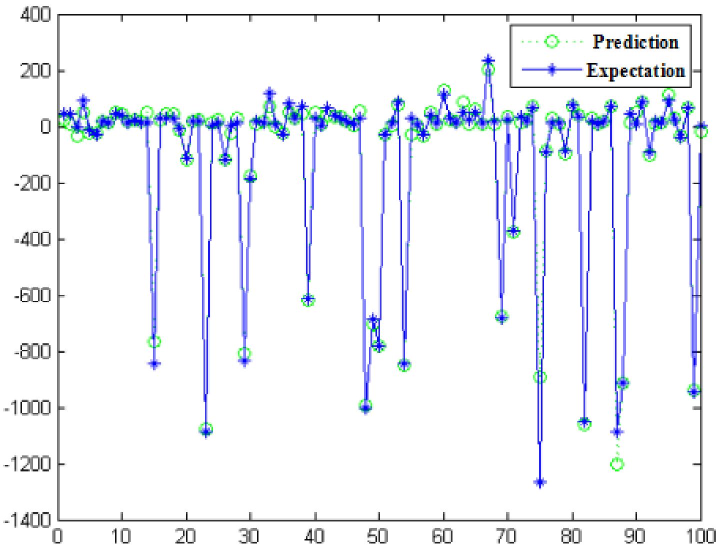

Figure 5. BP neural network prediction output graph

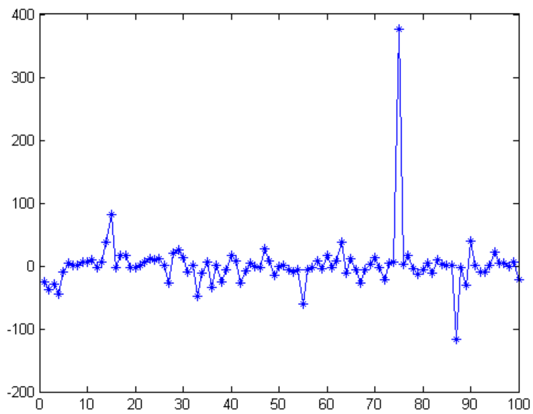

Figure 6. BP neural network prediction error graph

After evaluating the credit rating of enterprises without credit records, we came up with 4 enterprises with a credit rating of D, which in principle banks do not lend, 207 enterprises with a credit rating of C, and 91 enterprises with a credit rating of B. To verify the validity of the model, we used the data with known credit ratings as the measurement set, and used the algorithm to derive the value of the credit rating evaluation of the enterprises, and compared it

ISSN: 2688-9323

with the actual value, resulting in 15 The data evaluation discrepancy, thus we can see that the algorithm has certain reliability.

In order to explore the degree of impact of unexpected factors on different industries and different categories of enterprises, we classified the industries to which the enterprises belonged based on the national industrial classification standards and the names of industries in the annexes, and quantified the degree of impact on industries under the impact of the epidemic by referring to the data related to the announcement of the new crown epidemic in the National Statistical Yearbook, resulting in the processed data shown in the following table.

Table 3. Enterprise classification results

Industry name	Number of enterprises	The degree of impact of sudden factors on the industry
Agriculture, forestry, animal husbandry,fishery	4	0
Mining	7	0
Manufacturing	16	-0.2
Electricity, heat, gas,water production,supply industry	3	0
Construction	48	-0.2
Wholesale and retail trade	26	-0.2
Transportation, storage and postal services	11	-0.4
Accommodation and catering	3	-0.4
Information transmission, software,information technology	14	0.8
Finance	17	0.2
Real Estate Industry	3	0
Leasing and Business Services	24	0.2
Scientific Research and Technology Services	29	0.2
Water, Environment and Public Facilities Management	0	0
Residential services, repair and other services	15	-0.2
Education	2	0.6
Health and social work	7	0
Culture, sports and entertainment	17	0.2
Public administration,security , social organizations	0	0
Individual business	56	-0.6

If the degree of influence of the contingency factor is x, the total revenue of the enterprise under the influence of the contingency factor is y, and the total revenue of the enterprise without the influence of the contingency factor is w, then $y=w*(1+x)$. Using the corrected total revenue and total profit as the training set, the BP neural network is again applied to evaluate the composite score of each enterprise to derive the bank's credit strategy for the 302 enterprises. Due to the large amount of data, only part of the data is shown here.

ISSN: 2688-9323

Table 4. Results of banks' credit adjustment strategies

Enterprise	Whether loan (Yes: 1; No: 0)	Loan amount	Loan interest rate
E124	1	64.39003734	0.04
E125	1	74.83389032	0.04
E126	1	81.55862368	0.04
E127	1	71.02087405	0.04
E128	1	72.93881821	0.04
E129	1	80.13887036	0.04
E130	1	34.08249054	0.04
E131	1	57.65990375	0.04
E132	1	62.70835337	0.04
E133	1	47.73408489	0.04
E134	1	64.73993394	0.04
E135	1	62.03388192	0.04
E136	1	37.47881975	0.04
E137	1	48.2451051	0.04
E138	1	47.34749115	0.04

4. Conclusion

In this paper, for enterprises without credit rating, we extracted indicators about enterprise profitability and customer stability based on data, and established a BP neural network-based model to evaluate the credit rating of enterprises without credit rating. After the credit rating is obtained, the credit risk score is ranked for the enterprises to quantify their credit risk, and the bank's decision on the lending amount and interest rate preference for the enterprises is derived, i.e., 4 enterprises with credit rating of D are denied credit applications, 207 enterprises with credit rating of C are given 0.0425 interest rate preference, and 91 enterprises with credit rating of B are given 0.04 interest rate preference. In addition, the credit risk of each enterprise in the consideration and the impact of possible sudden factors on each type of enterprise, the classification of each enterprise according to industry combined with sudden change impact factors gave the bank's credit adjustment strategy when the total annual credit was 100 million yuan.

5. Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Xu Ying, Huang Yimiao,Ma Guowei. A beetle antennae search improved BP neural network model for predicting multi-factor-based gas explosion pressures[J]. Journal of Loss Prevention in the Process Industries,2020,{4} (prepublish).

[2] Yan-Wen Li,Ke Cao. Establishment and application of intelligent city building information model based on BP neural network model[J]. Computer Communications,2020,153{5}.

[3] Machine Learning - Intelligent Systems; Study Data from Dalian University of Technology Provide New Insights into Intelligent Systems (Activation Functions Selection for Bp Neural Network Model of Ground Surface Roughness)[J]. Journal of Robotics & Machine Learning,2020,{4}{5}.

[4] Chemistry; New Findings from Harbin Normal University Describe Advances in Chemistry (The Profit Manipulation of Chinese Listed Companies Based On Bp Neural Network Model and Dea Model)[J]. Computers, Networks & Communications,2019,{4}{5}.

ISSN: 2688-9323

[5] He Jiao,Pan Congyuan,Liu Yongbin,Du Xuewei. EXPRESS: Quantitative Analysis of Carbon with Laser-Induced Breakdown Spectroscopy (LIBS) Using Genetic Algorithm and BP Neural Network Models.[J]. Applied spectroscopy,2019,{4}{5}.

[6] Sustainability Research - Sustainability Research; Findings from China University of Petroleum Broaden Understanding of Sustainability Research (Comparison of Forecasting India's Energy Demand Using an MGM, ARIMA Model, MGM-ARIMA Model, and BP Neural Network Model)[J]. Computers, Networks & Communications,2018,{4}{5}.

[7] Wenxu Wang,Ruichun Tang,Cheng Li,Peishun Liu,Liang Luo. A BP neural network model optimized by Mind Evolutionary Algorithm for predicting the ocean wave heights[J]. Ocean Engineering, 2018, 162 {5}.

[8] Algorithms; New Algorithms Findings from H.F. Liu and Colleagues Discussed (Study of a Gray Genetic BP Neural Network Model in Fault Monitoring and a Diagnosis System for Dam Safety)[J]. Computers, Networks & Communications,2018,{4}{5}.

[9] Jia Zhou,Xinming Wan,Junping Zhang,Zhao Yan,Yang Li. Modeling of Constitutive Relationship of Aluminum Alloy Based on BP Neural Network Model[J]. Materials Today: Proceedings,2015,2(10).

[10] Jun Liang,TingXue Sun,JianAn Wang,BaoLuo Li. Using BP neural network Model to Design and Implementation of green building[J]. International Journal of Digital Content Technology and its Applications,2012,6(18).

[11] Lv Shuran,Lv Shujin. Applying BP Neural Network Model to Forecast Peak Velocity of Blasting Ground Vibration[J]. Procedia Engineering,2011,26{5}.

[12] Chang-Xue Jack Feng,Abhirami C. Gowrisankar. Practical guidelines for developing BP neural network models of measurement uncertainty data[J]. Journal of Manufacturing Systems,2006,25(4).