Financial Credit Risk Warning Based on Big Data Analysis

Dongmei Cui

Hunan International Economics University,
ChangSha, Hunan, China

Abstract
Big data era promotes customer data, trade data, management data, etc, in bank industry appearing bombing increase, but data analysis technology development will transform the pattern of bank marketing strategy. This paper aims to implement Internet financial ideas in commercial bank, designs and applies more scientific, flexible, matching policy requirement and aiming credit risk evaluation model to offer referential measures based on big data. Based on further study on national commercial bank, especially for the status and application characteristics of urban commercial bank credit risk evaluation, we integrate the characteristics of various data mining and neural network technology to put forward BP network-based financial credit risk evaluation model, improve and verify the proposed model to offer feasible solutions for credit risk evaluation in commercial bank. Experimental results data from empirical analysis illustrates rationality and effect of this model.
Key words: FINANCIAL CREDIT RISK, BP NEURAL NETWORK, BIG DATA, PARAMETER

1. Introduction

Most of financially stable risk evaluations in current large project rely on traditional ways to sample data such as artificial questionnaire distribution, finite times hearing, and suggestion feedback platform in government portal website platform[1]. However, because communication channel in these systems is limited in covered area, bad attainability and poor timeliness are gradually replaced by new media such as micro-message, micro-blog, forum, etc. Public opinion data with huge quantity and various shapes intensify the tendency of local problems globalization, simple questions complexion, individual problems popularization and common issues hotspots with powerful influence, seepage force, unique openness and interaction so that social risk is largely increasing. Because of different reasons, these constant information on Internet have not been completely paid attentions by relevant departments and effectively fused with the searched data in traditional forms[2]. Some disadvantages such as data loss, mutual separation and poor time effectiveness damage accurate evaluation results. They not only waste resources but are not beneficial to respond public opinion and maintenance of social order. Thus, it has become an urgently solving problem on how to occupy commanding point of information in socially stable risk evaluation of large project[3]. However, most current related researches focus on index system construction, evaluation mechanism establishment, evaluation pattern exploration, etc. They seldom involve data collection and mining as evaluation foundation. However, the appearance of big data definition explores the new idea for risk managers and researchers[4-6].

Though big data has not authoritative definition, it has changed humans' existence style. In particular, it has unique advantages in evaluating and predicting unknown events. For example, Google accurately predicts influenza virus spreading in the whole United States through appearing frequency of influenza and its related vocabulary in networked search. For another example, anthropologists and mathematician in American California University set up crime activity prediction platform together to seek sooner case time and region by means of analyzing the past 1300,0000 cases. This prophet ability is the kernel target of social stability risk evaluation in large projects and traditional information processing method is difficult to realize. Thus, concept, method and tool of big data are embedded in current evaluation system to improve capturing width, depth and real-time of information and apply scientific method to mine key value in massive data, which is helpful to improve risk evaluation and prediction[7].

In big data, many risk evaluations are up to behavior data collection and automatic generation. Data and search engine can reduce information asymmetry in a large degree and calculate its default probability and evaluation risk based on enterprise or individual history. It will be extremely visual and convenient. To organically integrate data base technology and mining technology and intelligently calculate risk index, measure different customers' potential value and risk degree and automatically provide useful information for management decision[8,9]. If various source system data can be loaded in data base at first and implement its powerful data extraction and loading ability, it can effectively save risk data time and improve processing efficiency of data collection.

In big data, many risk evaluations are up to behavior data collection and automatic generation. Data and search engine can reduce information asymmetry in a large degree and calculate its default probability and evaluation risk based on enterprise or individual history. It will be extremely visual and convenient. To organically integrate data base technology and mining technology and intelligently calculate risk index, measure different customers' potential value and risk degree and automatically provide useful information for management decision. If various source system data can be loaded in data base at first and implement its powerful data extraction and loading ability, it can effectively save risk data time and improve processing efficiency of data collection. In other words, from theory, technology and data resource, it has a condition to design an automatic bank credit risk evaluation model. Through inputting basic data and automatic calculation, it can generate a possible result of customer risk evaluation and debt repayment. This is the basic idea of us to design risk evaluation model. The rest of the paper is organized as follows: In section 2, we introduce the evaluation measures and rate assessments for financial credit risks. In section 3, we analyze the of evaluation indicators and relative mining methods based on big data analysis. Then a BP network-based financial credit risk evaluation model is proposed for implementation on corresponding problems. In section 4, we provide empirical analysis by actual cases and prove our scheme. Finally, a conclusion is provided in section 5.

2. Financial Risk Evaluation and rate Evaluation

2.1 Credit Risk Measuring

The characteristic of standard method is adopting rating in exterior rating agency to determine asset risk weight and it is appropriate for banks in low complexity degree. In standard method, credit risk weight asset (RWA) equalizes to multiplication between breaking contract (EAD, that is risk exposure) and customers' risk weight (RW) which is determined by exterior rating agency.

$$RWA=RW\times EAD \tag{1}$$

Customer's risk weight (RW) is related to the rating in exterior rating agency. To take the most common company in our national commercial bank as an example, table 1 depicts the risk weight in standard method.

Table 1. Risk weight adopted by standard method

Credit rating	AAA	A^+	BBB^+	BB^+	B^+	Lower	Not

	AA^{-}	A^{-}	BBB^{-}	BB^{-}	B^{-}	than B^{-}	rated
Company Loan	20	50	100	100	100	150	100

2.2 Credit Risk Rating

To establish a reliable internal credit rating system, commercial bank needs to set scientific credit risk rating at first. Scientific credit rating set usually has following features:

(1) The selected credit rating number is enough to distinguish risks;

(2) Each kind of risk rating can display sufficient credit risk information such as default probability or default loss. Standard & poor and famous Moody credit rating are the most widespread internationally exterior credit rating institution and credit rating system in these two institutions are broadly verified and admitted. Credit rating in most of our national commercial bank is set according to standard & poor. Moody credit rating and our nationally practical situation. For instance, table 2 displays a company credit rating set in our nationally commercial bank.

Table 2. Enterprise Credit Rate Setting

Rating	Credit level	Description
AAA	outstanding	Enterprise has high degree of credit, debt risk. The strength of the enterprise funds, asset quality, the indicators of advanced, business situation is good, strong profitability, obvious economic benefit, development prospect is broad, strong capacity to pay the minimum payment, uncertain factors influence on the operation and development of enterprises, into the possibility of financial distress.
AA	perfect	Enterprise credit degree is high, the debt risk. This kind of enterprise financial strength is strong, good asset quality, the indicators of advanced management, better economic benefits, the higher level of profitability, stability, development prospects are broad, liquidation and payment ability. The influence of uncertain factors on the operation and development of very small.
A	best	Enterprise credit degree good, to repay the debt in normal circumstances there is no problem. The strength of enterprise funds, asset quality is acceptable, the economic indicators in the upper level, operating in a virtuous circle, but there may be some uncertain factors affect its future development and operation, and weaken its profitability and solvency
BBB	better	Enterprise credit level, the ability to repay the debt in general. This kind of enterprise assets and financial situation, the economic indicators in the medium level, but the operating conditions, the level of profitability and future development are susceptible to the influence of uncertain factors, liquidity fluctuations.
BB	common	Enterprise credit level is poor, the lack of solvency. This kind of enterprise assets and financial condition, various economic indicators are at a low level, the future development prospects uncertain, settlement and payment capacity of poor, vulnerable to the uncertain factors, risk. This kind of enterprise has a lot of bad credit records, containing the speculative factors.
B	bad	The credit level of enterprise, debt paying ability is weak. The enterprise once in a bad economic environment, avoiding the possible risk, but at present, there are the ability to repay principal and interest.
CCC	worse	Enterprise credit is poor, corporate profitability and solvency is weak. This kind of investment security is small, there is great risk and instability, almost no debt paying ability.
CC	worst	Enterprise credit is poor, the enterprise has been in a state of loss, highly speculative investors, no debt paying ability.
C	none	Enterprise credit, enterprise is unable to repay debts, losses, on the verge of

		bankruptcy.

3. Credit Risk Evaluation Information Mining and Maintenance Based on Big Data Analysis

3.1 Neural Network Mining Method

The major mechanism in genetic BP neural network is adopting strongly global search ability of genetic algorithm. At first, it uses genetic algorithm to perform pre-learning of global coarse precision on weight and threshold in BP network to locate the optimal region. Therefore, weight and the closed value population accumulate some locations in parametric solution space. Then, BP algorithm is used again to perform gradient fine search in these small solution spaces on effectiveness of local search to finally seek the optimal solution. Therefore, the training in genetic BP neural network is divided into two parts. First, using genetic algorithm to optimize initial weight in network. Then, to using BP algorithm to train logging data to obtain network model. This paper proposes genetic algorithm to optimize BP neural network. The optimization does not change the direction of BP algorithm relying on gradient information to adjust networked weight but to implement global search of genetic algorithm to seek the most appropriate network connection weight and networked structure.

Neural network structure is defined by neural element characteristics and neural element connection characteristics in network. Towards credit risk evaluation model, its structure can be expressed by BP network model input, output number, hidden layer quantity and various hidden layer nodes. In network structure of evaluating model, input nodes quantity can be directly obtained and it is the quantity of credit risk evaluation index. Output node number of model can be one or more than one. Towards classification model, output nodes quantity is related to the classified classification quantity. It is supposed that there are m ratings in credit rating so output node quantity of evaluating model can be m or $\log_2 m$. Under most circumstances, in order to simplify network structure and improve training efficiency of model, one or more output network models can be transformed into many network models with one output. Under this rule, this paper simplifies the output of evaluative model. At first, according to business requirement, credit rating is divided into six ratings including AAA, AA, A, BBB, BB and B. Next, according to credit rating method, evaluating model output is transformed into a continuous scale variable and different value scope of variables corresponds to different credit rating, as shown in table 3.

Table 3. Transformation Relations between Evaluation Model Output and Credit Rating

Rating	AAA	AA	A	BBB	BB	B
Output at training phase	0.95	0.85	0.75	0.65	0.55	0.45
Output at application phase	[0.9,1.0]	[0.8,0.9]	[0.7,0.8]	[0.6,0.7]	[0.5,0.6]	[0.0,0.5]

It is very important for evaluation model performance to determine reasonable hidden layer quantity and various hidden layer nodes quantity. It usually believes that increasing hidden layer quantity can reduce network error and improve classification accuracy of evaluation model. Meanwhile, it will promote evaluation model structure to be more complex, increase model training time and appear over-fitting tendency. Moges et al, has proved that three-layer BP network can approximate random mapping relation fill under certain conditions (Moges Helen-Tadesse, *et al,* 2013). After experiment, this paper discovers that two hidden layer model is not obviously improving on credit risk classification accuracy improvement in comparison with one hidden layer evaluation model. Therefore, in order to simplify evaluation model and improve model training efficiency, this paper adopts three layer BP network structure to construct credit risk evaluation model. Hidden layer nodes selection not only largely affects the established evaluation model performance but it is also the direct reason to appear over-fitting in training.

It is complex to choose hidden layer nodes. There is not a scientific and common determining method in theory and it has only been determined according to some experience principles. Its fundamentally determining principle is try to implement the compact structure as much as possible and implement hidden layer nodes as little as possible under satisfying accuracy precondition. The research shows that hidden layer nodes are not only related to nodes at input and output layer but it also needs complexity and activation function of needing-to-solve questions and sample data characteristics. In order to obtain hidden layer node number of rational evaluation model, this paper determines common scope of hidden layer nodes based on following principle. Then, within this scope, expansion method is applied to determine reasonable hidden layer node number through repeated experiment.

1) Better number of hidden layer nodes is from 50% to 70% of the sum between input nodes and output nodes.

2) Hidden layer node number must be smaller than N-1. (N refers to training sample number).

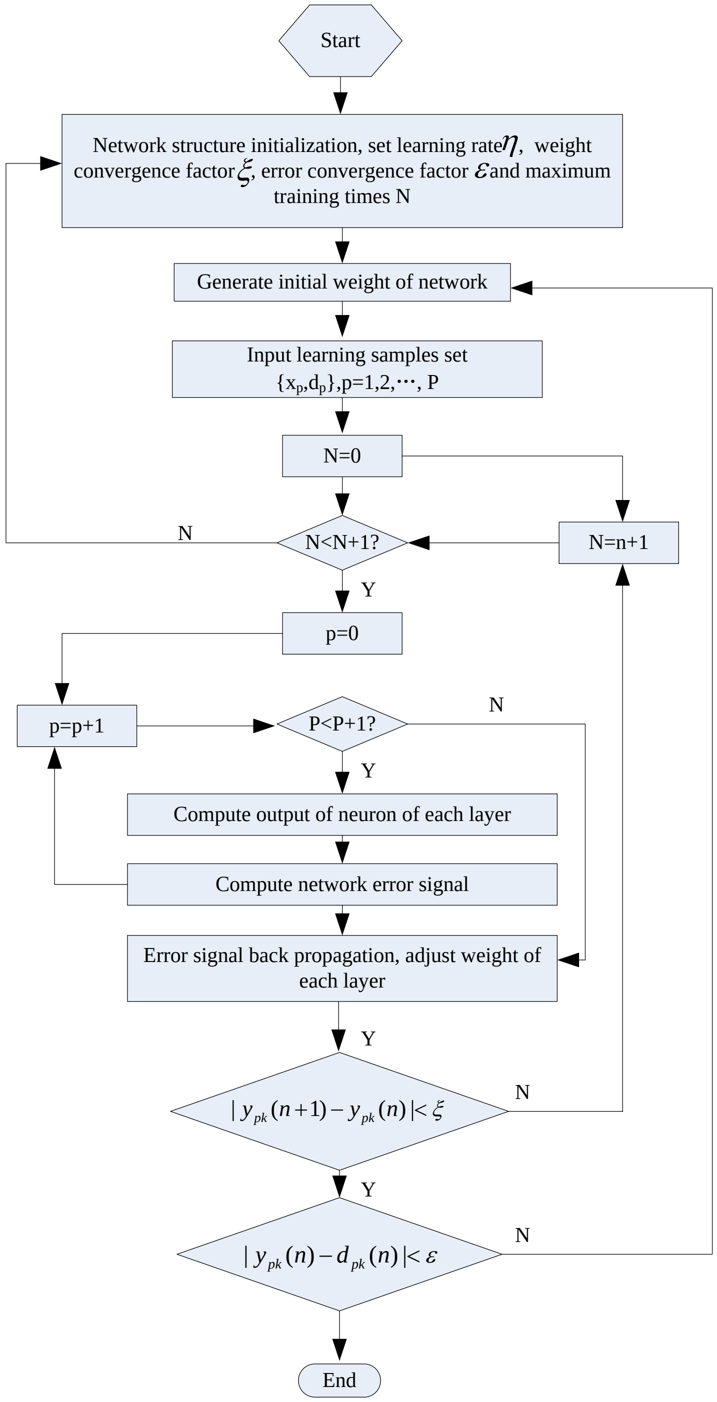

Figure 1. BP Learning Algorithm Flow

Some common parameters in BP learning algorithm are learning rate h, dynamic quantum a, convergence error critical value E, etc. These parameters are essential to training speed influence. In order to guarantee algorithm convergence, learning rate h must be smaller than one upper limit with usual $0 < \alpha < 1$. In order to improve learning speed of evaluation model, larger h should be adopted. However, if h is too large, it may lead to ascillation near stable point even algorithm non-convergence. The purpose to increase momentum term is to avoid that model training falls in shallower local minimal point. Its value should be related to weight correction value in theory but it usually gets constant with value scope $0 < \eta < 1$. It is usually larger than learning rate and determined according to practical situation. For specific network structure model and learning sample, there usually has the optimal learning rate h and momentum factor a0. In practical application, it needs repeated experiment, checks, verifies different h and a value and determines learning rate h and momentum factor a_0 of credit risk evaluation model appropriate for application environment in this paper.

Before evaluate the model training, we should determine error critical value E according to practical situation and the selection of error critical value E is completely based on evaluating convergence speed of model and learning accuracy of specific sample. When E is small, classification accuracy of evaluating model is high but convergence speed is slow and training times increase. If E value is large, it is opposite. Therefore, selection of value must be balanced between classification accuracy and training efficiency. This paper calculates error critical value of credit risk evaluation model based on table 2. To take AA credit rating as an example, during calculating output error of model, AA values 0.85. However, when evaluating model output is transformed into credit rate classification, value mapping in [0.8, 0.9) is AA. Model output error is calculated by means of the maximal value difference, we get:

$$E^p = \frac{1}{2}\sum_{k=1}^{l}(d_k^p - y_k^p)^2 = 0.00125 \quad (2)$$

3.2 Empirical Research of Credit Risk Evaluation

Established BP neural network model in this chapter aims to credit risk evaluation of enterprise customer in commercial bank. Therefore, neural element node in input layer corresponds 13 indicator variable in index system. In this paper, they respectively refer to return rate of net assets (node 1), the return rate of total assets (node 2), inventory turnover (node 3), accounts receivable turnover (node 4), total asset turnover (node 5), main business growth rate (node 8), asset-liability ratio (node 9), current ratio (node 10), quick ratio (node 11), cash flow debt ratio (node 12), and total ratio of cash flow (node 13). In practical operation, they do not need compile and only default input unit during data inputting unit.

Towards the above established credit risk evaluation model by networked structure and learning parameter, two evaluating model training of evaluation modes need to be performed. Due to initial casualty of connection weight, this paper performs multi-time model initialization and training towards each mode and gets the optimal accuracy model in test as the final model. Figure 5 refers to the final evaluating model training curve of two modes.

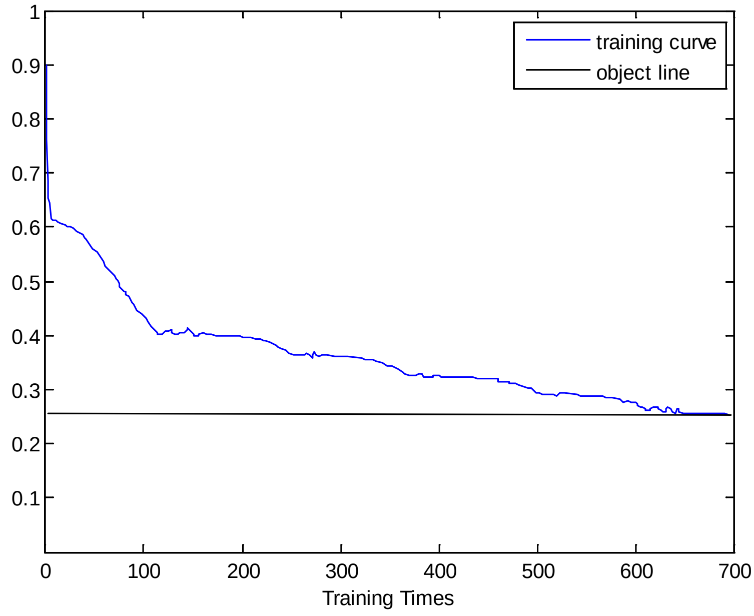

Figure 2. Training Graph of 0102 Evaluation Model

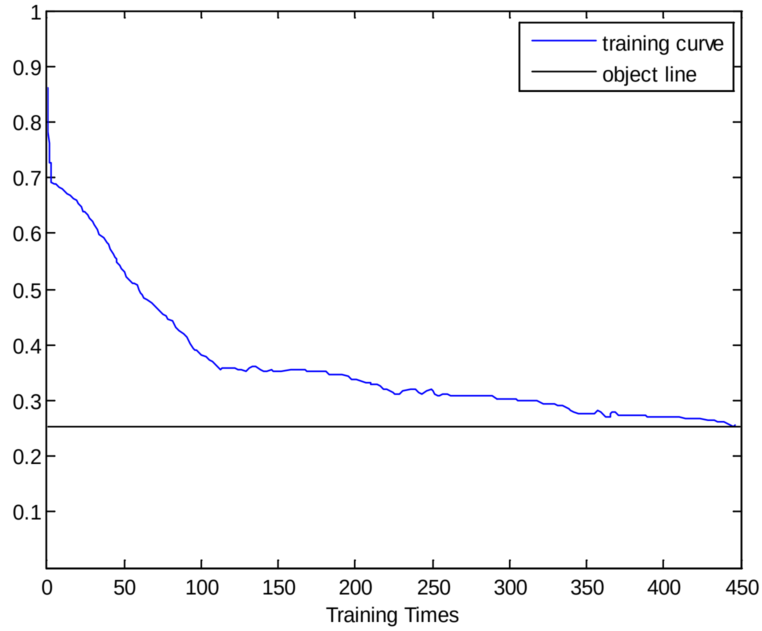

Figure 3. Training Graph of 0101 Evaluation Model

Two modes belong to predetermined learning step number range and they all reached the appointed error accuracy. Then, test sample is used to perform verification and case analysis of evaluating models. There are 7 test samples in 0101 evaluating mode and there are 9 test samples in 0102 evaluating model. Table 4 is test samples of two modes and table 5 refers to simulating test results of two modes.

Table 4. Qualitative Data of 0101 Evaluation Model Test Samples

	0101	0102	0103	0104	0105	0106	0107	0108	0109	0110	result
1	0.2	0.4	0.4	0.2	0.6	1	1	1	0.5	0	A

2	1	0.2	0.4	0.4	0.2	1	1	0	1	1	AA
3	0	0.4	0.8	0.2	0.2	0.5	1	0.5	0	0.5	B
4	0.6	0.5	04	0.2	0.6	1	1	0	0.5	0.5	BB
5	0.6	0.5	0.8	0	1	0.5	1	1	0.5	0.8	BBB
6	0.2	0.5	0.4	0.2	1	1	1	1	0.5	0	A
7	0.6	0.5	0.4	0.2	1	0.5	1	0	0.5	0.5	BBB

Table 5. Qualitative Data of 0102 Evaluation Model Test Samples

	0101	0102	0103	0104	0105	0106	0107	0108	0109	0110	result
1	0.6	0.4	0.4	0.4	0.6	1	1	0	1	1	A
2	1	0.4	0.8	0.4	1	1	1	1	0.5	0.6	AA
3	1	0.2	0.8	0.4	0.6	1	1	1	0.5	1	AA
4	0.6	0.4	0.8	0.8	1	1	1	1	1	1	AAA
5	0.6	0.4	0.8	0.4	0.6	1	1	1	0.5	0	B
6	0.6	1	0.4	0.4	1	1	1	0.5	0.5	0.8	BB
7	0.2	0.8	0.8	0	1	1	1	0.5	0.5	0	BBB
8	1	0.8	0.8	0.4	0.2	0.5	1	0	0.5	1	AA
9	0.6	0.4	0.8	04	1	1	1	0.75	1	0.75	A

Based on above tables, for 0101 evaluating mode, there are 5 cases are correctly classifying for evaluating model from 7 evaluating cases. Its classifying accuracy rate reaches 71.4%. Towards 0102 mode, there are 6 cases are correctly classifying for evaluating model in 9 evaluating cases. Its classifying accuracy rate is 66.7%.

In practice, during evaluating bank credit risk, accurate classification of credit rating is not the receivable lowest baseline for commercial banks. From bank perspective consideration, if low credit rating customer is evaluated as high credit rating customer, bank will afford extremely high risk. However, if high credit rating customer is evaluated as low credit rating customer, affordable risk of bank decreases.

Thus, towards commercial banks, accurate credit risk evaluation results and the decreased evaluating results of credit risk are all the expected credit risk evaluation results for banks. Based on above perspectives to analyze these tables, for 0101 evaluating mode, although number 5 sample case commits classification error, the result is to decrease customers' credit rating and it is an acceptable result. Towards 0102 mode, sample cases in number 2, 4 and 7 belong to classification error but these three cases are decreasing customers' credit rating. In this way, for 0101 evaluating mode, bank can accept 6 evaluating result cases and evaluating acceptable rate is 85.7%. However, towards 0102 evaluating mode, there are 9 cases to accept its evaluating result and evaluating acceptable rate is 100%.

Table 6. Simulation Results of Evaluation Model

NO.	Sample NO.	Credit rating	Expected output	Actual output	Classification result
0101	1	A	0.75	0.73	A
	2	AA	0.85	0.90	AA
	3	B	0.45	0.46	BB
	4	BB	0.55	0.52	BB
	5	BBB	0.65	0.59	BB
	6	A	0.75	0.76	A
	7	BBB	0.65	0.63	BBB
0102	1	A	0.75	0.76	A
	2	AA	0.85	0.78	A
	3	AA	0.85	0.88	AA
	4	AAA	0.95	0.82	AA

	5	B	0.45	0.44	B
	6	BB	0.55	0.59	BB
	7	BBB	0.65	0.48	B
	8	AA	0.85	0.80	AA
	0	A	0.75	0.70	A

4. Conclusions

This paper mainly analyzes the principle definition of big data, analyzes classical mining method of big data mining in machine learning field and discusses its application in financial credit safety risk pre-warning field. Then, this paper adopts BP neural network model and constructs credit risk evaluation model in commercial bank. Through learning historical credit data, it adjusts connecting weight of evaluating model and determines mapping relationship between credit risk evaluating index and credit rating so that evaluating model is capable to classify new credit data. By credit risk evaluation of this model, it weakens the determined human factors of index weight and improves evaluating scientific nature and evaluating results accuracy. After various evaluating mode verification, the designed evaluating model accuracy rate in this paper has been more than 65% while receivable rate of evaluating results has been more than 85%. Compared to 50% accuracy rate of traditional credit rating method, credit risk evaluation accuracy rate in neural network technology in this paper has been improved more than 10%. Based on above results, neural network in big data analysis-based financial credit risk evaluation model has higher accuracy and it can be taken as an important tool for loan decision in our national commercial banks.

Acknowledgements

This work is Supported by Social Science Funds of Hunan Province, Research on public security and the risk alarm Target System in the era of big data(14YBB051).

References

1. Fang Kuangnan, Huang Hong (2011) Variable selection for credit risk model using data mining technique. *Journal of Computers*, 6(9), p.p.1868-1874.
2. Li Jianping, Li Gang, Sun Dongxia (2012) Evolution strategy based adaptive Lq penalty support vector machines with Gauss kernel for credit risk analysis. *Applied Soft Computing Journal*, 12(8), p.p.2675-2682.
3. Moges Helen-Tadesse, Dejaeger Karel, Lemahieu Wilfried (2013) A multidimensional analysis of data quality for credit risk management: New insights and challenges. *Information and Management*, 50(1), p.p.43-58.
4. Piramuthu, Selwyn (2006) Expert Systems with Applications. *Intelligent Information Systems for Financial Engineering*, 30(3), p.p.489-497.
5. Ribeiro Bernardete, Chen Ning (2014) Hashing for financial credit risk analysis. *Lecture Notes in Computer Science*, 8835, p.p.395-403.
6. Sekerka Bohuslaw, Knaifl Oldrich, Beran Hynek (1995) Application of AI techniques in credit risk assessment. *Neural Network World*, 5(6), p.p.1045-1048.
7. Twala Bhekisipho (2013) Impact of noise on credit risk prediction: Does data quality really matter. *Intelligent Data Analysis*, 17(6), p.p.1115-1134.
8. Yang Jian, Zhou Yinggang (2013) Credit risk spillovers among financial institutions around the global credit crisis: Firm-level evidence. *Management Science*, 59(10), p.p.343-2359.
9. Yoon Jong Sik, Kwon Young S. (2010) A practical approach to bankruptcy prediction for small businesses: Substituting the unavailable financial data for credit card sales information. *Expert Systems with Applications*, 37(5), p.p.3624-3629.