

Integrated Learning-Based SME Credit Rating

Limei Wang, Zhe Wang, Yu Hu, Tian Bai

College of Computer Science and Technology, Jilin University, Changchun, China
Email: wanglm11@mails.jlu.edu.cn, wz2000@jlu.edu.cn, huyu11@mails.jlu.edu.cn, baitian@jlu.edu.cn

Received 28 February 2014; revised 3 April 2014; accepted 12 April 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

SME Credit rating index system becomes a significant research topic in recent years. So many researches have focused on this topic. However, the existing researches are only focused on one aspect of the SME Credit Rating problem. In order to resolve this problem, in this paper, we use the idea of ensemble learning, which integrated several basic machine learning algorithms to improve the learning result. Through further amendments, we build a set of SME corporate credit evaluation models which have higher forecast accuracy and stronger anti-jamming capability. Finally, we prove the effectiveness of our model through carrying out a set of experiments.

Keywords

SME; Ensemble Learning; Credit Rating

1. Introduction

SME, as an emerging research area, has attracted so many researchers in recent years [1] [2]. It also plays an increasingly important role for the development in China's socialist market economy.

However, the development process of SMEs suffered from many problems, especially the financing problem [3]. For example, SME financing costs are generally higher than bank's lending rates. And SME lending source is less than 20% of the whole society's lending rates. The most fundamental way to solve the above problems is to evaluate SME credit conditions. It will reduce the risk of financial intermediaries and the corresponding transaction costs. Through these measures, we can make capital markets more stable. On the other hand, capital can flow to SME initiatives.

Since the study of risk assessment started late relatively, and China currently does not have a specific SME credit evaluation model system. Therefore, in this paper, based on the previous works, we build a set of targeted credit evaluation models according to the characteristics of SME. We will rate the SME credit level by the me-

thod of ensemble learning. Our purpose is to response credit conditions of SME more comprehensive, accuracy and scientifically.

The rest of the paper is organized as follow: in Section 2, we briefly review some related works of SME credit rating. We study the SME credit evaluation model in Section 3. The proposed model is introduced in Section 4. In Section 5, we introduce the experimental result. We conclude our work in Section 6.

2. Background

In this section, we introduce some background knowledge and related works of our research.

2.1. Review the History of Credit Rating

Credit rating is a social service agency, which provides enterprise credit information to financial institutions. It also provides funding sources for expanding reproduction. Credit rating was originally produced in the early 20th century in United States by John Moody. And later, the credit rating was extended to a variety of financial products and a variety of evaluation objects, and it was also extended to other countries including Europe, Japan and Germany. In industry circle, the credit rating and firm size are not the same, so the credit rating model and content are also very different. Foreign credit rating can be roughly divided into three stages: Expert judgment stage, Credit scoring stage and Comprehensive evaluation stage. The credit rating method is the mainly qualitative analysis method in Expert judgment stage. In the stage of credit scoring, the existing methods are combine quantitative and qualitative analysis. And the key problem is the selection of weighting and feature vector. In the stage of Comprehensive Evaluation, the existing approaches are mainly based on quantitative analysis and supplemented on qualitative analysis. Now the rating method of this model is applied widely at home and abroad. Relatively speaking, our research in this area is relatively late. Affected by the planned economy, the development of SMEs in China has not been taken seriously. Until 1980, along with China's economic restructuring, it ushered in the spring of SME. Academic scholars are increasingly concerned about the linkages between SMEs credit conditions and economic development.

2.2. Related Works

In the context of enterprise credit evaluation index system designing, Yuanliang Liu *et al.* [4] added Supply Chain Finance to credit evaluation index system. This approach enhances the accuracy of the corporate credit rating. In [5], Shengxiang Bao *et al.* proposed several ways to improve SME financing capacity, such as minimize the short-term loans or increase their short-term liquidity. When researching the credit rating system of SME, Fang Gao *et al.* put the innovation capacity of enterprises, growth and development, product and market conditions into the construction of the index system [6]. In [7], Hong Ji *et al.* divided the growth stage of high-tech SMEs into entrepreneurship, growth, maturity, recession. According to the characteristics of the different growth stages of the high-tech SMEs, the authors suggest the composition factors of technological innovation capability for SME to improve it.

In designing the specific credit evaluation model and methods, Chinese scholars have made a lot of efforts. Chunfeng Wang *et al.* [8] started relatively early in the research aspects of credit evaluation model. They discussed the commercial bank credit risk assessment, and introduced the neural network technology into assessment model. The model has better classification performance and robustness than other assessment methods compared to traditional credit rating methods Fuzzy support vector machine method is applied firstly to corporate credit evaluation by Chunlin Fu *et al.* in [9]. This method was well adapted to the characteristic of ambiguity in credit assessment of the capital market. Because of that credit status of some corporate cannot be classified into absolute good or bad, and the corporate credit changes all the time, Yongqiao Wang *et al.* [10] applied SVM method to credit assessment. And they further improved the traditional support vector machine method and the method of fuzzy support vector. This method also solved the overlapping problem in business enterprise credit status. In [11], Jiexin Lin *et al.* analyzed the credit status of listing Corporation in China. They introduced Bayes discriminant method into the credit evaluation model. And they built classification model for two class according to the general covariance condition. In [12], Minjie Yu *et al.* combined the Decision Tree method and Genetic algorithm for credit rating. In [13], Sulin Pang *et al.* applied the BP neural network algorithm to 80 loan enterprises in state-owned banks in China. They divided the credit status of enterprises into two types through

the comparison and analysis of financial data, business status, history and other information.

However, the above researches are mainly used to solve the problem of credit evaluation of SMEs, and few of them are specialized on small and medium-sized enterprises. Small and medium-sized enterprises, as the mainstay and the new force of technological innovation field in China, need more attention. It is obviously that, completely copy the evaluation index system and evaluation model cannot meet the needs of small and medium-sized enterprises credit evaluation. Therefore, it is necessary to do in-depth study specifically for small and medium-sized enterprises, and providing new impetus for the benign development of the capital market.

3. The Construction of Credit Evaluation Index System of High Tech SMEs

In this research, we summarized the SME credit index factors and thought that indicators can be roughly divided into the growth of enterprises, innovation ability, the quality of enterprise and financial status (including debt paying ability, operating ability and profit ability). In this paper, financial index is suitable for most enterprises and essentials. This is also true in the SME credit evaluation. It is a quantitative index that reflecting the enterprise credit. We can query the data through the accounting statements of enterprises. And the data can be obtained easily. Then, the innovation ability and development ability of SME is the core in the whole life cycle of the enterprises, which related to the survival and death of enterprise. So these two points must be incorporated into the index factors. The index system includes the above factors as the basic study model in this paper. Based on the index system from Shengxiang Bao *et al.* [5] about the indicators of high-tech SME credit evaluation, we deleted some strong correlated coefficients. The constructed index system is showed in **Figure 1**.

Target layer	Criterion layer	Index layer	Index attribute
Credit evaluation index system of high tech SMEs	Innovation	Input intensity of R&D	Quantitative
		The proportion of technical persons The proportion of sales revenue of new products	
	Growth ability	The growth rate of net assets	Quantitative
		The growth rate of net profit	
		Operating profit growth rate	
	Future development	The prospect of the industry	Quantitative
		Industrial policy	
	The quality of enterprise	The quality of managers	Quantitative
		The quality of staff	
		Enterprise culture and strategy	
	Profitability	earning rate of the net assets	Quantitative
		Sales profit ratio	
		pe (price/earnings ratio)	
		liquidity ratio	
		asset-liability ratio	
	Debt paying ability	Ratio of liabilities and tangible assets	Quantitative
		cash flow ratio	
		Ratio of cash and total assets	
		interests covers	
		gross debt-to-Net cash ratio	
Ratio of Non-sexual raise cash flow and Current liabilities			
Operation ability	Account receivable turnover rate	Quantitative	
	Inventory turnover rate		
	Total assets turnover		
	Current assets turnover		
	Fixed assets turnover		
	Percentage of cash flows from operation activities in sales revenue		

Figure 1. Credit evaluation index system of high tech SMEs.

4. Study on the SME Credit Evaluation Model

In this paper, we select Ensemble learning method [14] for SME credit evaluation. Ensemble learning was pioneered by Hanson and Salamon. In [14], the authors found that combined model arranged from kinds of neural network model according to certain rules can significantly improve forecast accuracy in classifying and predicting. The Boosting algorithm is the most primitive ensemble learning algorithm. The algorithm has good generalization ability. Inspired by this thought, a large number of scholars have conducted research on the integrated learning research. Later scholars have improved the ensemble learning method that changes from the simply combination of multiple identical learning machines, which are used for solving a plurality of problem, to combination of multiple similar learning machines, which are used for classifying and multiple problems. The reason of ensemble better is obviously: if the two learning machines are the same, the credit evaluation result can only be one result, which is both right or both wrong. Obviously the test effect of single learning machine and multiple learning machines are the same. This combination has no meaning. The prediction results may have two different results when using two learning machines, which are similar. Finally model can determine what kind of classification is correct and the most close to the fact through determining the result again. The study of this paper is about classification. In this paper, three kinds of classification learning algorithms are selected as the basic algorithm for the ensemble learning model. And we use the ensemble model for the credit evaluation of high-tech SMEs. The combined model includes BP-Neural Networks, SVM, and Decision Tree. We brief introduce these three algorithms in the following.

The neural network method has a strong advantage in researching on complex problems [14]. It is a simulation and simplification of the biological neural network. The neural network consists of a large number of neurons and prominent. The network structure is showed in [Figure 2](#).

The interaction between neurons can realize the memory and storage of information. Neural network simulated by computer can achieve this function like biological neural network. On the other hand, the BP neural network system is a highly nonlinear mapping system. By using this classification method, the data does not need to be processed dimensionally. This technique simplifies the learning process. Neural network learning process can be divided into two stages: The first stage is the training and learning stage. In this stage, we need to set the network initial weights and bias. Then put the data into the model. The model will continue to adjust the weights and bias of the model until convergence is reached. The second stage is the prediction stage. This stage is a test of the model which formed in the first stage. The output data is obtained in this stage. The structure of neural network is variety. There is a schematic diagram of three layer network structure feedback neural network model showed in [Figure 3](#).

The Neural network model requires a large number of learning samples. On the other hand, we use the gradient descent method in adjusting the weight value and bias value. So the results of training will be trapped into local minima. In order to solve the above problems, we introduced the support vector learning method [14]. The theory and methods of SVM are introduced as follows:

The support vector machine method is a kind of popular statistical learning theory method. It is similar to the structural risk minimization principle. This method can solve problems of getting stuck in a local minimum, over learning or large sample demands etc. This paper makes up for the deficiency of the neural network by introducing SVM. The principle of support vector machine is shown in [Figure 4](#).

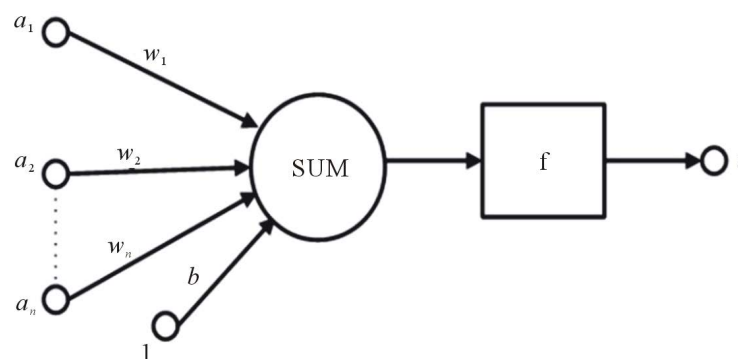


Figure 2. Neuron map from baidu.

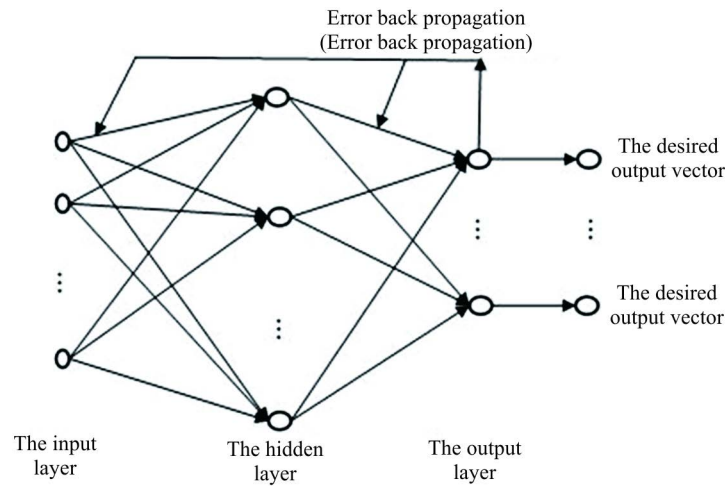


Figure 3. Three layer BP neural network structure diagram.

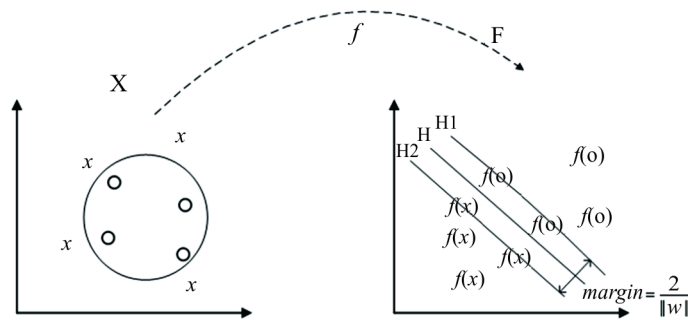


Figure 4. SVM nonlinear mapping.

SVM maps the data to a high dimensional space that the samples can be divided in that space. Then we can establish the optimal hyper plane in high-dimensional space. And in that space, the data has maximum margin and the minimum number of error classification that is the minimum structure risk. Assume that the training sample is $T = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$. Among them, $x_i \in X = R^n$; $y_i \in \{-1, +1\}$. The problem can be expressed as

$$(\omega \cdot x_i) + b] - 1 \geq 0, i = 1, 2, \dots, n \tag{1}$$

Adding constraints to the expressions, the problem can be converted into

$$Q(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + C \left[\sum_{i=1}^n \xi_i \right] \tag{2}$$

Among them, C is the penalty factor of error data, ξ_i is the number of error samples. Support vector machine will face with a complex problem in mapping the low dimensional data into high dimension. The process is not easy to realize. On the other hand, if the problem is converted into a dual problem we can use the theory of Mercer [14] to simplify the problem. So the decision function of SVM can be transformed into the following form:

$$f(x) = \text{sgn} \left(\sum_{i=1}^n a_i^* y_i K(x_i, x_j) + b^* \right) \tag{3}$$

The neural network method is sensitive to noise data. Therefore, we introduce the decision tree in order to overcome the disadvantages of the neural network. The decision tree method [14] simulates the process of classification using a tree structure. The method can approximate discrete values effectively and deal with the noise, which make up for the deficiency of the neural network. Learning process of decision tree can be divided into

two stages. The first stage is the tree construction and tree pruning, and the second stage is t sample prediction. The first stage is more complex and spends more time. The second stage is relatively simple. The typical decision tree algorithms are ID3, C4.5 and Cart etc. Construction process of decision tree is shown in **Figure 5**.

In this paper, we use three common algorithms, which is complementary advantages to classify and prediction. This combined model can improve the prediction accuracy. Inspired by the ensemble learning theory, the author constructed an integrated learning model through arranging the three kinds of machine learning parallel in the field of machine learning. The model is shown in **Figure 6**.

5. Empirical Analysis

In this section, we test our method using the real SME data.

5.1. Sample Acquiring and Processing

In this paper, we use the data which provided by the intermediary. In order to classify and predict, we selected 140 SMEs data. Test samples are selected randomly. Because accounting system of SME is not perfect. Missing data is a common phenomenon. In order to reduce impact of attributes missing on the final evaluation results and ensure the authenticity of final evaluation results, we have filtered the data before inputting data into the model. The data set is divided into four sub tasks in the stage of empirical analysis.

Task 1: Takes the first 10 data as the training sample, the latter 133 data as the test sample;

Task 2: Takes the first 30 data as the training sample, the latter 113 data as the test sample;

Task 3: Takes the first 50 data as the training sample, the latter 93 data as the test sample;

Task 4: Takes the first 70 data as the training sample, the latter 73 data as the test sample.

5.2. SME Credit Evaluation

The classical machine learning algorithms can be implemented using MATLAB software. Therefore, we use MATLAB to build the ensemble learning model. We compared three classical machine learning algorithms. Accurate rate was selected to evaluate the algorithms. Classification accuracy of four models as shown in **Table 1**:

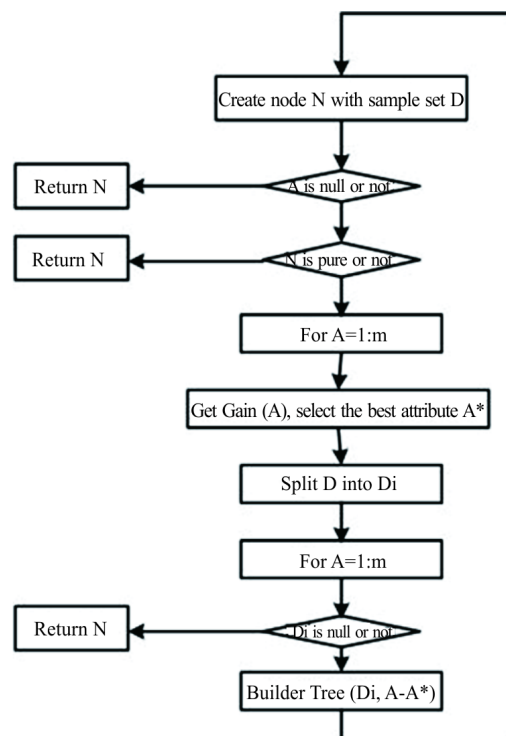


Figure 5. Construction process of decision tree.

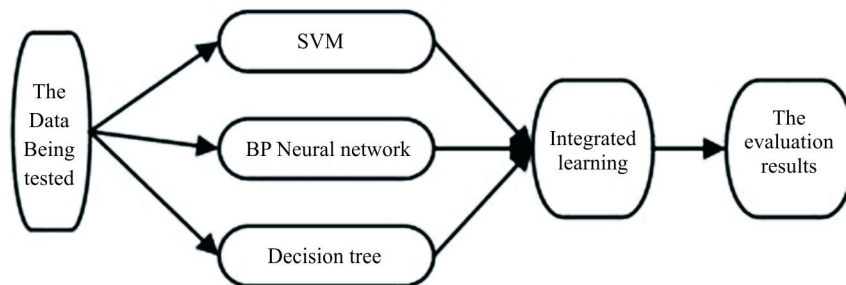


Figure 6. Integrated learning model.

Table 1. Comparison of prediction accuracy.

	Proposed Method	SVM	BP-Neural Network	Decision Tree
Task 1	0.6500	0.6250	0.5083	0.7833
Task 2	0.7600	0.7600	0.7700	0.7200
Task 3	0.7500	0.7250	0.5750	0.6500
Task 4	0.8000	0.7000	0.7333	0.7667

From the table, we can see that the ensemble learning method we proposed has higher forecasting precision.

6. Conclusions and Future Work

In this paper, we have investigated the credit rating problem by using ensemble learning theory. Three basic machine learning algorithms are selected to construct the ensemble model. The experimental result showed the effectiveness of our method.

However, when constructing models, we need to consider investment preference of the bank. On the other hand, corporate credit rating also needs to be divided into several grades. Therefore, the model needs further improvement to cover the above situation.

Acknowledgements

This work is supported by National Natural Science Foundation of China (NSFC) under Grant No. 61373051; the National Science and Technology Pillar Program (Grant No. 2013BAH07F05), Jilin Province Science and Technology Development Program (Grant No. 20111020); Project of Science and Technology Innovation Platform of Computing and Software Science (985 engineering), and the Key Laboratory for Symbolic Computation and Knowledge Engineering, Ministry of Education, China.

References

- [1] Li, X.F. and Li, J. (2003) Data Mining and Knowledge Discovery. Higher Education Press, Beijing.
- [2] Lin, X.-J. (2009) The Design of Medical Image Classification Based on Ensemble Learning. *Computer Knowledge and Technology*, **5**, 9515-9517.
- [3] Lu, Y. (2012) Research on the Credit Rating of SME. Anhui University, Hefei.
- [4] Liu, Y.L. and Liu, S.L. (2013) Research on Credit Risk Identification of Small Enterprises Based on Supply Chain Finance. *Financial Market*.
- [5] Bao, S.X. and Yin, Y.F. (2009) Credit Evaluation and Empirical Analysis of SME. *Science & Technology Progress and Polic*, **26**, 143-148.
- [6] Gao, F. (2010) Research on the Credit Rating System of Chinese High Tech SMEs. Hebei University, Baoding.
- [7] Ji, H., Cao, Z.C. and Xie, W. (2012) Research Based on the Growth Theory of Small and Medium Enterprises in Technological Innovation Ability. Scientific Research.
- [8] Wang, C.F. and Wang, H.H. (2010) Summary of Enterprise Credit Rating Index System Research. *Financial Forum*, **8**.
- [9] Lin, C.-F. and Wang, S.-D. (2002) Fuzzy Support Vector Machines. *IEEE Transactions on Neural Networks*, **13**, 464-

471.

- [10] Wang, Y.Q., Wang, S.Y. and Lai, K.K. (2005) A New Fuzzy Support Vector Machine to Evaluate Credit Risk. *IEEE Transactions on Fuzzy System*, **13**, 820-831.
- [11] Lin, J.X., Luo, W.Q. and Pang, S.L. (2005) Bayes Discriminant Credit Evaluation Model and Its Application. *Statistics and Decision*.
- [12] Ye, Z.X. and Yu, M.J. (2006) Genetic Algorithm and Classification Tree Method Based Credit Rating Method. *Journal of Systems Engineering*.
- [13] Pang, S.-L., Li, R.-Z. and Xu, J.-M. (2005) Application of BP Algorithm in Credit Risk Analysis. *Control Theory and Applications*.
- [14] Han, J.W. and Kamber, M. (2005) *Data Mining Concepts and Techniques*. (Fan, M. and Meng, X.F., Trans.), Mechanical Industry Press, Beijing.