2016 International Conference on Intelligent Transportation, Big Data & Smart City

Commercial Bank Credit Risk Assessment Method based on Improved SVM

Wang Zhen, Sun Wenjuan
(Harbin University of Science and Technology Rongcheng Campus;Shandong Rongcheng 264300,China)
(Harbin University of Science and Technology Rongcheng Campus;Shandong Rongcheng 264300,China)
402210224@qq.com

Abstract—**Credit risk evaluation is the basic work of commercial bank's credit risk management, which goal is to analyze the credit risk of the bank. Support vector machines ensemble has been proposed to improve classification performance recently. However, currently used fusion strategies do not evaluate the importance degree of the output of individual component SVM classifier when combining the component predictions to the final decision. A SVM ensemble method based on fuzzy integral is presented in this paper to deal with this problem. This method aggregates the outputs of separate component SVMs which is given different weights by means of fuzzy integral. The experiment results show that the result of fuzzy integral support vector machines is satisfactory.**

Keywords-Commercial Bank Credit Risk Evaluation Model; SVM; Fuzzy Integral

I. Introduction

As the main intermediary of financial transaction, the commercial bank is the weatherglass of states economic condition. Commercial bank also plays a important role in reducing economical risk and the unstable factor, guarantying the national economy healthily. The commercial bank itself undertakes various types risk in the operation, including credit risk, interest rate risk, fluid risk, management risk, capital risk and policy risk and so on. Credit risk holds the special important status in each kind of risk, which is the most primary factor which causes bank goes bankrupt[1, 2].

Since the 1980s, the technology of artificial intelligence such as expert system, neural network and decision support systems are introduced in the credit risk assessment, which overcome the disadvantage of strong hypothesis of statistical method and static reflection risk. Especially the neural network, it is a self-organizing, adaptive and self-learning characteristics, not only has a nonlinear mapping and generalization ability, but also has stronger robustness and higher prediction precision.

But the neural network has its own shortcomings to overcome. The first is that the network structure is difficult to determine. The second is that it is easy to fall into local extremum in the training, and training efficiency is not high. Aiming at the shortcomings of the neural network, many scholars have done a lot of beneficial attempt. So the low training efficiency of neural network due to local infinitesimal and slow convergence speed is improved in a certain extent. In addition, introducing fuzzy logic method, partial derivative of error function can help to increase the explaining ability of the neural network. Genetic algorithm is also used to study credit risk evaluation problem, and long encoding mechanism of genetic programming method is difficult to make better effect[3-5]. Measuring discreteness of the standard is the main cause of credit risk assessment using classification assessment model. The classification assessment model of the traditional credit risk play a more active role in a certain historical period and environment, but its inefficiency already cannot satisfy the needs of the increasingly complex credit risk decision[6]. And the key to realize effective shift of evaluation pattern is to establish a more scientific and effective credit risk measurement standard, then credit risk assessment prediction model is constructed. Based on the status of domestic credit evaluation that has less accumulation of data, statistical methods effect is bad, and the neural network is also difficult in effective learning, we introduce support vector machine learning algorithm[7,8] based on the theory of the small sample learning, and use it in commercial bank credit risk assessment, which has achieved good results.

In the next section, fuzzy integral SVM is set up to evaluate commercial bank credit risk assessment. In section 3, experiments are done to test the effect of proposed scheme. In the end, some remarks are given.

II. MODEL BUILDING BASED ON THE FUZZY INTEGRAL SVM

Support vector machine (SVM) integration process is shown in figure 1, and the experiment process is divided into the following steps:

Step 1: Generate the training set m from the source data set based on bagging method, and for each training set to train a support vector machine.

Step 2: Given the output probability of each support vector machine.

Step 3: Set fuzzy density $\{g(\mathrm{SVM}_i), k=1,2,\cdots,m\}$ and the importance of each support vector machine based on the execution performance on each training set .

Step 4: When given the newest test sample, we can get the SVM output probability.

Step 5: For $\omega_k, k=1,2,\cdots,m$ and C to compute the fuzzy integral e_k , and get the final decision making.

a) bagging individual generate method

The bootstrap sampling is the basic of the bagging method. In this method, the support vector classifier training set generate by random sample of several examples from the original training set. Training set size is usually same to the original training set and allows repeated selection. By this way, some examples from the original training sets may appear multiple times in the new training, and some may do not appear once. Bagging method by selecting training set increases the difference degree what improves the generalization ability.

978-1-5090-6061-0/17 $31.00 © 2017 IEEE
DOI 10.1109/ICITBS.2016.75

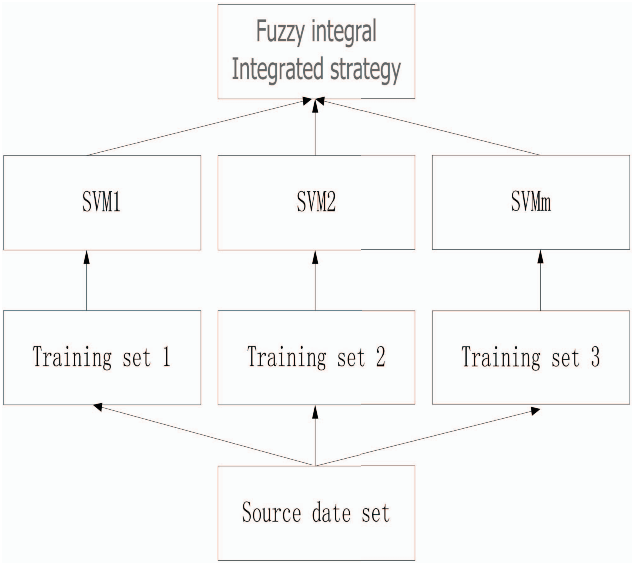

Figure 1. Support vector machine (SVM) integration process

b) SVMs integration based on fuzzy integral

The fuzzy integral theory will be introduced following. $X=\{x_1,x_2,\cdots,x_n\}$ is a limited set. And set function satisfies the following functions, g called fuzzy measurement.

$$g(\vartheta)=0$$

$$g(X)=1$$

$$g(A)\le g(B),\text{if}(A<B)$$

g_λ is Sugeno measurement.

$$g(A\cup B)=g(A)+g(B)+\lambda g(A)g(B)$$

Set $h:X\to[0,1]$ is the fuzzy set of X, and it be represented by $A_i\{x_1,x_2,\cdots,x_n\}$

$$g(A_1)=g(\{x_1\})=g_1$$

$$g(A_1)=g(\{x_1\})=g_1$$

$$g(A_i)=g_i+g(A_{i-1})+\lambda g_i g(A_{i-1})$$

$$,g_i=g(\{x_i\}),1<i\le n$$

$$\lambda+1=\prod_{i=1}^{n}(1+\lambda g_i)$$

We can get the value of λ from the above functions.

Assumes $h(x_1)\ge h(x_2)\ge\cdots\ge h(x_n)$, and we can get the fuzzy integral e:

$$e=\max_{i=1}^{n}[\min(h(x_i),g(A_i))]$$

If given the fuzzy density, we can compute the e, and then get the final decision making. The fuzzy density sets by the expert or generate from the training sets.

III. THE EMPIRICAL ANALYSIS

a) The establishment of the index system

In commercial bank the main credit risk can be divided into loans business risk, commercial bank risk, macroeconomic risk and other risk factors. To comprehensively consider the influence factors of credit risk, based on the principles of index selection, the Treasury statistical evaluation of enterprise performance evaluation index system, the industrial and commercial bank of China enterprise credit evaluation index system and domestic and foreign literatures related data, final we choose 16 indicators as a commercial bank credit risk assessment: Sales/total assets, total assets turnover ratio, current asset turnover, fixed asset turnover, inventory turnover, accounts receivable turnover ratio, liquidity ratio, working capital/total assets, quick ratio, speed ratio, asset-liability ratio, cost, profit margin, sales net interest rate, return on assets, return on equity, loans. These metrics can comprehensively reflect the enterprise's profit ability, debt paying ability, operation efficiency and profitability, etc [7, 8].

b) Sample acquired

In this paper, the data from the Harbin branch of industrial and commercial bank of China. When collecting data, we pay much attention on the sample's own industry characteristics, different industry enterprise management environment. Acquisition of data retrieval conditions are as follows:

(1) The sample range of industries: manufacturing;

(2) Types: short term loan a year (and within a year);

(3) The loan origination date: January 1, 1998 to January 31.

(4) The loan balance deadline: on August 13, 2001.

(5) The loan amount: loan actual amount;

(6) Loan balance: the loss of loan balance at August 13, 2001;

(7) Loan form: present form;

(8) Enterprise full name and code: identify the enterprise unique identification code;

(9) Report date and report: on December 31, 1997, the enterprise balance sheet and income statement;

(10) Collecting, sorting and get 176 samples, involving loan more than 5 billion yuan.

c) Sample data process

First of all to do samples robustness processing and get the 157 sample data final by twice and three times the standard deviation test for abnormal data. According to whether the loan default, commercial bank loans will be divided into two categories: normal and damaged. For data from this 157 companies has 80 companies financial situation is good, therefore the risk of bank loans is smaller, these enterprises in this paper, referred to "the performance enterprise"; and the rest of the 77 companies financial situation is bad, default probability, this paper referred to as "default". Two types of sample data is too close what can satisfy the requirements for the study of integrated SVMs.

Table 1. The factor load matrix

index	1	2	3	4
Sales/total assets	0.996			
total assets turnover ratio	0.996			
current asset turnover	0.953			
fixed asset turnover	0.843			
accounts receivable turnover ratio	0.819			
inventory turnover	0.755			
liquidity ratio		0.899		
working capital/total assets		0.891		
quick ratio		0.782		
speed ratio		0.750		
asset-liability ratio		-0.566		0.353
profit margin			0.879	
sales net interest rate			0.843	
return on assets			0.823	
return on equity			0.694	
loans				

From the table 1 and the table 2, it is not difficult to get the economic meaning of each factor, and each factor can be summed up in operating capacity factor, debt paying ability factor, profit ability factor and loan factor.

Table 2. The factor score matrix

index	1	2	3	4
Sales/total assets	0.195	0.052	-0.033	0.032
total assets turnover ratio	0.195	0.052	-0.033	0.032
current asset turnover	0.192	0.022	-0.033	-0.009
fixed asset turnover	0.170	0.076	-0.029	0.064
accounts receivable turnover ratio	0.165	0.025	-0.045	0.104
inventory turnover	0.152	-0.037	-0.014	-0.227
liquidity ratio	-0.032	0.276	0.009	-0.053
working capital/total assets	-0.040	0.273	0.014	-0.136
quick ratio	-0.038	0.240	0.004	0.249
speed ratio	-0.024	0.273	0.007	0.230
asset-liability ratio	0.012	0.230	0.004	0.338
profit margin	0.048	-0.174	0.325	0.037
sales net interest rate	0.050	0.005	0.312	0.045
return on assets	0.009	0.009	0.304	-0.117
return on equity	0.007	-0.022	0.25	-0.065
loans	-0.004	-0.028	0.037	0.786

d) Experiment result analysis

In the training sample, we obtain 30 samples through repeated random samples to replace the original training samples for bagging. Training 3 support vector machine independently from the 10 training sets by bagging method. 0 represents performance enterprise; 1 represents the default enterprise.

Table 3. Training set classification accuracy based on fuzzy integral Support vector machine ensemble

Observation	Forecast		
	0	1	accuracy rate(%)
0	27	6	81.8182
1	5	18	78.2609
total	32	24	80.3571

Table 4. Testing set classification accuracy based on fuzzy integral Support vector machine ensemble

Observation	Forecast		
	0	1	accuracy rate(%)
0	47	10	82.4561
1	12	42	77.7778
total	59	52	80.1169

Table 5. Accuracy of classification method

Algorithm	Accuracy rate(%)
single SVM	77.0655
SVM integration of the most vote	78.9214

principles	
fuzzy integral SVM ensemble	80.6735
single fuzzy neural network	77.0579
fuzzy neural network ensemble	75.5583

From above table we can conclude that the fuzzy integral SVM have been achieved best performance, it may be that the fuzzy integral in multiple classifier decision fusion when considering the classification results of each classifier and each classifier decision because of the importance of the final decision.

IV. Conclusions

Credit risk management has always been one of the most serious problems of the banks and other financial institutions. Domestic credit evaluation has less accumulation of data, statistical methods effect is bad, and the neural network is also difficult to get better effect in learning. We investigate support vector machine learning algorithm based on the theory of the small sample learning, and use it in commercial bank credit risk assessment. It not only combines each classification results, but also considers the relative importance of different SVMs classifier. According to the result of child support vector machine (SVM) classification, fuzzy integral method is used to give each child support vector machine different weights. The results show that this method has better performance.

References

[1] Wong K P. “On the determinants of bank interest margins under credit and interest rate risks”, Journal of Banking & Finance, 1997, 21(2): 251-271.

[2] Jiménez G, Ongena S, Peydró J L, et al. “Hazardous Times for Monetary Policy: What Do Twenty - Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk - Taking”, Econometrica, 2014, 82(2): 463-505.

[3] Hordern S V M, Wright J E, Umpleby A M, et al. “Comparison of the effects on glucose and lipid metabolism of equipotent doses of insulin detemir and NPH insulin with a 16-h euglycaemic clamp”, Diabetologia, 2005, 48(3): 420-426.

[4] Campbell W M, Sturim D E, Reynolds D A, et al. “SVM based speaker verification using a GMM supervector kernel and NAP variability compensation”, Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. IEEE, 2006, 1: I-I.

[5] Fine S, Scheinberg K. “Efficient SVM training using low-rank kernel representations”, The Journal of Machine Learning Research, 2002, 2: 243-264.

[6] Baltensperger E, Milde H. “PREDICTABILITY OF RESERVE DEMAND, INFORMATION COSTS, AND PORTFOLIO BEHAVIOR OF COMMERICAL BANKS”, The Journal of finance, 1976, 31(3): 835-843.

[7] Wu T F, Lin C J, Weng R C. “Probability Estimate for Multi-class Classification by Pairwise Coupling Journal of Machine Learning Reserch”, 2004,(5):975-1005

[8] R. Noori, A.R. Karbassi, A. Mogjiaddamnia, et aL “Assessment of Input Variables Determination on the SVM Model Techniques for Monthly Stream Flow Prediction”, Journal of Hydrology. 2011,401:177-189