

The Risk Assessment Model of Small and Medium-Sized Enterprises of Science and Technology Based On Boosting

X.F. Lei

Department of Computer Science and Technology
Jilin University
Changchun, China

Y.T. Li

Changchun WHY-E Science and Technology Co., Ltd
China

Z. Wang

Department of Computer Science and Technology
Jilin University
Changchun, China

Abstract—Break the financing-difficulties has always been the key to the development of small and medium-sized enterprises of science and technology (SMES), so the construction of a targeted credit evaluation model is of great significance to provide data support of financing. The paper proposed a credit evaluation model based on Boosting, which not only improves the accuracy but also avoids over fitting phenomenon. We also optimized the evaluation index by embodying the characteristics of SMES, which achieves the dynamic evaluation and meet the requirements of the more targeted model simultaneously, in order to ensure accuracy.

Keywords—SMES; financing difficulties; credit evaluation; boosting

I INTRODUCTION

SMES have become an important force in the market economy in terms of technological innovation, but the financing difficulties hinder its development seriously. The basic way to solve this problem is to build an evaluation model with high accuracy to reduce the impact of asymmetric information between enterprises and financial institutions, based on the specific credit rating index for SMES. The risk assessment of SMES is a classification problem in essence. We proposed a credit assessment model of SMES based on Boosting. Boosting is a way to improve the accuracy of weak learning algorithm. In 1990, Schapire proposed the original Boosting algorithm, which proved the equivalence between weak learning and the strong learning [1]. It makes a weak learning algorithm that is better than random guesses slightly can be promoted to the strong one. The algorithm achieves sample subset mainly through the operation for the sample set, then trains sample subset to generate base classifiers with weak classification algorithm. Boosting algorithm fuses these base classifiers by giving a low weight for good training results, while a high weight for bad one. It makes the poor training can get more attention in the next round of training, until produce the final classifier. In 1995, Freund and Schapire

improved Boosting algorithm and put forward the Adaboost algorithm [2], which does not require any prior knowledge about weak learning and more easily applied to practical problems. If no special instructions, the Boosting algorithm in this paper refers to the Adaboost algorithm.

II THE CREDIT RATING INDEX SYSTEM OF SMES

There was not specific credit rating index system for SMES inland. This paper optimized the predecessor's summary by fusing the elements of science and technology, more targeted. For qualitative indicators, we selected enterprise quality, enterprise technological, history credit record, and the development prospect, four aspects, 12 in total [3]. This paper takes enterprise technological aspect as characteristic qualitative indicators of SMES; for quantitative indexes, we selected profit ability, debt paying ability, operation ability, development ability, etc. [3], 19 in total, and optimized the innovation ability. So, there are 31 indexes in total in the index system.

III THE CONSTRUCTION OF THE RISK ASSESSMENT MODEL FOR SMES BASED ON BOOSTING

For SMES, accurate credit evaluation is an effective way to solve the financing difficulties, but the high accuracy of classification model is very difficult. This part we set up the risk assessment model based on Boosting for SMES. The thought of model building: first, carries on the data acquisition and data preprocessing, then improve the weak classifier by performing Boosting. The algorithm itself is achieved by changing the data distribution, every time it according to whether each sample is correct of the training set and the overall classification accuracy last time, to determine the weights of each sample set the new data that was modified to next classifier, like this, the trained classifier will be fused as a final decision classifier. According to the forecast function to

judge whether the enterprise has a good reputation, to give a decision support. Specific process is as follows (fig.1.):

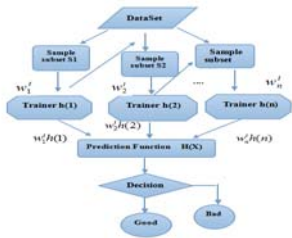


FIGURE I. THE PROCESS OF BOOSTING MODEL.

Main steps that mainly involved in the flow chart of Boosting as follows [4]:

Set the training sample set S contains N samples; the number of iterations is n;

$h(n)$ is the base classifier trained each iteration; w_n^i is the weight of samples i trained at one time; δ^n is the adjustment of the weighting factor.

(1)The initialization, give each training sample equal weight of 1/N, and use the selected weak classification method to train sample subset to generate the base classifier $h(n)$.

(2)According to the error rate of training sample in n rounds to construct a new data distribution, including the wrong sample of previous training and new sample, and update the weight of the (n+1) round by the weighting factor.

(3)According to the $h(1), \dots, h(n)$, get the prediction function, the final Weighted classifier is integrated by the base classifier. For the base classifiers, accuracy does not high individually, while the results of their joint are very high. Thus it improved the recognition rate of the weak classification algorithm.

IV THE EMPIRICAL ANALYSIS

This experiment tool is the R programming language, version for 3.1.0.

A. Data Acquisition

This study used the data of 44companies from 2010 to 2014 of sina finance network to ensure the validity and reliability. There are753 records, take a random sample of 551 as the training data, and 202 as test data. The "good" or "bad" enterprise distribution of sample set is roughly 2.5:1, which meet the demands of modeling.

B. Modeling by Boosting Algorithm

This experiment used R package" adabag" and function "Boosting" was used in the training data. Then return to a model to predict function. Function model is:Boosting(formula, data, boos=TRUE, mfinal=100, coeflearn='Breiman', control). Accuracy, as well as the importance of evaluation index ranking in the experiments will be influenced by mfinal, which is the number of iterations, specific in Table1 (The top 5 of the importance only was listed in the table).

TABLE I. INDEX ANALYSIS OF MFINAL ON THE IMPORTANCE OF THE ERROR RATE.

mfinal	NO.1(%)	NO.2(%)	NO.3(%)	NO.4(%)	NO.5(%)	Error
20	R&d staff ratio18.3	Rate of net profit growth 8.7	Rate of inventory turnover6.7	Rate of total assets Growth 6.5	Cash Ratio 6.3	0.129 3532
50	R&d staff ratio 14.8	Rate of total assets growth 7.5	Proportion of R&D 7.1	Rate of Inventory turnover .9	Cash ratio 6.6	0.119 403
70	R&d staff ratio 14.7	Rate of net profit growth 9.5	Rate of accounts receivable turnover7.3	R&d Staff Ratio 6.69	Cash Ratio 6.62	0.104 4776
100	R&d staff Ratio 16.7	Rate of accounts receivable turnover 6.8	Rate of accounts receivable turnover 6.8	Cash ratio 6.25	R&d staff ratio 5.4	0.084 57711

From the table above, first: the importance of evaluation index ranking is different, but R&D ratio is the highest overall,which proves that the index of model was optimized, namely infused the "technology innovation". Secend:the precision float is different as mfinal, but won't appear over fitting phenomenon, which is also a characteristics of Boosting.The details as shown in fig. 2.We default choose mfinal = 100, error is 0.08457711,namely the accuracy is 91.6%.

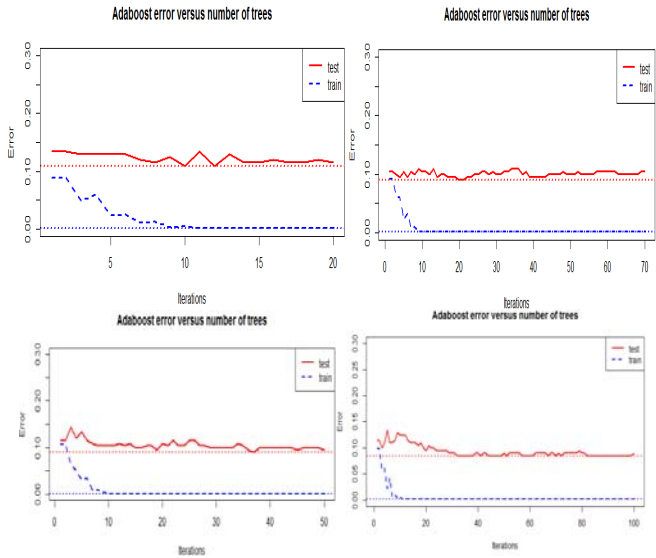


FIGURE II. INFLUENCE OF MFINAL ON ERROR (ACCURACY).

V CONCLUSION

1. In this paper, risk evaluation index for SMES is optimized by emphasizing R&D, from the results, we can see the enterprise innovation ability of research and development is one of the biggest in the importance of the evaluation index ranking. The assets and profits of such evaluation indexes is no longer the first choice, it is a disguised form to support the independent innovation of SMES.

2. In this paper, Boosting method has high accuracy and less complexity, it also can avoid over fitting phenomenon. In conclusion, the risk assessment model based on Boosting for SMES is feasible and effective. But we also found that there still have some room for improvement in accuracy. And on the other hand, taking the enterprise risk assessment only as simple binary classification problems is rough. The next step of work is the focus of the further improvement on the basis of this model, introducing Boosting to multi-classification of risk assessment, to achieve a more accurate assessment and to minimize the risk of financial institutions.

ACKNOWLEDGEMENTS

This work was financially supported by the National Science and Technology Support Projects-Jilin Financial Institutions Application Demonstration Project (ProjectNo.2013BAH07F05).

REFERENCES

- [1] Xuehua Shen, Zhihua Zhou, Jianxin Wu and Zhaoqian Chen. Boosting and Bagging review[J]. Computer Engineering and Application. (12): P31-33,2000.
- [2] Ying Cao,Qiguang Miao, Giachen Liu, Lin Gao. The research progress and prospect on Adaboost algorithm[J]. Journal of Automation, 39 (6) : 745-758, 2013.
- [3] Haitao Huo, The method of credit risk index system and evaluation for high-tech small and medium-sized enterprise [J]. Journal of Beijing institute of technology (social science edition), (1): P60-65, 2012.
- [4] Xia Li, Liyun He and Chao Liu, Boosting algorithm and its application in traditional Chinese medicine and health data classification [J]. China health statistics. (02): 158-162, 2008.