



Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates

Jie Sun^{a,*}, Jie Lang^b, Hamido Fujita^c, Hui Li^d

^a Business School, Tianjin University of Finance and Economics, Tianjin, PR China

^b School of Economics and Management, Zhejiang Normal University, Jinhua, Zhejiang, PR China

^c Faculty of Software and Information Science, Iwate Prefectural University, Iwate, Japan

^d College of Tourism and Service Management, Nankai University, Tianjin, PR China

ARTICLE INFO

Article history:

Received 5 May 2017

Revised 31 August 2017

Accepted 6 October 2017

Available online 9 October 2017

Keywords:

Enterprise credit evaluation

Decision tree ensemble

SMOTE

Bagging

Differentiated sampling rates

Imbalanced classification

ABSTRACT

Enterprise credit evaluation model is an important tool for bank and enterprise risk management, but how to construct an effective decision tree (DT) ensemble model for imbalanced enterprise credit evaluation is seldom studied. This paper proposes a new DT ensemble model for imbalanced enterprise credit evaluation based on the synthetic minority over-sampling technique (SMOTE) and the Bagging ensemble learning algorithm with differentiated sampling rates (DSR), which is named as DTE-SBD (Decision Tree Ensemble based on SMOTE, Bagging and DSR). In different times of iteration for base DT classifier training, new positive (high risky) samples are produced to different degrees by SMOTE with DSR, and different numbers of negative (low risky) samples are drawn with replacement by Bagging with DSR. However, in the same time of iteration with certain sampling rate, the training positive samples including the original and the new are of the same number as the drawn training negative samples, and they are combined to train a DT base classifier. Therefore, DTE-SBD can not only dispose the class imbalance problem of enterprise credit evaluation, but also increase the diversity of base classifiers for DT ensemble. Empirical experiment is carried out for 100 times with the financial data of 552 Chinese listed companies, and the performance of imbalanced enterprise credit evaluation is compared among the six models of pure DT, over-sampling DT, over-under-sampling DT, SMOTE DT, Bagging DT, and DTE-SBD. The experimental results indicate that DTE-SBD significantly outperforms the other five models and is effective for imbalanced enterprise credit evaluation.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

During the periods of the Asian financial crisis in 1997 and the global financial crisis in 2008, many enterprises fell into bankruptcy because of serious financial distress. For the economic society, a wave of enterprise bankruptcy often threatens the safety of bank system. Therefore, credit analysis and risk management become more and more important for financial market [26]. As a risk early warning tool for banks, enterprise credit evaluation is important for preventing banks from

* Corresponding author.

E-mail address: sj@tjufe.edu.cn (J. Sun).

suffering great loss due to the financial deterioration of their enterprise debtors. It helps banks make more scientific credit decisions and decrease the loss of non-performing loans. To some extent, any small improvement for the credit evaluation model would produce great profits for banks [15]. In addition, enterprise credit evaluation model can also help corporate inner managers be early conscious of the deterioration of financial condition and take measures to prevent it as early as possible. Therefore, how to improve the performance of credit evaluation model is always a very important research topic.

However, although many statistical and artificial intelligence models have been studied for enterprise credit evaluation, most researches are based on balanced data set and imbalanced enterprise credit evaluation modeling still needs to be further studied. In the real economic world, the high risky enterprises are much fewer than the normal enterprises, and enterprise credit evaluation belongs to the imbalanced classification problems. If the research data set is collected by 1:1 matching the high risky enterprises and the normal enterprises, the classification knowledge hidden in the enterprise credit evaluation model may not comprehensively reflect the real boundary between the two classes. Besides, it may also bring an overestimation of the model performance reported in the research. On the other hand, the classification algorithm like DT usually performs badly when it is trained directly on an imbalanced data set. The contribution of this research is to propose an effective DT ensemble (DTE) model based on SMOTE (synthetic minority over-sampling technique) and Bagging with DSR (differentiated sampling rates), which is abbreviated as DTE-SBD, so as to improve the performance of imbalanced enterprise credit evaluation.

2. Literature review

2.1. Enterprise credit evaluation approaches other than artificial intelligence

Enterprise credit evaluation approaches other than artificial intelligence are mainly categorized into three types. The first type is based on capital market theory or information economics [4,5,16,30,38], and this kind of models have rigorous theoretical foundation, but they do not output credit evaluation results automatically based on data driven mechanism.

The second type is based on data-driven statistical approaches, which construct enterprise credit evaluation models based on sample companies' historical data of multiple variables. The statistical approaches such as linear discriminant analysis [2,13] and logistic regression [22,34] are early applied to enterprise credit evaluation, and they have the advantage of simple model structure. However, their application is restricted by the statistical assumptions. For example, the linear discriminant analysis requires normal distribution of independent variables, and the logistic regression requires no colinearity among independent variables.

The third type is based on evaluation theory. The most classical approach of this type is the 5Cs expert scoring method, which is widely applied in the real-world practice of banks. More recently, Hu and Chen propose a PROMETHEE-based classification method using concordance and discordance relations, which is based on outranking relation theory, and apply it to predicting the bankruptcy of enterprises [18].

2.2. Enterprise credit evaluation approaches based on artificial intelligence

With the development of business intelligence, enterprise credit evaluation models based on artificial intelligence raise more and more concern in the recent three decades.

2.2.1. Enterprise credit evaluation approaches based on single classifier

Early application of artificial intelligence to enterprise credit evaluation is based on single classifier. Neural network is widely used for enterprise credit evaluation modeling [3,20,35]. Huang et al. find that enterprise credit evaluation models based on neural network usually outperform statistical models, and point out that the approach of neural network also has the problems such as difficulty of determining model structure, certain degree of working randomness, low efficiency of training, etc. [19]. Besides, Lee [23], Chen et al. [9], Li et al. [25] and Louzada et al. [27] apply support vector machine (SVM) to enterprise credit evaluation and empirically validate its effectiveness. Pang and Gong construct a C5.0 decision tree (DT) model for individual credit evaluation of banks by integrating C4.5 DT model with boosting technology and it obtains better accuracy than the primary DT model [29]. Chuang applies the hybrid case-based reasoning method for enterprise bankruptcy prediction and achieves enhanced performance [10]. It is obvious that various single classifier algorithms have been applied to enterprise credit evaluation and they usually can obtain better performance than statistical models. However, each single classifier algorithm has its own advantage and disadvantage, and one single classifier model may not contain the knowledge comprehensive enough for enterprise credit evaluation.

2.2.2. Enterprise credit evaluation approaches based on classifier ensemble

To overcome the weakness of single classifier, more and more studies focus on constructing ensemble models for enterprise credit evaluation, and find that ensemble methods perform better than the base single classifier approaches [24,28,43]. Hsieh and Hung introduce the concept of class-wise classification as a preprocessing step in order to obtain an efficient ensemble classifier and it works better than a direct ensemble of classifiers without the preprocessing step [17]. Zhang et al. propose a vertical bagging DT model for enterprise credit evaluation and validate its effectiveness [41]. Wang et al. compare the three ensemble learning methods of Bagging, Boosting and Stacking for credit evaluation, and find that multiple-classifier

Download English Version:

<https://daneshyari.com/en/article/6857042>

Download Persian Version:

<https://daneshyari.com/article/6857042>

[Daneshyari.com](https://daneshyari.com)