Multi-Class Imbalanced Corporate Bond Default Risk Prediction Based on the OVO-SMOTE-Adaboost Ensemble Model

Jie SUN[1] and Jingmei ZHU

School of Accounting, Tianjin University of Finance and Economics, Tianjin, China

Abstract. Corporate bond default risk prediction is important for regulators, issuers and investors in the bond market. We propose a new approach for multi-class imbalanced corporate bond risk prediction based on the OVO-SMOTE-Adaboost ensemble model, which integrates the one-versus one (OVO) decomposition method, the synthetic minority over-sampling technique (SMOTE) and the Adaboost ensemble method. We categorize corporate bond default risk into three classes: very low default risk, relatively low default risk and high default risk, which is more scientific than the traditional two-class bond default risk, and carry out empirical experiments by respectively using DT, SVM, Logit and MDA as basic classifiers. Empirical results show that the prediction performance of the OVO-SMOTE-Adaboost (DT) model is overall better than the other three ensemble models such as OVO-SMOTE-Adaboost (SVM), OVO-SMOTE-Adaboost (Logit) and OVO-SMOTE-Adaboost (MDA). In addition, the OVO-SMOTE-Adaboost (DT) model greatly outperforms the OVO-SMOTE (DT) model, which is a single classifier model based on OVO and SMOTE without Adaboost. Therefore, the OVO-SMOTE-Adaboost (DT) model has satisfying performance of multi-class imbalanced corporate bond default risk prediction and is of great practical significance.

Keywords. Corporate bond default risk prediction, multi-class imbalanced classification, OVO-SMOTE-Adaboost ensemble model

1. Introduction

The bond market is an indispensable part of the financial market system, which not only provides a platform for fund-raisers and investors to carry out financing and investment activities, but also plays an important role in deepening the reform of financial structure, promoting the marketization of interest rates and improving the government's function of macroeconomic regulation. In recent years, China's bond market has developed rapidly, with the total volume of bond issuance and the number of bond issuers both increasing, and the market size is growing, becoming the second largest market in the world. In 2014, China's bond market saw its first material default, then rigid payment was gradually broken in China's bond market. From 2014 to the present, the number of bond default cases has been increasing and the nature of bond default becomes more and

---

[1] Corresponding Author, School of Accounting, Tianjin University of Finance and Economics, Tianjin, China; E-mail: sunjiehit@gmail.com.

more complex and diverse, which not only harms the interests of investors, but also impedes the development of the bond market. Therefore, how to timely and effectively predict the default risk of corporate bonds is an important problem in the bond market. The Global Artificial Intelligence Technology Conference 2021 (GAITC2021), held in Hangzhou on June 5, aims to build a comprehensive platform that seamlessly links the global strength of artificial intelligence and presents intelligent technology in a three-dimensional manner, further promoting technological innovation, application innovation, achievement transformation and industrial chain collaboration in the field of artificial intelligence. In view of corporate bond default risk prediction, artificial intelligence technologies can extract more potentially valuable information from financial data than traditional statistical methods, and can predict the default risk of corporate bonds more effectively, which is of great significance to the healthy development of bond market.

2. Literature review

Prior literature has proposed many statistical or artificial intelligence models for predicting corporate financial distress or credit risk. Altman [1] used multiple discriminant analysis (MDA) to predict enterprise bankruptcy by introducing multiple financial ratios as variables, and constructed the Z-Score model, which was widely used for bond credit scoring. In recent years, many scholars have used artificial intelligence methods to achieve better prediction results. Li and Sun [2] found that the regression tree classification model shows better performance in financial distress prediction than MDA. Lee [3] introduced support vector machine (SVM) into enterprise credit rating prediction and used five-fold cross-validation method to determine the SVM model's best parameter values. They compared the performance of the SVM model with other models such as neural networks (NNs) and MDA, and the empirical results show that the SVM model has higher classification accuracy.

The application of single classifiers is easily affected by data structure. In order to solve the limitation of single classifier, some scholars have proposed multi-classifier ensemble methods, which refer to combining multiple single classifiers and finally integrating the output of multiple classifiers for prediction. Common ensemble methods include Bagging (Bootstrap aggregating), Boosting, Adaboost (adaptive Boosting), and so on. Sun et al. [4] proposed a financial distress prediction model based on weighted majority voting combination of multiple classifiers, which can greatly improve the accuracy and stability of financial distress prediction and outperform single classifiers. Tsai et al. [5] used the Bagging and Boosting algorithms to build ensemble models based on basic classifiers such as logistic regression, SVM, decision tree (DT), NNs, and so on. Li et al. [6] used NNs, SVM, random forest, Chi-squared automatic interaction detector and classification and regression tree (CART) to predict credit ratings, and the results showed that random forest has the highest accuracy. Sun et al. [7] proposed a time-weighted Adaboost-SVM ensemble model to predict dynamic financial distress, and empirical results show that its predictive performance is better than single SVM, batch-based ensemble with local weighted scheme, Adaboost-SVM and Timeboost-SVM. However, when the training data is highly class-imbalanced, both ensemble models and single classifier models usually have poor performance for predicting the minority class.

In order to solve the problem caused by imbalanced data, scholars have proposed different solutions. For example, Sun et al. [8] used SVM as the basic classifier, and

proposed a multi-classifier ensemble model. They divided the majority class into several subsets and combined each of them with the minority class to construct several class-balanced training datasets, and then used them to train multiple SVM classifiers for combination. The results show that the model can significantly improve the predictive ability for minority samples. Liu et al. [9] used a class-oriented feature selection method to solve the problem of multi-class imbalance. Sun et al. [10] proposed a DT ensemble model based on synthetic minority over-sampling technique (SMOTE) and Bagging with differentiated sampling rates, and their empirical results demonstrate the effectiveness of this method in credit assessment of enterprises in the circumstance of class imbalance.

In addition, most of the current literature simply divides corporate status into default and non-default and builds binary classification models to predict bond default risk, ignoring the diversity of corporate bond status. Hence, it is necessary to define the multi-class risk of corporate bond default and build multi-class bond default risk prediction models. However, multi-class prediction is more complex than the traditional two-class prediction, and we need to transform a multi-class classification problem into multiple two-class classification problems with a certain decomposition method. Common decomposition methods have one-versus-one (OVO), one-versus-all (OVA) and binary tree method. Li et al. [11] built a multi-class classifier through the OVO decomposition method. Zhou et al. [12] used the OVO and OVA decomposition methods for multi-class classification problems, and built different multi-class classification models based on three basic classifiers. Sun et al. [13] combined SVM with OVO, OVA, error correction output coding (ECOC) to build a multi-class financial distress prediction model, and empirical results show that OVO-SVM outperforms OVA-SVM and ECOC-SVM overall. However, the above studies mainly integrate decomposition methods with single classifiers to solve the multi-class classification problem, neglecting the advantage of ensemble classifiers. How to construct more effective ensemble models for multi-class corporate bond default risk prediction needs to be further investigated. This paper integrates the SMOTE oversampling method and the Adaboost ensemble method with the OVO decomposition method to build ensemble models for multi-class imbalanced corporate bond default risk prediction.

3. OVO-SMOTE-Adaboost Ensemble Model for Multi-Class Imbalanced Corporate Bond Default Risk Prediction

The framework of the OVO-SMOTE-Adaboost ensemble model for multi-class imbalanced corporate bond default risk prediction is shown in Figure 1, and the algorithm is described as follows.

(1) Decompose a multi-class sample into several two-class subsamples by OVO.

The initial training set *Train* contains three types of samples, the first with a very low risk of bond default, the second with a relatively low risk of bond default, and the third with a high risk of bond default. Based on the OVO decomposition method, the initial training set *Train* is decomposed into three two-class training subsets, $Train_1$ for the first class and the second class, $Train_2$ for the first class and the third class, $Train_3$ for the second class and the third class.

(2) Train the SMOTE-Adaboost ensemble model for two-class bond default risk prediction.

Use training sets *Train$_j$* (*j*=1, 2, 3) to train three two-class bond default risk prediction models based on the SMOTE oversampling method and the Adaboost ensemble method, respectively, as follows:

a) Initialize the sample weights, each with a weight of w_j^{1i}=1/n, and the initial weight distribution of the training set is represented as D_j^1=[$w_j^{11}, w_j^{12}, \ldots, w_j^{1n}$]=[$\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}$], where n is the sample number of the training set.

b) Iterate t=1...T

   i. Construct a training dataset *Train$_j^t$* by resampling from *Train$_j$* according to the sample weight distribution D_j^t;

   ii. Use the SMOTE oversampling method to balance the training dataset *Train$_j^t$* to get a balanced training dataset *Train-balance$_j^t$*;

   iii. Use the balanced training dataset *Train-balance$_j^t$* to train a weak classifier based on a certain classification algorithm (DT, SVM, Logit, and MDA will be used in our empirical study, respectively);

   iv. Test the weak classifier with *Train$_j$*, and calculate the error rate e_j^t=$P(h_j^t(x_i)) \neq y_i)$=$\sum_{i=1}^n w_j^{ti} I\left(h_j^t(x_i) \neq y_i\right)$, where y_i is the class label for the training sample and $h_j^t(x)$ is the basic classifier, and the following is the calculation rule for I.

$$I = \begin{cases} 1, & h_j^t(x_i) = y_i \\ 0, & h_j^t(x_i) \neq y_i \end{cases}$$

   vi. Calculate the voting weight α_j^t=$0.5\ln\frac{1-e_j^t}{e_j^t}$;

   vii. Update the sample weight $w_j^{(t+1)i} = w_j^{ti} \exp\left(-\alpha_j^t \mu_j^{ti}\right)/Z_j^t$, where Z_j^t is the normalization constant, $Z_j^t = 2\sqrt{e_j^t\left(1 - e_j^t\right)}$.

$$\mu_j^{ti} = \begin{cases} 1, & h_j^t(x_i) = y_i \\ -1, & h_j^t(x_i) \neq y_i \end{cases}$$

c) Constitute a strong ensemble classifier H_j through weighted voting combination of T weak classifiers obtained by iterative training.

(3) Combine the results of two-class bond default risk prediction to obtain the results of multi-class bond default risk prediction.

Input the feature values of a testing sample into the three strong ensemble classifiers for binary classification, $H_1(x)$, $H_2(x)$ and $H_3(x)$, respectively, and combine their results by the simple majority voting principle to finally output the multi-class prediction result $H(x)$.
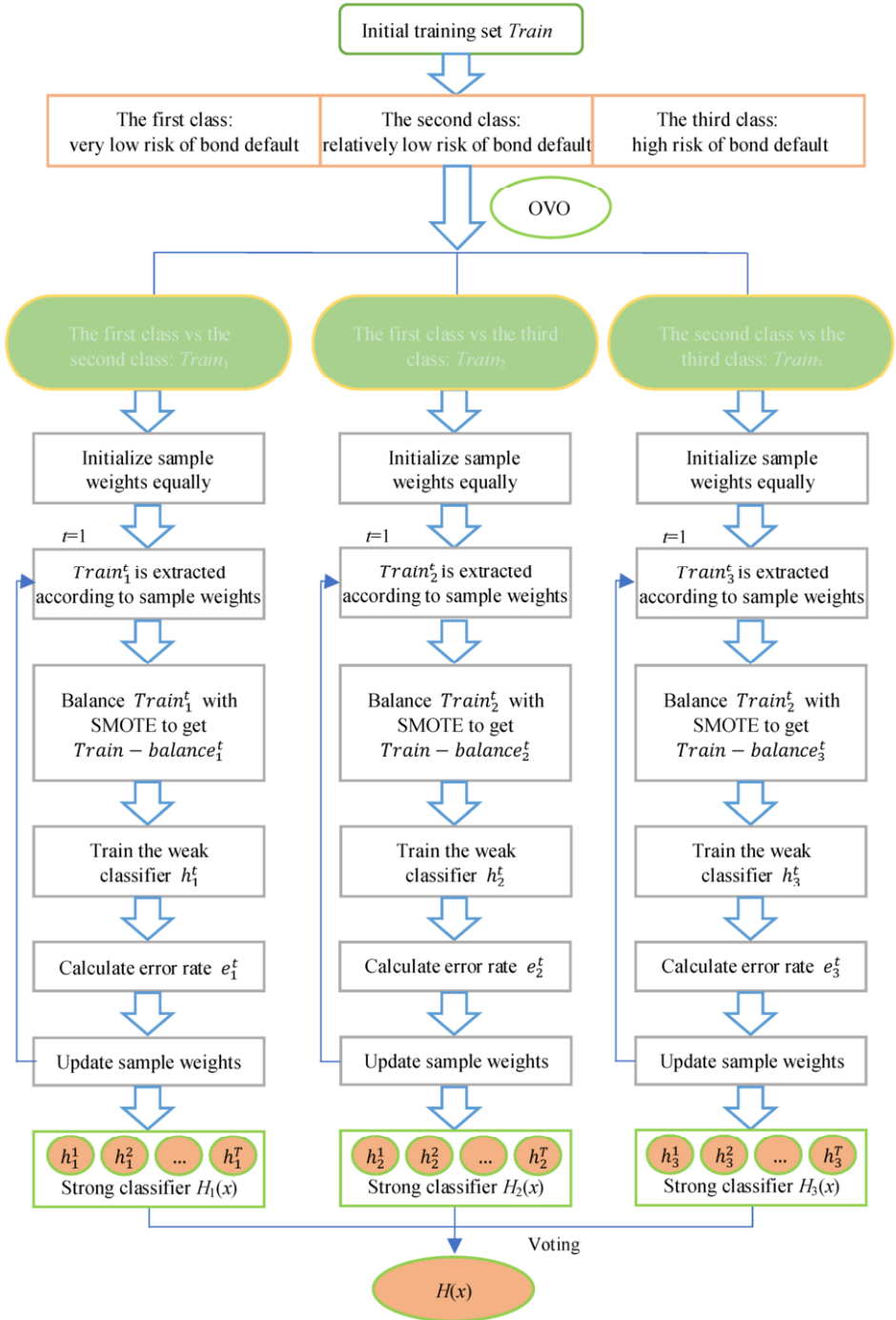
Figure 1. The framework of the OVO-SMOTE-Adaboost ensemble model.

4. Design of Empirical Research

4.1. Data Collection and Preprocessing

4.1.1. Data Collection

In 2014, China had its first default bond, breaking the rigid payment, so we collect data of enterprises that publicly issue bonds during 2014 to 2020 from the Wind database. Excluding the financial industry, we collect the credit ratings of bond issuers and their financial indicator data corresponding to the fiscal year that is two years earlier than the credit rating year. Then we delete the duplicate data of the companies issuing multiple bonds, the outliers and the data with missing values. We categorize the risk of bond default into three classes according to the credit ratings of the bond issuers: the first class is very low default risk, and the corresponding bond credit ratings are AAA and AA$^+$; the second class is relatively low default risk, and the corresponding bond credit ratings are AA, AA$^-$, A$^+$, A and A$^-$; the third class is high default risk, and the corresponding bond credit ratings are B, C and substantive default. The final sample consists of 3927 observations, among which 1534 belong to the first class, 2185 belong to the second class and 208 belong to the third class.

4.1.2. Data Preprocessing

In order to ensure the reliability of the results and improve the convergence speed and accuracy of the model, the original indicator data need to be normalized. The Min-Max normalization method is used to normalize the original data into a specific interval according to Eq. (1):

$$X' = \frac{X\text{-min}}{\text{max-min}} \tag{1}$$

where "max" is the maximum value of the sample data and "min" is the minimum value of the sample data. The normalization processing maps the data uniformly into the interval of [0,1].

4.2. Feature Selection

Initial features are composed of 26 financial indicators covering the five aspects of the operation capability, solvency, profitability, cash flow and development capability. We use the SPSS software to conduct multi-independent-sample Kruskal-Wallis test for the 26 financial indicators to verify mean differences among the three classes, and delete indicators with no significant mean differences among different classes. The correlation between each pair of indicators is analyzed, and the strong correlation indicators with a correlation coefficient greater than 0.5 is eliminated, resulting in 16 financial indicators as the model input variables, which were working capital turnover, accounts receivable turnover, fixed asset turnover, asset-liability ratio, current ratio, cash ratio, interest coverage, return on total assets, net profit margin on sales, ratio of operating profit to total operating revenue, ratio of net profit from operating activities to pre-tax profit, ratio of net cash flow from operating activities to operating revenue, proportion of net cash flows from operating activities, cash operating index, growth rate of operating profit and monetary capital growth rate.

4.3. Design of Empirical Comparison

For comparative analysis, we integrate the OVO decomposition method with the SMOTE oversampling method (OVO-SMOTE) to build single classifier models for multi-class imbalanced corporate bond default risk prediction, which are compared with the OVO-SMOTE-Adaboost ensemble models. We divide the preprocessed data set into the training set and the testing set in terms of the proportions of 80% and 20%. Then we use the training set to train the OVO-SMOTE single classifier models and the OVO-SMOTE-Adaboost ensemble models based on DT, SVM, MDA and Logit, respectively, and use the testing set to test their performance of multi-class bond default risk prediction. The training and testing process repeats 30 times, and we record 30 results and calculate the average for each model. The empirical research is implemented in the MATLAB 2016a software. We use the CART algorithm to build DT classifiers, and the minimum leaf tree is 50 after many experiments. The kernel function of SVM is set as the radial basis function (RBF), and the parameters c and g in the RBF SVM model are determined to be 32 and 2 respectively by grid optimization.

4.4. Model Evaluation Indicators

We use *Accuracy*, *Recall* (*R*), and *Macro-averaging Recall* (*Macro-R*) as indicators to evaluate model performance. For binary classification, *TP* refers to the number of positive samples that are correctly predicted as the positive; *FP* refers to the number of positive samples that are incorrectly predicted as the negative; *TN* refers to the number of negative samples that are correctly predicted as the negative; and *FN* refers to the number of negative samples that are incorrectly predicted as the positive. The indicators of *Accuracy* and *Recall* are calculated according to Eq. (2) and (3):

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{2}$$

$$R = \frac{TP}{TP+FN} \tag{3}$$

This study focuses on three-class corporate bond default risk prediction. Suppose N denotes the total number of testing samples, and $N1$, $N2$ and $N3$ are the testing sample numbers of the first, second and third class, respectively. $TC1$, $TC2$ and $TC3$ respectively represent the sample numbers of the first, second and third class that are correctly predicted as their true class. We should calculate the indicator of *Accuracy* as equation (4), the indicator of *Recall* for each class as Eq. (5)-(7), and the indicator of *Macro-R* as Eq. (8).

$$Accuracy = \frac{TC1+TC2+TC3}{N} \tag{4}$$

$$R1 = \frac{TC1}{N1} \tag{5}$$

$$R2 = \frac{TC2}{N2} \tag{6}$$

$$R3 = \frac{TC3}{N3} \tag{7}$$

$$Macro\_R = \frac{R1 + R2 + R3}{3} \tag{8}$$

where *R1* is the *Recall* for the first class, *R2* is the *Recall* for the second class, *R3* is the *Recall* for the third class.

5. Empirical Results and Analysis

5.1. Performance of Corporate Bond Default Risk Prediction based on OVO-SMOTE-Adaboost Ensemble Model

We repeat the whole experiment for 30 times to obtain 30 empirical results. In each experiment, we divide the initial data set into the training set and the testing set and use the training set to build the OVO-SMOTE-Adaboost ensemble models for corporate bond default risk prediction based on DT, SVM, Logit and MDA, respectively. We denote them as OVO-SMOTE-Adaboost (DT), OVO-SMOTE-Adaboost (SVM), OVO-SMOTE-Adaboost (Logit) and OVO-SMOTE-Adaboost (MDA), respectively, and use the testing set to test each model's performance. The average values of the model evaluation indicators obtained in 30 experiments by the OVO-SMOTE-Adaboost ensemble models are shown in Table 1.

As can be seen from Table 1, the OVO-SMOTE-Adaboost (DT) model has the highest *Accuracy* at 70.99%, showing overall better prediction performance than the other three models. According to the indicator of *Recall*, none of the above four models can achieve the highest value of *Recall* for all the three classes. For *R1*, the OVO-SMOTE-Adaboost (SVM) model has the highest value of 72.72%, followed by the OVO-SMOTE-Adaboost (DT) model with a value of 71.32%. For *R2*, the OVO-SMOTE-Adaboost (DT) model has the highest value of 70.75%, which is much higher than those of the other three models with a difference of 8.42%, 11.55% and 17.54%, respectively. For *R3*, the OVO-SMOTE-Adaboost (DT) model has the highest value of 71.14%, which is higher than those of the other three models with a difference of 5.21%, 1.95% and 0.57%, respectively. In terms of *Macro-R*, the OVO-SMOTE-Adaboost (DT) model is the highest at 71.07%. By comparing the results, the OVO-SMOTE-Adaboost (DT) model has optimal and stable performance for multi-class imbalanced corporate bond default risk prediction, and the OVO-SMOTE-Adaboost (SVM) model is overall better than the other two.

Table 1. The performance of the OVO-SMOTE-Adaboost ensemble models.

Ensemble model	*Accuracy*	*R1*	*R2*	*R3*	*Macro-R*
OVO-SMOTE-Adaboost (DT)	**70.99%**	71.32%	**70.75%**	**71.14%**	**71.07%**
OVO-SMOTE-Adaboost (SVM)	66.57%	**72.72%**	62.33%	65.93%	66.99%
OVO-SMOTE-Adaboost (Logit)	60.82%	62.03%	59.20%	69.19%	64.47%
OVO-SMOTE-Adaboost (MDA)	55.79%	57.48%	53.21%	70.57%	60.42%

5.2. Comparative Analysis between the Ensemble Models and the Single Classifier Models

For comparative analysis, we also train and test the single classifier models, i.e., OVO-SMOTE (DT), OVO-SMOTE (SVM), OVO-SMOTE (Logit) and OVO-SMOTE (MDA). The evaluation indicators of each model obtained in the 30 experiments are also recorded and averaged, and the results are compared with the empirical results of each ensemble model.
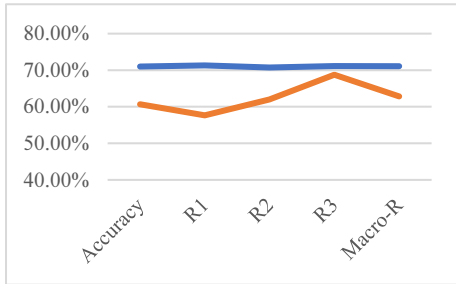
The results of OVO-SMOTE-Adaboost (DT) vs. OVO-SMOTE (DT) are shown in Figure 2. The OVO-SMOTE-Adaboost (DT) model outperforms the OVO-SMOTE (DT) model for each evaluation indicator, and the differences of *Accuracy*, *R1*, *R2*, *R3* and *Macro-R* between the two models are 10.33%, 13.65%, 8.76%, 2.36% and 8.26%, respectively. The empirical results of the two models show that the integration of the Adaboost ensemble method with the OVO decomposition method and the SMOTE oversampling method can greatly improve the performance of multi-class imbalanced corporate bond default risk prediction based on the DT classifier.

The results of OVO-SMOTE-Adaboost (SVM) vs. OVO-SMOTE (SVM) are shown in Figure 3. The OVO-SMOTE-Adaboost (SVM) model is 0.62% more accurate than the OVO-SMOTE (SVM) model, with 1.41% lower for *R1*, 2.02% higher for *R2*, 0.97% higher for *R3*, and 0.52% higher for *Macro-R*. The empirical results of the two models show that the integration of the Adaboost ensemble method with the OVO decomposition method and the SMOTE oversampling method slightly reduces the prediction ability for the first class, and slightly improves the prediction ability for the second and third class when SVM is used as the basic classifier. The two models have overall similar performance of multi-class imbalanced corporate bond default risk prediction.
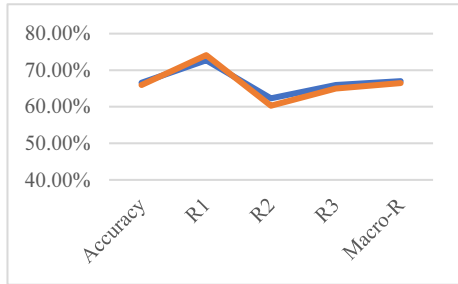
The results of OVO-SMOTE-Adaboost (Logit) vs. OVO-SMOTE (Logit) are shown in Figure 4. The two models show little difference of multi-class imbalanced corporate bond default risk prediction, with -0.02% for *Accuracy*, with 0.1% for *R1*, -0.08% for *R2*, -0.16% for *R3* and -0.05% for *Macro-R*. The empirical results of the two models show that the integration of the Adaboost ensemble method with the OVO decomposition method and the SMOTE oversampling method cannot enhance the performance of multi-class imbalanced corporate bond default risk prediction when Logit is used as the basic classifier.

The results of OVO-SMOTE-Adaboost (MDA) vs. OVO-SMOTE (MDA) are shown in Figure 5. The OVO-SMOTE-Adaboost (MDA) model was 1.72% higher than the OVO-SMOTE (MDA) model for *Accuracy*, 4.67% higher for *R1*, 0.26% lower for *R2*, 0.65% higher for *R3*, and 1.69% higher for *Macro-R*. The empirical results of the two models show that the integration of the Adaboost ensemble method with the OVO decomposition method and the SMOTE oversampling method can improve the overall

performance of multi-class imbalanced corporate bond default risk prediction when MDA is used as the basic classifier.
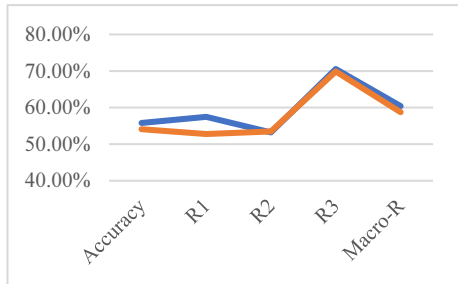

Figure 2. OVO-SMOTE-Adaboost (DT) vs. OVO-SMOTE (DT).

Figure 3. OVO-SMOTE-Adaboost (SVM) vs. OVO-SMOTE (SVM).

Figure 4. OVO-SMOTE-Adaboost (Logit) vs. OVO-SMOTE (Logit).

Figure 5. OVO-SMOTE-Adaboost (MDA) vs. OVO-SMOTE (MDA).

5.3. Discussion on Model Application

Corporate bond default has a negative impact on the development of bond market. Therefore, the corporate bond default risk prediction model based on integration of the OVO decomposition method, the SMOTE oversampling method and the Adaboost ensemble method is of great practical significance. Empirical results have shown that the OVO-SMOTE-Adaboost (DT) model has satisfying performance for multi-class imbalanced corporate bond default risk prediction. In practice, it can support the policy making of regulators and the decision makings of bond issuers and investors. First, for regulators, the OVO-SMOTE-Adaboost (DT) model for multi-class corporate bond default risk prediction is helpful for regulators to detect the risks in the bond market in advance, and to issue and implement regulatory policies in time to promote the healthy development of the bond market. Second, for bond issuers, the occurrence of bond default makes them harder to raise capital in the future, thus increasing their financing difficulty. The OVO-SMOTE-Adaboost (DT) model for multi-class corporate bond default risk prediction can help bond issuers to identify their own default risks in a timely manner and take management measures to prevent bond default. Finally, for bond investors, the OVO-SMOTE-Adaboost (DT) model for multi-class corporate bond default risk prediction can reduce the degree of information asymmetry between bond issuers and investors, and enable investors to identify and avoid investment risks in a timely manner.

6. Conclusion

In this paper, the OVO decomposition method, the SMOTE oversampling method and the Adaboost ensemble method are combined to establish the OVO-SMOTE-Adaboost ensemble models for multi-class imbalanced corporate bond default risk prediction. DT, SVM, Logit and MDA are used as the basic classifiers, respectively. The empirical results of the proposed ensemble models are compared with those of corresponding single classifier models, which only combine the OVO decomposition method and the SMOTE oversampling method without the Adaboost ensemble method. The conclusions are as follows:

(1) The OVO-SMOTE-Adaboost ensemble models achieve satisfying performance for multi-class imbalanced corporate bond default risk prediction. The OVO-SMOTE-Adaboost (SVM) model has better recognition ability for the bonds with very low default risk, the OVO-SMOTE-Adaboost (DT) model has better recognition ability for the bonds with relatively low default risk and the bonds with high default risk. The OVO-SMOTE-Adaboost (DT) model is overall superior to the other three ensemble models and is more suitable for multi-class imbalanced corporate bond default risk prediction.

(2) The OVO-SMOTE-Adaboost ensemble models show overall better performance for multi-class imbalanced corporate bond default risk prediction than the OVO-SMOTE single classifier models. In detail, the OVO-SMOTE-Adaboost (DT) model greatly outperforms the OVO-SMOTE (DT) model, and the performance of the OVO-SMOTE-Adaboost (MDA) model is slightly better than that of the OVO-SMOTE (MDA) model. There is little difference in performance of the other two groups of models. Therefore, it is of great significance to integrate the Adaboost ensemble method with the OVO-SMOTE model for multi-class imbalanced corporate bond default risk prediction, particularly when DT is used as the basic classifier.

This study still has the following limitations: (1) We only use financial features for multi-class imbalanced corporate bond default risk prediction, neglecting the impact of non-financial features on model performance. (2) We use cross-industry data to train and test the models and do not consider the different characteristics of corporate bond default risk prediction across different industries.

References

[1] Altman EI. Financial ratios discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance. 1968 Sep;23(4):589–609.
[2] Li H, Sun J, Wu J. Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. Expert Systems with Applications. 2010 Feb;37(8):5895-5904
[3] Lee YC. Application of support vector machines to corporate credit rating prediction, Expert Systems with Applications. 2007 Apr; 33(1):67-74.
[4] Sun J, Li H. Listed companies' financial distress prediction based on weighted majority voting combination of multiple classifiers. Expert Systems with Applications. 2008 Oct; 35: 818–827.
[5] Tsai CF, Sue KL, Hu YH, Chiu A. Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. Journal of Business Research. 2021 Mar; 130: 200-209.
[6] Li JP, Mirza N, Rahat B, Xiong D. Machine learning and credit ratings prediction in the age of fourth industrial revolution. Technological Forecasting and Social Change. 2020 Dec; 161: 120309.
[7] Sun J, Fujita H, Chen P, Li H. Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble. Knowledge-Based Systems. 2016 Mra; 120: 4-14.

[8] Sun J, Lee YC, Li H, Huang Q. Combining B&B-based hybrid feature selection and the imbalance-oriented multiple-classifier ensemble for imbalanced credit risk assessment. Technological and Economic Development of Economy. 2015 May; 21(3):351-378.

[9] Liu Z, Wang R, Tao W, Cai X. A class-oriented feature selection approach for multi-class imbalanced network traffic datasets based on local and global metrics fusion. Neurocomputing. 2015 Nov; 168: 365-381.

[10] Sun J, Lang J, Fujita H, Li H. Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates. Information Sciences. 2018 Jan; 425: 76-91.

[11] Li L, Yang C, Xie Q. 1D embedding multi-class classification methods. International Journal of Wavelets, Multiresolution and Information Processing. 2016 Apr,14(2).

[12] Zhou L, Tam KP, Fujita H. Predicting the listing status of Chinese listed companies with multi-class classification models. Information Sciences. 2016 Jan; 328: 222-236.

[13] Sun J, Fujita H, Zhenbbg Y, Ai W. Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods. Information Sciences. 2021 Jul; 559: 153-170.