Personal Credit Risk Assessment Based on Stacking Ensemble Model

Maoguang Wang, Jiayu Yu[(✉)], and Zijian Ji

School of Information,
Central University of Finance and Economics, Beijing, China
jiayuyl212@163.com

Abstract. Nowadays, compared with the traditional artificial risk control audit method, a more sophisticated intelligent risk assessment system is needed to support credit risk assessment. Ensemble learning combines multiple learners that can achieve better generalization performance than a single model. This paper proposes a personal credit risk assessment model based on Stacking ensemble learning. The model uses different training subsets and feature sampling and parameter perturbation methods to train multiple differentiated XGBoost classifiers. According to Xgboost's high accuracy and susceptibility to disturbances, it is used as a base learner to guarantee every learning "Good and different". Logistic regression is then used as a secondary learner to learn the results obtained by Xgboost, thereby constructing an evaluation model. Using the German credit data set published by UCI to verify this model and Compared with the single model and Bagging ensemble model, it is proved that the Stacking learning strategy has better generalization ability.

Keywords: Credit risk · XGBoost · Stacking

1 Introduction

Credit risk, also known as default risk, refers to the possibility of a default occurring when the borrower fails to repay the agreed debt on the basis of the credit relationship. The occurrence of credit risk mainly involves two aspects of repayment ability and repayment willingness. The repayment ability means that the borrower does not have enough funds to repay the previous loan for some reason; and the willingness to repay means the idea of the repayment of the borrower. The former is a concern and consideration in the field of credit risk assessment and measurement, while the latter mainly involves the identification of fraud.

Credit assessment allows the credit industry to benefit from improved cash flow, guaranteed credit collection, reduced potential risks, and better management decisions [1]. Therefore, the credit evaluation field has received a great deal of attention and it is of practical significance to carry out research on this.

Durand [2] first analyzed the credit risk and for the first time built a consumer credit evaluation model using statistical discriminant analysis. The results show that the use of quantitative methods can achieve better predictive ability. With the increase of data

volume and the development of machine learning algorithms, machine learning algorithms have been extensively studied in the field of credit evaluation.

Yu [3] et al. used SVM to construct a credit risk assessment model. However, SVM is sensitive to parameter settings and optimization methods. Khashman [4] studied the credit risk assessment system based on back propagation neural network. Experimental results show that the neural network can be effectively used for automatic processing of credit applications.

Ensemble learning which combines multiple learners allows the learner to learn different parts, making the ensemble model more generalized than a single model. Xiang [5] combined bagging with the C4.5 decision tree algorithm and applied it to German and Australian data to evaluate consumer credit. Chen [6] used XGBoost as a base classifier to construct a credit scoring model using data space, feature space, and algorithm parameter strategies, and then used a bagging method to learn a good integration model.

Because XGBoost has strong learning ability and generalization ability, it can guarantee accuracy; while the tree-based model belongs to the unstable learner, it is sensitive to parameter perturbation. Integrating XGBoost to ensure a "good and different" base learner improves the classification accuracy of the integrated model. This paper proposes a Stacking ensemble credit scoring model based on XGboost-LR. The model uses different training subsets and feature sampling and parameter perturbation methods to train multiple differentiated XGBoost classifiers; then the predicted probability obtained by each base classifier is used as a logistic regression feature variable to train logistic regression models, and According to the threshold, the probability is converted into the classification result. Using AUC as evaluation indicators, the effectiveness of the Stacking integration strategy was verified through comparison with a single model and a Bagging ensemble strategy. In the personal credit assessment, this method makes full use of the advantages of Stacking, and can solves the inadequacies of a single algorithm, providing a new research idea.

2 Model Strategy and Design

2.1 Model Strategy

GBDT (Gradient Boosting Decision Tree) was first proposed by Friedman [7]. The idea of GBDT optimizing its parameters is that each iteration is to fit the residual and approximate the residual using the negative gradient value of the loss function in the current model, and then fit a weak CART regression tree.

However, GBDT needs to iterate several times before the model have a better performance. Chen [8] have optimized this and proposed the XGBoost model. XGBoost uses a second-order Taylor expansion for the loss function.

According to the "Error-ambiguity decomposition" principle [9], the effectiveness of the ensemble model depends on the difference of the base learner. Therefore, constructing a xgboost with a large difference is the key to this model. The diversity of the base learner is mainly considered in the following three parts: Diversity of sample space, Diversity of feature space and Diversity of parameters.

The Bootsrap method [10] is a random sampling method with replacement. This method is conducive to the construction of different training sets, this paper will use the Bootsrap sampling method for sample space disturbance. The Random Subspace Method (RSM) proposed by Ho [11] refers to extracting some attribute subsets and then generating a base learner for each attribute subset training. This ensemble method based on feature partitioning can enhance the independence between the base learners. The difference in parameters of the base learner can easily generate a variety of base learners and enhance the generalization ability of the training model.

2.2   Model Design

Xgboost trains the model on the training set and then obtains the predicted probability of the trained Xgboost model on the validation set. The prediction probabilities are entered into the logistic regression model together with the original verification set labels as training samples of the secondary trainer, and the final model based on the Stacking ensemble strategy is trained.

For the accuracy of the Xgboost model, the influence of Learning rate, max_depth, and Column subsample ratio is greater. So in diversity of parameters aspect, Xgboost parameters are randomly selected in Table 1, making the model parameters diversified.

Table 1. Range of base classifiers parameters

Parameter	Range
Learning rate	[0.01, 0.05, 0.1, 0.2]
max_depth	[3, 4, 5, 7, 8]
Column subsample ratio	[0.5, 0.6, 0.7, 0.8]
reg_lambda	[0.4, 0.6, 0.8, 1, 2, 3]
Number of estimators	[70, 80, 90, 100]

3   Experiments

3.1   Data Pre-Processing

This paper uses UCI public the German credit data set [12] for model validation. The sample size in the data set is 1000, and the ratio of "good credit" customers to "bad credit" customers is 7/3. There are 20 variables in the original data set, including 7 numerical variables and 13 qualitative variables. Ordered qualitative variables can be directly divided into numbers. Unordered qualitative variables need to be transformed into dummy variables using one-hot encoding. In the end, there are 24 variables, 1 dependent variable, and dependent variable = 1 means "good credit", dependent variable = −1 means "bad credit".

Due to the need to compare the generalization capabilities of different models in the follow-up, the distance-based training model is susceptible to the gap between the data orders. Therefore, the data needs to be normalized. This article takes the maximum and minimum normalization method for the original data.

Randomly divide the initial sample data into disjoint two parts: the training set and the test set. 65% of the samples were used as training set, and 35% of the samples were used as test set. In constructing the xgboost-based learner, a training set is divided into a training subset and a validation subset in a 5-fold cross validation manner. So there will be 5 base classifiers.

3.2   Model Training

Because risk assessment is a two-category problem. Therefore, in the XGBoost model, the negative binomial log likelihood is used as a loss function:

$$L(y,f(x)) = \ln(1 + exp(-yf(x))), \quad y \in \{-1,1\} \tag{1}$$

As mentioned before, XGBoost belongs to a classifier with a large number of hyper-parameters. The value of the parameter is crucial. The performance of the model depends on the choice of parameters. However, so far, there is no theoretical method to guide the choice of parameters. In addition to the parameters in the spatial perturbations mentioned above, the remaining parameters are selected by the grid search method. The range of the value of the hyper-parameter is shown in Table 2.

Table 2. Range of other XGBoost parameters in grid search

Parameter	Grid search
min_child_weight	(1, 6, 2)
Subsample ratio	(0.5, 1, 0.1)
Gamma	(0, 0.1, 0.01)

Grid search uses 10-fold cross validation to perform parameter optimization. After parameter optimization, the min_child_weight of each leaf node was 2, and the Subsample ratio was 0.7, gamma was 0.

In terms of criteria, AUC on the test set are used as performance evaluation criteria for the model. AUC is an important indicator to measure the quality of different models. It is the area under the ROC curve. The closer AUC is to 1, the higher the degree of reality is.

3.3   Model Results

In order to verify the validity of the XGBoost-LR model using the Stacking ensemble strategy, this paper combines the three single models of Random Forest, XGBoost and Naive Bayes Algorithms, and these algorithms using the Bagging ensemble strategy model to compare the results. All models for comparison use the same data training set and test set for training and are trained 1000 epochs for robust. All single models apply the default values of the model in the scikit-learn package. In the bagging strategy, the sample space sampling rate is 0.8 and the feature space sampling rate is 0.6.

From Table 3, it can be seen that in the three single models using the original features, the XGBoost has the highest AUC value of 76.77%, and Naive Bayes is second only to XGBoost. The AUC of the learner applying Bagging strategy is higher than that of the single model, which shows the effectiveness of the ensemble strategy. As far as the ensemble strategy is concerned, it can be seen from this paper that the Stacking strategy has better generalization ability than the Bagging strategy and the AUC value increases from 77.81% to 78.48%. Although the increase in AUC is not significant, it is very promising because the ensemble training of 1000 samples is a very challenging classification task. Therefore, the classification AUC value proves the effectiveness of the proposed Stacking method.

Table 3. Performance comparisons for different classification and ensemble models

Model	AUC(%)
Random forest	74.34
Naive Bayes	75.28
XGBoost	**76.77**
RF-Bagging	77.66
Naive Bayes-Bagging	75.64
XGBoost-Bagging	**77.81**
XGBoost-LR-Stacking	**78.48**

4   Conclusion

This paper studies the ensemble strategy of models in the field of personal credit risk assessment and proposes an ensemble model based on Stacking. This model uses XGBoost as a base classifier and combines the random subspace and Bootstrap algorithms to increase the diversity between the base classifiers. The Logistic Regression model is used as a secondary learner, learning the results of each XGBoost as a feature variable to obtain an assessment model. The experiment confirmed that the model can have a certain degree of early warning for identifying customers with risks and has better generalization capabilities.

The future direction of work will focus on expanding the amount of data and constructing a more diverse base learner to verify the effectiveness of the model. And the use of domestic credit data to better build the XGBoost-LR ensemble model for China's real credit business.

References

1. Lee, T.S., Chiu, C.C., Lu, C.J., et al.: Credit scoring using the hybrid neural discriminant technique. Expert Syst. Appl. **23**(3), 245–254 (2002)
2. Durand, D.: Risk Elements in Consumer Instalment Financing, pp. 189–201. National Bureau of Economic Research, New York (1941)
3. Yu, L., Yao, X., Wang, S., Lai, K.K.: Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection. Expert Syst. Appl. **38**(12), 15392–15399 (2011)
4. Khashman, A.: A neural network model for credit risk evaluation. Int. J. Neural Syst. **19**(04), 285–294 (2009)
5. Xiang, H., Yang, S.: Clustering-based bagging ensemble consumer credit assessment model. Consum. Econ. **1**, 50–52 (2011)
6. Chen, Q., Wang, W., Ma, D., et al.: Class-imbalance credit scoring using Ext-GBDT ensemble. Appl. Res. Comput. **35**(2), 421–427 (2018). (in Chinese)
7. Friedman, J.H.: Stochastic gradient boosting. Comput. Stat. Data Anal. **38**(4), 367–378 (2002)
8. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, San Francisco (2016)
9. Krogh, A., Vedelsby, J.: Neural network ensembles, cross validation and active learning. In: International Conference on Neural Information Processing Systems, vol. 7, no. 10, pp. 231–238 (1994)
10. Efron, B., Tibshirani, R.: An Introduction to the Bootstrap, vol. 23, no. 2, pp. 49–54. Chapman & Hall, New York (1998)
11. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. **20**(8), 832–844 (1998)
12. UCI. Statlog (German credit data) data set[EB/OL]. https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german