APPLICATION OF BAYESIAN OPTIMIZATION AND STACKING INTEGRATION IN PERSONAL CREDIT DELINQUENCY PREDICTION

Jicong Yang and Hua Yin

Guangdong University of Finance and Economics, China

ABSTRACT

The national concept of consumption has changed to excessive consumption, and overdue debts have also increased. The surge of non-performing loans will not only lead to the liquidity difficulties of banks, but also lead to financial risks. Accurate prediction of personal credit overdue is one of the key issues to control financial risks. Traditional machine learning methods build classification models according to the characteristics of credit users, while ensemble learning can ensure high accuracy and prevent model overfitting, which is the mainstream of current application research. The Stacking method can fully combine the advantages of the base model and improve the model performance. The base model and hyperparameter selection have great influence on the prediction accuracy. Therefore, parameter selection according to the studied problem is the core of application. In this paper, the Stacking method is used to integrate multiple single models for credit user overdue prediction, and the parameters of the base model are optimized. The improved Bayesian optimization method is used to select appropriate parameter combinations to improve the model performance.

KEYWORDS

Credit overdue forecast, Stacking integrated learning, Bayesian optimization.

1. INTRODUCTION

With the change in consumer attitudes, the amount of consumer loans to our residents has also grown and the outstanding debt has increased. The proliferation of non-performing loans not only brings the problem of capital turnover difficulties to banks, but also constrains their development and may lead to financial risks, which in turn adversely affects domestic financial development; therefore, accurate prediction of personal credit overdue prediction is a key issue in controlling financial risks.

In this paper, the prediction of personal credit overdue is modeled as a classification problem. Through the personal and loan characteristics of previous credit users and overdue categories, a learning model is established to predict whether personal credit is overdue. The traditional classification algorithm for constructing a single model has the problems of uncertainty and weak generalization. The ensemble learning method integrates diversified weak classifier results to ensure high accuracy while preventing model overfitting [1]. In this paper, XGBoost, random forest and GBDT are used to construct the base learner, and Stacking method is used to integrate. However, the base learner usually needs to set hyperparameters, and the selection and setting of hyperparameters have a great impact on the prediction accuracy. Based on the above problems, this paper improved the Bayesian optimization algorithm and constructed an adaptive balance factor to improve the acquisition function, so that it could dynamically overcome the problem

David C. Wyld et al. (Eds): CST, NLMLT, DMS, CLBD, ITCS, VLSIE - 2022
pp. 79-92, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121908

that the Bayesian optimization algorithm would fall into the local optimum, optimize the hyperparameters of the base learners of random Forest, GBDT and XGBoost, and construct the optimization Stacking model. The overdue prediction is made based on the real customer data of UnionPay to verify the effect of the model.

Credit risk prediction has been an issue of importance to the financial industry, and in the past studies, researchers have been using various methods to construct credit risk models, and the specific work is as follows.

Wiginton[2] first proposed the use of logistic regression in corporate credit risk management problem and through experimental results it was concluded that logistic regression model has good prediction results in corporate credit risk management problem. Shin et al[3] selected the bankruptcy dataset of Korean listed companies to use SVM to predict the risk of corporate bankruptcy, and the analysis of the results obtained that SVM works better than MDA, Logit and NNs.Chen et al.[4] designed the XGBoost model with improved gradient boosting tree, second order Taylor expansion and also added regularization term to make the performance of the model improved significantly. After the introduction of XGBoost model, a large number of scholars started to apply XGBoost model to the field of risk control. Huang YP et al.[5] used XGBoost model with financial statements of listed companies in Taiwan as the research dataset. The analysis of the results concluded that XGBoost predicted the best results. chang YC et al.[6] used XGBoost models to predict credit risk problems and the results showed that XGBoost models have better results compared with logistic regression and SVM models.

In summary, from the traditional discriminant analysis method to the integrated learning XGBoost method, these models show good results in risk prediction. However, compared with the traditional method and machine learning method, the integrated learning method shows better prediction effect. This paper selects the base learner suitable for the problem studied in this paper to build the model based on Stacking method and referring to the studies of scholars. However, there is a very important factor in the construction of the model: the parameters of the model. Different parameter choices have different applicability to the problem. In view of this problem, this paper does further research.

2. THEORY

2.1. XG Boost

Based on gradient lifting tree algorithm, XGBoost algorithm adds regularization term to the objective function, which can reduce the complexity of the model and avoid overfitting[7]. Its objective function is shown in Equation (1) and Equation (2).

$$\text{obj}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1} \Omega(f_k) \tag{1}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\omega^2 \tag{2}$$

Where $\hat{y}_i$ is the predicted value, y_i is the true value, $\Omega(f_k)$ is the regular term, f_k is the decision tree, T represents the number of leaf nodes, ω represents the proportion of leaf nodes, γ controls the number of leaf nodes, and λ controls the proportion of leaf nodes.

XGBoost algorithm performs iterative operation and second-order Taylor expansion in the process of solving the objective function, as shown in formula (3) which improves the solving speed and the training speed of the model.

$$obj^{(t)} = \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_i) \tag{3}$$

Where g_i and h_i are the first and second derivatives of the loss function, respectively.

$$\begin{aligned} g_i &= \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \\ h_i &= \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \end{aligned} \tag{4}$$

2.2. Random Fores

Random forest is a kind of Bagging method, and decision tree model is used as the base model. The resampling method is used to select multiple sample sets with the same sample size as the given sample each time from the given sample and construct the decision tree based on it . In general, a decision tree divides nodes by selecting a feature from the set of features that can make the model result shift to the best direction. Random forest algorithm adopts the method of random feature selection. Specifically, when building each decision tree, firstly, a subset set containing M ($m \leq m$) features is selected from the feature set to which the node belongs, and the optimal features in this subset are divided. And $n = \log_2 N$ is a random parameter[8].

2.3. GBDT

GBDT is one of Boosting methods. GBDT mainly generates new decision trees, and takes the residuals of the results obtained from the decision trees in this stage as the input of the new decision trees in the next stage, and continues to iterate until the end of the iteration, the cumulative sum of the results of each decision tree is the result of the studied problem. At each iteration, the current decision tree needs to learn the prediction results and residuals of all decision trees in the previous iteration, and build the decision tree with the strategy to reduce the residuals in the subsequent iteration. Its advantage lies in the simple structure of GBDT, has a strong interpretability, the disadvantage is that there is no way to predict the development trend of a problem, that is, only in the scope of the prescribed prediction, can not exceed[9].

2.4. Stacking

The Stacking model fusion method selects multiple basic models and then combines the selected multiple models by specific methods. Because of the differences among models, the purpose of model fusion is to reflect the advantages of different models and make these weak models form strong models by certain methods. However, before adopting the method of model fusion, two criteria of model fusion should be followed. Firstly, the performance of the fused base learners should not be too different, and secondly, there should be discrimination between the learners. Only in this way can model fusion be adopted.

Figure 1 shows the algorithm flow. First, the given data set is divided into five parts, four of which are used for training and the other one is used for testing. Each time, the current training results are taken as the training set of the next layer model. It is also necessary to predict the test

set, take the arithmetic average of the results, and send them to the next layer for prediction. Then, the training results of the first-layer model are taken as the training set of the second-layer model, and the prediction results are taken as the test set of the second-layer model, and all of them are sent to the second layer for training and testing[10] .

According to the Stacking fusion criterion, the base model of the first layer fusion should have good performance, and the performance difference between the models should not be too big. From this perspective, XGBoost model and random forest model were selected as the base model of the first layer, and GBDT model was selected as the Stacking model of the second layer. The structure is shown in Figure 2.

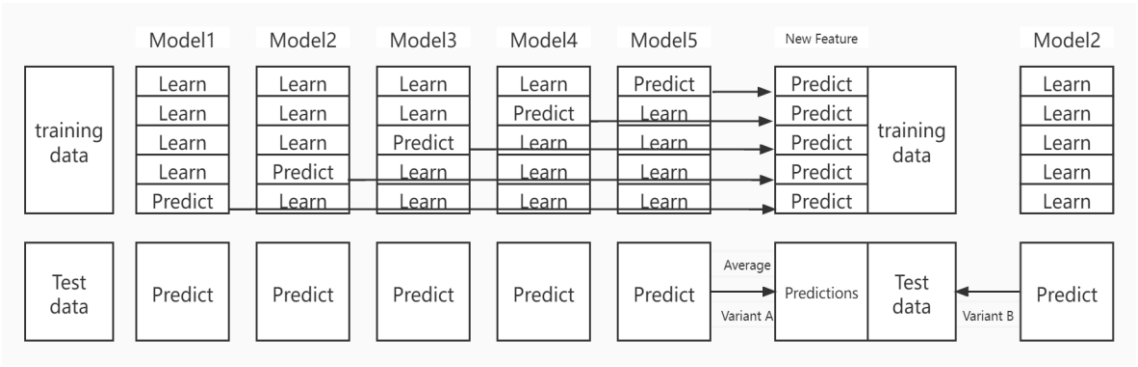

Figure 1. Stacking algorithm process

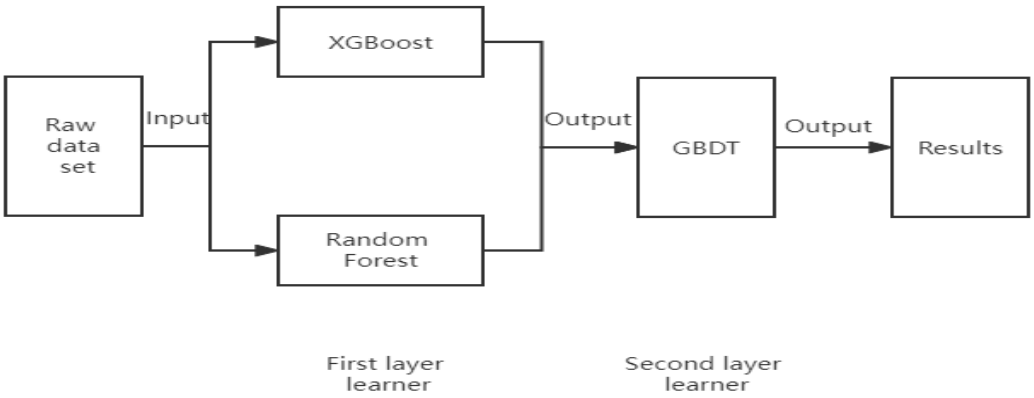

Figure 2. Stacking Model Structure

2.5. Bayesian Optimization Method and its Improvement

(1) Principle of Bayesian optimization algorithm

The idea of Bayesian optimization algorithm is to solve problems in global optimization by approximate approximation. There are two key steps in the execution of the Bayesian optimization algorithm. First, a priori function must be chosen to represent the distribution assumptions of the function being optimized. For this purpose, a Gaussian process is chosen because of its flexibility and ease of handling; second, a collection function must be constructed for determining the next point to be evaluated from the model posterior distribution[11].

In order to carry out Bayesian optimization, it is necessary to consider the establishment of distribution in the objective function, which is usually solved by Gaussian process.

A Gaussian process is an extension of the multidimensional Gaussian distribution to an infinite-dimensional stochastic process. It is defined by the mean value function $\mu(x)$ and the covariance function $k(x,x')$. the Gaussian distribution can be expressed as shown in Equation (5).

$$f(x) \sim GP(\mu(x), k(x,x')) \tag{5}$$

Where $\mu(x) = E(f(x))$, $E(f(x))$ is the mathematical expectation of $f(x)$, and the default value is 0; $f(x)$ denotes the mean absolute error; $k(x,x')$ denotes the covariance function of x.

Assuming that the past information $D_{1:t} = \{x_{1:t}, f_{1:t}\}$ has been obtained, where $f_t = f(x_t)$, then the next value to be searched for is $f_t = f(x_t)$ and the covariance matrix K is noted as：

$$K = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_t) \\ \vdots & \ddots & \vdots \\ k(x_t,x_1) & \cdots & k(x_t,x_t) \end{bmatrix} \tag{6}$$

From the Gaussian process, it follows that both f_t and f_{t+1} obey the joint Gaussian distribution. If we set the mean value of both to be 0, then the joint Gaussian distribution can be expressed as shown in Equation (7).

$$\begin{bmatrix} f_{1:t} \\ f_{t+1} \end{bmatrix} \sim N\left(0, \begin{bmatrix} k & k \\ k^T & k(x_{t+1},x_{t+1}) \end{bmatrix}\right) \tag{7}$$

where k can be expressed as：

$$k = [k(x_{t+1},x_1) \cdot k(x_{t+1},x_2) \cdots k(x_{t+1},x_t)] \tag{8}$$

The posterior probability of f_{t+1} is obtained by means of the edge density function is:

$$p(f_{t+1} \mid D_{1:t}, x_{t+1}) = N(\mu_t(x_{t+1}), \sigma_t^2(x_{t+1})) \tag{9}$$

where $\mu_t(x_{t+1})$ and $\sigma_t^2(x_{t+1})$ are calculated as follows:.

$$\mu_t(x_{t+1}) = k^T K^{-1} f_{1:t} \tag{10}$$

$$\sigma_t^2(x_{t+1}) = k(x_{t+1},x_{t+1}) - k^T K^{-1} k \tag{11}$$

From the above calculation it can be estimated that x_{t+1} satisfies a normal distribution at any interval, which in turn enables the sampling function to determine the next most dominant sample point.

By determining the next point to be evaluated through the sampling function, the number of iterations can be reduced and the evaluation cost can be lowered. Usually, the selection of sampling points is considered from two aspects: exploitation and exploration. exploitation is to search around the current optimal solution according to it, so as to find the global optimal solution; exploration is to try to explore the unevaluated sample points to avoid getting into the local optimal solution.

The acquisition function used in this paper is Probability of Improvement, and its acquisition function is shown in Equation (12).

$$PI(x) = \Phi(\frac{\mu(x) - y_{\max} - \delta}{\sigma_t(x)}) \tag{12}$$

where $y_{\max}$ is the current function optimal value, Φ is the standard normal distribution cumulative distribution function, and Φ is the equilibrium parameter that balances the relationship between development and exploration.

(2) Bayesian optimization algorithm improvement

However, influenced by the equilibrium parameter δ, the parameter value is too small will lead to the case of local optimal solution and too large will affect the exploration efficiency. Since the equilibrium parameter δ is a fixed value and cannot be dynamically adjusted according to the optimization condition, it can easily lead to the case of local optimal solution, therefore, this paper constructs the adaptive equilibrium factor to improve the acquisition function so that the acquisition function can avoid falling into the local optimal solution as much as possible. The improved collection function is shown in Equation (13) :

$$PI(x) = \Phi(\frac{\mu(x) - y_{\max} - \varepsilon}{\sigma(x)}) \tag{13}$$

In the formula $\varepsilon = 1 - 1/u$, $u = e^{y_{\max} - y}$, $y_{\max}$ represents the maximum value of the objective function in the current observed data, y represents the objective function value of the collection point in the last iteration, when y is close to $y_{\max}$, ε approaches 0, and the collection function tends to explore the state; when y is far from $y_{\max}$, ε approaches 1, and the collection function tends to develop the state.

2.6. Iv Value and WOE

When building a model, it is usually necessary to judge whether features have predictive ability, while IV refers to the value of information, which can be used to judge whether features can have predictive ability[12].The IV values are calculated as follows:

$$IV_i = (p_{y_i} - p_{n_i}) * WOE = (p_{y_i} - p_{n_i}) * \ln(\frac{p_{y_i}}{p_{n_i}}) = (\frac{y_i}{y_T} - \frac{n_i}{n_T}) * \ln(\frac{y_i / y_T}{n_i / n_T}) \tag{14}$$

Equation (14) is the IV value of a grouping in a variable, which is the sum of the IV values of each grouping , n is the number of variable groupings. In order to reflect the proportion of the sample size of a variable in the current subgroup to the overall, ($p_{yi} - p_{ni}$) is added here before WOE, so as to better reflect the contribution of a variable to the overall, the smaller the proportion, the smaller the contribution, and vice versa.

WOE in Equation (15) means weight of features. It is a way to encode the features. But the features need to be encoded after taking the corresponding grouping. After grouping, The WOE value for group i is calculated as follows:

$$\mathrm{WOE_i} = \ln(\frac{p_{y_i}}{p_{n_i}}) = \ln(\frac{y_i / y_T}{n_i / n_T}) = \ln(\frac{y_i}{y_T}) - \ln(\frac{n_i}{n_T}) \tag{15}$$

In Eq. (20), p_{yi} is the ratio of the number of past due in the group to the overall number of past due, p_{ni} is the ratio of the number of non-past due in the group to the overall number of non-past due, y_i is the number of past due in the group, n_i is the number of non-past due in the group, y_T is the number of all past due in the sample, and n_T is the number of all non-past due in the sample. Therefore, the meaning of WOE is the difference between "the number of past due in the group as a percentage of all past due" and "the number of non-past due in the group as a percentage of overall non-past due".

Usually an IV value less than 0.3 indicates no predictive power.

2.7. Peterson Correlation Coefficient Method

The Pearson correlation coefficient method is a measure of correlation between characteristics [13]. It is calculated as shown in Equation (16).

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 (y - \bar{y})^2}} \tag{16}$$

Where r indicates the correlation between two features. Usually r is less than 0.4 for weak correlation, greater than 0.6 for strong correlation, and greater than 0.8 for very strong correlation.

2.8. Evaluation Indicators

In order to enable comparison of training effects among different models, so the evaluation metrics taken in this paper include confusion matrix, accuracy, precision, recall, F1-score and AUC to measure the performance of a model [14].

The representation of the confusion matrix is shown in Table 1.

Table 1. Confusion Matrix

		Predicted results	
		1	0
True Category	1	TP	FP
	0	FN	TN

The accuracy rate is the proportion of correct samples to the total sample.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

The precision rate is the sample of all positive class samples with correct predictions.

$$Precision = \frac{TP}{TR + FP} \tag{18}$$

Recall is the fraction of all positive class samples that are correctly predicted.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{19}$$

The F1-score is the summed average of the recall and precision rates. It satisfies.

$$\frac{2}{\mathrm{F}_1} = \frac{1}{\mathrm{Precision}} + \frac{1}{\mathrm{Recall}} \tag{20}$$

$$\mathrm{F}_1 = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{TN}} \tag{21}$$

The ROC curve is a visual expression of the model effect. The dynamic relationship between TP and FP in the model is reflected by the drawn curve. To some extent, the differences between different learners can be understood through the ROC curve. The AUC value is the area under the curve, which is used to measure the generalization of the model.

3. DATA PRE-PROCESSING

3.1. Data Analysis

The data of this experiment are 11017 real data after desensitization provided by UnionPay. We built an extensive dataset with 199 credit characteristics.

3.2. Missing Value Handling

The source of the individual credit data widely miscellaneous, there may be repeat characteristics and lack of situation, and in the process of personal credit evaluation, the lack of some variable

values will affect the final prediction, if applied to the actual, may result in incalculable losses, so the first step to access to the data set needs to missing features the data set.

Table 2. Missing feature amount and proportion

Features	Missing amount	Missing percentage	Features	Missing amount	Missing percentage
X_121	10963	0.997906	X_110	10913	0.993355
X_120	10963	0.997906	X_063	10913	0.993355
X_119	10963	0.997906	X_071	10896	0.991808
X_118	10952	0.996905	X_072	10896	0.991808
X_102	10952	0.996905	X_073	10896	0.991808
X_103	10952	0.996905	X_107	10877	0.990078
X_104	10952	0.996905	X_115	10870	0.989441
X_111	10914	0.993446	X_116	10870	0.989441
X_064	10914	0.993446	X_117	10868	0.989259
X_062	10913	0.993355	X_108	10846	0.987257
X_109	10913	0.993355	…	…	…

As can be seen from Table 2, the missing ratio of X_062-X_073, X_081-X_087, X_092-X_120, X_128-X_130, X_133, X_135 and X_136 reaches more than 70%. Because the missing ratio is too high, if filling is adopted, It will affect the accuracy of the model. In order to reduce the deviation, the operation of deleting features is adopted in this paper.

3.3. Balanced Processing

As can be seen from the bar chart shown in Figure 3, the studied data set is unbalanced and not overdue: overdue = 4:1. Imbalanced data classification means that the proportion of categories in the data set is unbalanced. If the proportion of one category is large, the algorithm will favor the category with large proportion in classification. In order to eliminate the influence caused by the imbalance problem, this paper uses SMOTE algorithm to adjust the imbalance, so that the ratio of non-overdue class and overdue class in the processed data set reaches 1:1.

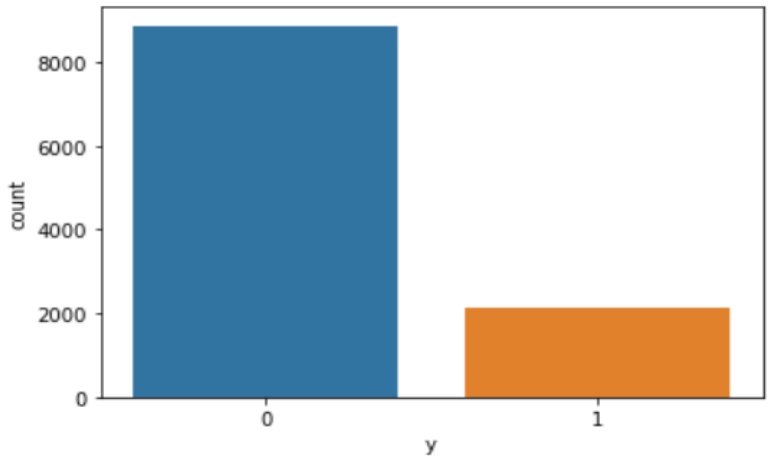

Figure 3. Positive and negative sample proportions

3.4. Feature Dimensionality Reduction

Credit data sets are characterized by high dimensionality and large redundancy among feature sets , so feature dimensionality reduction is needed. There are 146 features left in the dataset after processing for missing values. Feature dimensionality reduction is divided into two steps. Firstly, the IV value is calculated to remove the features with too low IV value. Secondly, the correlation analysis of the dataset was carried out to delete the features with high correlation.
Table 3 shows the statistics of some feature IV values：

Table 3. Value IV of the feature

Characteristics	Iv value	Characteristics	Iv value
X_125	0.640842	X_142	0.575979
X_146	0.631067	X_045	0.567916
X_078	0.622846	X_139	0.567685
X_127	0.615642	X_137	0.556560
X_126	0.606877	X_140	0.555608
X_141	0.601162	X_194	0.548917
X_144	0.593460	X_145	0.546987
X_131	0.590596	X_138	0.544725
X_059	0.582293	X_195	0.543420
X_079	0.576658	X_060	0.542415

After removing the features whose IV value was lower than 0.03 through the first step, 111 feature variables remained.

Correlation analysis was performed on the features to remove the variables with low IV value in the features with high correlation, and Pearson correlation coefficient method was used for processing.

After calculation by Pearson correlation coefficient method, features with correlation higher than 0.7 were removed to obtain the processed data set, which contained 40 feature variables.

4. EMPIRICAL STUDY

In this section, XGBoost, random forest and GBDT were first used to construct XRG-Stacking model and compared with XGBoost, random forest, GBDT, logistic regression and decision tree to verify the performance improvement of the fusion model compared with the single model. Furthermore, the improved Bayesian optimization method is used to optimize the parameters of XGBoost, random forest and GBDT. At the same time, it is compared with the optimization results of Bayesian optimization algorithm, grid search, random search, simulated annealing and genetic algorithm to verify the superiority of the improved Bayesian optimization algorithm. Finally, the IMPBO-XRG-Stacking model and the optimized base model were constructed by optimized XGBoost, random forest and GBDT to prove the improvement of problem accuracy by parameter optimization.

4.1. Experimental Verification

4.1.1. Comparative Analysis of XRG-Stacking Model

The experimental comparison results of XRG-Stacking and random Forest, XGBoost, GBDT, logistic regression and decision tree are shown in Table 4.

Table 4. Analysis of comparative results

	accuracy	precision	Recall	F1-score	AUC
logistic regression	0.6549	0.6683	0.6162	0.6412	0.6767
decision tree	0.7808	0.7630	0.8151	0.7882	0.8613
random Forest	0.7974	0.7938	0.8044	0.7991	0.8869
GBDT	0.8213	0.8457	0.7864	0.8150	0.9130
XGBoost	0.8607	0.9018	0.8267	0.8626	0.9408
XRG-Stacking	0.8779	0.9143	0.8353	0.8830	0.9508

According to the analysis in Table 4, the results of XRG-Stacking method had better effects compared with other basic models, indicating that the XRG-Stacking model that is combined with multiple models has better predictive effects on personal credit overdue problems than a single model. However, the XRG-Stacking model fusion method does not improve the prediction performance of the problem with a single model. By combining with real life, the number of resident loans is hundreds of millions and the number is very large. Therefore, the slight improvement in the performance of personal credit overdue prediction has a huge impact. It also has big implications for the financial industry. When it is difficult to further improve the performance of personal credit overdue problems by using a single model, the Stacking model fusion method can be considered to improve the ability of identifying whether users are overdue, so as to achieve better prediction effect.

4.1.2. Optimization Algorithm Optimization Base Model Comparison Experiment

The parameters to be optimized by the improved Bayesian optimization method for XGBoost, random Forest and GBDT models and the best parameter combination optimized by the improved Bayesian optimization algorithm are shown in Table 5

Table 5. Optimization parameters and optimal values

XGBoost	
parameter	**value**
learning_rate	0.07
n_estimator	177
min_child_weight	4.8
max_depth	10
gamma	0.31
subsample	0.83
colsample_bytree	0.72
Random Forest	
parameter	**value**
n_estimators	48
max_depth	10
min_samples_split	11
min_samples_leaf	14
GBDT	
parameter	**value**
n_estimator	70
learning_rate	0.1
subsample	0.8
max_depth	8
min_samples_split	150
min_samples_leaf	40

The prediction results obtained by feeding the optimization parameters into the model are compared with the optimization results of other optimization methods, as shown in Table 6.

Table 6. Comparative analysis of optimization models

	accuracy	**precision**	**Recall**	**F1-score**	**AUC**
XGBoost	0.8607	0.9018	0.8267	0.8626	0.9408
Random Forest	0.7974	0.7938	0.8044	0.7991	0.8869
GBDT	0.8213	0.8457	0.7864	0.8150	0.9130
Improved Bayesian-XGBoost	0.8781	0.9190	0.8388	0.8734	0.9498
Improved Bayesian-Random Forest	0.8323	0.8403	0.8211	0.8306	0.9181
Improved Bayesian-GBDT	0.8722	0.9107	0.8256	0.8661	0.9442
Bayesian-XGBoost	0.8721	0.9130	0.8301	0.8702	0.9478
Bayesian-Random Forest	0.8291	0.8310	0.8270	0.8290	0.9147
Bayesian-GBDT	0.8675	0.9023	0.8245	0.8616	0.9441

As can be seen from Table 6, compared with other optimization methods, the improved Bayesian optimization method has the best effect in optimizing the three base models.

4.1.3. Comparative Analysis of IMPBO-XRG-Stacking Model

The comparison and analysis results of the IMPBO-XRG-Stacking model constructed with optimized XGBoost, random forest and GBDT and the optimized base model are shown in Table 7.

Table 7. Comparison of evaluation results

	accuracy	precision	Recall	F1-score	AUC
Improved Bayesian-Random Forest	0.8323	0.8403	0.8211	0.8306	0.9181
Improved Bayesian-GBDT	0.8722	0.9107	0.8256	0.8661	0.9442
Improved Bayesian-XGBoost	0.8781	0.9190	0.8388	0.8734	0.9498
IMPBO-XRG-Stacking	0.8879	0.9243	0.8453	0.8830	0.9551

As can be seen from Table 7, the ImpBO-XRG-Stacking model is the highest compared with the single model in accuracy, accuracy, recall, F1-score and AUC.

5. CONCLUSION

The main research content of this paper is to use optimized XGBoost, random forest and GBDT to build Stacking model, select real desensitization data provided by UnionPay as data set, and send it into the model for training after data preprocessing. The comparison experiment with the optimized single model proves that the model fusion and parameter optimization can improve the accuracy of problem prediction.

ACKNOWLEDGEMENT

This work was supported by The Humanity and Social Science Youth foundation of Ministry of Education of China(21YJCZH20);Innovation Team Project of Higher Education of Guangdong "Intelligence Rule of Law Research Team" (2022WCXTD008).

REFERENCES

[1] Zehra W, (2021) Cross corpus multi-lingual speech emotion recognition using ensemble learning, Complex & Intelligent Systems, Vol. 7, No.4, pp1845-1854.
[2] J. C. Wiginton, (1980) A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior,Journal of Financial Quantitative Analysiss, Vol. 15, No.3, pp 757-770.
[3] Shin K S, (2005) An Application of Support Vector Machines in Bankruptcy Prediction Model,Expert Systems with Applications, Vol. 28, No.1, pp 127-135.
[4] T. Chen, (2016) XGBoost: A Scalable Tree Boosting System,KDD'16: Proceedings ofthe 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 785-794.
[5] Huang Y P, (2019)A New Perspective of Performance Comparison among Machine Learning Algorithms for Financial Distress Prediction, SSRN Electronic Journal.
[6] Chang Y C, (2018) Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions, Applied Soft Computing, pp73.
[7] Trizoglou P,(2021) Liu X, Lin Z. Fault detection by an ensemble framework of Extreme Gradient Boosting (XGBoost) in the operation of offshore wind turbines, Renewable Energy , Vol.179,pp 945-962
[8] Teles G, Rodrigues J J P C, Rabêlo R A L, et al.(2021) Comparative study of support vector machines and random forests machine learning algorithms on credit operation,Software: Practice and Experience,Vol. 51,No.12,pp 2492-2500
[9] Yang J S, Zhao C Y, Yu H T, et al.(2020) Use GBDT to predict the stock market, Procedia Computer Science,Vol.174,pp 161-171
[10] Khoei T T, Labuhn M C, Caleb T D, et al.(2021)A Stacking-based Ensemble Learning Model with Genetic Algorithm For detecting Early Stages of Alzheimer's Disease,2021 IEEE International Conference on Electro Information Technology,pp 215-222
[11] Xia Y, Liu C, Li Y Y, et al.(2017)A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring,Expert systems with applications,Vol.78,pp 225-241

[12] Wang Z, Zhang P, Sun W, et al.(2021) Application of data dimension reduction method in high-dimensional data based on single-cell 3D genomic contact data,ASP Transactions on Computers , Vol.1,No.2,pp 1-6
[13] Xie A, Yang H, Chen J, et al.(2021) A short-term wind speed forecasting model based on a multi-variable long short-term memory network,Atmosphere, Vol.12,No.5,pp 651
[14] Ahsan M M, Mahmud M A P, Saha P K, et al, (2021)Effect of data scaling methods on machine learning algorithms and model performance[J]. Technologies,Vol. 9,No.3,pp 52

AUTHOR

JICONG YANG is currently pursuing his master's degree at Guangdong University of Finance and Economics, majoring in big data analytics and applications.