Scientific Research Publishing

Open Journal of Business and Management, 2018, 6, 362-372
http://www.scirp.org/journal/ojbm
ISSN Online: 2329-3292
ISSN Print: 2329-3284

GBDT-SVM Credit Risk Assessment Model and Empirical Analysis of Peer-to-Peer Borrowers under Consideration of Audit Information

Zhou Li

School of Business Administration, South China University of Technology, Guangzhou, China
Email: gm_lizhou@foxmail.com

How to cite this paper: Li, Z. (2018) GBDT-SVM Credit Risk Assessment Model and Empirical Analysis of Peer-to-Peer Borrowers under Consideration of Audit Information. *Open Journal of Business and Management*, **6**, 362-372.
https://doi.org/10.4236/ojbm.2018.62026

Received: January 21, 2018
Accepted: April 23, 2018
Published: April 26, 2018

Abstract

With the rapid development of P2P (peer-to-peer) online lending industry, how to effectively evaluate the borrowers' credit risk in the platform has drawn more and more attention. In this paper, we propose a borrower credit risk assessment index system that includes basic information, work information, credit information, asset information, loan information and audit certification information, and come up with a credit risk assessment model that combines Gradient Boosting Decision Trees (GBDT) and support vector machine (SVM). Then, we select the data of P2P lending platform to carry out the empirical analysis of the credit risk assessment, and compare with the common four kinds of single prediction models such as logic regression (LR), artificial neural network (ANN), SVM and clustering algorithm. The results show that the increase of audit certification information helps to improve the forecasting effect of the model, and the credit risk assessment model of P2P lending platform based on GBDT and SVM has higher prediction accuracy and stability.

Keywords

P2P, Borrowers Credit Evaluation, GBDT, SVM

1. Introduction

As a new financing model, Peer-to-Peer Lending rapidly expanded and aroused widespread concern in the social community. According to Licaizhijia data: by the end of 2016, the accumulated volume of transactions in P2P lending industry has exceeded 3.36 trillion RMB, of which the transaction volume in 2016 is nearly 2 trillion, and the transaction scale in 2017 is expected to reach 4 trillion.

P2P lending platform in the rapid development also come up with some problems. According to the "P2P Lending Industry Report of the Whole Country in First Half Year 2016" published by Diyiwangdai platform, there were 2077 platforms that stopped operating, cash withdrawal difficulties, runways and other issues, with an increase of 559 over 2015. This will undoubtedly make investors and institutions face huge risks and will gradually reduce people's trust in P2P lending, which will hinder the development of P2P lending industry. It can be concluded from the above facts, that to find out how to accurately evaluate the credit risk of P2P borrowers is particularly important.

In recent years, some researches on credit risk assessment of P2P borrowers have been achieved initial results. Guo Zhongjin and Lin Haixia [1] studied the impact of P2P lending platform verification and authentication mechanism on the credit risk of P2P borrowers. The research shows that the certification of academic qualifications and household can reduce the bad loan rate of the P2P lending platform. Freedman and Jin [2] studied the impact of social networks on lending status in P2P lending. The study found that borrowers who provide social relationships are more likely to loan money and have lower interest rates. Chen and Han [3] classified credit information as "hard information" and "soft information" and compared Chinese and American investors. The results show that compared with the United States, Chinese investors pay more attention to "soft information". The artificial intelligence method is also gradually applied to the field of credit risk assessment. Malekipirbazari and Aksakalli [4] applied stochastic forest models to credit risk assessment and proved that stochastic forest classifiers who was discriminated with high reputation have a better effect than traditional FICO credit scores models. Harris [5] proposed a new credit risk assessment method—Cluster Support Vector Machine (CSVM), after comparing it with the original nonlinear support vector machine, and found CSVM has a better classification performance. Danenas and Garsva [6] successfully applied the particle swarm optimization algorithm to the traditional SVM, and applied the optimized model to the credit evaluation. At the same time, after comparing with the logistic regression and the neural network model, the results show that the optimization model increased classification accuracy significantly, but the stability of the model needs to be improved. Fang [7] and his colleagues used Lasso-logistic model to evaluate personal credit level. Compared with Logistic model and stepwise regression Logistic model, Lasso-logistic model can better extract the characteristic factors that affect credit risk, and improve the prediction accuracy greatly. Zhang *et al.* [8] introduced a hybrid model based on Logitsic regression model and SVM model. The model not only includes the advantages of a single model, but also analyzes the linear and non-linear features that affect the customer's credit risk. Wu Chong *et al.* [9] introduced the neural network into the field of credit risk assessment. Based on the fuzzy neural network, the credit risk of the borrower was evaluated, and the smaller network forecast error was obtained.

Based on the current researches, artificial intelligence methods have achieved some results in the field of personal credit risk assessment. However, the existing researches mainly focus on how to construct the evaluation model to evaluate the credit of individuals, but there is a lack of research on how to extract effective feature combinations, especially the lack of credit risk assessment of P2P borrowers. Credit data has the characteristics of complex distribution and various features, usually including discrete and continuous features [10]. And with the development of Internet technology, the portrayal of the credit level of P2P borrowers has also deepened and expanded. The characterization features of borrowers' credit will be more and more showing up. The selection of feature combination will be directly related to the credit risk assessment model for prediction accuracy and generalization ability [11].

Based on the existing researches both in China and oversea, this paper aims at the characteristics of complex distribution and various features of P2P borrowers' credit data and builds a P2P borrowers' credit evaluation index system, which is based on commonly used basic information, asset information, work information, credit information and loan information, with additional audit certification information to more objectively and truly reflect the borrower's credit level. Then, a credit risk assessment model for borrowers based on GBDT-SVM is constructed. By using GBDT model, the features of the original data of the borrower are extracted and a new feature combination is obtained, and then the SVM credit evaluation model is constructed based on the new feature combination. The model combines the advantages of a single model, not only to better capture the characteristics of the borrower credit, but also to simplify the structure of the forecasting model and improve the prediction accuracy and generalization ability of the model. It also reduces the labor and time costs of feature extraction. Finally, we collect the data of borrowers in RenRendai platform to conduct empirical research and compare with common prediction models such as Logistic Regression (LR), Neural Network (NN) and Support Vector Machine (SVM).

2. P2P Borrowers Credit Risk Index System

At present, the existing literature mainly focuses on five parts of information including personal basic information, work information, asset information, credit information and loan information when constructing P2P borrowers credit evaluation index system [12] [13] [14]. To enhance the authenticity and credibility of the borrower's information and reduce the fraud and default behavior, this article adds audit information on each borrower's conditions, including credit report certification, job certification and income certification. Audit process is completed by the P2P lending platform, and published in a timely manner on each loan page. Based on the credit evaluation index of borrowers in P2P lending platforms such as Renrendai and the commonly used credit evaluation indicators in the existing literature, this dissertation builds 6 first-level indicators and 47 second-level indicators. The indicator system is shown in **Table 1**.

Table 1. P2P borrowers credit evaluation index system.

First level indicators	Second level indicators	Indicator Description
basic information (X_1)	Age (x_1)	• 1 for 19 - 25 years old, 2 for 26 - 30 years old, 3 for 31 - 40 years old, 4 for 41 - 50 years old, 5 for above 50 years old.
	Gender (x_2)	• 1 for men, 2 for women.
	Marriage status (x_3)	• 1 for married, 2 for single, 3 for divorce.
	Education (x_4)	• 1 - 4 for master, postgraduate, high school, and below education.
Work information (X_2)	Company size (x_5)	• 1 - 4 for below 10, 10 - 100, 100 - 500, above 500.
	Industry (x_6)	• 1 - 19 for 19 different industries, including manufacturing, architecture, IT, real estate and so on.
	Working years (x_7)	• 1 for less than 1 year, 2 for 1 - 3 years, 3 for 3 - 5 years, 4 for above 5 years.
Assets information (X_3)	Revenue (x_8)	• 1 for monthly income less than 1000 RMB, 2 for monthly income between 1000 RMB to 2000 RMB, 3 for monthly income between 2000 RMB to 5000 RMB, 4 for monthly income between 5000 RMB to 10,000 RMB, 5 for monthly income between 10,000 RMB to 20,000 RMB, 6 for monthly income between 20,000 RMB to 50,000 RMB, 7 for monthly income more than 50,000 RMB.
	Real estate ownership (x_9) Mortgage (x_{10}) Car ownership (x_{11}) Car loan (x_{12})	• Refer to personal asset information, if there is a value of 1, otherwise the value of 0.
Credit information (X_4)	credits (x_{13}) loan times (x_{14}) overdue times (x_{15}) payoff times (x_{16}) successfully loan times (x_{17}) overdue revenue (x_{18}) bad debts (x_{19}) to be repayable ability (x_{20})	• Provided by the platform to represent to the previous personal credit information, which are all continuous variables.
	Credit level (x_{21})	• including AA, A, B, C, D, E, HR seven grades, respectively 1 - 7.
	Loan history (x_{22})	• 1 for having a loan value once, otherwise 0.
Loan information (X_5)	Loan purpose (x_{23})	• Including consumption, investment, education and other 9 major usages, ranging from 1 - 9.
	Loan description length (x_{24})	• This is the borrower's statement of the details of the loan, which reflects the borrower's attitude, expressed in words length of the statement.
	Request loan amount (x_{25}) loan amount (x_{26}) loan rate (x_{27}) payment term (x_{28})	• This is the borrower's loan related information, as a continuous variable.
Audit information certification (X_6)	Income certification (x_{29}) real estate certification (x_{30}) title certification (x_{31}) video certification (x_{32}) credit report certification (x_{33}) ID certification (x_{34}) mobile phone certification (x_{35}) phone bill certification (x_{36}) loan certification (x_{37}) bank account certification (x_{38}) ID scan copy certification (x_{39}) residence certification (x_{40}) work certification (x_{41}) education certification (x_{42}) details certification (x_{43}) marriage certification (x_{44}) car certification (x_{45}) children information certification (x_{46}) mobile certification (x_{47})	• This is the platform for the borrower to provide information on the audit certification. 1 for pass, otherwise 0.

3. Credit Risk Assessment Model Based on GBDT-SVM

3.1. Related Introduction of Theoretical Models

The Gradient Boosting Decision Trees (GBDT), first proposed by Friedman [15] in 2001, is a common artificial intelligence model. It makes use of the boosting thought in integrated learning [16], and iteratively reduces the training residuals

in the training process. Each iteration generates a decision tree. This idea enables GBDT capable of quickly capturing differentiated features combination. The decision tree structure is simple, the algorithm process is easy to understand, and with natural interpretability. It directly reflects the characteristics of the data and helps to understand the feature combination extracted by the model, which is not available in other algorithms [17].

Support Vector Machine (SVM) is a machine learning method proposed by Vapnik [18] and other researchers in the early 1990s. It is based on the principle of minimizing structural risk and ensures that the model can obtain smaller errors in both training and test sets [19]. In the dichotomous problem, we assume that the sample set $S=\{x_i,y_i\},i=1,\cdots,n$ is trained, which represents the sum of training samples. y_i is the corresponding output value x_i, and the value is 1 or −1. The classifier is a hyperplane $f(x)=\omega x-b=0$, if $f(x)>0$, then the point belongs to class 1, if $f(x)<0$, then the point belongs to class −1. Assuming that the sample becomes linearly separable after being mapped into a high-dimensional space through a non-linear function $\phi(x)$, it corresponds to solving the following optimization problem:

$$\begin{aligned}&\min\ \frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{n}\xi_i\\&\text{s.t.}\ \ y_i\left(\omega^{\mathrm{T}}\phi(x_i)+b\right)\geq 1-\xi_i\\&\qquad \xi_i\geq 0,\ i=1,2,\cdots,n\end{aligned}\tag{1}$$

Among them, C is for the penalty coefficient, ξ_i is for the relaxation of variables. According to the Lagrange function method, the problem can be translated into the following duality problem:

$$\begin{aligned}&\min\ Q(\alpha)=-\sum_{i=1}^{n}\alpha_i+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\Phi(x_i)^{\mathrm{T}}\Phi(x_j)\\&\text{s.t.}\ \ \sum_{i=1}^{n}\alpha_i y_i=0\\&\qquad C\geq\alpha_i\geq 0,\ i=1,2,\cdots,n\end{aligned}\tag{2}$$

α_i is the Lagrange factor, $\Phi(x_i)^{\mathrm{T}}\Phi(x_j)$ is the corresponding kernel function, available representation by $K(x_i,x_j)$

3.2. GBDT Constructs a New Feature Combination

GBDT is a combination model of decision trees. Using GBDT to construct feature combinations means that each leaf node in GBDT is regarded as a new feature. Therefore, the number of features in the new feature combination is the same as that of all leaf nodes in GBDT. And Each feature in the new feature combination has a value of 0 or 1. If the sample traverses a decision tree and falls on a leaf node, the value of the leaf node is 1, and the remaining leaf nodes of the tree take 0.

Because in Section 2, this paper constructs 47 second-level credit rating indicators, the original input sample is $X=\{x_1,x_2,\cdots,x_{47}\}$, After training GBDT model set of decision tree $T=\{T_1,T_2,\cdots,T_n\}$, n represents the total number of

decision trees in the GBDT. Assume: T_i refers to i decision tree, and m leaf nodes in this very i decision tree. $p=\{p_{i1}, p_{i2}, \cdots, p_{im}\}$ refers to the value of sample X in the leaf nodes of decision tree T_i, $p_{ij} \in \{0,1\}$. If sample X inputs to the decision tree T_i, and locate in j leaf node, then $p_{ij}=1$, otherwise $p_{ij}=0$. Input sample X in GBDT model, after training the responsive combinatorial features are $P=\{p_1, p_2, p_3, \cdots, p_{nm}\}$, mn refers to the sum of the numbers of all leaf nodes in GBDT model, and refers to the sum of all new features in the new combinatorial features. m and n refers to the parameters acquired through the training experiment during the decision process in GBDT model. The value of new combinatorial features $P=\{p_1, p_2, p_3, \cdots, p_{nm}\}$ coming from the previous sample $X=\{x_1, x_2, \cdots, x_{47}\}$ after GBDT model transformation is $n \times m$ dimension binary vector, n refers to the sum of decision trees in GBDT model, m refers to the number of leaf nodes in the decision tree, and X is responsive to P. Therefore, a schematic diagram of a new feature combination of GBDT construction can be obtained as follows (Figure 1).

3.3. GBDT-SVM Credit Evaluation Model Construction

Based on the new feature combination extracted by GBDT, this paper builds the credit risk assessment model of P2P borrowers by using SVM model. Therefore, GBDT-SVM composite model from the original SVM model (3) into the following model:

$$\begin{aligned} &\min\ Q_2(\alpha) = -\sum_{i=1}^{n}\alpha_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(P_i, P_j) \\ &\text{s.t.}\ \sum_{i=1}^{n}\alpha_i y_i = 0 \\ &\quad C \geq \alpha_i \geq 0,\ i=1,2,\cdots,n \end{aligned} \tag{3}$$

C is penalty factor, α_i is a Lagrange factor, $K(P_i, P_j) = \Phi(P_i)^{\mathrm{T}}\Phi(P_j)$ is a new kernel function, P_i is the new combinatorial feature after GBDT training

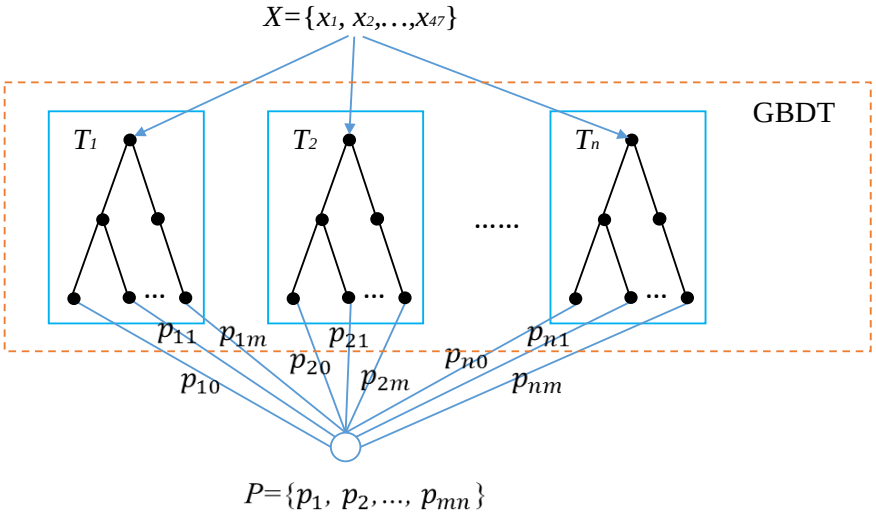

Figure 1. Combination schematic of GBDT new features construction.

model, $S^* = \{P_i, y_i\}$ is the new training sample set, n refers to the sum of training samples. The evaluation method of this paper is shown in Figure 2.

4. Empirical Analysis

4.1. Data Sources and Evaluation Criteria

Renrendai platform has anti-crawler restrictions on borrowers' credit data since 2015, so we can only crawl the borrower's data from 2011 to 2014 by web crawlers, including the credit information of 17,092 P2P borrowers. For the P2P lending platform, the biggest credit risk is default, so we use the default to measure the borrower's credit risk. In the sample data, there are 2205 default users and 14,887 non-default users. Default users are those who can not repay on time. Their credit will be affected and the next loan will be harder. Non-default users are those who repay on time. Their credit levels will rise, and the next loan will be easier. To facilitate the assessment of the predictive ability of the model, the default sample is marked as 1, regarded as a positive sample; non-default sample labeled as 0, as a negative sample. Correspondingly, the following definition is given:

TP: The number of borrowers who is actual default and also predicted to default

FP: The number of borrowers that are actually not in default and that are predicted to be in default

TN: The number of borrowers who have not actually defaulted and predicted no default

FN: The number of default borrowers who are predicted to be in not default

There are four general indicators for assessing the effect of model predictions: Accuracy, Precision, Recall, and F1-Measure. Specific to P2P lending issues can be displayed:

$$\text{Accuracy} = \frac{\text{sum of loaners who predicted accurately}}{\text{sum of all loaners}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{Precision} = \frac{\text{sum of loaners predicated as default and conduct default in fact}}{\text{sum of loaners predicted as default}} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{sum of loaners predicted as default and conduct default in fact}}{\text{sum of loaners conducting default in fact}} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

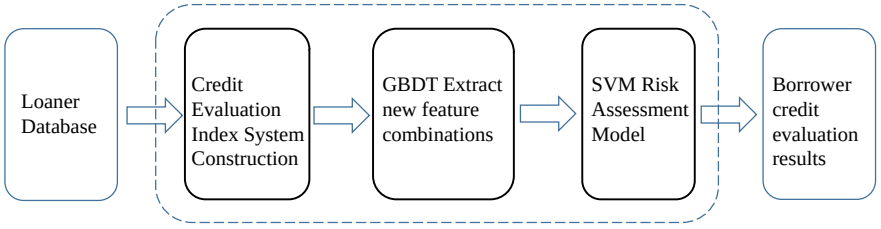

Figure 2. GBDT-SVM credit evaluation model diagram.

F1-Measure is the harmonic mean of the exact value P and the recall R, which is a comprehensive index to evaluate the effect of the model:

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TR}}{2\text{TR} + \text{FP} + \text{FN}}$$

Considering the peculiarity of P2P lending problems, we introduce two commonly used indicators—error 1 and error 2:

$$\text{Error 1} = \frac{\text{sum of loaners predicted default but don't conduct defact in fact}}{\text{sum of loaners who don't conduct default}} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

$$\text{Error 2} = \frac{\text{sum of loaners predicted no default but conduct default in fact}}{\text{sum of loaners who don't conduct default}} = \frac{\text{FN}}{\text{TP} + \text{FN}}$$

4.2. Data Preprocessing and Model Solving

Before the training of the model, the original data needs to be normalized to eliminate the influence of different dimension between different characteristic data so that the data can be compared. The standardization formula is as follows:

$$x_i' = \frac{x_i - \min x_i}{\max x_i - \min x_i} \tag{4}$$

To ensure the accuracy of the model prediction, this paper uses a randomized experiment for empirical analysis. The original data set is divided into training set and test set by 7:3 and conduct a total of 20 tests. The average date is set as the final prediction results, while using the standard deviations evaluate of the stability of the model. One common feature of P2P borrowers credit data is data imbalance. The sample of default used in this paper comparing to non-default is close to 7:1. According to the existing research [20], after the data is preprocessed by oversampling, the positive and negative samples in the training set are balanced, and then used in the training of the model. In the meantime, in order to show the advantages of the model, we compared it with Logistic Regression (LR) [7], Artificial Neural Network [9], Support Vector Machine (SVM) [6] and Clustering Algorithm [21].

4.3. Results Analysis

By comparing the general evaluation index of the prediction model, the prediction effect of the GBDT-SVM model is examined as shown in **Table 2**.

Accuracy and F1-Measure are comprehensive indicators of the evaluation model and should be given to priority. The comparison shows that the comprehensive prediction results of LR, ANN and GBDT-SVM models are relatively good, while the prediction results of clustering and SVM are poor. And GBDT-SVM combination model performs the best in these two aspects, especially with respect to SVM which has been greatly improved. LR and SVM per-

formed slightly better on Recall which shows the actual default. LR and SVM performed much less well than GBDT-SVM reflecting on prediction. It can be seen that the stability of LR and SVM models work not in a good way and fluctuates greatly, while the overall performance of GBDT-SVM is indeed better.

Then continue to investigate the model of the first type of error rate (Error 1) and the second type of error rate (Error 2), as shown in Table 3

Columns 1 and 2 in Table 3 indicate the second type of error rate and the first type of error rate, indicating the accuracy of the model predictions. The third and fourth columns are the corresponding standard deviations reflecting the model predictions of model stability. From the table, we can see that LR, SVM and GBDT-SVM models still have lower false-positive rates, while clustering and ANNs have higher false-positive rates. The GBDT-SVM composite model has a good performance in the first and second types of error rates. Although the second type error rate in LR and SVM model is better than the GBDT-SVM combination model, the performance of the first type of error rate is poor. From the perspective of model stability, the standard deviations of the two types of error rates of the GBDT-SVM composite model also performed the best. The standard error of the second kind of error rate in SVM model is smaller than that of GBDT-SVM combination model, but the standard error of the first kind of error rate is less than the combination model. Therefore, considering the combination of GBDT-SVM model does have a better performance on prediction and stability.

5. Conclusion

This article focuses on the assessment of the borrowers' credit risk on P2P

Table 2. Comparison of common indicators results.

	Accuracy	F1_measure	Precision	Recall
Clustering	0.7295	0.3789	0.2752	0.6096
LR	0.9480	0.8370	0.7271	0.9865
SVM	0.9368	0.8095	0.6832	0.9935
ANN	0.9635	0.8761	0.8103	0.9544
GBDT-SVM	0.9695	0.8968	0.8255	0.9819

Table 3. Error 1 and Error 2 in the model & the corresponding variance comparison.

	Error 2	Error 1	Error 2_std	Error 1_std
Clustering	0.3904	0.2516	0.0235	0.0097
LR	0.0135	0.0580	0.0054	0.0042
SVM	0.0065	0.0721	0.0024	0.0038
ANN	0.0456	0.0351	0.0172	0.0050
GBDT-SVM	0.0181	0.0325	0.0048	0.0023

lending platform. Based on the commonly used evaluation index, it adds audit certification index and constructs the borrower's credit evaluation index system. Then, by combining GBDT and SVM model, the GBDT model is used to extract the feature combination of the borrower's original data and the borrower's credit status is evaluated based on the new feature combination to build GBDT-SVM portfolio credit evaluation model. An empirical research is conducted on the borrowers' data in the Renrendai platform and compared with several commonly used machine learning models. The empirical results show that the addition of audit certification information effectively improves the prediction accuracy of the model. And the prediction accuracy and stability of the combined model are better than the single machine learning model. Therefore, the GBDT-SVM credit evaluation model proposed in this paper provides a new theoretical tool for the borrower's credit risk management for the P2P lending platform, which has obvious practical guiding significance for the operation decision of the platform. For example, based on the information provided by the borrower, if the GBDT-SVM credit evaluation model determines that the person is more likely to default, the platform will not lend to the person. If the GBDT-SVM credit evaluation model judges that the person is less likely to default, the person is allowed to borrow. This will reduce the borrower's default ratio and effectively control the borrower's credit risk.

Acknowledgements

This research has been supported by the Guangdong Natural Science Foundation Team Project (2017A030312001).

References

[1] Guo, Z.J. and Lin, H.X. (2013) Research on the Credit Mechanism of Internet Loan—Taking the Pat Loan as an Example. *Modern Management Science*, No. 5, 90-92.

[2] Freedman, S. and Jin, G.Z. (2017) The Information Value of Online Social Networks: Lessons from Peer-to-Peer Lending. *International Journal of Industrial Organization*, **51**, 185-222. https://doi.org/10.1016/j.ijindorg.2016.09.002

[3] Chen, D. and Han, C. (2012) A Comparative Study of Online P2P Lending in the USA and China. *Journal of Internet Banking & Commerce*, **17**, 1-15.

[4] Malekipirbazari, M. and Aksakalli, V. (2015) Risk Assessment in Social Lending via Random Forests. *Expert Systems with Applications*, **42**, 4621-4631. https://doi.org/10.1016/j.eswa.2015.02.001

[5] Harris T. (2015) Credit Scoring Using the Clustered Support Vector Machine. *Expert Systems with Applications*, **42**, 741-750. https://doi.org/10.1016/j.eswa.2014.08.029

[6] Danenas, P. and Garsva, G. (2015) Selection of Support Vector Machines Based Classifiers for Credit Risk Domain. *Expert Systems with Applications*, **42**, 3194-3204. https://doi.org/10.1016/j.eswa.2014.12.001

[7] Fang, K.N. and Zhang, G.J. (2014) Personal Credit Risk Early Warning Method Based on Lasso—Logistic Model. *The Journal of Quantitative & Technical Eco-*

nomics, No. 2, 125-136.

[8] Zhang, Q., Hu, L.Y. and Wang, Y. (2015) Study on Early Warning Model of Banking Credit Risk Based on Logit and SVM. *Theory and Practice—Systems Engineering*, **35**, 1784-1790.

[9] Wu, C., Lv, J.J., Pan, Q.S. and Liu, Y.X. (2004) Commercial Bank Credit Risk Assessment Model Based on Fuzzy Neural Network. *Theory and Practice—Systems Engineering*, **24**, 1-8.

[10] Ye, X.F. and Lu, Y.H. (2017) Credit Evaluation Model Based on Random Forest Fusion Naive Bayes. *Mathematics in Practice and Theory*, **47**, 68-73.

[11] Yao, X., Wang, X.D., Zhang, Y.X., *et al.* (2012) Summary of Feature Selection Methods. *Control and Decision*, **27**, 161-166.

[12] Su, Y. and Chen, C. (2017) Empirical Study on Factors Influencing Default Behavior of P2P Internet Borrowers. *Financial Development Research*, No. 1, 70-76.

[13] Li, B., Gao, Y. and Wang, J. (2016) Bank Personal Credit Evaluation Model Based on Gray Interval Correlation Analysis. *Mathematics in Practice and Theory*, **46**, 289-292.

[14] Fu, Y., Zang, D. and Qi, M. (2014) Risk Assessment of P2P Loan Credit. *Statistics and Decision*, No. 21, 162-165.

[15] Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, **29**, 1189-1232. https://doi.org/10.1214/aos/1013203451

[16] Zhou, Z.H. (2012) Ensemble Methods: Foundations and Algorithms. Taylor & Francis, Abingdon-on-Thames.

[17] Pang, S.L. and Gong, J.Z. (2009) Classification Algorithm and Its Application in Bank Personal Credit Rating. *Theory and Practice—Systems Engineering*, **29**, 94-104. https://doi.org/10.1016/S1874-8651(10)60092-0

[18] Vapnik, V.N. (1997) The Nature of Statistical Learning Theory. *IEEE Transactions on Neural Networks*, **8**, 1564-1564. https://doi.org/10.1109/TNN.1997.641482

[19] Li, L. and Zhou, Z. (2013) Enterprise Group Credit Risk Assessment SVM Integrated Classifier Construction and Application. *Technical Economy*, **32**, 65-70.

[20] Wu, L., Fang, B., Diao, L., *et al.* (2013) Over-Sampling and Under-Sampling Unbalanced Data Resampling Method. *Computer Engineering and Application*, **49**, 172-176.

[21] Zuo, Z. and Zhu, Y. (2004) Credit Rating Based on Data Mining Clustering. *Computer Applications and Software*, **21**, 1-3.