

Credit Risk Management of P2P Network Lending

Dongmei LI, Sanggyun NA*, Tao DING, Congchong LIU

Abstract: This article first studies the literature of P2P online loans, including online loans, credit risk factors and models, and summarizes the current status of P2P and credit risk assessment management in China. Based on the loan data of domestic P2P lending platforms, this paper conducts an empirical study on credit risk assessment. This study uses random forest importance assessment and logistic regression classification for credit risk assessment to identify loan targets with higher probability of default and improve overall loan quality. This research used 10,930 loan data, based on 26 fields, and finally selected 20 model variables to participate in credit risk quantification through feature structure and feature analysis. The final modelling test results show that the model screening accuracy rate is 73.3%, indicating that this model has a good performance in the credit risk quantification of borrowers.

Keywords: credit risk management; importance assessment; logistic regression; P2P network lending

1 INTRODUCTION

Since 2007, people's spending habits have changed greatly, including from cash to bank cards, from mobile phones to non-cash transactions, and from traditional finance to online finance. Many P2P financial companies were established due to the rapid rise of emerging industries, which has led to a huge increase in the number of P2P platforms [1].

Considering that many P2P borrowers have high non-performing loan ratios, investors and companies have suffered huge losses. Recently, P2P online lending has emerged. This microfinance market can provide some benefits to borrowers and lenders. There is a high correlation between the possibility of a loan default and the credit risk of the borrower. The higher interest rates charged to high-risk borrowers are not sufficient to cover the possibility of default on their loans [2]. Therefore, it is necessary to attract high credit scores and high credit borrowers to maintain their high quality business.

As the regulatory requirements for internet lending platforms have been frequently raised recently, many P2P platforms will encounter operational bottlenecks [3]. In particular, the restrictions on high interest rates have severely affected the platform's income level. Since most previous P2P platforms covered high risks through high interest rates, the platform itself could not control the risks well. However, in the future, it will be possible to operate well in a long-term strictly regulated environment, only if the P2P platform has good risk control, instead of relying on high interest rates to resolve the losses caused by high default rates.

A major problem with P2P network lending is the asymmetry of information between borrowers and lenders [4]. In other words, the lender does not know the credit quality of the borrower and the borrower is not aware of the real situation of the lender. There is no trust relationship between the two sides. This information asymmetry may lead to wrong choice and credit risks [5, 6]. In theory, the adverse consequences of these credit problems can be resolved or mitigated through pre-tests and regular monitoring. In the traditional bank loan market, banks can use collateral, certified accounts, regular reports, and even the existence of a board to increase trust in borrowers. However, this mechanism proposes a challenge to the P2P

platform environment. It is difficult to implement in the Internet lending platform and network environment, which will lead to significant transaction costs.

When the initial P2P platform was approved for loans, the evaluation basis of the platform was some scattered information of the borrower, such as the borrower's work unit, age, gender, marital status, monthly income and so on. The more common method of approval is to set a threshold criterion based on all the scattered information dimensions. When the borrower meets these conditions or can pass the personality assessment, the loan is granted.

In P2P Internet lending, borrowers who often fail to meet some of these thresholds, but are particularly good in the other conditions, are eventually strongly rejected. Lack of rules and inefficient approval standards are unreasonable. Therefore, how to integrate scattered and unstructured information into a scientific approval system has become an important research content of P2P network financial platforms.

Internet finance is an online financial service that uses an Internet network platform to conduct a series of financial activities [7]. It includes but is not limited to Internet financial companies that use the Internet, mobile Internet, streaming media and other network technologies to provide financial services. The Internet financial industry has continuously introduced new business products. At present, the main business items on the market include online loans, online banking, online payment, and Internet crowd funding [8].

P2P platform refers to a platform for financial peer-to-peer lending through the Internet, and is an abbreviation for P2P (P2P). P2P is called person-to-person (partner-to-partner) and also known as peer-to-peer online lending [9]. The P2P lending model is a private small-value lending model that collects a small amount of funds and then lends them to borrowers with capital needs [10]. This kind of financial service is a financial act of borrowing through the online credit platform of mobile internet technology. P2P network lending, as a part of P2P, refers to financial activities for online lending through the Internet. Because the funds come from investors, the default of borrowers is bound to affect the profitability of P2P companies and the property security of investors. Therefore, it is particularly important to prevent the borrower from defaulting and the borrower to carry out effective risk control.

Credit risk is the main type of P2P financial risk [11]. It is also called "default risk" in the P2P network lending industry. It means that the borrower has not fulfilled the repayment obligation according to the contract, which led to bad debts on the P2P platform and economic losses to the lender. This means that the borrowers with poor credit fail to fulfil their repayment obligations in accordance with the contract, resulting in the economic losses of P2P network lending platform or investors. At present, there is an important problem in P2P network lending, which faces the risk of greater uncertainty in its development process. The overdue default of P2P loan is becoming more and more difficult to control, so it is of practical significance to strengthen the research on enterprise credit risk management. It helps to reduce the overdue and malicious default caused by personal credit problems of enterprise loan customers. This can improve the efficiency of enterprises and protect effectively the interests of investors. In addition, it can also reduce industry risks and improve the confidence of enterprises. Applying credit scores can effectively improve or solve most of the problems existing in P2P lending platforms. It can standardize non-standardized personal risks with more direct evaluation criteria. Using credit risk score can make the overall credit asset quality become a quantitative indicator, making it easier to judge. Credit risk scoring can also effectively improve the low productivity of the approval process and make the approval more standardized [12].

The research is based on the actual risk management issues in the development of a domestic P2P company and studies the credit risk assessment and management of its borrowers. The data sources of the borrower's credit risk scoring model in this study are the data that the borrower needs to submit when applying for a loan, as well as data processed by some P2P lending platforms. The data source includes personal information of personal data, personal debt situation, default situation and other qualitative and quantitative data which reflect personal credit risk or default risk. The borrower's credit risk scoring model is used to predict the probability of default within a period repayment after the borrower applying for the loan contract again. The credit risk factors were screened and the credit risk was quantified. Then the corresponding credit risk assessment mechanism and policy management measures were proposed. It is effective to protect the financial security of investors and the interests of enterprises, which enhances the competitiveness of enterprises in the development of the industry.

2 RELATED WORKS

With the gradual popularization of P2P Internet financial lending, three major research directions have emerged. One of the research directions is to investigate the reasons for the emergence of the P2P Internet lending industry. The other research focuses on the factors related to the risk of default and another study mainly investigates the performance of P2P Internet lending at a certain risk.

2.1 P2P Internet Finance Lending

In the early stages of P2P Internet lending development, the platform lacked rigid and quantifiable

P2P-related financial data [13]. Therefore, the P2P platform began to require borrowers to provide more detailed information on borrowing. These accumulated application information are then disclosed.

The disclosed agencies included Prosper and Lending Club. Since then, more and more studies have used credit risk assessment models to study and assess default risks. Researchers use core regression to study and forecast the lender's return on investment in Lending Club and Prosper P2P lending data [14].

The P2P loan industry has been in various regions. Many researchers have paid more attention to innovative P2P loan. Ashta and Assadi investigated that Web 2.0 technologies were integrated to support advanced social interactions and associations with lower P2P loan costs [15]. The emergence of P2P Internet loans is a direct response to new social trends. Gregor Dorfleitner studies the determinants of the repayment behavior of the peer-to-peer microfinancing platform which provide loans to international charity lenders [16]. In addition, P2P Internet loans also can reflect a demand for new forms of relationship in the financial sector in the new information age.

2.2 Network Loan Credit Risk Model

There are integrated machine learning algorithms and pre-processing techniques to analyze the determinants of credit risk through Prosper's data. The credit risk of Lending Club borrowers can be predicted by analyzing loan data from Lending Club using the Logit model [17]. Emekter used binary logistic regression models to analyse the factors affecting default rate [13].

There are also factors such as interest rates and amounts [18]. Pope and Sydnor pointed out that race also has a certain impact on the combined power of lending [19]. In addition, some researchers have shown that some information such as social networks, facial features, language features, etc., to some extent, also affect the success rate of Internet loans [20-23]. Besides, Herzenstein and Michels also studied providing only P2P loan websites for borrowers without ID and other information to improve the success rate [24, 25]. A study by Duarte et al. shows that borrowers with credible characteristics will receive a better credit score, but the probability of default is lower [20]. In addition to the social relationships and friendships of loan applicants, the above factors are also important to explain the risk of default, reported by Freedman and Jin and Lin [22, 26]. The results show that when comparing American borrowers, lending club should continue to select the borrowers with high credit rating and attract the borrowers with the best reputation to greatly reduce the risk of default. Berkovich also reported that high-quality loans can provide excess returns [27].

2.3 Online Loan Credit Risk Factor

Gomez and Santor found that group loans offered lower default rates than traditional personal loans by analyzing Canadian microfinance data [28]. A study by Iyer et al. shows that lenders can use hard and soft data about borrowers to assess one-third of credit risk [29]. Lin et al. analyzed social relations and assessed the role of

credit risk in borrowing success rate and default risk. They discovered that the complex social relationships of borrowers are also important factors in determining the success of borrowing and reducing default risk. Lin et al. further reported that the applicant's friendship may increase the likelihood of successful financing and reduce the interest rate of financing loans [22]. Freedman and Jin also examined social relationships and found that they played a more important role in deciding whether to lend. The results show that borrowers with certain social relationships are more likely to obtain loans and receive lower interest rates. At the same time, a number of banks have already established certain risk factors for borrowers to participate in social networks.

A study by Herzenstein et al. shows that the financial strength of borrowers, whether they are listed, the level of publicity, and the attributes of people will affect the possibility of successful financing [30]. Duarte et al.'s further argue that a more trustworthy borrower can visually assess a higher credit score and a lower probability of default [20]. Larrimore et al. demonstrated that borrowers who use extended narratives to specifically describe and quantify vocabulary have a positive impact on the success of capital lending. However, humanized personal data and loan reasons have an impact on the success of capital loans [21]. Qiu et al. further revealed that in addition to personal information and social capital, the loan amount, the highest acceptable interest rate, and other variables such as the loan period set by the borrower also have a significant effect on the success or failure of the fund [31].

Galak et al.'s study further suggests that lenders prefer personal borrowing rather than group borrowing and group borrowers. They also found that lenders were more willing to borrow from their borrowers in terms of gender, occupation, and initials [20]. More interestingly, Gonzalez and Loureiro have similar findings. When recognizing that age can basically represent their repayment ability, the borrower's attractiveness to the lender has no positive impact on the success of the loan. When the lender and the borrower belong to the same gender, the attractiveness of this existence may lead to the failure of the loan.

3 ALGORITHM

3.1 Random Forest Importance Assessment

Considering that there are many features in the data set, how to select the features that have the greatest impact on the results to determine the number of features in the model? There are many similar methods, such as principal component analysis, Lasso and so on. This study used a random forest approach to screen for feature importance.

Random forest (RF) is an integrated machine learning method based on decision tree-based learners. The random forest approach is to build multiple decision trees and then vote to get the final result of the classification.

The RF method is simple and the results are easy to get. It also shows good performance in classification and regression. Therefore, the random forest is also known as "the method of representing the level of integrated learning technology."

Random forests combine tree predictors, so each tree depends on the value of a randomly sampled random vector, and all trees in the forest have the same distribution.

The random forest algorithm is more intuitive, as shown in the 3-2 random forest algorithm diagram.

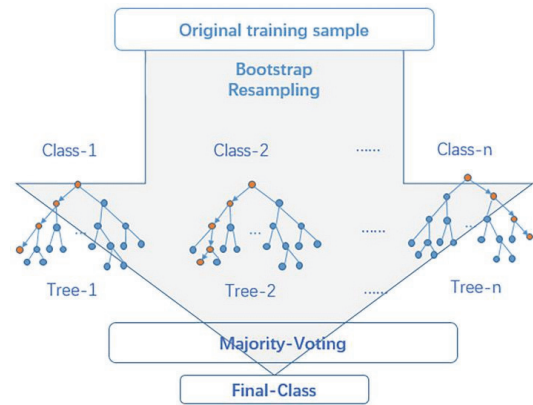


Figure 1 Random forest algorithm schematic

The algorithmic steps of the random forest are summarized:

The first step: using a sampling method with a return, select a certain number of samples from the overall sample set as a training set.

The second step: Generate a decision tree from the sample set obtained in the first step. Do the following at each node generated by the decision tree:

- (1) d features are randomly selected without repetition.
- (2) According to the d features, the sample data sets are sequentially divided, and then the best feature segmentation results can be obtained by using various methods such as the Gini coefficient.
- (3) Repeat the two steps from the first step to the second step, and the number of random forest decision trees is the K value.
- (4) Establish a random forest according to the sample set and use the obtained random forest to predict the test sample set, and use the voting method to determine the predicted result.

The idea of using random forest to evaluate the importance of features is to calculate the contribution value of each feature to each tree in the random forest, then average all the contributions, and finally compare the contributions of each feature.

The metric of the contribution value usually takes one of the Gini indices or out-of-bag data. The formula for calculating the Gini index is:

$$GI_m = \sum_{k=1}^{|k|} \sum_{k \neq k} p_{mk} p'_{mk} = 1 - \sum_{k=1}^{|k|} p_{mk}^2 \quad (1)$$

The meaning of each indicator in the formula: k means that there are k categories; p_{mk} means the proportion of the category k in the node m .

The importance of the feature x_j at the node m is the Gini exponential change before and after the node m branch, and is calculated as follows:

$$VIM_{jm}^{(Gini)} = GI_m - GI_l - GI_r \quad (2)$$

Among them, GI_l and GI_r respectively represent the Gini index of the two new nodes after branching.

When the node where the feature x_j appears in the decision tree i is in the set M , the calculation formula of the importance of x_j in the i -th tree is:

$$VIM_{ji}^{(Gini)} = \sum_{jm} VIM_{jm}^{(Gini)} \tag{3}$$

Assuming that there are n trees in the RF, then the importance of x_j in the n^{th} tree is:

$$VIM_j^{(Gini)} = \sum_{i=1}^n VIM_{ij}^{(Gini)} \tag{4}$$

Finally, the importance scores obtained through normalization are processed. The formula is as follows:

$$VIM_j = \frac{VIM_j}{\sum_{i=1}^C VIM_i} \tag{5}$$

The variable importance score is represented by VIM , and the Gini index is represented by GI . Assuming that there are m features X_1, X_2, X_m , the Gini index score of each feature X_i is now calculated. Features are ranked from high to low according to their importance, and the top N features are selected.

3.2 Logistic aAlgorithm

According to the characteristics of credit risk assessment, we use the logistic regression algorithm to quantify the credit risk. In the application of logistic regression, the first thought is the distribution of Bernoulli effort. From the perspective of probability, "overdue" is a random event. Bernoulli distribution is a kind of discrete distribution, which is used to express the probability of type 0 - 1 events. For overdue and non-overdue loan performance, the formula can be expressed as follows:

$$P(Y = y) = p^y (1 - p)^{1-y} \tag{6}$$

$$y = \begin{cases} 1, & \text{over due} \\ 0, & \text{non overdue} \end{cases} \tag{7}$$

Logistic regression is different for different applicants, and the overdue probability is different. The overdue probability can be expressed as a function:

$$P = f(x_1, x_2, \dots, x_k) \tag{8}$$

where $\{x_1, x_2, \dots, x_k\}$ is the personal qualification of the applicant, and P has the characteristics of being bounded and unobserved.

The logistic formula is:

$$E(p) = f(\beta_0 + \sum \beta_i x_i) \tag{9}$$

$$f(x) = \frac{\exp(x)}{1 + \exp(x)} \tag{10}$$

The overdue status of a group of applicants in the performance period is $\{y_1, y_2, \dots, y_n\}$, and $y_i \in \{0, 1\}$. The likelihood function and log likelihood function are:

$$L(p) = \prod P(Y = y_i) = \prod p^{y_i} (1 - p)^{1-y_i} \tag{11}$$

$$l(p) = \log(L(p))$$

$$l(p) = \sum (y_i \log(p) + (1 - y_i) \log(1 - p)) = \sum (y_i (\beta_0 + \sum \beta_i x_{ij}) - \log(1 + \exp(\beta_0 + \sum \beta_i x_{ij}))) \tag{12}$$

The parameter estimation formula is as follows:

$$\bar{p} = \arg \max l(p)$$

$$\bar{p} = \frac{\sum y_i}{n} \tag{13}$$

The parameter estimation formula is as follows:

$$\frac{\partial l}{\partial \beta_q} = \sum \left(y_i - \frac{1}{\exp(-\beta_0 - \sum \beta_i x_{ij})} \right) x_{iq} \tag{14}$$

Estimate the β_q by the gradient descent method, the formula is as follows:

$$\beta_q^{r+1} = \beta_q^r - h\delta \tag{15}$$

$$\delta = \frac{\partial l}{\partial \beta_q} \Big|_{\beta_q = \beta_q^r} \tag{16}$$

4 EXPERIMENTS

This study collected some of the loan records from June 2013 to October 2017 from the "TRUST SAVING" platform. Research is conducted using Python 2.7, an object-oriented, interpreted computer programming language.

4.1 Data Preprocessing

First, the acquired data features are filtered and characterized. Data is formatted and missing values are processed. In addition, the missing value is introduced into the model as a separate feature.

Then, feature data is supervised and consolidated by chi square method. Continuous feature variables are discretized. At the same time, discrete feature states are more combined and reduced. The effect of extreme values and data with meaningless fluctuations on prediction

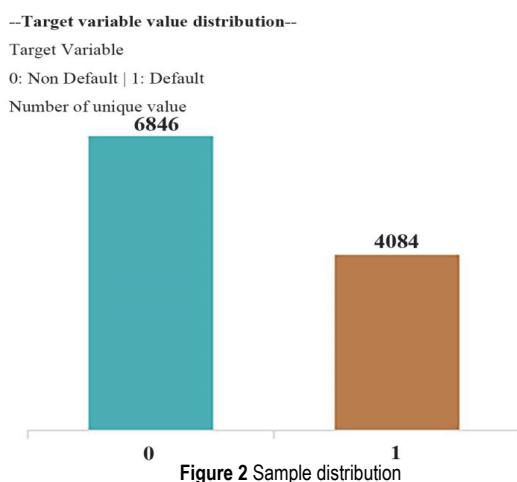
results is reduced to increase the stability and robustness of the model.

Last, a total of 26 features were selected from the variable analysis object of the study. Features and number of feature classes are shown in Tab. 1.

Table 1 Characterization of the data set

Feature number	Features	Feature class number
1	Highest education	5
2	Residence type	4
3	Unit type	5
4	Job type	4
5	Internal debt ratio	5
6	External liability	6
7	External debt ratio	5
8	Age	6
9	Day of birth	5
10	Customer signing bank card	5
11	Product name	4
12	Product period	4
13	Entry time (days)	5
14	Family know	3
15	Household register province	5
16	Household register area	5
17	Household register city	5
18	Month of birth	3
19	Sales Department	3
20	Salary method	3
21	Gender	2
22	Total debt ratio	5
23	Risk score	6
24	Monthly income	6
25	Year of birth	4
26	Marital status	4

Throughout the loan life cycle, it can be divided into loans that have not been reviewed, loan approvals, loan repayments, and loans that have been repaid in four states. Loans in repayments can be divided into non-default loans and default loans. Loans with repayments and loans that have been defaulted are used for modelling. According to the user's "repayment status" feature, the target variable is determined. If the loan has been successfully repaid (no default), the value is 0; if there is overdue loan (default), the value is 1. Finally, 10930 transaction data with repayments were selected as the sample set. Among them 6846 cases were successfully repaid, accounting for 62.6% of the total number of samples. 4084 cases were overdue, accounting for 37.4% of the total number of samples. The overall "0 - 1" distribution of the sample is shown in the sample distribution diagram of Fig. 2.



4.2 Feature Selection

This study ranks features from high to low using the Gini index evaluation method, according to the importance of the features. According to the importance score of the variables, the features with higher importance are selected.

This paper uses the random forest Gini index algorithm to analyze the feature data to obtain the feature importance results. We retain the three digits after the decimal point of the result, and sort them from highest to lowest according to their importance, while calculating the cumulative importance. The six characteristics of "gender", "total debt ratio", "risk score", "monthly income", "year of birth" and "marital status" have been deleted because they are obviously of low importance.

Based on the above analysis results, a total of 20 characteristics were selected to participate in credit risk forecast, including "Highest education", "Residence type", "Unit type", "Job type", "Internal debt ratio", "External liability", "External debt ratio", "Age", "Day of birth", "Customer signing bank card", "Product name", "Product period", "Entry time (days)", "Family know", "Household register province", "Household register area", "Household register city", "Month of birth", "Sales Department", "Salary method". The finally determined features are renumbered, as shown in Tab. 2.

4.3 Result

We distinguish data sets from test sets and use logistic regression model to quantify the credit risk of borrowers. Finally, the prediction results are compared with the actual results to verify the effectiveness of the credit risk assessment model.

Table 2 Final screening feature summary table

Features	Importance	Cum_Importance	Participation model
Household register area	0.156721	0.156721	1
Product name	0.153565	0.310286	1
Household register city	0.055818	0.366104	1
Product period	0.050435	0.416539	1
Family know	0.046723	0.463262	1
Internal debt ratio	0.043651	0.506913	1
External liability	0.039034	0.545947	1
Age	0.038635	0.584582	1
Entry time	0.038542	0.623123	1
External debt ratio	0.037582	0.660705	1
Day of birth	0.036527	0.697233	1
Customer signing ban card	0.033395	0.730628	1
Residence type	0.03146	0.762088	1
Unit type	0.030763	0.792851	1
Highest education	0.030523	0.823374	1
Job type	0.029756	0.85313	1
Household register province	0.029102	0.882232	1
Month of birth	0.023476	0.905707	1
Sales department	0.020211	0.925918	1
Salary method	0.020194	0.946112	1
Year of birth	0.020167	0.96628	0
Marital status	0.020148	0.986428	0
Gender	0.013572	1	0
Totaldebt ratio	0	1	0
Risk score	0	1	0
Monthly income	0	1	0

The hybrid matrix is shown in the hybrid matrix diagram of Fig. 3, where the first quadrant is the borrower whose model is predicted to be non-default and the

borrower that is actually not defaulting is 1440, the second and third quadrants are incorrectly predicted by the model. The numbers are 232 and 488, the fourth quadrant indicates that the model predicts that the default is also 535. The accuracy of the model is that the ratio of the predicted number to the total number is 73.3%, and the accuracy is higher.

4.4 Analysis

This study uses domestic P2P loan platform data to determine whether the borrower's credit risk can be effectively assessed. If the listing information of the P2P platform can effectively assess the credit risk of the borrower, the market is sustainable, and Chinese regulators restrict the P2P lending platform to information intermediaries without hindering their operational capabilities. First, we used the full loan data and use existing algorithms for feature filtering. Then, we compared the predicted results with the actual results to verify the validity of the credit risk management model.

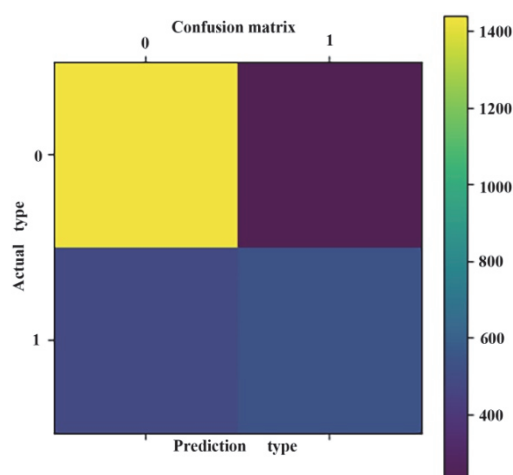


Figure 3 Confusion matrix

4.5 Recommendation Effect

In the theoretical part of the thesis, we discussed the credit risk assessment and credit risk management methods through comprehensive review and analysis. In addition, the existence of credit risk models is discussed. It also briefly introduces the methods used in the credit risk modelling process, especially the development methods used in the credit risk management model.

One of the purposes of this study is to establish a credit risk management model based on logistic regression algorithm, which will separate relatively high-quality customers with low credit risk and reduce credit losses in actual production. This analysis uses 10930 loan data from P2P in China. In the process of building the model, based on 26 domain definition standards, 20 model variables are selected to participate in credit risk quantification through feature structure and feature analysis. Finally, 73.3% of the customers are classified by the model, which proves the validity of the model.

5 CONCLUSION

In this paper, the logistic regression algorithm is applied to the credit risk management to improve the recognition ability of the loan object with high default probability and improve the overall loan quality. We collected the real data set from one of the domestic P2P platforms for feature processing and then performed model analysis. The model test results also show that the model has good performance in quantifying the borrower's credit risk.

Acknowledgements

This paper was supported by Wonkwang University in 2021.

6 REFERENCES

- [1] Yan, Y., Lv, Z., & Hu, B. (2018). Building investor trust in the P2P lending platform with a focus on Chinese P2P lending platforms. *Electronic Commerce Research*, 18(2), 203-224. <https://doi.org/10.1007/s10660-017-9255-x>
- [2] Ma, L., Zhao, X., Zhou, Z., & Liu, Y. (2018). A new aspect on P2P online lending default prediction using meta-level phone usage data in China. *Decision Support Systems*, 111, 60-71. <https://doi.org/10.1016/j.dss.2018.05.001>
- [3] Huang, R. H. (2018). Online P2P lending and regulatory responses in China: opportunities and challenges. *European Business Organization Law Review*, 19(1), 63-92. <https://doi.org/10.1007/s40804-018-0100-z>
- [4] Gharib, M., Malekimajid, M., & Movaghar, A. (2018). Slopcloud: an efficient solution for locality problem in peer-to-peer cloud systems. *Algorithms*, 11(10), 150-166. <https://doi.org/10.3390/a11100150>
- [5] Liao, H. H., Chen, T. K., & Lu, C. W. (2009). Bank credit risk and structural credit models: Agency and information asymmetry perspectives. *Journal of Banking & Finance*, 33(8), 1520-1530. <https://doi.org/10.1016/j.jbankfin.2009.02.016>
- [6] Amiram, D., Beaver, W. H., Landsman, W. R., & Zhao, J. (2017). The effects of credit default swap trading on information asymmetry in syndicated loans. *Journal of Financial Economics*, 126(2), 364-382. <https://doi.org/10.1016/j.jfineco.2016.10.001>
- [7] Wang, K., Tsai, S. B., Du, X., & Bi, D. (2019). Internet Finance, Green Finance, and Sustainability. *Sustainability*, 11(14), 3856-3861. <https://doi.org/10.3390/su11143856>
- [8] Yang, D., Chen, P., Shi, F., & Wen, C. (2018). Internet finance: Its uncertain legal foundations and the role of big data in its development. *Emerging Markets Finance and Trade*, 54(4), 721-732. <https://doi.org/10.1080/1540496X.2016.1278528>
- [9] Yoon, Y., Li, Y., & Feng, Y. (2019). Factors affecting platform default risk in online peer-to-peer (P2P) lending business: an empirical study using Chinese online P2P platform data. *Electronic Commerce Research*, 19(1), 131-158. <https://doi.org/10.1007/s10660-018-9291-1>
- [10] Liu, Z., Shang, J., Wu, S. Y., & Chen, P. Y. (2020). Social collateral, soft information and online peer-to-peer lending: A theoretical model. *European Journal of Operational Research*, 281(2), 428-438. <https://doi.org/10.1016/j.ejor.2019.08.038>
- [11] Havrylchyk, O. & Verdier, M. (2018). The financial intermediation role of the P2P lending platforms. *Comparative Economic Studies*, 60(1), 115-130. <https://doi.org/10.1057/s41294-017-0045-1>
- [12] Feng, X., Xiao, Z., Zhong, B., Qiu, J., & Dong, Y. (2018).

- Dynamic ensemble classification for credit scoring using soft probability. *Applied Soft Computing*, 65, 139-151. <https://doi.org/10.1016/j.asoc.2018.01.021>
- [13] Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics*, 47(1), 54-70. <https://doi.org/10.1080/00036846.2014.962222>
- [14] Guo, Y., Zhou, W., Luo, C., Liu, C., & Xiong, H. (2016). Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research*, 249(2), 417-426. <https://doi.org/10.1016/j.ejor.2015.05.050>
- [15] Ashta, A. & Assadi, D. (2009). Do social cause and social technology meet? impact of web 2.0 technologies on peer-to-peer lending transactions. *Cahiers du CEREN*, 29, 177-192. <https://ssrn.com/abstract=1281373>
- [16] Dorfleitner, G. & Oswald, E. M. (2016). Repayment behavior in peer-to-peer microfinancing: Empirical evidence from Kiva. *Review of Financial Economics*, 30, 45-59. <https://doi.org/10.1016/j.rfe.2016.05.005>
- [17] Serrano-Cinca, C., Gutiérrez-Nieto, B., and López-Palacios, L. (2015). "Determinants of Default in P2P Lending." *PLOS One* 10(10), e0139427. <https://doi.org/10.1371/journal.pone.0139427>
- [18] Puro, L., Teich, J. E., Wallenius, H., & Wallenius, J. (2010). Borrower decision aid for people-to-people lending. *Decision Support Systems*, 49(1), 52-60. <https://doi.org/10.1016/j.dss.2009.12.009>
- [19] Pope, D. G. & Sydnor, J. R. (2011). What's in a Picture? Evidence of Discrimination from Prosper. com. *Journal of Human Resources*, 46(1), 53-92. <https://doi.org/10.3368/jhr.46.1.53>
- [20] Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies*, 25(8), 2455-2484. <https://doi.org/10.1093/rfs/hhs071>
- [21] Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., & Gorski, S. (2011). Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, 39(1), 19-37. <https://doi.org/10.1080/00909882.2010.536844>
- [22] Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, 59(1), 17-35. <https://doi.org/10.1287/mnsc.1120.1560>
- [23] Riggins, F. J. & Weber, D. M. (2017). Information asymmetries and identification bias in P2P social microlending. *Information Technology for Development*, 23(1), 107-126. <https://doi.org/10.1080/02681102.2016.1247345>
- [24] Herzenstein, M., Sonenshein, S., & Dholakia, U. M. (2011). Tell me a good story and I may lend you money: The role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research*, 48(SPL), S138-S149. <https://doi.org/10.1509/jmkr.48.SPL.S138>
- [25] Michels, J. (2012). Do unverifiable disclosures matter? Evidence from peer-to-peer lending. *The Accounting Review*, 87(4), 1385-1413. <https://doi.org/10.2308/accr-50159>
- [26] Freedman, S. & 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>
- [27] Berkovich, E. (2011). Search and herding effects in peer-to-peer lending: evidence from prosper. com. *Annals of Finance*, 7(3), 389-405. <https://doi.org/10.1007/s10436-011-0178-6>
- [28] Galak, J., Small, D., & Stephen, A. T. (2011). Microfinance decision making: A field study of prosocial lending. *Journal of Marketing Research*, 48(SPL), S130-S137. <https://doi.org/10.1509/jmkr.48.SPL.S130>
- [29] Asim, K. D. (2009). Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending? Retrieved from <https://www.hks.harvard.edu/publications/screening-new-credit-markets-can-individual-lenders-infer-borrower-creditworthiness>
- [30] Herzenstein, M., Dholakia, U. M., & Andrews, R. L. (2011). Strategic herding behavior in peer-to-peer loan auctions. *Journal of Interactive Marketing*, 25(1), 27-36. <https://doi.org/10.1016/j.intmar.2010.07.001>
- [31] Qiu, J., Lin, Z., & Luo, B. (2011). *Effects of borrower-defined conditions in the online peer-to-peer lending market*. https://doi.org/10.1007/978-3-642-29873-8_16

Contact information:

Dongmei LI, PhD candidate
College of Business Administration, Wonkwang University,
No. 460, Iksandae-ro, Iksan, Jeonbuk 54538, Korea
E-mail: atlidongmei@163.com

Sanggyun NA, PhD, Professor
(Corresponding author)
College of Business Administration, Wonkwang University,
No. 460, Iksandae-ro, Iksan, Jeonbuk 54538, Korea
E-mail: nsghy@wku.ac.kr

Tao DING, PhD candidate
College of Business Administration, Wonkwang University,
No. 460, Iksandae-ro, Iksan, Jeonbuk 54538, Korea
E-mail: scyxtv@163.com

Congchong LIU, PhD candidate
College of Business Administration, Wonkwang University,
No. 460, Iksandae-ro, Iksan, Jeonbuk 54538, Korea
E-mail: liucc06@163.com