

The P2P Risk Assessment Model Based on the Improved AdaBoost-SVM Algorithm

Jianhui Yang, Dongsheng Luo

School of Business and Management, South China University of Technology, Guangzhou, China

Email: luo.dongsheng@foxmail.com

How to cite this paper: Yang, J. H., & Luo, D. S. (2017). The P2P Risk Assessment Model Based on the Improved AdaBoost-SVM Algorithm. *Journal of Financial Risk Management*, 6, 201-209.

<https://doi.org/10.4236/jfrm.2017.62015>

Received: May 31, 2017

Accepted: June 24, 2017

Published: June 27, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The improved AdaBoost-SVM algorithm is used to classify the safety and the risk from the Peers-to-Peers net loan platforms. Since the SVM algorithm is hard to deal with the rare samples and its training is slow, rule sampling is used to reduce the classify noise. Then, with the combinations of learning machine, P2P risks can be identified. The result shows that IAdaBoost algorithm can improve the risk platform classification accuracy. And the error of classification can be controlled in 5%.

Keywords

Peers-to-Peers, AdaBoost, SVM, The Combinations of Learning Machine, Rule Sampling

1. Introduction

In recent years, owing to the development of the domestic Internet financial business, the traditional financial industry has to reform rapidly. With the global integration process intensified, modern finance is showed a complex form. The complexity of the financial system makes the risk spread faster and faster, and the scope of the impact between the platforms is also growing.

As an important form of Internet finance, Peers-to-Peers loan, the risk infection and measurement are also of concern. Credit risk, which is the main problem faced by the P2P market, is largely associated with the fuzziness of risk factors. Measuring the credit risk is the inherent risk management requirement of P2P and bank market, and it is also an important basis for effective prevention of financial risk.

Domestic scholars on the network lending (P2P) were focused on the discussion of its platform operation mode and development trends, as well as network lending (P2P) industry risk control and risk management issues. From the new

perspective of “platform risk”, we have expanded the research of P2P domain (Ye, Li, & Xu, 2016). Wang mainly analyzes into the P2P network lending platform for risk regulation and prevention analysis and policy considerations (Wang, 2016). Liu analyzes the risk characteristics of China’s P2P industry from three different perspectives of lenders, investors and platforms, and constructs an improved debtor risk assessment model (Liu, 2013). Luo Chunyu, when studies the network P2P (P2P) risk assessment, builds quantitative methods and constructs the investor composition analysis model, as well as the borrower credit risk analysis model and multi-information source loan assessment model, supporting the investors to provide decision (Luo, 2012).

The foreign research of P2P network lending platform, mostly analyzes the main behavior of borrower transactions and platform development trends. Considering the current research on the credit characteristics and loan success factors of the main body of the transaction, we mainly analyze the risk problems and the dislocation of the network, and the lack of supervision. This is why China is not as good as Britain and the United States with complete and transparent credit system. What’s more, their network lending system (P2P) is developed into the scope of supervision. Compared to foreign, our network lending (P2P) still has to be improved.

In the context of this difference between domestic and foreign, P2P credit risk measurement and evaluation depends on the data screening and model establishment. In the machine learning algorithm, the commonly used algorithm models include perceptron, K-nearest neighbor, Decision Tree, Logistic regression, Support Vector Machine, AdaBoost algorithm, Hidden Markov, Conditional Random Field and so on. The machine learning algorithm is applied to the P2P risk assessment, which can effectively improve the evaluation and classification model. The traditional support vector machine algorithm training problem, in essence, is a convex secondary programming problem. Using the P2P risk measurement and risk assessment, we get the P2P platform indicators data, P2P network loan platform risk division, so as to filter the problem platform.

In view of the simple SVM algorithm, the sample set is required to be high, and the combined learning method generates multiple base classifiers by splitting learning and assembling them according to a certain strategy. The result of the combined classifier depends on the single base classifier. As a result of the determination, the error of the classification can be effectively reduced by the combination characteristics of the various base classifiers.

Boosting algorithm is a commonly used statistical learning method, which is widely used and effective. In the classification problem, it improves the classifier performance by changing the weight of training samples, combining multiple classifiers, and classifying these classifiers linearly. Applied to SVM, it can be enhanced for the separation and division of the sample set. It can change the probability distribution of training data, and call a weak learning algorithm for a series of training data distributions to learn a series of classifiers.

2. Statement of the Problem

Because of the huge risk of P2P platform, we focus on how to build the model and measure the P2P risk. As a result, the following article analyzes the P2P risk source and credit evaluation index system, and solves the risk assessment of P2P to avoid investing in bad P2P platform.

2.1. P2P Risk Assessment and Credit Index System

P2P network loan platform has faced many aspects of the risk source, including the platform itself and the risk of the risk of infection between platforms. Under the influence of many risk factors, the development and growth of P2P platform will be seriously constrained.

The current net loan is rating and there is no recognized standard and qualification; each rating agencies consider the dimensions and standards, and cannot really reflect the level of a platform. **Table 1** shows the evaluation index system used by the third-party rating agencies of each loan platform. For example, 360 large data research institute focus on the background strength of the platform; Dagong international rating report more focused on the debtor solvency of the inspection; Academy of Social Sciences Institute of Finance's evaluation system is focused on the level of risk control platform; Home of the comprehensive evaluation system set indicators and not for security (Yu, 2015).

The evaluation of the network borrowing (P2P) platform can be properly referred to the commercial bank credit rating method. **Table 2** below shows a comparison of the rating system for commercial banks. For the commercial bank credit rating, mainly the United States Federal Financial Institutions Regulatory Commission CAMELS rating system, the China Banking Regulatory Commission issued a joint-stock commercial bank risk rating system and Moody's as

Table 1. Evaluation index system adopted by the third-party rating agency of net credit platform.

Rating Agencies	Index system	Weight setting	Rating Agencies
360 Big Data Research Institute and Renmin University of China	Background Strength (30%), Platform Risk Control (25%), Operational Capacity (20%), Information Disclosure (15%), User Experience (10%)	The detailed weight is shown at left	360 Big Data Research Institute and Renmin University of China
Chinese Academy of Social Sciences Institute of Finance and Jinniu Financial Network	Basic indicators, operational capacity, risk control, social responsibility, information disclosure	Analytic Hierarchy Process (rating using the percentage system, 90 points or more is AAA, 80 points to 90 points is AA, 60 points to 80 points is A.)	Chinese Academy of Social Sciences Institute of Finance and Jinniu Financial Network
Net Loan Home	Trading volume (10%), revenue (10%), popularity (18%), income (6%), leverage (6%), liquidity (5%), dispersion (16%), transparency (11%), Brand (18%)	Analytic Hierarchy Process	Net Loan Home
Tiger Financial and Palm Tree Planning	Background strength, management team, wind control capability, partners guarantee strength, IT system support, customer experience, Operational capacity, major issues	Through field research, subjective rating.	Tiger Financial and Palm Tree Planning

Source: Net loan home, Yu Jiamin: network lending (P2P) platform quantitative monitoring research.

Table 2. Comparison of the rating system of commercial banks.

	United States Federal Financial Institutions Regulatory Commission	China Banking Regulatory Commission	Moody's	Standard & Poor's	Dagong International	In the integrity
1	Capital adequacy	Capital adequacy	Capital adequacy	Capital adequacy	Capital adequacy	Financial factors
2	Asset quality	Asset security	Asset quality	Credit Risk and Management	Asset quality	Risk management
3	Management	Management	Company structure	Company structure	Company structure	—
4	Profit level	Profit level	Profit level	Profit level	Profit level	—
5	Fluidity	Fluidity	Fluidity	Fluidity	Fluidity	—
6	Social sensitivity	Social sensitivity	Macro situation	Macroeconomic and Industry risk	Operating environment	External environment
7	—	—	Regulatory environment	Market risk and its management	Operating value	Operational factors
8	—	—	—	Management and its strategy	—	—

Source: Qi Fei (2012) Yu Jiamin: Network lending (P2P) platform quantitative monitoring research.

the representative of the rating agencies also have a mature commercial bank rating system. Six factors such as capital adequacy, asset quality and management level are summarized by the rating system adopted by regulators and international and domestic authorities, as shown in **Table 2**. It is of great significance for the research of this paper to summarize the rating system of commercial banks, especially the regulatory authorities (Li, Liu, & Chen, 2015).

2.2. SVM Algorithm Improvement

SVM technology mentioned showed below is the base classifier under the P2P network loan platform. The advantage of this method is that the number of classifiers is small, and the algorithm is simple and complicated (Ju, Wang, & Yao, 2012). But there are some drawbacks to this approach:

- (1) Base classifier learning needs to train all samples, its training is slow.
- (2) Poor treatment of rare classify.

Taking into account the above problems, the following selection of sampling is training methods. Training a data set with a subset of the samples can effectively avoid repetitive learning of the entire sample of the base classifier. Its advantages are as below:

- (1) The basis of the classifier to repeat the study only part of the training sample, its training speed can be effectively promoted
- (2) Sampling training covers most of the sample data, it can avoid the classifier to ignore the rare class phenomenon.

Therefore, P2P platform classification also uses a similar sampling training method to avoid the special platform data caused by the training set of unbalanced problems.

In this paper, AdaBoost is applied to SVM classification, and the sample set of each classifier is extracted from the original data set, and the improved AdaBoost-SVM classifier is obtained by multiple iterations.

2.3. AdaBoost Algorithm Detail

Now, $\{(x_i, y_i) | i = 1, 2, \dots, N\}$ represents a collection of N training samples. In the AdaBoost algorithm, the accuracy of the base classifier is closely related to its error rate. The initial sample is equal, and in each subsequent iteration, AdaBoost adjusts the weight on each sample, calculates the error rate of the classifier on the training set, and corrects the probability distribution of the training set.

Algorithm: Enters the sequence of N labeled instances, the distribution D on the N instances, such as $\langle (x_1, y_1), \dots, (x_n, y_n) \rangle$. The algorithm of the training base classifier, and the number of iterations T (Dong, Geng, & Zhou, 2007; Joshi et al., 2002; Ju, Wang, & Yao, 2012; Tan, Steinbach, & Kumar, 2006).

(1) Initialization: initialize the same weight for each sample: $\frac{1}{N}$;

(2) Adjust the distribution: $P_t = \frac{w_t}{\sum_{i=1}^D w_t}$;

(3) Passing the distribution to the base classifier training model, returning the prediction: $x \rightarrow [0, 1]$;

(4) Calculate the prediction error rate: $\varepsilon_t = \sum_{i=1}^N P_t |h_t x_i - y_i|$;

(5) the Importance of calculating the base classifier: $\alpha_t = \frac{\varepsilon_t}{1 - \varepsilon_t}$;

(6) Calculate the new weight vector: $w_{t+1} = w_t \alpha_t^{1 - |h_t x_i - y_i|}$.

In addition, IAdaBoost algorithm is based on the idea of AdaBoost algorithm, in order to avoid the base classifier to ignore the rare class, the initial weight of the sample with the sample size of the class to mark, to get a balanced sample classifier (Chew, Crisp, & Bogner, 2000; Wang & Le, 2005).

3. Outcome of Practice

Empirical data is from the Network Loan Home Platform (<http://www.wdzj.com/>), statistics from the September 21, 2016 to the February 21, 2017. It is a total of 6 months of P2P network loan platform data. **Table 3** below is the data classification table for the sample set. Data sets are all equally divided. Obviously, only 6 months is not enough for the training. But with the sampling, we re-established a sufficient sample set. The training set is the first five months of the data, the test set is the last month. Respectively, when using SVM, AdaBoost-SVM and IAdaBoost-SVM three algorithms to test, we compare the correct classification of the problem platform (the closed platform replaces with 1, the normal platform replaces with 0).

The results of IAdaBoost-SVM. SVM and AdaBoost-SVM are compared. The parameters σ are taken as the AdaBoost classifier with the fixed parameter

value, the parameter value is 6, the penalty parameter C is 100, the dimension of the dimension is 507, the number of iterations of the IAdaBoost algorithm and the AdaBoost algorithm is 5 (Li, Liu, & Chen, 2015).

3.1. Correct Classification Chart

From the classification results in **Figure 1** and **Figure 2**, we can see that in the range of small samples, IAdaBoost improved algorithm is higher than AdaBoost and SVM classification accuracy. At the same time, when the sample set reaches more than 1300 when the number of samples, the combination of learning performance is more excellent, in some cases, the test set the correct rate of about 90%. This allows us to correctly implement P2P risk measurement and risk forecasting.

Of course, we can see from the figure, IAdaBoost algorithm to improve the effect of rare data sets more effective.

3.2. Predictive Effect

As can be seen from **Figure 3** and **Figure 4**, under normalized conditions, the

Table 3. SVM, AdaBoost-SVM and IAdaBoost-SVM SampleSet (Divided into 5 Samples).

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Each month	20160921	20161021	20161121	20161221	20170121
Sample Number	507	507	507	507	507

The Source: Network Loan Home Platform (<http://www.wdji.com/>).

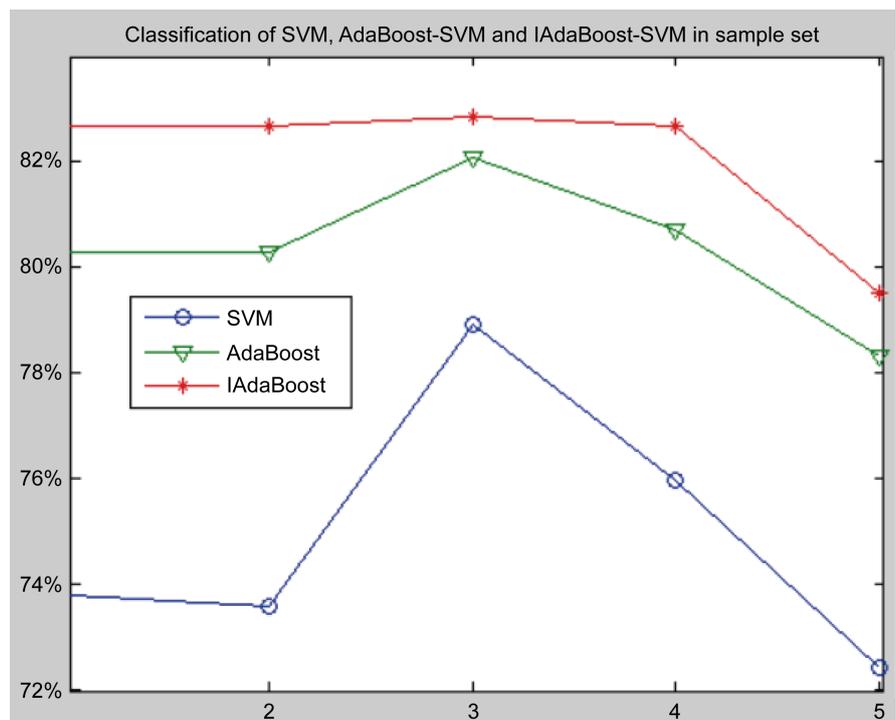


Figure 1. Classification of SVM, AdaBoost-SVM and IAdaBoost-SVM in sample set. Source: Network loan home platform (<http://www.wdji.com/>).

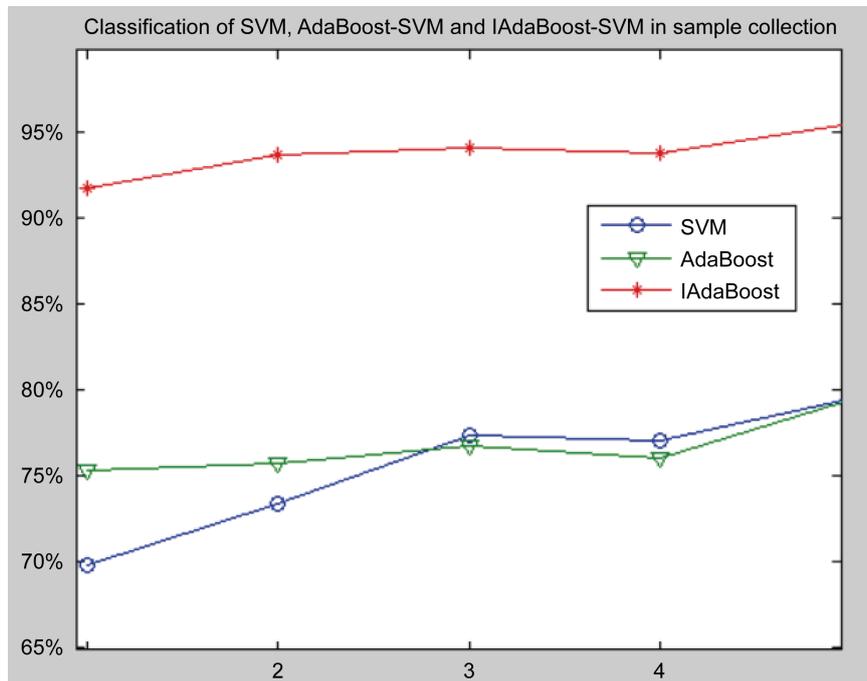


Figure 2. Classification of SVM, AdaBoost-SVM and IAdaBoost-SVM in the sample collection. Source: Network loan home platform (<http://www.wdji.com/>).

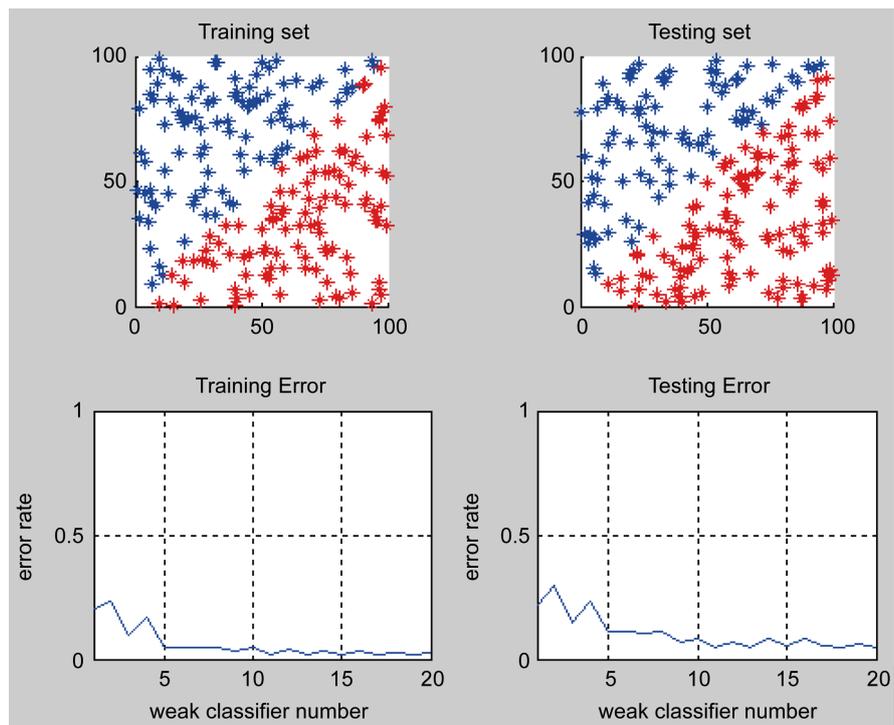


Figure 3. Scatter plot and error rate of the sample set A and test set A under the normalized AdaBoost algorithm. Source: Network loan home platform (<http://www.wdji.com/>).

simulated normal platform and the problem platform can be roughly super-plane classification (blue for the normal platform, red for the simulation of the problem platform). The combination learner can effectively control the error

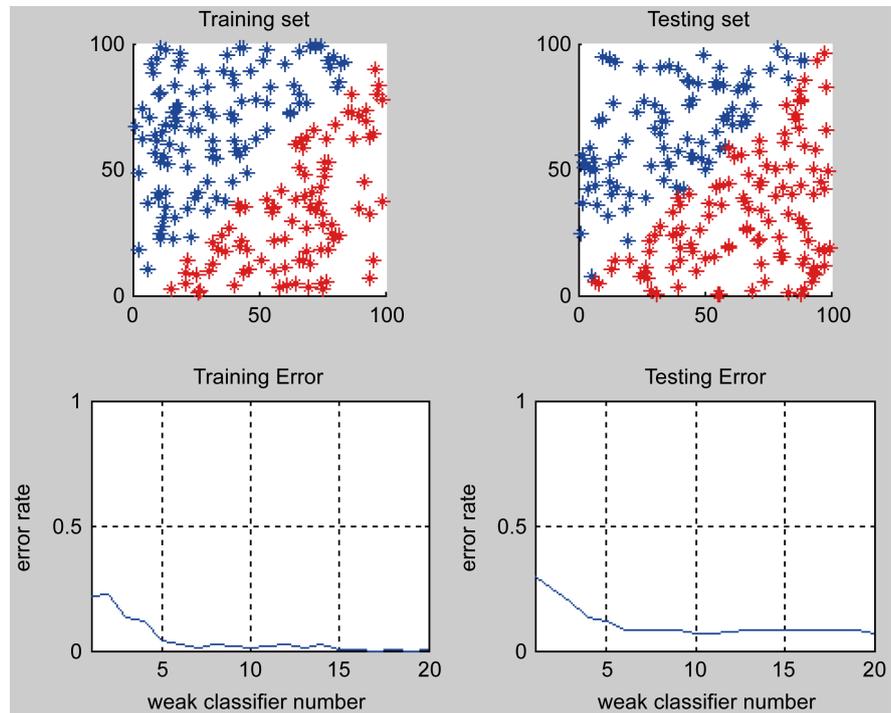


Figure 4. Scatter plot and error rate of the sample set B and test set B under the normalized IAdaBoost algorithm. Source: network loan home platform (<http://www.wdzt.com/>).

Table 4. Base classifier and its weight ratio.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Each month	20160921	20161021	20161121	20161221	20170121
Sample Number	507	507	507	507	507

Source: Network loan home platform (<http://www.wdzt.com/>).

rate within 5% of the learning process. The final base classifier and its weight are shown in **Table 4**. Boost 1, Boost 2 and Boost 5 have higher weights, more than 20%; the rest is lower weight.

4. Conclusion

The IAdaBoost algorithm proposed in this paper not only reduces the training sample, cuts the training range, deals with the unbalanced sample category, but also removes some of the noise data and selects the reliable sample points for training. In addition, the initialization of the improved algorithm can improve the weight of the rare samples, which is beneficial to the correct classification of rare samples. Application of the P2P network loan platform risk assessment can effectively screen out the problem platform, so as to carry out risk management. Of course, AdaBoost-SVM model also has its shortcomings. Sample sets and training set of data should be more detailed, and there is still room for improvement of sampling methods. In addition, the weights of the initial classification of the algorithm can be preprocessed to improve the processing speed of the model risk calculation.

References

- Chew, H.-G., Crisp, D. J., Bogner, R. E. et al. (2000). Target Detection in Radar Imagery Using Support Vector Machines with Training Size Biasing. In *Proceedings of the Sixth International Conference on Control, Automation, Robotics and Vision*, Singapore.
- Dong, L. H., Geng, G. H., & Zhou, M. Q. (2007). Design of Text Automatic Classifier Based on Boosting Algorithm. *Computer Applications*, No. 2, 384-386.
- Joshi, M. V., Agarwal, R. C., & Kumar, V. (2002). Predict Rare Classes: Can Boosting Make Any Weak Learner Strong? In *Proceedings of the Eighth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD2002)*, Edmonton, Canada.
- Ju, X., Wang, H., & Yao, H. L. (2012). Combinatorial Classifier Based on Boosting Support Vector Machine. *Journal of Hefei University of Technology*, No. 10, 1220-1222.
- Li, Y. J., Liu, X. X., & Chen, P. (2015). Improved AdaBoost Algorithm and SVM Combinatorial Classifier. *Computer Engineering and Applications*, 44-32.
- Liu, Z. T. (2013). *China's P2P Network Credit Risk Assessment*. Nanning: Guangxi University.
- Luo, C. Y. (2012). *P2P Network Lending in the Investment Decision Model*. Dalian: Dalian University of Technology.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining I: Posts & Telecom Pushers Inc.*
- Wang, F. (2016). China's P2P Network Lending Platform Risk Regulation and Prevention. *China Circulation Economy*, 121-127.
- Wang, Y. Z., & Le, S. B. (2005). Multiboot-Based Minimum Classification Error Algorithm. *Small Microcomputers*, No. 11, 1948-1950.
- Ye, Q., Li, Z. Q., & Xu, W. H. (2016). Study on Risk Identification of P2P Network Borrowing Platform. *Accounting Research*, 38-45.
- Yu, J. M. (2015). *Network Lending (P2P) Platform Quantitative Monitoring Research*. Guangzhou: South China University of Technology.



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.
 A wide selection of journals (inclusive of 9 subjects, more than 200 journals)
 Providing 24-hour high-quality service
 User-friendly online submission system
 Fair and swift peer-review system
 Efficient typesetting and proofreading procedure
 Display of the result of downloads and visits, as well as the number of cited articles
 Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jfrm@scirp.org