

Risk Monitoring and Early Warning of P2P Network Loan Platform Based on Machine Learning

Xiaokui Zhao*

Qinghai Normal University, Qinghai, China

304918110@qq.com

**corresponding author*

Keywords: Machine Learning, P2P Network, Lending Platform, Risk Monitoring

Abstract: With the rapid development of Internet finance, P2P online lending platform (LP) has also been growing rapidly. However, due to its own many risks, China's P2P online loan industry is facing huge challenges. Therefore, this paper intends to use machine learning related algorithms to predict and prevent the risk of the LP. This paper takes P2P network LP as the research object, and uses data analysis, case comparison and summary methods to sort out and discuss the current situation of the industry. This paper mainly uses the survey method and the analytic hierarchy process to study the suggestions of both lenders and borrowers and the norms of the platform. The experimental results show that about 24% of borrowers and lenders believe that integrity is the basis for maintaining interests. In order to reduce the risk of the LP, it is necessary to rely on the cooperation between the borrower and the borrower and the national regulation. Therefore, the risk warning of the LP is also an important part.

1. Introduction

In recent years, with the rapid development of the Internet, more and more P2P LPs have emerged. At present, the domestic financial market and regulatory mechanism are not perfect, and the construction of laws and regulations is not perfect, so there are often loan crises. Therefore, in order to better monitor and warn the risks of network loans and ensure the stable operation of social and economic order, it is very necessary. In this context, it is of great significance to study the design and implementation of credit system based on machine learning technology.

There are many theoretical achievements on machine learning and online LP. For example, some people say that P2P online lending market has become an important gathering place of Internet financial risks. It not only disturbs the order of China's financial market, but also causes unemployment of online LP practitioners and damage to the interests of investors [1-2]. Some scholars have also built an effective P2P online loan platform (LP) risk early warning model by

integrating mass data collection, Spark distributed platform computing, machine learning modeling and other big data technologies [3-4]. In addition, some scholars used big data web crawler technology to collect the public data of the third-party online loan website platform, and used machine learning model to warn the illegal fund-raising risk of the online LP [5-6]. Therefore, it can be seen from these studies that the loan network platform has a high utilization rate in this century, and it is inevitable to use modern technology to carry out risk early warning.

In this paper, machine learning is first studied, and its basic description is given, as well as the combination of online lending. Secondly, the P2P network LP is analyzed and systematically constructed. Then the basic model of risk early warning mechanism and the construction of risk evaluation index system are proposed. Finally, through data surveys and questionnaires, the problems and suggestions of the LP are discussed and relevant conclusions are drawn.

2. Risk Monitoring of P2P Network LP Based on Machine Learning

2.1. Machine Learning

Machine learning is to acquire knowledge information by collecting, sorting and analyzing existing data in the process of simulating biology, using computers as processing tools. With the continuous development of network technology and the trend of expanding application fields, many system software with artificial intelligence functions have emerged. Machine learning is a technology to establish the connection between computer networks and programs. It provides people with more convenient, efficient, accurate and accurate information by fusing and refining a large number of data related knowledge. It can use existing equipment to realize the cooperative work among multiple servers. It can realize the system design with strong pertinence, simple operation and high real-time processing capability according to different user requirements. The main content of machine learning is to establish a program library based on B/S structure, and then import it into clustering. Secondly, the existing data is used for analysis, training and reasoning to extract the sample feature information. In this process, a large amount of data needs to be processed and calculated to obtain results. Common algorithms include BP neural network (ANN), multi-layer perceptron, genetic algorithm and a series of intelligent machine learning [7-8].

For online lending, the most important thing is risk control. Here we need to use some statistical learning methods to minimize the bad debt rate. Statistical learning method is mainly to build some statistical models for a large number of data, and then use the models learned from training data to predict and analyze the test set data. Statistical learning includes a lot of content, in which supervised learning and unsupervised learning are used more frequently. Decision tree is a relatively traditional and basic classifier. Because this paper mainly makes risk prediction for credit card users, this paper mainly focuses on the decision tree that is also made for classification. In the process of building a decision tree, the most important thing is to select good features to classify data. Feature selection is to select a better feature that is capable of classifying data [9-10].

A random forest is actually a combination improvement of the decision tree. A big difference is that the left and right node selection characteristics of the tree are very different. For classification problems, it is more difficult to build a weak classifier than a strong one. Therefore, we generally build weak classifiers first, and then repeatedly learn to get many weak classifiers. However, when dealing with the problem that the number of positive class data and the number of negative classes differ greatly, Adaboost algorithm will continuously increase the weight of large class samples that are difficult to classify when it encounters large class samples that are difficult to classify. So that small class samples will be ignored to a certain extent. Hierarchical clustering is to first treat each data point as a cluster, then calculate the distance between each cluster and the cluster respectively, and then merge the clusters closest to each other. Finally, a tree like result will be presented [11-12].

2.2. P2P Online Lending Platform

First, the functionality is clarified. P2P online LP provides a trading place for borrowers and lenders. Secondly, the profitability is clarified. The majority of P2P online LPs are for profit, and obtain income by charging fees for information exchange, loan matching, credit evaluation and other services. Finally, the network externality is clarified. For P2P online LP, the more borrowers, the more lenders will be attracted. The more lenders, the more borrowers. The lender only needs to pay the handling fee and other management fees, while the borrower needs to pay the handling fee, toll collection and other management fees [13-14].

The working principle of P2P online LP is shown in Figure 1:

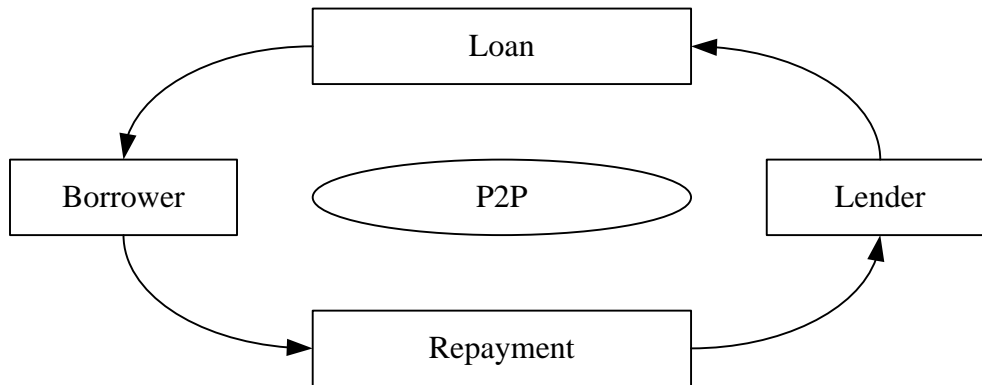


Figure 1. Working principle of P2P online lending platform

Based on the basic theory of the platform, P2P network LP should not only realize the social value with inclusive finance as the goal, but also realize the capital value with profitable operation as the goal. Its operation mode can be divided into profitable operation mode and public welfare operation mode. According to the role of P2P online LP in the transactions between borrowers and lenders, the profitable operation mode can be further divided into simple intermediary mode and compound intermediary mode. In the pure intermediary model, the platform only plays the role of lending information provider. In the composite intermediary model, the platform not only plays a role as a lending information provider, but also acts as a borrower's agent and other roles [15-16].

2.3. Basic Model of Risk Early Warning Mechanism

The research on risk early warning models is mostly based on statistical principles or artificial intelligence methods. Typical models include Probit probability model, KLR signal model, regression model and neural network model.

Probit model is a method to predict the occurrence probability of crisis by estimating the occurrence probability of crisis inducement indicators. The variable M represents the financial crisis and N represents the inducement of the financial crisis, so the Probit model can be expressed as shown in Formula (1):

$$\begin{cases} Q(M = 1) = G(a, \alpha) \\ Q(M = 0) = 1 - G(a, \alpha) \end{cases} \quad (1)$$

Where is the parameter vector corresponding to N , representing the probability coefficient of N factor. This model calculates the possibility of financial crisis in this year by making maximum likelihood estimation on historical data, determining the parameter vector, and then substituting it

into the crisis incentives to be evaluated.

The risk score, the duration of employment of the lender and the debt income ratio of the borrower are missing values. After learning mathematics, we know that we can find a polynomial of degree $x-1$ in a plane. Take the coordinates of the known x points into this polynomial to solve b_1, b_2, \dots, b_n . Lagrange polynomial is:

$$Q(a) = b_1 * \frac{(a - a_2)(a - a_3) \dots (a - a_x)}{(a_1 - a_2)(a_1 - a_3) \dots (a_1 - a_x)} + \Lambda + b_x * \frac{(a - a_1)(a - a_2) \dots (a - a_{x-1})}{(a_x - a_1)(a_x - a_2) \dots (a_x - a_{x-1})} \quad (2)$$

Take the index corresponding to the missing value into a in the above formula to get an estimate of the missing value. The KLR signal method is more theoretical. Although the indicators selected by this method are comprehensive and accurate, in an ideal state, it can help find abnormal variables, or even indicate the range of abnormal values, and timely find problems in the economy, so that decision-makers can formulate relevant policies to solve problems. The STV cross section regression model is also limited in practical application, mainly in two aspects: first, the model uses linear regression to predict the crisis, but in reality most cases are nonlinear. Secondly, the STV cross section model requires that the samples have the same type of properties, which is relatively demanding in the research. At the same time, it ignores the individuality of samples and only seeks for commonness, which easily leads to a single problem to be considered and affects the accuracy of prediction. The neural network model has no special requirements for data distribution and increases the predictable range. However, the internal network of neural network automatic learning is invisible. This part represents the relationship between the output value as a dependent variable and the input value as an independent variable, which limits the scope of the model in practical application [17-18].

2.4. Construction of Risk Evaluation Index System

In order to select more comprehensive, accurate and scientific early warning indicators for risk early warning analysis and research on P2P lending model of Internet finance, this paper mainly follows the following principles to determine early warning indicators:

As a new thing, Internet finance has the nature of both Internet and finance, that is, it has a dual nature. Therefore, this paper first gives full consideration to various factors that affect it, and completes the indicator setting on this basis. Different risk indicators reflect different risk characteristics and play different roles and influences in financial behavior. Internet finance is developed on the basis of traditional finance, and its essential attribute cannot be separated from finance. Therefore, the research on P2P risk early warning of Internet finance can refer to the early warning of traditional finance to a certain extent. In the ten years after the emergence of Internet finance, it has developed at a very high speed and quickly became the dominant player in the financial market. Because of its rapid development, the relevant credit reporting and regulatory system has not been followed up in time and is not perfect. Therefore, there are certain difficulties in selecting indicators and data. In order to carry out more effective analysis and research on Internet financial risks, this paper mainly takes the risk early warning under P2P network lending mode as the research object and goal, and selects indicators that are quantifiable and controllable. It is necessary to ensure that the selected indicators can reflect the economic and financial development to the greatest extent, that is, they are dynamic indicators. In addition, it is necessary to dynamically adjust the selected indicators and their classifications and values. Indicators reflecting capital security risk include whether the capital flow of online lending transactions intersects with the platform's own funds and the mode of fund custody.

3. Investigation on the Loan Platform

3.1. Development of Loan Platform

The development of P2P online LP has accelerated. However, with its continuous growth and the progress of Internet technology, people have a deeper understanding of the information industry and service industry and a higher degree of acceptance. At present, there are three main models in China. First, the license is issued by a third-party payment company. Second, the CBRC and other relevant departments approved the establishment of a "P2P online loan self-discipline organization" to manage private financing business. Third, cooperate with the Banking Association to jointly formulate policies and regulations and coordinate to solve the operation problems of the platform.

3.2. Sources of Related Data Investigation on the Loan Platform

This paper selects indicators such as the number of normal operating platforms, transaction amount, investors and borrowers, and the number and type of problem platforms in the P2P industry to visually reflect the current situation of the P2P industry. And through the questionnaire survey, the relevant suggestion data are obtained.

3.3. Investigation on the Loan Platform

This paper investigates relevant borrowers and platforms for the risk of LPs, including the current problems of lending, the amount of transactions, the number of borrowers and the number of LPs. For borrowers, investigate their general borrowing amount and borrowing frequency. For the LP, the investigation is about the frequency and amount of the platform's lending. In addition, the attitude of both parties and their opinions are investigated by questionnaire.

4. Analysis of Survey Results

4.1. Problems of the Loan Platform

This paper summarizes the problems of the LP through the official website data survey and questionnaire survey. There are 1200 LPs with problems investigated in this paper. The problems of the LP include mutual runs, suspected fraud, malicious escape, closure, loss of connection, economic investigation intervention, withdrawal difficulties, etc. The specific situation of the problem is shown in Table 1:

Table 1. Type and proportion of problem lending platforms

	Number	Proportion
Run out of business	24	2
Suspected of fraud	96	8
Malicious run away	312	26
Closed	384	32
Loss of communication	144	12
Economic investigation intervention	12	1
Withdrawal difficulty	208	19

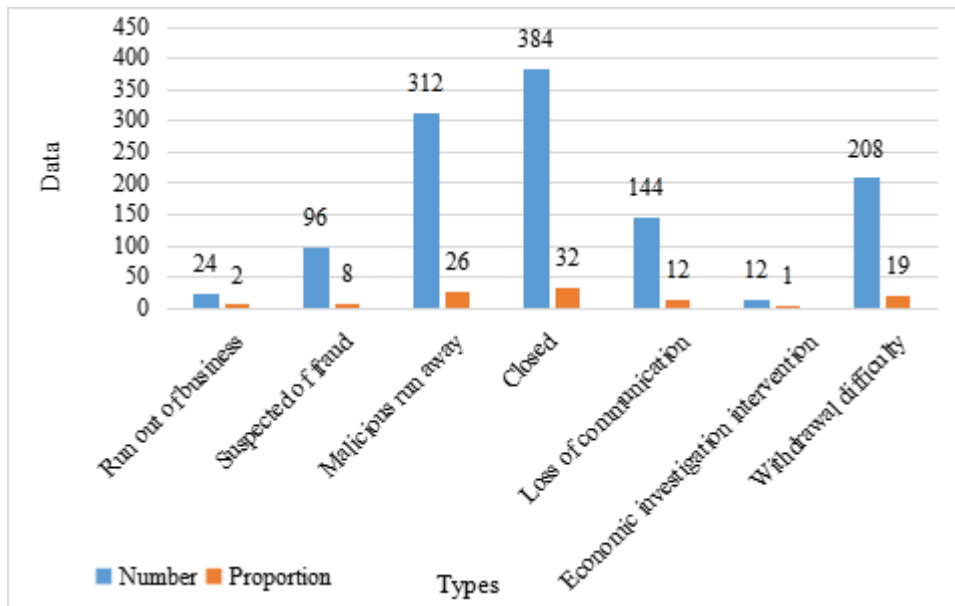


Figure 2. Type and proportion of problem lending platforms

As shown in Figure 2, we can see that closure accounts for 32%, the largest proportion, but most platforms have paid the principal for investors without causing huge losses. In contrast, malicious road running and fraud platforms account for 34%, involving serious crimes. The proportion of bank run failure and economic investigation intervention is relatively small, but they have seriously damaged the reputation of P2P.

4.2. Opinions of Both Parties

Through the questionnaire survey in this paper, both lenders and borrowers believe that online loan transactions should focus on the following points. The first is the credibility of both lenders and borrowers, the second is the effectiveness of both lenders and borrowers, the third is the security of the LP, the fourth is the repayment setting of the LP, and the fifth is the ability of both lenders and borrowers. There are different opinions on these situations, as shown in Table 2.

As shown in Figure 3, we can see that both lenders and borrowers believe that integrity is the first, so integrity can make the platform survive. In addition, both lenders and borrowers also pay attention to the transaction security of the LP, and the platform needs to improve its security performance. The actual effect problem is that the loan is timely and the repayment is timely. Only in this way can we achieve mutual benefit and win-win results. In fact, it is difficult to maintain the stability of the LP only by the lenders and borrowers, and the lending process needs to be standardized with the help of national laws. In this way, the loan risk will be much smaller.

Table 2. Opinions of both lenders and borrowers

	Borrower	Lender
Credibility and integrity	26%	23%
Actual effect	20%	18%
Safety	22%	23%
Repayment Settings	20%	16%
Lending ability	12%	20%

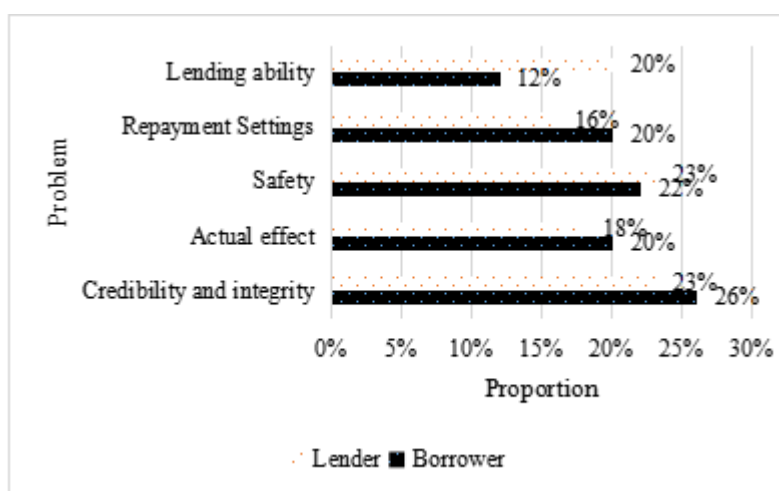


Figure 3. Opinions of both lenders and borrowers

5. Conclusion

In this paper, we discuss the feasibility and reliability of machine learning through the risk monitoring and early warning of P2P network LP and the method of case analysis. At present, China has established a relatively complete and mature risk assessment system and related systems. However, as the Internet is a new thing, its development is not perfect and there is no unified standard to regulate its operation process and supervision mode. In this paper, credit rating, repayment ability and ROI are used as monitoring objects. The risk of borrowing and lending is measured by the ability of both lenders and borrowers and the market environment.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

- [1] Federico Franzoni, Vanesa Daza: SoK: Network-Level Attacks on the Bitcoin P2P Network. *IEEE Access* 10: 94924-94962 (2020).
- [2] Emmanouil Krasanakis, Symeon Papadopoulos, Ioannis Kompatsiaris: p2pGNN: A Decentralized Graph Neural Network for Node Classification in Peer-to-Peer Networks. *IEEE Access* 10: 34755-34765 (2020).
- [3] Mohammad Shoab, Saad Aljubayrin: Intelligent Neighbor Selection for Efficient Query Routing in Unstructured P2P Networks Using Q-learning. *Appl. Intell.* 52(6): 6306-6315 (2020).
- [4] B. N. Jagdale, Jagdish W. Bakal: A Novel Authentication and Authorization Scheme in P2P Networking Using Location-Based Privacy. *Evol. Intell.* 15(2): 1251-1264 (2020).

- <https://doi.org/10.1007/s12065-020-00375-y>
- [5] Federico Franzoni, Vanesa Daza: Clover: An Anonymous Transaction Relay Protocol for the Bitcoin P2P Network. *Peer-to-Peer Netw. Appl.* 15(1): 290-303 (2020).
- [6] Dharmendra Kumar, Mayank Pandey: An Optimal Load Balancing Strategy for P2P Network Using Chicken Swarm Optimization. *Peer-to-Peer Netw. Appl.* 15(1): 666-688 (2020).
- [7] Sanjeev Kumar Singh, Chiranjeev Kumar, Prem Nath: Resource-Cardinality Based Scheme to Reduce Resource Lookup Cost in Structured P2P Networks. *Wirel. Pers. Commun.* 125(4): 3351-3377 (2020).
- [8] Sumit Kumar Tatarave, Somanath Tripathy: PJ-Sec: Secure Node Joining in Mobile P2P Networks. *CCF Trans. Pervasive Comput. Interact.* 3(1): 13-24 (2020). <https://doi.org/10.1007/s42486-020-00047-x>
- [9] Tommi Meskanen, Jarkko Kuusijärvi, Valtteri Niemi: Privacy-Aware Contact Sharing for Groups in P2P Networks. *Int. J. Embed. Real Time Commun. Syst.* 12(4): 36-54 (2020).
- [10] Lyes Badis, Mourad Amad, Djamil A ĩsani, Sofiane Abbar: P2PCF: A Collaborative Filtering Based Recommender System for Peer to Peer Social Networks. *J. High Speed Networks* 27(1): 13-31 (2020). <https://doi.org/10.3233/JHS-210649>
- [11] Rossano Gaeta, Marco Grangetto: On the robustness of three classes of rateless codes against pollution attacks in P2P networks. *Peer-to-Peer Netw. Appl.* 14(6): 3595-3604 (2020).
- [12] Prakasam Periasamy, Md Shohel Sayeed, J. Ajayan: Guest Editorials: P2P Computing for 5G, beyond 5G (B5G) Networks and Internet-of-Everything (IoE). *Peer-to-Peer Netw. Appl.* 14(1): 240-242 (2020). <https://doi.org/10.1007/s12083-020-01001-5>
- [13] Nitin Singha, Yatindra Nath Singh: New Incentive Mechanism to Enhance Cooperation in Wireless P2P Networks. *Peer-to-Peer Netw. Appl.* 14(3): 1218-1228 (2020).
- [14] Vamseedhar R. Reddyvari, Sarat Chandra Bobbili, Parimal Parag, Srinivas Shakkottai: Mode-Suppression: A Simple, Stable and Scalable Chunk-Sharing Algorithm for P2P Networks. *IEEE/ACM Trans. Netw.* 29(6): 2548-2559 (2020).
- [15] Sanjeev Kumar Singh, Chiranjeev Kumar, Prem Nath: Analysis and Modelling the Effects of Mobility, Churn Rate, Node's Life Span, Intermittent Bandwidth and Stabilization Cost of Finger Table in Structured Mobile P2P Networks. *Wirel. Networks* 27(2): 1049-1062 (2020). <https://doi.org/10.1007/s11276-020-02493-y>
- [16] Alexandra Igorevna Kononova, Larisa G. Gagarina: Dynamic Model of Growing File-Sharing P2P Network. *Autom. Control. Comput. Sci.* 54(7): 645-651 (2020). <https://doi.org/10.3103/S014641162007010X>
- [17] Kensworth Subratie, Saumitra Aditya, Vahid Daneshmand, Kohei Ichikawa, Renato Figueiredo: On the Design and Implementation of IP-over-P2P Overlay Virtual Private Networks. *IEICE Trans. Commun.* 103-B (1): 2-10 (2020). <https://doi.org/10.1587/transcom.2019CPI0001>
- [18] Pankaj Kumar Jadwal, Sonal Jain, Basant Agarwal: Spectral Clustering and Cost-Sensitive Deep Neural Network-Based Undersampling Approach for P2P Lending Data. *Int. J. Inf. Technol. Web Eng.* 15(4): 37-52 (2020). <https://doi.org/10.4018/IJITWE.2020100103>