

## Research on Financial Application Model of P2P Internet Based on Data Mining

Jining Li, Xin Wen

University of Science and Technology Liaoning, Anshan 114000, China

<sup>a</sup>lijining0316@163.com, <sup>b</sup>crazy\_poet@126.com

### Abstract

As a carrier of P2P network finance, P2P network financial platform has the characteristics of high transparency of information, high credibility of borrowers, low risk and simple operation. As the debit record of P2P network financial platform is increasing gradually, the historical data is becoming more and more huge. However, these historical data are only stacked in the database or data warehouse, and has not been effectively used. The online financial platform has little research on mining useful information behind the hidden data. The data mining technology should be introduced into the management of P2P network financial platform. In addition, the data mining technology should be used to analyze the existing data of P2P network financial platform, so as to provide services and suggestions for borrowers, payers and platform managers. It has become a new topic in the development process of P2P network financial platform. In view of the above problems and needs, this paper hopes to study the application of data mining technology in P2P network finance, and promote the construction and development of P2P network finance.

### Keywords

Data mining technology, P2P, network finance, application model.

### 1. Introduction

When some of the business in the financial industry, such as capital transactions, payment payments, information intermediaries, was achieved by Internet, the Internet financial model has emerged. This model is different from the current direct or indirect financing model [1]. At present, there are mainly three kinds of Internet financial enterprises in the financial market, namely, small loan enterprises, third party payment enterprises and financial intermediary enterprises [2]. Domestic commercial banks have launched a variety of mobile and Internet banking, which is also part of the internet banking. This fashionable financial model is a new model formed by the Internet and the traditional financial industry to promote each other and influence each other [3]. The characteristics of the Internet itself, namely openness and equality, are reflected in the financial industry. This has led to a qualitative change in the traditional financial sector and a fundamental impact on the human financial model. P2P network finance is also known as P2P credit, and it means person-to-person. It refers to the borrowing between individuals and individuals and individuals and small businesses. In general, this kind of lending business is done on the internet. The borrower publishes information on the Internet, such as the amount of the loan, the level of interest and the repayment plan [4]. The lender searches the information on the internet, understands the borrower's borrowing information, weighs the risks and gains, negotiates the borrower's content, and finally negotiates the borrower's content and completes the loan on the network platform. This is the self-service of financial markets. In addition to the traditional financial model, P2P network finance meets a part of the needs of social financing and investment.

Many small and micro businesses owners, entrepreneurs, wage earners, farmers and college students who are lack of collateral and social resources, they can obtain funds through the P2P network financial platform. Some investors who have an idle capital and have a willingness to invest are earning more than the traditional financial sector. P2P network financial income is relatively high [5].

In general, it can stay at 10% or higher. Compared to the traditional financial industry, its advantage is greatly improved the efficiency of borrowing services, and reduce the borrowing threshold and difficulty. Therefore, this effectively alleviates the difficulty of borrowing of some people [6]. In addition, P2P network finance can provide people with more flexible and more profitable financial way. P2P network finance provides a transparent and open platform for direct communication and information between borrowers. P2P network finance combines the Internet and small loans, and retains the advantages of the two. The P2P network financial platform solves the problem of complicated and inefficient procedures for borrowers to apply for loans from commercial banks. At the same time, a small amount of funds and resources are collected together to provide sufficient financial support for people in need. The openness of the Internet provides a platform for the collection of funds. The borrower not only gains higher returns while diversifying the risk of loss [7].

## 2. State of the art

In developed countries, data mining technology has penetrated into various fields. Among them, there are some typical applications. IBM, for example, uses NBA game data as a data source and digs out different combinations of players to optimize the tactics of the team [8]. The insurance industry should use data mining technology to predict the possibility of user fraud. In terms of user management, data mining techniques should be used to classify users. In the biological field, data mining techniques are used to analyze the permutations and combinations of DNA [9]. In the field of network teaching, through the use of data mining, teachers can analyze the student's learning situation, so as to further improve the quality of teaching. In the medical field, the application of data mining technology can analyze the characteristics of patients, the pharmacy management, predict the probability of disease and so on. At present, the latest development direction of foreign data mining is to further study and improve the algorithm model of knowledge discovery and to expand the application area of the model, such as the continuous improvement of genetic algorithms and support vectorization algorithms. Many Internet companies and research institutions on the development and application of data mining are very seriously. Amazon and Google have set up a corresponding research center [10]. Domestic research on data mining started a little late. Compared with foreign countries, domestic research is relatively backward. At present, there is no overall strength. However, after years of hard work, some progress has also been made. From 2001 to 2003, data mining related papers are very rare. However, in recent years, data mining has clearly become a hot topic in academia. The number of papers in this area to maintain high growth is the best evidence. Many enterprises have not only invested a great deal of money and manpower in data mining research, but also actively applied their research results to their products. At the same time, educational institutions also set up data mining courses. In the training of data mining personnel, enterprise spending is also continuing to rise. These phenomena show that data mining is a very attractive area, it can effectively promote the development of many industries of data. It is a practical application oriented technology. At home, it is now widely used in the fields of finance, banking, agriculture, manufacturing, retail, telecommunications, health care, education and bioscience.

In the information technology, intelligent search engine is the product of data mining technology in the field of search engine applications. It can give users more efficient and accurate web feedback. In the medical and health aspects, data mining can be used for cases, patient behavior analysis and pharmacy management. At the same time, it can also be used for the arrangement of the treatment plan and the judgment of the effectiveness of the prescription. In the retail industry, data mining technology is used to analyze and study supermarket customers to determine the specific customer interest, consumption habits, consumption trends and consumer demand, predict the customer's next purchase behavior. And for different types of customers, the appropriate marketing activities is carried out, so as to increase sales [11]. In agricultural construction, the use of data mining technology can provide solutions for rural informatization construction, and can effectively solve the problem of difficult collection of rural service information and the difficulty of information entering the villages and households. In telecommunications, data mining technology can be used to analyze the behavior

characteristics of telecom customers, and find the potential demand of users, and finally provide the basis for the design of packages. At the same time, customer management can attract new customers and reduce the loss of the old customers. In education, data mining technology can be applied to the distance education system, put forward personalized education services, thereby improving the quality of distance teaching. In the field of justice, data mining techniques can be used for case investigation, case analysis and crime monitoring, and it can also analyze the characteristics of criminal behavior. In terms of finance, data mining technology can be used to assess the credit rating of the account, analyze the usage patterns of credit card users, analyze the trend of stocks and explore the relationship between financial policies and financial markets. In e-commerce, data mining technology can give each visitor a personalized recommendation to stimulate the user's desire to buy, and increase sales.

In 2005, the first P2P network financial platform was established in the UK. The industry began to develop rapidly in Europe and the United States. In 2006, China's first P2P network financial platform was established. Therefore, P2P network finance can be regarded as a new thing. Just a few years, P2P network finance has gradually been accepted and developed rapidly. P2P network finance has aroused the attention and research of scholars at home and abroad. There are many qualitative researches on P2P network finance, and quantitative research is relatively few. It is mainly focused on the following aspects. When tracing the causes of Internet banking, both domestic and foreign scholars point to the imperfections of the traditional financial system. The traditional formal financial sector has limited coverage. It does not pay enough attention to the individual and the small enterprise, and the procedure is complicated and the process is tedious. Driven by the interests and driven by the reality of demand, the Internet banking has developed rapidly. At present, the domestic and foreign scholars generally believe that the traditional commercial banking business has been impacted to a certain extent by the emerging P2P network financial industry. Most scholars doubt that the P2P network finance will replace the traditional banking industry. At present, there is no unified conclusion [12].

In 2008, M.Klaffit explored the factors that affect the success of borrowing on the P2P network financial platform on the basis of data provided by the Prosper platform. He concluded that the detailed personal information is good for borrowing. He also points out that the success rate of borrowing is proportional to the borrower's credit rating and the bank's creditworthiness. AA and A-level borrowers are more likely to get loans. In contrast, the high-risk borrower's success rate is very low. In 2009, M.Lin, NRPrabhala, S.Viswanathan, et al. studied the world's largest P2P network financial platform Prosper. Through the social network, P2P network financial platform borrowers were investigated, which can effectively solve the problem of adverse selection. The results of the study show that the more detailed the social network information, the borrower is more likely to get loans, and the borrower's interest rate is relatively low. In 2010, Collier et al. studied the relationship between the borrowing rate and the borrower's borrowing amount, the borrower's financial condition and the auction method. The results show that bad financial conditions and huge borrowing will lead to higher borrowing rates [13]. In 2011, Chinese scholar Guo Qi conducted an empirical study on the factors that affect the success of borrowing by using the public historical data of P2P online financial platform. The results show that the amount of loans, personal credit rating, and the total number of bids are the most important positive factors. At the same time, different ways of repayment affect the success of the loan. If the borrower chooses a monthly repayment, the probability of a successful loan is greater. And the choice of the return of principal and interest in the end will reduce the probability of successful borrowing. In addition, the study also found that the borrower's previous borrowing interest rate on the success of the probability of borrowing is not significant. In addition, in the friendly loan model, the borrower's borrowing costs are negatively related to the borrower's relationship. When the borrower's relationship is closer, it can reduce the borrower's borrowing costs to some extent. In 2013, Chinese scholar Zhao Lefeng applied the regression analysis model to the P2P network financial platform. On the basis of historical transaction data from PPDAl, the following conclusions are drawn. As a whole, there is no obvious difference in the success rate of loan between

the central and western regions of China. The probability of successful borrower borrowing is positively related to the historical success of the borrower, the personal credit rating, the total number of bids, and the area in which the borrower belongs. And it is negatively related to the borrower's amount of borrowing, the term of borrowing and the interest rate of borrowing. A large number of studies have shown that the probability of success in borrowing on the P2P network financial platform is closely related to the borrower's credit rating and social relationship.

### **3. Methodology**

#### **3.1 Research Subjects on Data Mining Technology in P2P Network Financial Application**

The full name of the Bangyoudai is Qinhuangdao Bangyoudai of the loan Investment Advisory Co., Ltd., which was established in November 2013. It is not only one of the most abundant financial platform of P2P network in China, but also a representative P2P network financial platform. The operation model of Bangyoudai mainly refers to the foreign Prosper, which provides a fair, stable, transparent and efficient network financial platform for individuals with loan needs and financial needs [14]. First of all, users can upload various credit certificates to evaluate credit rating. The user then issues a loan request. Eventually, the financial needs of themselves or enterprises are met. By Bangyoudai loan platform, users can also lend their spare money to borrowers in the form of financial management. Users can obtain high returns. At the same time, this is also conducive to future borrowing.

Generally speaking, the borrower can choose the type of loan according to the characteristics of the borrowing and the demand of the borrower, and issue the loan information of the relevant loan type. Among them, the borrowing rate can be set in the range of 10%~24%. The borrower's borrowing amount is directly related to the credit rating. The higher the credit rating, the greater the amount of borrowing within a certain range. At present, the Bangyoudai loan platform is running well. Bangyoudai loan process can be divided into four stages: first of all, according to their own circumstances, the borrower chooses the type of loan, the amount of the loan, the time of the loan, and submit the loan application. Personal information is certified. After the audit is successful, the loan list is generated. Second, investors browse the list of loans. The risks and benefits are predicted according to the type of loan, the interest rate of the loan, the term of the loan, the credit rating and the relevant certification information of the borrower. The risk and earnings are predicted. Then, the appropriate borrowing is chosen to invest. Third, After the borrowing is full, the borrower borrows successfully. The funds will flow into the borrower's account. If the loan is not credited before the deadline, the loan will fail. Fourth, the borrower must pay off the funds in accordance with the terms agreed upon by both parties. The borrower should ensure that the funds in the account are sufficient to repay the money due for the current period. After all the money paid off, the borrowing business is completed. The system's loan business flow chart is as shown in Figure 1.

Bangyoudai loans are divided into six types: working class loan, net value loan, guarantee loan, enterprise loan, net business loan and card loan. The types of bids are classified as follows: the credit audit standard, the intelligent financial management standard, the field certification standard and the institution guarantee standard. In this paper, the Bangyoudai loan platform is taken as the research object, and the data source of data mining is also the real transaction data on the platform. There are two main reasons for using the Bangyoudai loan platform as the object of study. First, the Bangyoudai loan platform as a new platform, the operation model is more representative. Borrowers and investors are fully autonomous trading, which is conducive to the discovery of objective laws on the platform. Second, the data is accurate, which is the actual transaction data from the platform. The authenticity of data effectively ensures the reliability of data mining results.

#### **3.2 The Overall Demand of Data Mining Technology in P2P Network Financial Application**

Now the P2P network financial industry is developing rapidly in China, and its functions are becoming more and more perfect. With the gradual increase in the borrowing record of P2P network financial platform, transaction information is becoming increasingly large and complex. Compared

with the increasingly perfect business process, P2P network financial platform has little work to excavate the hidden knowledge behind the data. The platform's utilization of transaction data is still at the stage of development. P2P network financial platform has a huge user information and transaction records. Using the data mining technology to analyze the existing data on the P2P network financial platform, we can get the behavior characteristics of the drawer and the borrower. This helps the P2P network financial platform user management or provide effective support for the operation.

First of all, from a large number of transaction information mining users of different behavioral characteristics, which is the simplest, most direct and effective application of data mining in the platform. Data mining technology can help platform managers to build models, identify high-value high-quality financiers, and increase the loyalty of the contributor, which provides data and content support for targeted marketing campaigns. Secondly, the data mining analysis method can predict the full scale, extract the characteristics of successful loans, establish full scale predictive regression model, help borrowers to borrow successfully, and improve the transaction success rate of the platform. Finally, the model of successful borrowing is mined through association rules in data mining, so as to provide decision support for the borrower's loan settings. The personal credit risk assessment model roadmap is shown in Figure 2.

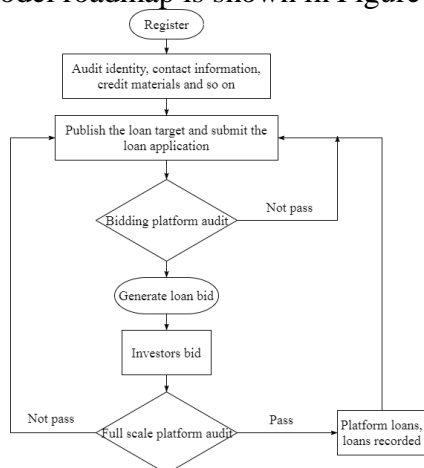


Figure 1. The system's loan business flow chart

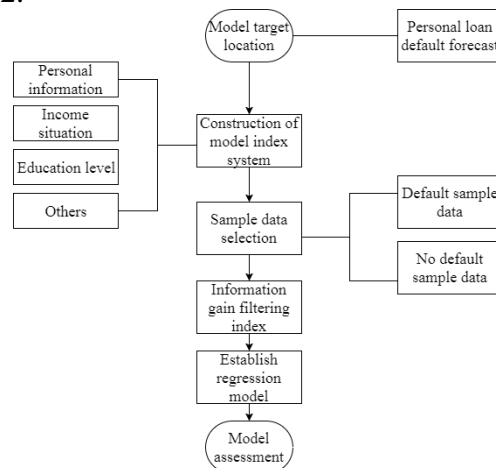


Figure 2. The personal credit risk assessment model roadmap

## 4. Experiments

### 4.1 The Overall Design of Data Mining Technology in P2P Network Financial Application

According to the data mining process, the data mining technology includes three basic steps in the overall design of P2P network financial application, which are defining the problem, preparing data and generating model. In P2P network financial applications, the overall flow chart of data mining technology is as shown in Figure 3:

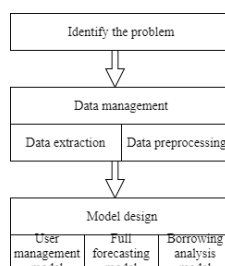


Figure 3. The overall design flow chart

The definition of the problem is to determine the data mining technology in the P2P network of financial applications. The purpose of this data mining is to carry out user management of the contributor, to predict the full standard and to analyze the loan. This study takes the real trading data and the user information as the data source. First, extract the relevant data from the platform database to form the target data set. Observe data and have an understanding of the data. Familiar with data,



and identify data quality problems. In the original data, the data noise is relatively large. The raw data must be preprocessed to produce a data set that is more applicable to the data mining method. Therefore, the data directly extracted from the background needs processing. It mainly includes solving noise data, dealing with missing information and abnormal points, data conversion and so on. High quality data preparation can greatly improve the quality of mined patterns and shorten the actual mining time. At this stage, different data mining techniques can be selected and applied as model setting parameters. Usually, the same kind of data mining problems can be solved by a variety of data mining methods and techniques. This stage is mainly about the selection and setting of data mining model.

### 4.2 Data Preprocessing

In general, in the original data, the data noise is relatively large. The raw data must be preprocessed to produce a data set that is more applicable to the data mining method. Therefore, it is necessary to process the data extracted directly from the background. It mainly includes data cleansing, data integration, data transformation and data reduction. High quality data preparation can greatly improve the quality of the excavated pattern and shorten the actual mining time.

Data cleanup is mainly through technical means to standardize the data format, clear the abnormal data, correct the error and clear the duplicate data. It mainly includes the filling of missing values, the smoothing of noise data, the identification of outliers or anomalies, and other methods. Data integration refers to the combination of data links in multiple data sources and unified storage.

Data transformation is the way to convert data into data models by data smoothing, data aggregation, data standardization, data generalization, data normalization and so on. The data reduction is to reduce the larger data set. However, the new data set does not destroy the integrity of the original data. The result of mining before reduction is basically the same as the result of mining after reduction. The main methods of data reduction include data cube aggregation, dimension reduction, data compression, numerical reduction, discretization and hierarchical generation of concepts.

### 4.3 Design of Borrowing Analysis Model

Through the filtering and integration of the borrowing information on the data management platform, the activity data is obtained, which includes the type of borrowing, the borrowing time, the borrowing amount and the borrowing rate. In this paper, we use association rules algorithm in data mining technology to analyze the combination of related factors of successful borrowing. The basic idea is as follow. From the data platform, we can extract all kinds of successful loans, the amount of the loan, the interest rate of the loan and the time of repayment of the loan. The variable values here are the statistical results on the data management platform. The data pattern is shown in Table 1.

Table 1. Examples of borrowing information data

Types of borrowing	Loan amount	Interest rate on borrowings	Time of payment
Working class loan	15000	17.6	6
Net business loan	15000	18	12
Enterprise loan	50000	19	12
Enterprise loan	98000	18	12

The data types extracted from the data platform are numeric. And the association rule algorithm requires a discrete type for the variable. The borrowing amount is divided into five different sections from low to high. In the Bangyoudai loan platform, the borrowing interest rate has strict control. The interest rate for all borrowings must be below 24%. The repayment time of the loan is also stipulated. The longest time is not more than 24 months, the shortest time is not less than 1 months. The association rules, minimum support and minimum confidence are set up, and the association rules algorithm is used to analyze the above variables. Finally, the loan analysis model is established. The results were shown and analyzed to identify the link between borrowing related factors. The process of determining the degree of completion of the borrowing is shown in Figure 4.

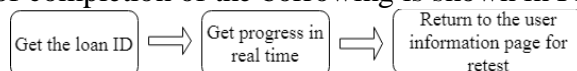


Figure 4. The process of determining the degree of completion of the borrowing

## 5. Conclusion

This paper expects to study the law of network lending market operation, and to explore hidden knowledge in P2P network finance by using data mining technology. Data mining technology is applied to the management of P2P network financial platform. Data mining technology is used to analyze the existing data of P2P network financial platform. The platform's understanding on users has been deepened. The full scale is forecasted and the borrowing is analyzed. The main points of this paper are as follows: First of all, the research status of P2P network finance at home and abroad is summarized, and the application of data mining technology in P2P network finance is expounded. Secondly, the functions and characteristics of the existing P2P network financial platform are analyzed. From the aspect of platform managers, the necessity and feasibility of data mining technology in P2P network finance are discussed. Finally, the model of P2P network finance is constructed by using data mining tools, and the results are explained.

## References

- [1] Al-Saggaf, Y., & Islam, M. Z. (2015). Data mining and privacy of social network sites' users: implications of the data mining problem. *Science and engineering ethics*, 21(4), 941-966.
- [2] Chen, Z., Li, Y., Wu, Y., & Luo, J. (2017). The transition from traditional banking to mobile internet finance: an organizational innovation perspective—a comparative study of Citibank and ICBC. *Financial Innovation*, 3(1), 12.
- [3] Edwards, M., Rashid, A., & Rayson, P. (2015). A systematic survey of online data mining technology intended for law enforcement. *ACM Computing Surveys (CSUR)*, 48(1), 15.
- [4] Gleasure, R., & Feller, J. (2016). Emerging technologies and the democratisation of financial services: A metatriangulation of crowdfunding research. *Information and Organization*, 26(4), 101-115.
- [5] Huang, H., Wu, X., Tao, G., Zhang, Y., Wang, Y., & Long, X. (2016). Research on influence factors of new rural sports culture development based on data mining technology. *RISTI (Revista Iberica de Sistemas e Tecnologias de Informacao)*, (E5), 1-13.
- [6] Jiang, Y., Yao, J., & Xu, J. (2016, July). Research on the application of data mining technology in user management of P2P network financial platform. In *Logistics, Informatics and Service Sciences (LISS)*, 2016 International Conference on (pp. 1-5). IEEE.
- [7] Jun, S., Park, S., & Jang, D. (2015). A technology valuation model using quantitative patent analysis: A case study of technology transfer in big data marketing. *Emerging Markets Finance and Trade*, 51(5), 963-974.
- [8] Jadhav, S., He, H., & Jenkins, K. W. (2017). An Academic Review: Applications of Data Mining Techniques in Finance Industry. *International Journal of Soft Computing and Artificial Intelligence*, 4(1), 79-95.
- [9] Koh, H. C., Tan, W. C., & Goh, C. P. (2015). A two-step method to construct credit scoring models with data mining techniques. *International Journal of Business and Information*, 1(1).
- [10] Li, N. (2016). Research on the influencing factors of rural financial development based on data mining technology. *RISTI (Revista Iberica de Sistemas e Tecnologias de Informacao)*, (E8), 106-117.
- [11] Natek, S., & Zwillling, M. (2014). Student data mining solution—knowledge management system related to higher education institutions. *Expert systems with applications*, 41(14), 6400-6407.
- [12] Ross, E. S. (2017). Nobody Puts Blockchain in a Corner: The Disruptive Role of Blockchain Technology in the Financial Services Industry and Current Regulatory Issues. *Catholic University Journal of Law and Technology*, 25(2), 7.
- [13] Sethi, S., Malhotra, D., & Verma, N. (2016). Data mining: current applications & trends. *International Journal of Innovations in Engineering and Technology*, 6(4), 586-589.
- [14] Xiao, M. J., Ou, Y. Y., & Li, Y. (2014). On the Influence Factors of Credit Risk of Online P2P Lending in China: Based on an Empirical Analysis by the Ranking Selection Model. *The Theory and Practice of Finance and Economics*, (193), 1-6.