

Mining on Public Sentiment Based on Net Loan Comments

Daohai Zhang¹, Gui Ren², Youcai Yang³, Barry Bishop⁴, Hiroshi Honda²

¹School of Management, Jiangsu University, China,

²School of Business and Information Technology, Northwestern Polytechnic University, USA,

³Department of Psychology, University of Wisconsin, USA,

⁴Department of Business Administration, Kyunsung University, Korea

Abstract:- With the growing dependency of network socialization, the information released by users has gradually become a valuable data source for enterprise competition and intelligence analysis. Recently, net loans have been developing rapidly, but several have gradually become a unique phenomenon in China. This research focuses on the comment label of net loans using the sentiment analysis and topic LDA mining method. It begins with data pre-processing, classification of comment label topic mining, and grasping the overall concern of users on the net loan platform. It then examines making the comment label for sentiment classification. Finally, the comments are mined again and analyzed by emotional theme distribution under different emotional words. The research results show that positive sentiment of the net loan platform includes investment cycle, the history of the anticipated annualized rate, payment methods of return, service fees, simultaneous capital security, income security, user experience, and platform supervision.

Keywords:- Sentiment analysis, net loan, text mining, LDA mining method, platform.

I. INTRODUCTION

Online lending (also known as P2P network lending) refers to as a point-to-point financial lending activity based on Internet platforms. According to the Chinese government, the development of inclusive finance was the development direction of the central government in support of Internet finance, from the first proposed online loan to enforced supervision, and then to standardizing development. Based on the Net Loan Industry Reports in 2017, as of the end of December 2016, the annual volume of the online banking industry has reached RMB 2,038.382 billion, and the number of investors and borrowers has increased by 134.64% and 207.37%, respectively, from 2015, with trillions of yuan. Funds poured into the online loan industry, and its popularity has also significantly increased annually. Online lending has provided a fundamental channel for personal credit, SME financing, and investors' financial management.

Nevertheless, the rapid growth of the platform has also caused much chaos, frequent risk events, and significantly damaged the interests of investors. For example, the Zubao

online was under investigation for illegal business, and Shengrong online was unable to withdraw more than RMB 900 million in 2015. In 2016, the Sudai e-loan was under investigation of illegal deposits, and Bao loan was involved in a naked loan scandal. There were eight online loan platforms which involved tens of thousands of people that were run out of business simultaneously in 2017. This phenomenon has also become one of the essential new incentives for the occurrence and instability of social mass events. Therefore, it must be given sufficient attention, research, and proposed similar policies.

Currently, the news media has played an increasingly important role in the information society we are. A significant amount of information comes from socialization annotations and texts. Traditional modeling tools can no longer process and quantify all this information in real time. Intelligent identification and processing through computers are in urgent need. In this new environment, the invisible touch trading activity is transferred to the visible network message and comment texts through the user experience. The user can achieve the purpose of risk warning through the mining and discovery of text knowledge and improve the prevention of online loan risks. Thus, this is a research topic with basic academic theory and practical significance.

II. LITERATURE REVIEW

At present, there are two significant aspects of research on P2P in the world. First of all, with the development of P2P research and its characteristics, the choice of P2P online loans (Klafft, 2009) would be more flexible, and the investment cost is low compared to traditional financing. The operation model of P2P and the fundamental process of its network lending has been studied (Lu, 2015), and the division of the P2P operation model into income type and zero return type has also been proposed. Secondly, there is the online loan risk and management research. Compared to traditional finance, Agarwal (Agarwal, 2008) has indicated that the reason for the high risk of P2P online loans was the low transaction cost.

Regarding network lending, Pan has studied the development status of P2P network lending (Pan, 2014), analyzed the sources of funding, the structure of P2P platforms, and has revealed the risks of developing P2P

network lending. According to the P2P financial market, Zhao has concluded there is a risk (Zhao, 2015) faced by the Chinese P2P market. These risks have also included credit risks, information technology risks, legal and operational risks, market risks, and other destabilizing factors. Based on the analysis of the risk mechanism of P2P online loan platform, the analytic hierarchy process to construct a P2P online loan platform risk assessment index system has adopted (Zhang, 2015).

The current literature mainly focuses on the qualitative and empirical research methods concerning the status quo and characteristics of China's P2P industry development and online loan risk research. There is little research based on the perspective of online loan users in China. This paper combines sentimental analysis and the LDA topic model to mine and analyze the data review information of online loan users, and obtain sufficient information with a significant amount of

review data. The purpose is to grasp the public sentiment of online loan users and further improve the customer service level for the online loan platform and to provide decision support.

III. DATA PREPARATION

A. Thematic Mining Process

Based on the topic mining of comment tags, the general steps are: First, pre-process the wording of the comment tag data source, and the stemming of the comment tag data source. The second step is to carry out an emotional classification of the commentary information. Finally, the LDA is modeled for the emotion after classification, and the user comment keywords under different emotion classifications are mined distributed. The process is shown in Figure 1.

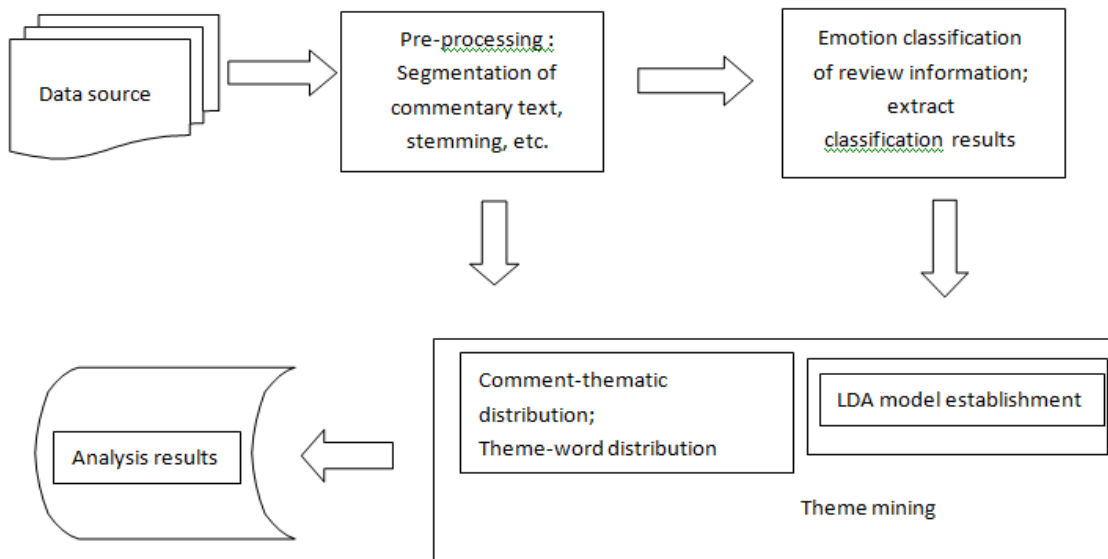


Fig 1:- Thematic mining process

B. Data Sources

Net loan company Tianye was the first portal of P2P wealth management industry in China and focuses on providing authoritative and real-time online loan platform data. It is the first gathering place for information on China's online banking industry. From the research on user sensation based on online loan reviews, we use online loan Tianye as the data collection platform. According to the platform comprehensive index ranking, 10,274 reviews of different levels of platforms were selected. The data review content is shown in Table 1.

Number	Comment
1	Personal loan depository is the most authentic on the big platform, and investment or withdrawal will jump to the Minsheng Bank's records. Earnings can also be used for novice red bonus, especially long-term benchmarks.
2	U plans, various types of funds available for different investment styles for users to choose, chose the medium risk of funds, the income is very satisfied.
3	The interest rate is reasonable, credit approval poor, and deposit and non-deposit are running at the same time. Need to improve.
4	Each investment is deducted from the proceeds of the network borrowing agency service fees 10%, and there is no need to recommend.
5	There are safe depositories, and it belongs to the venture capital department. It can be developed and it is excellent.
6	Interest rates are reasonable, but the depository and non-depository are not safe? Depository rates are low.
7	The annualized rate of the online is still reasonable. 10% of the interest rate is charged, and the overall interest rate is low. Both withdrawals exert handling fees.
8	All aspects of safety and compliance are functional, but interest rates need to be increased.

Table 1. Comment Data Sheet

The emotional and topic mining of comment labels is the mining and processing of the main information in the text. The data source contains the original review information, and the data must be pre-processed before use: (1) Be simple. Deleted all repeated invalid content. (2) To distinguish. Use the Chinese word segmentation tool Jieba to segment all the review data and the phrase. (3) Stemming treatment. Set "ta", "la", "ah", and punctuation as stop words and delete.

IV. EXPERIMENT PROCEDURE

A. LDA Model Structure

The Dirichlet Distribution LDA (Latent Dirichlet Allocation) is the primary model that generates topics based on word frequency in a set of documents. LDA is a very effective method for constructing the probability distribution of mixed words that can be used to accurately and reasonably find topics in a given document set. The general steps of the LDA topic model construction are:

1. After cleaning the word segmentation for all comment data, get corpus doc_clean.
2. Create a dictionary of words in a corpus:
dictionary = corpora.Dictionary(doc_clean)
Use each unique word as an index to create a dictionary for easy search.
3. Convert corpora to document word matrix doc_term_matrix:
doc_term_matrix = [dictionary.doc2bow(doc) for doc in doc_clean]

Document - The search word matrix is a matrix to describe the word frequency of a document. Each row corresponds to a document in the document set, and each column corresponds to a word.

4. Refer the above dictionary and document word matrix as parameters, input gensim. models. Ldamodel. LdaModel() function:

```
ldamodel = Lda(doc_term_matrix, num_topics=6, id2word = dictionary, passes=50)
```

Other related parameters are as follows:

Num_topics: The LDA model requires the user to specify how many topics to create. Users can make appropriate choices according to the size of the document set.

Passes: Number of times the model traverses the corpus. The more traverses, the more accurate the model. For large corpora, the longer it takes to traverse, the longer it takes to complete.

5. Last output topic model:

```
print(ldamodel.print_topics(num_topics=6, num_words=10))
```

The relevant parameters are as follows:

- num_topics : The number of theme models generated.
- num_words : The last n words output from each topic model.

B. Topic Mining

Based on the user sentiment research of online loan reviews, we input 10,274 review data into the LDA model, extract the top n words with the most massive probability distribution in each topic i, and generate the topic-word distribution matrix, i.e., the review topic model. For the data collected in this paper, four num_topics are set in this experiment, six num_words are printed, and the related topic vocabulary shown in Table 2 is obtained.

Topic1	Topic2	Topic3	Topic4
Depository	Interest rate	Good	Recharge
Risk control	Customer service	Red bonus	Novice
Online	Repayment	Guardance	Platform
Compliance	Withdraw	Withdraw	Withdraw
Interest	Hope	Website	Repayment
Income	Overdue	Experience	Voted

Table 2. Thesaurus

Topic1: Topics include depository, risk control, online, compliance and other keywords, indicating that users have a strong focus on platform security issues. According to the online Tianyen lending data, the number of P2P online lending platforms in China has reached 5,029, six new platforms were added, 93 new problematic platforms were added, 1.09% higher than the previous period, and the cumulative problematic platform has reached 3,287, online operating platform 1742, down 23.06%. With the gradual establishment of the supervision mechanism, the problematic platforms withdraws, users have more emphasis on issues such as fund custody, platform compliance, and risk control effectiveness relative to interest and earnings.

Topic2: Topics include interest rate, customer service, repayment, withdrawal, and other keywords, indicating that users are concerned about the platform's funding issues. Users pay attention to the investment cycle of each investment platform, historical expected annualized yield, payment return, service fees. These factors will directly affect the expected annualized return. According to the statistics of online lending, in the platform of national issues, cash withdrawal difficulties accounted for 20.55%, platform disconnection accounted for 62.90%, platform fraud accounted for 4.43%, and benign exits accounted for only 0.46%. Therefore, when there is a problem that the funds cannot be withdrawn for a long time and the user cannot contact the customer service, the so-called platform disappearance may have already appeared.

Topic3: Topics include keywords such as sound, red bonus, guardance, and withdrawals, indicating that users are concerned about the platform's investment income.

According to the statistics of online lending company Tianyen, in all online lending platforms, the interest rate of the comprehensive first-ranking, Yiren lending platform is between 5% and 9.6%, and the last, Sanban lending platform interest rate is between 1% and 16%. It shows that the highest rate of return of Sanban loans is higher than that of the Yiren loans, but the range of yield fluctuations of the Sanban loans is significant and unstable. Therefore, a platform with a high-interest rate may not necessarily have a strong overall strength. It is not only the revenue intensity of the platform that the user chooses to evaluate the investment on the online loan platform, but also the stability of the return of the platform.

Topic4: Topics include recharge, novice, repayment, voted, and other keywords, indicating that the user is not mature enough in the P2P network loan investment. At present, the market has a sophisticated environment for online loans, and there are various kinds of online loan platforms. It is difficult for novices to distinguish authenticity from others. Users are concerned about the difficulty of choice due to concerns about the safety of funds. Therefore, at this stage, it

is particularly crucial for the online loan industry market to provide proper guidance for novices.

Analyze the user concerns of the online loan review. The content of the user review mainly includes platform, depository, withdrawal, customer service, income, experience. It reflects that users are investing in various online loan platforms. The safety and profitability of funds are two major concerns for users. At the same time, users pay more attention to the customer experience, which is related to the risk of the platform.

C. User Emotion Classification

The authors used sentiment analysis to classify the data into emotions classification. 5851 positive emotions were scored, accounting for 56.95%; 2497 were neutral emotions, accounting for 24.30%; 1926 were negative emotions, accounting for 18.75%. The classification statistics are shown in Table 3.

Name	Positive emotions (%)	Neutral emotion (%)	Negative emotions (%)	Total numbers
Statistical results	56.95%	24.30%	18.75%	10,274

Table 3. Sentiment Classification Statistics

The results show that more than half of the users who maintain positive emotions on the online loan platform show that more than half of users recognize the development status of P2P online.

loans; but about 40% of users maintain neutral or negative emotions, indicating that the online loan industry is still a considerable development space. Active emotion segmentation statistics are shown in Table 4.

Grade	General (-10-0) (%)	Moderate (-20- -10) (%)	High (below -20) (%)	Total (%)
Statistical results	32.43%	17.36%	7.15%	56.95%

Table 4. Active Emotion Segmentation Statistics

The results show that the proportion of users' emotions from generally affirmative to highly affirmative decreases from high to low, indicating that the user's recognition of online loans has dramatically improved. The statistical results of adverse emotion segmentation are shown in Table 5.

Grade	General (-10-0) (%)	Moderate (-20- -10) (%)	High (below -20) (%)	Total (%)
Statistical results	13.69%	3.83%	0.34%	18.75%

Table 5. Negative Emotion Segmentation Statistics

The results show that the proportion of users' emotions from general negation to high negation diminishes from high to low, indicating that although users maintain negative sentiments towards online loans, they still maintain hopes.

D. User Emotion Topic Mining

After the comments were sentimentally categorized, the LDA topic mining was conducted again, and the topic vocabularies under different sentiment classifications were discovered, as shown in Table 6 and Table 7.

Topic1	Topic2	Topic3	Topic4
Investment	Income	Withdraw	Financial management
Hope	Good	Repayment	Investment
Experience	Safety	Speed	Escrow
Project	Interest rate	Customer service	Hope
Activity	Comparison	Expiry	Bank
Possible	Interest	Procedure	Expect

Table 6. Positive Affect Related Keywords List

Through the theme mining of users' positive emotions, it is shown that the user's recognition of the platform mainly includes investment, income, cash withdrawal, and wealth management.

Table 7. Negative Affect Topic List

Topic1	Topic2	Topic3	Topic4
Interest rate	Expiry	Withdraw	Website
Income	Now	Funds	Customer service
None	Withdraw	Repayment	Supervision
Waste	Funds	Safety	Phone
Investment	Safety	Income	Risk
Customer service	Trustee	Risk	Safety

Through the theme mining of user negative emotions, we can see that the online loan platform still has problems in the following aspects and needs further improvement.

V. RESULTS

A. Experimental Results and Analysis

The comments reflect the users' views and attitudes towards P2P online loans. The purpose of this paper's user opinion research based on online loan reviews is to extract valuable information from unstructured texts and analyze them to obtain valid conclusions. Thus, users receive public opinion monitoring, strengthening of the specification and management of the Internet financial investment and financing environment, and decision support.

Based on the LDA theme model, a user "hot topics" related topic vocabulary is established to determine the hot topics the user is concerned about in the P2P online loan behavior. The topic mining based on sentiment classification can obtain the problems of the online loan industry from the attitude of the users after performing online lending, to facilitate the avoidance of online loan risks.

According to the current problems in the online loan industry identified by users' topic mining under different emotion classifications, the following suggestions have been proposed simultaneously:

1. Users treat the network loan platform rationally. Due to the risk of online loans, users must not blindly invest in the immediate interests. Users need to strengthen their understanding of the network loan investment and financing model. The most critical point that distinguishes Internet finance from traditional finance is to de-authorize and at the same time increase the risk of investors. Personal experience is the most critical source of judgment and trustworthiness. The online loan platform has some subtle commonalities, and it collects reliable signals through the experience of details.
2. Online loan platform should enhance customer experience. As part of Internet finance, online loans are vulnerable to market public opinion. Excellent customer experience is a magic weapon for the success of the platform. The platform should strengthen the monitoring of user sentiment, and at the same time do a good job of risk control and make appropriate preparations in advance to cope with emergencies.
3. Government strengthens supervision of the platform. The online loan platform still needs to be included in the construction of the national credit system. As an integral part of the corporate and personal credit reporting system, the state should establish relevant systems, clarify the tasks of related departments, strictly examine the legitimacy of the online loan platform, and boycott irregular network loan platforms.

VI. CONCLUSION

With the growing dependency on the socialization of the Internet and the popularization of Internet finance, the review information published by users has gradually become a valuable data source for enterprise competition and intelligence analysis. This paper focuses on the user reviews of different levels of online loan platforms and has researched on the topic of emotions to understand the public opinion of online loan users. Through the processing of word segmentation, sentiment classification, the LDA model is used to generate related topic vocabularies, and valuable information in user reviews is discovered. Based on the hot topics of users' attention, the current problems of online loan platforms are analyzed and proposed. The proposal for further

development provides decision support for strengthening the regulation and management of China's Internet finance investment and financing environment.

VII. ACKNOWLEDGEMENT

The authors gratefully acknowledge the support by Chinese National Natural Science Foundation (71471076), Chinese Postdoctoral Science Foundation (2015M581749), Jiangsu Province Social Science Foundation (13JDB031), and Jiangsu University senior talent foundation (15JDG007).

REFERENCES

- [1]. Agarwal S, and Hauswald RBH (2008) The Choice between Arm's-Length and Relationship Debt: Evidence from E-loans. *Social Science Electronic Publishing* 488-500.
- [2]. Klafft M (2009) Online Peer-to-Peer Lending: A Lenders' Perspective. *International Conference on E-learning* 2(2): 371-375.
- [3]. Lu X and Li H (2005) P2P network lending operation mode and risk management and control *Reform* 2: 60-68.
- [4]. Pan Z, and Xing B (2014) Research on the development status and risk revealing of China's P2P network lending model. *Future and Development* 6: 86-89.
- [6]. Zhang Q, Zhang L (2015) Study on Risk Evaluation Indexes of P2P Online Loan Platform Based on Analytic Hierarchy Process. *Journal of Nanjing Audit University* 12(6): 85-94.
- [8]. Zhao S, Zheng Z, and Li Y (2015) The P2P Network Lending Risk in China under the Background of Big Data and Its Precaution. *Journal of Henan Commercial College* (3): 1-4.