

The Influence of Borrowers' Language Features on P2P Lending

Zhengwei Ma (✉ ma_zhengwei@163.com)

China University of Petroleum, Beijing

Dan Zhang

China University of Petroleum, Beijing

Research Article

Keywords: Peer-to-peer lending, Language features, Default risk

Posted Date: January 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-144354/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Under the background of the reshuffle of the P2P market in China, this paper investigates the influence of four borrower's language features on their funding and default rate based on language function theories. In our study, we use a logistic regression model and the empirical results show that: the more redundant the borrower's language expression is, the more open and objective the content is, and the more attention is paid to the punctuation details, the easier it is to obtain the loan successfully. When the borrower's description is more redundant and more attention is paid to the punctuation details, the probability of default would become lower. Taking the education level into consideration, we find that the negative relating effect between the description redundancy and the default rate would be lower with the increase of the borrower's education level. Therefore, we can conclude that the four linguistic features of borrowers which are defined in this paper can alleviate the information asymmetry problem of P2P lending to some extent and the borrower's linguistic features can be included into the risk control system.

Introduction

At present, the Chinese Internet financial business mainly includes the third-party payment, crowdfunding, P2P lending, digital currency, big data finance and others, among which the P2P lending has the biggest market scale. However, as a new mode of loan financing, P2P lending lacks the complete mechanism of risk control and early warning, creating increasingly serious risk exposure problems. The thunder tide, a phenomenon of collective debt evasion of P2P platforms, which started in June and broke out in July has shocked investors in 2018. Data from Wangdaizhijia show that in July 2018 a total of 197 P2P platforms closed. Cash crisis happened in a large number of P2P platform with large loan balance, affecting millions of investors. The total number of closed and problematic platforms in 2018 was 1279, involving a loan balance of 143.41 billion yuan, far exceeding the total loan balance of the problematic platform in previous years. The thunder tide of P2P in 2018 started in June and broke out in July. The centralized exposure to the online lending chaos in the 315 party in 2019, left the online lending in the firing line. The 2019 annual report data of Wangdaizhijia show that the internet lending in 2019 reached 96.4911 billion yuan, compared with 1794.801 billion yuan in 2018, which decreased 46.24%. After the rapid expansion of P2P and the current contraction, how to perfect the risk control system and reduce the information asymmetry between the borrowers and lenders has undoubtedly become an urgent problem to be solved.

Different from the traditional lending mediated by banks, the original definition of P2P lending is a kind of private lending without any intermediary and guarantee (Lin, Prabhala, & Viswanathan, 2013). As a third party, P2P platform only provides information exchange and credit evaluation services for borrowers and lenders. However, we find that many P2P platforms tend to launch field certification and institutional guarantee in order to reduce the borrower's default risk. In fact, in the case of no intermediary and guarantee, the process of the transaction between the borrower and the lender is also a game of information between the two sides (Wang & Kong, 2014). The main risk sources of P2P lending are moral hazard and adverse selection problems caused by information asymmetry (Simkins & Rogers, 2004). To date, many scholars have done a lot of research on this.

What kind of people are more likely to get loans? What kind of person will repay as promised? These two problems have been extensively studied. Based on the research on the demographic characteristics of borrowers, it is found that there are certain racial, gender and beauty discrimination in the lending market (Pope & Sydney, 2011; Laura et al. 2014). With the development of P2P online loan transaction, the borrower's historical behavior data has also become a reference for the lender. The research results of Puro et al (2010) and Chen and Ding (2013) show that the borrower's credit score, total repayment ratio, debt record, overdue repayment times and other previous data play an

important role in the borrower's success in obtaining the loan. These factors also have a certain role in indicating whether the borrower is possible to default (Tao & LIN, 2016).

In addition to individual information, the borrower's interpersonal relationship and social network relationship have been areas of extensive empirical research. A large body of empirical literature found that social capital and friendship among borrowers can improve the credibility of them to a certain extent so that to promote the success of loans and reduce the default rate based on the analysis of social network relationship (Cassar & Wydick, 2010; Godlewski, Sanditov, & Burger-Helmchen, 2012; Lin, Prabhala & Viswanathan, 2013). There also exists obvious herding behavior in the online lending market. For instance, Luo and Lin (2012) observed that friend bids and bid counts impose significant effects on the decision-making time of investors, which is considered as the evidence of herding. Lee et al (2012) also found strong evidence of herding and its diminishing marginal effect as bidding advances in P2P platform in Korea.

It is not difficult to find that the interpersonal relationship between the borrower and the lender has its unique particularity in the borrower's social relationship. The borrower takes "getting" as the purpose to obtain "giving" from the lender, and the interpersonal function of these two parts is usually embodied in the linguistic communication (He, 2018; Galema, 2019). Based on this, many scholars study the descriptive text information of the loan requests. Word use is generally considered to be a meaningful marker of cognitive and social processes (Pennebaker, Mehl, & Niederhoffer, 2003). The description can reflect a borrower's identity features, such as "success", "diligence" and "reliability", which will be related with the loan success and failure. (Iarrimore, 2011; Li, 2014; Wang, 2015). For example, Herzenstein et al. (2011) found that being trustworthy or successful are associated with increased loan funding but ironically they are less predictive of loan performance compared with other identities (moral and economic hardship).

However, due to the existence of the subjectivity of language interaction, different information receiver (investor) will have different perceptions on the features of information (Manuti, Traversa, & Mininni, 2012), so extracting the borrower's identity features by reading their description subjectively would cause some deviations. As a result, some scholars summarize the features from the text itself, and study the effect of intuitive text features such as punctuation, text length and misspelling on the prediction of the success and failure of loan (Dorfleitner, 2015; Chen, Huang, & Ye, 2018). Results show that these intuitive text features have an impact on the granting decision of the lender, but they cannot effectively predict the repayment probability of the borrower.

Cannot intuitive text features predict the default? Language has its discourse function, interpersonal function and experiential function (Halliday, 1983). Interpersonal function is often related to the personal experience and personality traits of the language's expression subject and the receiving subject (Frost, Czyzewska, & Kasch, 1992). It reflects the attitude of the expression subject to the content (Zhang, 2004). A major issue is the difference of language structures between English and Chinese. Chinese characters are ideographic characters, so they have a stronger ability to expand (Foo, Schubert, & Hui, 2004). The presentation of information in Chinese text is not divided by words, but by phrases as the main carrier of expressing information (Zagibalov & Carroll, 2008). Therefore, we believe that combined with each person's educational background, individuals will have different habits when they combine phrases to express information, which will reflect the differences in individual expression ability and thus reflect the differences in personality.

Based on the existing research perspective and results, we try to extract the language features from the intuitive text features of the borrower's loan requests to support its effect on the prediction of funding success and loan default.

Hypotheses And Method

Hypotheses development

In this paper, we summarize four borrower's language features, expression redundancy degree, objectivity degree, short-sentence preference degree and punctuation control degree. Expression redundancy measures whether the borrower is wordy and verbose during the process of expressing information; Objectivity is an indicator to measure how detailed and objective when the borrower discloses his personal information; Short-sentence preference is to measure the borrower's preference for short sentences; Punctuation control will show us borrower's attention and control to punctuation. The specific definition and calculation will be explained in detail in the later section. At present, we will make assumptions about how these features influence borrowing success and default rates based on the previous researches.

When Kahneman and Tversky (1973) studied on the Intuitive predictions, they concluded that in making predictions and judgments under uncertainty, people prefer to rely on a limited number of heuristics and representativeness. As a result, investors tend to read selectively to find key information or words which can be to representativeness of borrowers' personality to access the borrowers. Therefore, if the borrower's expression is too verbose in the loan description and loan title, and prefers long sentences which may cover the keyword information, it will disturb the lender's granting decision to a certain extent. Previous studies have also included basic information about the readability of the loan description (i.e., average word and sentence length) as control variables (Pope&Sydnor,2011).In addition, the usage of punctuation in the text reflects the borrower's control and cognitive ability. Excessive usage of punctuation makes loan description informal and reduces the readability of the text (Chen & Huang, 2018). Therefore, we suppose that the detail that whether the loan description sentence ends with a punctuation mark also has a certain impact on the readability of the loan request description. Research shows that lenders prefer borrowers who disclose their personal information proactively (Jiang,Wang,&Chen, 2018). For instance, some borrowers will use their true names, telephone number as their account names and disclose their detailed profession and revenue information by quote figures in their loan request description which are not indispensable in platforms. So it can be inferred that the objectivity of the borrower's expression has positive associations with funding success. Therefore, we propose Hypothesis 1 as follows:

H1a: Redundancy degree in loan requests has negative associations with funding success

H1b: Short-sentence preference degree in loan requests has negative associations with funding success

H1c: Objectivity degree in loan requests has positive associations with funding success

H1d: Punctuation control degree in loan requests has positive associations with funding success

The loan success rate measures the lender's feedback on the borrower's information disclosure, while the repayment probability reflects the borrower's financial status and credit situation, which is often closely related to personal characteristics. Previous studies have shown that personality quality and personal identity content play an important role in predicting loan default (Herzenstein et al, 2011; Larrimore et al, 2011). The larger the amount of key information in the loan description, the less likely the borrower is to default (Yu, 2017). According to the definition of this paper, a borrower's expression redundancy degree is high means that the same text length contains less key information. The borrower's expression exists largely to maintain his/her image to others, which has the possibility of self-whitewashing and intentional inducement (Baumeister, 1982). To a certain extent, objectivity may also be one of the tools for borrowers to obtain an objective image. We speculate that borrowers with such whitewashing are

often less honest than they express. The borrower's preference for sentence length is mainly related to the lender's reading experience. Generally speaking, borrowers who prefer to use short sentences are more colloquial in language expression. Their preciseness and self-control ability may not be as good as those who prefer writing long sentences, so we predict their default probability would be higher. As mentioned above, the punctuation control in loan requests reflects the borrower's self-control ability. We speculate that borrowers with strong self-control ability have stronger constraints on their own behavior and higher repayment probability. Therefore, we propose Hypothesis 2 as follows:

H2a: Redundancy degree in loan requests has negative associations with default probability

H2b: Objectivity degree in loan requests has positive associations with default probability

H2c: Short-sentence preference degree in loan requests has positive associations with default probability

H2d: Punctuation control degree in loan requests has negative associations with default probability

Method

Data source and sample selection

The data used in this study are obtained from Renrendai (renrendai.com), one of the largest peer-to-peer lending platforms in China. This study uses all loan requests created on Renrendai between June 1, 2016 and September 30, 2018. 13,485 loan requests are selected as our research samples by random sampling. Then manually extract other required texts from Renrendai web page according to the loan number. We preprocess the collected data as follows: first, exclude the institutional guarantee, field certification and credit assurance certification; second, exclude the loan titles, blank loan description or the texts without any language features (for example, "42378", "reserved zzxnejek", etc). There are 3,567 pieces of data remained in total, and 2,720 pieces of effective data are finally retained after the abnormal data is eliminated.

Variables explanation

In this paper, for the borrower's language features, expression redundancy, objectivity, short-sentence preference and punctuation control are defined as follows:

Redundancy: expression redundancy measures whether the borrower's language expression is wordy and verbose. It is calculated by dividing loan description words by number of key information. Higher redundancy means the borrower describes less key information in more words, that is, the more useless and wordy words in loan description, the higher the redundancy. The number of key information is obtained by manual marking one by one. The effective information includes personal identity, position, residence, income, loan purpose (unclear purposes such as daily consumption and turnover are not included), car and real estate situation, repayment source, etc. Some studies have shown that the lender does have a certain investment bias for the borrower's loan purpose and other effective information (Zhuang, Zhou, & Fu, 2015).

For example, if the loan application text is "The loan is mainly used for starting a business. I have real estate without mortgage and I have a good credit with no overdue. The source of repayment is my wage income, which is about 5000 yuan per month, so there is no pressure for repayment."

Here "starting a business", "have real estate without mortgage", "without overdue" and "The source of repayment is my wage income, which is about 5000 yuan per month." are four key information.

Objectivity: an indicator to measure whether the borrower's language expression actively discloses personal information and whether the information disclosed is detailed and objective. Objectivity degree is calculated by the number of figures quoted in the loan description, which is also manually marked one by one. The result of Larrimore (2011) show that the use of quantitative words that are likely related to one's financial situation had positive associations with funding success which was considered to be an indicator of trust. We reason that the number figures quoted can to some extent affect the lender's granting decision on whether the borrower is trustworthy or not.

For example, if the loan application text is "I have twice borrowed money from Renrendai, with a good repayment record. At present, I have built 4 houses, all of which have applied for real estate certificates. I want to decorate one set for my own use. But due to the shortage of funds, I take this platform to raise funds. " Here the number of figures quoted is 2.

Short-sentence preference: an indicator to measure the borrower's preference for short sentences. Short-sentence preference degree is calculated by dividing number of punctuation marks by number of loan description words. The larger the value is, the more the borrower prefers to use short sentences. Using short sentences often can increase the readability of the text.

Punctuation control: an indicator to measure the borrower's attention to punctuation. Here, it mainly uses whether the loan description ended with punctuation as the judgment standard. Although it is only a small detail, it can reflect the borrower's standard degree to the written language format to a certain extent. At the same time, the borrower remembers punctuation at the end of the loan description can make the description text more formal. Such a borrower may have the personality characteristics of doing a thing through from beginning. We believe that the characteristics may have a predictive effect on the performance of the borrower after successfully obtaining the loan. Other variables are shown in Table 1.

Model Construction

In our study, we employ logistic regression model to study the influence of borrower's language features (redundancy, objectivity, short-sentence preference and punctuation control) on the funding probability and default probability of P2P lending market. Our empirical model is as following:

$$P(Y_i = 1) = \beta_0 + \beta_1 Red + \beta_2 Obj + \beta_3 Sen + \beta_4 Pun + \sum_{i=1}^n \beta_i X_i + \varepsilon_i$$

where the dependent variable Y_i is a binary variable equal to 1 if the borrowers successfully have their loan requests granted (or default after receiving funding) and 0 otherwise. The main explanatory variables are **Red** (redundancy), **Obj** (objectivity), **Sen** (short-sentence), **Pun** (punctuation), the borrower's language expression features extracted in this paper: expression redundancy, objectivity, short-sentence preference and punctuation control. X_i is a vector of control variables, including amount of borrowing, interest rate, credit score of the borrower, age, education level, income, marriage status, working experience, etc., and ε is the random disturbance term.

Descriptive statistics

In this paper, first of all, descriptive statistics are conducted for the variables, and the results are shown in Table 2. As a result, our sample includes 2,720 loan requests, of which 2,474 were successfully funded while the remaining 246 were not funded. Among all requests that were successfully funded, there were 1,331 defaults and 1,143 loans were repaid on time. In addition, we can roughly see the demographic characteristics of the online loan market. Male borrowers are more active in the online loan market (the mean value of gender is 0.8739), mainly aged between 20 and 40 years old. Married people account for a large proportion (the mean value is 0.6235). The average academic background of the borrowers participating in the online loan is above college, and their educational level is higher than the social average. The average worktime of the borrowers is between 1 and 3 years, but they always have low income level (the average income is 2.9353).

Table 3 lists the correlation coefficients for all four key explanatory variables used in our analysis. Data shows that there exists correlation among the four key explanatory variables, but the correlation is relatively low. It also suggests that multicollinearity problems don't arise, so subsequent regression processing can be carried out.

Results

Language features and loan success rate

In this subsection, we examining the four variables that reflect the borrower's language features and success rate. A total of 836 samples were selected in the empirical regression process. We report the results in Tables 4. In Table 4, column (1) show that borrowers with higher credit rating, older age, longer worktime, higher educational background are more likely to be funded. Lenders are more willing to choose the requests with lower loan amount. In column (2), we mainly research the relationship between the four main independent variables and the loan success rate without excluding the mixed effect of control variables. Results show that the expression redundancy, objectivity and punctuation control are all significant at the level of 5%. In column (3), except the mixed effect of control variables, the three variables are still significant at the level of 10%. Compared with (1), the regression model (3) with the four variables of redundancy, objectivity, short-sentence preference and punctuation control degree has a certain improvement. As a result, H1c and H1d are supported but we failed to support H1a and H1b.

Language features and loan default rate

In Table 5, we mainly try to support Hypothesis 2 by study the relationship between the four language features variables and default rate. In the process of empirical regression, 2,474 samples were selected. Among them, column (4) shows that the lower the credit rating of the borrower is, the older the borrower is, and the lower education the borrower has, the more likely it is that the borrower will default. And borrowers with high income have higher default risk and so they have higher interest rate. In column (5), redundancy and punctuation control are negatively related to the default rate at the significance level of 5%. In column (6), except the mixed effect of the control variables, those two variables are still significant at the level of 5%, and the direction of significance remains unchanged, and they are still negatively related to the default rate. Compared with the regression model (4), regression model (6) which includes the four variables has improved the goodness of fit and prediction accuracy to a certain extent. Here, H2d is supported while H2a, H2b and H2c are failed to be supported.

Language features and education levels

After studying the influence of the borrower's language features on the loan success rate and default rate, we further trace the main factors that affect the borrower's language features. Previous studies have agreed that education has a significant effect on the loan success rate and loan default rate of P2P lending (Gathergood, 2012). Therefore, we add four interaction terms between language features and education levels respectively for further research.

The regression results in Table 6 show that in the prediction model of loan success rate, the adjustment among education, redundancy and punctuation control degree is not significant (column (8) and (9)), the interaction item of "*Obj*Edu*" is significant at the level of 5% (column (8)), and the interaction coefficient is negative, indicating that the higher the education level is, the weaker the positive effect of objectivity on the success rate of loan is. Education level has negative effect on the relationship between objectivity and the success rate. Similarly, in the prediction model of loan default rate, there is no significant moderating effect between education level and punctuation control (column (11)), but the regression coefficient before the interaction item "*Red*Edu*" is significant negative at the level of 10% (column (10)). That is to say, the higher the education level of the borrower, the stronger the negative effect of redundancy on the default rate. We can conclude that the education level of the borrower plays a certain role in regulating the relationship between language features of the borrower and the loan success rate or default rate.

Robustness test

This subsection performs two kinds of robustness tests to ensure the validity of the empirical results reported in the previous subsections. Method 1: Resampling. When verifying hypothesis 1, 836 samples are collected, among which the ratio of successful funding to failed funding is about 1:4. In the robustness test, we resample 681 sets of data, in which the ratio of successful funding to failed funding is about 1:2. Method 2: Adjusting the control variables. In the logistic regression model mentioned above, we mainly introduce nine control variables, such as gender, age, marital status. When verifying the robustness of the loan default prediction model, we add another two control variables, the number of loans and the number of successful loans, a total of 11 control variables. The verification results are consistent with the previous results. However, limited by the article space, we do not list the formula here, but our empirical results are robust.

Discussion

Actually, building trust and reduce information asymmetry is a critical issue to p2p online lending.

At present, a large number of studies focus on the borrowers' demographic characteristics (e.g, gender, age and income), social network relationship, and little attention is paid to their nicknames, loan titles, loan descriptions they use when applying for loans. This unverifiable information can show the borrower's personality and helps the lender judge the borrower's integrity. However, descriptive information is not specific and concrete enough which need to be further processed and extracted to build their relationship with loan success and loan default. This paper summarize four language features of borrowers to analyze their prediction effect on the success rate and default rate of loans in P2P network lending platform. As the non-quantized soft information of the borrower, the language features of the borrower play a certain role in alleviating the information asymmetry of P2P lending and can measure the credit level of the borrower.

Specifically, the results in terms of language features and loan success show that lenders prefer those borrowers who quote more figures in their loan description. Larrimore et al (2011) also concluded that quantitative words had positive associations with funding success. Our conclusion is consistent with his results. We find that borrowers who pay more attention to punctuation details are more likely to achieve a loan while Chen et al (2018) supposed that amount of punctuation is negatively associated with the funding probability. It is because that our focus on punctuation is different, we uses whether the loan description ended with punctuation as the judgment standard to access the reliability of borrowers not the amount of punctuation they used in loan description. Expression redundancy will reduce the readability to some extent, but it contains more key information to improve the credibility. This result is similar with the one of Dorfleitner's (2016) conclusions that the length of the description text in a loan application is positively related to funding success.

When it comes to language features and loan default, our results indicate that the borrower's control over the punctuation details in the loan description process can really reflect the borrower's credit degree. The attention to the statement format specification can also reflect the borrower's self-control to a certain extent, which indirectly reflects the borrower's control over the repayment behavior. We also conclude that the more redundant the language expression is, the lower the loan default rate will be, that means the borrower's 'redundancy' in the loan description and loan title can show the borrower's attention to the loan behavior through repeated description. This attention constitutes a self-realization of psychological expectation for the repayment behavior in the future, which drives the borrower to a success repayment (Sehlender & Weigold, 1992; Berger & Heath, 2007). In addition, the correlation between short-sentence preference, objectivity and the default rate is not significant. We reasoned that the sentence length preference in language description is just a kind of expression habit of the borrower, which has little effect on the default. Borrower's 'Objectivity' is an objective statement of their personal situation, but it can not guide and restrain their post-loan performance, so the correlation is not significant here.

Implications for Practice

Our research indicates that there are several strategies borrowers' should use to increase trust and thus the likelihood of getting funded. First, borrowers can extend the length of loan description and try to be more wordy and verbose so that it can convey more information for lenders to judge the borrowers' identity. An additional effective strategy to attract lenders is to quote more figures when describe the loan. Borrowers also should pay more attention to the use of punctuation. Remember to use punctuation at the end of loan description will give lenders a feeling of reliability. To lenders, they can give more chance to borrowers with higher education and learn to extract information form borrowers' loan title, loan description to access whether they are trustworthy. From our empirical results, the borrower who is verbose and has a good control of punctuation is less likely to default. There is no doubt our research conclusions can mitigate the issue of information asymmetry between borrowers and lenders.

Limitation and future study

This study is subject to several limitations. First, the number of failed loan samples in our selected study interval is relatively small, leading to some deviation in the research results. Researchers can further expand the study interval. Second, our data includes the personal information of borrowers and whether they fund successfully and default, we have no additional information about lenders. Different lenders will have a different preference on the subject of the loan. Therefore, additional research could capture the mental maps of lenders based on the perspective of decision makers. Third, in order to ensure the objectivity of the definition of language features, we did not pay much attention

to the content of description. However, as a rich and varied media, narrative language is far more complex than our exploration. The narrative content may affect the lender's mood, or the lender's evaluation of the borrower and the quantification of this impact is also one of the directions of future research.

Declarations

Acknowledgement

I would like to express my deepest appreciation to all who have helped me complete this study. And the project was sponsored by National Social Science Fund of China (ID: 18BJY251), The Ministry of education of Humanities and Social Science project (ID:17YJC790107).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Baumeister, R. (1982). A self-presentational view of social phenomena. *Psychological Bulletin*, 91(1), 3-26.
2. Berger, J., & Heath, C. (2007). Where Consumers Diverge from Others: Identity Signaling and Product Domains. *Journal Of Consumer Research*, 34(2), 121-134.
3. Cassar, A., & Wydick, B. (2010). Does social capital matter? Evidence from a five-country group lending experiment. *Oxford Economic Papers*, 62(4), 715-739.
4. Chen X, Ding X, Wang B. (2013). A Study of the Overdue Behaviors in Private Borrowing- Empirical Analysis Based on P2P Network Borrowing and Lending . *Finance Forum*, (11):65-72.
5. Chen, X., Huang, B., & Ye, D. (2018). The role of punctuation in P2P lending: Evidence from China. *Economic Modelling*, 68, 634-643.
6. Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., & Kammler, J. (2016). Description-text related soft information in peer-to-peer lending – Evidence from two leading European platforms. *Journal Of Banking & Finance*, 64, 169-187.
7. Frost, C., Czyzewska, M., & Kasch, L. (1992). Reality Construction and Presentational Language: A Social-Cognitive Perspective on Prejudice. *Humanity & Society*, 16(2), 197-216.
8. Foo, S., & Li, H. (2004). Chinese word segmentation and its effect on information retrieval. *Information Processing & Management*, 40(1), 161-190.
9. Gathergood, J. (2012). Self-control, financial literacy and consumer over-indebtedness. *Journal Of Economic Psychology*, 33(3), 590-602.
10. Halliday, M. (1983). On the transition from child tongue to mother tongue. *Australian Journal Of Linguistics*, 3(2), 201-215.
11. He Z. (2018). Interpersonal Pragmatics: the knowledge of using language to deal with interpersonal relationship . *Foreign Language Education*, 39(06):1-6.
12. Herzenstein, M., Sonenshein, S., & Dholakia, U. (2011). Tell Me a Good Story and I May Lend You My Money: The Role of Narratives in Peer-to-Peer Lending Decisions. *SSRN Electronic Journal*.

13. Jiang X, Wang B, Cheng J. (2018) The Effectiveness of Voluntary Information Disclosure- Based on the Research of PaiPai Loan. *Review of Economy and Management*, 34(06):95-105.
14. Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237-251.
15. Pennebaker, J., Mehl, M., & Niederhoffer, K. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review Of Psychology*, 54(1), 547-577.
16. Yu J. (2017). A Study on the Relationship between Descriptive Information and Default Behaviors: The Analyze Based on P2P Lending Platform. *Contemporary Economy & Management*, 39(05):86-92.
17. Galema, R. (2019). Credit Rationing in P2P Lending to SMEs: Do Lender-Borrower Relationships Matter?. *SSRN Electronic Journal*.
18. Godlewski, C., Sanditov, B., & Burger-Helmchen, T. (2012). Bank Lending Networks, Experience, Reputation, and Borrowing Costs: Empirical Evidence from the French Syndicated Lending Market. *Journal Of Business Finance & Accounting*, 39(1-2), 113-140.
19. Gonzalez, L., & Komarova Loureiro, Y. (2014). When Can a Photo Increase Credit?: The Impact of Lender and Borrower Profiles on Online P2P Loans. *SSRN Electronic Journal*.
20. Herzenstein, M., Sonenshein, S., & Dholakia, U. (2011). Tell Me a Good Story and I May Lend You My Money: The Role of Narratives in Peer-to-Peer Lending Decisions. *SSRN Electronic Journal*.
21. Hu, M., Li, X., & Shi, Y. (2018). Adverse Selection and Credit Certificates: Evidence from a P2P Platform. *SSRN Electronic Journal*.
22. Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., & Gorski, S. (2011). Peer to Peer Lending: The Relationship Between Language Features, Trustworthiness, and Persuasion Success. *Journal Of Applied Communication Research*, 39(1), 19-37.
23. Lee, E., & Lee, B. (2012). Herding behavior in online P2P lending: An empirical investigation. *Electronic Commerce Research And Applications*, 11(5), 495-503.
24. Li Y, Gao Y, Li Z, et al.(2014) The Influence of Borrower's Description on Investors' Decision- Analyze Based on P2P Online Lending . *Economic Research Journal*, 49 (Z1) :143-155.
25. Liao L, Ji L, Zhang W. (2015). Education and Credit: Evidence from P2P Lending Platform . *Journal of Financial Research*, (3):146-159.
26. Lin, M., Prabhala, N., & Viswanathan, S. (2013). Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending. *Management Science*, 59(1), 17-35.
27. Luo, B., & Lin, Z. (2011). A decision tree model for herd behavior and empirical evidence from the online P2P lending market. *Information Systems And E-Business Management*, 11(1), 141-160.
28. Manuti, A., Traversa, R., & Mininni, G. (2012). The dynamics of sense making: a diatextual approach to the intersubjectivity of discourse. *Text & Talk - An Interdisciplinary Journal Of Language, Discourse & Communication Studies*, 32(1), 39-61.
29. Pope, D., & Sydnor, J. (2011). What's in a Picture?: Evidence of Discrimination from Prosper.com. *Journal Of Human Resources*, 46(1), 53-92.
30. Puro, L., Teich, J., Wallenius, H., & Wallenius, J. (2010). Borrower Decision Aid for people-to-people lending. *Decision Support Systems*, 49(1), 52-60.
31. Schlenker, B., & Weigold, M. (1992). Interpersonal Processes Involving Impression Regulation and Management. *Annual Review Of Psychology*, 43(1), 133-168.

32. Simkins, B., & Rogers, D. (2004). Asymmetric Information and Credit Quality: Evidence from Synthetic Fixed-Rate Financing. SSRN Electronic Journal.
33. Tao, Q., & LIN, Z. (2016). Who Can Get Money? Evidence from the Chinese Peer-To-Peer Lending Platform. SSRN Electronic Journal.
34. Wang H, He L. (2015). An Empirical Study of Borrowing Description's Influence on P2P Lending. Journal of Finance and Economics, (1):77-85.
35. Wang S, Kong X, Xu Y. (2014). Internet trust in Internet Finance: formation mechanism, evaluation and improvement-A Study Based on Peer-to-Peer Lending. Financial Regulation Research, (05):67-76.
36. Zagibalov, T., & Carroll, J. (2008). Automatic seed word selection for unsupervised sentiment classification of Chinese text. Proceedings Of The 22Nd International Conference On Computational Linguistics - COLING '08.
37. Zhuang L, Zhou Q, Fu Y. (2015) P2P Network Lending Purposes, Investment Bias Effect and Soft Information Value. International Business, (06):77-85.

Tables

Table 1: Variables and definitions.

Variable	Name	Definition
Explained variable	<i>Success</i>	1 if a loan listing is fully funded and 0 otherwise
	<i>Default</i>	1 if the funded loan has been defaulted and 0 otherwise
Explanatory variable	<i>Redundancy</i>	The number of Chinese characters used by the borrower in a loan description/The number of key information
	<i>Objectivity</i>	The number of figures used in a loan description
	<i>Short-Sentence</i>	The number of punctuation marks used in a loan description/The number of Chinese characters
	<i>Punctuation</i>	1 if a loan description ended with punctuation and 0 otherwise
Control variables	<i>Credit</i>	Credit grade of the borrower at the time the listing was created. Credit grade takes on values between 1 (AA) and 7(high risk)
	<i>Age</i>	Age of the borrower in years
	<i>Gender</i>	1 if borrower is a man and 0 otherwise
	<i>Marriage</i>	0 if borrower is unmarried and 1 otherwise (married, divorced or widowed)
	<i>Education</i>	Education level of borrower, 1=middle/high school and below, 2=college,3=undergraduate, 4=graduate school and above
	<i>Income(in RMB)</i>	Income level of the borrower, 1=Less than 1000, 2=1000–1999, 3=2000–4999, 4=5000–9999, 5=10000–19999, 6=20000–49999, 7=50000 and more than 50000, per month
	<i>Worktime</i>	Borrower's working experience, 1=1 year and less than 1 year, 2=1–3 years (including 3 years), 3=3–5 years (including 5 years), 4=more than 5 years
	<i>Interest</i>	The rate that the borrower pays on the loan
	<i>Money</i>	Loan amount requested by the borrower

Table 2: Variables and descriptive statistics.

Variable	N	Min	Max	Mean	Sd
<i>Success</i>	2720	0	1	0.91	0.29
<i>Default</i>	2474	0	1	0.54	0.50
<i>Redundancy</i>	2720	-0.04	2	0.19	0.13
<i>Objectivity</i>	2720	0	19	1.18	1.62
<i>Short-Sentence</i>	2720	0	0.31	0.10	0.04
<i>Punctuation</i>	2720	0	1	0.63	0.48
<i>Credit</i>	2720	1	7	6.45	0.85
<i>Age</i>	2720	22	58	33.19	6.64
<i>Gender</i>	2720	0	1	0.87	0.33
<i>Marriage</i>	2720	0	1	0.62	0.48
<i>Education</i>	2720	1	4	2.15	0.78
<i>Income</i>	2720	1	8	2.94	1.07
<i>Worktime</i>	2720	1	4	2.85	1.01
<i>Interest</i>	2720	8	13	11.90	1.09
<i>Money</i>	2720	8.01	13.12	9.72	0.72

Table 3: Correlations between all explanatory variables.

	<i>Redundancy</i>	<i>Objectivity</i>	<i>Short-Sentence</i>	<i>Punctuation</i>
<i>Redundancy</i>	1			
<i>Objectivity</i>	-0.34***	1		
<i>Short-Sentence</i>	0.09***	-0.09***	1	
<i>Punctuation</i>	-0.14***	0.13***	0.43***	1

Notes: *p < 0.1, **p < 0.05, ***p < 0.01.

Table 4: Logit regression results on funding success.

	<i>Success</i> (1)	P	<i>Success</i> (2)	P	<i>Success</i> (3)	P
<i>Redundancy</i>			4.19*** [0.74]	0.00	1.85* [1.06]	0.08
<i>Objectivity</i>			0.10** [0.05]	0.05	0.16** [0.08]	0.04
<i>Short-Sentence</i>			1.91 [1.94]	0.33	2.21 [2.91]	0.45
<i>Punctuation</i>			0.59*** [0.17]	0.00	0.432* [0.255]	0.09
<i>Credit</i>	-1.59*** [0.21]	0.00			-1.59*** [0.21]	0.00
<i>Age</i>	0.05** [0.02]	0.01			0.05** [0.02]	0.01
<i>Gender</i>	-0.07 [0.35]	0.85			-0.11 [0.36]	0.75
<i>Marriage</i>	0.12 [0.25]	0.64			0.07 [0.25]	0.78
<i>Education</i>	0.84*** [0.16]	0.00			0.80*** [0.16]	0.00
<i>Income</i>	-0.04 [0.11]	0.74			-0.07 [0.12]	0.56
<i>Worktime</i>	0.24** [0.12]	0.04			0.17 [0.12]	0.16
<i>Interest</i>	-0.16 [0.12]	0.17			-0.15 [0.12]	0.23
<i>Money</i>	-1.74*** [0.18]	0.00			-1.73*** [0.18]	0.00
C	26.52*** [2.18]	0.00	-0.80 [0.24]	0.00	25.65*** [2.23]	0.00
N	836		836		836	
Pseudo R ²	0.65		0.09		0.66	
Percentage	87.10%		66.70%		88.20%	

correct

Notes: *p < 0.1, **p < 0.05, *** p < 0.01.

Table 5: Logit regression results on default rate

	<i>Default</i> (4)	P	<i>Default</i> (5)	P	<i>Default</i> (6)	P
<i>Redundancy</i>			-3.32*** [0.37]	0.00	-2.38*** [0.50]	0.00
<i>Objectivity</i>			0.04 [0.03]	0.17	-0.04 [0.038]	0.36
<i>Short-Sentence</i>			-1.34 [1.16]	0.25	0.71 [1.54]	0.64
<i>Punctuation</i>			-0.36*** [0.10]	0.00	-0.26** [0.13]	0.05
<i>Credit</i>	1.69*** [0.09)	0.00			1.68*** [0.09]	0.00
<i>Age</i>	0.07*** (0.01)	0.00			0.07*** [0.01]	0.00
<i>Gender</i>	0.14 [0.17]	0.42			0.09 [0.17]	0.61
<i>Marriage</i>	0.12 [0.13]	0.34			0.15 [0.13]	0.23
<i>Education</i>	-0.71*** [0.08]	0.00			-0.67*** [0.08]	0.00
<i>Income</i>	0.32*** [0.07]	0.00			0.32*** [0.07]	0.00
<i>Worktime</i>	-0.09 [0.06]	0.12			-0.08 [0.06]	0.22
<i>Interest</i>	0.70*** [0.06]	0.00			0.68*** [0.06]	0.00
<i>Money</i>	0.13 [0.11]	0.23			0.13 [0.11]	0.26
C	-21.80*** [1.28]	0.00	1.11*** [0.14]	0.00	-21.03*** [1.29]	0.00
N	2474		2474		2474	
Pseudo R ²	0.55		0.07		0.56	
Percentage	81.40%		59.70%		81.80%	

correct

Notes: *p < 0.1, **p < 0.05, *** p < 0.01.

Table 6: Interactive regression results.

	<i>Success</i>	P	<i>Success</i>	P	<i>Success</i>	P	<i>Default</i>	P	<i>Default</i>	P
	7		8		9		10		11	
<i>Redundancy</i>	-2.20 [3.27]	0.50	2.27** [1.09]	0.04	1.92* [1.07]	0.07	0.16 [1.36]	0.91	-2.33*** [0.50]	0.00
<i>Objectivity</i>	0.15 [0.08]	0.06	0.80*** [0.19]	0.00	0.15** [0.08]	0.04	-0.03 [0.04]	0.42	-0.03 [0.04]	0.47
<i>Short-Sentence</i>	2.37 [2.93]	0.42	1.83 [2.98]	0.54	2.49 [2.93]	0.40	0.83 [1.55]	0.59	0.76 [1.54]	0.62
<i>Punctuation</i>	0.44 [0.26]	0.08	0.59** [0.26]	0.02	-0.52 [0.72]	0.48	-0.27** [0.13]	0.04	-0.20 [0.37]	0.60
<i>Red*Edu</i>	1.83 [1.42]	0.20					-1.14* (0.59)	0.06		
<i>Obj*Edu</i>			-0.34*** [0.08]	0.00						
<i>Pun*Edu</i>					0.45 [0.32]	0.16			-0.03 [0.16]	0.86
<i>Credit</i>	-1.60*** [0.21]	0.00	-1.70*** [0.22]	0.00	-1.62*** [0.21]	0.00	1.69*** [0.09]	0.00	1.67*** [0.09]	0.00
<i>Age</i>	0.05** (0.02)	0.01	0.04** [0.02]	0.04	0.05*** [0.02]	0.01	0.07*** [0.01]	0.00	0.07*** [0.01]	0.00
<i>Gender</i>	-0.08 [0.36]	0.82	-0.07 [0.37]	0.84	-0.16 [0.36]	0.65	-0.08 [0.17]	0.65	-0.08 [0.17]	0.64
<i>Marriage</i>	0.09 [0.26]	0.71	0.19 [0.26]	0.46	0.05 [0.25]	0.84	-0.16 [0.13]	0.21	-0.16 [0.13]	0.21
<i>Education</i>	0.46 [0.31]	0.13	1.12*** [0.19]	0.00	0.55** [0.24]	0.02	-0.45*** [0.14]	0.00	-0.65*** [0.12]	0.00
<i>Income</i>	-0.06 [0.11]	0.62	-0.08 [0.12]	0.51	-0.09 [0.12]	0.46	0.31*** [0.07]	0.00	0.31*** [0.07]	0.00
<i>Worktime</i>	0.17 [0.12]	0.15	0.19 [0.12]	0.13	0.16 [0.12]	0.19	-0.07 [0.06]	0.24	-0.08 [0.06]	0.23
<i>Interest</i>	-0.13 [0.12]	0.27	-0.17 [0.12]	0.16	-0.15 [0.12]	0.22	0.68*** [0.06]	0.00	0.68*** [0.06]	0.00
<i>Money</i>	-1.78***	0.00	-1.74***	0.00	-1.72***	0.00	0.13	0.23	0.13	0.25

	0.19		0.19		0.18		0.11		0.11	
C	26.63***	0.00	26.10***	0.00	26.31***	0.00	-21.39***	0.00	-20.83***	0.00
	2.39		2.30		2.30		1.33		1.30	
N	836		836		836		2474		2474	
Pseudo R2	0.66		0.67		0.66		0.56		0.56	
	88.40%		88.20%		88.40%		81.70%		81.80%	

Notes: *p < 0.1, **p < 0.05, *** p < 0.01.