

An Effective Model between Mobile Phone Usage and P2P Default Behavior

Huan Liu¹, Lin Ma^{2,3*}, Xi Zhao^{2,4}, and Jianhua Zou¹

¹ School of Electrical and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

² School of Management, Xi'an Jiaotong University, Xi'an 710049, China

³ State key laboratory for manufacturing systems engineering, Xi'an 710049, China

⁴ Shaanxi Engineering Research Center of Medical and Health Big Data, Xi'an 710049, China

liuhuanli@stu.xjtu.edu.cn, malinxjtu@163.com, zhaoxi1@mail.xjtu.edu.com, jhzou@sei.xjtu.edu.cn

Abstract. P2P online lending platforms have become increasingly developed. However, these platforms may suffer a serious loss caused by default behaviors of borrowers. In this paper, we present an effective default behavior prediction model to reduce default risk in P2P lending. The proposed model uses mobile phone usage data, which are generated from widely used mobile phones. We extract features from five aspects, including consumption, social network, mobility, socioeconomic, and individual attribute. Based on these features, we propose a joint decision model, which makes a default risk judgment through combining Random Forests with Light Gradient Boosting Machine. Validated by a real-world dataset collected by a mobile carrier and a P2P lending company in China, the proposed model not only demonstrates satisfactory performance on the evaluation metrics but also outperforms the existing methods in this area. Based on these results, the proposed model implies the high feasibility and potential to be adopted in real-world P2P online lending platforms.

Keywords: P2P default behavior prediction · Mobile phone usage · Joint decision model.

1 Introduction

The P2P (peer-to-peer) online lending platforms provide micro-credit services by playing a mediating role between individual lenders and borrowers. Compared with traditional lending institutions, these platforms show lower costs, convenient conditions, and quick loan process. For above advantages, more and more individuals and investors are attracted by P2P platforms, especially in developing countries. In China, the online lending industry shows transaction size had reached 28 thousand billion RMB, increasing 137% over than 2015 [11]. The number of P2P platforms had grown to 2307 in 2016, which increase year-on-year by 2.81%.

However, the investment of lenders on P2P platforms may suffer a serious loss caused by default behaviors of borrowers, which may cause a critical customer churn problem to the platforms. In order to reduce risk in P2P lending, the platforms generally adopt risk control mechanism to filter some high default risk borrowers. Actually, the risk control mechanism may face serious challenges from several perspectives. First, to ensure profitability of platform, the cost on risk control must be as low as possible, which causes a high limit in restricting the facticity inspect of individuals information. Second, without other monitoring mechanism required by traditional banks, a pre-approval credit checking process is crucial to decrease the loss of default. Third, since the target customs are the mass individuals, the credit control mechanism must have the capability to handle users without or limited credit records in the credit behavior. All these challenges put forward for an automated risk control mechanism, which provides pre-approval credit estimate with high accuracy and reliable data source.

The growing need has motivated several studies in reducing the risk for P2P lending. Based on credit related records, such as FICO, credit history, etc., some researchers reduce the risk by rejecting loans with high potential default risk [5], by transferring the problem to a portfolio optimizing investment decision problem [8], or by replacing default loss as profit scoring to increase the overall income [22]. Other researchers try to find the connection between default behavior and soft information[26, 29, 3, 7]. All these aforementioned studies are effective to reduce the risk of P2P lending. However, there still exist several questions when applying on developing countries. Due to the immature credit system, not all borrowers have credit records. And the mass applicants make it difficult for platforms to verify off-line self-reported applications. These restrictions narrow the generality of the methods.

In this paper, we present a general and reliable joint decision model to predict default behaviors on P2P lending platform from mobile phone usage data. Mobile phone usage data contains a series of records from the call, message, data volume, and App usage. The great value of mobile phone usage data has already been discovered in analysing user behaviors, personality traits, socioeconomic status, consumption patterns, and economic characteristics [15, 16, 23, 28, 20, 17, 13], which are correlated with credit default behavior [26, 29, 3, 7, 6, 12]. Moreover, the ubiquity of mobile phones guarantees the extensive application of the proposed model, and the portability and versatility of smartphones ensure the data volume and multi-descriptions of each individual, and the automatic generating characteristic ensures the facticity of data. Supported by above conclusions, the proposed model using mobile phone usage data has great potential and advantages in predicting P2P default behavior.

The main contributions of this paper are threefold. (1) We present a risk control mechanism for P2P online lending platforms, which can realize automated and agile loan approval. (2) We propose a quantitative model to predict the default behavior of individuals, which can be implemented in the risk control mechanism of P2P online lending platforms. (3) We verify our proposed model on a real-world dataset, and gain satisfactory performance not only on

the evaluation metrics but also on the comparison with existing models in this area.

2 Related work

P2P online lending served as a marketplace for individuals to directly borrow money from others through Internet [1]. Benefit from the services with lower charge and without any confining of space [8, 30], P2P lending and platforms are growing rapidly. However, limited by information asymmetry and guarantee fund, platforms cannot perform precision default assessment for each loan applicant, which may lead to a high default rate. This situation attracts researchers to study increasing the profit of lenders and reducing the default rate of borrowers. In this work, we focus on the particular problem of building a quantitative model to predict individual default behavior on P2P loan repayment, which acts as a pre-approval credit checking in decreasing the risk for P2P lending.

Some researchers focus on recognizing default behavior of loan applicants by using financial and credit data. Emekter et al. [5] measured loan performances by credit records and historical data from LendingClub. Using the same data source, Polena and Regner [19] defined different ranks of loan risk. Different technologies also were used to predict defaults probability on borrowers, such as random forest classification [14], Bayesian network [27], logistic regression [21], decision tree [29], fuzzy SVM algorithm [25]. When data about individuals' credit is available, these methods achieved high precision on evaluating credit. However, limited by collecting credible individual data, the performance of the methods may decrease when applying on developing countries.

Other researchers try to understand the correlation between individual default behavior and soft information that can be correlated with the default probability. J Gathergood [6] inferred personality traits and socioeconomic status correlated with credit behavior. Lin et al. [12] found that the significant and verifiable relational network associated with a high possible on low default risk. Chen et al. [3] studied relationships between social capital and repayment performance, discovering that borrowers structural social capital may have a negative effect on his/her repayment performance. Zhang et al. [29] used social media information to constitute a credit scoring model. Wang et al. [26] studied the connection between borrowers self-report loan application documents and the risk of loans by text analysis. Gonzalez and Loureiro [7] focused on the characteristics of both lender and borrower on the P2P lending decision. These studies illustrate the existing relationship between soft information and credit scoring, especially prove that individuals' behaviors on other perspectives can affect default behavior.

Mobile phone usage data have been studied for modeling users and community dynamics in a wide range of applications. In [15, 16, 23], mobile phone usage data were used for modeling users, such as inferring personality traits and socioeconomic status. In [9, 10], phone usage data have already been used for analyzing behavior and psychology. Chiara Renso et al. [20] proposed meth-

ods on movement pattern discovery and human behavior inference. Parent et al. [17] summarized the approaches on mining behavior patterns from semantic trajectories. Mobile phone usage data can also reflect one’s purchase habits and natural attributes [28]. Liu et al. [13] proposed a model to extract factors from trajectories and construct the connection between these factors and rationality decisions. All these studies proved the close relationship between phone usage data and human reactions to socio-economic activities, which can affect default behavior as previously discussed. To the best of our knowledge, in the default behavior prediction on P2P online lending, we are the first to build a machine learning model to predict P2P default behavior using mobile phone usage data.

3 Mechanism Overview and Data Description

3.1 Mechanism Overview

The main purpose of risk control mechanism is to reduce the default rate of borrowers. According to the adoptive common mechanism on P2P lending platforms [30], we design the mechanism as demonstrated in Fig. 1. When a borrower applies for loans on a P2P platform, the risk control mechanism is triggered. Firstly, the loan approval process encrypts borrower’s ID and sends it to risk control service provider via API. Secondly, risk control service performs the default prediction and sends the result back. Thirdly, depending on the assessment result, loan approval process decides whether or not post the borrower’s loan application. Finally, if the loan application is posted online, lenders access the application and conclude the transaction. In order to preserve the privacy of borrowers, phone usage data are kept within risk control service providers. In this mechanism, the risk control service provider refers to a mobile carrier. As soon as risk control service received the loan request, it decrypts the encrypted ID and retrieves the applicant’s phone usage data. Then, the default prediction model analyses the borrower’s daily behavior and predicts the default probability of borrower and returns assessment consequence to the P2P platform. The detail of the prediction model is introduced in section 4.

3.2 Data Description

Mobile phone usage dataset Mobile phone usage data consists individuals demographic information and telecommunication services records, which contain detailed call, message, and data volume. These records are generated during the communication between a mobile phone and base transceiver stations (BTSs) of its carrier. Generally, a specific BTS, automatically selected according to the distance and signal strength, provides the requested services while logs detailed phone usage behaviors. Our mobile phone usage dataset is from one of the mobile carriers in China. Specifically, for message service, the recorded information includes the time stamp and the contact ID. For phone call service, the location and call duration is added to the aforementioned items. Both these records can

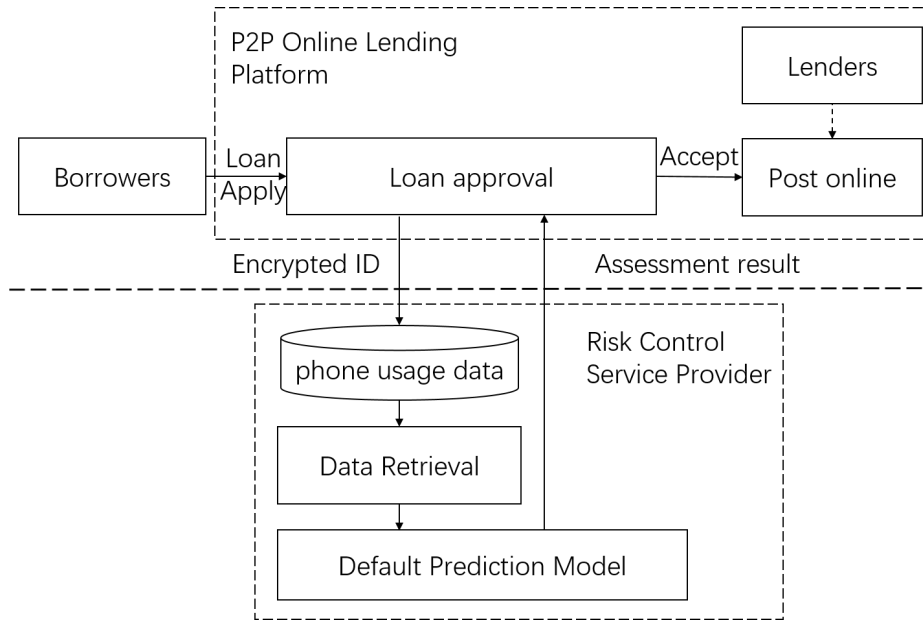


Fig. 1. A figure caption is always placed below the illustration. Please note that short captions are centered, while long ones are justified by the macro package automatically.

describe when and where individual contact others by phone or message. For data volume service, the detail information contains the time stamp, the location, and the data volume. In addition, we obtain the statistical data for each App on the frequency and data volume spend in every month. Besides these direct information from the records, users’ movement behaviors can be implied by locations of the selected BTSs. Despite losing a large volume of content in data such as message texts, voices during calls, and App data, these meta-level records reach a good balance between user privacy and behavioral representation power.

Actual default behavior dataset Our actual default behavior dataset of borrowers is from a P2P lending company in China, which contains 3027 subjects. Before advancing this study, the ethical problem of collecting and analyzing subjects’ behavior data requires careful consideration. The ethical and legal approval is granted by the contract we signed. The data has been anonymized on subjects’ name, ID, and phone numbers. Encryption techniques are applied by mobile carriers. It’s impossible for us to decrypt and identify the participants.

4 Methodology

In this section, we will discuss the default behavior prediction sub-process, as the decisive role of the risk control process. Based on the realization procedure, we separate the sub-process into two parts. First, we extract features from mobile phone usage data on five aspects. Second, we build a joint decision model for the default behavior prediction combining two popular machine learning algorithms.

4.1 Feature extraction

According to the existing feature pools on mobile data [9, 10, 18] and characteristics of our data, we extract a set of features conveying user behavioral information from 5 aspects, including consumption, social network, mobility, socioeconomic, and individual attribute. These features describe the phone usage behavior from different fields, as depicted in Table 1.

Table 1. Extracted Features from five different aspects.

Feature Set	Features Clusters	Records Type	Number
Consumption features	Communication Consumption	Calls & Messages	22
	MONET Consumption	Data Volume	6
	Telecommunication Consumption	Basic information	10
	Consumption Entropy	Calls & Messages	4
Social network features	Connections Quantity	Calls & Messages	2
	Connections Entropy	Calls & Messages	2
Mobility features	Mobility Sphere	Calls & Data Volume	2
	Mobility Quantity	Calls & Data Volume	8
	Mobility Entropy	Calls & Data Volume	3
Socioeconomic features	Age & Gender	Basic information	2
Individual attribute features	App Frequency	App usage data	8
	App Data Volume	App usage data	6
	Specific App Usage Behavior	App usage data	17

Consumption features Consumption features reflect the amount of usage on the communications network, and we provide a high-level view of the statistical criteria for calls, SMS, and internet usage.

Communication Consumption Statistics of usage time on call, SMS and Internet services, including the average, the maximum, the minimum number, the variance of usage frequency in one day, and the number of days that have records ,The number and the proportion of communications during the night(19pm to 7am of next day). The rate of communications occurred at home or at the workplace. The interval refers to the time interval between two interactions, including the average, the maximum, the minimum, the variance number.

MONET Consumption Statistical features focus on the Data Volume records occurred when the individuals using mobile internet, including the average, the maximum, the minimum number, the variance of usage frequency in one day, and the number of days that have records ,The number and the proportion of internet usage during the night(19pm to 7am of next day).

Telecommunication Consumption Individuals telecommunication service records, which consist of shutdown times in last year, total data volume used in last year, total expenditure on the mobile phone in last year, the number and cost of international and internal roams days in last year, time of network, star level.

Consumption Entropy We compute the number of call and SMS for different temporal partitions: by day, and by the time of the day (eight periods of time, 0 am to 3 am, 3 am to 6 am). We use Shannons entropy to compute communications day entropy and communications time entropy. The former can reflect the usage time regularity in every day of one month, and the latter reflects the usage time regularity in eight periods of one month.

Social network features Social network features are related to the characteristics of the graph of connections between different individuals, which can transmit information about social-related traits such as empathy of personality.

Connections Quantity The number of unique contacts from both calls and SMS, which can be used to measure the degrees in the Social network.

Connections Entropy We count the number of Connections time between the individuals and the unique contacts, and compute Shannons entropy to measure the contacts regularity.

Mobility features Mobility features focus on mobility patterns of the individuals in daily life, which can be inferred from the position of BTSs connected by the individuals.

Mobility Sphere The minimum radius which encompasses all the locations (BTS), and the distance between home and workplace of the individuals.

Mobility Quantity The record of Locations (BTSs) from both call and Data Volume services, including the average, the maximum, the minimum number, the variance of Locations in one day, and the number of days that have Locations, The number and the proportion of Locations during the night(19 pm to 7 am of next day). The number of locations where 80% of communications occurred.

Mobility Entropy We count the frequency for each location the individual stay on, and compute Shannons entropy to measure the locations regularity. Moreover, we compute the number of call and SMS for different locations (BTS), and use Shannons entropy to compute connections space entropy, which reflects the space regularity of connections.

Socioeconomic features Socioeconomic features are related to demographic information (age, gender), which required in specific P2P products. We get those features from the basic information of individuals.

Individual attribute features Individual attribute features refer to individuals' operation behaviors through an electronic device. In our data, the extracted features reflect individuals operation behaviors on mobile phones, which are mainly the App usage behaviors. These behaviors which have been proved can reflect differences on psychological level [10]. Specifically, payment, financial, and P2P online lending Apps usage features are extracted to compare the different operation preference on economic status related Apps of individuals.

App Frequency The number of installed Apps and the categories of Apps, statistics of usage frequency of Apps, including the total, the average, the variance, the maximum, the minimum usage frequency; the regularity on usage frequency.

App Data Volume Statistics of data volume spend on Apps, including the total, the average, the variance, the maximum, the minimum data volume spent; the regularity of data volume spent.

Specific App Usage Behavior The usage features on different categories of Apps, which consist of financial Apps, payment Apps, and the combination of financial and payment Apps. The feature set includes the number of installed Apps, the proportion of Apps, the number of Apps that belongs to the top5 frequently used Apps in different categories, and the number of Apps that belongs to the top5 frequently used Apps. Especially, for P2P online lending Apps, the total usage time on Apps, the total data volume spending on Apps, the regularity of usage frequency and the data volume are extracted.

4.2 Model Building

We select supervised learning to build our default behavior prediction model of P2P Online Lending. To this end, we represent individuals in the presented feature space, which we extract from the mobile phone usage data. Every presented

feature for an individual contains total 92 features. We select actual default behavior of 3027 subjects. After data pre-processing on aggregating to structural data, data cleaning i.e., 2999 subjects are included in the experiments and 28 subjects have been filtered due to missing data. To train and test the effect of our model, we randomly split the dataset into two parts, where 80% are used for training (2399 subjects) and 20% (600 subjects) are used for testing.

We try two different classification methods to compare their performance in this specific problem setting: Random Forests (RF) [2] and Light Gradient Boosting Machine (LightGBM) [24]. Random Forests algorithm is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest, which is widely used in classification problems. LightGBM is a highly efficient Gradient Boosted Decision Trees method proposed by Microsoft, which has faster training efficiency, low memory usage, higher accuracy, and support parallelization learning for processing large scale data.

Considering different methods have different advantages, we construct a joint decision model, which makes a default risk judgment through combining Random Forests with LightGBM. To build the proposed model, we train two independent submodels by using Random Forests algorithm and LightGBM algorithm separately. The final prediction result of the proposed model is determined by the average value of the two default possibilities, which are given by the two submodels. To give an example, if the default possibilities from the two submodels are 0.7 and 0.8, the ultimate default possibility judged by the proposed model is 0.75, which is the average value of 0.7 and 0.8. In order to tune the hyper-parameters automatically, we use grid-search strategy and fivefold cross validation over the entire training set for both of the two submodels. Finally, we get the optimal parameters of the Random Forests submodel and LightGBM submodel respectively, which make up the optimal parameters of the proposed model. According to the contrast result on the same testing phase, the proposed model has a better performance in the default behavior prediction for P2P Online Lending.

5 Experimental Results

In this section, we report the experimental results on real-world dataset as described in Section 3. Considering the unbalanced nature of the ground truth, we used the following four metrics to evaluate the prediction performance of default behavior, i.e., Precision, Recall, F1 score, AUCROC [4]. We use the AUCROC to measure the discriminatory ability. And the Precision, Recall, and F1 score are used to evaluate the correctness of the categorical predictions.

5.1 Feature Performance

In order to compare the performance of the features from mobile phone usage data as described in Section 4, we use three different features sets to build our

models, i.e., CSMS features set only, IA features set only, CSMS+IA features set. CSMS features set contains Consumption features, Social network features, Mobility features, and Socioeconomic features, which we extract from the daily CDR records data and basic information registered by the mobile carriers. IA features set contains Individual attribute features, which we extract from the special data of App usage. Using Random Forest and LightGBM methods, we build models on these three features sets respectively, and use AUCROC to measure the Classification performance. The compared results are depicted in Table 2. Obviously, the combination of CSMS+IA features has better performance in AUCROC on the two methods. Based on this conclusion, and we select CSMS+IA features set to build the default behavior prediction model.

Table 2. Classification performance (AUCROC) of different feature categories on different two methods.

Categories	Random Forests	LightGBM
CSMS features set	0.72	0.72
IA features set	0.69	0.69
CSMS+IA features set	0.76	0.77

5.2 Comparison of the methods

To accomplishing default behavior prediction, we adopt a joint decision model, which makes a default risk judgment through combining Random Forests with LightGBM as described in Section 4. We also use Random Forest method and LightGBM method individually to compare their performance with the proposed model in this specific problem setting. Three different models have been performed, and Figure 2 shows the performance of these models on four evaluation metrics. We found the proposed model achieving the best performance on Recall (0.885), F1 score (0.819), and AUCROC (0.774), which also has the better Precision (0.782), just 0.02 lower than LightGBM (0.784). According to the contrast result above, the proposed model has quantitative performance on the P2P default behavior Prediction.

5.3 Comparison against existing methods

The performance of the proposed method has also been compared with existing methods. In the state-of-art studies [14], random forest model has been trained on Lending Club dataset to assess the individual default risk. As depicted in Table 3, the proposed method has higher AUCROC (0.774), Recall (0.885) and Precision (0.782) than [14] with AUCROC of 0.71, Recall of 0.87 and Precision

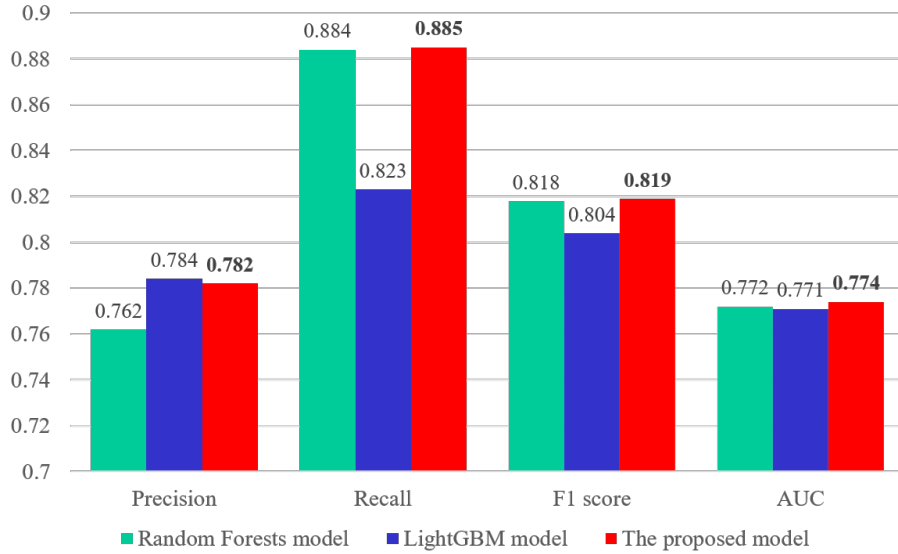


Fig. 2. A figure caption is always placed below the illustration. Please note that short captions are centered, while long ones are justified by the macro package automatically.

of 0.56, which depict the proposed method has better prediction performance. We also compare the performance with [21] following the same protocol on the division of test samples. They developed a logistic regression model to predict default also on data from Lending Club. As depicted in Table 3, our performance on Precision (0.782) are better than [21], which has Precision of 0.646. This shows that the proposed method is a more conservative model tending to reject more applicants to protect the P2P platforms from possible financial loss. These results demonstrate the feasibility of adopting the proposed method for P2P lending platforms. Moreover, we compare the performance with [18], where they build

Table 3. The performance comparison between our method and the existing methods

methods	AUCROC	Precision	Recall
[14]	0.71	0.56	0.87
[21]	-	0.646	-
[18]	0.725	0.29	-
Our method	0.774	0.782	0.885

a Gradient Boosted Trees (GBT) classifier model to assess the users financial

risk on credit card data, collected by a financial institution operating in the considered Latin American country. As depicted in Table 3, our proposed method has higher AUCROC (0.774) and Precision(0.782) than [18] with AUCROC of 0.725 and Precision of 0.29, These results demonstrate that the proposed method may have a better performance not only on P2P lending platforms but also on other financial risk platforms.

6 Conclusion

In this paper, we propose a risk control mechanism for P2P online lending platforms, which has a potential to be employed in countries lack of reliable personal credit evaluation system. We further propose a default behavior prediction model, which provides pre-approval credit estimate using mobile phone usage data in this mechanism. We extract features from five aspects, including consumption, social network, mobility, socioeconomic, and individual attribute. Specifically, we adopt a joint decision model, which makes a default behavior judgment through combining Random Forests with Light Gradient Boosting Machine. Lastly, we validate the proposed model using real-world dataset.

The experimental results demonstrate that the features combining all five aspects are most predictive for the future default behaviors of borrowers. Compared with other classifiers, the proposed model has achieved the best performance in terms of evaluation metrics. Moreover, the proposed model shows better performance when comparing to the existing methods in this problem setting.

In the future, we plan to measure the distinguishing power of the different features of our model in detail. Furthermore, we are interested in assessing how our risk control mechanism changes as a function of the P2P online lending products analyzed.

References

1. Boase, J., Ling, R.: Measuring mobile phone use: Self-report versus log data. *Journal of Computer-Mediated Communication* **18**(4), 508–519 (2013)
2. Breiman, L.: Random forests. *Machine learning* **45**(1), 5–32 (2001)
3. Chen, X., Zhou, L., Wan, D.: Group social capital and lending outcomes in the financial credit market : An empirical study of online peer-to-peer lending. *Electronic Commerce Research Applications* **15**(C), 1–13 (2016)
4. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: *ICML '06 : Proceedings of the International Conference on Machine Learning*, New York, Ny, Usa. pp. 233–240 (2006)
5. Emekter, R., Tu, Y., Jirasakuldech, B., Lu, M.: Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending. *Applied Economics* **47**(1), 54–70 (2015)
6. Gathergood, J.: Self-control, financial literacy and consumer over-indebtedness. *Social Science Electronic Publishing* **33**(3), 590–602 (2012)
7. Gonzalez, L., Loureiro, Y.K.: When can a photo increase credit? the impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance* **2**, 44–58 (2014)

8. Guo, Y., Zhou, W., Luo, C., Liu, C., Xiong, H.: Instance-based credit risk assessment for investment decisions in p2p lending. *European Journal of Operational Research* **249**(2), 417–426 (2015)
9. Harari, G.M., Lane, N.D., Wang, R., Crosier, B.S., Campbell, A.T., Gosling, S.D.: Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science* **11**(6), 838–854 (2016)
10. Harari, G.M., Müller, S.R., Aung, M.S., Rentfrow, P.J.: Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences* **18**, 83–90 (2017)
11. JiaZhuo, W., Hongwei, X.: China’s online lending industry in 2015. Tsinghua University Press (2015)
12. Lin, M., Prabhala, N.R., Viswanathan, S.: Judging borrowers by the company they keep: Social networks and adverse selection in online peer-to-peer lending. *Ssrn Elibrary* (2009)
13. Liu, S., Qu, Q., Wang, S.: Rationality analytics from trajectories. *ACM Transactions on Knowledge Discovery from Data (TKDD)* **10**(1), 10 (2015)
14. Malekipirbazari, M., Aksakalli, V.: Risk assessment in social lending via random forests. *Expert Systems with Applications* **42**(10), 4621–4631 (2015)
15. Montjoye, Y.A.D., Quoidbach, J., Robic, F., Pentland, A.: Predicting personality using novel mobile phone-based metrics (2013)
16. Oliveira, R.D., Karatzoglou, A., Cerezo, P.C., Oliver, N.: Towards a psychographic user model from mobile phone usage. In: *CHI 11 Extended Abstracts on Human Factors in Computing Systems*. pp. 2191–2196 (2011)
17. Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M.L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., et al.: Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR)* **45**(4), 42 (2013)
18. Pedro, J.S., Proserpio, D., Oliver, N.: Mobiscore: Towards universal credit scoring from mobile phone data **9146**, 195–207 (2015)
19. Polena, M., Regner, T., et al.: Determinants of borrowers default in p2p lending under consideration of the loan risk class. *Jena Economic Research Papers* **2016**, 023 (2016)
20. Renso, C., Baglioni, M., de Macedo, J.A.F., Trasarti, R., Wachowicz, M.: How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowledge and information systems* pp. 1–32 (2013)
21. Serrano-Cinca, C., Gutierrez-Nieto, B., López-Palacios, L.: Determinants of default in p2p lending. *PloS one* **10**(10), e0139427 (2015)
22. Serrano-Cinca, C., Gutierrez-Nieto, B.: The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (p2p) lending. *Decision Support Systems* **89**(C), 113–122 (2016)
23. Soto, V., Friasmartinez, V., Virseda, J., Friasmartinez, E.: Prediction of socioeconomic levels using cell phone records. *Lecture Notes in Computer Science* **6787**(2), 377–388 (2011)
24. Wang, D., Zhang, Y., Zhao, Y.: Lightgbm: An effective mirna classification method in breast cancer patients. In: *International Conference*. pp. 7–11 (2017)
25. Wang, M., Zheng, X., Zhu, M., Hu, Z.: P2p lending platforms bankruptcy prediction using fuzzy svm with region information. In: *e-Business Engineering (ICEBE), 2016 IEEE 13th International Conference on*. pp. 115–122. IEEE (2016)

26. Wang, S., Qi, Y., Fu, B., Liu, H.: Credit risk evaluation based on text analysis. *International Journal of Cognitive Informatics and Natural Intelligence* **10**(1), 1–11 (2016)
27. Wang, X., Zhang, D., Zeng, X., Wu, X.: A bayesian investment model for online p2p lending. In: *Frontiers in Internet Technologies*, pp. 21–30. Springer (2013)
28. Wu, S., Kang, N., Yang, L.: Fraudulent behavior forecast in telecom industry based on data mining technology. *Communications of the IIMA* **7**(4), 1 (2014)
29. Zhang, Y., Jia, H., Diao, Y., Hai, M., Li, H.: Research on credit scoring by fusing social media information in online peer-to-peer lending. *Procedia Computer Science* **91**, 168–174 (2016)
30. Zhao, H., Ge, Y., Liu, Q., Wang, G., Chen, E., Zhang, H.: P2p lending survey: Platforms, recent advances and prospects. *ACM Transactions on Intelligent Systems and Technology (TIST)* **8**(6), 72 (2017)