Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 91 (2016) 168 – 174

Procedia Computer Science

Information Technology and Quantitative Management (ITQM 2016)

Research on Credit Scoring by fusing social media information in Online Peer-to-Peer Lending

Yuejin Zhang[a,] *, Hengyue Jia[a], Yunfei Diao[a], Mo Hai[a], Haifeng Li[a]

[a]School of Information, Central University of Finance and Economics, Beijing 100081, China

Abstract

In recent years, online Peer-to-Peer (P2P) lending market is rapidly expanding in China. In this paper, we use public dataset from PPDai, a leading online P2P platform in China to study the loan default. We construct a credit scoring model by fusing social media information based on decision tree. The experimental result shows that our model has good classification accuracy. From the credit scoring model and classification rules, we get a conclusion that the loan information, social media information, and credit information are most important factors for predicting the default. However, the credit rating is not as important as the platform described.

Peer-review under responsibility of the Organizing Committee of ITQM 2016

Keywords: credit scoring; default risk; decision tree; Peer-to-Peer lending; social media information

1. Introduction

Online peer-to-peer (P2P) lending is a new financing channel which is based on electronic business platform and electronic commerce credit. In P2P lending, borrowers and lenders can use the internet platform to achieve online transactions. There is lower transaction cost, while the loan process is simple and easy to operate. Small and micro enterprises and individual borrowers that are difficult to get loans from the bank do not need loan guarantor and collateral in P2P, so they can get financing more easily. But it means higher credit risk to lenders. Credit risk is the possibility of loss that the bank will suffer after offering loan to the borrowers. It includes not only the actual risk of the borrowers failing to repay the loan on time, but also the potential default risk because of the downgrade of credit or decline of repayment ability of the borrowers. The main reason of credit risk is the asymmetric information between borrowers and lenders [1]. The main method to avoid credit risk in

* Corresponding author. Tel.: 8610-62288896; fax: 8610-62288896.
E-mail address: zhangyj@cufe.edu.cn.

Peer-review under responsibility of the Organizing Committee of ITQM 2016
doi:10.1016/j.procs.2016.07.055

financial institutions is to evaluate the risk by credit scoring [2]. In particular, personal credit scoring is to analyze large amount of data including personal information, credit history and credit behavior to find the characteristics of default and normal borrowers by using data mining technology and statistical analysis method, so as to forecast the credit risk. On the P2P platforms, investors would lend money to a stranger whose personal information may be incomplete, so there is serious information gap.

In this paper we study the online market for P2P in China. In order to reduce the serious problem of information asymmetry between both sides of P2P loans, we make use of social information to describe the behavior characteristics of the borrowers. A person's social behavior and language can reflect the characteristics of their behavior, which can be used as credit data. On the internet, the behavior and language of users can be obtained from social media. Social media refers to website and technology that allow people to write, share, evaluate, discuss, communicate with each other, including social networking sites, micro blog, twitter, micro letters, blogs, forums, and so on. Now, social media have covered almost all aspects of our life, through which we can know the latest developments of a person. Now some credit companies such as Lenddo have already use information of users from Facebook, LinkedIn and Twitter to evaluate credit risk of the consumers. Social data is most useful for people with little or no credit history. A person's social identity, online reputation and professional contacts circle, that should become a factor in the assessment of credit risk.

The main contribution of this paper is to add social behavior factors into the traditional credit scoring model, which are not considered in prior research of credit scoring models. We make an empirical study on PPDai. As a leading platform in China, PPDai has more than 6,000,000 members and 2 billion RMB transaction amount in loans in 2014. We use the public data of borrowers on the platform to create the credit scoring model including the personal information, loan information and social network data. The experimental results show that the credit scoring model based on decision tree can well distinguish between the default and the normal customers, and it has higher forecast accuracy comparing with the model based on logistic regression and neural network.

2. Related Studies

There is little prior research on the credit risk in P2P online lending. The existing literature mainly includes two aspects. The first aspect of research examines the factors in traditional credit scoring models that explain the funding success and default risk. The second aspect of research focuses on the relationship between social media information and default risk.

The main factors in the traditional credit scoring model include the basic information, repayment ability, life stability, credit record, guarantee and some other factors. Herzenstein et al. [3] find that the borrowers' financial strength and their effort when listing and publicizing the loan are more important than demographic attributes for funding success. Qiu et al. [4] show that the personal information, social capital, loan amount, acceptable maximum interest rate and loan period set by borrowers are all significant factors of funding success. Riza et al. [5] use Cox Proportional Hazard regression technique to evaluate credit risk and measure loan performances. They find that credit grade, debt-to-income ratio, FICO score and revolving line utilization play an important role in loan defaults.

Everett [6] analyzes the relationship between the social relationship and the default risk and the interest rate in online P2P lending, and concludes that there is a low default rate between the group members who have actual social relationship on the mutual financing platform. Collier et al. [7] make an empirical analysis of the financing behavior between members of community groups on P2P lending platform, and confirm that by combining individual reputation with the reputation of the group, the group members can supervise each other, which can effectively reduce the adverse selection phenomenon and control moral hazard. Using dataset from Prosper.com, Lin et al. [8] find that the social connections information can effectively solve the problem of asymmetric information in online P2P lending. Their research results show that the borrowers' friendships could increase the probability of successful funding, lower interest rates on funded loans, and these borrowers