



Information Technology and Quantitative Management (ITQM 2016)

Detecting the abnormal lenders from P2P lending data

Haifeng Li^{a,*}, Yuejin Zhang^a, Ning Zhang^a, Hengyue Jia^a

^aSchool of Information, Central University of Finance and Economics, Beijing, China

Abstract

Online peer-to-peer lending is a new but useful finance method for small enterprises that is conducted on the website. To exclude the risk of this method, we make a study on predicting the potential lenders that may have a bad credit score. We use an outlier detection method to find the abnormal lenders, and we find the detected outliers have bad credit scores with a high possibility.

© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

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

Keywords: trust model; credit score; classification; P2P

1. Introduction

Online peer-to-peer lending is a new but useful finance method for small enterprises. To finance small and micro enterprises in an effective method has attracted many attentions. This problem is very important especially in China. By the advances in information technologies, a new type of financing method, online peer-to-peer (P2P) lending has become an important issue for traditional financing. Online P2P lending allows people to lend and borrow funds directly through an online intermediary without the mediation of financial institutes.

1.1. Motivation

When a lender wants to acquire capitals from the online P2P companies, a risk will be raised. Traditional bank can audit the background of a lender with his application document, which, for the P2P companies or the borrowers, is an impossible task. Since a lender is never known has a good credit score or a bad one. Thus, how to find the lenders with bad credit score is a very challengeable question. Many researches have focused on this problem and proposed some useful method.

* Corresponding author. Tel.: +8613691380799
E-mail address: mydlhf@cufe.edu.cn.

1.2. Related Works

[1] represented an extension of the expansive credit risk and credit migration literature, prominent in the corporate bond and securities risk pricing literature, to an analysis of the drift of consumer credit scores. A rich data set of residential mortgages was used to observe credit score migration post loan origination and in a test of the ability of credit score transition to serve as a precursor to potential default and prepayment. The results indicated credit scores provide signals and information to investors and servicing agents in a fashion similar to credit ratings on commercial paper as to default potential. Soner[2] presented a proposed a three stage hybrid Adaptive Neuro Fuzzy Inference System credit scoring model, which was based on statistical techniques and Neuro Fuzzy. The performance of the proposed model was compared with conventional and commonly utilized models. The credit scoring models were tested using a 10-fold cross-validation process with the credit card data of an international bank operating in Turkey. Results demonstrated that the proposed model consistently performed better than the Linear Discriminant Analysis, Logistic Regression Analysis, and Artificial Neural Network (ANN) approaches, in terms of average correct classification rate and estimated misclassification cost. [3] addressed the question of what determines a poor credit score. The authors compared estimated credit scores with measures of impulsivity, time preference, risk attitude, and trustworthiness, in an effort to determine the preferences that underlie credit behavior. Data was collected using an incentivized decision-making lab experiment, together with financial and psychological surveys. Credit scores were estimated using an online FICO creditscore estimator based on survey data supplied by the participants. Preferences were assessed using a survey measure of impulsivity, with experimental measures of time and risk preferences, as well as trustworthiness. Controlling for income differences, the authors found that the credit score was correlated with measures of impulsivity, time preference, and trustworthiness. Based on trust theories, Chen et. al[4] the present study develops an integrated trust model specifically for the online P2P lending context, to better understand the critical factors that drive lenders' trust. The model is empirically tested using surveyed data from 785 online lenders of PaiPaiDai, the first and largest online P2P platform in China. The results show that both trust in borrowers and trust in intermediaries are significant factors influencing lenders' lending intention. Emerkter et. al[5] used data from the Lending Club, which is one of the popular online P2P lending houses, to explore the P2P loan characteristics, evaluate the credit risk and measures loan performances. They found that credit grade, debt-to-income ratio, FICO score and revolving line utilization played an important role in loan defaults. Loans with lower credit grade and longer duration were associated with high mortality rate. The result was consistent with the Cox Proportional Hazard test. Also, they found that higher interest rates charged on the high risk borrowers were not enough to compensate for higher probability of the loan default; thus, the Lending Club must find ways to attract high FICO score and high-income borrowers in order to sustain their businesses. Harris[6] investigated the practice of credit scoring and introduced the use of the clustered support vector machine (CSVM) for credit scorecard development. This algorithm was well known that as historical credit scoring datasets get large while highly accurate became computationally expensive. Accordingly, he compared the CSVM with other nonlinear SVM based techniques and shows that the CSVM can achieve comparable levels of classification performance while remaining relatively cheap computationally.

In this paper, we also addressed this problem and proposed a outlier detection method by the online documents of the lenders. This method can detect the abnormal lenders by their general features. The rest paper is organized as follows: Section 2 presents the data related the lenders. Section 3 introduces our detecting method. Section 4 concludes this paper.

2. Dataset Preparation and Data Processing

We use the data crawled from the website, which is a BBS that provide the lenders to discuss the issues related to P2P lending. We preprocess the data and get the dataset with 18 properties. We describe it with Table

Download English Version:

<https://daneshyari.com/en/article/488344>

Download Persian Version:

<https://daneshyari.com/article/488344>

[Daneshyari.com](https://daneshyari.com)