

文章编号: 1005-9679(2017)03-0111-06

P2P 平台出借人投资决策研究

王辛未, 姜睿思

(上海交通大学 安泰经济与管理学院, 上海 200030)

摘要: 随着 P2P 网贷平台的迅速发展, 一个关键的问题是投资者如何评估借款人的信用风险并有效投资。本文从网贷平台的真实贷款数据出发, 分别用逻辑回归和随机森林预测贷款的违约概率, 并通过距离度量模型与核权重相结合评估出新贷款需求的收益和风险。由于 P2P 网贷平台对放、贷款金额的限制, 本文同时解决了一个有约束的投资组合问题。数据结果显示, 与传统的信用分级相比, 本文的方法提高了贷款收益率预测的准确度, 选择出的投资组合收益率更高。

关键词: P2P 网贷; 随机森林; 核权重; 投资组合

中图分类号: F 224 **文献标志码:** A

Lenders' Investment Decision-making Research on P2P platform

WANG Xinwei, JIANG Ruisi

(Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai 200030, China)

Abstract: Along with the rapid development of P2P platform, one of the key question is how investors assess the borrowers' credit risk and get effective investment. This paper use the real loan data from a P2P platform, respectively, using logistic regression and random forest to predict the default probability of the loan, then combine the distance measurement model and the kernel regression to evaluate the return and risk of new demand for loans. Due to the amount limit to both the investors and the lenders on P2P platform, this paper also solved a constrained portfolio problem. Data shows that compared with traditional credit rating, the method in this paper improves the loans' yield prediction accuracy and can choose portfolios with higher return.

Key words: P2P platform; random forest; kernel regression; portfolio selection

本文对 P2P 平台上每一笔新增贷款需求进行了一个更详细的信用评估, 通过历史贷款数据核加权的方法评估出每一笔新贷款所能带来的风险和收益, 有助于个人投资者能基于此评估结果分配自己的投资, 实现更好的投资决策。

1 国内外文献综述

P2P 作为新兴的金融事务, 吸引了众多研究者

的关注。目前, 国内外的研究主要集中在网贷平台的运营模式、平台监管、借款人违约风险评估、投资者出借意愿等领域。

Hulme 等^[1]根据研究英国的 P2P 平台 zopa 得出 P2P 的出现是应对信息时代金融市场需求诞生的新型借贷者关系。在研究信息不对称问题时, 刘峙廷^[2]指出, 对于网贷平台的投资者来说, 在判断借款人的可信度时, 他们希望知晓借款人更多的“软

收稿日期: 2016-11-21

基金项目: 国家自然科学基金资助项目(71531010)

作者简介: 王辛未(1992-), 女, 硕士研究生, 研究方向: 互联网金融, 数据挖掘。Email: lamb_wxw@sina.com。

姜睿思, 男, 博士研究生, 研究方向: 运营管理、大数据与商务智能。

信息”和“硬信息”，且两者处于同等重要的地位。软信息指的是不能通过准确且具体的数值或者指标来表示的信息，与此相对应的，硬信息一般指的是可以准确具体表示的信息。Freedman 等^[3]利用 Prosper 平台提供的数据，研究了在 P2P 网贷中的信息不对称问题。由于在 Prosper 上提出贷款申请的很有可能是信用较低、在线下银行难以获得贷款的借款人，而投资者很多时候难以获得完整的资料去判断，这种信息不对称很容易就产生逆向选择和道德风险问题。他们提出可以利用借款人在社交网络中的个人信息，进而在某种程度上帮助贷款者去辨别借款者的违约风险。Dongyu 等^[4]从社会资源和借款人性格的角度去研究影响投资者对借款者信任的因素，数据显示，社交网络资源越丰富的借款人更容易得到投资者的信任因而更获得贷款。与此同时，借款人的性格与能否获得贷款也有较为明显的联系。Lin^[5]研究社会关系在借款者违约风险上的影响时发现，拥有稳定的朋友关系的借款人，尤其是拥有那些信用认证更高的优质朋友的借款人的违约概率能大幅降低，作者认为利用软信息去研究借款者信用是一个很有效的方向。Lee 等^[6]研究网贷平台上投资者行为，由于网贷平台上的投资者大多数并不是专业投资者，对借款者的信用状况缺乏理性的评判，故他们很容易将资金投放在已经拥有较多其他投资者投资的借款项目上，P2P 平台存在着“羊群行为”问题。雒春雨^[7]根据 P2P 网络借贷的一些特点，将投资者行为信息和借贷者的信用信息相结合进行挖掘，分别从投资者的构成、借贷者信用风险及多信息源 3 个角度进行综合的分析考量，并量化贷款评估及投资决策模型，帮助投资者进行投资决策。

近年来，数据挖掘人工智能方法在信用评估方面应用广泛。HaSH^[8]构建了一个由 Kohonen 神经网络模型和 Cox 比例风险回归模型组合的混合模型，Kohonen 神经网络模型用来聚类，将信贷违约者分为几个同质组，Cox 比例风险回归模型用来分析每组违约者的还款模式，综合来评估还款违约可能发生的阶段。研究结果，表明该模型能够取得较好的预测结果。对于银行来说，不良贷款数目过多严重影响银行的运营状况，Mandala 等^[9]对某家农村银行进行信用风险分析，以减少不良贷款数目为目标，通过利用决策树模型的 C5.0 算法对贷款进行信用评估，结果显示该评估方法能有效对不良贷款进行预警，可使不良贷款率从之前的 11.99% 下降到低于 5%。熊志斌^[10]借鉴传统 PCA 方法中的序数主成分概念，提出了基于顺序自联想神经网络

的非线性主成分分析法，结果显示该方法能提取更少的主成分表示出原始数据中更多的信息，并分别结合神经网络和逻辑回归模型，通过实证研究表明在相同的 PCA 方法处理数据下，结合神经网络模型的预测效果比逻辑回归要好。杨胜刚等^[11]研究德国某商业银行的真实客户信用数据，将决策树应用于变量的筛选过程，并将其与 BP 神经网络相结合构建了一个两阶段的组合模型。研究表明，和单一的 BP 神经网络模型相比，基于决策树和神经网络构建的个人信用评估组合模型能够提高分类预测的精度。姚潇^[12]利用两个公开的信用数据集实证研究发现，与其他的模型相比，所提出的模糊近似支持向量机能够显著地提高信用风险分类精度，具有较高的实用价值。Harris^[13]运用集群支持向量机(CSVM)进行信用风险评估，并且和传统的非线性支持向量机(SVM)比较得出集群支持向量机(CSVM)能够实现更好的分类效果。

2 模型建立

2.1 违约概率估计

借款人未来还款情况有两种，履约或者违约，我们可以通过借款人的一些信息变量比如贷款金额、贷款原因以及贷款人信用卡情况等个人信息进行分类预测，判断出借款人的违约概率。机器学习常用的进行分类预测的方法有逻辑回归、贝叶斯法、神经网络法、支持向量机法、随机森林法等，根据需求本文选择逻辑回归和随机森林这两种方法进行违约概率预测。

随机森林分为分类和回归两种技术，本文研究的是贷款是否会发生违约，属于分类预测，所以这里我们主要介绍随机森林的分类。Breiman^[14]把分类树组合成随机森林，即通过随机抽取变量(列)和样本数据(行)，生成很多分类树，再汇总分类树的结果，即成为随机森林算法。所谓的随机森林分类方法就是由许多决策树分类模型 $\{h(X, \theta_k), k=1, \dots\}$ 组成的组合分类模型，其中 $\{\theta_k\}$ 是相互独立同分布的随机变量，在给定的自变量 X 下，每个决策树分类模型均有一票投票权一起决策最优的分类结果。

随机森林进行 bootstrap 抽样时，从原始的训练集抽取 k 个样本，分别对这 k 个样本建立决策树模型，会得到 k 种分类结果，最终分类结果由这 k 种分类结果的投票表决结果决定。假设原始数据集 S 中有 N 个样本，那么每个样本未被抽中的概率为 $(1-1/N)^N$ 。当 N 足够大时， $(1-1/N)^N$ 将收敛于

$\frac{1}{e}$ 。随机森林之所以能够提高组合分类模型的外推预测能力,是因为它构造了不同的训练集来增加分类模型间的差异。通过 k 轮训练,得到一个分类模型 $\{h_1(x), h_2(x), \dots, h_k(x)\}$,再用他们构成一个多分类模型系统,然后使用多数投票确定最终的分类结果,所用公式为

$$H(x) = \arg \max_Y \sum_{i=1}^k I(h_i(x) = Y)$$

其中: $H(x)$ 表示组合分类模型结果; $h_i(x)$ 是单个决策树分类模型结果; Y 为因变量; $I(\cdot)$ 为示性函数。

逻辑回归是对定性变量的回归分析,是二分类因变量常用的统计方法。用逻辑回归根据贷款者的相关信息预测贷款在违约情况时,借款人违约可以表示为 1,借款人没有违约表示为 0。假设借款人 i 的相关信息即自变量为 n 个,即 $x_{i1}, x_{i2}, \dots, x_{in}$,由借款人的相关变量评估该贷款者违约概率为 p_i ,违约发生的概率与影响因素之间的关系可以表示为

$$p_i = 1/(1 + e^{-d_i})$$

$$d_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_n x_{in}$$

b_0, b_1, \dots, b_n 是回归系数, d_i 表示变量叠加后的线性权重之和, d_i 值越高,违约概率越高。

2.2 距离度量模型

根据历史数据去预测未来在很多领域应用很广泛,但是在 P2P 平台上,对于每一个借款人来说,记录下来相关历史数据条目很少,因而对于同一个借款人很难通过其过去表现来评估该借款人未来的还款表现。本文提出利用与某一给定借款人具有相似信息的其他借款人的贷款表现去评估该借款人的方法,根据违约概率的绝对距离去判断借款人相似度。

对于一笔给定的贷款 i ,可通过与其相似的 n 笔历史贷款真实收益(收益分别为 $R_j, j=1, 2, \dots, n$)加权平均预测投资贷款 i 可实现收益 u_i ,通过分类预测方法判断出贷款 i 和贷款 j 的违约概率分别为 p_i, p_j ,贷款 i 和贷款 j 的相似度量距离 $d_{ij} = |p_i - p_j|$ (1), d_{ij} 越小表示贷款 i 和贷款 j 的相似度越接近。用 w_{ij} 表示用于预测贷款 i 收益的贷款 j 的权重,对于和贷款 i 距离更近的贷款 j ,在评估贷款 i 的时候贷款 j 被赋予更高的权重,即相似距离 d_{ij} 越小,权重 w_{ij} 更大。反之,相似距离 d_{ij} 越大,权重 w_{ij} 越小。利用权重加权方法可以得到投资贷款 i 可实现的收益率为

$$u_i = \sum_{j=1}^n w_{ij} R_j \quad (2)$$

由贷款收益进而可得出贷款 i 的风险为:

$$\sigma_i^2 = \sum_{j=1}^n w_{ij} (R_j - u_i)^2 \quad (3)$$

2.3 核权重的确定

为了通过距离度量得到权重系数,本文引入了核回归的方法。非参数回归方法的一般模型为 $Y_i = m(x_i) + \epsilon_i$ (4),通过已知估计出回归函数,其中 Y_i 为从观测点 x_i 得到的观测值, $m(x_i)$ 是未知回归函数, ϵ_i 是独立同分布的零均值噪声。常见的非参数回归方法一般有局部和正交回归,局部回归又可以分为核回归、局部多项式回归、近邻回归和稳健回归。核回归中有 N-W 估计, P-C 估计, G-M 估计。

本文采用的是核回归(N-W 估计),N-W 估计为一种简单的加权平均估计。假设每个观测值通过两个维度进行评估,预测变量 x 和响应变量 y ,根据 n 个历史观测值 $\{(x_j, y_j) | j=1, 2, \dots, n\}$,可以预测给定某一预测变量 x 下响应变量 y 的值,即

$$y = \tilde{f}(x) = \frac{\sum_{j=1}^n K\left(\frac{x-x_j}{h}\right) y_j}{\sum_{j=1}^n K\left(\frac{x-x_j}{h}\right)} \quad (5)$$

$K(\cdot)$ 是高斯核函数, $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$ 。

对于贷款评估,需要评估的贷款 i 违约概率 p_i 即为预测变量,对应的收益率 u_i 是响应变量,根据 n 个历史贷款观测值 $\{(p_j, R_j), j=1, 2, \dots, n\}$ 预测贷款 i 可能带来的收益为

$$u_i = \frac{\sum_{j=1}^n K\left(\frac{x_i - x_j}{h}\right) y_j}{\sum_{j=1}^n K\left(\frac{x_i - x_j}{h}\right)} = \frac{\sum_{j=1}^n K\left(\frac{p_i - p_j}{h}\right) R_j}{\sum_{j=1}^n K\left(\frac{p_i - p_j}{h}\right)} \quad (6)$$

通过这样的方法,和贷款 i 距离近的历史贷款观测值被赋予更高的权重,距离远的赋予更低的权重:

$$w_{ij} = \frac{K\left(\frac{p_i - p_j}{h}\right)}{\sum_{j=1}^n K\left(\frac{p_i - p_j}{h}\right)} = \frac{K\left(\frac{d_{ij}}{h}\right)}{\sum_{j=1}^n K\left(\frac{d_{ij}}{h}\right)} \quad (7)$$

参数 $h(h>0)$ 被称为带宽,为全局光滑参数,最优的参数 h 可以通过训练集得到,它的选取要满足非参数估计法中的缺一交叉验证方法,通过求出下式最小值找出对应的 h 。

$$CV(h) = \frac{1}{n} \sum_{i=1}^n (\tilde{f}_h(x_{-i}) - y_i)^2 \quad (8)$$

其中,

$$\tilde{f}_h(x_{-i}) = \frac{\sum_{j=1, j \neq i}^n K\left(\frac{x_i - x_j}{h}\right) y_j}{\sum_{j=1, j \neq i}^n K\left(\frac{x_i - x_j}{h}\right)} \quad (9)$$

在基于违约概率间距离的收益评估模型中,可将式子重新表示为

$$CV(h) = \frac{1}{n} \sum_{j=1}^n \left[\frac{\sum_{k=1, k \neq j}^n K \left(\frac{p_j - p_k}{h} \right) R_k}{\sum_{k=1, k \neq j}^n K \left(\frac{p_j - p_k}{h} \right)} - R_j \right]^2 \quad (10)$$

即找到使该式达到最小值所对应的 h 便是最优的 h 。为了更有效地找到最优值 h , 根据 Silverman^[15] 里提到的方法, 将最优 h 的搜索范围缩小到 $0.25h_0$ 和 $1.5h_0$ 之间, h_0 值为 $h_0 = \left(\frac{4}{3n} \right)^{\frac{1}{5}} \sigma$, 在该区间内找到 $CV(h)$ 的最小值, 对应的 h 即为目标 h , σ 是训练集中所有贷款违约概率的标准差。

2.4 投资组合选择

P2P 平台上的投资者和传统银行的投资者不同, 由于 P2P 平台投资者的资金相对有限, 因而投资者不仅需要分辨出不良贷款, 还要合理有效地分配资金 M , 实现期望收益下的最小风险。因此, 可以将 P2P 平台的资金分配问题引申为一个典型的资产组合选择问题。Markowitz 的经典投资组合理论认为理性的投资者构建的投资组合应该在有效前沿上, 其目标是在给定风险时最大化期望收益, 或者在给定收益时最小化风险。P2P 平台上的投资者选择的最优组合在满足给定收益即 $R^* = \sum_i \lambda_i \mu_i$ 同时能够实现风险最小也即使得 $\sum_i \lambda_i^2 \sigma_i^2$ 最小。另外, 投资金额不可能为负数, 所以也要满足投资组合中每种投资比例不小于 0 即 $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$, λ_i 是代表最优投资组合中给第 i 笔贷款需求分配的资金比例。除此之外, P2P 平台上一一般都会限制每笔投资的最低额度 m , 同时结合实际情况可知投资者借给某一借款人 i 的资金也会不超过借款人的资金需求 c_i , 补充上述约束条件, 该投资组合选择问题即为

$$\begin{aligned} \min & \sum_{i=1}^s \lambda_i^2 \sigma_i^2 \\ \text{st.} & \sum_{i=1}^s \lambda_i \mu_i = R^*, \quad \sum_{i=1}^s \lambda_i = 1 \\ & \begin{cases} m \leq \lambda_i M \leq c_i & \text{贷款 } i \text{ 被投资} \\ \lambda_i = 0 & \text{贷款 } i \text{ 没有被投资} \end{cases} \end{aligned}$$

3 实证分析

本文的研究选取了美国网贷平台 lending club 公司 2015 年 1 月 1 日到 2015 年 12 月 1 日的借款人公开数据, 从中选择出还款状况为“Fully Paid”“Charged Off”“Default”的贷款, 即“履约”“违约”状

况共计 33 283 条。为了满足实验的需求, 对上述数据进行了预处理, 主要经过数据抽取、缺失值处理、属性构造等过程。选出如表 1 所示共计 15 类属性用于贷款者违约概率的评估。实验运算通过 R 软件和 Matlab 进行。

表 1 预测收益精确度比较

解释变量	含义	
grade	贷款者信用评级	
home_ownership	贷款者房屋所有权	
loan_amnt	贷款金额	
annual_inc	贷款者年收入	
dti	贷款者债务收入比	
tot_cur_bal	所有账户的总余额	
total_rev_hi_lim	所有账户的总信用额	
acc_open_past_24mths	在过去 24 个月内的交易数目	
avg_cur_bal	所有账户的平均余额	
bc_util	所有账户的总余额与所有账户的总信用额度的比率	
mo_sin_old_rev_tl_op	第一张有记录信用卡开户距今时长	
mort_acc	贷款者拥有的贷款账户数量	
num_il_tl	分期付款账户有赊销的账户数	
num_rev_accts	信用卡数量	
num_rev_tl_bal_gt_0	在使用中的信用卡余额大于 0 的数量	
预测变量	loan_status	贷款者履约违约状况
收益率	loan_amnt	贷款金额
计算	total_pymnt	总还款额

将数据以 7 : 3 的比例分为训练集和测试集, 基于训练集数据得出逻辑回归和随机森林的模型参数, 进而预测出训练集和测试集中各贷款的违约概率。由于本文侧重点在于距离度量模型和核回归部分, 则在逻辑回归和随机森林预测违约概率的具体步骤和结果就不一一展示了。通过距离度量模型, 计算出测试集中各贷款和训练集中贷款的相似度, 通过求核权重的方法, 从训练集中用缺一交叉验证方法得到最优带宽即找 (10) 式最小值, $CV(h)$ 和 h 的关系如图 1 所示, R 软件运算结果显示最优带宽 $h = 0.35h_0$ 。

由于训练集中的贷款收益率都是已知的, 通过式 (6)、(7) 对训练集中的贷款通过核权重加权得出测试集中每笔贷款的收益, 进而由式 (3) 得到方差。分别用随机森林、逻辑回归和信用分级 3 种方法预

测贷款的收益值以及方差值,对于传统信用分级方式,处于同一个信用级别的贷款被认为具有相同的风险和收益值,收益与贷款违约概率的关系如图 2 所示,方差与违约概率的关系如图 3 所示。由于本文研究的理性投资者一般不会选择违约概率很高、极易发生违约的贷款,故图 2、3 中展示的关系选择出的贷款违约概率以 0.4 为分界。图中显示随着贷款违约概率的增长,贷款的风险增大,收益减少,从曲线可以看出,同评级方法相比,逻辑回归和随机森林预测出的收益和方差更加平滑,不同贷款之间有区分度,方便更精确地判断投资组合。

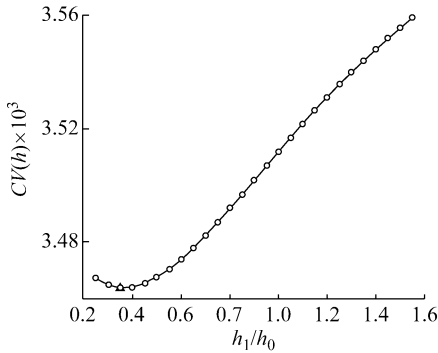


图 1 核权重最优参数 h 选择

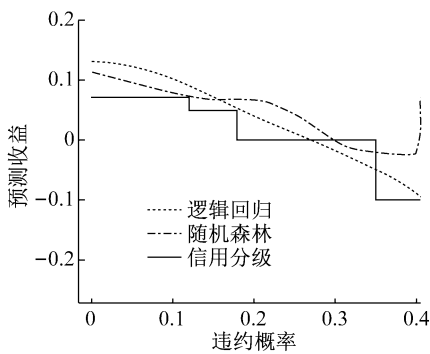


图 2 预测收益和违约概率的关系

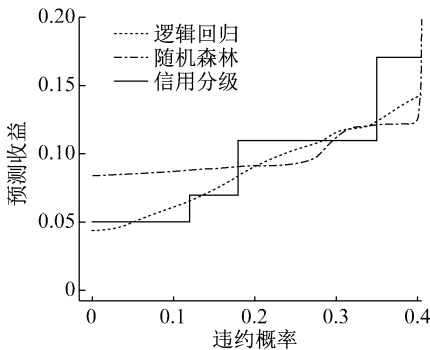


图 3 预测风险和违约概率的关系

由于有测试集贷款的真实收益值,故可以将预测收益和真实收益数据进行比较算出各模型预测收益的欧氏距离,如表 2 所示。结果显示,与传统的信用分级方法相比,随机森林和逻辑回归在预测准确

度上有所提高,且随机森林预测效果更好。

表 2 预测收益精确度比较

名称	欧氏距离
信用分级	14.898 88
逻辑回归	14.367 23
随机森林	13.666 11

在测试集中每个贷款可能实现的收益和风险已被评估出来的情况下,通过有约束的投资组合选择求解出最优的资产配置,即投资者投分配给各借款人的比例,结合测试集中的各贷款的真实收益值、比较信用分级、逻辑回归以及随机森林这 3 种方法选出的投资组合的真实收益状况,即比较哪种方法的投资收益率更高、夏普率更低。为了验证模型的适用性,将数据分成 8 组按照上面所述方法进行测试,由于网贷平台借款的平均还款期限相对较短,故选取 15 年美国三年期的国债到期平均收益率作为无风险利率即 $r=1.05\%$ 。由于 lending club 上的最小投资额为 25,即模型中的 $m=25$ 。假设投资者想要实现的目标收益率 $R^*=0.07$,投资金额 M 为 5 000,通过 Matlab 运算得出每组数据下各项贷款的投资比例,并结合每个贷款的真实收益率数据,得到模型选择的投资组合所能实现的真实收益率。结果显示,随机森林判断出的投资组合收益率、夏普率最高,其次是逻辑回归方法,都优于基于信用分级的方法。8 组结果如表 3 所示。

表 3 投资组合收益比较

	信用分级	逻辑回归	随机森林
1	0.013 4	0.026 9	0.058 6
2	0.025 3	0.069 3	0.040 2
3	0.026 9	0.033 9	0.060 3
4	0.042 7	0.056 2	0.067
5	0.039 1	0.050 1	0.073 3
6	0.069 5	0.023 6	0.047 6
7	0.054 7	0.060 2	0.030 1
8	0.034 5	0.046 7	0.069 9
均值	0.038 263	0.045 863	0.055 875
标准差	0.017 711	0.016 382	0.015 258
夏普率	1.567 557	2.158 65	2.973 85

4 结论与建议

本文研究中提出了一个通过 P2P 网贷平台各贷款违约概率的距离去度量贷款间的相似关系,进而用相似的历史贷款去评估新进的贷款需求,为投

投资者提供了更详细的贷款收益风险信息,帮助投资者有效地提高了投资组合的收益率。本文的研究方法有以下优势:首先,相比于传统方法,提出的随机森林方法能够更精确的预测贷款的违约风险;其次,在没有同一借款人的之前的贷款信息的情况下,提出的距离度量模型可以利用与其相似的借款人的历史贷款数据信息去评估新贷款需求的收益风险状况。此外,历史贷款的权重使用核权重,能够在不需要太多统计假设的情况下保证结果的有效性。8 组数据测试结果显示,通过上述方法选择出来的投资组合比传统的根据借款人信用分级方法选择出的投资组合收益率更高,为 P2P 平台的投资者进行投资决策提供了更有效更明确的方法,进而实现更高收益,使投资者更有积极性地将资金应用到网贷平台,有助于网贷平台更健康地发展。由于国内网贷平台数据尚未公开,故本文选择了美国网贷平台的数据进行实证分析,由于本文模型的相对普适性,故研究内容对投资人投资决策有很重要的参考意义。随着国内网贷平台数据渐渐公开,之后的工作将基于国内的真实数据展开更进一步的研究。

参考文献:

- [1] Hulme M K, Wright C. Internet based social lending: Past, present and future[J]. Social Futures Observatory, 2006, 115
- [2] 刘峙廷. 我国 p2p 网络信贷风险评估研究[D]. 广西: 广西大学, 2013
- [3] Freedman S M, Jin G Z. Learning by doing with asymmetric information: Evidence from prosper. com [R]. New York: National Bureau of Economic Research, 2011
- [4] Dongyu C, Gezhi C, Jie D, *et al.* Antecedents of initial trust in the online peer-to-peer lending market-
place[C]//ICSSSM11. IEEE, 2011: 1-4
- [5] Lin M, Prabhala N R, Viswanathan S. Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending[J]. Management Science, 2013, 59(1): 17-35
- [6] Lee E, Lee B. Herding behavior in online P2P lending: An empirical investigation[J]. Electronic Commerce Research and Applications, 2012, 11(5): 495-503
- [7] 雒春雨. P2P 网络借贷中的投资决策模型研究[D]. 大连:大连理工大学, 2012
- [8] HaSH. Behavioral assessment of recoverable credit of retailer's customers[J]. Information Sciences, 2010, 180(19): 3703-3717
- [9] Mandala I G N N, Nawangpalupi C B, Praktikto F R. Assessing credit risk: an application of data mining in a rural bank[J]. Procedia Economics and Finance, 2012, 4: 406-412
- [10] 熊志斌. 基于非线性主成分分析的信用评估模型研究[J]. 数量经济技术经济研究, 2013, 30(10): 138-150
- [11] 杨胜刚, 朱琦, 成程. 个人信用评估组合模型的构建—基于决策树—神经网络的研究[J]. 金融论坛, 2013(2): 57-67
- [12] 姚潇, 余乐安. 模糊近似支持向量机模型及其在信用风险评估中的应用[J]. 系统工程理论与实践, 2012, 32(3): 549-554
- [13] Louveau A, Smirnov I, Keyes T J, *et al.* Structural and functional features of central nervous system lymphatic vessels[J]. Nature, 2015, 523: 337-341
- [14] Breiman L. Random forests[J]. Machine learning, 2001, 45(1): 5-32
- [15] Silverman B W. Density estimation for statistics and data analysis[M]. Boca Raton: CRC press, 1986