

基于风险和剩余价值的在线 P2P 借贷投资推荐方法

朱梦莹 郑小林 王朝晖

(浙江大学计算机科学与技术学院 杭州 310027)

(mengyingzhu@zju.edu.cn)

Investment Recommendation Based on Risk and Surplus in P2P Lending

Zhu Mengying, Zheng Xiaolin, and Wang Chaohui

(College of Computer Science & Technology, Zhejiang University, Hangzhou 310027)

Abstract Online peer-to-peer (P2P) lending, which is a newly personal wealth distribution and management system, has become a new type of financing mode for Internet users. P2P lending platform allows borrowers to create borrow listing and investors to bid and invest borrowers' listing directly. In the P2P lending, there is a significant issue that is how to reasonably match borrowers and investors and then allocate the amount of investors, so as to recommend low risk and high rate investment decisions to the investors. This paper proposes a recommendation framework risk based total surplus risk total surplus maximize (RTSM), which can solve the problem of allocating the investment amount into borrowers' listings. Specifically, we first propose to adapt various methods of regression to evaluate default risk. Then, we give the hypothesis the surplus of investors and borrowers under default risk which is based on the theory of surplus in economics. And based on this hypothesis, we combine the risk assessment and investment recommendation to maximize the total surplus under default risk. We apply the recommendation framework RTSM into two real-world datasets (Prosper and PPDai). Finally, experiments and analysis indicate that RTSM can reduce risk and improve the overall benefits of both investors and borrowers.

Key words online P2P lending; risk assessment; total surplus maximize; recommendation system; investment recommendation

摘要 在线 P2P(peer-to-pear)借贷是一种新兴的在线个人财富分配和管理系统,它允许投资人直接对借款人创建的借款标的进行竞标和投资.在 P2P 借贷平台中,存在一个重要的问题即如何合理分配投资人的投资金额给合适的借款人.针对该问题,提出了一种基于风险和剩余价值最大化的投资推荐框架 RTSM(risk total surplus maximize).RTSM 首先对借款标的进行风险评估,然后基于经济学中的剩余价值理论,使用投资人和借款人在有风险情况下的剩余价值假设,将风险评估与投资推荐结合在一起,为投资人推荐高收益低风险的投资决策.实验在风险评估和投资推荐 2 个阶段对美国和中国知名的 P2P 借贷平台(Prosper、拍拍贷)的真实数据进行分析验证.从实验结果可以看出:RTSM 可以更好地降低风险和提高投资人与借款人的整体利益.

收稿日期:2016-08-15;修回日期:2016-10-27

基金项目:国家自然科学基金项目(61379034,U1509221);国家科技支撑计划基金项目(2014BAH28F05,2015BAH07F01)

This work was supported by the National Natural Science Foundation of China (61379034, U1509221) and the National Key Technology R&D Program of China (2014BAH28F05, 2015BAH07F01).

通信作者:郑小林(xlzheng@zju.edu.cn)

关键词 在线 P2P 借贷；风险评估；最大化剩余价值；推荐系统；投资推荐

中图法分类号 TP391

P2P(peer-to-peer)借贷是指无担保借款,借贷双方通过没有金融机构中介的网络平台进行竞标和借贷行为^[1]. P2P 借贷作为互联网技术在金融领域的新应用^[2-3],能有效地促进信息发布和搜索,并提供必要的功能来完成交易^[4].

2005 年第 1 个借贷平台 Zopa 成立于欧洲(英国),2006 年 2 月美国第 1 个借贷平台(Prosper.com)成立. 2005 年以来,在线 P2P 借贷在许多国家,包括美国、加拿大、英国、日本、意大利和中国,以不同的形式经历了快速增长时期. 最成功的在线 P2P 平台是英国的 Zopa(Zopa.com)、美国的 Prosper(Prosper.com)、Kiva(Kiva.org)、Lending Club(lendingclub.com). 例如 Prosper 在 2009 年成功完成 1.7 亿美元的借款. 到 2016 年 7 月为止,总部设在旧金山的非盈利组织 Kiva 通过其平台获得的借款总额已经达到 8.82 亿美元^①.

尽管起步较晚,在线 P2P 借贷在中国也有相当

大的发展,产生了拍拍贷(PPDai.com)、宜人贷(yirendai.com)、人人贷(We.com)等具有一定影响力的平台. 国内最大的 P2P 借贷平台拍拍贷成立于 2007 年,在一年半的时间里就积累了超过 8 万名用户,使得这种金融投资新式迅速在国内发展. 而宜人贷作为中国第 1 家在美国纳斯达克上市的 P2P 平台,成为中国互联网金融在海外上市的首家公司. 截止 2015 年年底,中国一共拥有 3 858 家 P2P 借贷平台^②.

P2P 借贷在蓬勃发展的同时,也相应地产生了许多问题,例如借贷资金的安全仍会存在隐患,借款人的信息可信度如何保证,一笔借款产生后缺乏行之有效的对于还款的监管手段等. 这其中,如何评估 P2P 借贷过程中借贷风险,建立一个安全有效的 P2P 借贷投资推荐体系也是急需解决的问题之一.

P2P 借贷平台中,一部分平台直接连接借款人和投资人,而其他平台则通过第三方机构(通常是银行或其他支付机构)连接,如图 1 所示:

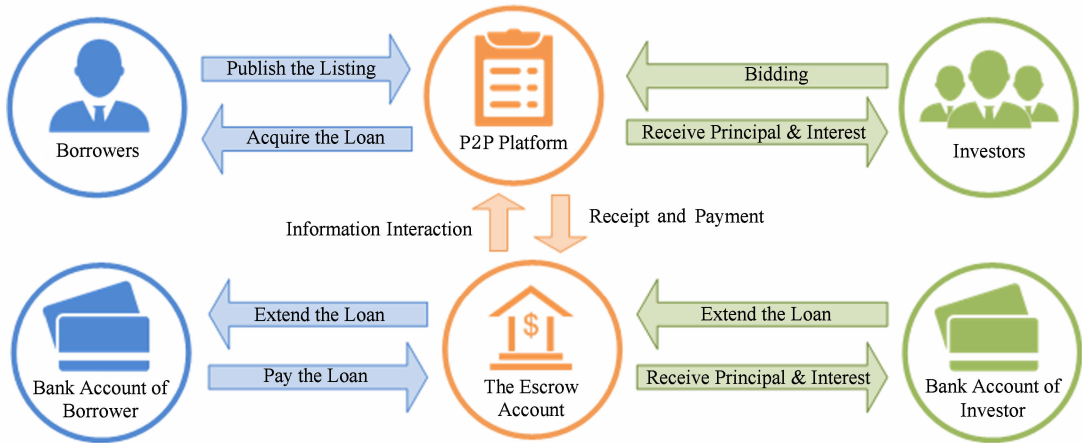


Fig. 1 The framework of P2P lending platform.

图 1 P2P 借贷平台架构图

作为借款人和投资人的信息发布媒介,P2P 借贷平台可以筛选出优质的借款标的,为投资人推荐高收益低风险的投资决策. 可是,现如今 P2P 借贷平台往往只对借款标的进行简单的排序,推荐给投资人较高收益的借款标的.

尽管现在的 P2P 借贷平台都表明自己拥有专业的风险评估团队,可以对所有借款标的进行风险

评估,但由于 P2P 借贷的风险评估环节不如银行严谨,违约风险成了 P2P 借贷系统中最严重的问题.

如图 2 所示,我们分析了拍拍贷和 Prosper 两个平台上借款标的期望投资收益和违约风险的关系. 从图 2 中可以看出,违约风险随着期望投资收益的增加而增加,所以,仅仅考虑投资收益进行投资推荐是不够的.

① <http://www.kiva.org/about>.

② <http://shuju.wdjl.com/industry-list.html>.

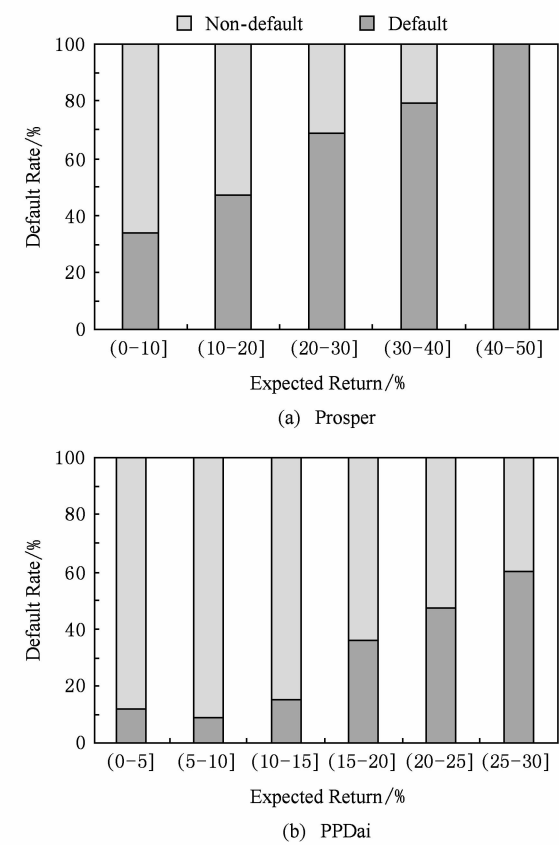


Fig. 2 The framework of P2P lending platform.
图2 收益与违约风险关系

本文将基于经济学中的剩余价值理论,结合风险评估与投资推荐解决 P2P 借贷过程中投资人的投资金额分配问题,为投资人推荐高收益低风险的投资决策.

1 风险评估和投资推荐相关工作

1.1 P2P 借贷中的风险评估

近年来,有 2 类工作大量出现在关于 P2P 借贷的风险研究中,分别是 P2P 借贷风险的关键因素研究和 P2P 借贷的风险评估方法研究.

研究者们致力于研究 P2P 平台中影响违约风险的共有的因素.在 P2P 借贷的环境下,投资人很难对借款人获得全面的信息,导致信息不对称的问题十分严重^[5].在 P2P 借贷中,借款人的个人信息和借款标的信息被认为是评估借款人是否违约的重要信号,可以用来评估借款人的违约风险和设置投资利率^[6-7].通过分析从 Prosper 收集的数据,Lin^[7]发现信用评级较低的借款请求不太可能被资助,并且这样的借款请求更有可能违约或以很高的利率结束.Emekter 等人^[8]则分析了另一家美国知名 P2P

平台 Lending Club,得出拥有更高的 FICO 分数和高收入的借款人在该平台违约率更低,平台需要想办法去吸引这样的借款人.在中国,Chen 等人^[9]表明,对拍拍贷平台的数据分析,信用分数对获得借款的概率产生了部分影响,但利率是决定因素.然而,违约率较高的借款人信用水平要低得多.小额信贷理论表明,社交网络可以帮助减少借款过程中的信息不对称,并且可以激励借款人偿还借款^[10].Freedman 和 Jin^[11]发现有借款推荐人或是 bid 是由借款人的朋友提交的,很少会拖欠还款,并且有更高的回报率.他们的结论是,借款人的朋友能更好地识别风险和拥有可信赖性,因为他们拥有更多的额外信息,而且他们认为社交网络的监控为还款提供了更强大的动力.

另一类工作则是致力于降低信息不对称情况下的借贷违约风险.研究表明,机器学习算法和大数据分析可以很好地应用在不同的商业应用中. Bahrammirazee^[12]通过对比 3 种人工智能技术:人工神经网络、专家系统、混合智能系统和传统方法在金融问题上的表现,得出人工智能技术比传统方法表现更好. Zhao 等人^[13]利用了 Logiest Regression 预测出借款标的的违约风险,并将风险利用在个性化推荐中. Baesens 等人^[14]使用神经网络技术分析了 3 个信用违约的数据集,构建了智能且可解释的风险评估系统. Byanjankar 等人^[15]也使用了神经网络方法预测 P2P 借贷中的风险,并且认为基于信用分数的神经网络在筛选违约标的时有非常有效的表现. 李斌等人^[16]使用了属性归纳和聚类方法对业务过程中的风险因素进行分析,并且采用 Markov 模型构建可传导的风险评估模型.

然而,这些研究工作大部分都只是致力于研究违约风险,却很少有将违约风险应用到其他应用场景或数据分析中,例如 Zhao 等人^[13]利用违约风险完成投资组合的个性化推荐以及利用对贷款 loan 的多视角风险分析结果制订投资组合策略^[17].

1.2 投资推荐

推荐系统根据用户的兴趣特点和购买行为,向用户推荐用户偏爱的信息和商品.推荐系统一般包括基于内容的推荐系统^[18]和协同过滤^[19].基于内容的推荐主要通过对用户交互的目标项目信息进行挖掘,为用户推荐相似度较高的目标项目;协同过滤方法则是基于相似用户对目标项目的历史交互行为来估计当前用户对目标项目的喜好程度,这种交互行为可以是点击、浏览、购买、评分.传统的推荐系统的改进是针对用户进行个性化的物品推荐,并且追求

推荐准确度的提高. 但是, 在 P2P 借贷或者其他的金融领域, 高准确度并不像在其他传统推荐领域那样重要, 例如电子商务领域^[20]、广告领域^[21]、电影传媒领域^[22].

在 P2P 借贷中, 我们希望投资人可以投资高收益的借款标的, 而低违约风险的借款人可以获得借款. 在文献[23]中, 作者提出可以引入经济学中的剩余价值, 平衡生产者和消费者双方的利益, 使得整个市场更有效率.

同时, 我们认为风险在 P2P 借贷的推荐中也是非常重要的衡量因素. 我们希望能推荐给投资人高收益低风险的借款标的. 在文献[13]中, 作者将 P2P 借贷的风险评估和协同过滤相结合, 为投资人提供个性化的借款标的投资组合推荐.

2 经济剩余价值

在本节中, 我们将介绍一些关键的概念和经济学中的定义, 这些将会成为第 3 节中描述的推荐框架的理论基础.

2.1 经济剩余

经济剩余是经济学中的概念, 用来衡量买卖商品或服务时的市场收益^[24]. 通常, 经济剩余可分为消费者剩余、生产者剩余.

消费者剩余(consumer surplus)是指购买者的支付意愿减去购买者的实际支付量. 消费者剩余衡量了购买者自己感觉到所获得的额外利益. 例如在 P2P 借贷场景中, 我们将借款人发布的标的作为一项产品, 消费者指的就是投资人. 那么投资人希望无风险获得的收益即他的支付意愿, 而借款标的实际的借款利息则是他实际支付的量. 如果尊重买者的偏好, 消费者剩余不失为经济福利的一种好的衡量标准.

生产者剩余(producer surplus)是指卖者出售一种物品或服务得到的价格减去卖者的成本. 同样在 P2P 借贷系统中, 生产者是借款人. 那么借款标的实际借款利息为借款人出售这一服务的价格, 借款人发布的最大愿意支付的借款利息为借款人出售借款标的来缓解经济负担的成本. 生产者剩余衡量生产者参与市场所得到的额外利益.

总剩余(total surplus)是指消费者剩余和生产者剩余的总和. 总剩余可以衡量社会的经济福利. 消费者剩余是买者从参与市场活动中得到的利益, 而生产者剩余是卖者得到的利益. 因此, 常把总剩余作为社会经济福利的衡量指标.

2.2 市场效率与总剩余价值

在微观经济学理论中, 福利经济学研究资源配置如何影响经济福利. 福利经济学分析市场的买者(消费者)和卖者(生产者)参与市场中得到的利益, 并且考虑如何使这种利益尽可能达到最大.

假设存在在一个消费者和生产者的分配矩阵 $\mathbf{Q}$, 则生产者剩余为 $PS(\mathbf{Q})$, 消费者剩余为 $CS(\mathbf{Q})$. 那么, 我们可以根据总剩余的定义得到总剩余价值 TS :

$$TS(\mathbf{Q}) = CS(\mathbf{Q}) + PS(\mathbf{Q}). \tag{1}$$

那么要使得市场的分配达到最有效率(efficiency), 我们需要使得总剩余最大化, 即:

$$\begin{aligned} & \underset{\mathbf{Q}}{\text{maximize}} \sum_{i=1}^m \sum_{j=1}^n TS_{ij}(\mathbf{Q}_{ij}), \\ & \text{s. t.} \quad \mathbf{1}^T \mathbf{Q} \leq \mathbf{M}, \mathbf{Q}_{ij} \in \mathbb{R}_+, \end{aligned} \tag{2}$$

其中, $\mathbf{M} = [M_1, M_2, \dots, M_n]$, M_j 是商品 j 可以被提供的总数量, 在 P2P 借贷系统中指借款标的 j 的总借款金额.

然而, 在这样的市场中还存在一个问题, 根据理性人假设^[25]的描述, 要满足充分有效的市场, 每个参与经济决策的主体必须是充满理智的, 他们的唯一目标就是追求自身经济利益的最大化. 例如消费者追求满足程度最大化, 生产者追求利润最大化. 可是, 在任何经济活动中, 这样的完全“合乎理性的人”是不可能存在. 尤其是在互联网金融的经济活动中, 由于存在类似自动投标的机制, 机器经济人^[26]也参与了这一经济活动. Zhang 等人^[23]在他们的文章中引入了概率, 认为消费者可能以更大概率选择最佳的分配决策, 但他们也可能以一个较小的概率选择一个非理性的决策, 他们定义分配矩阵 $\mathbf{Q}$ 中的每一个 $\mathbf{Q}_{ij}$ 是一个满足一定概率分布的随机变量. 在这种情况下, 最大化总剩余将被定义为

$$\begin{aligned} & \underset{\Theta(\mathbf{Q})}{\text{maximize}} \sum_{i=1}^m \sum_{j=1}^n \int TS_{ij}(\mathbf{Q}_{ij}) p(\mathbf{Q}_{ij}) d\mathbf{Q}_{ij}, \\ & \text{s. t.} \quad \mathbf{1}^T \int \mathbf{Q} p(\mathbf{Q}) d\mathbf{Q} \leq \mathbf{M}, \mathbf{Q}_{ij} \in \mathbb{R}_+, \end{aligned} \tag{3}$$

其中, $p(\mathbf{Q}_{ij})$ 是每一笔金额 $\mathbf{Q}_{ij}$ 的概率密度函数, $p(\mathbf{Q}) = [p(\mathbf{Q}_{ij})]_{m \times n}$, $\Theta(\mathbf{Q})$ 是所有 $\mathbf{Q}_{ij}$ 的参数集.

3 基于风险与剩余价值的投资推荐

在本节中, 我们首先介绍了风险评估, 然后描述了 P2P 借贷中投资人和借款人的剩余价值, 最后提出了基于风险与剩余价值的投资推荐框架.

3.1 风险评估

对于每一个借款标的, 投资收益和投资风险是

衡量标的的质量的最重要的 2 个因素^[27]. 在标的募集过程中, 借款者会发布自己愿意支付的最大借款利率($rate_j^{\max}$), 投资者则可以根据这个借款利率来衡量标的 p_j 的投资收益. 但是, 标的 p_j 的投资风险却很难被直接衡量. 因此, 我们需要根据标的和借款人的一些特征, 使用风险评估算法得到 1 个借款标的 p_j 的风险($risk_j$). 这样, 对于每一个借款标的 p_j , 我们用 1 个四元组表示: $p_j = (rate_j^{\max}, rate_j, risk_j,$

$M_j)$, 其中 $rate_j^{\max}$ 为借款者愿意支付的最大借款利率, $rate_j$ 为借款标的 p_j 的实际成交利率, $risk_j$ 为借款标的 p_j 的违约概率, M_j 为借款标的 p_j 所需要募集的资金上限.

在这里, 我们总结了 3 种类型的特征: 标的特征(features of listing, FOL)、借款者的用户特征(features of borrower, FOB)、借款者的社交特征(features of social, FOS), 具体的特征如表 1 所示:

Table 1 The Feature and Properties of PPDai and Prosper
表 1 拍拍贷和 Prosper 特征和属性

Type of Features	Subtype of Features	PPDai	Prosper
FOL		Amount Funded	Amount Funded
		PPDaiIR	Borrower Rate
		Loan Category	Category
		Payment Option	
		Loan Duration	Loan Term
		Vouch	
FOB	Identity Characteristics	ID Verification	Images
		Cellphone Certification	Verification Stage
		Education Certification	
		Vedio Certification	
	Financial Characteristics		Is Borrower Homeowner
			Debt To Income
	Demographic Characteristics	Gender	
		Age	
		EduStat	
		MrgStat	
FOS	Credit Characteristics	CldStat	
		Credit Grade	Credit Grade
		CdtBrw	Prosper Score
		CdtLnd	
		Friend Num	Group Rating
		A_FrdListN	Friend Num
		A_FrdListSucPct	A_FrdListN
		A_FrdBCreditS	A_FrdListSucPct
		A_FrdLCreditS	A_FrdCreditS

表 1 中标的特征指标的本身财务特征包括标的利率、募集的资金数量、借款期限、借款类型(房贷、车贷、商业贷等)、还款方式等.

借款者的用户特征分为 4 个部分:

1) 借款者的身份证明特征, 由于 P2P 借贷是互联网借贷, 借款者和投资人并不直接见面, 所以一些验证借款者身份的的必要措施被广泛使用在 P2P 借

贷平台的流程中, 例如身份证验证、照片验证、视频验证、学历证明验证等, 这里我们根据 Prosper 和拍拍贷 2 个平台的特点, 选取了符合它们各自平台流程的一些身份属性. 在 Prosper 数据集中我们选取了是否有头像图片这一属性; 在拍拍贷数据集中我们选择了是否有身份证验证、视频验证、学历验证、婚姻验证 4 个属性.

2) 借款者的财务特征,主要用于证明借款者的还款能力,在 Prosper 中我们选取借款者是否有房屋,借款者的收入负债比;但拍拍贷中并没有提供这方面的特征属性。

3) 借款者的人口学特征,在拍拍贷中,我们选取借款者的性别、年龄、学历、婚姻状态等;但在 Prosper 中我们并没有找到相关的特征。

4) 借款者的信用特征,在 Prosper 平台上,我们选择使用 FICO 给出的 Credit Grade 和平台自身给用户的信用分 Prosper Score 作为该平台的信用特征;在拍拍贷平台上,拍拍贷自己建立了信用分数评价体系,我们使用拍拍贷的用户信用分和信用评级作为该平台的信用特征。

借款者的社交特征是 P2P 借贷平台特有的一种特征类型。大多数 P2P 平台愿意在 P2P 借贷服务以外构建线上社区,用户在社区中存在“朋友”、“群组”等社交关系。近几年,有不少研究者研究了社交关系对 P2P 借贷信用体系和风险体系的影响。在这其中,我们选择 Prosper 的群组和朋友属性与拍拍贷的朋友属性作为借款者的社交特征。

使用历史借贷数据,我们可以根据以上这些特征,估计每个标的 p_j 的投资风险 $risk_j$ 。在这里我们并不打算提出某一个具体的模型算法,而是选择提出一种风险评估的框架,不同的平台、不同的标的特

征都可以选择不同的模型算法来评估投资风险,例如使用逻辑回归(logistic regression, LR)^[13]、GBDT(gradient boosting decision tree)^[17]、人工神经网络(artificial neural network, ANN)^[15]。

3.2 P2P 借贷中的剩余价值

在 P2P 借贷系统中,我们认为借款人是一笔借款请求的生产者,而借款请求被视作是一个金融产品;投资人则作为这个金融产品的消费者,他们购买这一笔借款中的一部分或全部来获得投资收益。

在这一问题中,我们主要需要解决的是投资人怎么样将自己的资金分配到每个金融产品,即每个借款请求中去,使得整个市场达到最有效率的配置,即满足最大化总剩余价值。

3.2.1 投资人剩余价值

在一个标准的在线 P2P 借贷流程中,如图 3 所示,借款人(借款请求的生产者) b_k 启动一个借款请求 p_j 。一旦借款人发布了借款请求,投资人(借款请求的消费者) u_i 可以对借款请求投标,他需要提供自己愿意出的金额和他希望得到的利率,这个利率会低于或等于 $rate_j^{\max}$ 。当所有金额在规定的时间内募集完成,借款请求就会成立,中标的投标记录是按照所有投标记录中利率从低到高排序后,择优中标。在所有中标的投标记录中,最高的利率会作为这个借款请求 p_j 的实际利率 $rate_j$ 。最后,借款人收到借款,借款标的进入还款程序^[28]。



Fig. 3 The borrowing process of P2P lending.

图 3 P2P 借贷平台借款流程

在文献[23]中,他们解决了不考虑风险控制的剩余价值最大化问题。在不考虑风险的情况下,投资人的剩余价值就是他们可以获得的投资收益减去无风险利率带来的收益。若无风险利率为 $\hat{r}$,那么我们可以得到:

$$CS_{ij}(Q_{ij}) = (rate_j - \hat{r})Q_{ij}. \quad (4)$$

3.2.2 借款者剩余价值

借款者的剩余价值指借款者愿意支付的借款利息减去实际支付的利息。通常的观念上,借款者不需要设置一个愿意支付的最大借款利率 $rate_j^{\max}$,他们总是希望这个利率越低越好。但是,借款者可能会因此无法吸引到足够的投资者。通过了解借款者发布

借款请求的原因,我们可以知道有些借款者是因为需要减轻当前的金融负担,例如还信用卡消费、还房贷、生意周转等。在这种情况下,借款者愿意支付的借款利息有可能是他们减轻或消除金融负担的最低成本。

同样地,在文献[23]中,借款者剩余价值可以定义为

$$PS_{ij}(Q_{ij}) = (rate_j^{\max} - rate_j)Q_{ij}. \quad (5)$$

因此,总剩余价值为

$$TS_{ij}(Q_{ij}) = CS_{ij}(Q_{ij}) + PS_{ij}(Q_{ij}) = (rate_j^{\max} - \hat{r})Q_{ij}. \quad (6)$$

3.3 基于风险与剩余价值的投资推荐

在本节中,我们将风险评估的结果引入 P2P 借

贷的剩余价值中,提出 RTSM 框架。

3.3.1 风险与投资人剩余价值

本文在考虑剩余价值时,考虑引入违约风险来平衡市场的效率与风险。

在不违约的情况下,即 $risk_j=0$ 的情况下,投资人的剩余价值如式(4)所示;而在违约的情况下,投资人的剩余价值是损失的本金减去无风险收益,所以我们认为投资人在 $risk_j$ 违约风险下的剩余价值为

$$CS_{ij}(Q_{ij})=((1-risk_j)rate_j-risk_j-\hat{r})Q_{ij}. \quad (7)$$

3.3.2 风险与借款者剩余价值

同理在不违约的情况下,借款者的剩余价值如式(5);但在违约的情况下,我们认为借款者愿意支付的收益是 0,即他发布的最大愿意支付的利率是一种为了骗取投资人本金的“欺诈”数据。在这种情况下,我们认为违约风险为 $risk_j$ 时,借款者的剩余价值为

$$PS_{ij}(Q_{ij})=((1-risk_j)rate_j^{max}- \\ (1-risk_j)rate_j+risk_j)Q_{ij}. \quad (8)$$

因此,在存在违约风险的情况下,总剩余价值为

$$TS_{ij}(Q_{ij})=CS_{ij}(Q_{ij})+PS_{ij}(Q_{ij})= \\ ((1-risk_j)rate_j^{max}-\hat{r})Q_{ij}. \quad (9)$$

3.3.3 基于风险和最大化总剩余的推荐算法 RTSM

我们假设分配矩阵 Q 满足高斯分布 $Q_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij})$,我们将 $TS_{ij}(Q_{ij})$ 带入式(3)中,可以得到最大化总剩余价值:

$$\begin{aligned} & \underset{U, \Sigma}{\text{maximize}} \sum_{i=1}^m \sum_{j=1}^n \int \frac{((1-risk_j)rate_j^{max}-\hat{r})Q_{ij}}{\sqrt{2\pi}\sigma_{ij}} \cdot \\ & \exp\left(-\frac{(Q_{ij}-\mu_{ij})^2}{2\sigma_{ij}^2}\right) dQ_{ij} \\ & \text{s. t. } \mathbf{1}^T \int \frac{Q}{\sqrt{2\pi}\Sigma} \exp\left(-\frac{(Q-U)^2}{2\Sigma^2}\right) dQ \leq \mathbf{M}, \\ & Q_{ij} \in \mathbb{R}_+, \end{aligned} \quad (10)$$

其中, $U=[\mu_{ij}]_{m \times n}$, $\Sigma=[\sigma_{ij}]_{m \times n}$.

式(10)可以被化简为

$$\begin{aligned} & \underset{U}{\text{maximize}} \sum_i \sum_j \int \mu_{ij} ((1-risk_j)rate_j^{max}-\hat{r}) d\mu_{ij} \\ & \text{s. t. } \mathbf{1}^T U \leq \mathbf{M}, \mu_{ij} \in \mathbb{R}_+. \end{aligned} \quad (11)$$

我们的 RTSM 推荐算法表述如下:

算法 1. RTSM 算法。

输入: P2P 数据 $P(\text{feature}, rate^{max}, rate, M)$;

输出: 投资人推荐金额分配矩阵 Q 。

步骤 1. P2P 借贷数据预处理。对 P2P 借贷数据的特征集进行归一化和标准化处理;

步骤 2. 构建风险预测模型,并且对模型进行训练测试,得到 $risk'_j=risk_model(\text{feature})$,在这里

风险预测模型可以是任何风险模型算法,例如 LR, GBMT, ANN;

步骤 3. 假设分配矩阵满足高斯分布 $Q \sim \mathcal{N}(\mu, \sigma)$,计算在总剩余价值最大化时的分配矩阵:

$$\begin{aligned} & \underset{U}{\text{maximize}} \sum_{i=1}^m \sum_{j=1}^n \int \mu_{ij} ((1-risk'_j)rate_j^{max}-\hat{r}) d\mu_{ij} \\ & \text{s. t. } \mathbf{1}^T U \leq \mathbf{M}, \mu_{ij} \in \mathbb{R}_+, \end{aligned} \quad (12)$$

其中, $U=[\mu_{ij}]_{m \times n}$;我们使用 linear programming 可以得到 U 的最优解;

步骤 4. 我们将推荐金额分配矩阵 Q 的期望值作为分配矩阵 Q 中的值,投资人推荐金额分配矩阵的期望值为

$$\bar{Q}_{ij}=\mu_{ij}. \quad (13)$$

4 实验结果与分析

4.1 数据集介绍

4.1.1 Prosper 数据集

Prosper 数据主要分为 5 种表: Listings, Loans, Bids, Groups, Members.

1) Listings. 借款人创建借款标的来募集贷方提供的资金。Listings 表中会说明借款人的情况以及募集资金的原因。当足够多的资金被募集, Listing 就会变成一笔借贷。

2) Loans. 当借款标的募集到足够的资金后,借款标的将会转变成 Loan。Loans 表中说明了借款人的还款情况。

3) Bids. 当借款人发布借款标的后,投资人可以对借款标的进行竞标,所有竞标的利率信息和金额信息都存在 Bids 表中。

4) Groups. 群组是由一些拥有相同兴趣或类似好友关系的用户组建的。每个群组都有一个组长,管理群组里的人员,介绍给组员一些借款标的。群组的分数代表了群组的可信程度。

5) Members. Members 表记录了在 Prosper 网站上注册的用户信息。用户可能拥有一个或多个身份,并且可能拥有一个或多个朋友。

Prosper 数据集包含了从 2005-11—2012-02 的所有交易记录 and 用户信息。

为了分析完成借贷流程的 P2P 借贷数据,本文剔除了标的状态为流标的全部数据和部分不相关的特征变量,数值化了非数值的特征变量。有关时间、地理信息、个人描述等特征在本文中被忽略。而信用等级(AA-HR)被转换成数值(7~0),同样地, Group

Rating(0~5 Stars)、是否拥有房产(No, Yes)被转换成数值(0,1)、Prosper Rating 等特征也被转换成了相应的数值.

最终,经过数据预处理的 PROSPER 数据集的结构如表 2 描述. 表 3 是 Prosper 数据集的指标性数据.

Table 2 Prosper Dataset Feature Description

表 2 Prosper 数据集字段描述

Type of Feature	Variable	Type	Description
FOL	Amount Funded	Decimal	The monetary amount of the listing which has been funded.
	Borrower Rate	Decimal	The final interest that the borrower will pay if the listing becomes a loan.
	Term	Integer	The length of the loan in months.
	Category	Integer	The category of this listing. The category is one of the following numerical values; 0—Not available; 1—Debt consolidation; 2—Home improvement; 3—Business; 4—Personal loan; 5—Student use; 6—Auto; 7—Other; 8—Baby & Adoption Loans; 9—Boat; 10—Cosmetic Procedures; 11—Engagement Ring Financing; 12—Green Loans; 13—Household Expenses; 14—Large Purchases; 15—Medical/Dental; 16—Motorcycle; 17—RV; 18—Taxes; 19—Vacation; 20—Wedding Loans.
	Listing Status	Boolean	The status of the listing is one of the following values; 0—Paid;1—Defaulted.
FOB	Images	Boolean	Photograph certification; 0—no;1—yes.
	Verification Stage	Integer	A three-stage indicator based on Prosper's verification of the borrower's information,0~3.
	Debt To Income Ratio	Decimal	The debt to income ratio of the borrower at the time the listing was created. This value is capped at 10.01.
	Is Borrower Homeowner	Boolean	Specifies whether or not the borrower has is a verified homeowner at the time the listing was created; 0—no;1—yes.
	Credit Grade	String	Credit grade of the borrower at the time the listing was created.
FOS	Group Rating	Integer	The group rating of a group following values,0~5 stars.
	FrdN	Integer	Number of friends of the borrower.
	A_FrdListN	Decimal	Average number of listing from friends.
	A_FrdListSucPet	Decimal	Average ratio of success listing from friends.
	A_FrdBCredit	Decimal	Average score of friends' credit grade.

Table 3 Statistics of the Selected Prosper Dataset

表 3 Prosper 数据集经济指标

# Listings	# Lenders	# Biddings	# Total Amount/US\$	Expected Return			Default Rate/%
				Minimum Rate/%	Maximum Rate/%	Average Rate/%	
49 789	60 373	23 857 678	304 338 314	0.01	49.75	19.89	56.7

4.1.2 拍拍贷数据集

拍拍贷数据主要分为 4 张表: List Info, User Info, Bid Info, Friend Info.

1) List Info. 借方创建借款标的来募集贷方提供的资金. 借款标的中会说明借方的情况和募集资金的原因. 当足够多的资金被募集, 借款标的就会变成一笔借贷, 进入还款程序, 在拍拍贷数据中还款情况也记入 List Info 表中.

2) User Info. 记录在拍拍贷网站上注册的用户信息. 用户可能拥有一个或多个身份, 并且可能拥有一个或多个朋友.

3) Bid Info. 当借款人发布借款标的后, 投资人可以对借款标的进行竞标, 所有竞标的标的利率信息和金额信息都存在 Bid Info 表中.

4) Friend Info. 在拍拍贷网站上用户可以互加好友, 这张表中包含用户和他的朋友 ID、创建时间等.

拍拍贷数据集包含了从 2007-06—2011-06 的所有交易记录 and 用户信息.

同样为了分析 P2P 借贷数据中已经完成借款流程的部分, 本文剔除了标的状态为流标的全部数据和部分不相关的特征变量, 数值化了非数值的特征变量. 有关时间、地理信息、个人描述等特征在本

文中被忽略. 其中, 借款类型被我们处理成一个 8 b 的二元组, 其他变量也都被归一化.

最终, 经过预处理的拍拍贷数据集的特征如表 4 描述, 表 5 是拍拍贷数据集上的指标性数据.

Table 4 PPDai Dataset Feature Description

表 4 拍拍贷数据集字段描述

Type of Feature	Variable	Type	Description
FOL	Amount Funded	Decimal	The monetary amount of the listing which has been funded.
	Payment Option	Boolean	The payment option is one of the following numerical values; 1—monthly repayment; 0—due repayment.
	PPDaiIR	Decimal	The final interest that the borrower will pay if the listing becomes a loan.
	Loan Duration	Integer	The length of the loan in months.
	Vouch	Boolean	The guarantee state is one of the following values; 0—no, 1—yes.
	Loan Category	Integer	The category of this listing. The category is one of the following numerical values; 0—no category; 1—debt consolidations; 2—home improvement; 3—personal loans; 4—business loans; 5—student loans; 6—automobile loans; 7—other loans.
	Listing Status	Boolean	The status of the listing is one of the following values; 0—Paid; 1—Defaulted.
FOB	Gender	Boolean	The gender of borrower, 0—male; 1—female.
	Age	Integer	The age of borrower.
	EduStat	Integer	The education level, the higher the score, the higher the degree.
	MrgStat	Boolean	The marital status, 0—unmarried; 1—married.
	CldStat	Boolean	Indicate whether the borrower owns a child, 0—no; 1—yes.
	IDVerification	Boolean	The identity verification, 0—no; 1—yes.
	Cellphone Certification	Boolean	The mobile phone certification, 0—no; 1—yes.
	Education Certification	Boolean	The education certification, 0—no; 1—yes.
	Video Certification	Boolean	The video certification, 0—no; 1—yes.
	Credit Grade	Integer	Credit grade of the borrower at the time the listing was created.
	LCreditS	Decimal	The score of investing credit grade if the borrower has a role of investor.
	BcreditS	Decimal	The score of borrowing credit grade.
FOS	FrdN	Integer	Number of friends of the borrower.
	A_FrdListN	Decimal	Average number of listing from friends.
	A_FrdListSucPct	Decimal	Average ratio of success listing from friends.
	A_FrdBCreditS	Decimal	Average score of friends' borrowing credit grade.
	A_FrdLCreditS	Decimal	Average score of friends' investing credit grade.

Table 5 Statistics of the Selected PPDai Dataset

表 5 PPDai 数据集经济指标

# Listings	# Lenders	# Biddings	# Total Amount/CNY	Expected Return			Default Rate/%
				Minimum Rate/%	Maximum Rate/%	Average Rate/%	
10 428	3 200	294 031	67 392 379	0. 01	27	17. 27	32. 7

4.2 风险评估实验结果与分析

首先, 我们将已经经过数据预处理的数据集分为训练样本和测试样本, 这里训练样本随机取整个数据集的 80%, 而 20% 为测试样本. 接着我们分别使用了 KNN(*k*-nearest neighbor), DT(Decision Tree), NB (Naive Bayesian), LR, ANN, RF (Random

Forests), GBDT 等方法对训练样本进行学习训练, 并对测试样本进行预测.

其中, 我们将采用 AUC 值和 *Accuracy* 值作为风险评估实验的评价指标, 具体计算方式如式(14)和式(15).

$$Accuracy = \frac{1}{N} \sum_{i=1}^N I(y_i = \hat{f}(x_i)), \quad (14)$$

其中, N 是数据容量; I 是指示函数, 即 $y = \hat{f}(x)$ 时为 1, 否则为 0.

$$AUC = \frac{\sum_{i \in \text{positiveClass}} \text{rank}_i - \frac{M(1+M)}{2}}{M \times N}, \quad (15)$$

其中, M 为正类样本的数目, N 为负类样本的数目.

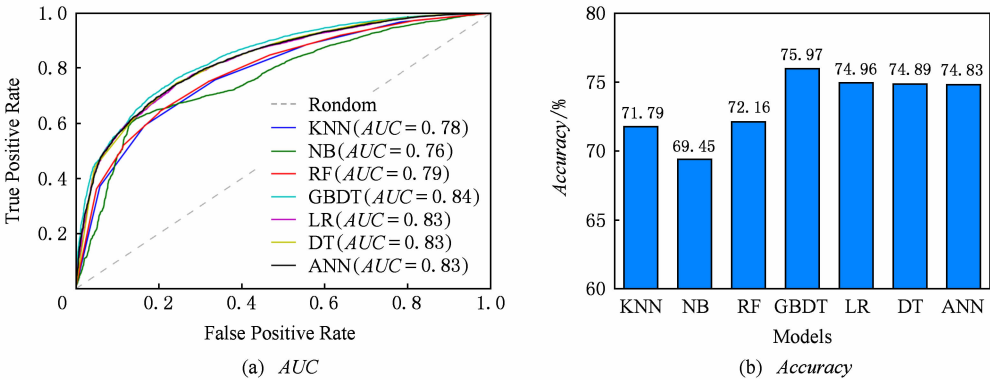


Fig. 4 The prediction performance of assessment Prosper dataset.

图 4 在 Prosper 数据集上各模型的预测表现

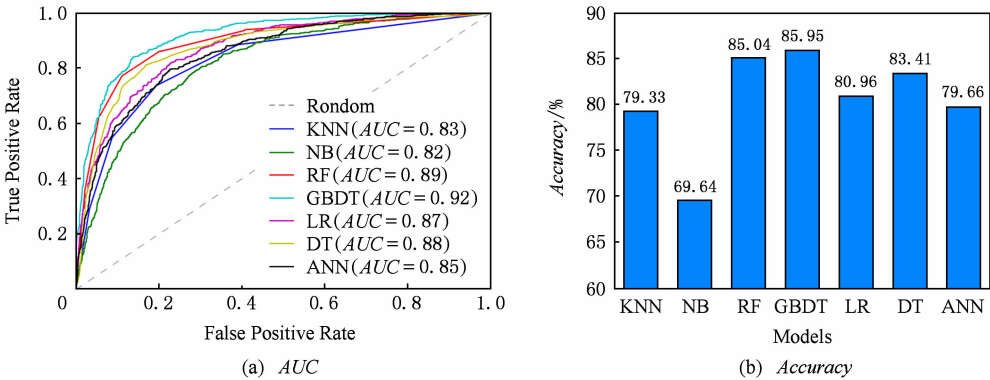


Fig. 5 The prediction performance of assessment on PPDai dataset.

图 5 拍拍贷数据集上各模型的预测表现

Table 6 The Accuracy and AUC on Prosper Dataset and PPDai Dataset				
表 6 Prosper 数据集和拍拍贷数据集上各模型的 Accuracy 和 AUC 值				
Dataset Evaluation	Prosper		PPDai	
	Accuracy/%	AUC	Accuracy/%	AUC
NB	69.45	0.76	69.64	0.82
KNN	71.79	0.78	79.33	0.83
ANN	74.89	0.83	79.66	0.85
LR	74.96	0.83	80.96	0.87
DT	74.83	0.83	83.41	0.88
RF	72.16	0.79	85.04	0.89
GBDT	75.97	0.84	85.95	0.92

当我们得到所有模型的最优参数后, 我们可以比较每一个模型的预测表现. 图 4 和图 5 显示了在 Prosper 数据集和拍拍贷数据集上各个模型的 Accuracy, AUC 值. 这其中, GBDT 算法无论从 Accuracy 还是 AUC 来看, 在 2 个数据集上都是最优的.

如表 6 所示, 对于 Prosper 数据集, GBDT 算法是表现最好的. 在 Accuracy 方面, 比其他模型算法提升 1.3%~9.4%; 在 AUC 方面, 比其他算法提高了 1.2%~10%. 对于拍拍贷数据集, GBDT 算法同样是表现最好的, 在 Accuracy 方面, 比其他提升了 1.1%~23.4%; 在 AUC 方面, GBDT 算法比其他算法提高了 3.3%~12.2%.

4.3 投资组合推荐实验结果

我们分别使用了来自中国和美国最知名的 P2P 借贷平台拍拍贷和 Prosper 的 2 个数据集来评价我们的推荐框架的表现.

在这里, 我们使用总剩余价值来衡量市场效率. 我们选取了真实的金额分配 (Actual) 作为基准. 分别计算了 RSTM 算法和 TSM 算法^[23]的结果, 并且

对它们进行了比较和分析.

为了计算总剩余价值,我们使用了每年的中国1年期定期存款利率和美国联邦基金利率作为各自的无风险利率,具体数据如表7所示. P2P 借贷的总剩余价值 TS 应该被计算成:

$$TS=\sum_{i=1}^m\sum_{j=1}^n(Q_{ij}(1-Default_j)rate_j^{\max}-\hat{r}_{\text{Year}}), \quad (16)$$

其中, $\hat{r}_{\text{Year}}$ 为每一年的无风险利率; $Default_j$ 为借款列表实际违约与否,若 $Default_j=1$ 则表示违约,若 $Default_j=0$ 则表示没有违约.

经过计算, Actual, TSM, RTSM 的市场表现(总剩余价值)的结果如表8所示.

表8的结果可以看出我们的RTSM框架在Prosper数据集中,总剩余价值和每一笔/每一单位金额的剩余价值都比实际值提升了28.46%. 每一单位金额可以获得的剩余价值从US\$0.0434提升到US\$0.0558,这表明我们提升了资金的效率. 在拍拍贷数据集中,总剩余价值和每一笔/每一单位金额的剩余价值都比实际值提升了9.44%. 每一单位

金额可以获得的剩余价值从CNY0.1019提升到CNY0.1116,同样也提升了资金的效率. 而与不带风险评估的TSM框架相比,我们的RTSM显然可以更好的适应有风险时的市场,并且带来更高效的市场配置.

Table 7 Risk Free Interest Rate in China and US

表7 中美无风险利率表

Year	Risk Free Interest Rate/%	
	US	China
2005	3.22	2.07
2006	4.97	2.52
2007	5.02	3.465
2008	1.92	3.06
2009	0.16	2.25
2010	0.18	2.50
2011	0.10	3.25
2012	0.14	3.25
2013	0.11	3.00

Table 8 Results on Total Surplus with RTSM, TSM and Without Framework

表8 Actual、TSM、RTSM 框架下的总剩余价值和单位剩余价值

Dataset Evaluation	Prosper				PPDai			
	$TS/US\$$	$(TS/Listing)/US\$$	$(TS/capita)/US\$$	Default Rate/%	TS/CNY	$(TS/Listing)/CNY$	$(TS/capita)/CNY$	Default Rate/%
Actual	13216721	265.45	0.0434	56.7	6870562	658.86	0.1019	32.7
TSM	16865283	338.74	0.0554	56.5	7517381	720.88	0.1115	32.5
RTSM	16977759	340.99	0.0558	45.3	7519358	721.07	0.1116	25.3

同样,我们还分析了使用RTSM推荐框架配置资产以后,市场的风险的情况. 我们计算了在TSM和RTSM框架配置下借款标的违约率. 在Prosper中,实际违约率为56.7%,TSM的违约率为56.5%,而RTSM的违约率只有45.3%,比实际的违约率降低了11.4%;而在拍拍贷中,实际违约率为32.7%,TSM的违约率为32.5%,而RTSM的违约率只有25.3%,降低了7.4%,说明RTSM有效地降低了市场中的风险.

5 结束语

大部分现有的推荐系统想办法引入新的算法来提升评分矩阵的预测准确性,例如RMSE,MAE等,但是很少有论文聚焦在什么才是最好的目标矩阵,使

得目标矩阵更符合市场经济,并且更有可解释性.

本文基于风险控制和微观经济学原理构建了一个新的推荐系统框架. 在RTSM框架中,我们充分考虑了P2P借贷市场的特殊性,利用了金融市场最重要的2个特点——风险与收益. 我们认为只是追求不考虑风险的情况下,收益最大化还不足以满足真实的市场经济的规则,所以引入了风险评估. 不同于传统的风险评估使用专家经验和简单的统计结论,我们采用机器学习的思想来完成风险评估,通过实验可以看出,GBDT模型可以很好地提高风险评估的结果. 在引入风险的情况下,我们首次定义了投资人剩余价值和借款人剩余价值,并且基于这一定义得出了总剩余价值. 在真实的数据集上,我们也验证了考虑风险的情况下,引入风险评估对整个市场的有效配置有所提升.

P2P 借贷系统作为近年来兴起的全新的投资借贷模式,它存在很多不同于传统的线下借贷的优势和特点.但由于 P2P 借贷平台本身的风险控制不够严谨和投资人获得的信息不对称,P2P 借贷产生了不少问题和争议.我们认为 3 个方向值得我们进一步研究:

1) 基于大数据分析的 P2P 借贷金融风险控制

P2P 借贷中的风险不仅仅单指借款人的违约风险,它是一个多层次复杂的风险,包括信用风险、平台风险、技术风险等.基于大数据分析构建符合 P2P 借贷模式的风险控制,将会是提升 P2P 借贷平台资产质量的重要手段.

2) 基于行为金融学的 P2P 借贷投资推荐

羊群行为一直是行为金融学中的重要话题,在信息不对称的 P2P 借贷领域,这一现象尤为典型.通过对 P2P 借贷中信息传递与羊群行为的分析,可以构建投资推荐模型.这一方向可以对投资人非理性投资做出很好的解释和修正.

3) P2P 借贷投资组合推荐

投资组合可以很好地分散投资人的风险,经常被应用在证券资产中.但在有着较高风险的 P2P 借贷投资中,投资组合也许会有助于帮助投资人做出高收益低风险的投资推荐.所以,投资组合推荐将会是投资推荐中一个重要的研究方向.

参 考 文 献

- [1] Bachmann A, Becker A, Buerckner D, et al. Online peer-to-peer lending—A literature review [J]. *Journal of Internet Banking & Commerce*, 2011, 16(2): 1-18
- [2] Collier B C, Hampshire R. Sending mixed signals: Multilevel reputation effects in peer-to-peer lending markets [C] //Proc of the 2010 ACM Conf on Computer Supported Cooperative Work (CSCW 2010). New York: ACM, 2010: 197-206
- [3] Lin Mingfeng, Prabhala N R, Viswanathan S, et al. Social networks as signaling mechanisms: Evidence from online peer-to-peer lending [R/OL]. Washington: Smith School of Business, University of Maryland, 2009. [2016-08-01]. http://people.stern.nyu.edu/bakos/wise/papers/wise2009-p09_paper.pdf
- [4] Iyer R, Khwaja A I, Luttmer E F P, et al. Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending?, 15242 [R]. Rochester, NY: SSRN's eLibrary, 2009
- [5] Lin Mingfeng, Prabhala R, Viswanathan S. Judging borrowers by the company they keep 1: Social networks and adverse selection in online peer-to-peer lending [J]. *SSRN Elibrary*, 2009, 59(1): 17-35
- [6] Herzenstein M, Andrews R L. The democratization of personal consumer loans? Determinants of success in online peer-to-peer loan auctions [J]. *Bulletin of the University of Delaware*, 2008, 15(3): 274-281
- [7] Lin Mingfeng. Peer-to-Peer lending: An empirical study [C] //Proc of the AMCIS 2009. Atlanta, GA: AIS, 2009: 17-25
- [8] Emekter R, Tu Yanbin, Jirasakuldech B, et al. Evaluating credit risk and loan performance in online peer-to-peer (P2P) lending [J]. *Applied Economics*, 2015, 47(1): 54-70
- [9] Chen Dongyu, Hu Huiling, Lou Hao, et al. Herding behavior in online microloan markets: Evidence from China [C/OL] //Proc of the WHICEB 2015 Proceedings. Atlanta, GA: AIS, 2015: 1-14 [2016-08-01]. <http://aisel.aisnet.org/whiceb2015/84/>
- [10] Krumme K A, Herrero S. Lending behavior and community structure in an online peer-to-peer economic network [C] //Proc of the 4th Int Conf on Computational Science and Engineering. Los Alamitos, CA: IEEE Computer Society, 2009: 613-618
- [11] Freedman S, Jin G Z. The information value of online social networks: Lessons from peer-to-peer lending, 19820 [R]. Cambridge, MA: National Bureau of Economic Research, 2014
- [12] Bahrammirzaee A. A comparative survey of artificial intelligence applications in finance: Artificial neural networks, expert system and hybrid intelligent systems [J]. *Neural Computing & Applications*, 2010, 19(8): 1165-1195
- [13] Zhao Hongke, Wu Le, Liu Qi, et al. Investment recommendation in P2P lending: A portfolio perspective with risk management [C] //Proc of the IEEE Int Conf on Data Mining. Los Alamitos, CA: IEEE Computer Society, 2014: 1109-1114
- [14] Baesens B, Setiono R, Mues C, et al. Using neural network rule extraction and decision tables for credit-risk evaluation [J]. *Management Science*, 2003, 49(3): 312-329
- [15] Byanjankar A, Heikkila M, Mezei J. Predicting credit risk in peer-to-peer lending: A neural network approach [C] //Proc of 2015 IEEE Symp Series on Computational Intelligence. Los Alamitos, CA: IEEE Computer Society, 2015: 719-725
- [16] Li Bin, Xie Feng, Chen Zhong. A business oriented risk assessment model [J]. *Journal of Computer Research and Development*, 2011, 48(9): 1634-1642 (in Chinese)
(李斌, 谢丰, 陈钟. 一种面向业务的风险评估模型[J]. *计算机研究与发展*, 2011, 48(9): 1634-1642)
- [17] Zhao Hongke, Liu Qi, Wang Guifeng, et al. Portfolio selections in P2P lending: A multi-objective perspective [C] //Proc of the 20th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 2075-2084
- [18] Pazzani M J, Billsus D. Content-Based Recommendation Systems [M]. Berlin: Springer, 2007

[19] Su Xiaoyuan, Khoshgoftaar T M. A survey of collaborative filtering techniques [J/OL]. Advances in Artificial Intelligence, 2009, 2009(4): 1-19 [2016-08-01]. <https://www.hindawi.com/journals/aai/2009/421425>

[20] Musat C C, Faltings B. Personalizing product rankings using collaborative filtering on opinion-derived topic profiles [C] // Proc of the 29th Int Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2015: 830-836

[21] Theocharous G, Thomas P S, Ghavamzadeh M. Ad recommendation systems for life-time value optimization [C] //Proc of the 24th Int Conf on World Wide Web. New York: ACM, 2015: 1305-1310

[22] Wang Ruiqin, Jiang Yunliang, Li Yixiao, et al. A collaborative filtering recommendation algorithm based on multiple social trusts [J]. Journal of Computer Research and Development, 2016, 53(6): 1389-1399 (in Chinese)
(王瑞琴, 蒋云良, 李一啸, 等. 一种基于多元社交信任的协同过滤推荐算法[J]. 计算机研究与发展, 2016, 53(6): 1389-1399)

[23] Zhang Yongfeng, Zhao Qi, Zhang Yi, et al. Economic recommendation with surplus maximization [C] //Proc of the 25th Int Conf on World Wide Web. New York: ACM, 2016: 73-83

[24] Baran P. Economic progress and economic surplus [J]. Economic Review, 1954, 17(4): 289-317

[25] Coleman J S, Fararo T J. Rational choice theory: Advocacy and critique [J]. Contemporary Sociology, 1994, 23(4): 617-618

[26] Parkes D C, Wellman M P. Economic reasoning and artificial intelligence [J]. Science, 2015, 349(6245): 267-272

[27] Luo Chunyu, Xiong Hui, Zhou Wenjun, et al. Enhancing investment decisions in P2P lending: An investor composition

perspective [C] //Proc of the 17th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2011: 292-300

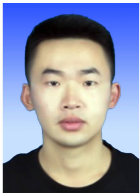
[28] Garman S R, Hampshire R C, Krishnan R, et al. Person-to-person lending: The pursuit of (more) competitive credit markets [C/OL] //Proc of the 29th Int Conf on Information Systems. Atlanta, GA: AIS, 2008: 1-16 [2016-08-01]. <http://aisel.aisnet.org/icis2008/160/>



Zhu Mengying, born in 1992. PhD candidate. Student member of China Computer Federation. Her research interests include artificial intelligence, data mining, recommendation system, and Internet finance.



Zheng Xiaolin, born in 1977. PhD. Associate professor, PhD supervisor, deputy director of Institute of Artificial Intelligence, Zhejiang University. Senior member of IEEE and China Computer Federation, committee member in Service Computing of China Computer Federation. His main research interests include data mining, recommendation system and Internet finance.



Wang Chaohui, born in 1993. Master candidate. His main research interests include data mining, recommendation system and Internet finance (zjuwch@zju.edu.cn).