

考虑投资者朋友关系和预期效用的 P2P 借贷个性化投资推荐方法

万常选 游运 江腾蛟 刘喜平 廖国琼 刘德喜

(江西财经大学信息管理学院 南昌 330013)

(江西财经大学数据与知识工程江西省高校重点实验室 南昌 330013)

(wanchangxuan@263.net)

Personalized Investment Recommendation in P2P Lending Considering Friend Relationships and Expected Utilities of Investors

Wan Changxuan, You Yun, Jiang Tengjiao, Liu Xiping, Liao Guoqiong, and Liu Dexi

(School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330013)

(Jiangxi Key Laboratory of Data and Knowledge Engineering, Jiangxi University of Finance and Economics, Nanchang 330013)

Abstract With the rapid development of Internet technology, online P2P lending market investment recommendation has become an important research direction in the field of online finance. For potential investors in P2P lending market, there are two key issues which need to be solved. One is how to choose the right investment projects considering their investment needs and preferences, the other is how to reasonably allocate their investment amount to these projects. Previous studies on these two questions mainly focused on default risk prediction of lending projects, global optimal product recommendation and portfolio optimization. With further research, limitations of recommendation model which was designed based on historical transaction data and utility indifference of investors have become increasingly prominent. It does not meet the investment decision needs of investors with different risk preferences. In view of this, firstly, based on the historical data of Prosper platform, this paper establishes the P2P relationship network model and calculates the concept features of the lending projects and investors separately, and then obtains the corresponding conceptual model. Secondly, we build friend network model to capture the mutual influence of investment behavior among investors, excavate investment behavior influence factor and take it as indicator variables of investment interest prediction, then generate a candidate investment recommendation list. Thirdly, in order to optimize the shares of each recommended candidate, individualized portfolio recommended framework is constructed based on expected utility theory considering the influence of risk preference difference on investment demand of investors. Finally, the recommendation results of our model are compared with other benchmark models to comprehensively evaluate its recommend effect. We implement experiments on real datasets of Prosper platform,

收稿日期:2017-09-01;修回日期:2018-03-16

基金项目:国家自然科学基金项目(61562032,61662027,61363010,61363039,61462037);江西省自然科学基金重大项目(20152ACB20003);江西省研究生创新专项资金项目(YC2016-B062)

This work was supported by the National Natural Science Foundation of China (61562032, 61662027, 61363010, 61363039, 61462037), the Grand Natural Science Foundation of Jiangxi Province (20152ACB20003), and the Special Funds for Graduate Students of Jiangxi Province(YC2016-B062).

experimental results demonstrate that our method has better recommendation quality than traditional investment recommendation method.

Key words P2P lending; P2P relationship network; friend relationship; expected utility; investment recommendation

摘 要 随着互联网技术的快速发展,在线 P2P 借贷市场投资推荐已经成为网络金融领域的重要研究方向. 对于 P2P 借贷市场潜在投资者来说,需解决的关键问题包括 2 个方面:1)如何选择真正符合自己投资需求和偏好的投资项目;2)如何将自己的投资金额在这些投资项目中进行合理分配. 以往关于这两者的研究主要是侧重在借贷项目的违约风险预测、投资项目全局推荐及投资组合优化等方面. 而随着研究的深入可以发现,仍在投资者效用无差异假设及投资者历史交易数据的基础上设计推荐模型,将难以满足不同风险偏好投资者的投资决策需求,保证推荐的有效性. 鉴于此,1)基于 Prosper 平台历史数据建立 P2P 关联网络模型,并分别计算借贷项目和投资者的概念特征,得出相应的概念模型;2)进一步考察 P2P 关联网络模型中的投资者朋友关系,以捕获投资者之间投资行为的相互影响,发掘投资者投资行为的影响因子,并将其应用于借贷项目兴趣度的预测,以提高投资项目推荐的有效性;3)在此基础上,从预期效用理论出发,进一步考虑投资者风险偏好对投资需求的影响,建立个性化投资组合推荐框架,以提高其投资的满意度和经济性能;4)将其推荐结果与其他基准模型的推荐结果进行对比分析,以综合评价其推荐效果. 在 Prosper 平台真实数据的基础上进行了详细的实验测试,结果表明:该方法相较于传统的投资推荐方法具有更好的推荐效果.

关键词 P2P 借贷;P2P 关联网络;朋友关系;预期效用;投资推荐

中图法分类号 TP311

P2P 借贷是基于因特网的一种新型金融模式,该模式将拥有闲散资金的社会成员与资金的需求方联系起来,不仅为中小企业融资和小额借贷提供了一个新的解决方案,也为社会成员提供了一种新的投资途径. 由于 P2P 借贷市场具有能拓宽传统金融机构的融资渠道,降低借贷交易成本,提高借贷交易效率和投资者投资回报等特点,该类市场发展迅猛,也吸引着越来越多的潜在投资者. 对于潜在的投资者来说,要保证其投资效益,其中关键的决策问题就是如何从众多的投资项目中选择真正符合自己投资偏好的投资项目,并在此基础上进行投资额的合理分配.

在 P2P 平台中,贷款人(也称为投资者)在进行投资项目的选择时,贷款人和借贷项目的风险水平往往是其决策的重要市场信号. 目前,关于贷款人和借贷项目风险的研究较多^[1-3]. 这些研究通过对平台中的贷款人、借贷项目的风险及其相关因素进行研究,以帮助潜在的投资者进行有效决策. 然而,随着平台发展规模越来越大,借贷项目品种的日益繁多,信用风险的相关信息量剧增.“信息过载”将使得投资者,特别是众多的非专业的投资者陷入决策困境.

文献[4]设计了一个基于实例的信用风险评估模型来评估每个借贷项目的风险和回报,将 P2P 借贷投资决策问题设计成一定约束条件下的优化投资组合问题,以提高 P2P 借贷市场投资者的投资性能. 但该方法是根据 P2P 借贷市场数据驱动的投资决策总体框架,即向所有的投资者推荐若干全局最优产品,因此忽视了不同投资者在投资偏好、风险类别等方面的差异性对其投资行为的影响.

文献[5]针对不同投资者,重构投资决策过程,首先,根据借贷项目信用评估结果及投资者历史交易数据,为每一位投资者产生一份推荐投资项目列表;其次,根据投资组合理论,对该推荐列表中的每一个投资项目的投资份额进行优化. 该方法考虑了不同投资者当前的投资状态的差异性,提出了基于风险管理的个性化推荐策略. 然而,该方法存在 2 个局限性:1)在确定推荐投资项目列表时,主要是从借贷项目与投资者之间的投资关系、贷款人与借贷项目之间的从属关系等角度出发进行算法设计,忽视了 P2P 平台在线社交网络中贷款人之间的朋友关系及其对投资者投标行为的影响,这将降低投资项目推荐的准确率和召回率,影响推荐结果的有效性;

2) 忽视了投资者风险偏好差异性对其投资决策的影响。

文献[6]通过在大型的 P2P 借贷网站上进行研究发现,朋友网络可带来管道效应、棱镜效应和羊群效应,会极大地影响相关投资者的投标行为及投资偏好,即当一个投资者的亲密朋友对项目 A 投标后,该投资者对项目 A 的投资偏好将会受到一定程度的影响,导致其可能也随之投标。

文献[7]发现投资者的投资组合与他们的投标偏好密切相关。因此,在确定推荐投资项目时可在前人研究的基础上,进一步考虑 P2P 平台在线社交网络中投资者之间的朋友关系及其对投资者投标行为的影响,以便准确地把握投资者真正感兴趣的借贷项目,改善投资组合推荐的最终效果;二是在进行 P2P 投资项目投资组合优化分析时,文献[7]与大多数 P2P 投资推荐相关研究类似,忽视了不同投资者风险偏好的差异性对其投资决策的影响。在金融领域中,投资者对待风险的态度一般通过投资者的效用函数来测定,在面临不确定的环境状态时,投资者往往更愿意选择使其预期效用最大化的决策方案。

基于以上分析,为了能够提高组合投资推荐的各方面性能,本文根据来自 Prosper 平台的相关数据,构建 P2P 关联网模型,分析其中的投资者之间的朋友关系,基于朋友关系网络计算投资者的朋友投标行为影响因子,并在此基础上设计基于预期效用理论的个性化推荐方案。

本文的贡献主要包括以下 4 个方面:

1) 根据 Prosper 平台相关数据,构建了反映平台各对象之间相互关联的 P2P 关联网模型,并在此基础上分别计算借贷项目和投资者的概念特征,得出相应的概念模型;

2) 分析 P2P 关联网模型中投资者之间的朋友关系,并在此基础上探讨投资者直接朋友的投标行为对其投资偏好的影响,设计基于投资者朋友关系的朋友投标行为影响因子,并将之应用于投资者的项目投资兴趣度的预测,以提高投资推荐的有效性;

3) 在 Prosper 平台历史交易数据基础上,考察投资者风险厌恶系数及风险承受能力,结合预期效用理论,构建基于预期效用理论的 P2P 借贷平台投资组合优化模型,实现投资者投资额在推荐投资项目列表中的合理分配;

4) 整体设计了基于投资者朋友关系和预期效用最大化的投资组合推荐算法,并提出了新的评价

指标,以评估投资推荐模型的推荐效果以及投资决策满意度、经济效益和效用。

1 相关研究

对于 P2P 借贷平台来说,如何为潜在的投资者提供有效的投资信息,帮助他们进行投资项目的选择及投资额的分配,以降低他们的投资风险,提高他们的投资回报,关系到 P2P 借贷平台的长远而有效的发展,关于这方面的研究主要体现在以下 3 个阶段。

1.1 借贷项目违约风险研究

通过借贷项目违约风险评估,可在一定程度上帮助投资者摆脱由于信息不对称等信息问题所产生的不利选择困境,规避投资决策风险。目前关于这方面的研究较多,大多从 2 个方面着手:1) 从借贷项目违约风险的相关特征出发,研究传统的财务指标^[8-9]、非财务指标^[10]及社会网络信息^[11-13]等对借贷项目违约风险的影响;2) 从借贷项目违约风险预测方法的角度展开研究。关于市场风险预测的方法有很多,大致可分为 4 类:1) 统计模型,如逻辑斯蒂回归、 k 近邻方法和生存分析等;2) 人工智能方法,如神经网络、支持向量机和遗传算法等;3) 运筹方法,如线性规划法、二次规划法等;4) 混合方法。以上这些研究工作主要集中于研究借贷项目违约风险预测,却很少有将预测结果应用到其他应用场景或投资决策分析。

1.2 基于违约风险的投资组合策略研究

随着 P2P 借贷市场的飞速发展,借贷项目品种变得日益繁多,而投资者非专业性及处理信息的能力有限。因此,对于投资者而言,仅拥有关于各个借贷项目违约风险的相关信息远不足以保证其投资决策的有效性,因为信息过量往往会使决策者陷入困境,因此有必要提高投资者信息分析的效率,并为之提供有效的投资策略。

文献[14]根据来自多目标项目的静态特征和动态特征,对借贷项目进行多视角风险分析,并在此基础上为投资者提供 2 种投资组合优化策略,即加权目标优化策略和多目标优化策略。文献[15]提出数据驱动的投资决策框架,利用投资者的投资组合理论来增强 P2P 投资决策性能。文献[4]通过设计一个基于实例的信用风险评估模型来评估每个借贷项目的风险和回报,并在此基础上将 P2P 借贷投资决策问题设计成一定约束条件下的优化投资组合问题,

以提高 P2P 借贷市场的投资性能. 以上方法是根据 P2P 借贷项目风险评估建立投资决策总体框架, 向所有的投资者推荐若干全局最优产品, 因此忽视了不同投资者投资偏好的差异对其投资行为的影响.

1.3 个性化投资推荐

推荐系统按所使用的数据来分类, 可以分为内容过滤、协同过滤和社会化推荐系统等. 已有的推荐算法虽然取得了一定的推荐效果, 但主要局限于一些特定的领域, 如旅游业、电子商务、电影传媒推荐等. 关于 P2P 借贷领域的个性化投资推荐的研究较少.

文献[16]通过对借贷项目进行风险评估, 并在此基础上, 结合经济学剩余价值理论, 提出剩余价值最大化投资推荐框架, 具有一定的有效性; 但对 P2P 在线平台各对象之间的相互关联研究较少, 同时也忽视了投资者之间的朋友关系及其对投资者投资偏好的影响, 导致难以准确把握投资者的当前投资需求, 影响推荐的效果.

文献[5]根据借贷项目信用评估结果及投资者历史交易数据, 从借贷项目与投资者之间的投资关系、借款人与借贷项目之间的从属关系等角度出发为投资者推荐投资项目列表, 并在此基础上结合投资者当前的投资环境, 对推荐列表中的每一个投资项目的投资份额进行优化. 该方法考虑了不同投资者投资环境的差异性, 提出了基于风险管理的个性化推荐策略. 但该方法忽视了 P2P 平台在线社交网络中投资者之间的朋友关系及其对投资者投资偏好的影响, 导致难以准确把握投资者真正感兴趣的借贷项目, 影响投资推荐的有效性. 同时, 该方法忽视了投资者风险偏好差异性对其投资决策的影响. 由于不同的投资者对待风险的态度差异, 如风险厌恶程度不一样, 甚至风险类别不同, 在面对风险不同的同样货币预期值的投资收益时, 产生的效用水平往往不同, 而这很有可能会影响投资者的投资组合策略.

目前, 在 P2P 借贷领域社交网络的研究大多集中在风险预测、融资效率及经济行为影响等方面; 关于 P2P 借贷领域投资推荐的相关研究中, 在进行个性化投资推荐算法设计时, 往往忽视了投资者社交关系对投资者投标行为的影响. 由于 P2P 借贷市场存在信息不完全与非对称、历史投资数据稀疏、投资者与投资项目特征难以有效匹配等问题, 传统的推荐方案往往难以满足该市场的个性化投资需求.

基于在线社交网络的数据分析结果表明, 考虑

社会网络用户之间的直接或间接的社会关联信息的社会化推荐方法, 可有效提高推荐系统性能^[17]. 随着当前在线社交网络的不断发展与应用, 结合投资者社交网络信息进行推荐算法设计, 将有望成为提高 P2P 借贷市场个性化投资服务性能的重要途径之一.

因此, 本文根据投资者的历史交易数据, 构建反映平台各对象之间相互关联的 P2P 关联网络模型, 分析其中的投资者之间的朋友关系及其对投资者投资决策行为的影响; 同时, 考虑投资者历史风险承受能力及风险系数, 从投资者的预期效用最大化出发, 设计个性化投资组合推荐方案, 以期提高投资推荐的有效性和投资者的决策满意度.

2 借贷项目和投资者的概念模型

本节将依据 P2P 借贷市场中各对象之间的相互关系, 构建 P2P 关联网络模型, 并在此基础上分别分析借贷项目和投资者的概念特征, 得出相应的概念模型, 为后续设计个性化投资推荐模型提供支持.

2.1 构建 P2P 关联网络模型

微观社会资本理论认为, 社会网络中的个体行动者的社会地位状况及关系指向特征将影响其决策行为与经济结果. 朋友网络是 P2P 借贷市场中的一种重要的在线社交网络, 有别于传统的社会网络, 该网络中的注册用户之间可以不受时空的限制自由地组织联系、快速地进行信息分享和市场交易, 并由此建立各种关联. 基于微观社会资本理论和 P2P 平台历史交易信息, 通过分析 P2P 借贷市场中各对象之间的相互关联, 构建相应的 P2P 关联网络模型, 如图 1 所示, 以分析其中各对象之间的关联特征.

图 1 定义了 P2P 关联网络中互为关联的相关对象的关系图 $G=(C,P,E)$, 其中 P 表示借贷项目的集合, 其属性特征包括借款类别、借款数量、借款期限、借款利率、信用等级等内容; $C=\{L,B\}$ 表示 P2P 借贷平台贷款人(即投资者) L 和借款人 B 的集合, 其属性特征包括成员角色、关键字、债务收入比、所在城市、受教育程度等内容; $E=\{E_1,E_2,E_3\}$ 表示借款人、借贷项目、贷款人之间的 P2P 平台参与人(主要指借款人和贷款人)之间的直接朋友关系 E_1 、借款人-项目之间的从属关系 E_2 、贷款人-项目之间的投资关系 E_3 的集合.

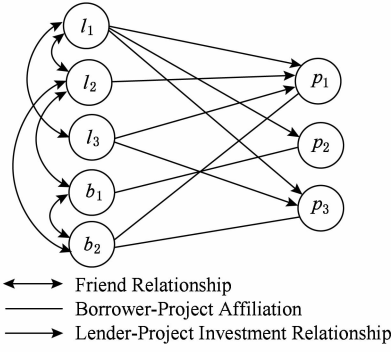


Fig. 1 Example of P2P relationship network

图 1 P2P 关联网络示例

如图 1 示例所示, $l_1, l_2, l_3 \in L, b_1, b_2 \in B, p_1, p_2, p_3 \in P$, 其中, $(l_2, b_1) \in E_1$ 表示贷款人 l_2 与借款人 b_1 之间的直接朋友关系, 对此可用一个变量 R_1 描述 P2P 平台参与人之间是否是直接朋友关系, 如 $R_1(l_2, b_1) = 1$ 表示贷款人 l_2 与借款人 b_1 之间是直接朋友关系; $(l_2, p_1) \in E_2$ 表示贷款人 l_2 对项目 p_1 进行了投资, 即存在贷款人-项目之间的投资关系, 可用一个 3 维向量 $\mathbf{R}_2$ 描述贷款人与项目之间的投资关系特征, 如是否投资、投资数额、投资回报率, 如 $\mathbf{R}_2(l_2, p_1) = (1, 50, 2.2\%)^T$ 表示贷款人 l_2 对项目 p_1 进行了投资、投资额为 50、投资回报率为 2.2%; $(b_2, p_1) \in E_3$ 表示项目 p_1 属于借款人 b_2 , 即存在借款人-项目之间的从属关系, 可用变量 R_3 描述借款人与项目之间的从属关系特征, 如 $R_3(b_1, p_1) = 0$, 则表示项目 p_1 不属于借款人 b_1 .

2.2 基于 P2P 关联网络模型的借贷项目概念模型

在投标期间, 对于 P2P 关联网络的借贷项目集 P 中的任一项目 $p_i (i=1, 2, \dots, n)$, 项目预期回报率 p_i^E 和项目违约风险 p_i^{risk} 是可反映其质量的 2 个最为基本的概念特征; 而项目实际回报率 p_i^{real} 和项目期望回报率 p_i^* 是可反映其质量的 2 个复合概念特征.

其中, 假设 p_i^E 用项目的借款利率表示; 当 p_i 违约, 则 $p_i^{\text{real}} = 0$, 否则 p_i^{real} 为项目的借款利率; 而 p_i^{risk} 难以直接衡量. 可利用历史数据训练违约风险预测模型, 然后根据项目 p_i 的属性特征 (如借款数量、借款利率、借款期限和信用等级) 和与之存在从属关系 E_3 的借款人的属性特征 (如债务收入比) 及相应的已训练的项目违约风险预测模型求得 p_i^{risk} . 这样, 假设项目集 P 中的任一项目 $p_i (i=1, 2, \dots, n)$, 其违约风险预测特征可用向量表示为 $\mathbf{p}_i^{\text{RF}} = (1, p_i^{\text{amt}}, p_i^{\text{rate}}, p_i^{\text{term}}, p_i^{\text{gat}}, p_i^{\text{boi}})^T$, 其中 p_i^{amt} 表示借款数量, p_i^{rate}

表示借款利率, p_i^{term} 表示借款期限, p_i^{gat} 表示信用等级, p_i^{boi} 表示借款人的债务收入比, 我们采用常用的逻辑斯蒂回归方法^[5]进行项目违约风险预测, p_i^{risk} 可表示为

$$p_i^{\text{risk}} = \frac{\exp(\boldsymbol{\beta}_i^T \mathbf{p}_i^{\text{RF}})}{1 + \exp(\boldsymbol{\beta}_i^T \mathbf{p}_i^{\text{RF}})}, \quad (1)$$

其中, $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \dots, \beta_{i(|\mathbf{p}_i^{\text{RF}}|-1)})^T$ 是系数向量, 根据已完成的历史数据和最大似然估计方法可训练得到 $\boldsymbol{\beta}_i$; T 表示向量转置.

p_i^* 反映项目的期望回报率, 与项目的预期回报率和违约风险相关, 即 $p_i^* = p_i^E (1 - p_i^{\text{risk}})$.

这样, 基于 P2P 关联网络模型, 借贷项目 p_i 的概念模型可表示为四元组 $p_i^{\text{CM}} = (p_i^E, p_i^{\text{risk}}, p_i^{\text{real}}, p_i^*)$, 分别表示借贷项目的预期回报率、违约风险、实际回报率和期望回报率.

2.3 基于 P2P 关联网络模型的投资者概念模型

与借贷项目概念模型类似, 对于 P2P 关联网络中投资者集合 L 中的任一投资者 $l_j (j=1, 2, \dots, m)$, 其概念模型可表示为六元组 $l_j^{\text{CM}} = (l_j^E, l_j^{\text{risk}}, l_j^{\text{Prisk}}, l_j^{\text{real}}, l_j^b, l_j^n)$, 分别表示投资者的预期回报率、预期投资风险、历史投资风险、实际回报率、风险厌恶系数和投资能力.

投资者的预期回报率 l_j^E 可定义为投资者 l_j 过去所投资的项目预期回报率的加权平均值, 即 $l_j^E = \frac{|S_j^p|}{\sum_{i=1}^{|S_j^p|} \omega_{ji} p_i^E}$; 投资者的预期投资风险 l_j^{risk} 可定义为投资者 l_j 过去所投资的项目违约风险的加权平

均值, 即 $l_j^{\text{risk}} = \sum_{i=1}^{|S_j^p|} \omega_{ji} p_i^{\text{risk}}$, 可反映投资者在历史投资过程中的投资风险承受能力; 投资者的历史投资

风险为 $l_j^{\text{Prisk}} = \sqrt{\sum_{i=1}^{|S_j^p|} (\omega_{ji})^2 (p_i^{\text{risk}})^2}$; 投资者的实际回报率可定义为投资者 l_j 过去所投资的项目真实

回报率的加权平均值, 即 $l_j^{\text{real}} = \sum_{i=1}^{|S_j^p|} \omega_{ji} p_i^{\text{real}}$; 其中 ω_{ji} 表示投资者 l_j 投资项目 p_i 的资金占其总投资额的比例, S_j^p 表示投资者 l_j 过去所投资的项目集合.

文献[5]指出, 投资者的投资能力与其最终所能获得的实际回报率有关, 获得的实际回报率越大, 其投资能力越强, 因此, 投资者的投资能力 l_j^n 可以表示为 l_j^{real} 的单调递增函数, 即

$$l_j^n = 1/[1 + \exp(-l_j^{\text{real}}/h)],$$

其中, h 为调节参数.

贷款人的风险厌恶系数 l_j^b 难以直接衡量. 但组合投资理论和实践表明, 投资者在不确定市场环境中对风险持谨慎保守态度, 即都是风险厌恶者, 一般假设效用函数为指数函数, 可表示为 $U(x) = 1 - \exp(-bx)$, 其中 b 表示投资者的风险厌恶系数, x 表示投资者投资期望收益率, 假设 x 服从均值为 l_j^E , 标准差为 l_j^{Prisk} 的正态分布. 对于 P2P 关联网络中投资者集合 L 中的任一投资者 l_j , 其预期效用函数可表示为

$$E(U_j) = E(U_j(x)) = 1 - E(\exp(-bx)) = 1 - \exp\left[-b_j l_j^E + \frac{1}{2}(b_j)^2 (l_j^{\text{Prisk}})^2\right]. \quad (2)$$

在预期效用一定的情况下, 式(2)也反映了该投资者在一定风险系数下, 过去投资预期回报率与所承担的历史投资风险之间的关系:

$$l_j^E = \frac{1}{2} b_j (l_j^{\text{Prisk}})^2 - \frac{1}{b_j} \ln[1 - E(U_j)]. \quad (3)$$

将式(3)对 l_j^E 求导, 可得投资者的风险厌恶系数, 可表示为

$$l_j^b = b_j = \frac{\Delta l_j^E}{l_j^{\text{Prisk}} \Delta l_j^{\text{Prisk}}} = \frac{l_j^E - r^*}{(l_j^{\text{Prisk}})^2}, \quad (4)$$

其中, $\Delta l_j^E / \Delta l_j^{\text{Prisk}}$ 表示投资者 l_j 单位风险下的风险溢价, r^* 表示投资者在无风险情况下的收益率, 如购买国家公债或银行储蓄收益率.

由式(4)可知, 投资者的风险厌恶系数与其单位风险溢价成正比, 与投资者历史投资风险成反比.

2.4 基于投资者朋友关系的朋友投标行为影响因子

文献[18]指出投资者进行投资决策时往往更多地受到社会资本因素引起的羊群行为的影响. 文献[19]通过研究发现, 在 P2P 借贷市场中, 由于投资者面临异构偏好或者因获取及分析相关决策信息需要花费大量成本, 会出现羊群行为, 此时朋友投标将对其决策行为产生重要影响. 因此, 投资者 l_j 是否愿意对借贷项目 p_i 投标, 除了与其初始投资偏好有关, 还会受到其朋友(特别是直接朋友)投资行为的影响, 一般投资者的直接朋友投资能力越强, 其投资决策对该投资者的投资行为的影响会越大.

本节基于 P2P 关联网络模型, 分析投资者之间的朋友关系, 发掘用于反映投资者对某一借贷项目的投资偏好受到其直接朋友投标行为的影响程度的相关因素, 称为投资者的朋友投标行为影响因子, 并将其引入个性化投资推荐模型的设计中, 以期能更为准确地把握投资者的当前投资偏好, 提高投资项目推荐的有效性.

基于 P2P 关联网络模型的投资者朋友关系网

络可定义为 $G = (L, \bar{E}_1)$, 其中 $L = \{l_1, l_2, \dots, l_j, \dots, l_m\}$ 为该平台的投资者(即贷款人)集合, 构成投资者朋友关系网络的节点, $\bar{E}_1 \subset E_1$ 为投资者朋友关系网络中任意 2 个节点之间的朋友关系集合. 假设投资者朋友关系网络中的某一投资者 l_j 的直接朋友的集合为 L_j , 即投资者 l_j 与任一投资者 $l_k \in L_j$ 之间存在 $\{l_j, l_k\} \in \bar{E}_1$; 同时, 如果投资者 l_k 与借贷项目 $p_i \in P$ 之间存在 $(l_k, p_i) \in E_2$, 则 $A_{ki} = 1$, 否则 $A_{ki} = 0$. 对于借贷项目 p_i , 投资者 l_j 的朋友投标行为影响因子可定义为

$$l_{ji}^n = \frac{\sum_{k=1}^{|L_j|} l_k^{\text{real}} A_{ki}}{\sum_{k=1}^{|L_j|} A_{ki}}, \quad (5)$$

其中, $|L_j|$ 为投资者 l_j 的直接朋友个数.

3 个性化投资推荐模型设计

首先从基于用户协同过滤(user-based collaborative filtering, UCF)的个性化推荐方法出发, 综合考虑项目投资相似度和项目投资能力^[5], 结合投资者的朋友投标行为影响因子, 确定每一位投资者的推荐投资项目列表; 其次, 考虑投资者风险厌恶系数和历史投资过程中所反映的风险承受能力, 从投资者预期效用最大化出发设计投资推荐优化方案, 确定其投资额在推荐投资项目列表中的分配.

3.1 确定个性化推荐投资项目列表

在传统的 UCF 推荐方法中, 对于借贷项目 p_i , 投资者 l_j 在时刻 t 的投资兴趣大小主要是通过计算其与截至时刻 t 已投资该项目 p_i 的投资者之间的投资相似度来确定:

$$\gamma_t(l_j, p_i) = \frac{\sum_{k \in S_{it}^1} s(l_j, l_k)}{|S_{it}^1|}, \quad (6)$$

其中, S_{it}^1 表示截至时刻 t 已投资该项目 p_i 的所有投资者的集合; $\gamma_t(l_j, p_i)$ 为投资者 l_j 在时刻 t 对项目 p_i 的兴趣度; $s(l_j, l_k)$ 表示投资者 l_j 和投资者 l_k 的相似度:

$$s(l_j, l_k) = \frac{|S_j^p \cap S_k^p|}{|S_j^p \cup S_k^p|}, \quad (7)$$

其中, S_j^p, S_k^p 分别表示投资者 l_j, l_k 所投资项目的集合.

文献[5]综合考虑项目投资相似度和项目投资能力对投资者投资偏好的影响, 将投资者 l_j 对项

目 p_i 的兴趣度表示为

$$\gamma_i(l_j, p_i) = \frac{\sum_{k \in S_{it}^1} s(l_j, l_k) l_j^n}{|S_{it}^1|}. \quad (8)$$

然而,投资者的投资偏好不仅与项目投资相似度和项目投资能力相关,而且在很大程度上还会受到其朋友(特别是有投资能力的直接朋友)投资决策的影响.因此,本文在文献[5]所提出方法的基础上,进一步考虑投资者朋友关系网络中投资者 l_j 的直接朋友的投资能力及其对项目 p_i 的投资情况,分析其对投资者 l_j 的投资偏好的影响,并将投资者 l_j 对投资项目 p_i 的兴趣度调整为

$$\gamma_t(l_j, p_i) = \frac{\sum_{k \in S_{it}^1} s(l_j, l_k) l_j^n}{|S_{it}^1|} (1 + l_{ji}^a). \quad (9)$$

一般认为, $\gamma_t(l_j, p_i)$ 越大,表示在时刻 t 投资者 l_j 对项目 p_i 的兴趣度越大,投资者 l_j 在进行投资项目选择时越有可能选择项目 p_i . 对于每一个投资者 l_j ,可由 $\gamma_t(l_j, p_i)$ 值的大小排序,产生一个粗略的推荐投资项目列表 T_j^1 .

对于当前的投资者较少的借贷项目或新借贷项目带来的推荐“冷启动”现象,可先判断该借贷项目是否为投资者 l_j 的支持组项目^①,如果是,则将该借贷项目添加到推荐投资项目列表 T_j^1 中,并安排在与之最为相似的项目前面,形成推荐投资项目列表 $T_j^{2[5]}$.

3.2 确定推荐投资项目的投资额

经济学中往往将市场参加者的风险偏好分为三大类,即风险厌恶、风险中性和风险爱好.预期效用理论常用于分析在风险条件下“理性人”的决策行为,即在面临不确定性的决策环境时,投资者的决策往往是追求财富的预期效用最大化过程.在现实生活中,绝大多数的投资者是理性的风险厌恶者,但不同的投资者风险厌恶程度有所不同,因此其对应的效用函数中的风险厌恶系数将有所不同.对于一个投资者来说,其风险厌恶系数越大,一个单位的风险为其所带来的效用损失将越大,反之越小.针对不同风险厌恶程度的投资者,文献[20]引入了指数效用函数,用于对投资者行为决策过程进行分析,取得了显著的效果.在此,本文将在指数效用函数假设基础上,结合 2.3 节所分析的投资者风险厌恶系数及其风险承受能力,构建基于预期效用最大化的 P2P 借贷平台的投资组合优化模型.

对于投资者 l_j ,假设其总投资金额为 M_j ,由 3.1 节所确定的推荐投资项目列表为 T_j^2 ;并将投资者 l_j 的总投资金额 M_j 在推荐投资项目列表 T_j^2 中的分配金额列表记为

$$\mathbf{v}_j = (v_{j1}, v_{j2}, \dots, v_{j|T_j^2|}),$$

其中, $\sum_{i=1}^{|T_j^2|} v_{ji} = M_j$.

基于预期效用的 P2P 借贷平台投资优化模型的目标是:在 3.1 节所确定的投资者推荐投资项目列表的基础上,考虑投资者风险厌恶系数和历史投资风险,确定推荐列表中每一个投资项目的投资额,以使得该投资者预期效用最大化:

$$\begin{aligned} \max E(U(x)) &= 1 - E(\exp(-bx)) = \\ \max &\left(\sum_{i=1}^{|v_j|} \frac{v_{ji}}{M_j} p_i^E - \frac{1}{2} b_j \sum_{i=1}^{|v_j|} \left(\frac{v_{ji}}{M_j} \right)^2 (p_i^{\text{risk}})^2 \right). \end{aligned} \quad (10)$$

根据历史数据,通过 2.3 节可计算得出 $b_j = l_j^b$,因此,依据约束条件下的效用最大化原则,可得模型:

$$\begin{aligned} \max E(\mathbf{v}_j) &= \sum_{i=1}^{|v_j|} \frac{v_{ji}}{M_j} p_i^E - \frac{1}{2} l_j^b \sum_{i=1}^{|v_j|} \left(\frac{v_{ji}}{M_j} \right)^2 (p_i^{\text{risk}})^2, \\ \text{s. t.} \quad &\sum_{i=1}^{|v_j|} \frac{v_{ji}}{M_j} p_i^{\text{risk}} = l_j^{\text{risk}}, \sum_{i=1}^{|v_j|} v_{ji} = M_j, \\ &v_{ji} \geq 0, i = 1, 2, \dots, |v_j|. \end{aligned} \quad (11)$$

3.3 投资推荐过程

本节描述基于投资者朋友关系和预期效用最大化的投资推荐过程,如算法 1 所示,包含 2 个输入数据集和 2 个参数.其中, $SetH$ 是训练数据集,由过去的是否违约已知的借贷项目组成; $SetT$ 是测试数据集,在该数据集中包含了当前潜在投资者可投标的借贷项目,这些借贷项目是否违约是未知的; r^* 表示投资者在无风险情况下的收益率,如购买国家公债或银行储蓄收益率,本文假设 $r^* = 0.025$.

算法 1. 基于投资者朋友关系和预期效用最大化的投资推荐.

输入:训练数据集 $SetH$,测试数据集 $SetT$,以及分别包含的借贷项目样本数量 n_1, n_2 ;无风险情况下的投资收益率 r^* ;投资者的投资能力与实际回报率之间的调节参数 h (实验时设为 0.05);

输出:个性化的推荐投资项目列表及其投资金额分配比例 $(T_j^2, \mathbf{v}_j)_{j=1}^L$.

/ * 初始化及模型训练 */

① 对于某一位贷款人,其支持组是指该贷款人特别关注或较为信任的借款人的集合,贷款人的支持组项目指其支持组中成员的借贷项目.

① $\{H_i\}_{i=1}^{n_1}, \{T_i\}_{i=1}^{n_2}; / *$ 对 $SetH$ 和 $SetT$ 进行预处理 $*$ /

② $\{\beta_i\}_{i=1}^{n_1}; / *$ 训练各借贷项目 p_i 的违约风险预测模型 $*$ /

③ $\left\{p_i^{\text{Hrisk}} = \frac{\exp(\beta_i^T p_i^{\text{RF}})}{1 + \exp(\beta_i^T p_i^{\text{RF}})}\right\}_{i=1}^{n_1}; / *$ 预测 $SetH$ 中各借贷项目 p_i 的违约风险 $p_i^{\text{Hrisk}} *$ /

$/ *$ 建立 $SetH$ 和 $SetT$ 中各项目 p_i 的概念模型 $*$ /

④ $\{p_i^{\text{HE}}\}_{i=1}^{n_1}, \{p_i^{\text{TE}}\}_{i=1}^{n_2}; / *$ 分别计算 $SetH$ 和 $SetT$ 中各借贷项目 p_i 的预期回报率 $*$ /

⑤ $\{p_i^{\text{Hreal}}\}_{i=1}^{n_1}, \{p_i^{\text{Treal}}\}_{i=1}^{n_2}; / *$ 分别计算 $SetH$ 和 $SetT$ 中各借贷项目 p_i 的实际回报率 $*$ /

⑥ $\left\{p_i^{\text{Trisk}} = \frac{\exp(\beta_i^T p_i^{\text{RF}})}{1 + \exp(\beta_i^T p_i^{\text{RF}})}\right\}_{i=1}^{n_2}; / *$ 预测 $SetT$ 中各借贷项目 p_i 的违约风险 $p_i^{\text{Trisk}} *$ /

$/ *$ 建立投资者集合 L 中各投资者 l_j 的概念模型 $*$ /

$$\textcircled{7} \left\{l_j^{\text{E}} = \sum_{i=1}^{|S_j^{\text{P}}|} w_{ji} p_i^{\text{HE}}\right\}_{j=1}^{|L|};$$

$$\textcircled{8} \left\{l_j^{\text{real}} = \sum_{i=1}^{|S_j^{\text{P}}|} w_{ji} p_i^{\text{Hreal}}\right\}_{j=1}^{|L|};$$

$$\textcircled{9} \left\{l_j^{\text{n}} = \frac{1}{1 + \exp(-l_j^{\text{real}}/h)}\right\}_{j=1}^{|L|};$$

$$\textcircled{10} \left\{l_{ji}^{\text{a}} = \frac{\sum_{k=1}^{|L_j|} l_k^{\text{real}} A_{ki}}{\sum_{k=1}^{|L_j|} A_{ki}}\right\}_{j=1}^{|L|};$$

$$\textcircled{11} \left\{l_j^{\text{risk}} = \sum_{i=1}^{|S_j^{\text{P}}|} w_{ji} p_i^{\text{Hrisk}}\right\}_{j=1}^{|L|};$$

$$\textcircled{12} \left\{l_j^{\text{b}} = (l_j^{\text{E}} - r^*) / (l_j^{\text{risk}})^2\right\}_{j=1}^{|L|};$$

$$\textcircled{13} \left\{l_j^{\text{Prisk}} = \sum_{i=1}^{|S_j^{\text{P}}|} (w_{ji})^2 (p_i^{\text{Hrisk}})^2\right\}_{j=1}^{|L|};$$

$/ *$ 确定 $SetT$ 中接受投资推荐的各投资者 $l_j (j=1, 2, \dots, m_{\text{ac}})$ 的推荐投资项目列表 $\mathbf{T}_j^2 *$ /

⑭

$$\left\{ \left\{ r_i(l_j, p_i) = \frac{\sum_{l_k \in S_{it}} s(l_j, l_k) l_j^{\text{n}}}{|S_{it}^{\text{L}}|} (1 + l_{ji}^{\text{n}}) \right\}_{i=1}^{n_2} \right\}_{j=1}^{m_{\text{ac}}};$$

$$\text{其中}, s(l_j, l_k) = \frac{|S_j^{\text{P}} \cap S_k^{\text{P}}|}{|S_j^{\text{P}} \cup S_k^{\text{P}}|}.$$

$/ *$ 确定 $SetT$ 中各投资者 $l_j (j=1, 2, \dots, m_{\text{ac}})$ 的推荐投资项目列表 $\mathbf{T}_j^2$ 的投资额分配列表 $\mathbf{v}_j *$ /

⑮

$$\left\{ \begin{array}{l} \max \sum_{i=1}^{|v_j|} \frac{v_{ji}}{M_j} p_i^{\text{TE}} - \frac{1}{2} l_j^{\text{b}} \sum_{i=1}^{|v_j|} \left(\frac{v_{ji}}{M_j} \right)^2 (p_i^{\text{Prisk}})^2 \\ \text{s. t. } \sum_{i=1}^{|v_j|} \frac{v_{ji}}{M_j} p_i^{\text{Trisk}} = l_j^{\text{risk}}, \\ \sum_{i=1}^{|v_j|} v_{ji} = M_j, \\ v_{ji} \geq 0, i=1, 2, \dots, |v_j|. \end{array} \right\}_{j=1}^{m_{\text{ac}}}.$$

基于投资者朋友关系和预期效用最大化的投资推荐过程可描述如下:

步骤 1. 初始化及模型训练过程(行①~③). 首先, 对训练数据集 $SetH$ 和测试数据集 $SetT$ 进行预处理; 其次, 利用 $SetH$ 基于逻辑斯蒂方法训练得到各借贷项目 p_i 的违约风险预测模型的系数向量 β_i ; 最后, 基于训练好的违约风险预测模型预测 $SetH$ 中各借贷项目 p_i 的违约风险 p_i^{Hrisk} .

步骤 2. 建立借贷项目概念模型(行④~⑥). 首先, 计算 $SetH$ 和 $SetT$ 中各借贷项目 p_i 的预期回报率($p_i^{\text{HE}}, p_i^{\text{TE}}$)和实际回报率($p_i^{\text{Hreal}}, p_i^{\text{Treal}}$); 其次, 基于步骤 1 训练得到的违约风险预测模型对测试数据集中各借贷项目 p_i 的违约风险 p_i^{Trisk} 进行预测.

步骤 3. 建立投资者概念模型(行⑦~⑬). 首先, 针对每一位投资者 l_j , 结合历史交易数据及 P2P 关联网信息, 分别计算投资者的投资预期回报率 l_j^{E} 、预期投资风险 l_j^{risk} 、历史投资风险 l_j^{Prisk} 、实际回报率 l_j^{real} 、风险厌恶系数 l_j^{b} 和投资能力 l_j^{n} ; 其次, 针对 $SetT$ 中的借贷项目 p_i , 结合历史交易数据及投资者朋友关系网络信息, 分析投资者 l_j 对项目 p_i 的投资兴趣受其直接朋友投标行为的影响程度, 即基于朋友关系网络计算投资者的朋友投标行为影响因子 l_{ji}^{a} .

步骤 4. 基于投资者朋友关系确定投资者的推荐投资项目列表(行⑭). 首先, 针对每一位投资者 l_j , 结合 $SetT$ 中借贷项目 p_i 的投标情况, 分别计算在时刻 t 投资者 l_j 与项目 p_i 的其他投资者 $l_k (l_k \in S_{it}^{\text{L}})$ 的相似度 $s(l_j, l_k)$, 其中 S_{it}^{L} 为截至时刻 t 已投标了项目 p_i 的所有投资者的集合. 其次, 综合考虑项目投资相似度 $s(l_j, l_k)$ 、项目投资能力 l_j^{n} 和基于朋友关系网络的投资者的朋友投标行为影响因子 l_{ji}^{a} , 针对每一位投资者 l_j , 计算其对 $SetT$ 中借贷项目 p_i 的投资兴趣度; 最后, 根据投资者 l_j 对 $SetT$ 中各借贷项目 p_i 的投资兴趣度及投资者支持组项目, 确定推荐投资项目列表 $\mathbf{T}_j^2$.

步骤 5. 基于投资者风险偏好和预期效用理论, 确定投资者的推荐投资项目的投资额(行⑮). 针对每一位投资者 l_j , 根据步骤 3 所估计的投资者风险厌恶系数 l_j^b 和预期投资风险 l_j^{risk} , 从预期效用最大化出发, 建立投资组合优化模型, 确定投资者 l_j 的总投资额 M_j 在推荐投资项目列表 T_j^2 中的分配金额列表 v_j .

4 实验研究及结果分析

4.1 数据来源

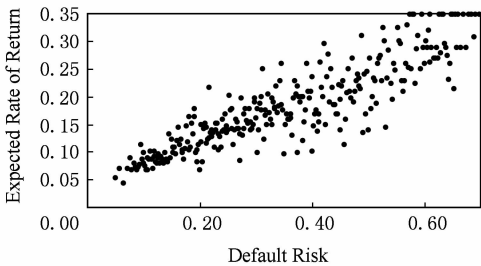
实验数据来源于美国 Prosper 网络借贷平台 2006-01—2012-01 期间的开放数据. 该数据集包含相互关联的 7 类对象信息, 分别是项目类别、群组、注册成员、列表项目、投资项目、投标、角色, 信息量巨大. 其中, 注册成员 1309510 个, 列表项目 371896 个, 投标数为 9638888.

为了便于训练和测试, 首先, 对该数据集进行了初步过滤, 选取已完结且违约状态明确的 25229 个借贷项目及与这些项目相关的列表项目、投资项目、投标和注册成员等信息, 针对每一个借贷项目, 根据列表项目数据、平台投标数据和注册成员数据统计计算其借款人和投资者的相关信息, 同时, 针对每一位投资者, 根据平台投标数据、注册成员数据中的朋友属性列表分别统计其投标信息和直接朋友关联信息, 以此作为构建 P2P 朋友关系网络的基础数据

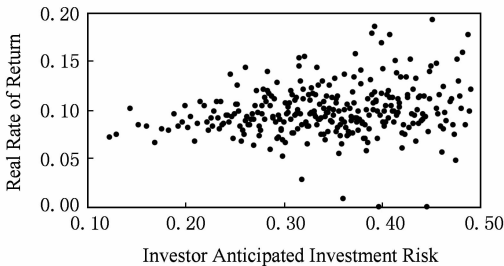
集; 其次, 考虑到数据的完整性, 从基础数据集的 25229 个借贷项目中选取了最终违约状态明确且相关数据完善的 23488 个借贷项目, 针对每一个借贷项目, 计算传统财务指标特征值, 并进行数值化、规范化处理, 以此作为样本数据集; 最后, 将样本数据集按时间划分为 2 个数据集, 其中, 2006-01—2008-12 期间的借贷项目的集合构成训练数据集 $SetH$, 在该数据集中包含了 19156 个借贷项目, 将 2009 年 1 月之后最终违约状态明确的 4332 个借贷项目构成测试数据集 $SetT$. 拟采用文中所设计的投资推荐方法计算对应的投资者对测试数据集 $SetT$ 中的 4332 个借贷项目的投资兴趣度, 确定其推荐投资项目列表及投资额的分配.

4.2 基于 P2P 关联网络模型的概念模型分析

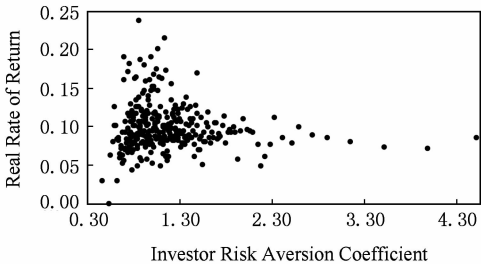
通过随机选取 Prosper 平台上的 300 个借贷项目进行统计分析发现, 借贷项目的预期回报率与其违约风险显著正相关, 如图 2(a)所示, 它们的 Pearson 相关系数为 0.889. 这表明对于潜在的投资者来说, 如果他们想赚得越多, 就需要试着选择高风险的借贷项目. 根据金融经济学理论, 风险厌恶系数越大, 投资者则越厌恶风险, 这对其风险投资的妨碍就越大. 如果投资者风险厌恶系数较大, 投资太过保守, 则其实际投资收益往往较低, 如图 2(c)所示, 当投资者风险厌恶系数大于 2.3, 其实际回报率则低于 0.11. 然而通过对平台投资者的投资预期风险和实际回报率进行分析发现, 并非投资者承担较大的



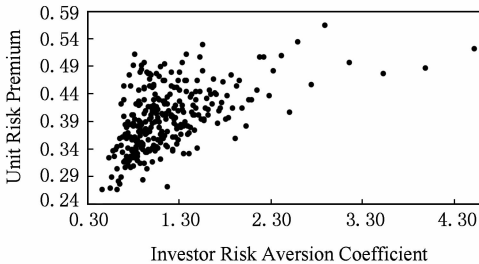
(a) Loan project default risk and expected rate of return



(b) Investor anticipated investment risk and real rate of return



(c) Investor risk aversion coefficient and real rate of return



(d) Investor risk aversion coefficient and unit risk premium

Fig. 2 Conceptual association between lending projects and investors

图 2 借贷项目和投资者相关概念关联

风险就一定能获得更高的投资收益,如图 2(b)所示,很多的投资者所承担的风险很大,但它们的实际回报率却很低.从图 2(c)中也可以发现,对于大多数 P2P 借贷平台的投资者,其风险厌恶系数较低,介于 0.46~2.1 之间,表示愿意投资风险较高的借贷项目,但是他们中仅有部分人实际回报率较高,获得较好的投资收益,究其原因,可能是由于经验不足而冒险选择了单位风险溢价较低的投资项目.

4.3 基于投资者朋友关系的投资推荐的效果分析

本文所提出的基于投资者朋友关系的投资推荐为 FNI_OF 模型,该模型按照 3.1 节的式(9)计算每一位投资者针对测试数据集 *SetT* 中的 4 332 个借贷项目的投资兴趣度,确定其推荐投资项目列表.为了检验 FNI_OF 模型的有效性,以 LR 模型^[21]、UCF 模型^[22]、REC_G 模型^[5](简称为 RGP 模型)作为基准模型,即按照 3.1 节的式(8)计算每一位投资者针对 *SetT* 中的 4 332 个借贷项目的投资兴趣度,确定其推荐投资项目列表;其中,LR 模型是依据文献[21]所用的逻辑斯蒂回归方法对 *SetT* 中的 4 332 个借贷项目进行违约风险评估,并在此基础上

确定全局推荐投资项目列表.为了消除由于推荐投资项目数量的不同选择及随机取样所带来的偏差,保证验证结果的可信度,我们设置了不同的推荐投资项目数量,并分别计算每一位投资者的相应指标值,然后再求平均值.

首先,为了检验 FNI_OF 模型所确定的推荐投资项目列表与投资者实际投资决策的切合程度,本节采用投资推荐中较为常用的 *precision*,*recall*,*F-measure* 指标对上述 2 种方法的推荐结果进行对比,实验结果如图 3 所示.根据图 3(a)、图 3(b)和图 3(c)不难发现,随着推荐投资项目数量 *k* 的增加,4 种模型的推荐准确率(即 *precision_LR*,*precision_UCF*,*precision_RGP*,*precision_FNI*)都随之下降,但 4 种模型的推荐召回率(即 *recall_LR*,*recall_UCF*,*recall_RGP*,*recall_FNI*)和 *F-measure* 指标(即 *F_LR*,*F_UCF*,*F_RGP*,*F_FNI*)都随之上升;与 LR,UCF,RGP 基准模型相比,考虑投资者朋友关系的 FNI_OF 模型的各项指标都要更优,这表明本文提出模型所确定的推荐投资项目列表更能符合投资者的投标兴趣和偏好,从而有助于提高推荐的有效性.

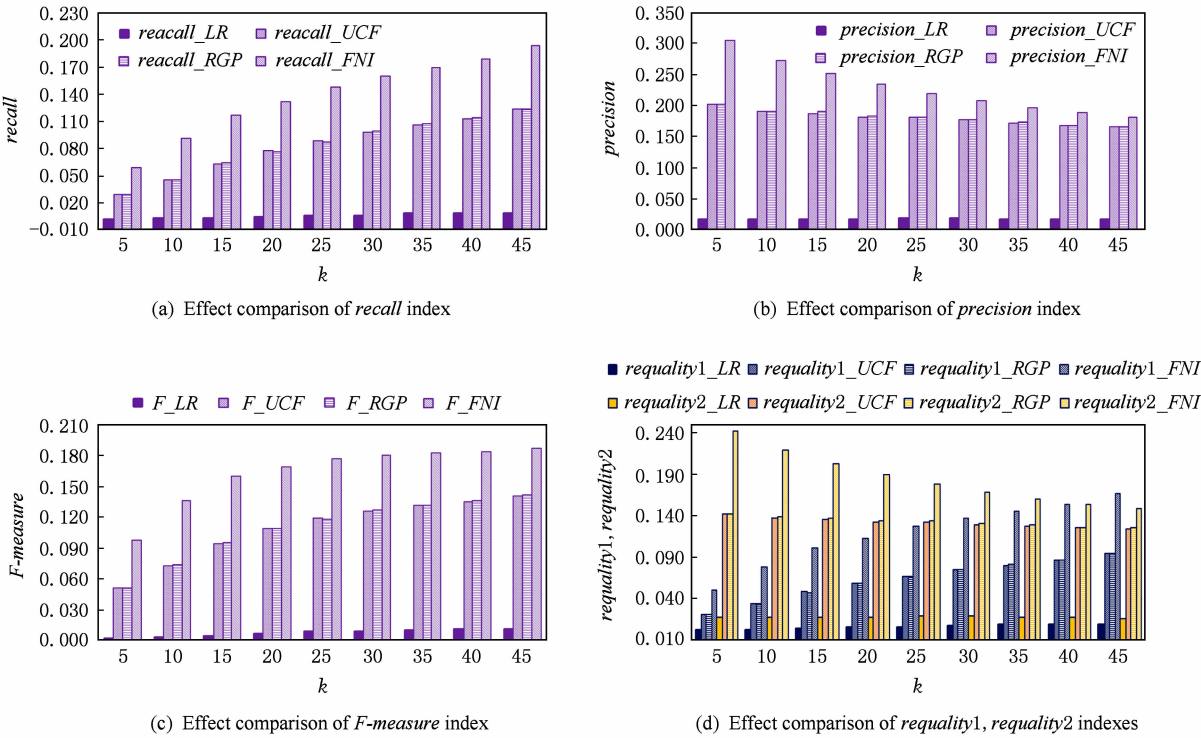


Fig. 3 Comparison about investment recommendation effect

图 3 投资推荐的效果比较

其次,为了检验推荐模型所推荐结果的投资决策效果,即向投资者推荐符合其投资需求的非违约项目的的能力,本节提出 *requality1* 指标、*requality2*

指标进行实验评估.一般认为,模型所确定的推荐投资项目列表中非违约的项目越多,该模型越有助于投资者规避风险;模型所确定的推荐投资项目列表

中投资者最终投标且没有违约的项目越多,该模型越有助于投资者正确地选择满足自身投资需求的项目。其中, *requality1* 指标用于反映推荐模型所确定的推荐投资项目列表中投资者最终投标且没有违约的项目数占其投资项目数的比例; *requality2* 指标反映投资者最终投标且没有违约的项目数占推荐投资项目数的比例。

4 种模型针对 *requality1*, *requality2* 指标的实验评估结果如图 3(d) 所示。根据图 3(d) 可以发现, 与 LR, UCF, RGP 基准模型相比, 对于不同的推荐投资项目数量 k , 考虑投资者朋友关系的 FNI_OF 模型的 *requality1* (即 *requality1_FNI*)、*requality2* (即 *requality2_FNI*) 指标都要更优, 这表明本文提出模型所确定的推荐投资项目列表在提高推荐性能的同时, 还有助于投资者规避投资决策风险, 保证投资决策的有效性。

4.4 投资组合推荐的经济性能分析

本节以文献[5]中所提出的 REC_G 模型(简称为 RGP 模型)作为基准模型, 该模型首先通过 3.1 节的式(8)计算每一位投资者(即贷款人)针对测试数据集 *SetT* 中的 4 332 个借贷项目的投资兴趣度, 确定其推荐投资项目列表, 并在此基础上利用传统的投资组合推荐方法进行投资额的分配; FNI_OF 模型是在 RGP 模型的基础上进一步考虑投资者朋友关系的投资组合推荐方法, 即该方法首先利用 3.1 节的式(9)计算每一位投资者针对 *SetT* 中的 4 332 个借贷项目的投资兴趣度, 确定其推荐投资项目列表, 并在此基础上利用传统的投资组合推荐方法进行投资额的分配; FNI 模型则是在 FNI_OF 模型确定的推荐投资项目列表的基础上, 进一步考虑投资者预期效用最大化进行投资额的分配。针对以上 3 种模型, 本节采用 5 个指标对各种模型的推荐经济效果进行对比分析, 以便对本文所提出的 FNI 模型的推荐经济效果进行全面评价。

第 1 个指标是投资推荐的真实回报率, 这是衡量投资组合推荐经济效果的一个重要指标。

在投资组合推荐过程中, 如果投资者在推荐结果中的实际回报率与其预期回报率之间的差异越小, 同时推荐的准确率和召回率越高, 投资者投资的经济效益将越好。第 2 个指标 PAQ (precision and return quality) 和第 3 个指标 RAQ (recall and return quality) 是在文献[5]所提出相关指标的基础上改进得到, 目的是考察投资者在推荐结果中的实际回报率与其预期回报率之间的差异对其投资满意度及经济效益的影响:

$$PAQ(j) = precision(j)(CP_j^{real} - CP_j^E), \quad (12)$$

$$RAQ(j) = recall(j)(CP_j^{real} - CP_j^E), \quad (13)$$

其中, CP_j^{real} 是投资者 l_j 在推荐结果中的实际回报率, 即 $CP_j^{real} = \sum_{i=1}^{|v_j|} \omega_{ji} p_i^{real}$, 其中, ω_{ji} 表示投资者 l_j 投资项目 p_i 的资金占其总投资额的比例; CP_j^E 是投资者 l_j 在推荐结果中的期望回报率, 即 $CP_j^E = \sum_{i=1}^{|v_j|} \omega_{ji} p_i^*$; *precision*(j) 和 *recall*(j) 分别是面向投资者 l_j 的推荐投资项目列表的推荐准确率和召回率。

在投资组合推荐过程中, 如果投资者承担单位风险所获投资实际回报率与其预期回报率之间的差异越小, 同时推荐的准确率和召回率越高, 投资者决策的经济效用将越好。因此, 除了以上介绍的 3 个指标之外, 本文提出了第 4 个指标 RPAQ (under the unit risk precision and return quality) 和第 5 个指标 RRAQ (under the unit risk recall and return quality), 目的是考察投资者承担单位风险所获投资实际回报率与其预期回报率之间的差异对其决策满意度及经济效用的影响:

$$RPAQ(j) = precision(j)(RP_j^{real} - RP_j^E), \quad (14)$$

$$RRAQ(j) = recall(j)(RP_j^{real} - RP_j^E), \quad (15)$$

其中, RP_j^{real} 是投资者 l_j 在推荐结果中承担单位风险的实际回报率, RP_j^E 是投资者 l_j 在推荐结果中承担单位风险的期望回报率。

第 4 个指标 RPAQ 定义为投资推荐准确率乘以投资者承担单位风险所获投资实际回报率与其预期回报率之间的差异。第 5 个指标 RRAQ 定义为投资推荐召回率乘以投资者承担单位风险所获投资实际回报率与其预期回报率之间的差异。这 2 个指标值越大, 表明投资者的决策满意度越高、经济效用越大。

为了消除由于推荐投资项目数量的不同选择以及随机取样所带来的偏差, 保证验证结果的可信度, 我们设置了不同的推荐候选投资项目数量, 并分别计算每一位投资者的相应指标值, 然后再求平均值, 最终结果如图 4 和图 5 所示。

首先, 分析各模型的推荐结果为投资者带来的实际回报率。在图 4 中, *real_RGP*, *real_FNI_OF*, *real_FNI* 分别表示 RGP, FNI_OF, FNI 这 3 种模型的推荐结果为投资者带来的实际回报率。根据图 4 可以发现, 在本文实验数据集中, 对于不同的推荐投资项目数量 k , 3 种模型的推荐结果为投资者带来的实际回报率都要高于投资者自行投资决策所获得的

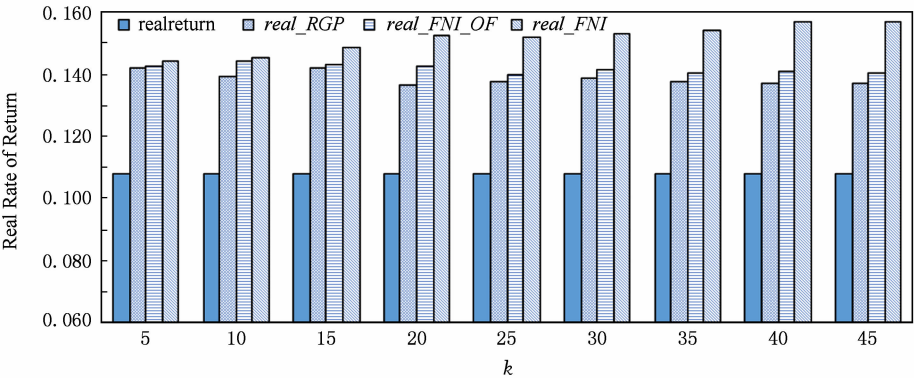


Fig. 4 Actual return rate of investors under different circumstances

图 4 不同情况下投资者实际回报率的比较

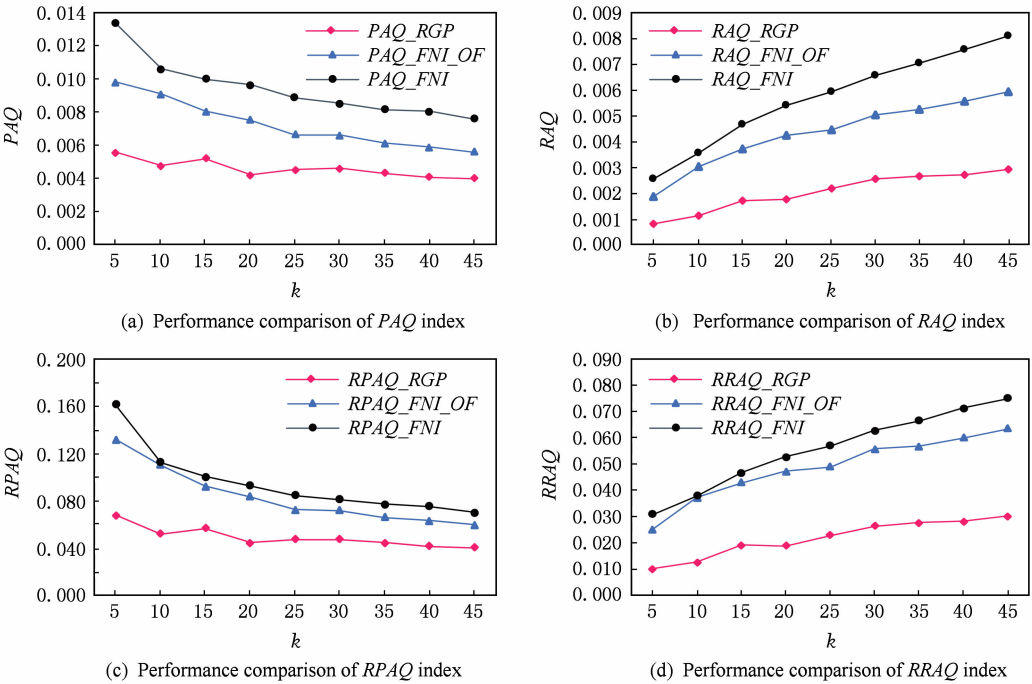


Fig. 5 Recommendation economic performance of different models

图 5 不同模型推荐的经济性能比较

实际回报率. 而与 RGP 基准模型相比, FNI 模型的推荐结果为投资者带来的实际回报率最高, 其次是 FNI_OF 模型. 这表明本文所提出的基于投资者朋友关系和预期效用最大化的投资组合推荐方法能有效地提高投资者的实际收益率, 使其投资组合推荐的经济效果更好.

其次, 评价各模型的投资组合推荐的质量及经济意义. 在图 5(a) 和图 5(b) 分别对比了在不同的推荐投资项目数量的情况下, 各模型的 PAQ 指标和 RAQ 指标的性能. 分析发现, FNI 模型的投资组合推荐的 PAQ 指标和 RAQ 指标的性能都是最好的, 其次是 FNI_OF 模型. 这表明在投资推荐过程中, 考虑投资者朋友关系对其投资偏好的影响, 同时从

投资者预期效用最大化出发建立投资组合优化模型, 可有效地提高模型的投资组合推荐质量, 使其推荐结果更具经济意义.

最后, 评价各模型的投资组合推荐为投资者带来的决策满意度及经济效用. 在图 5(c) 和图 5(d) 分别对比了在不同的推荐投资项目数量的情况下, 各模型的 RPAQ 指标和 RRAQ 指标的性能. 分析发现, FNI 模型的投资组合推荐的 RPAQ 指标和 RRAQ 指标的性能都是最好的, 其次是 FNI_OF 模型. 这表明在投资组合推荐过程中, 考虑投资者朋友关系对其投资偏好的影响, 同时从投资者预期效用最大化出发建立投资组合优化模型, 可有效提高投资者的决策满意度, 使其推荐结果更具经济效用.

5 总结与展望

根据美国 Prosper 市场中的交易数据,在分析 P2P 借贷市场中各对象之间的相互关联关系的基础上构建了 P2P 关联网络模型,并基于 P2P 关联网络模型分别计算借贷项目和投资者的概念特征,得出相应的概念模型;基于多维社会资本理论,在前人研究的基础上进一步考察了 P2P 关联网络模型中的投资者朋友关系网络,分析了投资者的直接朋友投标行为对其投资偏好的影响;在计算投资者对借贷项目的投资兴趣度时,除了考察投资者与借贷项目已投标者的平均投资相似度和借贷项目已投标者的投资能力对投资者投资偏好的影响之外,同时还引入了基于朋友关系的投资者的朋友投标行为影响因子,并根据该影响因子调整投资者对借贷项目的投资兴趣度,结合投资者支持组中的新项目信息,以确定面向投资者的个性化的推荐投资项目列表;在指数效用函数假设基础上,考察投资者风险厌恶系数及历史风险承受能力,构建基于预期效用最大化的 P2P 借贷平台投资组合优化模型,确定投资者的总投资额在个性化的推荐投资项目列表中的分配,并对基于投资者朋友关系和预期效用最大化的投资组合推荐算法进行了整体设计;提出了新的评价指标,以评价投资推荐模型的推荐效果以及投资决策满意度、经济效益和效用。

实验结果表明:在投资推荐效果方面,与 LR 模型、UCF 模型和 RGP 模型等基准模型相比,本文提出的考虑投资者朋友关系的 FNI_OF 模型,不仅在考察推荐情况与投资者实际投资决策的切合程度的传统指标 *precision*, *recall*, *F-measure* 上都要更优,而且在本文提出的考察推荐模型向投资者推荐符合其投资需求的非违约项目能力的 *requality1*, *requality2* 指标上也都要更优,这表明考虑投资者朋友关系的投资推荐方法更能反映投资者的真实投标偏好和投资需求,从而更有助于提高推荐的有效性;在投资组合推荐的经济性能方面,与 RGP 基准模型相比,考虑投资者朋友关系的投资组合推荐的 FNI 模型的投资者实际回报率以及 *PAQ*, *RAQ*, *RPAQ*, *RRAQ* 指标都要更优,这表明基于投资者朋友关系和预期效用最大化的投资组合推荐方法在优化模型投资组合推荐效果的同时,还能有效地提高投资者的实际投资收益和投资决策满意度,即其推荐的结果更具经济效益和效用。

另外,本文虽然是基于美国 Prosper 市场交易数据进行研究,但该方法具有一般的借鉴意义。P2P 作为互联网金融的主流模式之一,都存在平台信息不完全、“硬”信息质量不高、总体安全性低、平均风险性高的共性,因此,针对不同的 P2P 社交网络模式,只要适当地调整对应的 P2P 社交网络关联模型及概念特征指标的计算公式,则可将其应用于其他的 P2P 借贷市场中的投资项目投资组合推荐,以提升推荐的效果及推荐的经济性能。

在未来的工作中,我们打算进一步分析非 P2P 平台的社会资本信息,以便提取更多的影响投资者行为的社交网络特征指标,并将之应用于 P2P 平台的个性化投资组合推荐中,结合算法的改进,进一步提高投资推荐的效果及经济性能,为投资决策提供更好的服务。

参 考 文 献

- [1] Klafft M. Online peer-to-peer lending: A lenders' perspective [J]. SSRN Electronic Journal, 2008, 2(2): 371-375
- [2] Rosenberg E, Gleit A. Quantitative methods in credit management: A survey [J]. Operations Research, 1994, 42(4): 589-613
- [3] Tao Qing, Wang Jue. A new fuzzy support vector machine based on the weighted margin [J]. Neural Processing Letters, 2004, 20(3): 139-150
- [4] Guo Yanhong, Zhou Wenjun, Luo Chunyu, et al. Instance-based credit risk assessment for investment decisions in P2P lending [J]. European Journal of Operational Research, 2015, 249(2): 417-426
- [5] Zhao Hongke, Wu Le, Liu Qi, et al. Investment recommendation in P2P lending: A portfolio perspective with risk management [C] //Proc of the 15th IEEE Int Conf on Data Mining. Piscataway, NJ: IEEE, 2015: 1109-1114
- [6] Liu De, Brass D J, Lu Yong, et al. Friendships in online peer-to-peer lending: Pipes, prisms, and relational herding [J]. MIS Quarterly, 2015, 39(3): 729-742
- [7] Ceyhan S, Shi Xiaolin, Leskovec J. Dynamics of bidding in a P2P lending service: Effects of herding and predicting loan success [C] //Proc of the 20th Int Conf on World Wide Web. New York: ACM, 2011: 547-556
- [8] Serranocinca C, Gutiérreznieta B, Lópezpalacios L. Determinants of default in P2P lending [J]. Plos One, 2015, 10(10): e0139427
- [9] Liao Li, Li Mengran, Wang Zhengwei. The intelligent investor: Not-fully-marketized interest rate and risk identify—Evidence from P2P lending [J]. Economic Research Journal, 2014, (7): 125-137(in Chinese)

- (廖理, 李梦然, 王正位. 聪明的投资者: 非完全市场化利率与风险识别——来自 P2P 网络借贷的证据[J]. 经济研究, 2014, (7): 125-137)
- [10] Pope D G, Sydnor J R. What's in a picture? Evidence of discrimination from Prosper. com [J]. Journal of Human Resources, 2011, 46(1): 53-92
- [11] Chen Dongyu, Han Chaodong. A comparative study of online P2P lending in the USA and China [J]. Journal of Internet Banking & Commerce, 2012, 17(2): 1-15
- [12] Everett C R. Group membership, relationship banking and loan default risk: The case of online social lending [J/OL]. Banking & Finance Review, 2015 [2017-08-01]. <http://dx.doi.org/10.2139/ssrn.1114428>
- [13] Lin Mingfeng, 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
- [14] Zhao Hongke, Liu Qi, Wang Guifeng, et al. Portfolio selections in P2P lending: A multi-objective perspective [C] //Proc of the 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 2075-2084
- [15] 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
- [16] Zhu Mengying, Zheng Xiaolin, Wang Chaohui. Investment recommendation based on risk and surplus in P2P lending [J]. Journal of Computer Research and Development, 2016, 52(12): 2708-2720(in Chinese)
- (朱梦莹, 郑小林, 王朝晖. 基于风险和剩余价值的在线 P2P 借贷投资推荐方法[J]. 计算机研究与发展, 2016, 52(12): 2708-2720)
- [17] Quijano-Sanchez L, Recio-Garcia J A, Diaz-Agudo B, et al. Social factors in group recommender systems [J]. ACM Trans on Intelligent Systems and Technology, 2013, 4(1): No. 8
- [18] Shen Dawei, Krumme C, Lippman A. Follow the profit or the herd? Exploring social effects in peer-to-peer lending [C] //Proc of the 2nd IEEE Int Conf on Social Computing. Piscataway, NJ: IEEE, 2010: 137-144
- [19] Luo Binjie, Lin Zhangxi. A decision tree model for herd behavior and empirical evidence from the online P2P lending market [J]. Information Systems & e-Business Management, 2013, 11(1): 141-160
- [20] Çanakoglu E, Özekici S. Portfolio selection in stochastic markets with exponential utility functions [J]. Annals of Operations Research, 2009, 166(1): 281-297

- [21] Dong Gang, Lai Kin Keung, Yen Jerome. Credit scorecard based on logistic regression with random coefficients [J]. Procedia Computer Science, 2010, 1(1): 2463-2468
- [22] Resnick P, Iacovou N, Suchak M, et al. Group Lens: An open architecture for collaborative filtering of netnews [C] //Proc of ACM Conf on Computer Supported Cooperative Work. New York: ACM, 1994: 175-186



Wan Changxuan, born in 1962. Professor and PhD supervisor at Jiangxi University of Finance and Economics. Senior member of CCF. His main research interests include Web data management, sentiment analysis, data mining and information retrieval, etc.



You Yun, born in 1979. PhD candidate. Lecturer at East China University of Technology. Her main research interests include big data analysis and P2P investment decision-making, etc(youyun_rose@163.com).



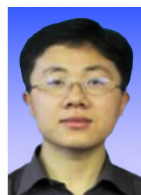
Jiang Tengjiao, born in 1976. PhD, lecturer. Her main research interests include sentiment analysis, data mining and Web data management, etc(tj_jiang@163.com).



Liu Xiping, born in 1981. PhD, associate professor. Member of CCF. His main research interests include information retrieval, big data analysis and Web data management, etc(lewislxp@gmail.com).



Liao Guoqiong, born in 1969. Professor and PhD supervisor at Jiangxi University of Finance and Economics. Senior member of CCF. His main research interests include recommendation in social networks, data mining and mobile computing, etc(liaoguoqiong@163.com).



Liu Dexi, born in 1975. Professor and PhD supervisor at Jiangxi University of Finance and Economics. Senior member of CCF. His main research interests include information retrieval and natural language processing, etc(dexi.liu@163.com).