

IDEA: 一种基于 P2P 借贷网络的投资决策分析算法^①

周雅慧, 张一舟, 米晋宏

(复旦大学 计算机科学技术学院 上海市数据科学重点实验室, 上海 201203)

(上海财经大学 上海市金融信息技术研究重点实验室, 上海 200433)

摘要: 互联网金融 P2P 借贷平台上存在着较大的贷款投资风险, 为协助投资人获得更佳的贷款收益, 本文综合考虑贷款坏账风险、流标风险、利率和投资人风险偏好等要素, 提出投资决策算法 IDEA(Investment DEcision Analysis): 构建投资人-贷款网络, 充分利用网络中的贷款投资行为信息来度量贷款坏账风险, 利用图半监督学习方法度量贷款流标风险, 为投资决策提供依据. 在真实数据集上的实验结果表明, 相对于现有算法, 我们的算法不仅可以取得更佳的投资收益, 而且能够协助具有不同风险偏好的投资人进行投资决策.

关键词: P2P 投资决策; 图半监督学习; 数据挖掘; 风险评估

IDEA: An Investment Decision Analysis Algorithm for P2P Lending

ZHOU Ya-Hui, ZHANG Yi-Zhou, MI Jin-Hong

(Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University, Shanghai 201203, China)

(Shanghai Key Laboratory of Financial Information Technology, Shanghai University of Finance and Economics, Shanghai 200433, China)

Abstract: There has been high risk on internet financial P2P lending platform. In order to assist investors to obtain a better loan proceeds, this paper comprehensively takes the risk of bad debts, failure of bidding, interest rate and investor risk preferences and other factors into consideration, and proposes the investment decision-making algorithm IDEA (investment Decision Analysis): building the investor-loan network, taking advantage of investment behavior of loan information in the network to measure the risk of bad debts and make use of the graph based semi-supervised learning method to metric the risk of failed bidding to provide the basis for investment decisions. Experiment results on real datasets show that, compared with existing algorithms, our algorithm can not only get a better return on investment, but also assist investors with different risk appetite to make investment decisions.

Key words: P2P investment decisions; graph-based semi-supervised learning; data mining; risk evaluation

1 引言

随着互联网金融的普及, P2P(peer to peer)借贷模式得到快速发展^[1], 其操作方式是借款人发布一笔贷款, 由多名投资人合资借出. 若贷款没在平台规定时间(筹款期)内筹集到全部资金则贷款流标, 反之贷款成功. 若贷款最终无法还清, 则贷款成为坏账. 由于 P2P 贷款利率较高, 吸引了很多人进行投资. 然而伴随高利率的同时, 投资人也面临着较大的投资风险. 如何协助投资人在此平台上进行投资决策是一个极具

意义的研究方向.

进行 P2P 贷款投资决策研究的关键在于对贷款存在的风险进行精准度量. 现有的研究大都对贷款的坏账风险和借款人的信用状况进行度量, 并以此为依据, 帮助投资人进行决策. 常用的技术包括逻辑回归, 贝叶斯网络, 神经网络等. 例如: Xubo W 基于贝叶斯网络进行贷款坏账风险预测, 选择出低坏账风险的贷款^[2]. Dong G 提出利用逻辑回归预测借款人信用情况, 选择出优质借款人^[3]. Guo Y 通过回归方法学习坏账的属性

^① 基金项目:国家自然科学基金(91546105,71331005,7150316)

收稿时间:2016-03-01;收到修改稿时间:2016-03-28 [doi:10.15888/j.cnki.csa.005412]

权重,提出一种基于实例的坏账风险评估方法^[4]. Chunyu L基于贷款已有投资人的历史投资表现作为依据对贷款风险进行评估,发现跟着投资经验丰富、拥有较高收益的投资者投标,收益可提高^[5]. 本文对现有投资决策研究进行了总结和分析,发现这些研究存在如下两点问题:

一是现有的投资决策研究在对贷款的风险进行研究时,只考虑了贷款的坏账风险. 在真实的 P2P 平台上,投资人在进行投资决策时还需要面对较大的贷款流标风险. 例如,知名 P2P 平台——拍拍贷,自 2007 年成立至今,每年平台上被流标的贷款占 75% 以上. 贷款流标使得投资人不但浪费了投资时间,而且无法获得任何收益. 文献[6]指出如何量化流标风险概率是一项极具挑战和有意义的研究方向^[6]. 然而现有研究主要集中于对贷款流标风险的影响因素进行定性研究^[7-9],并没有将其考虑入贷款投资决策模型中.

二是现有的投资决策算法在对贷款的风险进行度量时,大多数仅从贷款及借款人本身的属性出发,忽略了平台上另一类重要参与者——投资人的投资行为信息. 在 P2P 平台上,每个投资人都具有自己的投资风险偏好,投资人的投资行为反映了这种偏好,一笔贷款的风险可以由其投资人的投资风险偏好进行量化^[10]. 因此,本文通过引入投资人的投资行为信息,帮助我们提升贷款风险的度量准确率.

综上,本文综合考虑贷款坏账风险、流标风险、利率等要素,提出了基于 P2P 借贷网络的投资决策算法——IDEA. 将投资人的贷款投资行为映射为投资人-贷款网络,充分利用网络中的投资行为信息来度量贷款坏账风险,利用图半监督学习方法度量贷款流标风险,并在此基础上提出贷款期望投资收益率公式,计算每笔贷款的期望收益以生成 topN 贷款候选集,协助投资人进行决策.

2 投资决策算法IDEA

本章首先在 2.1 节构建投资人-贷款网络,基于该借贷网络,我们在 2.2 节和 2.3 节分别提出本文的流标风险度量模型和坏账风险度量模型. 之后综合考虑贷款的流标和坏账风险及投资人风险偏好,提出了本文的期望投资收益率公式. 并基于该公式在 2.4 节给出本文的贷款投资决策算法——IDEA.

2.1 投资人-贷款网络

在这一小节,我们介绍将投资人投资贷款的行为映射为投资人-贷款网络. 通过挖掘网络中投资人的风险偏好,有助于衡量贷款风险:我们发现被同一投资人投资的贷款之间往往具有相近的利率和风险,故可以利用投资人的历史投资行为来帮助更好的评估其风险偏好. 另一方面,对同一笔贷款,我们也可以利用该贷款所有投资人的风险偏好,来衡量该笔贷款的风险.

下面我们给出投资人-贷款网络的形式化定义:

定义 1(投资人-贷款网络). 构建投资人-贷款网络 $G = \{I, L, E\}$. 其中 $I = \{I_1, I_2, \dots, I_m\}$ 表示投资人集合. $L = \{l_1, l_2, \dots, l_n\}$ 表示贷款集合, $E = \{e_{11}, e_{12}, e_{13}, \dots, e_{ij}, \dots, e_{mn}\}$ 表示投资人与贷款之间的投资关系,其中 e_{ij} 对应第 i 个投资人对第 j 笔贷款的投资金额.

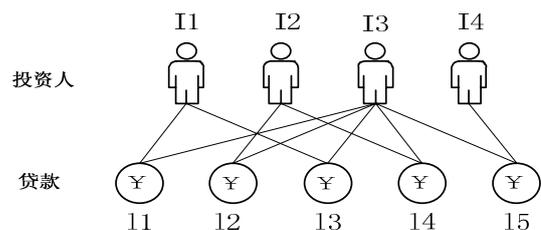


图1 投资人-贷款网络 G

如图 1 所示,投资人和贷款构成二部图网络. 在网络 G 中,投资人 I_1 投资了贷款 l_1 和 l_3 , 则 l_1 和 l_3 更可能具有相近的利率和风险,可通过 l_1 和 l_3 度量投资人 I_1 的风险偏好. 而贷款 l_3 的风险可通过其投资人 I_1 和 I_3 的投资风险偏好进行度量. 之后我们将给出基于此网络的贷款流标风险模型和坏账风险模型

2.2 贷款流标风险模型

贷款流标风险是指一笔贷款无法在规定时间内筹集到所需资金的风险. 若该笔贷款未能在筹款期内筹集到足够资金,则该笔贷款被判为流标,对投资人而言,其在该笔贷款的筹款期内无法获得任何收益,浪费了投资时间成本.

现有的流标风险研究主要利用贷款的属性进行评估,通常假设贷款间相互无影响,即满足独立同分布假设. 常用方法如:线性回归、神经网络等. 其基本研究思路可表示为公式(1).

$$P(Y|X) = \prod_{j=1}^n P(y_j|x_j) \tag{1}$$

其中 y_j 为第 j 笔贷款的流标风险标识, $y_j \in \{0,1\}$ (满标 $y_j=1$, 流标 $y_j=0$). $Y=\{y_1,y_2,\dots,y_n\}$, Y 表示所有贷款的流标标识向量. $X=\{x_1,x_2,\dots,x_n\}$, X 表示所有贷款相关属性的集合.

如前所述, 在贷款流标风险研究中, 通过引入投资人-贷款网络, 我们可以利用贷款间的相似性(体现在贷款通过相同投资人形成的连接关系), 来提高流标风险预测的准确性. 假设 N_j 表示与贷款 l_j 存在相似性的贷款集合, 我们使用 $Y_{N_j} = \{y_i | l_i \in N_j\}$ 表示所有属于 N_j 的贷款的流标标识向量, 我们通过 Y_{N_j} 可以获得贷款 l_j 的网络属性, 则上述研究思路可表示为公式 (2).

$$P(Y|X) = \prod_{j=1}^n P(y_j|Y_{N_j}, x_j) \tag{2}$$

基于公式(2)的研究思路, 本文提出一种基于借贷网络的图半监督学习^[11,12]方法 PNICA(P2P Lending Network based Iterative Classification). 通过已有流标标识的 L_{train} 和待预测的 L_{test} 中的贷款属性信息和贷款间的相关关系来预测 L_{test} 中贷款的流标风险, 基本步骤如下所示, 伪代码如方法 1 所示.

首先, 我们利用贷款之间的共同投资人将投资人-贷款网络 G 转换为贷款网络 G' , 图 1 转化为如图 2 所示, 方法为若两笔贷款被同一个人投资过, 则认为这两笔贷款存在相似性, 即两笔贷款之间产生连接.

其次, 我们对 L_{train} 中的贷款属性 x_j 进行扩展: 通过方法 $Aggregation(l_j, Y)$ 获得贷款 l_j 的网络属性值 X_{N_j} (见定义 2), 并加入到 x_j 中. 接下来使用扩展后的训练集来训练基分类器 f .

然后, 我们利用训练完成的基分类器 f 为测试集 L_{test} 中每笔贷款的流标风险进行初始化(初始贷款的网络属性值 X_{N_j} 用 0 向量表示).

最后, 我们使用初始化时标记的结果, 为 L_{test} 中每笔贷款通过 $Aggregation$ 方法计算其网络属性值, 由于网络属性的计算依赖于与贷款相关的样本集合 N_j , 而在 N_j 中属于 L_{test} 的贷款每次迭代均可能发生变化, 故每一次预测结果可能会发生变化, 故不断重复该过程, 直到结果收敛, 我们将得到目标预测结果 \hat{Y}_{test} .

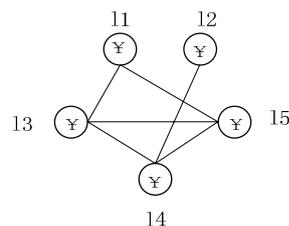


图 2 贷款网络

方法 1: 流标风险度量模型 PNICA

输入:

X : 贷款属性集合, f : 基分类器, G : 投资人-贷款网络,
 L_{train} : 历史贷款集, L_{test} : 当前待预测贷款集.
 Y_{train} : 历史贷款集流标标识向量, MAX_It : 最大迭代次数.

输出:

$\hat{Y}_{test} = (\hat{y}_1, \dots, \hat{y}_j, \dots), \forall l_j \in L_{test}$: 测试集预测结果

```

1  D ← {}; i ← 0
2  使用 G 构建贷款网络 G'
3  for l_j in L_train
4      x_j' = (x_j, Aggregation(x_j, Y_train))
5      D.add(x_j', y_j)
6  end for
7  f = f(D) // 训练基分类器
8  for l_j in L_test
9      y_j_hat = f(x_j, 0)
10 end for
11 repeat
12     k ← k + 1
13     for l_j in L_test
14         x_j' = (x_j, Aggregation(l_j, Y_train ∪ {y_j_hat | l_j ∈ L_test}))
15         y_j_hat = f(x_j')
16     end for
17 until y_hat_test values converge or k = MAX_It
18 return y_hat_test

```

定义 2(相邻贷款流标率). 提取网络属性值的 $Aggregation$ 方法, 通常使用聚集函数来完成, 如均值方法 $mean$, 加和方法 sum 等. 利用该方法可以将贷款网络中与贷款 j 相连的 N_j 规范化为固定数值的属性. 此处我们选择均值方法. 并定义为相邻贷款流标率: 表示在贷款网络 G' 中与第 j 笔贷款相连的所有贷款

N_j 中,流标贷款所占的比率.

2.3 贷款坏账风险模型

坏账风险是指贷款没在规定的时间内进行偿还的风险.如同流标风险,现有的坏账风险研究大多数仅考虑贷款属性和借款人属性.然而由于这类信息公开的并不充分,使得获得的模型预测结果不够理想.考虑到贷款风险可以通过这笔贷款的投资人风险偏好进行度量(若一笔贷款是由高风险偏好的投资人投资的,那么一般风险较高,反之,则风险较低),而投资人的风险偏好可以通过其历史投资行为衡量.故我们利用投资人-贷款网络,通过研究投资人历史投资行为,来衡量投资人的历史投资风险表现,以此评估一笔贷款的坏账风险.首先给出下列定义.

定义 3(投资人投资权重 ω_{ij}).表示投资人 I_i 对贷款 l_j 的投资金额占 I_i 历史总投资金额的比例.

定义 4(贷款投资权重 φ_{ij}).表示投资人 I_i 对贷款 l_j 的投资金额占 l_j 总筹资金额的比例.

由上述定义,我们可以通过公式(3)计算投资人的历史投资收益率 \bar{R}_i .其中 R_j 为投资人投资过的每笔贷款的实际收益率.

$$\bar{R}_i = \sum_{j=1}^n \omega_{ij} R_j \quad (3)$$

统计学中通常使用标准差来度量风险,类似的,我们定义投资人的历史投资收益风险表现 PI_i 为实际投资收益率 R_j 的标准差,如公式(4)所示:

$$PI_i = \sqrt{\sum_{j=1}^n \omega_{ij} (R_j - \bar{R}_i)^2} \quad (4)$$

本文将一笔贷款的坏账风险定义为该笔贷款已有投资人的历史投资风险表现的标准差,如公式(5)所示:

$$P_{bad_j} = \sum_{i=1}^m \sum_{k=1}^m \sqrt{J_{ik} \varphi_{ij} PI_i \varphi_{kj} PI_k} \quad (5)$$

其中 J_{ik} 表示投资人 I_i 和 I_k 的 Jaccard 相似度.以下给出贷款坏账风险度量方法 RBE(risk of bad debtsevaluation),如方法 2 所示.

方法 2: 坏账风险度量模型 RBE

输入:

L_{train} :历史贷款集, L_{test} :当前待投资贷款集

输出:

$P_{bad} = (P_{bad_1}, \dots, P_{bad_j}) \forall l_j \in L_{test}$: 贷款坏账风险

1 for I_i in I

```

2    $\omega_{ij} = \frac{e_{ij}}{\sum_{j=1}^n e_{ij}}$ 
3    $\bar{R}_i = \sum_{j=1}^n \omega_{ij} R_j$ 
4    $PI_i = \sqrt{\sum_{j=1}^n \omega_{ij} (R_j - \bar{R}_i)^2}$ 
5   end for
6   for  $I_i$  in  $I$ 
7     for  $I_k$  in  $I$ 
8        $J_{ik} = \frac{D_i \cap D_k}{D_i \cup D_k}$  ( $D_i$  为  $I_i$  投资过的贷款集合)
9     end for
10  end for
11  for  $l_j$  in  $L_{test}$ 
12     $\varphi_{ij} = \frac{e_{ij}}{\sum_{i=1}^m e_{ij}}$ 
13    利用公式(5)计算  $P_{bad_j}$ 
14  end for
15  return  $P_{bad}$ 

```

2.4 贷款投资决策算法

本节综合考虑了贷款流标、坏账风险和投资人的风险偏好等因素提出贷款预期收益计算方法,并提出贷款投资决策算法——IDEA:以贷款预期收益为依据进行决策.

假设贷款的预期流标风险为 P_{fail} ,预期坏账风险为 P_{bad} ,若贷款流标,则投资人实际投资收益率为 0;若贷款成功筹集到足够资金,当贷款正常还款时,投资人实际投资收益率为贷款发布利率 r ;当贷款成为坏账时,投资人实际投资收益率为-1.根据数学期望的计算,我们得到如公式(6)所示的期望投资收益率 ROI(Return On Investment).

$$ROI = 0 \times P_{fail} + (1 - P_{fail}) \times [(1 - P_{bad}) \times r + (-1) \times P_{bad}] \\ = (1 - P_{fail}) \times [(1 - P_{bad}) \times r - P_{bad}] \quad (6)$$

公式(6)计算得到的高期望投资收益率的贷款,即为推荐给投资人的优质贷款.考虑到实际投资决策中,不同投资人拥有不同的投资风险偏好,期望能获得不同的投资建议.基于 2.3 节的工作,我们使用投资人历史投资风险表现 PI_i 表示投资人风险偏好,为计算方便,我们对 PI_i 进行规范化: $PI'_i = 1 - (PI_i - PI_{min}) / (PI_{max} - PI_{min})$.考虑投资人风险偏好后,贷款期望投资收益率 ROI 表达为公式(7).

$$ROI = (1 - P_{fail}) \times [(1 - P_{bad} \times PI') \times r - P_{bad} \times PI'] \quad (7)$$

基于上述公式, 将贷款期望投资收益率作为决策依据, 我们给出本文的贷款投资决策算法——IDEA. 算法 1 描述了 IDEA 算法的计算过程. 算法包含两份输入数据集: (1)历史贷款集 L_{train} , 包括投资人的历史投资行为和贷款的相关属性. (2)当前待投资贷款集当前贷款集 L_{test} , 包括当前投资人投资行为和待被投资的贷款的相关属性. IDEA 算法首先建立投资人-贷款网络, 基于此网络通过 PNICA 模型评估贷款集 L_{train} 中贷款的流标风险 P_{fail} , 通过 RBE 模型评估贷款集 L_{train} 中贷款的坏账风险 P_{bad} . 结合投资人风险偏好. 通过公式(7)计算得到贷款集 L_{train} 中每笔贷款的期望投资收益率 ROI . 并按照 ROI 从大到小进行排序, 输出 $topN$ 笔贷款, 即为贷款候选集.

算法 1 贷款投资决策算法 IDEA

输入:

L :贷款数据集, f :基分类器, X :贷款属性集合,
 MAX_It :最大迭代次数, $topN$:候选贷款数目

输出:

候选 $topN$ 贷款集:Recommendationset

```

1 Recommendationset ← {}
2 使用贷款数据集 L 建立投资人-贷款网络 G
3  $P_{fail} \leftarrow PNICA(L, X, f, G, MAX\_It)$ 
4  $P_{bad} \leftarrow RBE(L)$ 
5 for  $l_j$  in  $L_{test}$ 
6   利用公式(7)计算  $ROI_j$ 
7 end for
8 对  $ROI$  进行降序排序
9 for  $K \leftarrow 1$  to  $topN$ 
10  Recommendationset.add( $ROI, k$ )
11 end for
12 return Recommendationset
    
```

3 实验

本文实验采用 Matlab 编程实现, 实验平台配置 8 台相同配置的 PC 机(通过局域网连接), CPU 为 Intel(R) Xeon(R) CPU E5-2609 v2 @ 2.50GHz, 内存 128G, 操作系统为 CentOS release 6.5.

3.1 实验数据

本文在使用拍拍贷数据集进行实验, 该数据集通过网络爬虫技术从拍拍贷网站上获得. 实验中将 2012 年到 2013 共 98000 多条贷款数据作为历史贷款集 L_{train} , 2014 年 1 月到 2 月共 16000 多条贷款数据作为

当前待投资贷款集 L_{test} . L_{train} 和 L_{test} 包括投资人的投资行为和每笔贷款的相关属性. 考虑到在当前待投资贷款集中, 一笔贷款需有投资人, 才可以利用投资人的历史投资表现进行风险度量. 本文选择在贷款设定的筹集时间 T_{loan} 的前 10% 内的投资人. 通过 L_{train} , 可统计到平台上用户的投资次数(投资经验)分布, 如图 3 所示, 只有少部分投资人拥有较为丰富的历史投资经验. 为了选择有经验的投资人对贷款风险进行度量, 本文选择投资经验在 30 笔及以上的投资人构建网络.

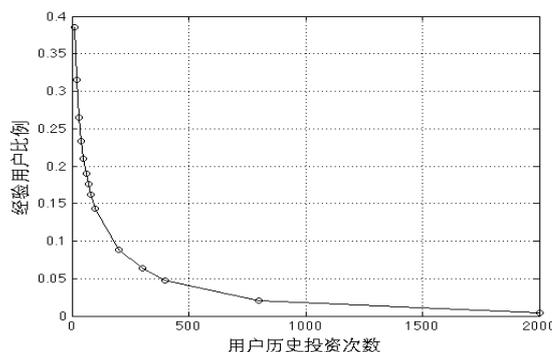


图 3 投资人历史投资经验分布

3.2 实验结果

本文共包含四组实验, 以对提出的算法进行验证, 分别是: (1)IDEA 算法的对比实验; (2)基于 PNICA 方法的流标风险实验; (3)投资人风险偏好实验; (4)基于 RBE 方法的坏账风险实验

3.2.1 IDEA 对比实验结果

在第 2 章介绍了 IDEA 算法, 为验证该算法是否能协助投资人获得更高收益贷款候选集, 在本节将设置两组对照方法, 方法 1 是 REC_ICP 算法^[5]: 使用贷款已有投资人的历史的投资回报表现与投资风险的比值来衡量一笔贷款的投资价值. 方法 2 是通过投资平台上所有贷款所获得的平均收益.

本实验将 IDEA 中投资人风险偏好 PI' 设置为 0.9, 实验结果如图 4 所示. (在 PI' 范围内的任意取值下, 均可以返回令人满意的贷款候选集, 将在实验 3.2.3 节介绍)可以发现不同 $topN$ 下的贷款候选集合中, IDEA 算法均获得比其它的对比算法更高的实际投资收益率, 而不使用任何推荐算法的实际投资收益率仅 2.69%.

REC_ICP 算法理想化了贷款投资情况, 其只选取了中标贷款数据集. 这在实际的决策中是不实际的. 因为在真实的 P2P 平台中, 存在 70% 到 80% 的贷款被

流标. 在 REC_ICP 算法的推荐结果中较多的流标贷款会被推荐出来, 而 IDEA 算法综合衡量了贷款的流标和坏账风险及收益, 因此可以得到比 REC_ICP 更好的实际投资收益率.

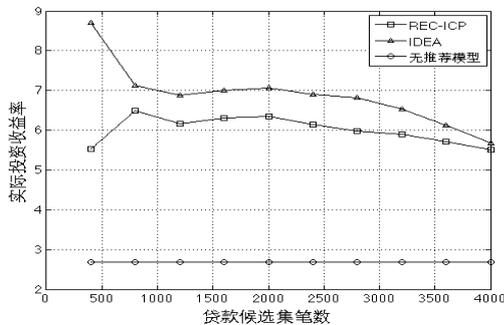


图 4 实际投资收益率对比实验

3.2.2 流标风险实验

在 2.2 节我们提出的 PNICA 方法中, 贷款的相关属性集包括贷款的利率、期限、金额和借款人性别、借款人职业、借款人信用分、借款人年龄. 网络属性为相邻贷款流标率.

本文将属性进行了离散化处理, 包括: 借款人性别(男 1, 女 0); 借款人职业(分为工薪族、网店卖家和私营业主及其他, 分别取值 1 到 4); 借款人年龄(按照 $d = (\text{年龄}_{\max} - \text{年龄}_{\min}) / 5$, 分别取值 1 到 6); 以及借款人信用分(按照是否进行了身份认证、视频认证、学历认证和手机认证取值 1 到 4); 借款的期限、金额和利率参照原始数据.

为使最终的贷款流标风险 P_{fail} 其值介于 0 到 1 之间, 本文在分类过程中的的基分类器 f 选择多元线性回归模型.

将不考虑投资人-贷款网络的网络属性的流标预测模型设置为参照组, 我们得到实验结果如表 1 所示. PNICA 提升了 8% 的准确率.

表 1 流标预测准确率

	参照组	PNICA
流标预测准确率	70%	78%

通过模型训练, 得到每个属性的相关系数. 其中网络属性的相关系数取值 0.8734, 对流标与否的影响最大. 也验证了引入投资人-贷款网络的有效性.

3.2.3 投资人风险偏好实验结果

通过 2.4 节提出的投资人风险偏好的度量方法, 得到 L_{train} 中投资人的历史投资风险表现 PI 的分布图,

横坐标是投资人历史投资风险表现分数, 纵坐标是投资人比例, 如图 5 所示, 符合正态分布.

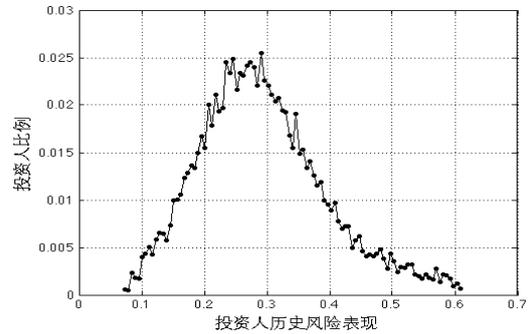


图 5 用户历史投资风险表现分布图

将投资人风险偏好 PI' 应用于 IDEA, 设置 PI' 分别为 0.1、0.3、0.5、0.7、0.9, 并得到不同 $topN$ 贷款候选集合下的实际投资收益率, 如图 6 所示, 各种风险偏好的投资人均可在 IDEA 算法下得到较高的实际投资收益率, 而高风险偏好的投资人可得到相对更高的实际收益率, 如图 7 所示, 发现随着 PI' 设置的越高, 也就是越风险厌恶, 贷款候选集中的贷款发布的利率越低, 如图 8 所示, 发现随着 PI' 设置的越高, 贷款候选集中的贷款实际出现坏账的笔数也越低, 验证了引入投资人风险偏好的有效性.

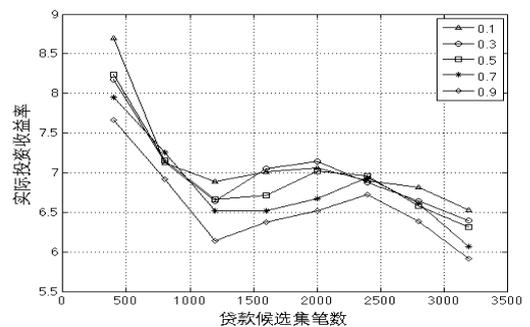


图 6 贷款候选集中实际投资收益率

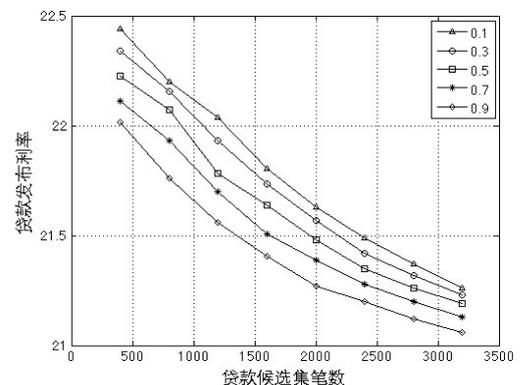


图 7 贷款候选集中贷款发布的利率

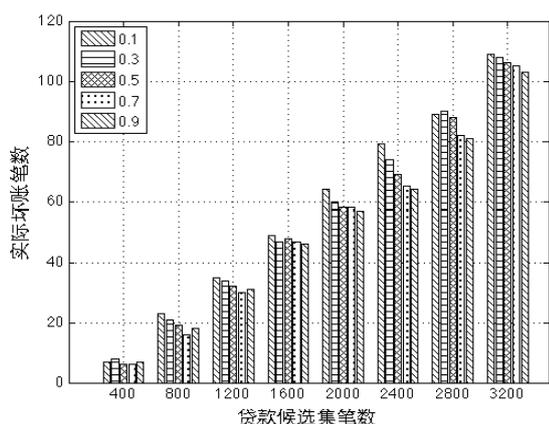


图 8 贷款候选集中实际坏账笔数

3.2.4 坏账风险实验结果

通过 2.3 节提出的贷款的坏账度量方法可计算得到 L_{test} 中每笔贷款的预期坏账风险, 并可以计算得到在每个坏账风险区间实际的坏账率, 如图 9 所示, 可发现实际的贷款坏账率与预期坏账风险正相关. 验证了坏账风险度量方法的有效性.

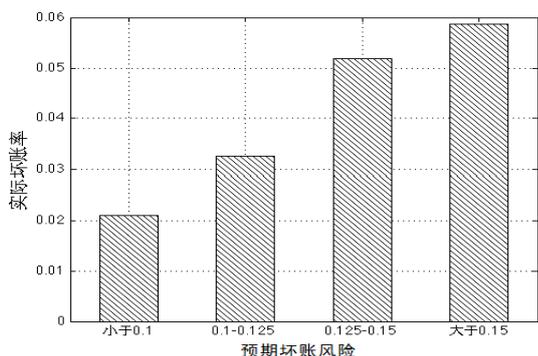


图 9 各预期坏账风险区间对应的实际坏账率

4 结论

本文通过建立投资人-贷款网络, 提出 PNICA 模型和 RBE 模型分别对贷款流标风险和坏账风险进行量化. 并在此基础上提出投资决策算法 IDEA. 通过拍拍贷数据集验证 IDEA 算法不仅可以提高风险度量的准确度, 取得更佳实际投资收益, 而且能够协助具有不同风险偏好的投资人进行投资决策.

参考文献

1 Zhao H, Wu L, Liu Q, et al. Investment recommendation in P2P lending: A portfolio perspective with risk management. IEEE International Conference on Data Mining. IEEE. 2014.

1109–1114.
 2 Wang X, Zhang D, Zeng X, et al. A Bayesian investment model for online P2P lending. *Frontiers in Internet Technologies*. Springer Berlin Heidelberg, 2013: 21–30.
 3 Dong G, Lai KK, Yen J. Credit scorecard based on logistic regression with random coefficients. *Procedia Computer Science*, 2010, 1(1): 2463–2468.
 4 Guo Y, Zhou W, Luo C, et al. Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research*, 2015.
 5 Luo C, Xiong H, Zhou W, et al. Enhancing investment decisions in P2P lending: an investor composition perspective. *Proc. of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2011. 292–300.
 6 Li Y, Rakesh V, Reddy CK. Project success prediction in crowdfunding environments. *WSDM*, 2016.
 7 Ceyhan S, Shi X, Leskovec J. Dynamics of bidding in a P2P lending service: Effects of herding and predicting loan success. *Proc. of the 20th International Conference on World Wide Web*. ACM. 2011. 547–556.
 8 Sonenshein S, Herzenstein M, Dholakia UM. How accounts shape lending decisions through fostering perceived trustworthiness. *Organizational Behavior & Human Decision Processes*, 2011, 115(1): 69–84.
 9 Li S, Qiu J, Lin Z, et al. Do borrowers make homogeneous decisions in online P2P lending market? An empirical study of PPDai in China. 2011 8th International Conference on Service Systems and Service Management (ICSSSM). IEEE. 2011. 1–6.
 10 Liu ZM. An Analysis of the lending behavior in P2P network credit-based on the dual persuasion model. *Finance Forum*, 2014.
 11 Jebara T, Wang J, Chang SF. Graph construction and b-matching for semi-supervised learning. *International Conference on Machine Learning*. ACM. 2009. 441–448.
 12 Su X. Collaborative filtering: A survey. 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions). IEEE. 2015.