

doi: 10.7690/bgzdh.2025.02.004

基于多目标进化算法的 P2P 贷款推荐

李宇远, 杜 杰

(南方电网数字电网集团有限公司南方电网数字企业科技(广东)有限公司, 广州 510670)

摘要: 针对目前 P2P 平台贷款组合中传统推荐算法存在的缺点, 提出一种基于多目标进化算法的组合贷款推荐模型。建立推荐组合预测评级和一致性目标函数, 平衡推荐贷款组合的准确性和回报率; 基于改进的多目标进化推荐算法对目标函数进行求解; 为解决决策变量空间维数过高问题, 提出决策空间降维和改进的初始化策略, 加快种群的收敛速度, 提升算法搜索效率。将所提模型分别与协同过滤(collaborative filtering, CF)、神经协同过滤(neural collaborative filtering, NCF)、物质扩散(probabilistic spreading, ProbS)、粒子群优化(particle swarm optimization, PSO)、动态规划算法(dynamic programming algorithm, DPA)、混合多目标进化和物质扩散算法(MOEA-ProbS)等模型进行对比, 结果表明: 所提模型性能有所提升, 平均准确率为 0.104 7, 平均利润系数为 0.154 2, 平均风险系数为 0.002 3。

关键词: P2P 平台; 组合贷款; 多目标进化; 风险评估

中图分类号: TP393 **文献标志码:** A

P2P Loan Recommendation Based on Multi-objective Evolutionary Algorithm

Li Yuyuan, Du Jie

(China Southern Power Grid Digital Enterprise Technology (Guangdong) Co., Ltd.

China Southern Power Grid Digital Digital Grid Corporation Co., Ltd., Guangzhou 510670, China)

Abstract: In view of the shortcomings of the traditional recommendation algorithm in the loan portfolio of P2P platform, a loan portfolio recommendation model based on multi-objective evolutionary algorithm is proposed. To balance the accuracy and the rate of return of the recommended loan portfolio, the objective function is established to predict the rating and consistency of the recommended loan portfolio. The objective function is solved based on the improved multi-objective evolutionary recommendation algorithm. To solve the problem of high dimension of the decision variable space, the dimension reduction of the decision space and the improved initialization strategy are proposed to accelerate the convergence speed of the population and improve the search efficiency of the algorithm. The proposed model is compared with collaborative filtering (CF), neural collaborative filtering (NCF), probabilistic spreading (ProbS), particle swarm optimization (PSO), dynamic programming algorithm (DPA), MOEA-ProbS and other models, and the results show that the performance of the proposed model is improved, the average accuracy is 0.104 7, the average profit coefficient is 0.154 2, and the average risk coefficient is 0.002 3.

Keywords: P2P platform; portfolio loan; multi-objective evolution; risk assessment

0 引言

计算机、网络、大数据、物联网、通信等技术^[1-2]的不断发展为金融领域^[3]的改革与创新带来了契机, 特别是为 P2P 贷款^[4-6]的发展奠定了基础。P2P 贷款是一种新兴的交易机制, 其将少量资金集中, 为有金融需求的人提供贷款。P2P 借贷服务主要体现在满足个人资金需求、发展个人信用体系、提高小额资金利用率等方面。

从 P2P 借贷市场, 贷款人可以获得比银行更高的利息收入; 而对于借款人来说, 除了高效和方便之外, 还可以获得更低的还款率。随着越来越多的借款人加入这些市场, 每天都有数以千计的贷款申请发布在这些平台上; 因此, 贷款人很难选择满足

其理想需求的贷款。近些年, 大量研究人员对 P2P 贷款进行研究, 已广泛应用于贷款风险评估^[7-8]、投资行为分析^[9]等领域。然而, 针对贷款组合推荐研究较少, 已有的研究要么仅考虑推荐的准确性(如个性化推荐^[10]), 要么仅考虑回报和风险管理^[11]。考虑到贷款推荐存在一些特殊需求, 如组合多样性、风险与收益偏好(低风险低回报、高风险高回报、风险与收益折中)等, 这些需求将导致推荐解空间存在海量搜索空间, 从而为推荐的效率及性能带来挑战。

为改善上述问题, 笔者提出一个多目标优化模型, 在确保推荐给贷款人的贷款符合其风险偏好基础上, 帮助贷款人选择贷款组合, 从而实现收益和

收稿日期: 2024-07-09; 修回日期: 2024-08-15

第一作者: 李宇远(1985—), 男, 广东人。

风险之间的权衡。首先,在分析 P2P 贷款特点基础上,建立了推荐组合预测评级和一致性目标函数,平衡推荐贷款组合的准确性和回报率;其次,为求解 P2P 贷款推荐组合多目标优化问题,提出一种基于改进的多目标进化推荐算法对目标函数进行求解;最后,提出决策空间降维和改进的初始化策略,加快种群的收敛速度,提升算法搜索效率,解决决策变量空间维数过高问题。

1 问题描述

笔者通过深入了解贷款交易的场景来定义 2 个目标函数:首先,为衡量推荐贷款组合的准确性和回报率,笔者提出目标推荐预测评级概念;其次,为度量推荐贷款组合的风险,提出目标一致性概念。为更好地说明这 2 个目标的计算,介绍基于指数表示的贷款组合的编码方法。令 P2P 平台中的每笔贷款都有一个唯一的索引号,并且贷款组合长度 L 设置为 N ,则任意一笔贷款组合可以表示为 $(L_1, L_2, \dots, L_N)$,其中 $L_i, i=1, 2, \dots, N$ 为整数,表示特定的贷款编号。

1.1 推荐预测评级

一般情况下,推荐系统的目标是获得一个效用函数,该函数试图预测用户对某个项目的“评级”或“偏好”。换句话说,推荐的最终目标是预测用户评级矩阵 R 中缺失的值。令系统中所有用户构成的集合为 U ,其中 $U=\{u_1, u_2, \dots, u_n\}$ 。系统中所有项目构成的集合为 V ,其中 $V=\{v_1, v_2, \dots, v_n\}$ 。此外,令 $R=(i, j)$ 为用户 u_i 对项目 v_j 的偏好。需注意,项目通常是非负整数或规一化的实数。推荐过程可描述为:将具有最高预测值 $R(i, j)$ 的 $v_j \in V$ 推荐给用户 u_i 。进一步,用户 u_i 的推荐项目集 V_i 可通过以下方式获得:

$$V_i = \arg \max_{V_i \subset V} \sum_{v_j \in V_i} R(i, j) \quad (1)$$

对于上式中的推荐问题,最常用的方法有基于内容的过滤和协同过滤等。这些方法只适用于具有明确评级的推荐系统。考虑到 P2P 借贷平台中的推荐系统,其评级矩阵 R 中的评级是隐式的即 R 中的每个元素 $R(i, j)$ 非 1 即 0,这意味着贷款人 u_i 是否投资了贷款 v_j 。

考虑传统物质扩散^[12](ProbS)算法无明确贷款推荐等级,且仅通过准确率指标而无法评估收益等问题,笔者提出了改进的 ProbS 算法。该算法对权

重进行初始化,并通过考虑 P2P 贷款问题中每个贷款的效用计算权重;因此,改进的 ProbS 算法将回报率和准确率组合为一个目标加权预测评级目标函数。首先,需要获取评级矩阵 R ,其中 $R=(i, j) \in R$ 为贷款人 u_i 对相应贷款 v_j 的预测得分;其次,预测评级目标函数可根据评级矩阵 R 计算。需注意,当预测评级的值较高时,相应的贷款组合更精确,或者回报率较高。对于贷方 u_i ,预测评级的计算如下:

$$R_p = \sum_{j=1}^L r_{ij} / L \quad (2)$$

式中: L 为推荐列表/投资组合长度; r_{ij} 为由改进 ProbS 获得的 u_i 给出的 v_j 的评级权重。可以看出:在改进 ProbS 方法中的 2 个分配步骤之后,由每笔贷款的权重确定的回报隐含地包含在偏好中。度量 R_p 不仅可以测量用户的兴趣爱好,还可以捕获推荐贷款组合的回报。

1.2 目标一致性

由于 P2P 贷款中的贷款人有自己的风险偏好,为确保推荐给贷款人的贷款符合其风险偏好,笔者提出了第 2 个目标函数:目标一致性。当一致性的值较高时,推荐的贷款组合更符合贷款人的风险偏好;因此,贷款人的对贷款风险更可接受。对于贷款人 u_i ,一致性的具体计算如下:

$$C_f = \sum_{j=1}^L c_{ij} / L \quad (3)$$

式中 c_{ij} 为贷款项 v_j 和贷款人 u_i 之间的一致程度。此外,可通过模拟高斯分布对 c_{ij} 进行量化,其定义如下:

$$c_{ij} = \frac{1}{\sqrt{2\pi}\sigma} e^{- (r_j/E_i - \mu)^2 / 2\sigma^2} \quad (4)$$

式中: σ 和 μ 为高斯分布可调参数,分别为标准差和均值; E_i 为贷款人 u_i 的回报偏好,即通过分析其历史投资记录,贷款人 u_i 的平均预期回报率。在投资方面,贷款人的偏好主要反映在其回报预期中。在 P2P 借贷平台中,笔者将贷款利率通常视为回报;因此,贷款人对回报率 E_i 的偏好可通过其历史投资获得,具体定义如下:

$$E_i = \sum_{j=1}^H r_j / H \quad (5)$$

式中: H 为贷款人 u_i 已投资的贷款数量; r_j 为贷款 v_j 的申报利率。

为归一化 c_{ij} 的值, 本文中令 $\sigma=1/\sqrt{2\pi}$ 。注意到每个贷款人都有自己的预期利率, 其中某些贷款人的预期利率较高, 该项贷款可能具有高风险; 而一些贷款人的预期利率较低, 该贷款可能具有低风险。为满足不同贷款人的不同利率风险, 对于每个贷款人 u_i , 笔者旨在推荐最接近贷款人预期利率 E_i 的贷款 v_j 。换言之, 当 $r_j/E_i=1$ 时, 一致性程度应最大; 因此, 当 c_{ij} 达到峰值时, 设置 $\mu=1$ 以满足贷款 v_j 和贷款人 u_i 的最大一致性。

2 多目标进化推荐算法

考虑到贷款推荐存在一些特殊需求, 如组合多样性、风险与收益偏好(低风险低回报、高风险高回报、风险与收益折中)等, 这些需求将导致存在海量搜索空间, 从而为推荐的实时性及有效性带来挑战。笔者提出一种多目标进化推荐算法, 并通过降维策略和初始化策略, 加快算法的收敛速度, 提高算法的性能。

2.1 算法执行过程

令 N_p 为初始种群大小, T_g 为种群进化的最大迭代次数。算法具体执行过程可描述如下:

步骤 1 基于决策空间降维策略去除一些劣质解(贷款), 从而降低种群进化过程中的搜索难度, 并获得降维搜索空间 O 。

步骤 2 对于每个贷款人 u_i , 基于初始化策略生成初始种群 P 。

步骤 3 种群进化首先随机选择 2 个个体进行交叉从而生成新个体 g , 并对新个体进行变异从而生成最终个体 v 。

步骤 4 重复上述交叉、变异操作最终生成新的种群 P_n 。对于交叉操作, 笔者选择将 2 个个体中相同的贷款保留给子代, 然后随机选择剩余的不同贷款来填充其余的子代。对于变异操作, 笔者为交叉生成的个体的每个基因位置生成一个随机数。如果随机数小于突变概率, 则从搜索空间 O 中随机选择一个不在推荐列表中的贷款来替换原始基因位置。需注意, 每个贷款人 u_i 都有一个对应于自己的候选推荐列表 O_i 。

步骤 5 种群进化计算种群中所有个体的目标函数值, 然后采用非支配排序和拥挤距离排序获得的新种群 P_n 替换初始种群 P 。

步骤 6 当种群进化算法到达最大迭代次数 T_g 时, 终止算法, 并输出推荐结果。

接下来对多目标进化推荐算法 2 个主要策略: 决策空间降维和种群初始化策略进行介绍。

2.2 决策空间降维

由于 P2P 贷款市场上的贷款数量巨大, 决策变量的空间维数过高, 使得传统搜索算法在搜索时效率较低。笔者提出降低决策空间的维数策略, 提升算法搜索效率。

为得到降维后的最终搜索空间 O , 笔者利用改进 ProbS 获取带权重的评级矩阵 R_P 和一致性矩阵 C_f ; 同时, 分别为每个贷款人 u_i 计算 3 个长度为 L_e 的候选列表。具体而言, 候选名单 X_{i1} 、 X_{i2} 和 X_{i3} 分别通过获得 r_{ij} 和 c_{ij} ($j \in [1, 3]$) 中最大值的贷款获得。将 3 个候选列表(X_{i1} 、 X_{i2} 和 X_{i3})组合, 得到最终决策空间 O_i , 即:

$$O_i = X_{i1} \cup X_{i2} \cup X_{i3}。 \quad (6)$$

式中 $\cup$ 为组合算子。

进一步, 重复上述过程, 降维后的总决策空间 O 可描述如下:

$$O = \{O_1, O_2, \dots, O_{N_m}\}。 \quad (7)$$

式中 N_m 为贷款人的数目。

为更好地确定 L_m 的值, 考虑到最终推荐贷款组合的长度为 L ; 因此, 在降维和进化搜索中保持相似的选择比率, 即:

$$N_e/L_e = L_e/L; \quad (8)$$

$$L_e = \left\lfloor \sqrt{N_e \times L} \right\rfloor, N_e \gg L。 \quad (9)$$

2.3 初始化策略

值得注意的是, 传统多目标进化算法中常用的初始化策略是随机初始化。在 P2P 借贷中, 即使降维后, 决策空间维数仍然很大; 因此, 随机初始化得到的初始种群质量较差, 这使得在进化中很难找到更好的解。为解决上述问题, 笔者提出一种改进的初始化策略, 不仅可以加快种群的收敛速度, 而且可以提高算法的性能。

改进的初始化策略执行过程中, 对任意的贷款人 u_i 、贷款 v_j 都有 2 类目标, 即 r_{ij} 和 c_{ij} 。进一步, 笔者在初始化中使用调和平均值平衡 r_{ij} 和 c_{ij} , 因此有:

$$h_{ij} = 2 \times r_{ij} \times c_{ij} / (r_{ij} + c_{ij})。 \quad (10)$$

式中 h_{ij} 为经调和平均值平衡后的矩阵。

改进的初始化策略主要包括 2 个步骤: 1) 对于每个贷款人 u_i , 可通过 r_{ij} 、 c_{ij} 和 h_{ij} 确定 3 个个体

P_1 、 P_2 和 P_3 ; 2) 对于每笔贷款, 采用竞争进行选择。也就是说, 对于贷款人 u_i , 从 O_i 中随机选择 2 笔贷款 m 和 n , 并将 h_{im} 与 h_{in} 进行比较。如果 h_{im} 较大, 则将 m 添加到推荐列表中; 否则, 将 n 添加到推荐列表中。通过该方式, 最终选取 L 笔贷款, 然后将其合并为一笔贷款。重复上述过程, 则可以得到第 i 个贷款人的种群 P_i 。最后, 将初始种群 P 的所有个体进行合并, 则有:

$$P = \{P_1, P_2, \dots, P_{N_p}\}。 \quad (11)$$

式中 N_p 为初始种群大小。

3 仿真与分析

3.1 数据集与实验设置

实验所用数据集主要由某信贷公司提供的从 2015 至 2019 年的所有交易记录构成。数据集中收集了有关贷款的一些基本信息, 如提交时间、截止日期、金额、回报率和投资状态等。其中, 投资状态主要包括贷款人信息、投资时间、预期回报率、投资金额和成功与否等信息。对数据集中数据进行整理, 清洗部分缺失数据, 最终数据集相关统计情况如表 1 所示。此外, 数据中贷款共包括 4 类标签, 分别用来描述异常贷款(-1)、终止贷款(0)、延期贷款(1)、有效贷款(2)。可以看出: 该数据集中每条记录包含 9 个属性, 共计 4 397 730 条记录, 数据集具有纬度高、特征复杂、数据海量等特点。

表 1 数据集相关统计情况

描述	数值	描述	数值
贷款人数	10 033	贷款人未完全偿还贷款	-1
贷款笔数	39 993	贷款资金不足	0
投资记录数	4 397 730	贷款人延迟还款	1
属性长度	9	贷款人按期还款	2

同时, 仿真软件环境为 Matlab 2019A, 算法运行硬件环境为 i7-6700@3.4Ghz CPU, 内存为 128 Gb 的 P900 联想服务器, 操作系统为 Windows 10。仿真过程涉及的相关参数如表 2 所示。

表 2 仿真过程涉及的相关参数

参数	数值
种群大小	100
最大生成个数	100
最大迭代次数	1 000
交叉概率	{0.5, 0.48, 0.46, 0.44, 42}
变异概率	{0.1, 0.05, 0.03, 0.025, 0.02}
推荐列表 L 长度	{10, 20, 30, 40, 50}

3.2 指标选取

为验证所提模型的有效性, 将所提模型与协同

过滤(CF)、神经协同过滤(NCF)、ProbS、粒子群优化(PSO)、动态规划算法(DPA)、混合多目标进化和物质扩散算法(MOEA-ProbS)等模型进行比较。需注意, 除 PSO 和 DPA 通常用于解决单目标优化问题, 其余算法如 CF、NCF、ProbS、MOEA-ProbS 均为多目标优化方法。对 2 个单目标优化方法(PSO 和 DPA)采用权重进行处理, 将对目标问题转换为单目标函数, 则 PSO 和 DPA 中目标函数可简化为: $\alpha R_P + (1-\alpha)C_f$, 其中 α 为权重, 主要用来平衡推荐预测评级和一致性目标所占比重(本实验选取 $\alpha=0.5$)。其余多目标优化算法目标函数均与笔者所提模型相同。

同时, 实验时选取检索准确率(accuracy, ACC)、利润系数(profit coefficient, PCO)和风险系数(risk coefficient, RCO)等指标用于评估模型检测性能。各指标计算公式如下所示:

$$ACC = \sum_{i=1}^{|U|} A_i(L) / |U|; \quad (12)$$

$$A_i(L) = h_i(L) / L; \quad (13)$$

$$PCO = \sum_{i=1}^{|U|} P_i(L) / |U|; \quad (14)$$

$$P_i(L) = \sum_{j=1}^L r_j \times s_j / L; \quad (15)$$

$$RCO = \sum_{i=1}^{|U|} R_i(L) / |U|; \quad (16)$$

$$R_i(L) = \sqrt{\frac{1}{L} \sum_{j=1}^L (r_j \times \bar{r})^2}。 \quad (17)$$

式中: U 为所有贷款人的集合; $h_i(L)$ 为推荐列表中贷款人 u_i 实际投资的贷款数量; r_j 为贷款 v_j 的利率; s_j 为贷款 v_j 的状态标签; $\bar{r}$ 为推荐列表中贷款的平均利率。

3.3 性能对比与结果分析

表 3—5 分别为不同模型在不同推荐列表长度时准确率、利润系数和风险系数统计结果。

表 3 不同推荐列表长度准确率统计

模型	10	20	30	40	50	均值
CF	0.045 1	0.043 9	0.038 8	0.035 6	0.034 5	0.039 6
NCF	0.051 8	0.050 6	0.048 9	0.047 6	0.046 5	0.049 1
ProbS	0.108 3	0.089 1	0.084 6	0.079 8	0.071 1	0.086 6
PSO	0.022 5	0.021 8	0.020 3	0.019 2	0.018 9	0.020 5
DPA	0.128 5	0.114 3	0.108 4	0.096 7	0.090 3	0.107 6
MOEA-ProbS	0.068 9	0.049 8	0.043 3	0.041 8	0.040 1	0.048 8
所提模型	0.143 2	0.112 8	0.098 7	0.089 2	0.082 6	0.104 7

表 4 不同推荐列表长度利润系数统计

模型	10	20	30	40	50	均值
CF	0.048 9	0.050 3	0.051 1	0.052 7	0.053 8	0.051 4
NCF	0.060 3	0.061 4	0.062 5	0.063 3	0.064 1	0.062 3
ProbS	0.068 2	0.066 7	0.063 5	0.062 8	0.061 0	0.064 4
PSO	0.160 8	0.138 5	0.126 7	0.119 2	0.117 7	0.132 6
DPA	0.157 8	0.140 1	0.130 2	0.123 3	0.120 5	0.134 4
MOEA-ProbS	0.138 8	0.119 2	0.108 6	0.102 1	0.100 3	0.113 8
所提模型	0.183 8	0.157 7	0.148 3	0.142 5	0.138 9	0.154 2

表 5 不同推荐列表长度风险系数统计

模型	10	20	30	40	50	均值
CF	0.035 6	0.036 8	0.037 2	0.037 6	0.037 9	0.037 0
NCF	0.032 1	0.032 7	0.033 2	0.033 6	0.034 0	0.033 1
ProbS	0.028 8	0.029 5	0.030 5	0.031 1	0.031 5	0.030 3
PSO	0.030 2	0.043 8	0.048 9	0.052 3	0.056 5	0.046 3
DPA	0.005 1	0.005 5	0.005 9	0.006 3	0.006 6	0.005 9
MOEA-ProbS	0.002 1	0.002 4	0.002 7	0.003 1	0.003 4	0.002 7
所提模型	0.001 8	0.002 1	0.002 3	0.002 5	0.002 7	0.002 3

从上表可以看出：DPA 算法平均准确率最高 0.107 6，所提模型次之 0.104 7，PSO 算法平均准确率最低 0.020 5；所提模型平均利润系数指标最高 0.154 2，DPA 算法次之 0.134 4，PSO 算法最低 0.020 5；所提模型平均风险系数指标最低 0.002 3，MOEA-ProbS 算法次之 0.002 7，PSO 算法最高 0.046 3。所提模型、CF、NCF、ProbS、PSO、DPA、MOEA-ProbS 等模型在不同指标性能各有优劣，然而所提模型在牺牲部分准确率指标下(和平均准确率最高的 DPA 算法相比，仅减少 0.29%)，平均利率系数和风险系数表现最优。其中，平均利润系数较排名第 2 的 DPA 算法相比增加了 1.98%，较性能最低的 PSO 算法相比增加了 13.37%；平均风险系数较排名第 2 的 MOEA-ProbS 算法相比减少了 0.04%，较性能最低的 PSO 算法减少了 4.4%。

4 结论

笔者基于多目标优化对 P2P 平台组合贷款优化问题进行分析，提出一种基于多目标进化算法的组合贷款推荐模型。首先，为衡量推荐贷款组合的准确性和回报率，提出目标推荐预测评级和一致性 2

个目标函数；其次，将这 2 个目标函数作为多目标进化的输入并搜索解空间；最后，通过降维和初始化策略，加快算法的收敛速度和高性能。所提模型为 P2P 平台组合贷款推荐提供了一定借鉴。

参考文献：

[1] 毛龙灿, 杨南. 基于大数据背景的皮革人才培养优化研究[J]. 中国皮革, 2021, 50(9): 38-41.

[2] 杨涛. 互联网时代下皮革行业电商物流体系研究分析[J]. 中国皮革, 2021, 50(8): 82-85.

[3] 王玲琳. 大数据技术在金融信用卡领域的应用研究[J]. 商展经济, 2021(22): 75-78.

[4] 伍思雨, 冯骥. 基于去参数聚类方法的 P2P 网贷平台分析[J]. 现代信息科技, 2020, 4(5): 32-34, 37.

[5] 王萧萧, 王亭雯, 马玉玲, 等. 基于深度森林的 P2P 网贷借款人信用风险评估方法[J]. 计算机科学, 2021, 48(S2): 429-434.

[6] GUO Y, JIANG S, ZHOU W, et al. A predictive indicator using lender composition for loan evaluation in P2P lending[J]. Financial Innovation, 2021, 7(1): 1-24.

[7] 肖灵云, 刘军库. 基于相似度组合的主观题评分方法研究[J]. 贵州大学学报(自然科学版), 2021, 38(5): 64-68.

[8] PENA M V T, BREIDBACH C F. On emergence in service platforms: An application to P2P lending[J]. Journal of Business Research, 2021, 135: 337-347.

[9] 周先波, 欧阳梦倩. P2P 投资经验与甄别违约风险的能力—基于学习的视角[J]. 统计研究, 2019, 36(12): 40-54.

[10] 万常选, 游运, 江腾蛟, 等. 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[J]. 计算机研究与发展, 2018, 55(10): 2307-2320.

[11] 刘驹展. 基于 P2P 风险性分析的推荐系统[J]. 通讯世界, 2019, 26(4): 1-2.

[12] 李普聪, 王顺, 钟元生, 等. 融合信任关系的物质扩散与热传导混合推荐算法[J]. 小型微型计算机系统, 2021, 42(10): 2044-2052.