计算金融与仿真

课程论文

论文题目：计算金融与仿真

学生姓名：李晶晶　　张璐　　朱冯婧

指导老师：邓智斌

1. 投资组合选择介绍

P2P（点对点）借贷是一种个人之间直接借贷的方式, 通过在线平台进行, 无需传统银行中介. 这种创新的融资方式提供了一个市场, 借款人可以向多个贷方申请贷款, 贷方可以选择为这些贷款提供全部或部分资金. 这一模式利用技术手段简化了借贷过程, 通常为借款人带来更低的利率, 同时为投资者提供更高的回报.

目前, 全球几大平台主导着 P2P 借贷领域, 主要包括 LendingClub, Prosper 和 Funding Circle 等. 这类平台通过评估贷款项目和借款人的 FICO 分数, 贷款金额和期限, 借款人的资产, 债务状况, 就业类型等数据, 为每笔贷款提供风险评级. 其中, LendingClub 是美国最大的 P2P 借贷平台之一, 提供个人贷款和投资机会. 本文使用来自 LendingClub 的公共数据集（涵盖 2007–2018 年的贷款记录）, 将贷款属性转化为违约概率作为后续建模依据.

为了实现贷款投资组合的多元化, 投资者可依据自身的风险偏好从不同风险等级的贷款中进行选择. 与传统金融机构贷款不同, P2P 贷款中的每位个人投资者其可用于投资的资金较为有限. 因此, 如何在资金约束下, 基于借款人的风险特征评估潜在回报与风险, 选择合适的贷款项目, 并进行最优资金配置, 是 P2P 投资者亟需解决的问题.

大量学者针对 P2P 网络贷款的投资组合优化问题开展了深入研究. Wan 等人将贷款投资组合决策转化为一个在特定时点下实现收益最大化与风险最小化的优化问题, 并引入混合治愈模型 (mixture cure model,MCM) 以提升投资效果, 构建实例驱动的模型对投资者的组合决策进行优化 [1]. Guo 等人则提出一种基于实例的 P2P 投资组合决策模型, 从风险最小角度出发, 利用 LendingClub 和 Prosper 数据集实现投资组合配置优化 [2]. Ajay 等人则将该问题转化为多目标优化模型, 在计算贷款相似度时将期望值框架与传统核方法结合, 优化结果优于既有模型 [3].

贷款投资组合优化（Loan Portfolio Optimization）与传统投资组合优化具有相似性, 但更关注与贷款相关的特定风险, 例如借款人信用评分, 贷款金额, 借款用途等因素引发的借款人风险, 违约风险以及利率波动风险. 优化目标是通过合理配置不同借款人贷款项目, 在控制风险的同时实现收益最大化.

在风险量化方面, VaR（Value at Risk）与 CVaR（Conditional Value at Risk）被广泛用于衡量投资组合的风险水平. 其中, VaR 表示在置信水平 α 下, 金融资产或组合在未来特定持有期内的最大可能损失；而 CVaR 则刻画了超过 VaR 截止点的极端损失的期望, 即对尾部风险的加权平均, 是更稳健的风险衡量指标. 在贷款组合优化中, CVaR 能够帮助投资者更有效地控制整体尾部损失风险, 从而实现收益与风险的权衡优化.

2. 模型建立

决策变量

$$x_i \in \{0,1\}, \quad i = 1,2,\ldots,N$$

其中,

- $x_i = 1$ 表示选择资助第 i 个贷款对象；
- $x_i = 0$ 表示不选择.

参数说明

- A_i：第 i 个贷款的金额；
- r_i：第 i 个贷款的利率；
- P_i：第 i 个贷款的违约概率；
- B：总投资预算；
- $R_{\max}$：允许的最大违约风险；
- G_k：第 k 个信用等级的贷款集合；
- α_k：第 k 个信用等级的最大投资比例；
- m：最多选择的贷款数量（Top-m）.

模型形式

$$\begin{aligned}
\max_{x_i \in \{0,1\}} \quad & \sum_{i=1}^{N} x_i A_i \left[r_i (1 - P_i) - P_i \right] \\
\text{s.t.} \quad & \sum_{i=1}^{N} x_i A_i \leq B \quad \text{(预算限制)} \\
& \sum_{i \in G_k} x_i A_i \leq \alpha_k B, \quad \forall k \quad \text{(信用等级比例约束)} \\
& \sum_{i=1}^{N} x_i A_i P_i \leq R_{\max} \quad \text{(风险控制)} \\
& \sum_{i=1}^{N} x_i \leq m \quad \text{(Top-}m\text{ 选择)}
\end{aligned} \tag{1}$$

$$\begin{aligned}
\max_{x_i \in \{0,1\}} \quad & \sum_{i=1}^{N} x_i A_i \left[r_i (1 - P_i) - P_i \right] \\
\text{s.t.} \quad & \sum_{i=1}^{N} x_i A_i \leq B \quad \text{(预算限制)} \\
& \sum_{i \in G_k} x_i A_i \leq \alpha_k B, \quad \forall k \quad \text{(信用等级比例约束)} \\
& \sum_{i=1}^{N} x_i A_i \cdot \tilde{P}_i^{(s)} - \eta \leq \mathcal{M} z_s, \quad \forall s = 1, \ldots, S \quad \text{(VaR 限制)} \\
& \sum_{s=1}^{S} z_s \leq (1 - \beta) S \\
& \sum_{i=1}^{N} x_i \leq m \quad \text{(Top-}m\text{ 选择)} \\
& z_s \in \{0,1\}, \quad \eta \in \mathbb{R}
\end{aligned} \tag{2}$$

$$
\begin{aligned}
\max_{x_i \in \{0,1\}} \quad & \sum_{i=1}^{N} x_i A_i \left[r_i (1 - P_i) - P_i \right] \\
\text{s.t.} \quad & \sum_{i=1}^{N} x_i A_i \leq B \quad \text{(预算限制)} \\
& \sum_{i \in G_k} x_i A_i \leq \alpha_k B, \quad \forall k \quad \text{(信用等级比例约束)} \\
& \xi_s \geq \sum_{i=1}^{N} x_i A_i \cdot \tilde{P}_i^{(s)} - \eta, \quad \forall s = 1, \ldots, S \quad \text{(场景损失)} \\
& \eta + \frac{1}{S(1-\beta)} \sum_{s=1}^{S} \xi_s \leq \text{CVaR}_{\max} \quad \text{(CVaR 限制)} \\
& \sum_{i=1}^{N} x_i \leq m \quad \text{(Top-}m\text{ 选择)} \\
& \xi_s \geq 0, \quad \eta \in \mathbb{R}
\end{aligned} \tag{3}
$$

3. 算法设计

Algorithm 1: 启发式算法：贷款组合优化（含 CVaR 控制）

Input: 贷款数据 $\{A_i, r_i, P_i\}_{i=1}^{N}$；预算 B；Top-m 限制；CVaR 上限 $\mathrm{CVaR}_{\max}$；置信水平 β；模拟场景矩阵 $\tilde{P}_i^{(s)} \in \{0,1\}^{S\times N}$

Output: 最优选择向量 $x^* \in \{0,1\}^N$

$x_{\text{best}} \leftarrow 0$； ; // 初始为空解

根据评分 $\text{score}_i = A_i[r_i(1-P_i) - P_i]$ 降序排列贷款

$total\_budget \leftarrow 0$, $total\_selected \leftarrow 0$, $x_{\text{current}} \leftarrow 0$

foreach i *in* 排序后的贷款列表 **do**

    if $total\_budget + A_i > B$ 或 $total\_selected + 1 > m$ **then**

        continue;

    $x_{\text{current}}[i] \leftarrow 1$;

    $total\_budget \leftarrow total\_budget + A_i$;

    $total\_selected \leftarrow total\_selected + 1$;

if `Feasible`($x_{current}$) **then**

    $x_{\text{best}} \leftarrow x_{\text{current}}$;

Function `Feasible`(x):

    for $s \leftarrow 1$ **to** S **do**

        $L_s \leftarrow \sum_{i=1}^{N} x_i A_i \cdot \tilde{P}_i^{(s)}$;

    $\eta \leftarrow \beta$ 分位点的 $\{L_s\}$;

    for $s \leftarrow 1$ **to** S **do**

        $\xi_s \leftarrow \max(L_s - \eta, 0)$;

    $\mathrm{CVaR}_\beta(x) \leftarrow \eta + \frac{1}{S(1-\beta)} \sum_{s=1}^{S} \xi_s$;

    return 是否满足 $\mathrm{CVaR}_\beta(x) \le \mathrm{CVaR}_{\max}$ 且满足其他约束;

return x_{best};

4. 案例研究

4.1. 数据集描述

本研究使用的数据集来自 Lending Club 平台, 原始数据由 Kaggle 网站https://www.kaggle.com/datasets/wordsforthewise/lending-club公开提供. Lending Club 是美国最大的网络借贷平台之一, 提供了详尽的个人借款申请及其还款情况的数据, 广泛应用于学术界和工业界进行信贷风险评估, 违约预测及投资组合优化等研究.

该数据集包含了从 2007 年至 2018 年的借款记录, 共计数百万条样本. 每条记录对应一笔贷款申请, 涵盖了包括贷款金额, 利率, 借款人信用等级, 债务收入比, 贷款期限, 还款状态, 就业年限, 收入, 地址状态, 房屋所有权, FICO 评分区间等在内的多维度信息.

在本研究中, 我们主要筛选并保留以下变量用于建模分析：

- **loan_amnt**: 借款人申请的贷款金额, 作为 A_i；
- **int_rate**: 借款合同中约定的年利率, 用于计算收益率 r_i；
- **grade**: 借款人的信用等级（A 至 G）, 用于分组限制；
- **loan_status**: 实际贷款的还款状态（如 Fully Paid, Charged Off）, 用于推断违约情况；
- **annual_inc**: 借款人年收入；
- **dti**: 债务收入比, 用于辅助风险刻画；
- **term**: 贷款期限（如 36 months 或 60 months）；
- **emp_length**: 借款人工作年限；
- **addr_state**: 借款人所在州；
- **fico_range_high, fico_range_low**: 借款人 FICO 信用评分区间；

为了满足模型中对违约概率 P_i 的需求, 我们将 “loan_status” 字段中状态为 “Charged Off” 的贷款视为违约样本, 其余如 “Fully Paid”, “Current” 等状态作为非违约样本, 并基于历史频率法估算每一类贷款的违约概率.

此外, 为模拟贷款违约的风险场景, 我们以借款人的信用等级, FICO 评分和历史违约频率为依据, 构建了 S 个 Monte Carlo 风险场景, 用于后续 CVaR 优化模型的风险评估.

通过上述数据处理步骤, 最终形成了一个结构规范, 信息完备, 适用于组合优化问题的数据集, 为后续实证分析与建模提供了坚实的数据基础.

4.2. 使用机器学习预测违约概率 P_i

尽管本研究使用的数据集中所有贷款均已获得实际资助, 但在资金有限, 需进行优选配置的情境下, 我们仍需要对这些已发放贷款的还款风险进行再评估. 为此, 我们引入机器学习方法, 对每笔贷款的未来违约概率 P_i 进行预测建模, 以作为后续优化模型中的风险输入参数.

建模目标 预测函数的目标是：对于每笔已发放贷款 i, 根据其已知特征向量 x_i, 估计其在未来发生违约的概率 $P_i = \mathbb{P}(y_i = 1 \mid \mathrm{x}_i)$. 其中, $y_i = 1$ 表示贷款最终发生违约（如状态为 `Charged Off`）, $y_i = 0$ 表示贷款最终还清（如状态为 `Fully Paid`）.

特征构造 我们基于借款人基本属性, 贷款合同信息以及信用评级等信息构建预测特征集, 涵盖但不限于以下变量：

- **loan_amnt**：贷款金额；
- **term**：贷款期限；
- **int_rate**：贷款利率；
- **grade** 和 **sub_grade**：Lending Club 信用评级；
- **emp_length**：工作年限；
- **home_ownership**：住房类型；
- **annual_inc**：年收入；
- **dti**：债务收入比；

- **purpose**：贷款用途；
- **fico_range_high / low**：信用评分；

建模方法 考虑到目标变量仍是二元状态（违约 / 未违约）, 我们采用监督学习的二分类方法进行建模. 尝试的模型包括逻辑回归（Logistic Regression）, 随机森林（Random Forest）, 梯度提升树（GBDT）, 极端梯度提升（XGBoost）等.

由于样本中违约样本占比相对较小, 我们在训练过程中采用类别加权, 欠采样等方式处理类别不平衡问题.

训练与评估 我们将全部已发放贷款随机划分为训练集（70%）与测试集（30%）, 使用交叉验证调优参数, 并基于测试集评估模型表现. 评价指标包括准确率（Accuracy）, AUC 值（Area Under the ROC Curve）, F1 分数等.

最终, 我们选择 AUC 表现最优的模型用于对所有贷款样本生成违约概率预测值 $\hat{P}_i$, 作为后续优化模型中的输入.

说明 虽然原始数据中每笔贷款都已实际放款, 但我们的建模任务是为现实中的“再选择”提供依据. 即在预算受限, 资源不足时, 如何在这些真实已放款的贷款中优先挑选违约概率低, 预期收益高的子集, 构建一个更稳健的投资组合. 因此, $\hat{P}_i$ 的预测并非用于决定放款与否, 而是作为组合优化的“风险估计量”, 用于构建期望收益与 CVaR 等风险指标.

4.3. 使用算法求解原始模型, VaR 模型与 CVaR 模型

在本节中, 我们分别构建并求解三类贷款筛选优化模型：原始模型, VaR（Value at Risk）模型和 CVaR（Conditional Value at Risk）模型. 三者均以最大化投资收益为目标, 并在此基础上逐步引入风险控制手段, 以模拟实际投资场景中对风险的不同管控需求.

原始模型（期望损失约束） 原始模型以最大化期望净收益为目标, 同时设置多个线性约束, 确保预算控制, 信用等级平衡与风险控制等现实要求. 该模型为典型的 0-1 整数规划问题, 可使用 Gurobi 等商业求解器在中小规模下获得最优解.

VaR 模型（引入分位损失约束） VaR 模型进一步考虑极端情境下的最大可能损失, 将分位数损失 VaR_β 作为新的风险控制手段. 通过构造 Monte Carlo 风险场景, 我们对贷款可能发生的违约路径进行模拟, 并引入辅助变量 $z_s \in \{0,1\}$ 表示某个场景是否超过 VaR 阈值 η.

由于 VaR 对应的非凸约束结构和 0-1 变量使得模型求解复杂度较高, 通常只适用于中小规模问题, 且求解结果可能存在不连续性.

CVaR 模型（引入条件期望损失约束） 为克服 VaR 模型不可微, 不可凸等缺点, CVaR 模型通过引入辅助变量 $\xi_s \geq 0$ 和 VaR 估计值 η, 以线性结构实现对尾部风险的精确控制. 其约束具有良好的可解性与可扩展性, 广泛应用于实际金融优化问题中.

该模型可转化为混合整数线性规划（MILP）形式, 使用 Gurobi 等求解器可以在合理时间内获得精确解, 特别适合场景规模较大的优化问题.

求解策略 在三类模型中, 原始模型为最简形式, 适用于风险可控或追求高收益的情境；VaR 模型在表达风险容忍边界方面具有直观优势；而 CVaR 模型则提供了更稳健的风险控制能力和更优的优化特性.

考虑到实际求解效率与解的稳定性, 本研究在实现中主要采用 CVaR 模型进行主模型求解, 并通过 Gurobi 求解器实现精确建模与最优解求取. 对于大规模场景（如 $S > 1000$）, 可结合启发式算法进行初解生成与变量预选, 从而提升整体求解效率.

4.4. 结果分析

5. 结论

参考文献

[1] 万洁 and 郝俊. 基于混合治愈模型的网络借贷投资组合决策模型与实证研究. *系统科学与数学*, 43(5):1138–1156, 2023.

[2] Yiyang Guo, Wei Zhou, Chen Luo, and et al. Instance-based credit risk assessment for investment decisions in p2p lending. *European Journal of Operational Research*, 249(2):417–426, 2016.

[3] A Byanjankar, J Mezei, and M Heikkila. Data-driven optimization of peer-to-peer lending portfolios based on the expected value framework. *Intelligent Systems in Accounting, Finance and Management*, 28(2):119–129, 2021.

小组分工

姓名	学号	分工
李晶晶	202428016443014	论文撰写, 模型设计
张璐	202428016446002	数据处理, 模型求解
朱冯婧	202428016443008	论文撰写, 算法设计