

基于不平衡数据的AdaFocal-XGBoost集成信用评分模型研究

郭楷, 范宏*

东华大学旭日工商管理学院, 上海

收稿日期: 2024年11月6日; 录用日期: 2024年12月2日; 发布日期: 2024年12月9日

摘要

随着大数据时代的到来, 信用风险管理在金融领域的重要性日益凸显, 信用评分作为其核心工具, 面临着海量增长的客户信用数据和个体信用画像动态变迁的挑战。传统的信用评估方法在适应性和灵活性上存在不足, 尤其是在处理不平衡数据时。本文提出了基于不平衡数据的AdaFocal-XGBoost集成信用评分模型, 旨在提高信用风险预测的准确性和适应性。AdaFocal-XGBoost模型结合了XGBoost的高效计算和AdaFocalLoss的自适应损失调整, 特别针对样本不平衡问题进行了优化。通过在UCI数据库中的四个信贷数据集(Australian、German、Japan和Taiwan地区)上的实验, 本研究全面评估了AdaFocal-XGBoost模型的性能, 并与其他多种信用评分模型进行了对比。结果表明, AdaFocal-XGBoost在AUC、准确率、F1分数、Gmean、KS、精确率和召回率等关键指标上均优于其他模型, 特别是在处理严重不平衡数据集时表现出色。本研究不仅验证了集成学习与自适应损失函数结合的有效性, 也为信用评分领域提供了新的解决方案, 有助于金融机构提高融资效率和管控风险敞口。

关键词

信用评分, 不平衡数据, 集成学习

Research on AdaFocal-XGBoost Integrated Credit Scoring Model Based on Unbalanced Data

Kai Guo, Hong Fan*

Glorious Sun School of Business and Management, Donghua University, Shanghai

Received: Nov. 6th, 2024; accepted: Dec. 2nd, 2024; published: Dec. 9th, 2024

*通讯作者。

文章引用: 郭楷, 范宏. 基于不平衡数据的 AdaFocal-XGBoost 集成信用评分模型研究[J]. 统计学与应用, 2024, 13(6): 2204-2214. DOI: 10.12677/sa.2024.136214

Abstract

With the advent of the big data era, credit risk management has become increasingly important in the financial field, and credit scoring, as its core tool, faces the challenge of massive growth of customer credit data and dynamic changes in individual credit profiles. Traditional credit assessment methods are deficient in adaptability and flexibility, especially when dealing with unbalanced data. In this paper, we propose an AdaFocal-XGBoost integrated credit scoring model based on unbalanced data, aiming to improve the accuracy and adaptability of credit risk prediction. The AdaFocal-XGBoost model combines the efficient computation of XGBoost and the adaptive loss adjustment of AdaFocalLoss, which is especially optimized for the sample imbalance problem. Through experiments on four credit datasets (Australian, German, Japan, and Taiwan region) from the UCI database, this study comprehensively evaluates the performance of the AdaFocal-XGBoost model and compares it with various other credit scoring models. The results show that AdaFocal-XGBoost outperforms other models in key metrics such as AUC, accuracy, F1 score, Gmean, KS, precision, and recall, especially when dealing with severely unbalanced datasets. This study not only verifies the effectiveness of integrative learning combined with adaptive loss function, but also provides a new solution in the field of credit scoring, which can help financial institutions to improve the financing efficiency and control the risk exposure.

Keywords

Credit Scoring, Unbalanced Data, Ensemble Learning

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

信用风险管理对金融领域的高质量发展有着不可替代的作用, 信用评分作为银行和金融机构的关键工具, 是信用风险评估不可或缺的步骤。然而, 随着大数据时代的来临, 客户信用数据的海量增长与个体信用画像的动态变迁, 对传统静态信用评估方法构成了严峻挑战, 暴露出其在适应性与灵活性上的明显短板。根据 2024 年信贷市场行业风险调研, 部分信贷产品不良率上升, 如小额现金分期或消费分期贷款平均不良率为 3.98%, 较 2023 年上升了 1.97 个百分点。而信用评分作为银行和金融机构的关键管理工具, 在信用风险管理中占据核心地位。因此建立评判精准、适应信贷场景的信用风险预测模型既有助于提高融资效率, 也有助于管控金融机构的风险敞口。

最早应用于信用评分领域的统计方法基于数据服从特定概率分布的假设, 其中判别分析法和回归分析法是两种常见的代表性方法。章彤等[1]为筛选最优信用特征组合, 首先以各指标数值区间中违约比例与非违约比例的差值总和构建总信息值统计量, 继而基于弹性网回归模型进行特征选择。在此基础上, 运用线性判别分析(Linear Discriminant Analysis, LDA)实现信用评分, 以期获得更为精准的信用风险评估结果。Zhou 等[2]将单一特征选择扩展到特征组合选择, 并结合 LDA 算法推导出最优信贷特征组合。传统的统计方法因其简单、易于理解和实现等优点而被广泛应用于信用评分领域。然而, 这些方法通常建立在一些强假设前提之上, 例如多元正态性和同方差性等。在实际的信用评分场景中, 数据往往难以满足这些假设条件, 导致基于统计方法构建的模型预测性能受到限制, 难以有效应对日益复杂的信用风险

评估挑战。为了适应信贷业务快速发展的需求, 并克服传统参数模型在信用评级领域的局限性, 研究者们开始探索基于机器学习的信用评级方法。早期的研究主要关注单一的机器学习模型, 如支持向量机[3]、 k 近邻[4]和朴素贝叶斯[5]等。这些模型通过构建复杂的决策边界, 能够自动提取信用数据中的关键特征, 在一定程度上提升了信用评分的预测性能。然而一些统计模型假设特征与信用违约之间是线性关系, 但实际情况可能存在非线性关系, 这样的模型就无法准确捕捉特征与违约之间的复杂关系。因此本文将使用几类传统的统计方法作为对比的基准模型。

基于决策树及其集成算法的信用评级模型凭借其优异的可解释性、对数据质量的鲁棒性、较高的计算效率以及融合领域知识的能力, 在信用评估领域得到了广泛应用。康海燕等[6]利用 LightGBM 算法剔除冗余特征, 随后利用基分类器为决策树的数个集成学习算法构造 Stacking 个人信用评级模型。王重仁等[7]将贝叶斯优化应用于 XGBoost 算法中, 使得预测效果优于其他树集成模型算法。Xia 等[8]开发了一种基于树的过拟合谨慎异构集成模型, 根据过拟合度量动态地为基础模型分配权重, 从而提高了集成模型在信贷数据集中的表现。Mushava 等[9]提出了一种基于广义极值和指数逻辑分布的损失函数并将其嵌入 XGBoost 中, 降低了信用评级场景下的错分成本。Rao 等[10]在成本函数中采用熵权法引入加权马氏距离, 对成本敏感决策树基分类器采用加权投票从而评估借款人的信用风险。Wang 等[11]提出基于深度和浅层特征选择以及快照集成技术的可解释深度学习框架, 旨在提高金融市场预测的准确性。Zhang 等[12]尝试通过学习决策树来量化解释卷积网络做出的预测。由于 XGBoost 算法在信用评级领域的泛用性与性能都得到了广泛的认可, 因此本文将从 XGBoost 算法出发, 对基于 XGBoost 的集成模型做出改进以适应不平衡场景下的信用评级任务。

不平衡学习大致可以分为两类: 事前重采样法和算法内部改进方法。在事前层面, 常见的两种技术是随机过采样(ROS)和随机欠采样(RUS)。它们的目的是通过随机复制少数类中的一些实例, 从多数类中移除一些样本来平衡类分布。简单的 ROS 可能会导致过拟合, 因此各种改进的采样方案被开发出来。一种典型的方法是合成少数类过采样技术(SMOTE), 它通过随机选择一些少数类样本作为采样种子来增加少数类实例。然而, SMOTE 可能会生成错误的少数类实例。基于如何选择采样种子和合成有效的少数类实例, SMOTE 衍生出许多变体。刘婧怡等[13]使用 Borderline-SMOTE 算法, 将少数类样本划分为安全, 危险, 噪声区域, 针对危险区域样本合成新的少数类样本, 降低了过拟合的风险。周万珍等[14]使用自适应合成采样(ADASYN)自适应地增强或减弱少数样本作为采样种子的机会, 并将此方法与生成对抗网络(GAN)结合, 提高了生成的数据样本的质量并在不平衡指标上优于其他过采样方法。Douzas 等[15]使用自组织映射过采样(SOMO)产生少数类样本的二维表示, 以指导有效数据的生成。而 RUS 一方面忽略了大量训练数据并丢弃可能有用的观察结果, 另一方面, 这种信息丢失可能会扭曲多数样本的基本分布。为了克服 RUS 的缺点, 业界提出了众多智能欠采样技术。Batista 等[16]提出 Edited Nearest Neighbours (ENN) 算法, 通过对数据集中的待欠采样样本进行一系列编辑, 删除与其邻近样本相似度较低的样本以降低多数类样本比例。Kim 等[17]使用遗传算法训练了聚类模型, 其中远离聚类质心的样本被视为噪声, 随后删除了来自多数类的这些噪声示例, 从而产生了一种高效的欠采样方法。但此类技术依赖于复杂的技术来执行数据预处理, 在面对大规模数据集时, 这些技术无法很好地扩展。基于上述考量, 本文将使用算法内部改进方法对集成学习模型进行针对不平衡任务的改进。

本文的贡献主要体现在: (1) 基于自适应损失函数的信用评级模型, 显著提升了对少数类样本的识别能力。将 AdaFocalLoss 损失函数嵌入树集成算法 XGBoost 算法构建了的 AdaFocal-XGBoost 模型能够根据模型在验证集上的表现动态调整损失函数的参数, 从而更有效地关注难以识别的违约用户, 显著提升了模型对这些高风险用户的识别能力。(2) 严谨的实验设计和多维度指标评估, 验证了 AdaFocal-XGBoost 模型的性能优势。选取了多个具有代表性的公开信贷数据集, 并采用了 AUC、KS 值等多个常用信用评

分模型评价指标,对 AdaFocal-XGBoost 模型与逻辑回归、支持向量机等传统统计学习方法,以及 GBDT、LightGBM、AdaBoost 等主流集成学习模型进行了全面的比较分析。实验结果表明,AdaFocal-XGBoost 模型在多个数据集上均取得了显著的性能提升,特别是在违约用户识别方面表现更为突出。(3) 为信用风险管理实践提供了新的方法论指导和可操作的工具。AdaFocal-XGBoost 模型的提出,不仅丰富了信用评分模型的理论体系,更重要的是为金融机构的信用风险管理提供了新的方法论指导和可操作的工具。该模型能够更有效地识别高风险客户,从而帮助金融机构更精准地进行风险定价、授信额度控制等风险管理决策,最终提升其风险管理的整体效能。

2. 模型及理论基础

2.1. XGBoost

XGBoost 的核心思想是将一个强预测模型建立为多个弱预测模型的线性组合,这些弱预测模型在信用评分场景下一般设置为决策树。通过迭代地添加新的弱模型,模型的性能逐步提高。XGBoost 定义了一个目标函数(Objective Function),它由两部分组成:训练数据的损失函数(Loss Function)和正则化项(Regularization Term)。目标函数如下:

$$Obj = L(\theta) + \Omega(\theta) \quad (1)$$

其中, $L(\theta)$ 是损失函数,衡量模型预测值与真实值之间的差异; $\Omega(\theta)$ 是正则化项,用于控制模型的复杂度,防止过拟合。

对于第 i 个训练样本,损失函数可以表示为:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) \quad (2)$$

其中, y_i 是真实值, \hat{y}_i 是模型预测值, l 是损失函数,常用的损失函数包括均方误差(MSE)和对数似然损失(Logistic Loss)。正则化项通常包括模型中树的数量和每棵树的复杂度:

$$\Omega(\theta) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (3)$$

其中, T 是树的数量, w_j 是第 j 棵树的权重, γ 和 λ 是正则化参数。

XGBoost 使用贪心算法来选择每一步的分裂点,以最大化负梯度的减少量。对于第 k 轮迭代,XGBoost 会添加一个新的树 $\phi(x; \theta_k)$ 来更新模型:

$$\theta_{k+1} = \arg \min_{\theta} \left[L(\theta_k) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 + \sum_{i=1}^n \ell(\hat{y}_i + \phi(x_i; \theta)) \right] \quad (4)$$

其中, ℓ 是梯度提升的目标函数, $\phi(x_i; \theta)$ 是新加入的树模型。由于目标函数不可微分, XGBoost 使用泰勒展开来近似二阶导数:

$$\ell(\hat{y}_i + \phi(x_i; \theta)) \approx \ell(\hat{y}_i) + \ell'(\hat{y}_i) \phi(x_i; \theta) + \frac{1}{2} \ell''(\hat{y}_i) \phi(x_i; \theta)^2 \quad (5)$$

这样, XGBoost 就可以通过求解一个二次优化问题来找到最佳的分裂点。

2.2. AdaFocalLoss

AdaFocalLoss 是基于 FocalLoss 改良的损失函数。FocalLoss 最初是为了提高分类器性能而提出的,通过减少交叉熵损失中易于分类样本的权重,允许对难以分类样本进行强化训练。FocalLoss 通过引入一

个动态缩放因子, 可以动态降低训练过程中易区分样本的权重, 从而将重心快速聚焦在那些难区分的样本上。基础的 FocalLoss 损失函数可以表示为:

$$FL(p_i) = -\alpha_i (1 - p_i)^\mu \log(p_i) \quad (6)$$

其中 p_i 是模型对于一个样本属于某个类别的预测概率, 如果该样本是正样本, 则 p_i 是模型预测为正样本的概率; 如果该样本是负样本, 则 p_i 是模型预测为负样本的概率。 α_i 是平衡正负样本的权重系数, 用于处理类别不平衡问题。对于正样本, α_i 通常是一个常数(如 0.25), 对于负样本, α_i 是 $1 - \alpha_i$ 。 μ 是调节易分类样本权重的聚焦参数(focal parameter), 其值通常在 0 到 5 之间。当 $\mu = 0$ 时, Focal Loss 退化为标准的交叉熵损失。FocalLoss 的核心思想是当 p_i 接近 1 时, 即样本容易被分类时, 通过 $(1 - p_i)^\mu$ 这一项减少该样本的损失贡献, 而当 p_i 接近 0 时, 即样本难以分类时, 增加该样本的损失贡献。通过这种方式, 模型被鼓励更多地关注难分类的样本。

AdaFocalLoss 的核心在于自适应地调整 FocalLoss 中的超参数 μ 。AdaFocalLoss 会根据验证数据中观察到的校准误差调整 γ :

$$\mu_t = \mu_{t-1} \cdot \exp(\varepsilon(C_{val,b} - A_{val,b})) \quad (7)$$

其中, μ_t 是当前步骤的 μ 值, μ_{t-1} 是前一步骤的 μ 值, $C_{val,b}$ 是验证集中第 b 个样本组的平均概率, $A_{val,b}$ 是第 b 个样本组中正确答案的百分比, ε 是一个超参数, 决定了每次更新对 μ 的调整程度。

3. AdaFocal-XGBoost 集成信用评分模型

以 XGBoost 为代表的基于树模型的集成学习模型在信用评分领域取得了较好的结果, 但在样本严重失衡的情况下, 集成模型对于少数类样本的分类效果存在一定不足。而在不平衡处理时常用的事前重采样法较为依赖超参数的选取, 在面对不同数据集时需要频繁调整超参数, 技术无法很好地进行扩展。因此本文提出将集成算法与自适应调整损失函数相结合的 AdaFocal-XGBoost 信用评分模型, 模型主要包括三个部分: (1) 使用 XGBoost 首先处理不平衡数据, 得出验证集结果; (2) 根据验证集数据中观察到的误差调整 AdaFocalLoss 中的 γ 值; (3) 在下一迭代中使用更新过参数的 AdaFocalLoss 继续进行训练, 直至模型完全收敛, 得出最终的预测结果。

模型训练过程如下:

输入: 初始样本集;

(1) 初始化模型参数 θ_0 ;

(2) 对于第 k 次迭代, 给定当前模型参数 θ_{k-1} , 添加一个新的树模型 $\phi(x; \theta_k)$ 来更新模型:

$$\theta_k = \theta_{k-1} + \sum_{i=1}^n \omega_i f_i(x; \theta) \quad (8)$$

其中, $f_i(x; \theta)$ 是第 i 棵树的预测函数, ω_i 是该树模型的权重;

(3) 将迭代后的模型用于验证集, 调整模型的损失函数:

$$FL(p_i) = -\alpha_i (1 - p_i)^{\mu_k} \log(p_i) = -\alpha_i (1 - p_i)^{\mu_{k-1} \cdot \exp(\varepsilon(C_{val,b} - A_{val,b}))} \log(p_i) \quad (9)$$

(4) 根据增加的子模型和计算得出的梯度更新目标函数:

$$Obj = \sum_{i=1}^n FL(p_i) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (10)$$

通过梯度下降法和贪心算法结合生成新的树补充入集成模型; 对各个弱学习器分配各自的权重, 累

加得到最终的结果:

$$\hat{y} = \sum_{k=1}^K \sum_{i=1}^{n_k} \omega_i f_i(x; \theta_k) \quad (11)$$

输出: 违约概率, 大于等于 0.5 则判定为违约, 小于 0.5 则判定为不违约。

4. 性能实验与结果分析

4.1. 数据描述与处理

本文共采用 UCI 数据库中的 4 个信贷数据集: Australian、German、Japan 和 Taiwan 地区。关于四个信用评分数据集的详细信息见表 1。

Table 1. Dataset information

表 1. 数据集信息

数据集	样本个数	特征数	履约/违约数	数据集
Australian	690	14	383/307	Australian
German	1000	20	700/300	German
Japan	690	14	383/307	Japan
Taiwan region	6000	17	5000/1000	Taiwan region

为了全面评估 AdaFocal-XGBoost 模型在信用评分任务上的有效性, 本文对多个公开的信贷数据集进行了实验研究。为确保模型评估的客观性和可重复性, 每个数据集均按 8:1:1 的比例随机划分为训练集、验证集和测试集。AdaFocal-XGBoost 模型利用 80% 的训练集进行参数学习和优化, 并通过 10% 的验证集进行超参数优化, 以防止过拟合。最终, 模型在剩余 10% 的测试集上进行评估, 以衡量其在信用评分任务上的泛化性能和实际效果。

4.2. 对比模型及性能评估指标

为全面评估各类信用评分模型的性能与适用性, 本文构建了一个广泛且深度的模型对比框架, 旨在从不同角度洞察模型在信用评分领域中的表现与潜力。基准模型包括基于统计的信用评分方法: 线性判别分析 LDA 和逻辑回归 LR; 基于传统机器学习算法: 决策树 DT、K 近邻 KNN、支持向量机 SVM 和神经网络 NN; 基于集成学习的信用评分技术: AdaBoost、梯度提升树 GBDT、LightGBM 以及本文模型用到的基础模型 XGBoost; 基于不平衡采样的集成模型: BalanceCascade、KmeansSMOTEBoost 以及 OverBagging。线性判别分析和逻辑回归是统计学中的经典方法, 它们在信用评分领域有着悠久的历史, 选择这两种模型是为了提供一个基准, 展示传统统计方法在当前数据集上的表现; K 近邻、决策树以及神经网络是简单易实现的机器学习模型, 比较这些模型的性能可以评估在没有复杂特征工程和模型调优的情况下的机器学习算法在信用评分任务上的表现; 集成学习模型的引入是为了研究在信用评分中, 集成方法是否能够提供比单一模型更好的预测能力。而本文使用的三种不平衡集成对比算法都基于事前重采样方法, 通过比较这三个模型与 AdaFocal-XGBoost 的性能, 可以探索事中自适应调整损失函数与事前重采样方法在处理不平衡数据时的优劣势。

本文使用七种信用评分常用的评价指标: AUC 值、准确率 ACC、F₁ 分数、Gmean、KS、精确率 Prec 以及召回率 Recall。AUC 和 ACC 适合评估二分类问题下的模型的整体性能, 而 F₁、Gmean 适用于类别不平衡问题的总体表现; KS 用于衡量模型区分不同风险等级(如好坏客户)的能力, KS 值越大, 模型的区

分能力越强; 精确率和召回率适用于错判成本较高的场景, 在信用评分中漏筛一个违约样本的成本远高于对一个不违约样本的误判。

4.3. 实验结果分析

为了进行直观反映各信用评分算法的性能, 首先通过 ROC 进行图形化的性能比较。ROC 曲线是一种广泛用于评估信用评分性能的图形指标, 其 x 轴表示假阳性率(False Positive Rate, FPR), y 轴表示真阳性率(True Positive Rate, TPR)。图 1 展示了不同信用评分模型在四个数据集(Australian、German、Japan 和 Taiwan 地区)上的 ROC 曲线。AdaFocal-XGBoost 模型在 Taiwan 地区数据集上表现显著优于其他模型; 在 Australian 和 Japan 数据集上, AdaFocal-XGBoost 虽优势相对较小, 但其 ROC 曲线下面积(AUC)仍略大于其他模型。在 German 数据集上, AdaFocal-XGBoost 与 BalanceCascade 的性能相近, 但 AdaFocal-XGBoost 的 AUC 值分别略高于 BalanceCascade。值得注意的是, KNN 模型在 German、Japan 和 Taiwan 地区数据集上的 AUC 值排名均较为靠后。这可能是由于 KNN 基于相似度进行预测, 而信贷数据集本身存在类别不平衡的问题, 导致其预测结果偏向于多数类无风险贷款, 从而影响了对于少数类风险贷款的识别能力。

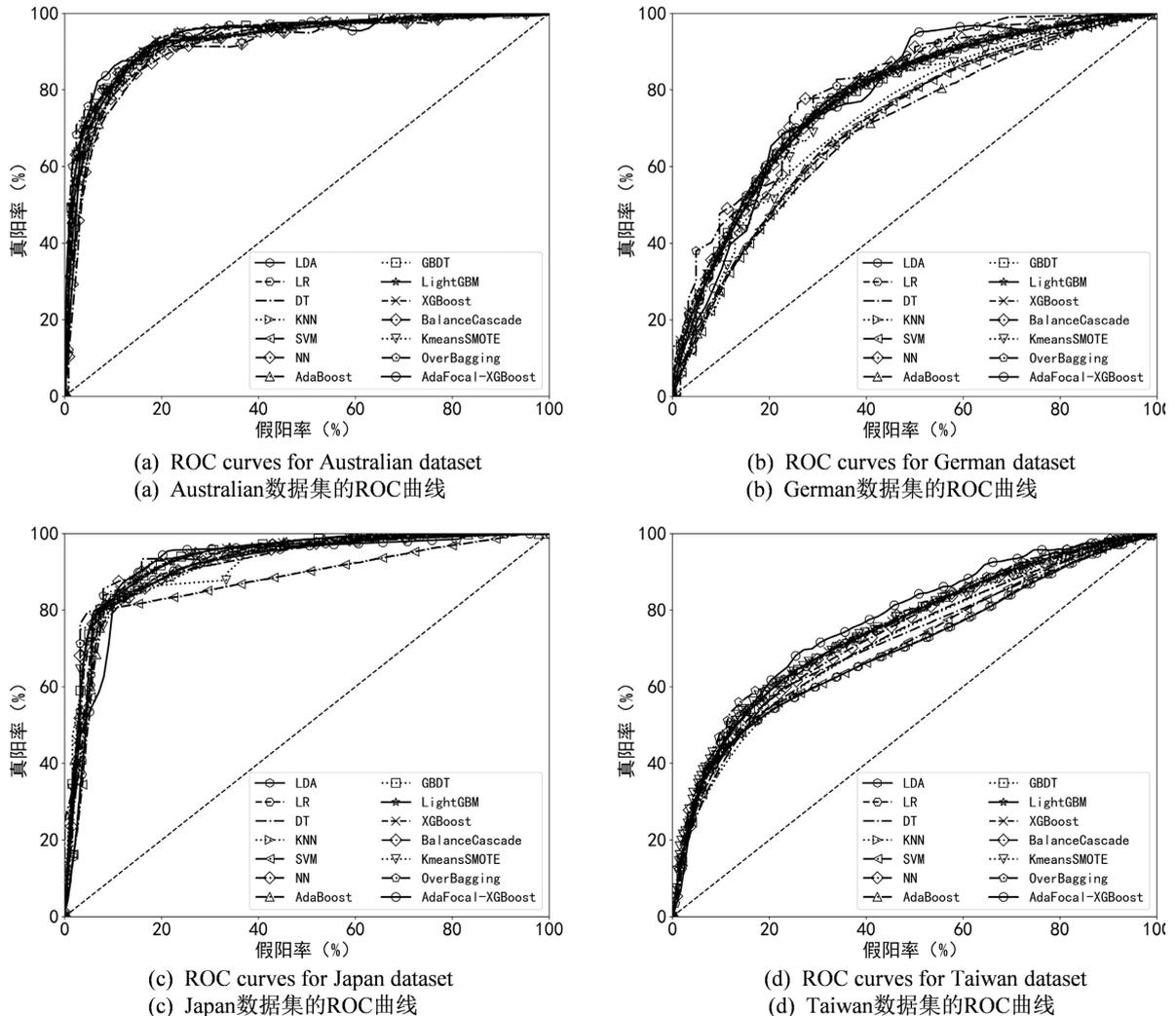


Figure 1. Comparison of AUC on different datasets
图 1. 不同数据集上的 AUC 对比

为深入探究模型性能差异, 表 2 列出了各信用评分模型在 Australian 数据集上的量化对比结果。AdaFocal-XGBoost 模型在 AUC、Acc、Gmean 和 KS 指标上都取得了最佳性能, 证明其在信用评分方面具有显著优势。LDA 在 F1 和 Rec 指标上取得了最优, 但其 Prec 指标是唯一低于 0.8 的, 这表明模型对非违约样本的识别性能不足, 导致模型在 AUC、Acc 等模型整体性能指标上表现不佳。

Table 2. Performance comparison on Australian dataset
表 2. 在 Australian 数据集上的性能比较

模型	AUC	Acc	F1	GMean	KS	Prec	Rec
LDA	0.92688	0.85942	0.85339	0.86369	0.73078	0.79607	0.91961
LR	0.92977	0.86487	0.85199	0.86575	0.73158	0.83094	0.87414
DT	0.91402	0.84368	0.82360	0.83549	0.68270	0.82704	0.82020
KNN	0.91336	0.84939	0.82265	0.82049	0.68602	0.86399	0.78508
SVM	0.92622	0.86255	0.84461	0.85257	0.72054	0.84973	0.83954
NN	0.91476	0.85017	0.83133	0.84778	0.69631	0.83282	0.82984
AdaBoost	0.92733	0.85554	0.84963	0.85986	0.72332	0.79128	0.91726
GBDT	0.93922	0.86371	0.84778	0.86260	0.72530	0.84260	0.85303
LightGBM	0.93713	0.86238	0.84485	0.86018	0.72074	0.84756	0.84215
XGBoost	0.93942	0.86333	0.84676	0.86178	0.72376	0.84487	0.84866
BalanceCascade	0.92847	0.85990	0.82209	0.85363	0.70811	0.81707	0.82716
KmeansSMOTEBoost	0.92367	0.86957	0.82581	0.85289	0.71076	0.86486	0.79012
OverBagging	0.94023	0.88406	0.83636	0.86628	0.73280	0.85185	0.85185
AdaFocal-XGBoost	0.94827	0.87440	0.84849	0.87646	0.75309	0.82716	0.82753

表 3 展示了各种信用评分模型在 German 数据集上的性能比较。AdaFocal-XGBoost 在该数据集上的表现优异, 在五个评价指标上取得最佳结果。AdaBoost 在召回率指标上表现出色, 但其 Prec、AUC、Acc、KS 指标均为最差, 对非违约样本预测能力极差。

Table 3. Performance comparison on German dataset
表 3. 在 German 数据集上的性能比较

模型	AUC	Acc	F1	GMean	KS	Prec	Rec
LDA	0.77954	0.75846	0.83719	0.63756	0.34534	0.79256	0.88714
LR	0.78084	0.76012	0.83813	0.64133	0.35080	0.79421	0.88720
DT	0.70962	0.72324	0.81020	0.59916	0.28562	0.77911	0.84389
KNN	0.73835	0.72798	0.83146	0.42669	0.14850	0.73412	0.95857
SVM	0.71123	0.70650	0.78813	0.36659	0.31523	0.79660	0.77983
NN	0.77994	0.76592	0.83961	0.66778	0.38602	0.80674	0.87529
AdaBoost	0.70350	0.70208	0.82368	0.14356	0.01482	0.70314	0.99409
GBDT	0.77916	0.75870	0.83877	0.62583	0.33346	0.78790	0.89666
LightGBM	0.77758	0.76152	0.84095	0.62723	0.33749	0.78865	0.90069

续表

XGBoost	0.78113	0.75816	0.84204	0.59045	0.29943	0.77567	0.92083
BalanceCascade	0.79360	0.77500	0.84099	0.70760	0.44296	0.82069	0.86232
KmeansSMOTEBoost	0.74626	0.75500	0.82310	0.70213	0.42286	0.82014	0.82609
OverBagging	0.80306	0.77000	0.84354	0.65938	0.38242	0.79487	0.89855
AdaFocal-XGBoost	0.80486	0.78500	0.83004	0.76117	0.53099	0.83395	0.84677

表 4 所示为各种信用评分模型在 Japan 数据集上的性能对比。AdaFocal-XGBoost 在五个指标中都取得了最优, 对正负类样本的预测性能都较为优秀。不平衡集成模型在预测违约样本上性能均较为优秀, 召回率数值较高。BalanceCascade 的 F_1 值在所有模型中最高, 同样是较为值得信赖的模型。LDA 对多数类样本识别较为准确, 但召回率较低, 对多数类和少数类样本的判别能力较为不均衡。

Table 4. Performance comparison on Japan dataset

表 4. 在 Japan 数据集上的性能比较

模型	AUC	Acc	F1	GMean	KS	Prec	Rec
LDA	0.91265	0.86058	0.86427	0.86542	0.73623	0.94020	0.79969
LR	0.91560	0.85493	0.86132	0.85891	0.72057	0.91748	0.81164
DT	0.91344	0.84901	0.86292	0.84596	0.69627	0.86981	0.85614
KNN	0.91109	0.84867	0.85959	0.85027	0.70084	0.88621	0.83452
SVM	0.86821	0.85658	0.86109	0.86108	0.72696	0.93115	0.80084
NN	0.91773	0.84868	0.85926	0.85032	0.70180	0.88950	0.83102
AdaBoost	0.92105	0.85484	0.85940	0.85945	0.72344	0.92933	0.79927
GBDT	0.93621	0.86415	0.87590	0.86436	0.72873	0.88977	0.86245
LightGBM	0.93494	0.86340	0.87620	0.86259	0.72523	0.88289	0.86961
XGBoost	0.93621	0.86784	0.87887	0.86849	0.73702	0.89586	0.86251
BalanceCascade	0.93481	0.87681	0.88435	0.87771	0.75556	0.90278	0.86667
KmeansSMOTEBoost	0.92053	0.85507	0.85915	0.85783	0.71810	0.91045	0.81333
OverBagging	0.93788	0.86232	0.87075	0.86312	0.72635	0.88889	0.85333
AdaFocal-XGBoost	0.93866	0.88406	0.88060	0.88368	0.76765	0.89394	0.86765

表 5 给出了各种信用评分模型在 Taiwan 地区数据集上的性能对比。Taiwan 地区数据集的数据不平衡现象较为严重, 基于统计的算法和传统机器学习算法表现较差, 而集成模型、不平衡集成模型以及 AdaFocal-XGBoost 在不平衡场景下保持了良好的表现, AdaFocal-XGBoost 在六个指标中都位列第一, 展现了模型对于不平衡数据集较强的预测能力。

Table 5. Performance comparison on Taiwan region dataset

表 5. 在 Taiwan 地区数据集上的性能比较

模型	AUC	Acc	F1	GMean	KS	Prec	Rec
LDA	0.69845	0.65116	0.63327	0.64933	0.30234	0.66763	0.60228
LR	0.69988	0.64862	0.63449	0.64747	0.29726	0.66120	0.60986

续表

DT	0.71990	0.66657	0.64910	0.65889	0.33315	0.68511	0.61669
KNN	0.71690	0.66755	0.62541	0.65800	0.33513	0.71635	0.55495
SVM	0.70582	0.67314	0.61010	0.63649	0.34634	0.75608	0.51136
NN	0.73771	0.68058	0.65426	0.67619	0.36118	0.71314	0.60437
AdaBoost	0.71697	0.67278	0.60701	0.65163	0.34561	0.75990	0.50533
GBDT	0.74961	0.69477	0.66976	0.69063	0.38956	0.72968	0.61893
LightGBM	0.74942	0.69540	0.67123	0.69150	0.39082	0.72922	0.62179
XGBoost	0.75038	0.69453	0.66907	0.69026	0.38909	0.73006	0.61749
BalanceCascade	0.75138	0.68667	0.66607	0.68397	0.37179	0.70093	0.63452
KmeansSMOTEBoost	0.75232	0.68932	0.65909	0.68972	0.39671	0.74839	0.58883
OverBagging	0.75676	0.69333	0.67318	0.69066	0.38513	0.70841	0.64129
AdaFocal-XGBoost	0.76761	0.69833	0.68483	0.70620	0.41876	0.73346	0.64225

综合以上四个数据集的表现可以看出, 引入 AdaFocalLoss 后, AdaFocal-XGBoost 模型的整体表现要优于原始 XGBoost 方法, 尤其是在召回率、KS 等针对少数类样本分类性能的指标上提升较大。同时, 在正负类不平衡现象较轻的 German、Japan 数据集上的性能差异要小于 Taiwan 地区这类不平衡现象较严重的数据集。这表明 AdaFocalLoss 的引入增强了对于不平衡场景下模型的表现, AdaFocal-XGBoost 是信用评分领域值得信赖的解决办法。而通过对比框架中四类不同算法的对比可以得出, 统计和传统机器学习模型的整体表现不及集成模型以及非平衡集成模型, 集成模型技术在信用评分场景下表现优秀。

5. 总结

随着金融科技的迅速发展, 信用风险管理在银行和金融机构中的重要性愈加凸显。本文提出的基于不平衡数据的 AdaFocal-XGBoost 集成信用评分模型, 旨在应对传统信用评分方法在处理不平衡数据时的局限性。通过引入自适应调整的损失函数, 模型能够更有效地关注难以分类的少数类样本, 从而提升信用评分的准确性和可靠性。实验结果表明, AdaFocal-XGBoost 模型在多个公开信贷数据集上的表现均优于传统统计方法和其他机器学习模型, 尤其在 AUC、F1 分数、召回率等关键指标上展现出显著优势。这一成果不仅验证了集成学习与自适应损失函数结合的有效性, 也为信用评分领域提供了新的思路和方法, 在实际应用中能够为金融机构的风险管理和决策提供更为精准的支持。

未来的研究可以进一步探索 AdaFocalLoss 在其他复杂信贷数据集、企业风险评估等场景中的应用。此外, 随着数据量的不断增加和模型复杂性的提升, 如何在保证模型性能的同时, 提升模型的可解释性, 将是一个值得深入探讨的方向。

基金项目

国家自然科学基金(71971054), 项目名称: 宏观经济波动下的动态复杂银行网络系统稳定性及宏观审慎监管研究。

参考文献

- [1] 章彤, 迟国泰. 基于最优信用特征组合的违约判别模型——以中国 A 股上市公司为例[J]. 系统工程理论与实践, 2020, 40(10): 2546-2562.
- [2] Zhou, Y., Chi, G., Liu, J., Xiong, J. and Wang, B. (2022) Default Discrimination of Credit Card: Feature Combination

- Selection Based on Improved FDAF-Score. *Expert Systems with Applications*, **206**, Article 117829. <https://doi.org/10.1016/j.eswa.2022.117829>
- [3] 陆爱国, 王珏, 刘红卫. 基于改进的 SVM 学习算法及其在信用评分中的应用[J]. 系统工程理论与实践, 2012, 32(3): 515-521.
- [4] Shekhar, S., Hoque, N. and Bhattacharyya, D.K. (2022) PKNN-MIFS: A Parallel KNN Classifier over an Optimal Subset of Features. *Intelligent Systems with Applications*, **14**, Article 200073. <https://doi.org/10.1016/j.iswa.2022.200073>
- [5] Kim, T. and Lee, J. (2023) Maximizing AUC to Learn Weighted Naive Bayes for Imbalanced Data Classification. *Expert Systems with Applications*, **217**, Article 119564. <https://doi.org/10.1016/j.eswa.2023.119564>
- [6] 康海燕, 胡成倩. 基于特征提取和集成学习的个人信用评分方法[J]. 计算机仿真, 2024, 41(1): 311-320.
- [7] 王重仁, 韩冬梅. 基于超参数优化和集成学习的互联网信贷个人信用评估[J]. 统计与决策, 2019, 35(1): 87-91.
- [8] Xia, Y., Zhao, J., He, L., Li, Y. and Niu, M. (2020) A Novel Tree-Based Dynamic Heterogeneous Ensemble Method for Credit Scoring. *Expert Systems with Applications*, **159**, Article 113615. <https://doi.org/10.1016/j.eswa.2020.113615>
- [9] Mushava, J. and Murray, M. (2024) Flexible Loss Functions for Binary Classification in Gradient-Boosted Decision Trees: An Application to Credit Scoring. *Expert Systems with Applications*, **238**, Article 121876. <https://doi.org/10.1016/j.eswa.2023.121876>
- [10] Rao, C., Liu, M., Goh, M. and Wen, J. (2020) 2-Stage Modified Random Forest Model for Credit Risk Assessment of P2P Network Lending to “Three Rurals” Borrowers. *Applied Soft Computing*, **95**, Article 106570. <https://doi.org/10.1016/j.asoc.2020.106570>
- [11] Wang, J. and Dong, Y. (2024) An Interpretable Deep Learning Multi-Dimensional Integration Framework for Exchange Rate Forecasting Based on Deep and Shallow Feature Selection and Snapshot Ensemble Technology. *Engineering Applications of Artificial Intelligence*, **133**, Article 108282. <https://doi.org/10.1016/j.engappai.2024.108282>
- [12] Zhang, Q., Yang, Y., Ma, H. and Wu, Y.N. (2019) Interpreting CNNs via Decision Trees. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 6254-6263. <https://doi.org/10.1109/cvpr.2019.00642>
- [13] 刘婧怡, 卢胜男. 基于自适应 Borderline-SMOTE 过采样的 LightGBM 不平衡数据分类算法[J]. 信息技术与信息化, 2024(6): 205-208.
- [14] 周万珍, 盛媛媛, 张永强, 等. 基于 ADASYN 和 WGAN 的混合不平衡数据处理方法[J]. 河北工业科技, 2024, 41(4): 291-298.
- [15] Douzas, G., Rauch, R. and Bacao, F. (2021) G-SOMO: An Oversampling Approach Based on Self-Organized Maps and Geometric Smote. *Expert Systems with Applications*, **183**, Article 115230. <https://doi.org/10.1016/j.eswa.2021.115230>
- [16] Batista, G.E.A.P.A., Prati, R.C. and Monard, M.C. (2004) A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. *ACM SIGKDD Explorations Newsletter*, **6**, 20-29. <https://doi.org/10.1145/1007730.1007735>
- [17] Kim, H., Jo, N. and Shin, K. (2016) Optimization of Cluster-Based Evolutionary Undersampling for the Artificial Neural Networks in Corporate Bankruptcy Prediction. *Expert Systems with Applications*, **59**, 226-234. <https://doi.org/10.1016/j.eswa.2016.04.027>