

基于 BO-XGBoost 与集成学习方法的供应链金融信用评价研究

顾天下, 刘勤明, 叶春明

(上海理工大学 管理学院, 上海 200093)

摘要: 针对供应链金融领域中小企业融资的信用风险控制问题, 提出了一种在 Bagging 算法框架下结合贝叶斯优化和 XGBoost 算法的集成学习模型 BO-XGBoost-Bagging(BXB)。首先, 基于 XGBoost 特征重要度进行特征筛选, 建立供应链金融信用评价指标体系。其次, 通过贝叶斯优化获得 XGBoost 的最优超参数, 并结合 Bagging 算法得到集成模型 BXB。最后, 在中小企业数据集上进行预测, 通过实证研究验证信用评价模型的有效性。实证结果表明, BXB 模型相比其他模型具有更好的预测效果, 能够更加准确、全面地对中小企业的信用风险进行评估, 更好地区分风险企业和正常企业, 最大程度减少违约损失, 在供应链金融信用评价方面有着较高的应用价值。

关键词: 信用评价; XGBoost 算法; 贝叶斯优化; 集成学习; 中小企业

中图分类号: F 832.4 **文献标志码:** A

Credit evaluation of supply chain finance based on BO-XGBoost and ensemble learning method

GU Tianxia, LIU Qinming, YE Chunming

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: To solve the problem of the credit risk control of the small and medium-sized enterprises(SMEs) financing in the supply chain finance field, an ensemble learning model BO-XGBoost-Bagging (BXB) that combines Bayesian optimization and XGBoost was proposed under the Bagging framework. Firstly, based on the XGBoost feature importance, the feature screening was carried out, and the financial credit evaluation index system for the supply chain was established. Secondly, the optimal super parameters of XGBoost were obtained by Bayesian optimization, and the integrated model BXB was obtained by bagging. Finally, the prediction was performed on SMEs data set, and the effectiveness of the credit evaluation model was verified by empirical research. The empirical results show that the BXB model has a better predictive effect than other models and can

收稿日期: 2021-10-27

基金项目: 国家自然科学基金资助项目(71632008, 71840003); 上海市自然科学基金资助项目(19ZR1435600); 教育部人文社会科学研究规划基金资助项目(20YJAZH068); 上海理工大学科技发展项目(2020KJFZ038); 2020 年上海理工大学大学生创新创业训练计划项目(SH2020067)

第一作者: 顾天下(1998-), 男, 硕士研究生。研究方向: 机器学习、供应链金融等。E-mail: gtx980801@163.com

通信作者: 刘勤明(1984-), 男, 教授。研究方向: 维护调度、人工智能等。E-mail: lqm0531@163.com

evaluate the credit risk of SMEs more exactly and comprehensively. The model can better distinguish between risky companies and normal companies, and minimize default losses. It has high application value in the credit evaluation of supply chain finance.

Keywords: credit evaluation; XGBoost; Bayesian optimization; ensemble learning; SMEs

近几年来,供应链金融呈现出巨大的发展潜力,成为中小企业解决融资难、融资贵、融资慢等问题的有效途径,受到我国政府的高度关注和支持^[1]。中小企业是国民经济发展的基础,但其发展过程中融资难的问题一直比较突出^[2]。供应链金融可有效解决处于供应链上下游的中小企业基数大却融资难的问题,成为中小企业的重要融资方式之一,但各种金融风险也随之产生。其中,最大的风险问题是中小企业的违约问题,因此对中小企业的信用风险进行控制至关重要。目前,信用评价已成为金融风险管理的重要方式,通过建立评价模型,根据融资企业的历史数据进行预测,可有效进行风险评估和控制^[3]。

当前,越来越多的信用风险研究借助机器学习方法建立模型,例如逻辑回归^[4-7]、K近邻^[8]、随机森林^[9-10]、支持向量机^[11-13]以及集成学习算法^[14]等。Yang^[15]基于支持向量机对企业信贷进行风险控制,节约了成本并提高了效率。Ye^[16]结合逻辑回归、决策树和支持向量机,建立集成模型进行风险控制,实证表明集成模型具有更好的稳定性和预测效果。然而目前大部分关于信用评价的研究都仅关注算法的应用,基于超参数优化的研究较少。李欣等^[17]提出了改进的网格搜索法优化XGBoost算法,缩短了调参时间,同时也提高了模型的F1值和几何平均值。王重仁等^[18]通过实证研究发现,贝叶斯优化优于传统的超参数优化算法,建立的分类模型在预测效果方面也优于对比模型。

中小企业的违约风险问题是供应链金融发展的重要阻碍, Liu等^[19]用支持向量机结合AdaBoost算法建立模型,并通过实证表明其提出的算法在中小企业信用风险预测方面具有高准确率。刘平山等^[20]结合医药行业的特点,构建了基于GBDT (gradient boosting decision tree) 的信用风险评估模型,通过数据分析证明该模型在医药供应链金融的信用评价中具有优越性。Zhang等^[21]基于改进的随机森林建立信用评价模型,进一步

提高模型性能,对中小企业进行信用风险控制。Zhang等^[22]利用萤火虫算法优化支持向量机,建立预测模型,并验证了其有效性。

目前,大部分信用评价研究都关注于个人信贷领域,或脱离供应链金融情况下的企业信贷领域,研究供应链金融模式下中小企业融资问题的较少,对供应链金融风险控制的研究也过于简单和片面。因此,需要建立完善的信用风险控制体系,对风险评估和控制制订有效的措施。

1 理论基础

1.1 XGBoost 算法

XGBoost (eXtreme gradient boosting) 算法为Chen等^[23]提出的基于Boosting算法改进的机器学习模型。该模型具有准确率高、运行速度快等优势,也可以在不进行插补预处理的情况下处理缺失值^[24]。XGBoost在求解过程中运用了二阶泰勒公式,可以更加准确、迅速地得到目标函数最优解。XGBoost将每颗树分值的和作为样本的预测值,最终预测函数为

$$\hat{y}_i^{(n)} = \sum_{k=1}^n f_k(x_i) = \hat{y}_i^{(n-1)} + f_n(x_i) \quad (1)$$

XGBoost的目标函数为损失函数和正则化惩罚项的结合,在一定程度上防止了过拟合,具体函数为

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

$$L^{(n)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(n-1)} + f_n(x_i)) + \Omega(f_n) + C \quad (3)$$

式中: $l(\cdot)$ 为损失函数; $\Omega(\cdot)$ 为正则项; C 为常数。

正则化项由下面两部分组成:

$$\Omega(f_i) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (4)$$

式中: ω_j 为叶子节点权重; λ 和 γ 为惩罚系数;

T 为树的个数。

通过二阶泰勒公式展开，得到如下函数：

$$L^{(n)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(n-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_i) + C \tag{5}$$

式中， g_i ， h_i 分别为一阶和二阶导数。

将正则化项代入式(5)，并去掉常数项，得到

$$L^{(n)} \approx \sum_{j=1}^T \left[\left(\sum g_i \right) \omega_j + \frac{1}{2} \left(\sum h_i + \lambda \right) \omega_j^2 \right] + \gamma T \tag{6}$$

令 $G_j = \sum g_i$ ， $H_j = \sum h_i$ ，当 $\omega_j = -\frac{G_j}{H_j + \lambda}$ 时，目标函数取得如下最小值：

$$L^{(n)} = \frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \tag{7}$$

XGboost 树生成的分裂点的方式由式(8)确定，即

$$\mathcal{L}_{\text{split}} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \tag{8}$$

式中： $\mathcal{L}_{\text{split}}$ 为结构评分结果， $\mathcal{L}_{\text{split}}$ 越大，表明切分后的目标函数值越低，即效果越好； G_L ， H_L ， G_R ， H_R 表示分裂后不同子树的相应值。

1.2 集成学习

集成学习的主要思想是将多个基础模型作为基分类器组合起来进行集成，从而提高模型的性能。集成学习可以将一系列相对弱的分类器进行组合，从而生成一个强分类器，提高模型的准确性。集成学习主要有 3 个类型：Bagging、Boosting 和 Stacking。

a. Bagging。

在 Bagging 算法中，通过 bootstrap 方法重抽样生成 M 个子数据集，在每个子数据集上继续训练预测，最后的预测结果根据 M 个模型的输出，采取投票的方式获得。

b. Boosting。

Boosting 算法是一种减小偏差的集成算法。其中最常用的是 AdaBoost 算法，其根据基分类器的表现不断更新权重，在训练下一个基分类器时，增加错误分类样本的权重，从而迭代生成 Boosting 集成模型。

c. Stacking。

Stacking 算法是一种分层的集成算法。大多数情况下分为两层，第一层是多个基分类器，通过

对原始样本的训练预测进行输出，将输出值作为第二层模型的输入再进行一次分类预测，从而生成 Stacking 集成模型，进一步提高模型的预测能力。

2 基于 BO-XGBoost-Bagging 的信用评价模型

本文针对中小企业的信用风险问题进行研究，提出了在 Bagging 框架下结合贝叶斯优化和 XGBoost 的信用评价模型 BO-XGBoost-Bagging (BXB)，从高维数据中筛选重要信息，从而减小中小企业违约带来的损失，对信用风险进行有效监管。该模型的大致流程如图 1 所示。

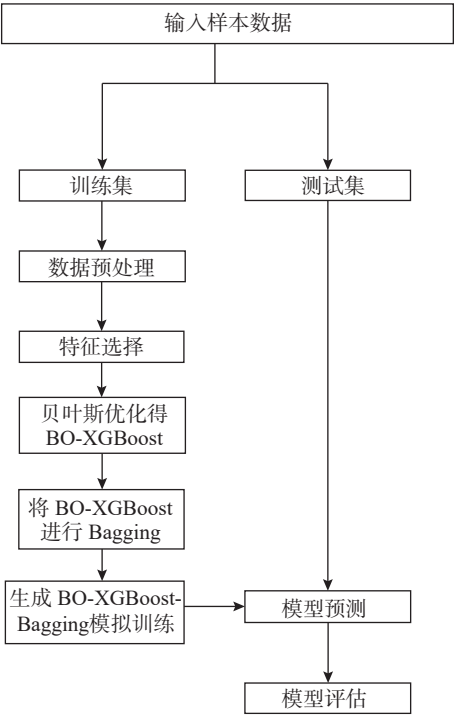


图 1 模型流程图
Fig.1 Model flow chart

2.1 信用评价指标体系的构建

2.1.1 信用评价指标的选取

供应链金融的诞生为中小企业和银行发展带来新机遇的同时，也带来了新挑战。在多元化和复杂化的背景下，对供应链金融进行风险管控也显得尤其重要。而信用风险评价指标不仅是信用评价的依据，还是衡量信用评价结果是否公平公正的重要标准。何平均等^[25]引入客户集中度指标，完善指标评价体系。徐秋慧等^[26]结合企业财务指标和所在供应链的发展情况构建信用风险指标体系，对供应链金融农业企业进行信用评价。

Yang 等^[27] 通过研究, 选取了财务指标和交易数据对中小企业进行风险评估。Xuan 等^[28] 通过选取财务数据、交易数据和物流数据构建供应链金融信用评价指标体系, 为供应链金融信用风险控制提出了新方案。

本文借鉴以上研究, 选取传统信用评价方法中的 5C 指标作为指标选取的依据。其中: 品质(character)指的是企业的名誉; 能力(capacity)主要是指企业偿还债务的能力; 资本(capital)代表股东的投资情况; 担保(collateral)主要考虑企业无法偿还债务时能被抵押的资产; 环境条件(condition)主要分析企业所处的环境情况。参考以往研究, 在常规财务指标体系中加入了供应链集中度指标。中小企业供应链集中度越高, 与供应链关系越密切, 则其所在供应链完整性和稳定性越高, 表明中小企业供应链金融能力越高。本文根据 5C 指标最终选取 9 个一级指标、25 个二级指标作为最初的指标体系, 具体内容如表 1 所示。

2.1.2 特征选择

特征选择是从收集到的高维数据特征中删除冗余或者无关的特征, 选择最优的特征集合, 从而降低模型复杂度, 提高模型准确率。周传华等^[29] 提出根据是否依赖机器学习, 特征选择方法主要分为两种: 过滤式(filter)和封装式(wrapper)。过滤式具有高效率, 但其最终选择结果不理想, 比较常用的方法主要有 Fisher 得分^[30]、Laplacian 得分^[31] 和 Pearson 相关系数^[32] 等。封装式是在特征空间内进行搜索寻优, 虽然效率低, 但选择的最优特征子集效果更好。封装式主要通过决策树^[33]、支持向量机^[34]、随机森林^[35] 等机器学习算法计算特征重要度并进行排序, 从而选择最优子集。本文使用封装式特征选择方法, 通过 XGBoost 的特征重要度进一步选择最优特征, 从而作为本文最终的信用评价指标体系。

2.2 贝叶斯优化

贝叶斯优化(Bayesian optimization)是一种自动化机器学习方法, 可以自动搜索出模型的超参数, 适合于参数较多的模型进行超参数优化。相关研究表明, 贝叶斯优化优于其他优化算法^[36], 同网格搜索、随机搜索相比, 贝叶斯优化更加高效准确, 迭代次数更小。贝叶斯优化主要是在目标函数方程未知的情况下根据已有采样点预估函数的最值, 其中先验函数和采集函数是最重要的

表 1 指标体系
Tab.1 Index system

衡量内容	一级指标	二级指标
品质(character)	企业自身素质	上市公司信息披露考评结果
能力(capacity)	偿债能力指标	长期负债权益比率
		长期借款与总资产比
		资产负债率
		流动比率
		长期债务与营运资金比率
		息税折旧摊销前利润 / 负债合计
		权益乘数
	营运能力指标	流动资产周转率
		总资产周转率
	盈利能力指标	净资产收益率
投资收益率		
流动资产净利润率		
研发费用率		
非经常性损益		
现金与利润总额比		
现金流量指标	每股现金净流量	
发展能力指标	销售费用增长率	
	总资产增长率	
	营业收入增长率	
	资本积累率	
资本(capital)	股东获利能力	每股收益
		每股净资产
担保(collateral)	长期资产	企业固定资产净额
环境条件(condition)	供应链状况	供应链集中度

两部分。高斯过程指的是服从联合正态分布的随机变量的集合, 先验函数则利用高斯过程回归。采集函数主要有探索(explore)和利用(exploit)两个方向: 探索是指针对未知区域, 选取未评估的参数组合, 防止陷入局部最优; 利用是指挖掘已知最优点周围的点, 迅速找到最优解。本文所用采集函数表示为

$$P(x) = \Phi\left(\frac{\mu(x) - f(x^+) - \varepsilon}{\sigma(x)}\right)$$

(9)

式中: $\Phi(\cdot)$ 为正态累计分布函数; $\mu(x)$, $\sigma(x)$ 分别为期望和方差; ε 用来调节搜索方向; $f(x^+)$ 表示目前的最优值。

贝叶斯优化的具体步骤如下:

- a. 生成随机的初始化点, 根据输入样本使用高斯过程更新目标函数的后验概率分布;
- b. 根据后验概率分布找到使采集函数最大化的参数集合;
- c. 计算当前参数集合的目标函数值, 若目标函数值达到要求则停止迭代输出, 否则迭代至最大次数。

2.3 BO-XGBoost-Bagging 模型

XGBoost 实质上是一种 Boosting 算法, 主要通过树模型不断串行来拟合残差。在拟合残差的迭代中, 根据上一次的迭代结果进行调整, 从而逐步降低偏差。由于串行方式不能显著降低模型的方差, 而 Bagging 是并行集成方式, 能从原数据集中进行重抽样, 并对各基分类器分别训练并采取投票机制, 从而减小方差。为了权衡偏差与方差, 进一步提高精度, 本文提出在 Bagging 框架下结合贝叶斯优化和 XGBoost 的集成学习模型 BO-XGBoost-Bagging, 具体过程如下:

- a. 基于 XGBoost 的特征重要度, 进行特征筛选, 得到模型的指标体系;
- b. 通过 BO 进行优化, 得到 XGBoost 的最优超参数集, 得到 BO-XGBoost 模型;
- c. 将 BO-XGBoost 进行 Bagging, 得到 BO-XGBoost-Bagging 模型;
- d. 对于每个 BO-XGBoost 基分类器的预测结果, 采取投票机制将预测结果进行组合, 得到最终的结果。

3 算例分析

本文从国泰安数据库中选取 253 家中小企业 2020 年的数据进行实证研究, 经过数据预处理剔除了严重缺失信息的企业, 将剩余企业的缺失信息用均值代替。由于破产数据难以获得, 参考李成刚等^[37]、李哲等^[38]的研究, 采用惯常做法, 将 st 标记的企业作为风险企业, 最终选择了 210 家中小企业。其中, 正常企业有 137 家, 风险企业有 73 家。为了全面评价所建立模型的分类预测性能, 首先选取准确率 (accuracy) 为模型评价指标。准确率用来评估模型对整体样本的分类准确度, 但对预测性能的评价不够全面。通常情况下, 风险企业的预测更加重要, 对风险企业进行准确识别最为关键。因此, 选取 AUC 值和 F1 值对模型

整体性能进行进一步评估。AUC 值是衡量模型优劣的重要指标, F1 值同时考虑召回率和精确率之间的关系, 主要反映风险企业的预测效果。精确率表示在预测为风险企业的样本中实际风险企业的占比, 召回率表示在实际为风险企业的样本中预测正确的风险企业占比。基于上述 3 个评价指标, 对所有建立的模型进行预测评价, 并比较 3 个指标的结果。

3.1 特征选择

3.1.1 基于 XGBoost 的特征选择

本文总共选取了 25 个候选指标, 通过 XGBoost 度量各个指标的重要性程度, 对其进行特征选择, 按照表 1 二级指标顺序排列, 结果如图 2 所示。通过特征选择最终保留特征重要性前 10 的特征, 其中包括上市公司信息披露考评结果、流动比率、总资产周转率、净资产收益率、流动资产净利润率、每股现金净流量、总资产增长率、资本积累率、每股净资产、供应链集中度。评价指标的具体明细如表 2 所示。

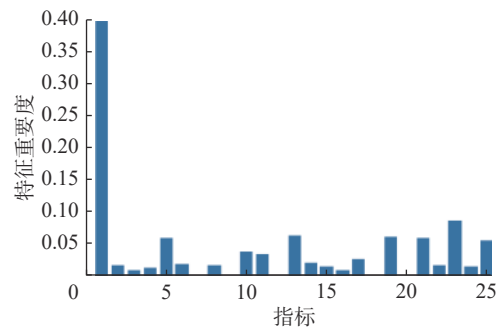


图 2 特征重要度
Fig.2 Feature importance

3.1.2 指标相关性分析

根据筛选出的 10 个指标, 进一步分析特征之间的相关性, 分别计算 10 个特征之间的相关系数, 并用相关系数热力图直观表现不同特征之间的关系, 如图 3 所示, 热力图中颜色极深或极浅表示变量之间相关性很高。由热力图可以得出所有变量的相关系数都小于 0.5, 且大部分相关系数小于 0.3。由于变量之间的相关系数都不大, 可以不考虑它们之间的共线性问题, 并将这 10 个评价指标确定为最终的指标体系。

3.2 超参数优化

本文以 AUC 值为优化目标, 使用贝叶斯优化分别对逻辑回归 (logistic regression, LR)、随机森林 (random forest, RF)、支持向量机 (support

表 2 指标名称及描述
Tab.2 Index name and description

一级指标	二级指标	英文缩写	指标说明
企业自身素质	上市公司信息披露考评结果	IDER	衡量上市公司信息披露的质量
偿债能力指标	流动比率	CR	流动资产/流动负债
营运能力指标	总资产周转率	TROTA	营业收入/资产总额期末余额
盈利能力指标	净资产收益率	ROE	净利润/股东权益余额
	流动资产净利润率	NPMOCA	净利润/流动资产余额
现金流量指标	每股现金净流量	NCFPS	现金及现金等价物净增加额/总股数
发展能力指标	总资产增长率	GROTA	(期末总资产-期初总资产)/期初总资产
	资本积累率	CAR	(期末股东权益-期初股东权益)/期初股东权益
股东获利能力指标	每股净资产	NAPS	股东权益总额/普通股股数
供应链状况	供应链集中度	SCC	衡量前5大供应商采购比例和客户销售比例之间的关系

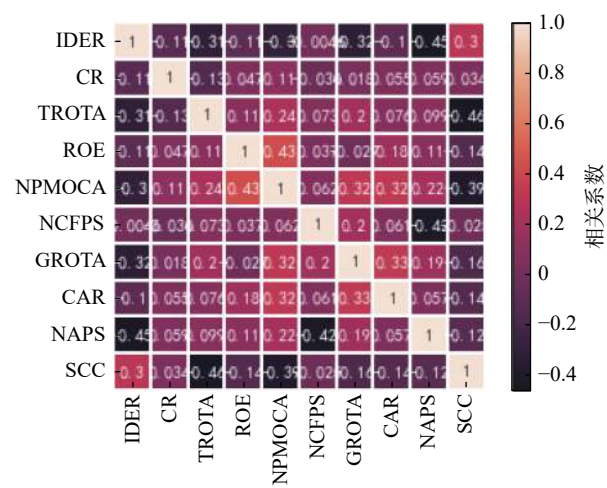


图 3 相关系数热力图

Fig. 3 Thermodynamic diagram of correlation coefficients

vector machine, SVM)、 K 近邻(K -nearestneighbor, KNN)和 XGBoost 进行调参。由于各模型的特点与参数均不相同,本文以表现最佳的 XGBoost 模型为例,超参数设置如表 3 所示。

3.3 结果比较

首先选取了 4 个在信用评价领域中较为常用的机器学习算法与 XGBoost 算法进行对比,分别为逻辑回归、随机森林、支持向量机和 KNN。这些对比算法和 XGBoost 算法使用相同的特征子集,同时采用贝叶斯优化进行调参,将原始数据集中随机选取的 70%作为训练集对模型进行训练,原始数据集另外的 30%作为测试集对模型性能进行测试,并与其他模型在提出的 3 个评价指标上进行对比,实验得到的测试集预测效果如表 4

所示。

从表 4 可以得出,BO-XGBoost 的准确率最高,而其他模型准确率均低于 0.9。在整体预测效果 AUC 值方面,BO-XGBoost 仅略低于 BO-RF 模型 0.0033,但相比其他模型仍有着较小的优势。在风险企业预测效果 $F1$ 值方面,BO-XGBoost 仍然有着显著优势,明显高于其他对比模型。总体来说 BO-XGBoost 相比其他主流分类器,具有更好的分类性能,说明了 BO-XGBoost 模型能够较为准确、快速地对中小企业进行信用评价。

BO-XGBoost 相比其他模型具有更好的分类效果,将 BO-XGBoost 进行集成,分别进行 Bagging, Boosting 和 Stacking。第一个模型训练 m 个差异化的 BO-XGBoost 模型进行 Bagging,通过 bootstrap 采样,生成 BO-XGBoost-Bagging(BXB)模型。第

表 3 XGBoost 参数说明及调参结果
Tab.3 XGBoost parameter description and parameter adjustment results

参数名称	参数含义	参数结果
max_depth	指数的最大深度,用来控制树的复杂度	4
n_estimators	指模型中树的个数	131
gamma	用于调节惩罚项	0.49
lambda	调节L2正则化项	0.24
subsample	用来控制每棵树中随机采样的比例	0.86
colsample_bytree	对每颗树使用的特征进行随机采样	0.88
min_child_weight	用来限制叶子节点继续分割,以防止过拟合	6
eta	更新过程中的收缩步长	0.01
alpha	调节L1正则化项	0.59

表 4 BO-XGBoost 与其他算法的结果对比

Tab.4 Comparison between BO-XGBoost and other algorithms

模型	准确率	AUC值	F1值
BO-LR	0.873 0	0.954 5	0.800 0
BO-RF	0.888 9	0.964 5	0.844 4
BO-SVM	0.857 1	0.937 9	0.780 5
BO-KNN	0.841 3	0.924 6	0.750 0
BO-XGBoost	0.904 8	0.961 2	0.863 6

二个模型用 Boosting 的经典算法 AdaBoost 实现，训练 m 个 XGBoost 作为基分类器，生成 BO-XGBoost-AdaBoost (BXA) 模型。第三个模型将 LightGBM、随机森林、GBDT、XGBoost 作为 Stacking 模型的第一层，逻辑回归为第二层，训练生成 BO-XGBoost-Stacking (BXS) 模型。用新建立的 3 个模型在随机选取的 70% 原始数据集上进行训练，对另外 30% 的数据集进行预测，得到如表 5 所示的结果。

表 5 不同集成模型的结果对比

Tab.5 Comparison of results of different integration models

模型	准确率	AUC值	F1值
BXB	0.920 6	0.962 3	0.888 9
BXA	0.650 8	0.500 0	0
BXS	0.904 8	0.925 3	0.875 0

由表 5 可知，将 BO-XGBoost 进行 Bagging 得到的 BXB 模型在各个指标上达到最优。该模型各项指标相比 BO-XGBoost 模型分别提升了 1.75%，0.11% 和 2.93%。BXB 模型相比 BO-XGBoost 预测性能有显著提升，分类效果也优于其他集成模型。BXA 模型在预测时，所有预测概率都为 0.5，以至于所有样本都分为一个种类。BXS 模型相比 XGBoost 原模型没有太大的提升，但相比其他基础模型在准确率和 $F1$ 值上略有提升。因此，相比其他集成学习方式，强分类器更加适合进行 Bagging。单一的 XGBoost 模型容易受到噪声干扰，所以 Bagging 集成方法从一定程度上降低了泛化误差，并通过投票机制来避免极端情况的出现，从而进一步提高了模型的分类精度。本文提出的 BXB 模型在对中小企业信用风险识别预测方面具有较好的效果，显著优于其他分类模型，能够更加准确、全面地对中小企业的信用风险进行评估，在供应链金融的风险控制中具有实际意义。

4 结 论

针对供应链金融领域中小企业融资的信用风险控制问题，提出了一种在 Bagging 框架下结合贝叶斯优化和 XGBoost 的集成学习模型 BO-XGBoost-Bagging。该模型结合了 XGBoost 出色的预测能力、贝叶斯优化高效的超参数优化能力和 Bagging 集成方法的强泛化能力。与其他模型相比，BXB 在模型准确度、泛化能力等指标上都有着更好的表现，可以有效区分风险企业与正常企业，并准确、高效地预测中小企业的信用风险。本文提出的基于 BXB 的信用评价方法有助于银行等金融机构在供应链金融模式下对中小企业的融资问题进行信用风险控制，一定程度上减少了违约损失，促进了供应链金融的健康发展。

参考文献：

[1] 宋华. 困境与突破: 供应链金融发展中的挑战和趋势 [J]. 中国流通经济, 2021, 35(5): 3–9.

[2] 刘露, 李勇建, 薛克雷. 融资选择冲突与可变参数风险分担机制 [J]. 系统工程理论与实践, 2021, 41(3): 649–666.

[3] ZHU Y, ZHOU L, XIE C, et al. Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach[J]. International Journal of Production Economics, 2019, 211: 22–33.

[4] E SILVA E C, LOPES I C, CORREIA A, et al. A logistic regression model for consumer default risk[J]. Journal of Applied Statistics, 2020, 47(13/15): 2879–2894.

[5] 郭林, 李战江, 孔文婷. 基于二阶段 Logistic 回归的小企业信用指标判别模型 [J]. 数学的实践与认识, 2020, 50(20): 35–45.

[6] 王小燕, 张中艳, 马双鸽. 基于文本先验信息的贷款信用风险评估模型 [J]. 中国管理科学, 2021, 29(5): 34–44.

[7] 许文彬, 张佳韬. 我国城投债信用风险测算——基于 Logistic 回归的评分卡模型 [J]. 数理统计与管理, 2020, 39(4): 705–720.

[8] HAN L, SU Z, LIN J. A hybrid KNN algorithm with Sugeno measure for the personal credit reference system in China[J]. Journal of Intelligent & Fuzzy Systems, 2020, 39(5): 6993–7004.

[9] ATA O, HAZIM L. Comparative analysis of different distributions dataset by using data mining techniques on credit card fraud detection[J]. Tehnički Vjesnik, 2020, 27(2): 618–626.

[10] TANG L X, CAI F, OUYANG Y. Applying a nonparametric random forest algorithm to assess the credit

- risk of the energy industry in China[J]. *Technological Forecasting and Social Change*, 2019, 144: 563–572.
- [11] 王润华. 基于改进支持向量机的消费信贷中个人信用评估模型 [J]. 统计与决策, 2010(11): 54–56.
- [12] YONTAR M, NAMLI Ö H, YANIK S. Using machine learning techniques to develop prediction models for detecting unpaid credit card customers[J]. *Journal of Intelligent & Fuzzy Systems*, 2020, 39(5): 6073–6087.
- [13] SHI J, ZHANG S Y, QIU L M. Credit scoring by feature-weighted support vector machines[J]. *Journal of Zhejiang University SCIENCE C*, 2013, 14(3): 197–204.
- [14] 莫赞, 张灿凤, 魏伟, 等. 基于 Bagging 集成的个人信用风险评估方法研究 [J]. 系统工程, 2019, 37(1): 143–151.
- [15] YANG J H. Research on the forecasting of enterprise credit scoring based on SVR model[J]. *Academic Journal of Engineering and Technology Science*, 2020, 3(1): 55–64.
- [16] YE K T. Research on personal credit risk assessment based on machine learning algorithm[J]. *World Scientific Research Journal*, 2020, 6(11): 146–156.
- [17] 李欣, 俞卫琴. 基于改进 GS-XGBoost 的个人信用评估 [J]. 计算机系统应用, 2020, 29(11): 145–150.
- [18] 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.
- [19] LIU Y, HUANG L H. Supply chain finance credit risk assessment using support vector machine-based ensemble improved with noise elimination[J]. *International Journal of Distributed Sensor Networks*, 2020, 16(1): 1–10.
- [20] 刘平山, 曾梓铭. 基于 GBDT 的医药供应链金融信用风险评价 [J]. 会计之友, 2021(16): 24–31.
- [21] ZHANG H, SHI Y X, TONG J Y. Online supply chain financial risk assessment based on improved random forest[J]. *Journal of Data, Information and Management*, 2021, 3(1): 41–48.
- [22] ZHANG H, SHI Y X, YANG X R, et al. A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance[J]. *Research in International Business and Finance*, 2021, 58: 101482.
- [23] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. California, San Francisco: ACM, 2016: 785–794.
- [24] RUSDAH D A, MURFI H. XGBoost in handling missing values for life insurance risk prediction[J]. *SN Applied Sciences*, 2020, 2(8): 1336.
- [25] 何平均, 李菁菁. 客户集中度与供应链金融信用风险——基于中小企业板制造业上市公司的实证研究 [J]. 征信, 2018, 36(7): 21–26.
- [26] 徐秋慧, 王迪, 邓之音. 供应链金融模式下中小企业融资问题分析——以 44 家农业中小企业为例 [J]. 农村金融研究, 2020(9): 63–71.
- [27] YANG Y B, CHU X J, PANG R Q, et al. Identifying and predicting the credit risk of small and medium-sized enterprises in sustainable supply chain finance: evidence from China[J]. *Sustainability*, 2021, 13(10): 5714.
- [28] XUAN F. Regression analysis of supply chain financial risk based on machine learning and fuzzy decision model[J]. *Journal of Intelligent & Fuzzy Systems*, 2021, 40(4): 6925–6935.
- [29] 周传华, 柳智才, 丁敬安, 等. 基于 filter+wrapper 模式的特征选择算法 [J]. 计算机应用研究, 2019, 36(7): 1975–1979, 2010.
- [30] 安起光, 于晓静. 基于 Fisher 判别方法的信用风险评估实证研究 [J]. 现代管理科学, 2012(4): 67–69.
- [31] 钱晓亮, 左开中, 接标. 新的基于 Laplacian 的特征选择方法 [J]. 计算机工程与应用, 2016, 52(15): 79–82.
- [32] ZHANG Y. Assessing fair lending risks using race/ethnicity proxies[J]. *Management Science*, 2018, 64(1): 178–197.
- [33] 董纪阳. 基于决策树自动化特征选择的基金客户流失预测研究——后疫情时代下的思考 [J]. 山东社会科学, 2020(9): 74–80.
- [34] ZHOU T, LU H L, WANG W W, et al. GA-SVM based feature selection and parameter optimization in hospitalization expense modeling[J]. *Applied Soft Computing*, 2019, 75: 323–332.
- [35] BOU-HAMAD I. Bayesian credit ratings: a random forest alternative approach[J]. *Communications in Statistics-Theory and Methods*, 2017, 46(15): 7289–7300.
- [36] SHAHRIARI B, SWERSKY K, WANG Z Y, et al. Taking the human out of the loop: a review of Bayesian optimization[J]. *Proceedings of the IEEE*, 2016, 104(1): 148–175.
- [37] 李成刚, 贾鸿业, 赵光辉, 等. 基于信息披露文本的上市公司信用风险预警——来自中文年报管理层讨论与分析的经验证据 [J/OL]. 中国管理科学 [2021-12-21]. <https://doi.org/10.16381/j.cnki.issn1003-207x.2020.2263>.
- [38] 李哲, 迟国泰. 基于最大指标区分度与最优相对隶属度的上市公司信用风险研究 [J]. 中国管理科学, 2021, 29(4): 1–15.

(编辑: 丁红艺)