基于集成学习算法的金浸出率智能预测模型设计研究

周一帆[1,2]

(1. 驻马店职业技术学院 信息工程学院,河南 驻马店　463000;
2. 鸿蒙生态开发重点实验室,河南 驻马店　463000)

摘要:针对金浸出率预测准确率不高及噪声数据较大的问题,提出了一种高效的智能预测方法。将改进的XGBoost和 LightGBM 模型相结合,并采用动态学习率、正则化优化和贝叶斯优化算法,设计了一个集成学习模型对金浸出率进行预测。结果表明:相较于传统单一模型,集成学习模型预测精度明显提升,其均方误差较 XGBoost 和 LightGBM 分别降低 28.8%和 22.9%左右,在真实生产环境中稳定性较高。该集成学习模型能够有效提高金浸出率的预测精度,具有较强的实际应用价值。

关键词:金浸出率;预测;集成学习;XGBoost 模型;LightGBM 模型;数值仿真

中图分类号:TTF803.21　　**文献标识码**:A　　**文章编号**:1009-2617(2025)05-0698-08

DOI:10.13355/j.cnki.sfyj.2025.05.016

浸出率是湿法冶金工艺中评估金属回收效果和工艺优化的关键指标,其准确预测对于提高资源利用效率、降低生产成本和减少环境污染具有重要意义[1]。传统方法难以处理复杂多变的工艺参数和非线性动态特性[2],而基于智能预测的研究可以利用实时数据和机器学习算法,实现对浸出过程的精准建模与优化控制,从而为湿法冶金工业提供科学决策支持,提升工艺的智能化水平和经济效益。目前虽已有一些关于智能预测在湿法冶金浸出过程中的应用研究[3-5],但浸出过程受多因素影响,使得数据采集与处理工作仍面临诸多挑战。因此,如何有效地利用数据进行精准预测,成为了研究的关键。

为提高浸出率预测精度,以金浸出率预测为例,提出了一种基于集成学习的湿法冶金智能预测模型。首先,针对金浸出率数据的复杂非线性关系和噪声特性,改进 XGBoost 模型[6-7],引入动态学习率和正则化优化,并结合贝叶斯优化调整超参数,提升模型的鲁棒性与泛化能力;其次,优化 LightGBM 基学习器[8-9],采用多尺度特征重要性评估策略和 SMOTE(Synthetic Minority Over-sampling Technique)算法[10]处理特征不均衡与类别不平衡问题,同时改进预排序算法以增强实时性和计算效率;最后,设计基于 Boosting 的集成方法[11],将改进的 XGBoost 和 LightGBM 模型相结合,充分发挥其互补优势,构建高效的集成学习模型。

1　基于改进 XGBoost 的浸出率预测基学习器

XGBoost(Extreme Gradient Boosting)是一种高效的梯度提升树(GBDT)算法[12],广泛应用于回归和分类任务中。为提升其在金浸出率预测任务中的准确性和鲁棒性,提出了一种基于改进 XGBoost 的浸出率预测基学习器,主要通过引入动态学习率、正则化优化及贝叶斯优化超参数调整等方法进行 XGBoost 基学习器的改进。

1)模型框架与损失函数改进。XGBoost 的基本框架通过逐步建立决策树的方式来进行集成学习,其损失函数可表示为:

$$L(\theta)=\sum_{i=1}^{n} l(y_i,\hat{y}_i)+\sum_{k=1}^{K}\Omega(f_k)。$$

式中:$l(y_i,\hat{y}_i)$—第 i 个样本的损失函数;$\hat{y}_i$—预

收稿日期:2024-11-27

作者简介:周一帆(1989—),男,硕士,讲师,主要研究方向为大数据、人工智能及计算机技术。

测值；y_i—实际值；$\Omega(f_k)$—对模型复杂度的正则化项。正则化项用来控制模型的复杂度，防止过拟合：

$$\Omega(f_k)=\gamma T_k+\frac{1}{2}\lambda\sum_{j=1}^{T_k}w_{kj}^2。$$

式中：γ、λ—正则化参数；T_k—第 k 棵树的叶节点数；w_{kj}—第 k 棵树的叶节点的权重。

2）动态学习率调整。在传统的 XGBoost 模型中，学习率 η 是一个固定值。为了增强模型的适应性，提出采用动态学习率调整策略。具体地，学习率根据模型在训练过程中的表现动态调整，采用的策略如下：

$$\eta_t=\frac{\eta_0}{1+\alpha t}。$$

式中：η_0—初始学习率；α—学习率衰减因子；t—训练迭代的次数。通过这种动态调整，模型能够在训练初期快速收敛，在后期逐步减小步长，以避免过拟合。

3）贝叶斯优化调整超参数。为进一步提升模型的预测精度和泛化能力，通过贝叶斯优化方法调整超参数。贝叶斯优化是通过建立代理模型来搜索最优超参数组合。通过贝叶斯优化，超参数（如树的最大深度、学习率、正则化参数 λ 和 γ）会在训练过程中自动调整，以最小化损失函数。这种方法能有效减少人工调参的工作量，并确保模型在复杂数据集上的优异性能。

4）模型训练与迭代优化。模型训练采用梯度下降法来优化目标函数。每次迭代时，XGBoost 通过最小化损失函数的泰勒展开来更新决策树的结构：

$$\mathcal{L}(f)=\sum_{i=1}^{n}\left[l(y_i,\hat{y_i})+\frac{1}{2}w_i^2\right]。$$

对于每一棵树，在每个节点分裂时，XGBoost 通过计算分裂增益来决定最优分裂点，增益计算公式为：

$$\text{Gain}(t)=\frac{1}{2}\left[\frac{(\sum_{i\in L}g_i)^2}{\sum_{i\in L}h_i+\lambda}+\frac{(\sum_{i\in R}g_i)^2}{\sum_{i\in R}h_i+\lambda}-\frac{(\sum_{i\in S}g_i)^2}{\sum_{i\in S}h_i+\lambda}\right]。$$

式中：g_i、h_i—样本 i 的一阶、二阶梯度信息；L、R—左、右子树；S—当前节点的样本集合。

基于改进 XGBoost 的金浸出率预测模型，通过动态调整学习率、优化正则化项及结合贝叶斯优化调整超参数[13]，有效提高模型的鲁棒性和泛化能力。通过上述方法，模型能够在处理复杂数据集时更好地拟合目标变量，从而实现对金浸出率更为准确的预测。

改进的 XGBoost 基学习器的结构如图 1 所示。

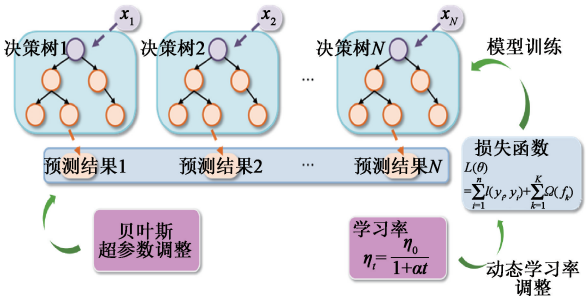


图 1　改进的 XGBoost 基学习器的结构

Fig. 1　Structure of improved XGBoost-based learner

2　基于改进 LightGBM 的浸出率预测基学习器

LightGBM（Light Gradient Boosting Machine）是一种高效的梯度提升决策树（GBDT）算法，广泛应用于大规模数据的回归与分类任务。为在金浸出率预测中获得更高的精度与效率，对 LightGBM 模型进行了显著改进，具体体现在多尺度特征重要性评估策略、SMOTE 算法的引入来处理特征不均衡与类别不平衡问题，以及改进预排序算法以提高计算效率和实时性。

1）LightGBM 模型基本框架。LightGBM 的基本框架与 XGBoost 相似，通过树的集合（Boosting）来进行学习。其目标函数可表示为：

$$L(\theta)=\sum_{i=1}^{n}l(y_i,\hat{y_i})+\sum_{k=1}^{K}\Omega(f_k)。$$

式中：$l(y_i,\hat{y_i})$—第 i 个样本的损失函数；$\hat{y_i}$—预测值；y_i—实际值；$\Omega(f_k)$—对模型复杂度的正则化项，常采用 L2 正则化来控制模型的复杂度。

LightGBM 使用直方图算法对数据进行预排序，以提高计算效率。

$$\Omega(f_k)=\gamma T_k+\frac{1}{2}\lambda\sum_{j=1}^{T_k}w_{kj}^2。$$

式中：γ—叶节点数的惩罚参数；λ—正则化参数；

T_k—树 k 的叶节点数；w_{kj}—树 k 的叶节点的权重。

2)多尺度特征重要性评估策略。为更好地处理多尺度特征数据，提出了一种基于分层信息的特征重要性评估策略。传统的 LightGBM 使用单一尺度进行特征划分，而在改进模型中，采用多尺度特征评估，通过在不同尺度上评估特征的重要性，以增强模型对多尺度特征的适应能力。具体来说，特征的重要性计算由以下两部分组成：

$$\text{Importance}_{\text{total}}(f)=\sum_{t=1}^{T}\text{Importance}_t(f)。$$

式中：$\text{Importance}_t(f)$—在第 t 层次（或子空间）上评估的特征 f 的重要性；T—尺度层次的数量。每个层次上，特征重要性计算公式为：

$$\text{Importance}_t(f)=\frac{1}{N_t}\sum_{i=1}^{n_t}\frac{g_i^2}{h_i+\lambda}。$$

式中：g_i、h_i—样本 i 的一阶、二阶梯度信息；n_t—第 t 层次上的样本数量；N_t—该层次的总样本数；h_i—该样本的二阶梯度。

通过这种多尺度的特征评估方法，模型能够更全面地捕捉到数据中不同特征层次上的信息，从而增强模型对复杂数据关系的学习能力。

3)SMOTE 算法处理特征不均衡与类别不平衡。由于金浸出率预测任务中可能存在特征不均衡和类别不平衡问题，传统的 LightGBM 模型在训练时容易受到影响，导致模型精度降低。为解决这一问题，本研究引入了 SMOTE 算法，该算法通过生成合成样本来平衡训练集中的样本类别。SMOTE 算法通过在特征空间内生成新的合成样本来平衡数据集。具体地，给定一个样本点 x_i 和其 k 个近邻，SMOTE 会通过以下方式生成合成样本：

$$x_{\text{new}}=x_i+\lambda\cdot(x_j-x_i)。$$

式中：x_j—样本 x_i 的一个随机近邻；λ——个随机选择的系数，$\lambda\in[0,1]$；x_{new}—生成的合成样本。

通过 SMOTE 算法，数据集的类别分布得到有效平衡，从而提高模型的稳定性和预测精度。

4)改进预排序算法。传统的 LightGBM 使用直方图算法来对数据进行预排序，但在大规模数据处理时，仍然会存在较高的计算成本。为提高计算效率和实时性，对预排序算法进行了改进，采用基于分层抽样和即时学习[14]的预排序方法。

在每次训练迭代中，首先对训练数据进行分层抽样，确保每一层数据分布均衡，接着利用即时学习算法对分层数据进行预排序。具体步骤如下：

1)对训练数据进行分层抽样，确保每一层中的样本数量与目标变量分布一致。

2)使用即时学习算法对分层数据进行排序和分裂评估，每次只对当前数据子集进行排序。

3)使用动态更新机制对排序结果进行优化，避免全局排序的计算瓶颈。

这种改进的预排序算法能够在保证模型准确性的同时，提高训练效率，适应实时预测任务的需求。

5)模型训练与迭代优化。LightGBM 使用增量学习的方式通过梯度提升（Gradient Boosting）[15]来训练模型。在每次迭代中，利用梯度信息和二阶梯度信息来构建新的决策树，通过最小化目标函数来更新模型：

$$\mathcal{L}(f)=\sum_{i=1}^{n}\left[l(y_i,\hat{y_i})+\frac{1}{2}w_i^2\right]。$$

树的分裂增益计算公式为：

$$\text{Gain}(t)=\frac{1}{2}\left[\frac{\left(\sum_{i\in L}g_i\right)^2}{\sum_{i\in L}h_i+\lambda}+\frac{\left(\sum_{i\in R}g_i\right)^2}{\sum_{i\in R}h_i+\lambda}-\frac{\left(\sum_{i\in S}g_i\right)^2}{\sum_{i\in S}h_i+\lambda}\right]。$$

式中：g_i和 h_i—样本 i 的一阶和二阶梯度信息；L 和 R —左右子树；S —当前节点的样本集合。

基于改进的 LightGBM 模型，通过引入多尺度特征重要性评估、SMOTE 算法处理类别不平衡问题，以及改进预排序算法，显著提升了模型的精度和计算效率。这些改进使得 LightGBM 能够更好地适应金浸出率预测任务中的复杂数据结构，为工业过程中的实时预测和优化提供了有效支持。改进的 LightGBM 基学习器的结构如图 2 所示。

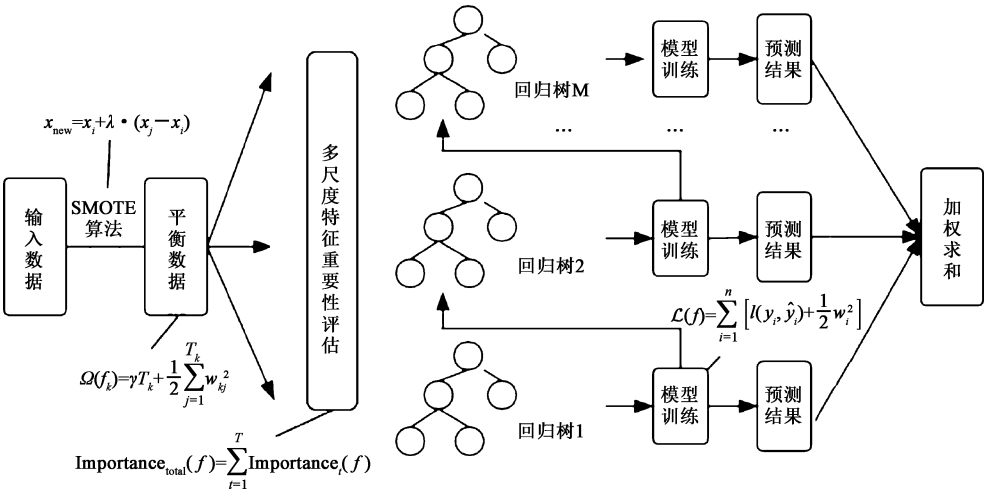


图 2　改进的 LightGBM 基学习器的结构

Fig. 2　Structure of improved LightGBM-based learner

3　基于 Boosting 集成学习的金浸出率智能预测

在金浸出率预测任务中，模型面临数据的非线性、噪声及高维性等挑战。因此，研究设计一种基于 Boosting 的集成学习方法，将改进的 XGBoost 和 LightGBM 模型相结合，充分发挥其互补优势，构建一个高效的集成学习模型，以提升金浸出率预测的精度和鲁棒性。具体来说，通过集成多个基学习器（改进的 XGBoost 和 LightGBM），并通过 Boosting 算法优化模型的性能。

1）Boosting 集成学习方法基本框架。Boosting 是一种集成学习方法，它通过加权组合多个弱分类器（或回归器）来形成一个强学习器。Boosting 模型的核心思想是在每一轮迭代中训练一个新的弱学习器，该学习器关注的是上一轮学习器未能正确预测的样本。最终，所有弱学习器的预测结果会加权汇总得到最终的预测结果。Boosting 模型的目标函数可以表示为：

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y_i}) + \sum_{k=1}^{K} \Omega(f_k)。$$

式中：$l(y_i, \hat{y_i})$ —第 i 个样本的损失函数；$\hat{y_i}$ —模型的预测值；y_i—真实值，$\Omega(f_k)$—对模型复杂度的正则化项；f_k—第 k 个弱学习器。

2）改进的 XGBoost 和 LightGBM 模型集成。为提高金浸出率的预测精度，研究结合了改进的 XGBoost 和 LightGBM 模型，利用 Boosting 方法进行集成。XGBoost 和 LightGBM 分别是基于梯度提升（GBDT）的模型，它们通过多次迭代训练树模型并结合各树的结果来提高预测精度。其基本结构可通过加权组合表示为：

$$\hat{y_i} = \sum_{k=1}^{K} \alpha_k \hat{y}_{ik}。$$

式中：$\hat{y}_{ik}$ —第 k 个基学习器对第 i 个样本的预测值；α_k—基学习器的权重系数；K—基学习器的总数。

3）集成学习中的加权组合与优化。Boosting 方法的关键是如何选择合适的加权系数 α_k，以使得每个基学习器的贡献最大化。传统的 Boosting 方法通过使用梯度信息来调整加权系数。对于每一轮的迭代，基学习器的权重可按下式更新：

$$\alpha_k = \frac{1}{2}\ln\left(\frac{1-\varepsilon_k}{\varepsilon_k}\right)。$$

式中，ε_k—第 k 个基学习器的错误率，表示为：

$$\varepsilon_k = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(y_i \neq \hat{y}_{ik}),$$

式中，$\mathbb{I}$—指示函数，当预测值与真实值不相等时，为 1，否则为 0。

为进一步提高预测效果，本研究对基学习器进行了改进。对于 XGBoost，改进了其正则化项，使得模型更加适应复杂数据的训练，具体如下：

$$\Omega(f_k) = \gamma T_k + \frac{1}{2}\lambda \sum_{j=1}^{T_k} w_{kj}^2。$$

式中：γ—叶节点的惩罚参数；T_k—第 k 个基学习

器的树的叶节点数；w_{kj}—k 个基学习器叶节点的权重。

对于 LightGBM，采用了多尺度特征重要性评估策略，针对每一层的样本，通过计算其梯度信息的平方和来调整权重，从而提升模型对不同特征的敏感度。特征重要性评估公式为：

$$\text{Importance}_t(f)=\frac{1}{N_t}\sum_{i=1}^{n_t}\frac{g_i^2}{h_i+\lambda}。$$

式中：g_i、h_i—样本 i 的一阶、二阶梯度；n_t—第 t 层次上的样本数量；h_i—样本的二阶梯度；λ—正则化参数。

4)集成模型的预测过程与优化。在集成学习过程中，基学习器的预测结果会被加权组合，最终得到金浸出率的预测值。假设第 i 个样本的真实值为 h_i，预测值为 $\hat{y}_i$，则集成学习模型的预测结果为：

$$\hat{y}_i=\sum_{k=1}^{K}\alpha_k\hat{y}_{ik}。$$

此外，为进一步提升集成模型的预测性能，采用贝叶斯优化算法对超参数进行调优。贝叶斯优化通过构建代理模型(如高斯过程)来估计目标函数的分布，并利用这一信息来选择最优超参数组合，从而有效提高模型的预测精度。

5)模型训练与迭代优化。在每轮 Boosting 训练中，基学习器会根据当前的误差对模型进行调整。通过引入正则化项和梯度信息，改进的 XGBoost 和 LightGBM 模型能够更好地拟合金浸出率预测任务中的复杂数据关系。

6)模型评估。模型评估使用均方误差(MSE)作为评价指标：

$$\text{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2。$$

通过与传统的单一模型进行比较，评估集成学习方法在金浸出率预测中的优势。

基于 Boosting 集成学习方法，通过结合改进的 XGBoost 和 LightGBM 模型，充分利用两者的互补优势，在金浸出率预测任务中实现了更高的预测精度和更强的鲁棒性。通过加权组合基学习器的预测结果，并对超参数进行优化，最终构建了一个高效且精准的预测模型。改进的 Boosting 集成学习方法的结构如图 3 所示。

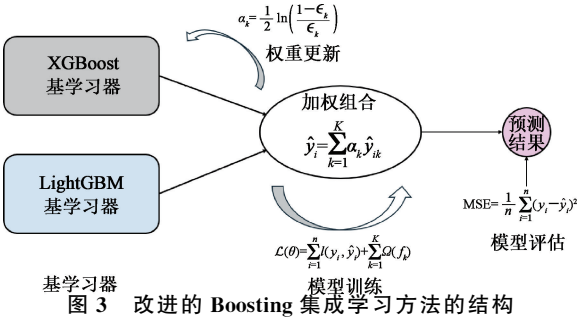


图 3 改进的 Boosting 集成学习方法的结构

Fig. 3 Structure of the improved Boosting integrated learning method

4 数值仿真

4.1 试验环境与数据来源

集成学习模型数值仿真试验在一个高性能计算环境中进行。试验平台配备了 16 核的 AMD Ryzen 9 5900X 处理器，64GB 的 RAM 内存，以及配备 NVIDIA RTX 3090 显卡的计算机。

数据采集自某金属冶炼厂的实际生产数据集，数据集包含了不同金属原料在浸出过程中的多种工艺参数，包括温度、酸度、浸出时间、氧气浓度、矿浆浓度等多维特征。数据集共有约 50 000 条样本，其中每条样本对应一个金浸出率的真实值，作为回归任务的目标变量。此外，数据集还存在特征不均衡问题，部分特征在样本中分布较为集中，因此采用 SMOTE 算法对数据进行过采样处理。

4.2 数值模拟结果及分析

1)集成学习模型与单一模型对比试验。首先，分别训练基于 XGBoost、LightGBM 和集成学习模型的 3 个不同模型；然后，使用相同的训练集和测试集进行验证，通过均方误差(MSE)和决定系数(R^2)作为评价指标，验证集成学习模型在预测精度和泛化能力方面的优势。本研究提出的基于 Boosting 的集成学习模型与单一的 XGBoost 和 LightGBM 模型对金浸出率的预测效果对比结果见表 1。

表 1 集成学习模型与单一模型的预测效果对比

Table 1 Comparison of predictive effectiveness of integrated learning model and single model

模型	MSE	R^2
XGBoost	0.045	0.91
LightGBM	0.042	0.92
本研究模型	0.037	0.94

从表 1 看出：本研究提出的集成学习模型的 MSE 和 R^2 均优于XGBoost和 LightGBM 单一模型；集成学习模型的 MSE 为 0.037，较 XGBoost (0.045) 和 LightGBM (0.042) 分别降低 17.8% 左右。

2)集成学习模型在不同数据平衡策略下的性能对比试验。通过两种数据平衡策略进行比较：一种是原始数据集，另一种是采用 SMOTE 算法对数据进行过采样处理。分别在这两种数据集上训练集成学习模型，并对比其预测精度(MSE 和 R^2)和训练时间，分析数据不均衡对模型性能的影响，结果见表 2。

表 2　集成学习模型在不同数据平衡策略下的性能对比

Table 2　Performance comparison of integrated learning models under different data balancing strategies

数据平衡策略	MSE	R^2	训练时间/s
原始数据集	0.044	0.91	120
SMOTE 过采样数据	0.037	0.94	150

从表 2 看出：SMOTE 过采样后的数据集在集成学习模型上的表现明显优于原始数据集；经过采样后的数据集，集成学习模型的 MSE 为 0.037，比原始数据集的 MSE(0.044)降低17.6%左右。

3)集成学习模型与贝叶斯优化超参数调优效果对比试验。目的是通过贝叶斯优化算法对集成学习模型的超参数进行调优，以提升模型的性能。试验将对比使用默认超参数和经过贝叶斯优化后的超参数两种情况，对比二者在金浸出率预测中的差异。

表 3　集成学习模型经贝叶斯优化超参数调优前后的效果对比

Table 3　Comparison of effect of ensemble learning model before and after hyperparameter tuning by Bayesian optimization

超参数优化	MSE	R^2	训练时间/s
调优前	0.042	0.92	130
贝叶斯调优后	0.036	0.95	180

从表 3 看出：经过贝叶斯优化超参数调优后的集成学习模型的性能显著提升；优化后模型的 MSE 为 0.036，相比于默认超参数的 0.042，下降 14.3% 左右。

5　验证试验

5.1　试验环境与数据采集预处理

试验环境：集成学习模型试验在某金冶炼厂的实际生产环境中进行，试验硬件配置采用高性能服务器集群。

数据采集：试验数据来源于该冶炼厂金浸出过程的实时监控系统。数据采集设备有温度传感器、pH 传感器、氧气浓度探测器、矿浆浓度测量仪、浸出时间记录器等。数据记录频率为每分钟一次，采集的主要特征为反应温度、酸度、矿浆浓度、氧气浓度、浸出时间等。总数据集包含约70 000条样本。

数据预处理：在数据采集的基础上，首先对缺失值进行处理，使用均值插补法填补缺失数据。为保证数据质量，去除异常值和不符合实际生产条件的数据样本。之后对特征进行标准化处理。

5.2　验证试验结果及分析

基于集成学习模型与传统的单一 XGBoost 和 LightGBM 模型在真实生产环境中的预测精度对比结果见表 4。试验使用冶炼厂提供的实时数据集，包括温度、酸度、矿浆浓度、氧气浓度、浸出时间等输入特征。每个模型分别在相同的数据集上进行训练和测试，并以 MSE 和 R^2 为评价指标，对比各模型的预测效果，从而评估集成学习模型在实际应用中的优势。

表 4　不同模型在真实生产环境中的预测精度对比

Table 4　Comparison of prediction accuracy of different models in real production environment

模型	MSE	R^2
XGBoost	0.052	0.89
LightGBM	0.048	0.90
本研究模型	0.037	0.94

从表 4 看出：集成学习模型的 MSE 和 R^2 均优于传统的 XGBoost 和 LightGBM 模型；集成模型的 MSE 为 0.037，较 XGBoost(0.052) 和 LightGBM (0.048)分别降低约 28.8% 和 22.9% 左右；同时，集成学习模型的 R^2 达 0.94，表现出更强的拟合能力。由此可知，集成学习模型能更好地处理采集所得真实数据中的复杂非线性关系，从而提高金浸出率预测精度，且稳定性较好。

为探讨集成学习模型在不同数据集规模下的预测性能，使用冶炼厂采集的不同规模的数据集

(10 000、30 000、70 000 条样本),分别训练集成学习模型并进行验证。评价指标为 MSE、R^2 和训练时间,分析集成学习模型在小规模和大规模数据集上的适应性、精度和计算效率,试验结果见表 5。

表 5 集成学习模型在不同数据集规模下的预测性能对比

Table 5 Comparison of predictive performance of ensemble learning models under different data set sizes

数据集规模/条	MSE	R^2	训练时间/s
10 000	0.046	0.91	85
30 000	0.039	0.93	130
70 000	0.037	0.94	180

从表 5 看出:对于 10 000 条样本的数据集,MSE 为 0.046,而随数据集增至 30 000 和 70 000 条,MSE 分别降至 0.039 和 0.037,R^2 也从 0.91 分别升至 0.93 和 0.94,可见随数据集规模增大,集成学习模型的 MSE 逐渐降低,R^2 逐渐升高,说明金浸出率预测精度随数据集规模增大而逐渐升高;训练时间从 10 000 条样本的 85 s 延长至70 000条样本 180 s,试验中发现,训练时间虽随数据集规模增大而延长,但在大数据集上的精度有所提升,说明集成学习模型能够有效处理大规模数据,保持良好的预测能力和稳定性。

为探讨集成学习模型与传统的单一 XGBoost 和 LightGBM 模型在超参数优化下的预测精度差异,从而评价超参数调优对模型精度的提升作用,通过贝叶斯优化算法对每个模型的关键超参数进行调优,对比不同优化方法下的模型性能,评价指标为 MSE、R^2 和训练时间,试验结果见表 6。

表 6 不同模型在超参数优化下的性能对比

Table 6 Performance comparison of different models under hyperparameter optimization

模型	MSE	R^2	训练时间/s
XGBoost(优化)	0.039	0.93	95
LightGBM(优化)	0.037	0.94	110
本研究模型	0.033	0.96	160

从表 6 看出:经过贝叶斯优化后的集成学习模型的 MSE 为 0.033,R^2 为 0.96,均优于单一的 XGBoost 和 LightGBM 模型,说明集成学习模型在超参数优化后表现出更高的准确性和稳定性;集成学习模型经超参数调优后训练时间为 160 s,比单一的 XGBoost 和 LightGBM 模型的训练时间长,但从精度提升的角度来看,优化后的集成模型在实际应用中具有更高的价值,能够提供更精确的金浸出率预测结果。

6 结论

研究提出了一种基于集成学习的金浸出率智能预测模型,通过将改进的 XGBoost 和 LightGBM 模型相结合,充分发挥二者的互补优势,提升了金浸出率的预测精度和稳定性。集成学习模型在处理复杂非线性关系和噪声数据方面表现优异,相比传统模型,具有更低的均方误差和更高的决定系数。研究结果可为金浸出率的智能预测提供一种高效、精确的解决方案。

参考文献:

[1] 胡广浩,毛志忠,何大阔. 浸出过程浸出率预测与在线优化[J]. 系统工程学报,2011,26(4):524-530.
HU Guanghao, MAO Zhizhong, HE Dakuo. Prediction of leaching rate and on-line optimization operation of leaching process [J]. Journal of Systems Engineering, 2011, 26(4): 524-530.

[2] 白璐. 湿法冶金浸出过程非线性预测控制方法的研究[D]. 沈阳:东北大学,2019.

[3] 李淑芹,王宇斌,毛欣钰,等. 基于不同算法的磁化水水洗去除飞灰中可溶性氯盐浸出率预测模型[J]. 化工矿物与加工,2022,51(1):44-48.
LI Shuqin, WANG Yubin, MAO Xinyu, et al. A prediction model for the leaching rate of soluble chlorides from washing fly ash with magnetic water based on different algorithms[J]. Industrial Minerals & Processing, 2022, 51(1): 44-48.

[4] 郭计敏. 金氰化浸出过程非线性预测控制方法及仿真算法分析研究[J]. 湿法冶金,2023,42(6):659-665.
GUO Jiming. Nonlinear predictive control method and simulation algorithm analysis of gold cyanide leaching process [J]. Hydrometallurgy of China, 2023, 42(6): 659-665.

[5] 莫文水. 金氰化浸出过程混合建模及仿真算法分析研究[J]. 湿法冶金,2023,42(4):429-435.
MO Wenshui. Hybrid modeling and simulation algorithm of gold cyanide leaching process[J]. Hydrometallurgy of China, 2023, 42(4): 429-435.

[6] 刘尊方,雷浩川,盛海彦. 基于 XGBoost 模型的湟水流域耕地土壤养分遥感反演[J]. 干旱区地理,2023,46(10):1643-1653.
LIU Zunfang, LEI Haochuan, SHENG Haiyan. Remote

sensing inversion of soil nutrient on farmland in Huangshui River Basin based on XGBoost model [J]. Arid Land Geography,2023,46(10):1643-1653.

[7] 郝若飞,赵松.基于 XGBoost 模型的缺血性脑卒中患者中医药治疗复发预测研究[J].中国药物应用与监测,2023,20(6):441-447.
HAO Ruofei, ZHAO Song. Prediction of recurrence in ischemic stroke patients treated with traditional Chinese medicine based on XGBoost model [J]. China Drug Application and Monitoring,2023,20(6):441-447.

[8] BHUTTA A A,NISA M U,MIAN A N. Lightweight real-time WiFi-based intrusiondetection system using LightGBM [J]. Wireless networks,2024,30(2):749-761.

[9] 何哲,石玉玲,李富春,等.基于 LightGBM 模型的甘肃省临夏县滑坡易发性评价[J].水资源与水工程学报,2024,35(1):197-205.
HE Zhe,SHI Yuling,LI Fuchun,et al. Landslide susceptibility assessment based on LightGBM model in Linxia County,Gansu Province[J]. Journal of Water Resources and Hydraulic Engineering,2024,35(1):197-205.

[10] 曲羽姗.基于 SMOTE 和 LR-RF-AdaBoost-XGBoost 模型的银行代发工资客户流失预测研究[D].大连:东北财经大学,2023.

[11] 李赟波,王士同.多源域分布下优化权重的迁移学习 Boosting 方法[J].计算机科学与探索,2023,17(6):1441-1452.
LI Yunbo,WANG Shitong. Transfer learning Boosting for weight optimization under multi-source domain distribution[J]. Journal of Frontiers of Computer Science & Technology,2023,17(6):1441-1452.

[12] 牛庆威,张如玉,白雨昊,等.基于梯度提升决策树(GBDT)的低阻油层识别[J].计算机与数字工程,2023,51(6):1428-1432.
NIU Qingwei, ZHANG Ruyu, BAI Yuhao, et al. Low resistivity reservoir identification based on gradient boosting decision tree(GBDT) [J]. Computer & Digital Engineering,2023,51(6):1428-1432.

[13] 李亚茹,张宇来,王佳晨.面向超参数估计的贝叶斯优化方法综述[J].计算机科学,2022,49(增刊 1):86-92.
LI Yarou, ZHANG Yulai, WANG Jiachen. Survey on Bayesian optimization methods for hyper-parameter tuning [J]. Computer Science,2022,49(Sup. 1):86-92.

[14] 卫升,王艳,纪志成.多工况生产过程下的即时学习能耗预测建模方法[J].系统仿真学报,2024,36(6):1378-1391.
WEI Sheng,WANG Yan,JI Zhicheng. Just-in-time learning energy consumption predictive modeling method in multi-condition production process[J]. Journal of System Simulation,2024,36(6):1378-1391.

[15] AHN J M,KIM J,KIM K. Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting[J]. Toxins,2023,15(10). DOI:10.3390/toxins15100608.

Design of Intelligent Prediction Model for Gold Leaching Rate Based on Ensemble Learning Algorithms

ZHOU Yifan[1,2]

(1. *School of Information Engineering, Zhumadian Vocational and Technical College, Zhumadian* 463000, *China*;
2. *Key Laboratory of HarmonyOS Ecosystem Development, Zhumadian* 463000, *China*)

Abstract: To address the problem of low prediction accuracy and large noise data in gold leaching rate prediction, a high-efficiency intelligent prediction method was proposed. Combining the improved XGBoost and LightGBM models, a dynamic learning rate, regularization optimization, and Bayesian optimization algorithm were used to design an ensemble learning model for gold leaching rate prediction. The results show that the ensemble learning model has a significantly improved prediction accuracy compared to traditional single models, and its MSE is about 28.8% and 22.9% lower than XGBoost and LightGBM, respectively. The model has high stability in real production environments. The ensemble learning model can effectively improve the prediction accuracy of gold leaching rate and has strong practical application value.

Key words: gold leaching rate; prediction; ensemble learning; XGBoost model; LightGBM model; numerical simulation