

基于 XGBoost 算法的摇床分选指标预测研究^①

何微¹, 戴昊霖^{1,2}, 陈雯¹

(1.长沙矿冶研究院有限责任公司, 湖南长沙 410012; 2.中南大学 自动化学院, 湖南长沙 410083)

摘要:提出了一种基于极限梯度提升(XGBoost)算法的摇床分选指标预测方法。该方法通过颜色矩量化矿带的颜色特征,图像矩量化矿带的形态特征,灰度共生矩阵评价指标的对比度、同质性、相关性和ASM能量量化矿带的纹理特征,根据分选过程中矿带颜色、形态和纹理的变化有效提取摇床矿带特征;然后将图像特征作为输入、选矿指标作为输出,利用XGBoost算法筛选矿带特征并构建预测模型,训练模型并在测试集上预测精矿品位、回收率和精矿产率,实现对分选指标的准确预测。通过与决策树模型和随机森林模型的精矿品位、精矿产率和回收率预测结果对比,基于XGBoost算法的模型对精矿回收率、产率预测精度高。

关键词:摇床; 特征提取; 品位预测; XGBoost; 特征选择; 图像识别; 智能控制; 分选指标; 预测模型; 矿带

中图分类号: TD922; TP274

文献标志码: A

doi:10.3969/j.issn.0253-6099.2025.01.013

文章编号: 0253-6099(2025)01-0070-06

Prediction of Separation Indices of Shaking Table Based on XGBoost

HE Wei¹, DAI Haolin^{1,2}, CHEN Wen¹

(1.Changsha Research Institute of Mining and Metallurgy Co., Ltd., Changsha 410012, Hunan, China; 2.School of Automation, Central South University, Changsha 410083, Hunan, China)

Abstract: A prediction method for mineral processing indices of shaking table was proposed based on eXtreme Gradient Boosting (XGBoost). As for the separation product zone, it can quantify the colour difference by colour moments, the shape characteristics by image moments, and the texture characteristics by evaluation indices of grey-level co-occurrence matrix, including contrast, homogeneity, correlation and ASM energy. The features of the separation product zone can be effectively extracted according to the difference in its colour, shape and texture during separation process. Subsequently, those features can be filtered by employing XGBoost, with which a prediction model can be then constructed. After being trained with the test set to predict the grade, recovery rate and yield of concentrate, this model can achieve accurate prediction of separation indices. It is found that the XGBoost-based model outperforms the decision tree model and random forest model in terms of accuracy when applied for predicting recovery and yield of concentrate.

Key words: shaking table; feature extraction; grade prediction; XGBoost; feature selection; image recognition; intelligent control; separation index; prediction model; separation product zone

摇床处理重金属矿物(包括稀土、钨、锡、钽和铌矿石)具有富集比高、分选效率高、床面矿物分带清晰等优点,但实际生产中,摇床存在自动化水平低、工人工作强度大等问题。利用图像识别技术检测矿带位置,标定精矿带、中矿带和尾矿带位置,调整接矿装置,可实现摇床的智能控制^[1-3]。正确提取可以反映摇床分选效率的矿带特征,依据矿带特征建立分选指标预测模型对实现摇床智能化至关重要。

基于提取方式不同,特征提取算法可分为两大类:一类是需要人工提取图像特征的算法,如基于颜色和形态特征的传统方法;另一类是使用基于深度神经网络的自动特征提取方法^[4]。传统图像特征提取方法通常基于图像的局部结构、纹理、颜色等特征,这些特征在图像中的表达相对稳定,不受噪声和变形的影响,具有较好的稳定性。基于颜色、形态提取图像特征的方法已在浮选行业得到广泛应用。摇床的分选机理是

① 收稿日期: 2024-09-06

作者简介: 何微(1998—),男,甘肃定西人,硕士研究生,主要研究方向为选矿过程智能控制。E-mail: 1208146347@qq.com

通信作者: 陈雯(1965—),女,四川资阳人,博士,正高级工程师,主要从事复杂铁矿资源综合利用计算研发工作。E-mail: chen_wen@vip.sina.com

通过矿物密度差异得到精矿带、尾矿带,以矿带颜色、形态、纹理特征作为模型输入建立预测模型,可以较好解释分选指标的波动原因。在采集的彩色床面图像中,矿带是众多像素点的集合,需要对该集合内所有点的颜色特征进行定量提取,颜色矩方法可以定量描述区域内的颜色信息。矿带形态特征能反映矿物颗粒聚集数量与运动情况。矿带纹理特征可反映不同分选条件下精矿区和尾矿区图像的纹理差异^[5]。综上所述,摇床的图像特征较为复杂,特征维度高。若要实现摇床产品品位预测,需利用合适的方法提取摇床图像特征,利用合适的机器学习模型建立矿带图像与分选指标之间的映射关系。

针对选矿指标难以预测、矿带图像难以检测、矿带特征众多的问题,在构建机器学习模型时,为了达到对数据的合理利用,应当选取样本容量小、特征维度高、准确性高的模型。极限梯度提升(eXtreme Gradient Boosting, XGBoost)算法处理小样本、高维度数据,具有模型优越性以及较强的泛化能力和预测能力^[6]。本文运用 XGBoost 算法建立精矿产率、精矿品位、精矿回收率预测模型并依据特征重要性选取特征,利用矿带图像特征及产率、品位、回收率数据对模型进行训练,然后用于相关预测。

1 数据采集与处理

1.1 试验平台与数据采集

摇床试验平台及图像采集模块如图 1 所示。以某钨矿的摇床分选试验作为研究对象,在摇床稳定运行后采集图像,并将摇床产品送样分析化验钨品位,计算钨精矿产率和钨回收率。试验共采集到摇床床面分带图像 530 张,分析并计算相应的摇床试验指标 530 组。通过图像处理对图片进行增强、去噪和分割,提取矿带颜色特征,形态特征和纹理特征,建立图像与分选指标的数据集。

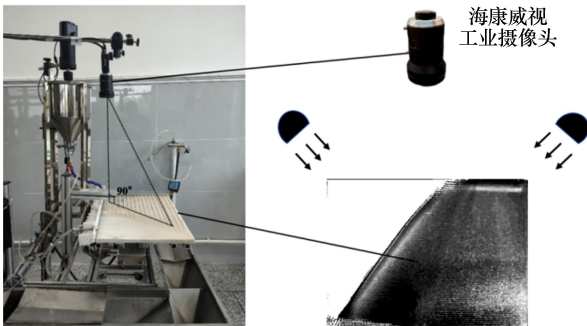


图 1 试验平台及图像采集模块

Fig. 1 Experimental platform and image acquisition module

1.2 图像处理

1.2.1 颜色特征提取

颜色矩是一种简单而有效的颜色特征表示方法。颜色信息主要分布于低阶矩中,因此本文使用颜色的一阶、二阶和三阶矩表达图像的颜色分布。其中一阶颜色矩采用一阶原点矩,反映图像的整体明暗程度;二阶颜色矩采用二阶中心矩的平方根,反映图像的颜色分布范围;三阶颜色矩采用三阶中心矩的立方根,反映图像颜色分布的对称性。对应的计算公式见式(1)~(3)。

一阶颜色矩:

$$E_i = \frac{1}{N} \sum_{j=1}^N P_{ij} \quad (1)$$

二阶颜色矩:

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (P_{ij} - E_i)^2} \quad (2)$$

三阶颜色矩:

$$S_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (P_{ij} - E_i)^3} \quad (3)$$

式中 P_{ij} 为第 i 个颜色通道中第 j 个像素的颜色值。

1.2.2 形态特征提取

矿带形态特征是关于位置、面积和倾斜程度的描述。例如,横向冲洗水量增加时,矿物颗粒的横向运动速度加快,矿带面积减小,倾斜程度减小,同时矿带中心位置可能会向给矿端移动。精矿带形态如图 2 所示,从(a)到(b)再到(c),横向冲洗水量依次增大,矿带中心位置逐渐向床头移动;精矿带向上倾斜时,其产率随之增加。为了量化上述特征,采用图像矩对矿带形态特征进行提取。

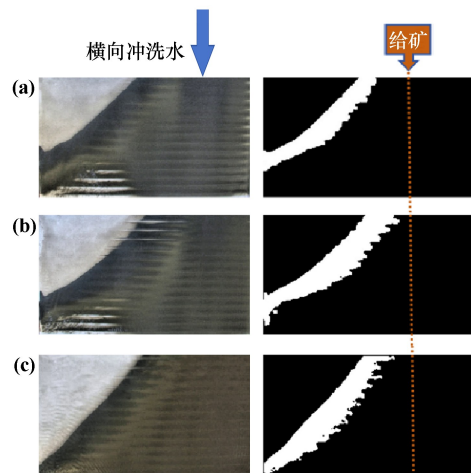


图 2 精矿带形态示意图

Fig. 2 Morphology of concentrate zone

在图形处理领域中,图像矩是一种对图形像素点

灰度值进行加权处理的方法。图像矩通常用来描述分割后的图像对象,是提取图像形状和方向的描述。图像矩包括面积(或总体亮度)以及有关几何中心和方向的信息,其公式为:

$$m_{ji} = \sum_{x,y} (P(x,y) \cdot x^j \cdot y^i) \quad (4)$$

式中: m_{ji} 为第 $j+i$ 阶图像矩; x 与 y 均为像素坐标值; P 为像素的坐标强度。

对于经过0和1二值化的矿带图像,图像一阶矩可表示矿带面积;由于矿带区域像素值均为1,二阶矩计算得到的质心即为矿带几何中心——形心;利用三阶矩可计算得到矿带的主轴与水平方向的夹角。利用各矿带像素点集合,构建原图矿带的0-1二值化图像,计算其一阶、二阶、三阶图像矩,即可提取矿带区域的面积、几何中心以及斜度。图3为矿带分割-二值化示意图。

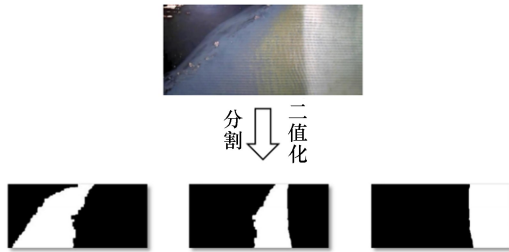


图3 矿带分割-二值化示意图

Fig.3 Segmentation and binarization of separation product zone

为验证图像矩形态特征与分选状态的联系,计算矿带面积、中心点 x 及 y 坐标、斜度与回收率、选矿效率最大信息系数(the maximal information coefficient, MIC)相关系数,结果见表1。

表1 图像矩 MIC 相关系数

Table 1 MIC correlation coefficient of image moments

区域	特征名称	MIC 相关系数	
		回收率	选矿效率
精矿带	面积	0.425	0.365
	中心点 x 坐标	0.196	0.247
	中心点 y 坐标	0.233	0.226
	斜度	0.324	0.246
尾矿带	面积	0.214	0.168
	中心点 x 坐标	0.193	0.158
	中心点 y 坐标	0.195	0.176
	斜度	0.186	0.183

通过灰度共生矩阵描述矿带图像纹理特征。不同分选条件下精选区矿层刻槽纹理不同,表示此时矿层厚度不一。矿层厚时,露出刻槽浅,纹理弱;矿层薄时,露出刻槽深,纹理强,此处矿物颗粒受床面往复运动影响,呈现波纹状或拖曳状运动轨迹,尖灭线左侧由于没

有刻槽束缚,此种轨迹更加明显。受不同床面运动情况与冲洗水影响,此种波纹状运动轨迹不同,图像中纹理也有差异。为定量描述床面图像中的纹理特征,图4给出了摇床给矿量500g和750g时的摇床刻槽及运动纹理对比。给矿量500g时,床面矿带分布均匀,床条裸露;给矿量750g时,床面矿带分布均匀,但床条明显被矿物没过,从而影响床面图像的纹理特征。

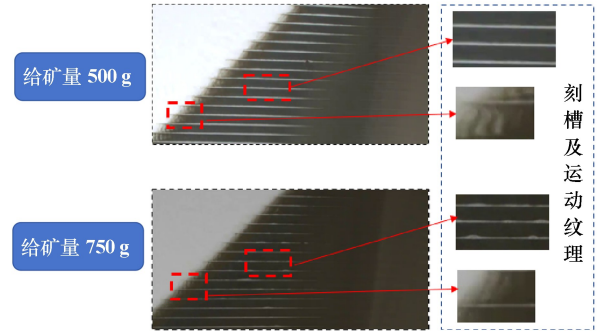


图4 不同给矿量条件下刻槽及运动纹理对比图

Fig.4 Comparison of grooves and textures under different conditions

灰度共生矩阵共有14种评价统计量,根据床面图像中的纹理特点,使用4种评价灰度共生矩阵的统计量,分别为对比度、同质性、相关性以及ASM(angular second moment)能量。根据纹理大小,设置计算时的像素对距离为 $d=4$ 和 $d=2$,从 0° 、 45° 、 90° 、 135° 共4个方向进行计算。

颜色矩一阶至三阶矩H、S、V分量3个通道共9个特征;形态特征包含面积以及中心点 x 、 y 坐标和斜度共4个特征;纹理特征值包括对比度、相关性、同质性和ASM能量4个分量,在2个距离和4个方向上共32个特征。因此精矿带和尾矿带图像共有 $2 \times (9+4+32) = 90$ 个特征。特征数据集部分数据如表2所示,分选指标数据集部分数据如表3所示。

表2 图像特征数据集部分数据

Table 2 Partial data in image characteristics dataset

样本数	一阶颜色矩	二阶颜色矩	三阶颜色矩	...	灰度共生矩阵
1	0.042	101	140	...	0.182
2	0.042	102	139	...	0.191
3	0.044	98	141	...	0.194
			⋮		
530	0.043	93	144	...	0.185

2 基于XGBoost算法的图像特征筛选

在机器学习中,有效选择特征用于建立特征子集对提高模型准确性至关重要^[7]。经典的决策树算法

表 3 分选指标数据集部分数据

Table 3 Partial data in separation index dataset

样本数	精矿产率/%	精矿品位/%	回收率/%
1	21.90	0.34	46.43
2	37.94	0.27	60.01
3	54.76	0.30	79.06
		⋮	
530	12.31	0.73	48.24

(classification and regression tree, CART) 利用基尼系数、信息增益和特征分裂次数来添加权重计算特征重要性^[8]。将特征作为分裂特征次数的这种方法考虑了特征在树节点分裂中的使用频率,即参与构建树的次数。通常来说,越早参与决策树的节点分裂的特征重要程度越高,可以采用权重加权的方式计算其累计分值。

XGBoost 算法内置了特征重要性函数,用于计算特征的重要性,并剔除贡献度相对较低的特征,从而增强模型的鲁棒性,并实现特征降维的目标。XGBoost 算法内置的 cover 特征重要性计算方法适用于枚举特征,且不会过度拟合目标函数,不受目标函数数量纲的影响。cover 方法计算公式为:

$$V(k) = \frac{\sum_{t=1}^T \sum_{i=1}^{N(t)} I[\beta(t,i) = k] H_y(t,i)}{\sum_{t=1}^T \sum_{i=1}^{N(t)} I[\beta(t,i) = k]} \quad (5)$$

式中: T 为所有树的数量; $N(t)$ 为第 t 棵树的非叶子节点个数; $\beta(t,i)$ 为第 t 棵树第 i 个非叶子节点的划分特征; $H_y(t,i)$ 为落在第 t 棵树第 i 个非叶子节点上所有样本的二阶导数之和; $I(\cdot)$ 为指示函数。

本文借助 XGBoost 算法,采用 cover 特征重要性计算方法计算特征重要性。将提取的 90 个特征与精矿品位进行相关性分析,得到与精矿品位相关的特征 58 个。用皮尔逊相关系数表示特征与精矿品位的相关性,图像特征与精矿品位相关性分析结果如图 5 所示。对采集的 58 个相关特征数据进行特征重要性筛选,特征重要性贡献度如图 6 所示。

对模型选取不同数目时预测的准确度进行检验。在二分类问题中,准确度表示的是被分类为正例的样本中真正属于正例的比例,它是衡量分类器在预测正例时准确性的指标之一,取值范围为 $[0, 1]$ 。准确度越接近 1,表示分类器预测的正例中有更多的真正例,即分类器的预测更加准确;准确度接近 0 时,表示分类器预测的正例中有更多的误判负例,即分类器的预测不够准确。

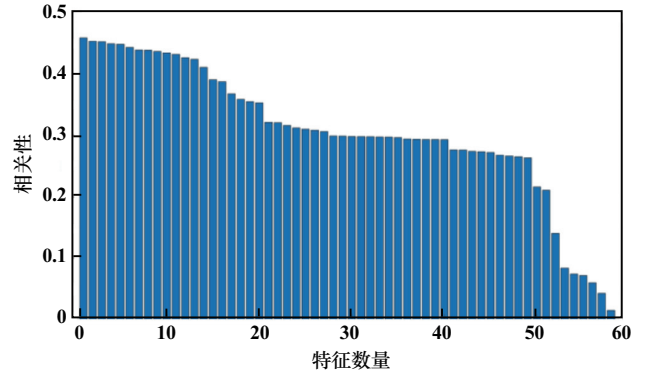


图 5 图像特征与精矿品位相关性分析结果

Fig. 5 Correlation between image characteristics and concentrate grades

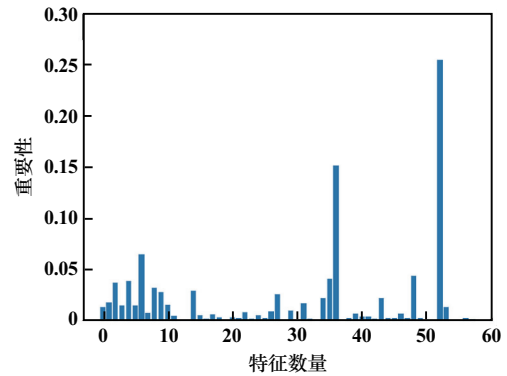


图 6 特征重要性贡献度

Fig. 6 Contribution degree of feature importances

计算得到不同数量特征下 XGBoost 算法的准确度如图 7 所示,特征数量与程序运行时间的关系如图 8 所示。由图 7 可知,特征数量为 16、18、22 时,模型预测准确度均可达到 92.8% 左右。由图 8 可知,随着特征维度逐渐增加,模型计算量增多,程序运行时间增加,特别是特征数量超过 21 后模型运行时间大大增加。特征数量为 16 时,模型预测准确度达到要求且模型运行时间更短,筛选后的特征数量选择 16 个。筛选得到的特征数据集包含 16 个特征,见表 4。

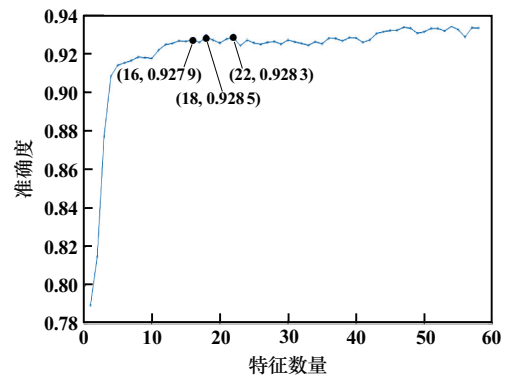


图 7 不同数量特征下 XGBoost 算法的准确度

Fig. 7 Precision of XGBoost algorithm with different feature number

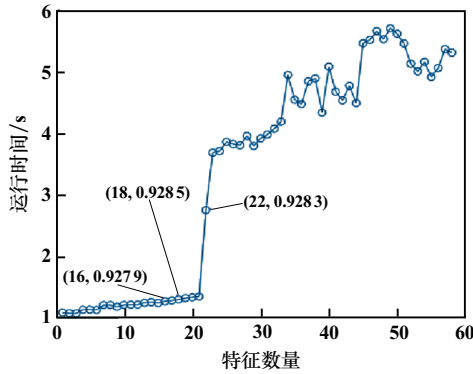


图8 特征数量与程序运行时间关系

Fig.8 Relationship between feature number and running time of programs

表4 特征筛选结果

Table 4 Results of feature screening

特征编号	特征重要性	特征名称
1	0.021	精矿带 H 分量一阶矩
2	0.041	精矿带 S 分量一阶矩
4	0.043	精矿带 H 分量二阶矩
6	0.072	精矿带 V 分量二阶矩
8	0.037	精矿带 S 分量三阶矩
9	0.035	精矿带 V 分量三阶矩
10	0.015	尾矿带 H 分量一阶矩
14	0.035	尾矿带 S 分量二阶矩
27	0.035	精矿带面积
31	0.022	精矿带中心点 x 坐标
34	0.029	精矿带中心点 y 坐标
35	0.048	精矿带斜度
36	0.150	尾矿带面积
43	0.028	尾矿带中心点 x 坐标
48	0.049	尾矿带中心点 y 坐标
52	0.255	灰度共生矩阵 ASM($d=2$)

3 摇床分选指标预测

3.1 特征筛选前后分选指标预测实验

选取相关系数 R^2 衡量模型对数据的拟合程度,选取平均绝对误差 (E_{MAE}) 和均方误差 (E_{MSE}) 衡量模型的预测效果:

$$R^2 = 1 - \frac{S_{res}}{S_{tot}} \quad (6)$$

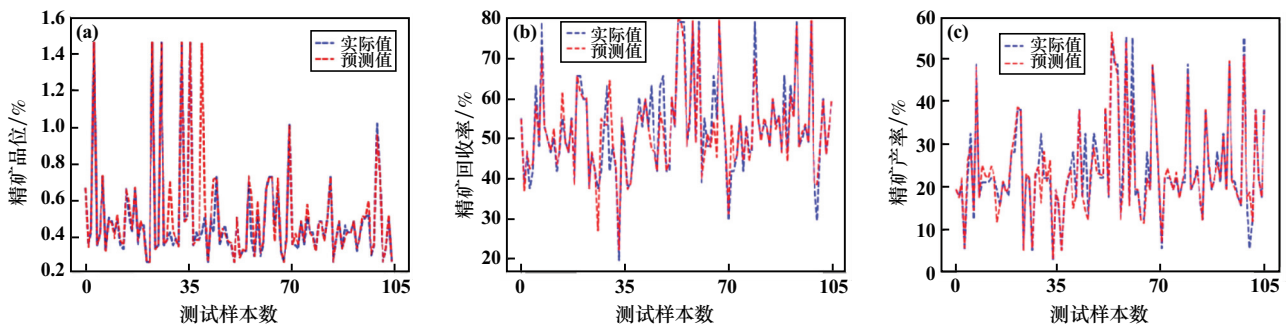
$$E_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

$$E_{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

式中: S_{res} 为残差平方和; S_{tot} 为总平方和; y_i 为观测值; \hat{y}_i 为对应的预测值; n 为样本数量。

R^2 越接近 1, 表示模型对数据的拟合程度越好; E_{MSE} 和 E_{MAE} 越小, 表示预测效果越好。

特征筛选后 XGBoost 算法模型预测效果如图 9 所示, 分别对精矿品位、精矿回收率、精矿产率进行预测, 预测效果评价见表 5。特征筛选后, 精矿产率 R^2 进一步提升, E_{MAE} 和 E_{MSE} 分别下降 6.5% 和 21.5%; 精矿品位 R^2 提升 2.3%, E_{MAE} 和 E_{MSE} 分别下降 24.9% 和 12.5%; 精矿回收率 R^2 提升 6.7%, E_{MAE} 和 E_{MSE} 分别下降 19.0% 和 37.6%。特征筛选后的 XGBoost 算法模型对于数据的拟合效果更好, 模型的预测精度得到提升, 表明采用 XGBoost 算法对特征预先进行筛选可以大幅提高模型的预测精度, 验证了根据模型预测精度确定重要特征的合理性。



(a) 精矿品位; (b) 精矿回收率; (c) 精矿产率

图9 XGBoost 算法模型预测效果

Fig.9 Prediction results of XGBoost algorithm model

3.2 预测性能对比

选取经典的决策树模型 (decision tree regression, DTR) 和随机森林模型 (random forest, RF) 构建预测模型, 选取经过特征筛选后的特征作为模型输入, 与

XGboost 算法模型预测效果进行对比, 结果如表 6 所示。从表 6 可以看出, 多次对比试验下, XGBoost 模型对数据的拟合程度较好, 精矿产率和回收率预测模型的 R^2 明显高于 DTR 模型和 RF 模型, E_{MSE} 低于 DTR

模型和 RF 模型,表明 XGBoost 算法模型在精矿产率和回收率预测方面具有良好的应用前景。

表 5 预测效果评价

Table 5 Evaluation of prediction results

模型	评价指标	R^2	E_{MAE}	E_{MSE}
特征筛选前	精矿产率	0.871 2	1.964 5	18.318 4
	精矿品位	0.792 3	0.042 1	0.014 4
	回收率	0.792 6	3.050 1	34.811 1
特征筛选后	精矿产率	0.892 6	1.836 6	14.374 6
	精矿品位	0.810 5	0.031 6	0.012 6
	回收率	0.845 8	2.472 0	21.719 0

表 6 模型预测结果

Table 6 Prediction result of models

指标名称	评价模型	R^2	E_{MAE}	E_{MSE}
精矿产率	DTR	0.836 3	1.918 8	21.908 8
	RF	0.888 3	2.010 2	14.953 2
	XGBoost	0.892 6	1.836 6	14.374 6
精矿品位	DTR	0.808 4	0.037 9	0.012 7
	RF	0.821 0	0.037 6	0.021 9
	XGBoost	0.810 5	0.031 6	0.012 6
回收率	DTR	0.643 8	2.564 3	50.169 6
	RF	0.766 3	3.284 6	32.918 1
	XGBoost	0.845 8	2.472 0	21.719 0

4 结论

提出一种基于 XGBoost 特征重要性筛选的摇床分选指标预测方法。通过 XGBoost 模型对特征的重要性进行评估,筛选从摇床矿带图像提取到的特征从而确定特征子集,以此来提升模型的泛化能力和预测精度,同时避免输入模型的数据维度过高导致“维度灾难”、陷入局部最优。利用实验室摇床图像数据构建预测模型,得到以下结论:

1) 经过特征筛选后的 XGBoost 模型对于精矿产率、精矿品位、精矿回收率预测精度都有所提高。精矿产率的 R^2 进一步提升, E_{MAE} 和 E_{MSE} 分别下降 6.5% 和 21.5%; 精矿品位的 R^2 提升 2.3%, E_{MAE} 和 E_{MSE} 分别下降 24.9% 和 12.5%; 精矿回收率的 R^2 提升 6.7%, MAE 和 MSE 分别下降 19.0% 和 37.6%。

2) XGBoost 模型对于精矿回收率、产率的预测拟合程度较高,且 XGBoost 模型的 E_{MAE} 和 E_{MSE} 值较小,说

明 XGBoost 模型对于精矿回收率、产率预测精度高,为精矿回收率、产率预测提供了新途径。

参考文献 (References):

- [1] 杨文旺,何庆浪,兰希雄,等. 选矿摇床智能巡检机器人开发与应用[J]. 有色金属(选矿部分), 2020(5):102-106.
YANG Wenwang, HE Qinglang, LAN Xixiong, et al. Development and application of intelligent inspection robot for table concentrator[J]. Nonferrous Metals (Mineral Processing Section), 2020(5):102-106.
- [2] 杨文旺,武涛,赵玉华,等. 重选摇床智能控制系统的设计与应用[J]. 黄金, 2020,41(8):57-60.
YANG Wenwang, WU Tao, ZHAO Yuhua, et al. Design and application of intelligent control system for gravity separation table concentrator[J]. Gold, 2020,41(8):57-60.
- [3] 李驰骏. 基于视觉的摇床自动接矿系统设计[D]. 昆明:昆明理工大学, 2020.
LI Chiqin. Design of table automatic ore receiving system based on vision[D]. Kunming: Kunming University of Science and Technology, 2020.
- [4] 史世阳. 基于改进深度森林的图像分类算法研究[D]. 郑州:中原工学院, 2024.
SHI Shiyang. Research on image classification algorithm based on improved depth forest[D]. Zhengzhou: Zhongyuan University of Technology, 2024.
- [5] 戴昊霖,陈雯,蒋朝辉,等. 基于纹理去噪的摇床多矿带分割方法[J]. 有色金属(选矿部分), 2023(6):183-190.
DAI Haolin, CHEN Wen, JIANG Zhaohui, et al. Multiple ore zones segmentation method of shaking table based on texture denoising[J]. Nonferrous Metals (Mineral Processing Section), 2023(6):183-190.
- [6] LIU W, CHEN Z, HU Y. XGBoost algorithm-based prediction of safety assessment for pipelines[J]. International Journal of Pressure Vessels and Piping, 2022,197:104655.
- [7] HE H J, DENG G M. Grouped feature screening for ultra-high dimensional data for the classification model[J]. Journal of Statistical Computation and Simulation, 2022,92(5):974-997.
- [8] 王婕婷,李飞江,李珏,等. 缓解随机一致性的基尼指数与决策树方法[J]. 中国科学:信息科学, 2024,54(1):159-190.
WANG Jieting, LI Feijiang, LI Yu, et al. Gini index and decision tree method with mitigating random consistency[J]. Scientia Sinica (Informationis), 2024,54(1):159-190.

引用本文:何微,戴昊霖,陈雯. 基于 XGBoost 算法的摇床分选指标预测研究[J]. 矿冶工程, 2025,45(1):70-75.

HE Wei, DAI Haolin, CHEN Wen. Prediction of separation indices of shaking table based on XGBoost[J]. Mining and Metallurgical Engineering, 2025,45(1):70-75.