

基于 OOA-CatBoost 的烧结矿粒度预测模型

李 喆^{a,c}, 王 猛^{a,c}, 董 振^{a,c}, 姜娟娟^{b,c}, 李 杰^{a,c}, 杨爱民^{b,c}

(华北理工大学 a. 冶金与能源学院; b. 理学院; c. 铁矿石优选与铁前工艺智能化河北省
工程研究中心, 河北 唐山 063210)

摘要: 在烧结生产过程中, 烧结矿粒度是评价烧结矿质量的重要指标之一。为提高烧结生产水平, 针对烧结过程检测的滞后性和粒级占比调控预测问题, 建立基于整体生产参数的烧结矿粒级占比预测模型; 将经过特征选择之后的 12 个关键参数作为输入变量, 对应的烧结矿粒级占比作为输出变量; 通过缺失数据填补、数据增强以及异常点替换等数据预处理方法, 获取用于预测烧结矿粒级占比的高质量数据集; 采用鱼鹰优化算法 (OOA) 和类别特征梯度提升算法 (CatBoost) 构建预测模型。结果表明, OOA-CatBoost 算法模型的平均绝对误差 (MAE) 为 0.276 9, 均方误差 (MSE) 为 0.043 3, 决定系数 (R^2) 为 0.949 9。对比侏儒猫鼬优化 (DMO) 算法、麻雀搜索算法优化 (SSA)、鱼鹰优化算法优化的随机森林 (RF)、轻量梯度提升机 (LightGBM) 以及极限梯度提升算法 (XGBoost) 等其他 11 个机器学习模型, 本文模型取得良好的预测效果。基于工业实测数据, OOA-CatBoost 算法对烧结矿粒级占比的平均预测误差达到 0.085 2, 可为优化原料配比、混合料参数和烧结机参数调控提供理论指导, 从而提升优质烧结矿的粒级占比。

关键词: 烧结矿粒级; 鱼鹰优化算法; CatBoost; OOA-CatBoost 算法; 预测

中图分类号: TF046.4

文献标志码: A

文章编号: 1000-8764(2025)03-0047-12

doi: 10.13403/j.sjqt.2025.03.041

Sinter grade prediction model based on OOA-CatBoost

LI Zhe^{a,c}, WANG Meng^{a,c}, DONG Zhen^{a,c}, JIANG Juanjuan^{b,c}, LI Jie^{a,c}, YANG Aimin^{b,c}

(North China University of Science and Technology: a. Institute of Metallurgy and Energy; b. College of Science;
c. Hebei Engineering Research Center for Iron Ore Optimization and Pre-Iron Process Intelligence,
Tangshan 063210, Hebei, China)

Abstract: In the process of sintering production, sinter grade is one of the important indicators to evaluate the quality of sinter. In order to improve the level of sintering production, a prediction model of sinter grade proportion based on the overall production parameters is established in order to solve the problems of hysteresis and grade proportion control and prediction in the sintering process. The 12 key parameters after feature selection are taken as input variables, and the corresponding sinter grade proportion is used as the output variable. Through data preprocessing methods such as missing data filling, data augmentation, and outlier point replacement, high-quality datasets for predicting the proportion of sinter fraction are obtained. The Osprey Optimization Algorithm (OOA) and Categorical Feature Gradient Boosting Algorithm (CatBoost) are used to construct the prediction model. The results show that the mean absolute error (MAE) of the OOA-CatBoost algorithm model is 0.276 9, the mean square error (MSE) is 0.043 3, and the coefficient of determination (R^2) is 0.949 9. Compared with 11 other machine learning models, such as the dwarf mongoose optimization (DMO) algorithm, the sparrow search

收稿日期: 2024-09-05; 修回日期: 2024-10-19

基金项目: 国家自然科学基金资助项目(52074126); 河北省自然科学基金资助项目(E2022209110)

作者简介: 李 喆 (1999—), 女, 硕士研究生, 从事冶金智能制造、工业大数据等方面的研究。

通信作者: 姜娟娟 (1995—), 女, 讲师, 从事烧结球团智能制造、矿物结构特征定量化表征、钢铁大数据深度挖掘及建模、冶金资源高效利用等方面的研究。

algorithm optimization (SSA), the random forest (RF) optimized by the osprey optimization algorithm, the lightweight gradient booster (LightGBM) and the extreme gradient boosting algorithm (XGBoost), the proposed model achieves good prediction results. Based on the industrial measured data, the average prediction error of the OOA-CatBoost algorithm for the proportion of sinter grade can reach 0.085 2, which can provide theoretical guidance for optimizing the ratio of raw materials, mixing parameters and sintering machine parameters, so as to improve the grade proportion of high-quality sinter.

Key words: sinter grade; osprey optimization algorithm; CatBoost; OOA-CatBoost algorithm; prediction

烧结矿是高炉炼铁的关键原料, 其在整个冶炼过程中占比超过 70%^[1-2]。确保烧结矿质量是高炉生产的基础^[3]。现阶段, 对于烧结矿质量的标准评估主要集中于还原性能、低温还原粉化率 (RDI)、软熔性、化学成分及其稳定性、转鼓强度等指标^[4-7]。然而, 随着技术的不断进步以及炉料结构的持续优化, 高炉炼铁对于烧结矿的粒度也提出了更高的标准。烧结矿粒度是影响高炉煤气利用率和燃料比的重要因素。不同容积的高炉对烧结矿粒度有不同要求, 不能简单认为粒度越大透气性越好或者粒度越小越好^[8]。准确控制烧结矿的粒度范围可为现场操作人员提供有价值的预测数据, 提高高炉冶炼效率, 同时可降低能源消耗^[9-10]。但由于烧结过程机理复杂, 涉及大量的物理化学变化, 其参数变量具有强耦合、非线性、大滞后等特点^[11], 难以实时获取, 从而影响对烧结矿粒度的调控。近年来, 随着机器学习的发展, 大量研究人员在预测烧结矿的化学成分和冶金性能方面进行了深入且广泛的探讨, 并获得了显著的研究成果^[12-16]。但相对而言, 对烧结矿粒度的预测研究较少。刘月明等^[17]提出了一种基于深度神经网络算法 (DNNRegressor) 的烧结矿粒度预测模型; 王诗歌^[18]基于像素点个数, 标定真实尺寸图像的像素, 求取了烧结矿的真实粒径。

随着智能化的发展, 具有自适应能力强、结构简单和鲁棒性高等特点的群体智能优化算法开始逐渐应用于冶金工业的回归预测问题中^[19-22]。针对烧结矿粒度预测精度低、模型鲁棒性差等问题, 本文提出一种鱼鹰优化算法 (osprey optimization algorithm, OOA) 与类别特征梯度提升算法 (categorical boosting, CatBoost) 相结合的方法 OOA-CatBoost, 作为烧结矿粒度预测模型。研究具体工作: ① 对采集数据进行数据预处理, 包括缺失数据填补、相关性分析、数据增强以及异常点

替补; ② 构建 OOA-CatBoost 模型, 对输入烧结矿粒度数据集进行训练预测; ③ 选取侏儒猫鼬优化算法 (DMO)、麻雀优化算法 (SSA) 分别与随机森林 (RF)、轻量梯度提升机 (LightGBM) 以及极限梯度提升算法 (XGBoost) 随机组合, 作为 OOA-CatBoost 模型的对比模型。本文旨在提高模型预测的精度, 并增强模型的鲁棒性与泛化性, 为烧结工艺生产提供理论与技术指导。

1 数据处理

烧结是将细粒矿粉转化为块状矿石的过程。在整个造块过程中, 通过对原料参数、操作参数和设备参数的细微调整, 确保所有性能指标和状态参数都达到最优, 从而生产出满足入炉标准的高质量烧结矿^[23-24]。其工艺流程如图 1 所示。

本研究基于国内某钢铁厂的数据信息, 详细统计了包括原材料的化学组成、混合料水分和粒度以及烧结机的关键参数等共计 27 个参数。具体数据详见表 1。由于条件的限制, 仅收集了 191 条数据, 并且其中部分数据存在缺失。为了确保试验结果的科学性, 需要对数据进行缺失值补充以及增强处理^[25]。

1.1 缺失数据填补

针对数据缺失问题, 可采用统计学方法和机器学习两类方法。统计学方法包括插值法、回归填充等, 其对于缺失数据预测有很高的精确度。但此类方法参数繁多, 很难准确地建立高维数据变量的联合分布^[26-27]。随着人工智能的发展, 通过机器学习的方法可以有效建立高维数据序列, 挖掘数据中潜在规律, 进而对缺失数据进行填补^[28]。本研究采用 2018 年 YOON 等^[29]提出的 Gain 算法对烧结生产的缺失数据进行填补, 其结构如图 2 所示。

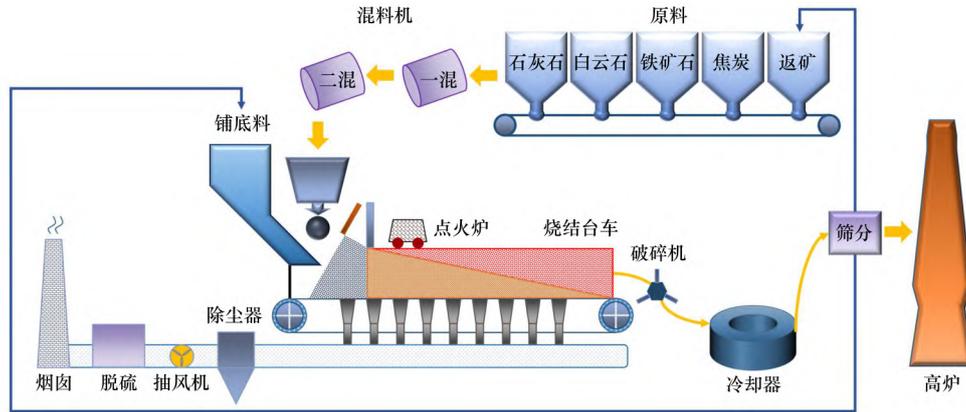


图 1 烧结工艺流程

Fig. 1 Flow chart of sintering process

表 1 特征参数汇总

Table 1 Gain model training parameter

参数类别	特征参数名称
化学成分	$w(\text{TFe})$ 、 $w(\text{CaO})$ 、 $w(\text{MgO})$ 、 $w(\text{SiO}_2)$ 、 $w(\text{Al}_2\text{O}_3)$ 、 $w(\text{TiO}_2)$ 、 $w(\text{S})$ 、 $w(\text{P})$ 、 $w(\text{K}_2\text{O})$ 、 $w(\text{Na}_2\text{O})$ 、 $w(\text{Zn})$ 、 $w(\text{MnO})$ 、 $w(\text{FeO})$ (%)、铝硅比 [$w(\text{Al}_2\text{O}_3) / w(\text{SiO}_2)$]、碱度 (R)、目标碱度 (R)
原料参数	混合料水分、混合料 > 8 mm、(5~8) mm、(3~5) mm、≤ 3 mm 粒级占比
烧结相关参数	烧结时间 (min)、废气温度 (°C)、烧结速度 (mm/min)、转鼓指数 (%)、成品率 (%)、烧损 (%)

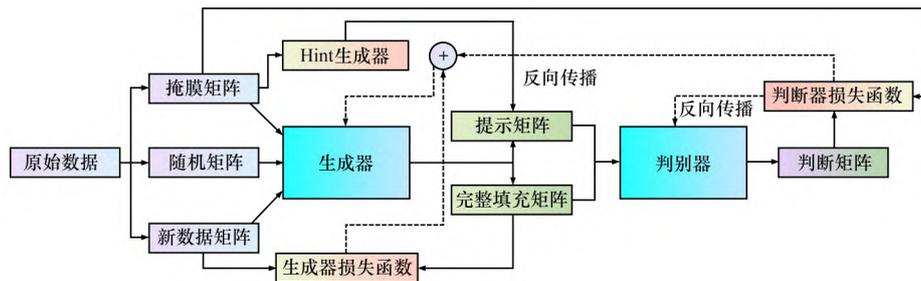


图 2 Gain 原理示意

Fig. 2 Gain principle schematic

该算法的核心思想在于: 生成器根据真实数据对缺失数据进行预测与填充; 判别器在一个完整的数据内判断哪些数据是填补数据, 并给予生成器反馈, 如此反复, 使得生成器与判别器不断优化, 得到最优解^[30-31]。Gain 训练参数如表 2 所

示。经训练后决定系数 (R^2) 可达到 0.971 2, 缺失数据的预测精度较高。

1.2 相关性分析

根据实际烧结生产数据, 选择适当的特征参数可以增强模型预测的准确性和可解释性。在预

表 2 Gain 模型训练参数

Table 2 Gain model training parameter

参数符号	参数名称	参数设定值	参数含义
miss_rate	缺失数据的概率	0.02	表示数据中缺失数值的可能性
batch_size	每个小批量训练样本的数量	128.00	影响训练过程中的批量大小
hint_rate	提示概率	0.90	表示在填充缺失数据时给出提示的可能性
alpha	超参数	100.00	用于调整算法的学习速率或者正则化项
iterations	训练迭代次数	10 000.00	指定训练过程中的迭代次数

测烧结矿粒度方面,研究数据涵盖了 27 个不同维度,包括原料成分占比、混合料的粒度及水分含量和烧结过程控制参数。为了提升模型的性能、减少计算的复杂性,需要挑选最相关且预测能力最强的特性^[32-33]。本研究结合了 MIC 和 Pearson 的相关性分析来确定预测模型的最终特征变量集。为了更便于绘图,特征变量用 $C_1 \sim C_{27}$ 表示。

最大信息系数(maximal information coefficient, MIC)是一种用于探测变量间非线性关联的新方法^[34]。MIC 取值范围在 0 到 1 之间,数值越高代表相关性越明显^[35]。通过对原始数据与 16 ~ 40 mm 粒度烧结矿的 MIC 相关性进行分析,结果如图 3 所示。以 MIC 指标 0.3 为参考,在所有相关特征变量中,筛选后的输入特征量可选用下述 13 种: $w(\text{TFe})$ 、 $w(\text{Al}_2\text{O}_3)$ 、 $w(\text{TiO}_2)$ 、 $w(\text{Na}_2\text{O})$ 、 $w(\text{MnO})$ (%)、目标碱度(R)、混合料(3~5] mm 粒度占比、混合料 ≤ 3 mm 粒度占比、烧结时间(min)、废气温度($^{\circ}\text{C}$)、烧结速度(mm/min)、成品率(%)、烧损(%)。

Pearson 相关性分析:采用统计手段来衡量两个变量之间的线性关系强度。其结果通常在 -1 到 1 之间。当相关系数达到 1 时,表示完全正相关;而 -1 则表示完全负相关;0 则意味着不存在线性联系。采用 Pearson 做相关性分析可以减少不相关或冗余的特征,27 组参数的 Pearson 相关性分析结果如图 4 所示。由图 4 可知,选择相关系数 $|r| > 0.3$ 的靠前的 17 个特征,具体包括 $w(\text{TFe})$ 、 $w(\text{CaO})$ 、 $w(\text{SiO}_2)$ 、 $w(\text{Al}_2\text{O}_3)$ 、 $w(\text{TiO}_2)$ 、 $w(\text{K}_2\text{O})$ 、 $w(\text{Na}_2\text{O})$ 、 $w(\text{Zn})$ 、 $w(\text{MnO})$ 、混合料(3~5] mm

粒度占比、混合料 ≤ 3 mm 粒度占比、烧结时间、废气温度、烧结速度、转鼓指数、成品率、烧损。

选取 Pearson 相关性和 MIC 相关性的共同特征,将样本数据集转化为 12 个维度,分别为 $w(\text{TFe})$ 、 $w(\text{Al}_2\text{O}_3)$ 、 $w(\text{TiO}_2)$ 、 $w(\text{Na}_2\text{O})$ 、 $w(\text{MnO})$ 、混合料(3~5] mm 粒度占比、混合料 ≤ 3 mm 粒度占比、烧结时间、废气温度、烧结速度、成品率、烧损。

1.3 数据增强与异常点替换

现有 191 条数据难以支撑目标模型的训练任务,还需要进行数据增强。生成对抗网络(GAN)在图像、文本等数据增强上均有显著效果^[36-37]。本文选定 GAN 对数据进行增强处理,以提高数据可用量。以烧结原料的 $w(\text{TFe})$ 、 $w(\text{Al}_2\text{O}_3)$ 、 $w(\text{TiO}_2)$ 、 $w(\text{MnO})$ 为例,图 5 展示了数据增强前后的对比情况。由图 5 可以观察到,增强后数据点的分布均匀,模型的泛化能力增强。

此外,比较原始数据和生成数据的分布,以 TFe 质量分数的数据增强为例,如图 6 所示。由图 6 可以看出,TFe 质量分数的生成和原始数据分布规律大致相同,且处于烧结生产正常数值范畴。

箱型图用于展示数据分散情况,反映原始数据的分布特征,便于检测异常值^[38]。它由分别代表 25%、50% 和 75% 四分位数(Q_1)、中位数(Q_{MD})以及上四分位数(Q_3)组成^[39]。四分位数距离(Q_{IQR})表示了 Q_3 和 Q_1 之间的距离。对于本研究的烧结数据集,将小于 $Q_1 - 3Q_{\text{IQR}}$ 或大于 $Q_3 + 3Q_{\text{IQR}}$ 的数据点定义为异常点^[40]。通过此定义筛选出经数据增强处理后烧结数据中的异常点。图 7

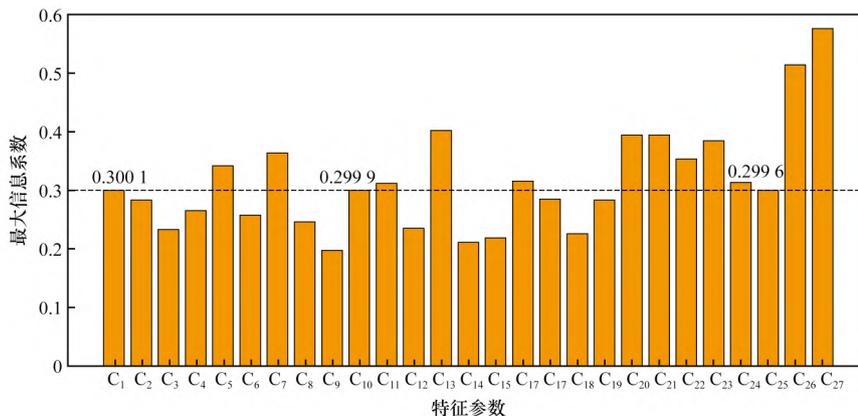


图 3 MIC 相关性分析

Fig. 3 MIC correlation analysis

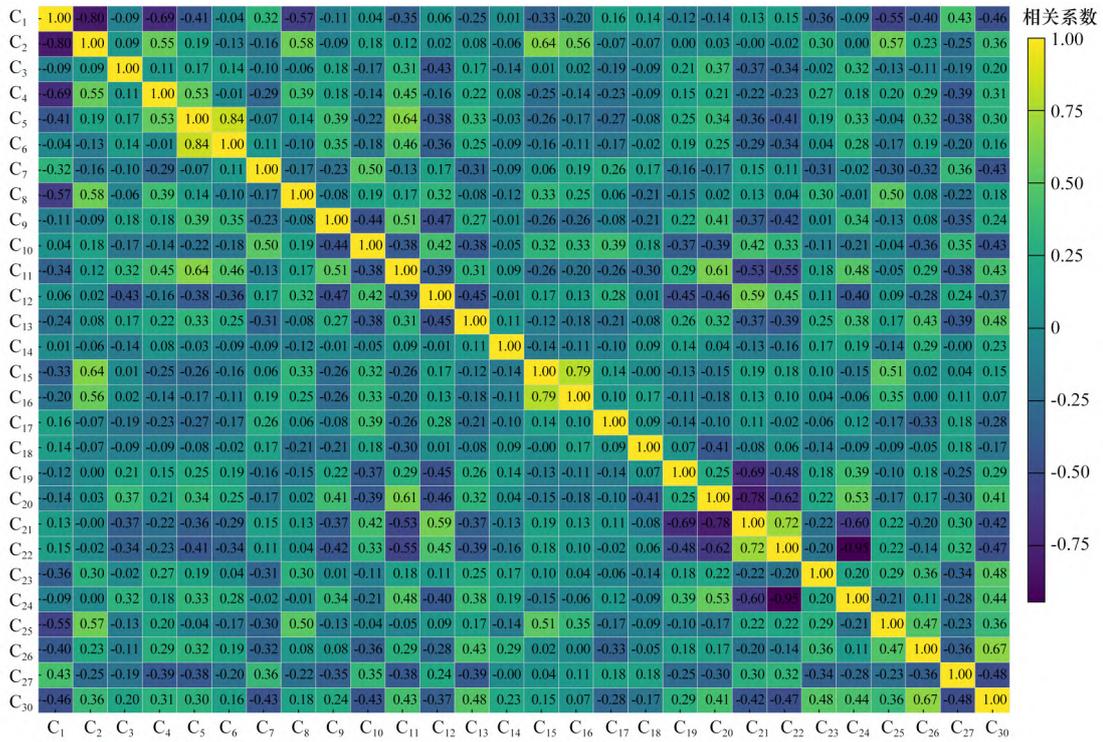
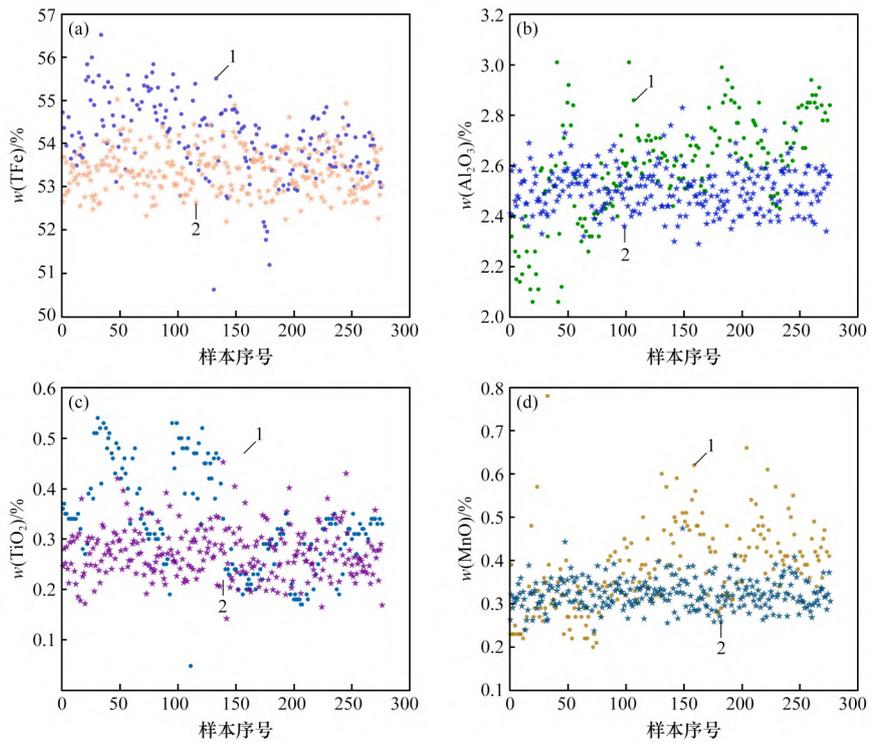


图4 Pearson 相关性分析

Fig.4 Pearson correlation analysis

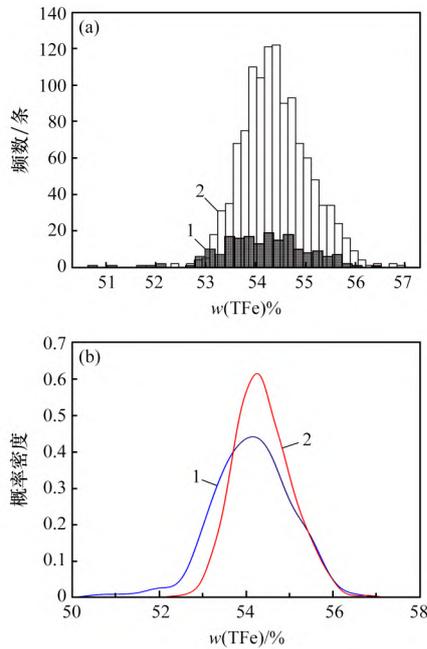


(a) $w(\text{TFe})\%$; (b) $w(\text{Al}_2\text{O}_3)\%$; (c) $w(\text{TiO}_2)\%$; (d) $w(\text{MnO})\%$

1—原数据; 2—增强数据。

图5 数据增强前后示意

Fig.5 Schematic before and after data enhancement



(a) 直方图; (b) 概率密度图
1—真实数据; 2—生成数据。
图 6 数据增强前后数据分布

Fig. 6 Data distribution before and after data enhancement

展示了异常值检测结果,空心圆圈标注的点代表数据中的异常点。在经过数据增强处理的数据集中,利用多项式回归替换异常数据,获得满足模型训练标准的烧结矿粒度数据集。在数据预处理后,总计获得 500 条数据,并按 7:1:2 的比例将数据集划分为训练集、测试集和验证集。

2 OOA-CatBoost 模型构建

2.1 CatBoost 算法

本文采用 CatBoost 模型作为基础模型,用于烧结矿粒度的预测。CatBoost 是 2018 年由 PROKHORENKOVA 等^[41]提出的一种新型梯度提升树(GBDT)算法。其工作步骤如下。

首先,使用遗忘树作为 CatBoost 预测变量,并对所有数值特征进行二值化处理。然后,对模型训练集 $D = \{x_i, y_i\}$ 进行随机排序,结果表示为 σ 。接下来,通过不包含 x_i 的样本集训练多个决策树模型 $m_i (i = 1, 2, 3, \dots, n)$ 。这种渐进式训练策略使模型能够从不断增长的训练数据中学习。每轮训练使用残差 (R_i) 估算公式:

$$R_i = y_i - m_{\sigma(n)} \quad (1)$$

式中: $n = i - 1$ 。

模型利用残差作为梯度估计,能够有效减少目标泄漏,解决预测偏移问题。通过不断调整基学习器,进而生成最终模型。

2.2 OOA 算法

为了优化训练参数,提升 CatBoost 模型的收敛效果,引入鱼鹰优化算法(OOA)进行超参数寻优。OOA 是一种模拟鱼鹰捕猎行为的智能群体算法,由 DEGHANI 等^[42]提出。该算法首先在 CatBoost 的参数范围内定义鱼鹰种群,然后进入全局勘探阶段,通过鱼鹰识别目标位置进行搜索。当鱼鹰识别到目标时,其位置会显著变化,从而增强在最优区域的搜索能力。在这个阶段,鱼鹰的位置更新由以下公式决定:

$$a_{ij}^{\text{new}} = a_{ij} + e_{ij} \times (F_{ij} - I_{ij} \times a_{ij}) \quad (2)$$

式中: a_{ij}^{new} 为第 i 只鱼鹰在第 j 个维度的新位置; e_{ij} 是在 $[0, 1]$ 范围内的随机数; F_{ij} 是第 i 只鱼鹰第 j 个维度中选择的鱼; I_{ij} 是在 $[1, 2]$ 范围内的随机数; a_{ij} 是第 i 只鱼鹰在第 j 个维度上的当前位置。

如果更新后的位置超出边界,则根据以下公式进行修正:

$$a_{ij}^{\text{new}} = \begin{cases} \mu_j & a_{ij}^{\text{new}} > \mu_j \\ l_j & a_{ij}^{\text{new}} < l_j \\ a_{ij}^{\text{new}} & l_j \leq a_{ij}^{\text{new}} \leq \mu_j \end{cases} \quad (3)$$

式中: μ_j 代表上限; l_j 代表下限。

随后,算法进入局部开采阶段,鱼鹰在当前位置进行小范围移动,位置更新如下。

$$a_{ij}^{\text{new}} = a_{ij}^{\text{new}} + \frac{l_j + e_{ij} \times (\mu_j - l_j)}{t}$$

$$t = 1, 2, \dots, T = \begin{cases} a_{ij}^{\text{new}} & l_j \leq a_{ij}^{\text{new}} \leq \mu_j \\ l_j & a_{ij}^{\text{new}} < l_j \\ \mu_j & a_{ij}^{\text{new}} > \mu_j \end{cases} \quad (4)$$

式中: a_{ij}^{new} 为第二阶段中第 i 只鱼鹰在第 j 维度中的新位置; t 为迭代次数; T 为最大迭代次数。

每次移动后,鱼鹰会比较前后位置的适应度,优选适应度更好的位置进行更新。更新公式如下。

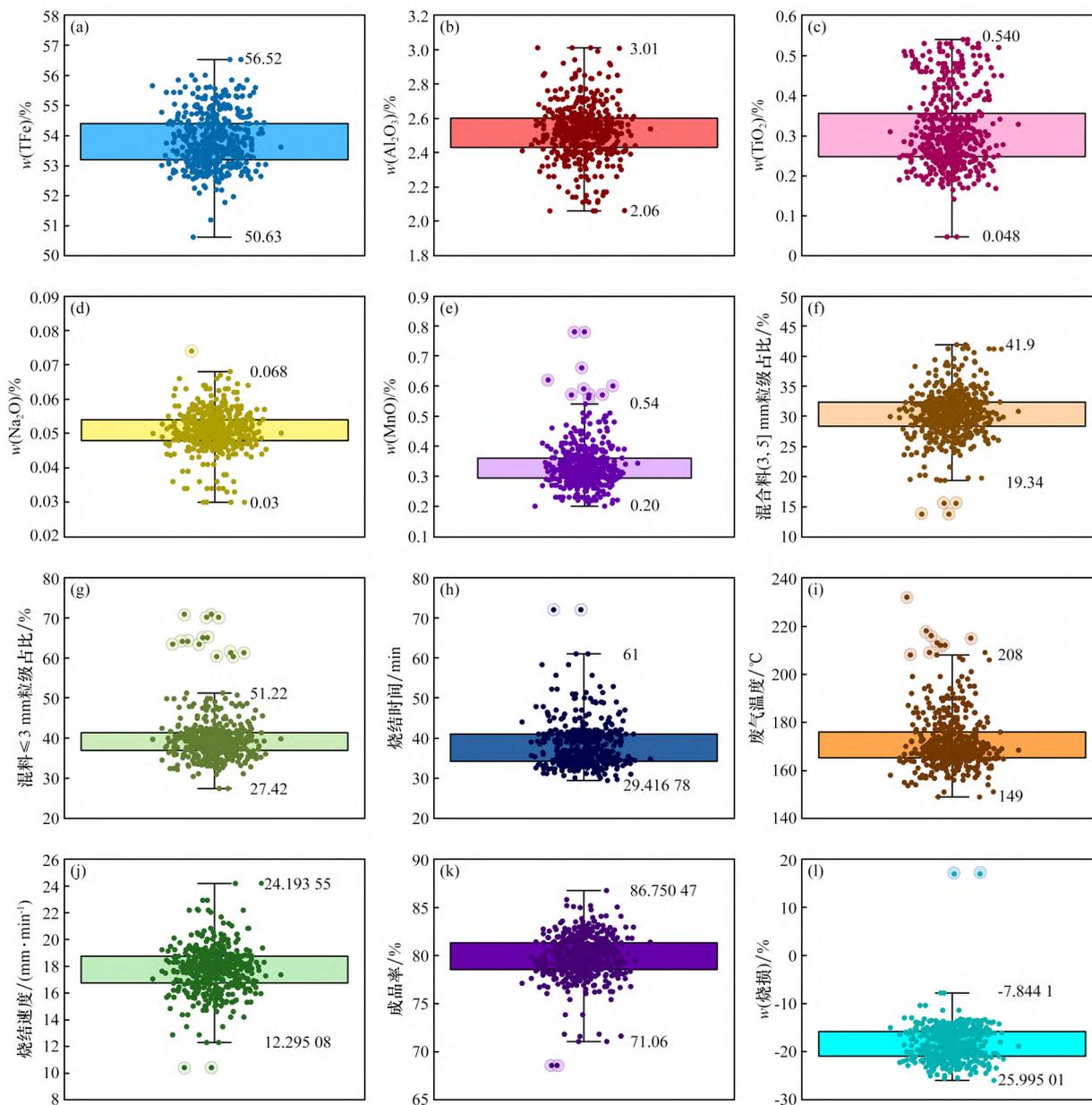
$$A_i = \begin{cases} a_i^{\text{new}} & F_i^{\text{new}} < F_i \\ A_i & \text{其他情况} \end{cases} \quad (5)$$

式中: F_i 为鱼鹰更新前的适应度值; F_i^{new} 为鱼鹰位置更新后的适应度值; A_i 为更新后的位置。

2.3 模型参数选定与构建

OOA-CatBoost 方法将 OOA 算法与 CatBoost 模型相结合, 实现对烧结矿粒级占比的精准预测。首先, 利用 CatBoost 算法处理烧结矿的复杂非线性关系(如预测成分、混合料粒度及烧结参数等),

并处理过程中可能出现的类别特征与非线性问题; 然后, 采用鱼鹰优化算法(OOA)对 CatBoost 模型的超参数进行优化, 通过模拟鱼鹰捕食行为, 自动搜索最优参数组合, 提升模型的收敛速度和全局优化能力, 最终提高模型的预测精度。该过程综合考虑了 CatBoost 的类别特征处理能力与 OOA 的全局优化能力, 极大提高了烧结矿粒度预测的



(a) $w(\text{TFe})$; (b) $w(\text{Al}_2\text{O}_3)$; (c) $w(\text{TiO}_2)$; (d) $w(\text{Na}_2\text{O})$;
 (e) $w(\text{MnO})$; (f) 混合料(3, 5] mm 粒级占比; (g) 混料 ≤ 3 mm 粒级占比; (h) 烧结时间;
 (i) 废气温度; (j) 烧结速度; (k) 成品率; (l) 烧损

图7 特征数据箱型图

Fig. 7 Characteristic data box plots

准确性和泛化能力。具体步骤如下。

(1) 数据收集。收集烧结矿生产过程中的监测数据,包括烧结原料成分、混合料粒度组成及烧结工艺参数,对收集到的数据进行预处理,确保数据的准确性,并将数据分为训练集和测试集。

(2) CatBoost 模型构建。使用 CatBoost 算法处理数据,通过随机排序训练集并使用逐步学习策略,以减少模型偏差并提高泛化能力。此步骤中,CatBoost 能够有效捕捉数据中的非线性关系,对复杂工艺条件下的烧结矿粒度组成进行初步预测。

(3) OOA 参数优化。使用 OOA 算法对 CatBoost 模型的关键超参数(如迭代次数、树的深度、学习率等)进行优化。首先,确定需要优化的 CatBoost 参数范围;然后,设定适应度函数,以模型的均方根误差(RMSE)作为适应度函数,随机初始化鱼鹰种群。通过模拟鱼鹰的捕食过程,不断迭代,更新每个个体的参数组合,直到达到设定的迭代次数或满足终止条件,最终得到全局最优的 CatBoost 参数组合。

(4) 模型训练。将 OOA 算法得到的最优超参数输入到 CatBoost 模型中,输入训练集数据,训练得到最佳的模型参数,提升模型的泛化能力和预测精度。

(5) 预测分析。将测试集数据输入至训练好的 CatBoost 模型,完成烧结矿粒度预测。通过对预测结果的验证与分析,评价模型的预测效果,并为生产过程中的反向调控提供数据支持。模型流程如图 8 所示。

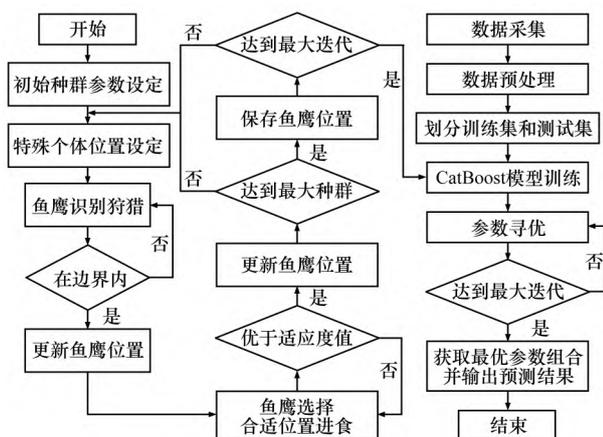


图 8 OOA-CatBoost 模型流程

Fig. 8 Flowchart of the OOA-CatBoost model

3 结果和讨论

3.1 模型对比

为了验证 OOA-CatBoost 模型的优越性,控制智能群体算法的种群数量与最大迭代次数一定,选取 DMO、SSA 分别与 RF、XGBoost 以及 LightGBM 模型随机组合,作为 OOA-CatBoost 模型的对照组。本文选取三种常用的评价指标来对所有模型进行定量评估和比较,以综合评价各模型的预测效果,分别为平均绝对误差(mean absolute error, MAE)、均方误差(mean square error, MSE)和决定系数(R^2),如式(6)、(7)以及式(8)所示。

$$\sigma_{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y_p| \quad (6)$$

$$\sigma_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - y_p)^2 \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_p)^2}{\sum_{i=1}^n (y_i - y_a)^2} \quad (8)$$

式中: σ_{MAE} 为平均绝对误差; σ_{MSE} 为均方误差; n 为样本个数; y_i 为真实值; y_p 为预测值; y_a 为真实值的平均值。

对于决定系数,其越靠近 1,模型准确度越高。其余两项指标,取值越小,模型准确度越高。在经过相同次数迭代实验后,得到的结果如表 3 所示。通过不同的智能群体算法与机器学习模型

表 3 不同模型预测结果

Table 3 Predicted results from different models

模型名称	种群数量	迭代次数	σ_{MAE}	σ_{MSE}	R^2
DMO-RF	10	40	0.698 4	0.056 6	0.746 0
DMO-XGBoost	10	40	0.718 9	0.053 7	0.771 5
DMO-CatBoost	10	40	0.664 3	0.051 3	0.791 2
DMO-LightGBM	10	40	0.841 3	1.187 5	0.699 8
SSA-RF	10	40	0.427 3	0.063 0	0.893 9
SSA-XGBoost	10	40	0.341 8	0.061 1	0.900 3
SSA-CatBoost	10	40	0.297 8	0.046 3	0.942 8
SSA-LightGBM	10	40	0.329 1	0.055 5	0.917 7
OOA-RF	10	40	0.419 6	0.062 9	0.894 3
OOA-XGBoost	10	40	0.353 4	0.057 8	0.910 7
OOA-CatBoost	10	40	0.276 9	0.043 3	0.949 9
OOA-LightGBM	10	40	0.368 3	0.055 1	0.918 9

随机组合进行,对比分析可得,OOA-CatBoost模型相对DMO、SSA以及OOA组合的其他预测模型,在MAE以及MSE上都达到了最低,分别为0.276 9和0.043 3,同样 R^2 也达到了最高值0.949 9。在 R^2 上,OOA-CatBoost算法相对DMO以及SSA优化的CatBoost模型分别提高了15.87%和0.71%。

综上所述,OOA-CatBoost在烧结矿粒度预测上,预测误差达到最小,预测精度达到最高。基于OOA-CatBoost模型,在不同的粒度范围下进行OOA-CatBoost模型的模拟仿真试验,得到烧结矿不同粒度组成的预测仿真效果如图9所示。由图9可见,尽管验证集数值差异较大,但曲线的拟合程度仍然较高。本文选定100条试验数据作为验证集,得到模型对 ≤ 5 mm、(5,10] mm、(10,16] mm、(16,25] mm、(25,40] mm、 >40 mm 粒度烧结矿的预测误差分别为1.091%、1.054%、0.996%、2.332%、1.300%和2.159%;而对于全部粒度范围预测值的平均最大误差为1.489%。

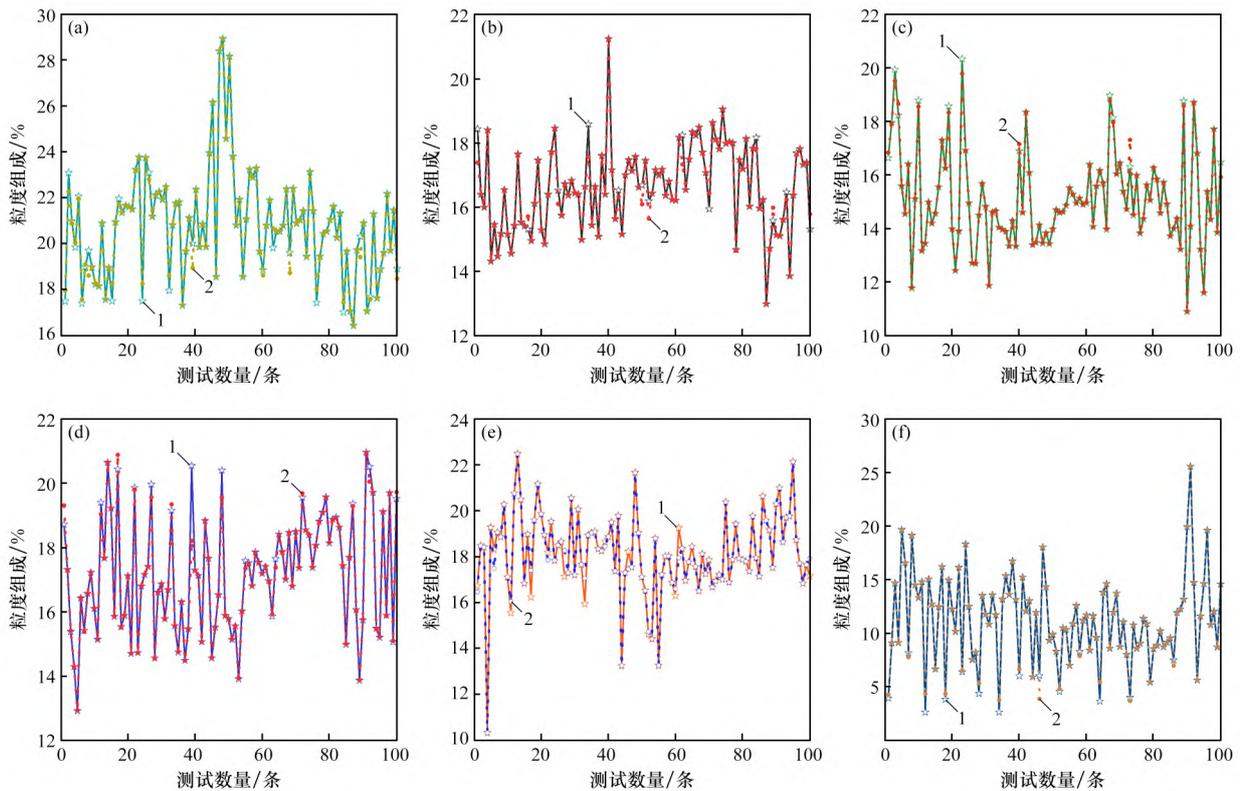
3.2 工业验证

将基于OOA-CatBoost的烧结矿粒度预测模型应用于实际生产中,得到结果如图10所示。由图10可见,实际值与预测值最大误差为1.384%,这表明该模型在工业生产中具有一定的实用价值。模型基于烧结原料成分、混合料参数以及烧结过程等参数预测烧结矿粒度情况,可为烧结生产调节原料配比、优化烧结过程参数提供技术指导。

4 结论

(1) 本文提出的烧结生产数据预处理方法,在为模型提供高质量且可靠数据的同时,也为钢铁工业的大数据预处理提供了策略。基于烧结生产实测数据结合智能群体算法以及机器学习的迭代计算方法建立的烧结矿粒度预测模型,可预测不同条件下烧结矿的粒度组成情况,为实际生产提供指导。

(2) 将OOA算法与CatBoost模型相结合,对



粒度范围/mm: (a) ≤ 5 ; (b) (5,10]; (c) (10,16]; (d) (16,25]; (e) (25,40]; (f) >40

1—实际值; 2—OOA-CatBoost 预测值。

图9 不同粒度组成预测仿真效果

Fig. 9 Simulation effect of the prediction of the percentage of different grain sizes

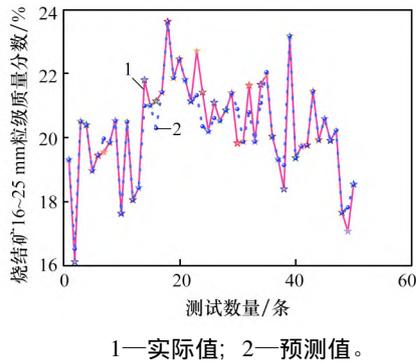


图 10 烧结矿 16 ~ 25 mm 粒级质量分数预测结果

Fig. 10 Predicted share of sinter ore with grain size 16 ~ 25 mm

比 DMO、SSA 算法优化的 CatBoost 模型，模型的预测精度分别提高 15.87%、0.71%。通过仿真试验，模型得到烧结矿粒度预测实际值与预测值的最大平均误差为 1.49%，为预测提供有效反馈。模型在工业生产中预测最大误差为 1.38%，具有一定的推广应用价值。

参考文献:

- [1] 石泉,唐珏,储满生. 基于工业大数据的智能化高炉炼铁技术研究进展[J]. 钢铁研究学报, 2022, 34(12): 1314-1324.
- SHI Quan, TANG Jue, CHU Mansheng. Research progress of intelligent blast furnace ironmaking technology based on industrial big data[J]. Journal of Iron and Steel Research, 2022, 34(12): 1314-1324.
- [2] FENG S L, WANG B, ZHOU Z X, et al. Research on multi-decision sinter composition optimization based on OLS algorithm[J]. Metals, 2023, 13(3): 548.
- [3] 刘欢. 高硅铁精矿粉高碱度烧结实验研究[D]. 西安: 西安建筑科技大学, 2017.
- LIU Huan. Experimental study on high basicity sintering of high silicon iron ore[D]. Xi'an: Xi'an University of Architecture and Technology, 2017.
- [4] 张朝晖,徐凯强,邢相栋,等. 烧结矿低温还原粉化影响因素研究进展[J]. 钢铁研究学报, 2021, 33(6): 453-460.
- ZHANG Zhaohui, XU Kaiqiang, XING Xiangdong, et al. Research progress of influence factors on low-temperature reduction degradation[J]. Journal of Iron and Steel Research, 2021, 33(6): 453-460.
- [5] 张国鹏,沈峰满,高强健,等. 降低 MgO 含量对高 Al₂O₃ 烧结矿冶金性能的影响[J]. 钢铁研究学报, 2022, 34(8): 749-757.
- ZHANG Guopeng, SHEN Fengman, GAO Qiangjian, et al. Effect of reducing MgO content on metallurgical properties of high Al₂O₃ sinter[J]. Journal of Iron and Steel Research, 2022, 34(8): 749-757.
- [6] LI Z X, LI T L, SUN C Y, et al. The effects of MgO and Al₂O₃ content in sinter on the softening-melting properties of mixed ferrous burden[J]. Materials, 2023, 16(15): 5490.
- [7] 单长冬,张建良,王耀祖,等. 富氧条件下高精粉配比对烧结指标及冶金性能的影响[J]. 钢铁, 2023, 58(1): 22-30.
- SHAN Changdong, ZHANG Jianliang, WANG Yaozu, et al. Effect of high magnetite concentrates ratio on sintering index and metallurgical properties under oxygen-enriched conditions[J]. Iron & Steel, 2023, 58(1): 22-30.
- [8] 许满兴. 烧结矿质量及其对高炉冶炼主要操作指标的影响[C]//2017 第五届炼铁对标、节能降本及新技术研讨会论文集. 扬州: 河北省金属协会, 2017: 10.
- XU Manxing. Sinter quality and its influence on main operation indexes of blast furnace smelting[C]//2017 5th Ironmaking Benchmarking, Energy Saving and Cost Reduction and New Technology Seminar Papers. Yangzhou: Hebei Metal Society, 2017: 10.
- [9] 张志刚,沈小峰. 梅钢提高五号机烧结矿平均粒径的生产实践[J]. 甘肃冶金, 2018, 40(3): 8-10.
- ZHANG Zhigang, SHEN Xiaofeng. Practice of improving the average particle size of sinter in No. Five machine by meisteel[J]. Gansu Metallurgy, 2018, 40(3): 8-10.
- [10] HOLLIDAY M, LAI Y F, HOBBS M, et al. Continuous measurement of ferrous sinter size distributions using an optical sensor system[J]. Ironmaking & Steelmaking, 2023, 50(8): 1104-1111.
- [11] 储满生,王茗玉,唐珏,等. 基于大数据的智能化烧结技术研究进展[J]. 钢铁, 2023, 58(9): 26-38.
- CHU Mansheng, WANG Mingyu, TANG Jue, et al. Process of intelligent sintering technology based on big data[J]. Iron & Steel, 2023, 58(9): 26-38.
- [12] WANG S H, LI H F, ZHANG Y J, et al. A hybrid ensemble model based on ELM and improved AdaBoost. RT algorithm for predicting the iron ore sintering characters[J]. Computational Intelligence and Neuroscience, 2019,

- 2019: 4164296.
- [13] KUMAR V ,PATIL D ,MRUNMAYA K P ,et al. Impact of sinter basicity and alumina on softening melting behavior in blast furnace [J]. Transactions of the Indian Institute of Metals ,2020 ,73(2) :365 – 375.
- [14] LI Y F ,ZHANG Q W ,ZHU Y ,et al. A model study on raw material chemical composition to predict sinter quality based on GA-RNN[J]. Computational Intelligence and Neuroscience ,2022(1) :3343427.
- [15] 李泽政,刘卫星,李飞,等. 基于数据增强的烧结矿转鼓强度预测研究[J]. 烧结球团 ,2023 ,48(6) :62 – 68.
LI Zezheng ,LIU Weixing ,LI Fei ,et al. Research on prediction of drum strength based on data enhancement sinter [J]. Sintering and Pelletizing ,2023 ,48(6) :62 – 68.
- [16] REN X Y ,YANG B ,LUO N ,et al. The prediction of sinter drums strength using hybrid machine learning algorithms [J]. Computational Intelligence and Neuroscience , 2022(1) :4790736.
- [17] 刘月明,刘小杰,吕庆,等. 基于烧结大数据预测小于 10 mm 烧结矿含量模型 [J]. 中国冶金,2019 , 29(11) :31 – 38.
LIU Yueming ,LIU Xiaojie ,LÜ Qing ,et al. Prediction model of sinter content less than 10mm based on sintering big data [J]. China Metallurgy ,2019 ,29(11) : 31 – 38.
- [18] 王诗歌. 基于图像识别的烧结矿粒度温度分布研究 [D]. 马鞍山: 安徽工业大学 ,2021.
WANG Shige. Research on particle size temperature distribution of sinter based on image recognition [D]. Ma' anshan: Anhui University of Technology ,2021.
- [19] 王丽敬,胡支滨,韩阳,等. 高炉利用系数提升的鼓风制度自适应调控模型研究 [J]. 冶金自动化 ,2023 , 47(2) :57 – 65; 88.
WANG Lijing ,HU Zhibin ,HAN Yang ,et al. Research on adaptive control model of blasting system for improvement of blast furnace utilization coefficient [J]. Metallurgical Industry Automation ,2023 ,47(2) :57 – 65; 88.
- [20] XU Z H ,WANG Z J ,QI X W ,et al. Prediction of green properties of flux pellets based on improved generalized regression neural network [J]. Metals ,2022 ,12(11) : 1840.
- [21] DAI F L ,FAN X H ,HUANG X X ,et al. Fuel-appropriate distribution of the material layer based on numerical model of sintering with particle swarm optimization algorithm [J]. Minerals ,2023 ,13(4) :511.
- [22] 伍铁斌,朱红求,龙文,等. 改进的鲸鱼优化算法及其在烧结配料中的应用 [J]. 中南大学学报(自然科学版) ,2020 ,51(1) :103 – 111.
WU Tiebin ,ZHU Hongqiu ,LONG Wen ,et al. Improved whale optimization algorithm and its application in sintering blending process [J]. Journal of Central South University (Science and Technology) ,2020 ,51(1) : 103 – 111.
- [23] 杜胜. 烧结过程多时间尺度工况识别与热状态参数智能控制 [D]. 武汉: 中国地质大学 ,2022.
DU Sheng. Operating mode recognition with multiple time scales and intelligent control of thermal state parameters for sintering process [D]. Wuhan: China University of Geosciences ,2022.
- [24] 李一帆,李锦祥,杨锦堂,等. 基于 BO-RF 的烧结矿化学成分预测模型研究 [J]. 烧结球团 ,2023 ,48(6) : 109 – 115; 138.
LI Yifan ,LI Jinxiang ,YANG Jintang ,et al. Research on chemical composition prediction model of sinter based on BO-RF [J]. Sintering and Pelletizing ,2023 ,48(6) : 109 – 115; 138.
- [25] 李星彤,龚伟,李帝阅. 基于 Boosting 算法的转炉终点预测模型 [J/OL]. 材料与冶金学报 ,2025: 1 – 17 [2025 – 03 – 24]. <http://kns.cnki.net/kcms/detail/21.1473.TF.20240407.1828.002.html>.
LI Xingtong ,GONG Wei ,LI Diyue. End point prediction model of converter based on Boosting algorithm [J]. Journal of Materials and Metallurgy ,2025: 1 – 17 [2025 – 03 – 24]. <http://kns.cnki.net/kcms/detail/21.1473.TF.20240407.1828.002.html>.
- [26] ZHAO Y X ,UDELL M. Missing value imputation for mixed data via Gaussian copula [C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM , 2020: 636 – 646.
- [27] 赵志文,高敏. 缺失数据下随机系数自回归模型的参数估计 [J]. 统计与决策 ,2022 ,38(1) :16 – 20.
ZHAO Zhiwen ,GAO Min. Parameter estimation of random coefficient autoregressive model with missing data [J]. Statistics & Decision ,2022 ,38(1) :16 – 20.
- [28] 郭艳,宋晓祥,袁昊,等. 物联网缺失数据在线填充方法 [J]. 陆军工程大学学报 ,2023 ,2(6) :8 – 16.
GUO Yan ,SONG Xiaoxiang ,YUAN Hao ,et al. Missing

- values online imputation for Internet of Things data [J]. Journal of Army Engineering University of PLA, 2023, 2(6): 8–16.
- [29] YOON J, JORDON J, SCHAAR M. Gain: Missing data imputation using generative adversarial nets [Z/OL]. arXiv, (2018–06–07). <http://arxiv.org/abs/1806.02920>.
- [30] 李翔, 高辉, 陈良亮. 基于 GAIN 的数据插补及 Bi-GRU 在充电桩预警中的应用 [J]. 广东电力, 2022, 35(12): 22–31.
LI Xiang, GAO Hui, CHEN Liangliang. Data imputation based on GAIN and application of Bi-GRU in charging pile early warning [J]. Guangdong Electric Power, 2022, 35(12): 22–31.
- [31] 周备, 张莹, 张生瑞, 等. 基于生成对抗网络的追尾事故数据填补方法研究 [J]. 交通运输系统工程与信息, 2024, 24(1): 132–137; 198.
ZHOU Bei, ZHANG Ying, ZHANG Shengrui, et al. Rear-end crash data imputation methods using generative adversarial networks [J]. Journal of Transportation Systems Engineering and Information Technology, 2024, 24(1): 132–137; 198.
- [32] KUZUDISLI C, BAKIR-GUNGOR B, BULUT N, et al. Review of feature selection approaches based on grouping of features [J]. PeerJ, 2023, 11: e15666.
- [33] 王翔, 胡学钢. 高维小样本分类问题中特征选择研究综述 [J]. 计算机应用, 2017, 37(9): 2433–2438; 2448.
WANG Xiang, HU Xuegang. Overview on feature selection in high-dimensional and small-sample-size classification [J]. Journal of Computer Applications, 2017, 37(9): 2433–2438; 2448.
- [34] 方军, 王兴东, 汪洋, 等. 基于 MIC 和 IPSO-RELM 的带钢热镀锌板锌层厚度预测 [J]. 热加工工艺, 2024, 53(22): 62–68; 82.
FANG Jun, WANG Xingdong, WANG Yang, et al. Thickness prediction of zinc layer of hot-dip galvanized sheet in strip based on mic and ipso-rerm [J]. Hot Working Technology, 2024, 53(22): 62–68; 82.
- [35] 江德文, 王振阳, 戴建华, 等. 基于支持向量机的高炉煤气利用率预测建模 [J]. 中国冶金, 2021, 31(4): 55–63; 67.
JIANG Dewen, WANG Zhenyang, DAI Jianhua, et al. Forecast modeling of gas utilization rate of blast furnace based on support vector machine [J]. China Metallurgy, 2021, 31(4): 55–63; 67.
- [36] 耿延兵, 周国宾, 靳杰, 等. 基于数据增强的可分离卷积轴承寿命预测方法研究 [J]. 控制工程, 2024, 31(3): 503–509.
GENG Yanbing, ZHOU Guobin, JIN Jie, et al. Research on the method of life prediction of separable convolutional bearings based on data enhancement [J]. Control Engineering of China, 2024, 31(3): 503–509.
- [37] 徐金鹏, 郭新峰, 王瑞波, 等. 基于 GAN 数据增强的软件缺陷预测聚合模型 [J]. 计算机科学, 2023, 50(12): 24–31.
XU Jinpeng, GUO Xinfeng, WANG Ruiibo, et al. Aggregation model for software defect prediction based on data enhancement by GAN [J]. Computer Science, 2023, 50(12): 24–31.
- [38] 陈晓, 刘长华, 刘志亮, 等. 2010—2019 年北黄海海域长序列海量温盐数据分析与处理方法 [J]. 海洋与湖沼, 2022, 53(1): 49–61.
CHEN Xiao, LIU Changhua, LIU Zhiliang, et al. Analysis and processing of long sequence and massive temperature and salinity data of the north Yellow Sea from 2010 to 2019 [J]. Oceanologia et Limnologia Sinica, 2022, 53(1): 49–61.
- [39] 刘馨, 张卫军, 石泉, 等. 基于数据挖掘与清洗的高炉操作参数优化 [J]. 东北大学学报(自然科学版), 2020, 41(8): 1153–1160.
LIU Xin, ZHANG Weijun, SHI Quan, et al. Operation parameters optimization of blast furnaces based on data mining and cleaning [J]. Journal of Northeastern University (Natural Science), 2020, 41(8): 1153–1160.
- [40] LIU S, LIU X J, LYU Q, et al. Comprehensive system based on a DNN and LSTM for predicting sinter composition [J]. Applied Soft Computing, 2020, 95: 106574.
- [41] PROKHORENKOVA L, GUSEV G, VOROBEV A, et al. CatBoost: Unbiased boosting with categorical features [Z/OL]. arXiv, (2017–06–28) [2019–01–20]. <https://arxiv.org/abs/1706.09516>.
- [42] DEGHANI M, TROJOVSKY P. Osprey optimization algorithm: A new bio-inspired metaheuristic algorithm for solving engineering optimization problems [J]. Frontiers in Mechanical Engineering, 2023, 8: 1126450.