

基于决策树算法的焦炭 CSR 和 CRI 性能预测模型

刘晓航¹, 史先菊^{1a,2}, 许德明¹, 贺 铸¹, 李光强¹, 王 强¹

(1. 武汉科技大学 a. 钢铁冶金及资源利用省部共建教育部重点实验室; b. 省部共建耐火材料与冶金国家重点实验室, 湖北 武汉 430081; 2. 宝钢股份中央研究院 武钢有限技术中心, 湖北 武汉 430080)

摘要: 焦炭的 CSR(焦炭反应后强度)和 CRI(焦炭反应性指数)性能指标对提高高炉冶炼效率、降低生产成本和促进绿色生产等起到至关重要的作用。传统依赖人工经验调控存在响应慢、误差大等问题,难以实现对 CSR 和 CRI 的精准实时预测。本文基于焦炭的灰分、挥发分、硫分和固定碳等基础属性,采用大数据拟合与机器学习技术,构建预测 CSR 和 CRI 的决策树算法;通过网格搜索结合交叉验证来优化超参数,筛选出最优的决策树,并利用特征相关热图、特征相关性及 SHAP 值解释模型的预测机制。结果表明:当树深度为 19、随机种子数为 44 时,CSR 模型的预测效果最佳,精度达 98.543%;当树深度为 18、随机种子数为 75 时,CSR 模型的预测效果最佳,精度达 96.825%;现场测试结果显示,封装后在线实时预测软件的单次预测时间仅为 0.1~0.3 s,软件具备良好的实时性与稳定性。本文开发的预测系统能有效支持高炉生产中焦炭的质量决策,显著提升预测效率与准确性,推动炼铁过程向智能化、绿色化和高效化发展。

关键词: 高炉炼铁; 焦炭反应后强度; 焦炭反应性指数; 机器学习; 决策树算法

中图分类号: TF526⁺.1; TP181

文献标志码: A

文章编号: 1000-8764(2025)04-0151-10

doi: 10.13403/j.sjqt.2025.04.069

Coke CSR and CRI performance prediction model based on decision tree algorithm

LIU Xiaohang¹, SHI Xianju^{1a,2}, XU Deming¹, HE Zhu¹, LI Guangqiang¹, WANG Qiang¹

(1. Wuhan University of Science and Technology a. Key Laboratory for Ferrous Metallurgy and Resources Utilization of Ministry of Education; b. The State Key Laboratory of Refractories and Metallurgy, Wuhan 430081, Hubei, China; 2. Technical Center of Wuhan Iron and Steel Co., Ltd., Academia Sinica, Baosteel, Wuhan 430080, Hubei, China)

Abstract: The CSR and CRI performance index of coke play a crucial role in improving blast furnace smelting efficiency, reducing production costs and promoting green production. Traditional reliance on manual experience regulation has problems such as slow response and large errors, making it difficult to achieve accurate real-time prediction for CSR and CRI. Based on the basic properties of coke, such as ash, volatile content, sulfur content and fixed carbon, big data fitting and machine learning technology are used to construct a decision tree algorithm for predicting CSR and CRI. Grid search combined with cross-validation is used to optimize hyperparameters, the optimal decision tree is screened, and the prediction mechanism of the model is explained by using feature correlation heat map, feature correlation and SHAP value. The results show that when the tree depth is 19 and the number of random seeds is 44, the prediction effect of the CSR model is the best, with an accuracy of 98.543%. When the tree depth is 18 and the number of random seeds is 75, the prediction effect of the CSR model is the best, with an accuracy of 96.825%. The field test results show that the single

收稿日期: 2024-06-25; 修回日期: 2024-07-08

基金项目: 国家自然科学基金重点资助项目(U22A20173);湖北省重点研发计划资助项目(2022BAA021&2022BAD043)

作者简介: 刘晓航(1998—),男,博士研究生,从事冶金领域机器学习方面的研究。

通信作者: 王 强(1989—),男,博士,教授,从事智慧冶金和冶金固废资源化方面的研究。

prediction time of the online real-time prediction software after packaging is only 0.1 ~ 0.3 s, and the software has good real-time performance and stability. The prediction system developed in this paper can effectively support the quality decision of coke in blast furnace production, significantly improve the prediction efficiency and accuracy, and promote the development of the ironmaking process to be intelligent, green and efficient.

Key words: blast furnace ironmaking; CSR; CRI; machine learning; decision tree algorithm

在当代高炉生产过程中, 焦煤既是还原剂也是骨架结构的核心材料, 其特性对高炉的运作效率、焦炭使用量、经济成本及环境影响方面有着直接作用^[1-3]。焦炭是在高温环境中将焦煤经干馏工艺转化形成, 其为铁矿还原提供了热量和还原性气体(CO), 而且它的品质还决定了高炉内物料层的透气性和底部工作区的稳定性^[4-5]。其中, 评价焦炭质量的两个关键指标为反应后强度(CSR)和焦炭反应性指数(CRI)^[6]。CSR反映的是焦炭的机械强度, 这对维持高炉料柱的良好气流通道至关重要, 可避免因焦炭碎裂引发的气流分布不均问题^[7]。过高的 CSR 值可以使焦炭抵抗高温下的热冲击和机械应力, 确保良好的通风条件和支撑功能, 从而提升炼铁效率及产品质量; 过低的 CSR 值则会降低焦炭品质, 干扰钢铁生产的稳定性和经济效益。CRI 衡量了焦炭的反应活性^[8], 它直接影响高炉内的间接还原效率及焦炭消耗速率。过高的 CRI 值破坏炉内热量均衡, 从而影响生产效率; 过低的 CRI 值表明焦炭过硬不易参与反应^[9-10]。精确测量这两个参数, 对优化焦炭制造工艺流程和高炉冶炼有着现实意义, 可促进节能减排和资源高效利用, 并推动冶金工业实现低碳发展^[11-12]。

早期对于焦炭 CSR 和 CRI 两大性能指标的研究主要利用工厂生产数据进行数学建模, 以回归分析为主找寻数据之间的联系^[13-15]。在数学推论方面, 么秋香等^[16]以灰分、挥发分、机械强度等为输入变量建立焦炭质量的多元回归预测模型, 实现对配煤和焦炭质量的优化控制; 程榆^[17]也建立了焦炭质量的多元线性回归法预测方程, 得出 CSR 和 CRI 成反比的规律。但回归分析方法存在的计算速度慢、不具备实时性等问题, 而神经网络算法可以很好解决这些问题。苏鑫等^[18]利用 GA-BPNN 算法预测焦炭产量, 预测结果的平均误差为 2.94%, 新模型解决了 BP 神经网络预测结

果不稳定的问题; 阎少宏等^[19]利用级联 BP 神经网络预测焦煤质量, 将误差控制在 15% 以内, 证明了级联 BP 神经网络具有优越的拟合能力与稳定的收敛特性; 李爱莲等^[20]利用量子遗传算法优化 BP 神经网络实现焦炭质量的预测, 该模型相较于传统的 BP 网络模型预测精度更高; 周洋等^[21]利用加入聚类分析的神经网络模型预测高炉焦比指标, 与传统神经网络相比, 新模型的误差降低了 3.13 kg/t; 詹艳艳等^[22]采用 GA 优化径向基函数(RBF)神经网络预测焦炭的抗碎强度、耐磨强度、反应性指数和反应后强度, 新模型具有较强的适应性; 刘剑等^[23]采用混沌菌群算法优化 BP 神经网络过程, 建立延迟焦化过程生焦高度的斜率预测算法, 使算法的预测精度得到了较大的提升。神经网络算法的训练速度慢、缺乏解释性, 而决策树算法不仅训练速度快、解释性强, 而且还有预测效率高、优化目标明确等优点。

为此, 本文基于决策树算法搭建焦炭 CSR 和 CRI 预测模型, 采用某钢铁厂的现场数据, 以焦炭的灰分、挥发分、硫分、水分、固定碳、M40、M25、M10 和焦末含量为输入特征; 通过特征相关热图、特征相关性和 SHAP 值等分析 CSR 和 CRI 与输入特征之间的变化规律, 利用网格搜索和交叉验证方法找寻两种性能预测模型的最佳超参数, 最后通过 Tkinter 和 Cv2 库将机器学习代码进行封装形成在线预测软件, 为工业生产提供指导。

1 机器学习模型

1.1 模型搭建

决策树算法回归模型采用 Python 语言搭建, 其构建流程如图 1 所示。焦炭数据集由某钢铁厂提供, 数据按 9:1 划分为训练集和验证集。采用特征选择方法来降低数据的维度, 通过网格搜索和交叉验证方法确定适应于数据库训练的模型最佳

参数,生成完整的决策树回归算法。使用预剪枝方法建立决策树模型,通过特征相关热图、特征相关性和SHAP值进行数据分析,确定模型适用性。采用决策树算法进行数据预测,与试验数据对比确定预测精度并进行模型评估。当模型预测准确率未达到预期目标时调整模型参数进一步优化,直至达到预期为止。将优化好的训练模型进行线上工业推广。

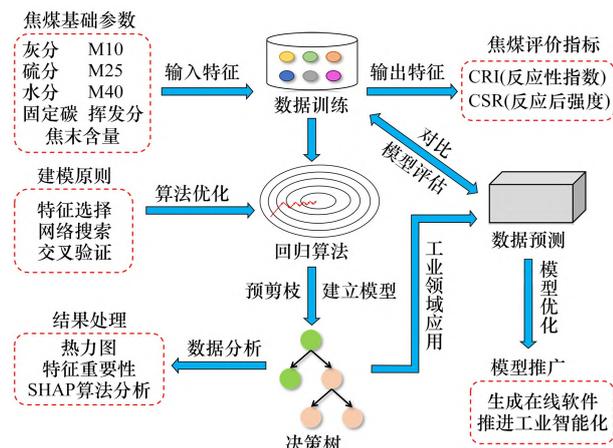


图1 机器学习流程

Fig. 1 Diagram of machine learning

1.2 数据训练

将数据进行划分,确定输入特征为焦炭的基础属性参数,包括焦炭的灰分、挥发分、硫分、水分、固定碳、M40和M25(两者均为焦炭抗碎强度,表示焦炭冷态机械强度指标中抵抗破碎的能力大小)、M10(为焦炭耐磨强度,它的好坏反映了焦炭耐磨能力的大小)和焦末含量,设置输出特征为焦炭的CRI和CSR指标。同时对数据集进行预处理,去除重复值和缺失值,并将数据进行标准化和归一化处理,得到510组数据集。利用Python语言Pandas库中data.describe()函数计算得到连续型数据的均值、最大值、最小值、中位数、四分位数、总数、标准差,如表1所示。

1.3 SHAP分析原理

SHAP(shapley additive explanations)工具包在解释决策树模型的输出中起着至关重要的作用,其将每个特征的影响从预测值中分解出来进行分析。这个过程可以表示如下。

$$y_i = y_{\text{base}} + f(x_{i1}) + f(x_{i2}) + \dots + f(x_{ij}) \quad (1)$$

式中: x_{ij} 表示第*i*个样本的第*j*个特征; y_i 表示

样本的预测值, y_{base} 为决策树模型的基线(所有样本的目标变量的平均值); $f(x_{ij})$ 为当前特征的SHAP值。

表1 数据信息表(质量分数)

项目	均值	标准差	最小值	下四分位数	中位数	上四分位数	最大值
灰分	13.27	0.41	12.10	13.00	13.20	13.50	14.40
挥发分	1.31	0.24	0.80	1.10	1.30	1.50	1.90
硫分	0.76	0.05	0.60	0.72	0.75	0.79	0.92
水	3.22	3.79	0.10	0.20	0.90	6.70	13.50
固定碳	85.47	0.56	83.90	85.00	85.60	85.90	86.80
M40/M25	92.30	0.98	87.20	91.60	92.20	93.00	95.20
M10	5.36	0.96	2.60	4.80	5.40	6.00	8.00
焦末含量	9.68	4.11	3.20	7.30	8.70	11.00	38.40
CRI	26.65	2.86	19.50	24.80	26.60	27.90	36.30
CSR	62.41	3.61	52.80	60.50	62.40	64.90	71.30

注:下四分位数为25%的数据小于或等于该值;中位数为50%的数据小于或等于该值,即数据的中位点;上四分位数为75%的数据小于或等于该值。

SHAP值是第*i*个样本中当前特征对 y_i 最终预测值的贡献。在 $f(x_{ij}) > 0$ 时,特征增强预测值,具有正向作用。相反,如果 $f(x_{ij})$ 为负,则表示该特征降低了预测值,表明存在负作用。

1.4 模型评价指标

1.4.1 平均绝对值误差

回归决策树中,MAE是评价模型预测精度的常见指标,衡量的是预测值与试验值之间的差距的绝对值的平均数。计算方法如下。

$$\sigma_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

式中: σ_{MAE} 为平均绝对值误差; N 为样本数量; y_i 表示第*i*个样本的试验值; \hat{y}_i 表示第*i*个样本的预测值。

σ_{MAE} 越小表明误差越低,预测值与试验值更接近,模型性能越好。其优点在于计算不受正负号影响,易于理解,且对于异常值较为稳健。

1.4.2 决定系数

决定系数(R^2)是用于评估统计模型对观测数据拟合优度的一个重要指标,它反映了模型能够解释的因变量(响应变量)变异所占的比例。计算

方法如下。

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2} \quad (3)$$

式中： \bar{y} 表示全部样本试验值的平均值。

R^2 的取值范围在 0 到 1 之间，反映了因变量被自变量通过回归模型所解释的比例。当 R^2 趋近于 1 时，模型对数据的拟合程度较高；而当 R^2 接近 0 时，模型对因变量变化的解释能力有限，拟合效果较差。

2 决策树分析

利用 Pandas 数据分析库和机器学习中 SHAP 解释库进行数据分析。5 种分析方法：① sns. heatmap() 函数分析输入特征与输出特征之间的关系；② model. feature_importances_() 函数分析 8 种基础参数对 CRI 和 CSR 数据预测影响的重要性，总结出对结果影响的重要指标；③ shap_interaction_values() 函数解释基础参数之间的交互的关系，分析特征组合对试验结果的影响；④ shap_values() 函数反应每个基础样本的影响力，同时表现出影响的正负性；⑤ shap.decision_plot() 函数清晰直观地识别出每组决策主要影响的大小和方向。

2.1 特征相关热图

输入与输出特征相互关系的特征相关热图如图 2 所示。图 2 中，右侧刻度颜色深浅表示数据之间的相关程度，正值表示正相关，负值表示负相关，数值的绝对值越大表示相关程度越高。由图 2 可以看出：固定碳与挥发分、硫分、水分都有很大程度的相关性，呈负相关；固定碳与灰分之间的相关性较强，为 -0.91，即存在很强的多重共线性，这两个参数在进行特征工程时可以剔除其中之一，避免因多重共线性造成数据过拟合。焦末含量与 CRI 之间的热力值为 0，表示无相关性，该参数在预测 CRI 特征时可以剔除；焦末含量与 M40/M25 之间的热力值同样为 0，这两个参数在进行特征工程时同样可以剔除其中之一。

2.2 相关性分析

8 种基础参数对 CRI、CSR 指标的特征重要性分析如图 3 所示。由图 3 可见，CRI 值的影响因素

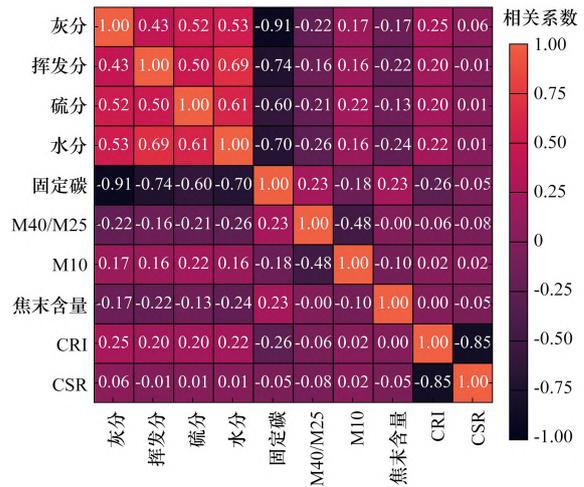


图 2 8 种基础参数与 CRI、CSR 指标的特征相关热图

Fig. 2 Heatmap of the characteristics of 8 basic parameters and CRI, CSR indicators

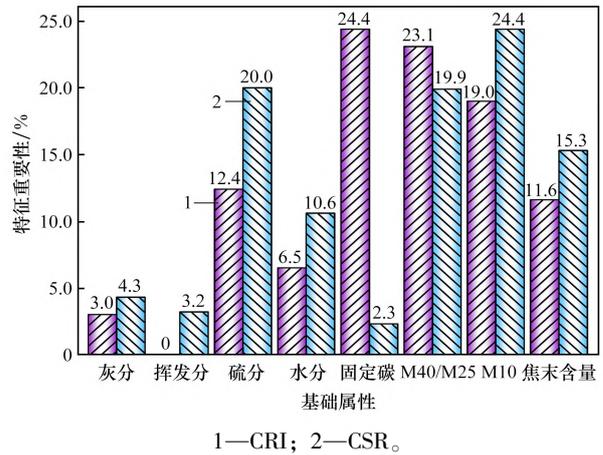


图 3 8 种基础参数对 CRI、CSR 两种指标的相关性分析

Fig. 3 Correlation analysis of 8 basic parameters for CRI and CSR indicators

按重要性从大到小的排序：固定碳 > M40/M25 > M10 > 硫分 > 焦末含量 > 水分 > 灰分 > 挥发分。其中，固定碳、M40/M25 以及 M10 这三项占据了大约 20% 的比例，是主要的影响因素。高固定碳含量意味着焦炭更为纯净和稳定，有助于降低 CRI 值，从而有利于维持高炉内焦炭骨架的稳定性并减少燃料消耗。相比之下，灰分和水分的影响较小，占比不足 10%，而挥发分对 CRI 几乎无影响。至于 CSR 值，影响因素的重要性排序：M10 > 硫分 > M40/M25 > 焦末含量 > 水分 > 灰分 > 挥发分 > 固定碳。M10、M40/M25 与硫分构成了约 20% 的重要影响部分。较高的 M10 值表明焦炭的耐磨性能较差，其在运输或使用过程中更易破碎，

这会间接影响到焦炭在高炉内的填充结构及其支撑作用, 进而影响 CSR。灰分、挥发分及固定碳对 CSR 的影响相对较小, 占比低于 10%。上述特征重要性的结论符合焦炭性能的一般经验判断, M40/M25 和 M10 是影响这两种指标的主要因素, 而灰分和挥发分对两者的影响都较为有限。固定碳对 CRI 有显著影响, 但对 CSR 的作用则微乎其微。

为了更加直观地体现出 CSR 和 CRI 的相关关系, 本文以 CRI 为自变量, CSR 为因变量进行分析。拟合方程如式(3)所示, 拟合结果如图 4 所示。由图 4 可以看出, CRI 与 CSR 的拟合曲线斜率为负值, 具有明显的负相关关系, 且拟合曲线的 R^2 为 0.723 5, 说明二者具有较强的负相关性。

$$y = -1.076 5x + 91.10 \quad (4)$$

式中: y 为 CSR 值, %; x 为 CRI 值, %。

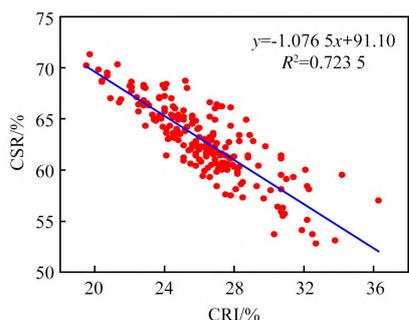


图 4 焦炭 CSR 与 CRI 指标的相关性分析

Fig. 4 Correlation analysis of CSR and CRI indicators of coke

2.3 SHAP 值分析

SHAP 交互图将 SHAP 值推广到更高阶, 实现快速、精确地交互计算, 它将每个预测返回至 1 个矩阵。其中, 主要影响在对角线上、交互影响

在对角线外, 用于揭示基础参数之间的交互关系, 如图 5 所示。每个子图的横轴表示 SHAP 值, 子图的宽带表示特征组合对结果影响。如 M10 与 M40/M25 特征组合(第一排第二组子图)中, 随着 SHAP 值的不断增大, 特征组合对结果的影响先变强后变弱; M10、水分与焦末含量、挥发分、灰分的矩阵重合于基线上, 表明特征组合对结果的影响很弱。

2.3.1 对 CRI 的预测分析

SHAP 值分析图的核心优势在于能够揭示每个样本中各特征的具体贡献程度, 不仅展示影响的大小, 还能体现其作用方向(正向促进或负向抑制), 从而实现对模型决策过程的细粒度解释。CRI 预测过程中基础参数对结果的影响如图 6 所示。由图 6 可见: 随着 SHAP 值的增大, 固定碳的特征值减小, 说明两者存在着负向线性关系。根据 SHAP 值的宽度来看, 固定碳含量与 CRI 之间的线性关系较强, 因此在工况中, 随着固定碳含量增加, CRI 有减弱的趋势; 在 M10 对 CRI 的影响中发现, 随着 SHAP 值的增大, M10 的特征值也随之增大, 说明两者存在着正向线性关系, 但对其 SHAP 宽度值进行比较发现, 其线性关系弱于固定碳, 所以在工况中随着 M10 值的增加, CRI 有增大的趋势。

图 7 决策图展现了决策树模型在处理 510 组样本时每一步的决策流程, 可揭示各重要因素的作用力度及方向。图 7 中, 中间灰色竖线代表模型的基础预测值, 而彩色线条则显示了实际预测结果。由图 7 可见, 8 种基本参数从下至上对 CRI 预测结果的影响逐级增强。固定碳、焦末含量对

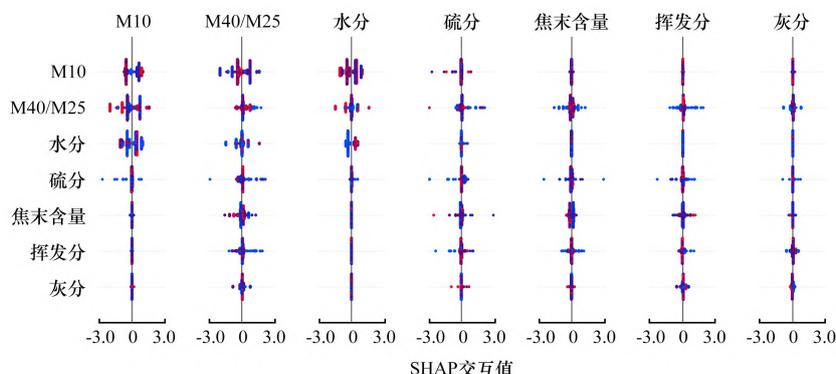


图 5 8 种基础参数的 SHAP 交互图

Fig. 5 SHAP interaction plot of 8 basic parameters

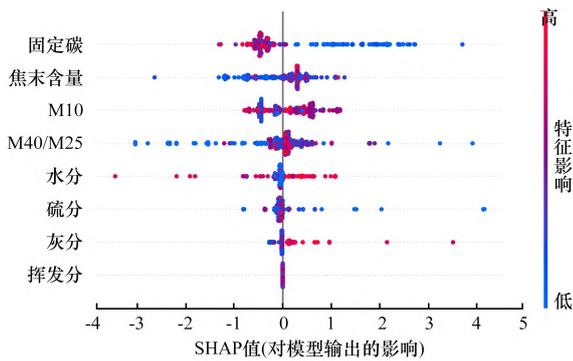


图 6 CRI 指标的 SHAP 值分析

Fig. 6 The SHAP value analyze CRI index

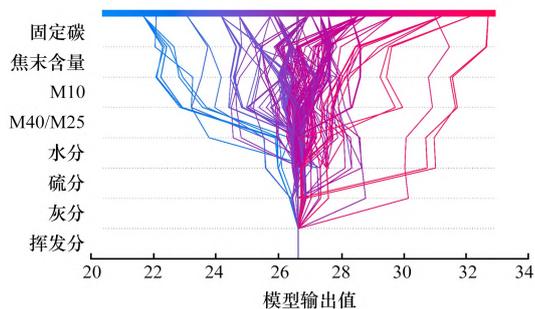


图 7 CRI 指标的决策分析

Fig. 7 Decision analysis CRI index

CRI 性能影响程度最大，硫分、灰分、挥发分对 CRI 性能影响程度最小。结合 SHAP 值分析发现，改善固定碳、焦末含量对改善 CRI 性能更为显著，固定碳对结果影响呈负向线性相关，焦末含量对结果影响呈负向线性相关。从整体决策来看，比例较多的决策曲线集中在基础预测值附近，表明自下而上各参数对 CRI 的决策逐渐增强，但不会有太大的波动，个别曲线波动较大。值得一提的是，挥发分的决策曲线始终与基础线重合，表明其未对预测结果产生显著影响。

2.3.2 对 CSR 的预测分析

CSR 预测的 SHAP 值分析如图 8 所示。由图 8 可见，随着 SHAP 值的增大，水分和固定碳的特征值不断增大，说明水分和固定碳对 CSR 值存在着正向线性关系。由坐标点的宽度发现，水分与结果的线性关系更强，表明在预测 CSR 性能的过程中，水分值增加比固定碳值增加对 CSR 性能的改变更为显著。数据点的稀疏程度表明水分和固定碳对预测影响不占主导因素。随着 SHAP 值的增大，其他基础参数的数值减小，表明其对 CSR 值存在着负向线性关系。分析数据点的密集程度发

现，其线性相关的强弱： $M10 > M40/M25 > 硫分 > 焦末含量 > 挥发分 > 灰分$ 。

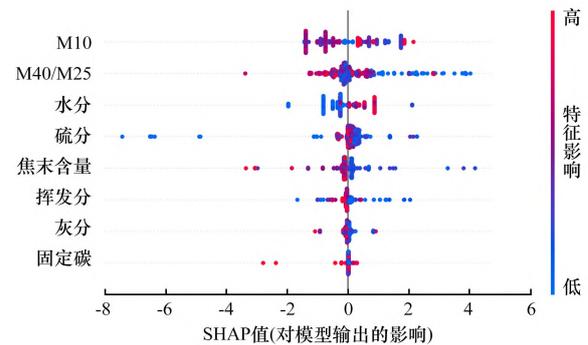


图 8 CSR 指标的 SHAP 值分析

Fig. 8 Analysis of SHAP Values for CSR Indicators

8 种基础参数在 CSR 预测中的决策过程如图 9 所示。由图 9 可见，510 组决策曲线中的基线值为 62.5，M10、M40/M25 对 CSR 性能影响程度最大，挥发分、灰分、固定碳对 CSR 性能影响程度最小。结合 SHAP 值分析发现，改善 M10 和 M40/M25 的含量对改善 CSR 性能更为显著，其呈负向线性相关。从整体曲线布局发现，有极少数的曲线在决策过程中出现较大波动，大部分的决策具有很强的规律性。与 CRI 的决策过程对比发现，8 种基础参数对 CSR 的预测结果更稳定。

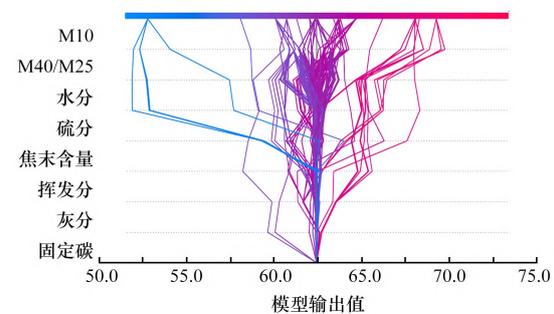


图 9 CSR 指标的决策分析

Fig. 9 Decision analysis CSR index

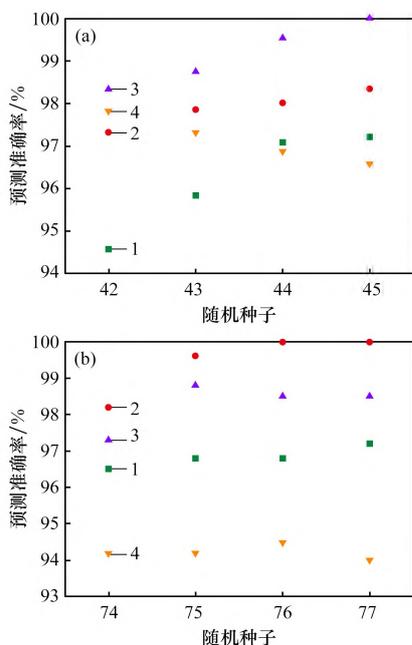
决策树模型分析结果总结出了 8 种基础属性对焦炭 CSR 和 CRI 性能的影响，以及各个属性之间的相互影响。特征相关热图、相关性和 SHAP 值分析图得到的重要性基本一致。根据现场经验综合判断，相关性分析的准确程度更高。同时，特征相关热图解释了各种参数之间的正负相关影响，SHAP 值分析图解释了特征对模型输出的影响，这也为在工业生产中调整 CSR 和 CRI 性能数值提供了参考。

3 工业应用

3.1 数据预测

将工业数据读取到机器学习算法模型中,通过网格搜索和交叉验证方法得到决策树模型的最佳匹配参数:①决策树模型1,利用灰分、挥发分、硫分、水分、固定碳、M40、M25、M10和焦末含量输入特征预测焦炭的CRI特性,设置决策树高度为19、随机种子数为44;②决策树模型2,利用灰分、挥发分、硫分、水分、固定碳、M40、M25、M10和焦末含量输入特征预测焦炭的CSR特性,设置决策树高度为18、随机种子数为75。

为验证机器学习决策树模型已达到最佳参数设置情况,通过算法修改随机种子数和决策树深度,统计 4×4 组预测准确率,如图10所示。由图10(a)的CRI预测结果可见:当决策树深度达到19时,预测准确率达到最高,依次增加随机种子数后预测准确率不断发生变化,但当随机种子数达到45时,预测准确率达到100%,说明训练模型已出现过拟合行为,这与网格搜索得到的最佳参数匹配。模型最佳参数的深度为19、随机种子数为44时,预测结果如图11(a)所示。



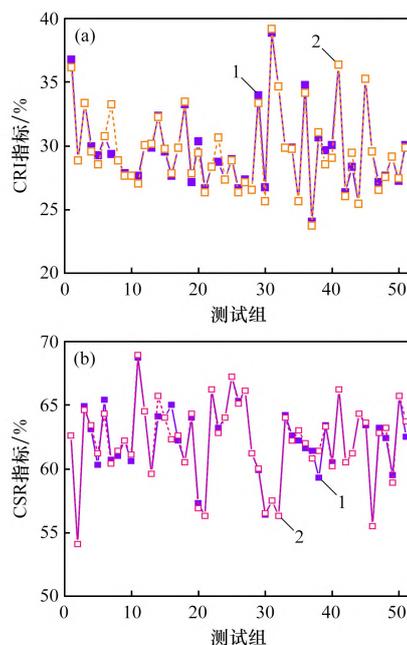
(a) CRI; (b) CSR

深度: 1—17; 2—18; 3—19; 4—20。

图10 不同参数下模型的预测准确率

Fig. 10 Prediction accuracy under different parameters

由图10(b)的CSR预测结果可见,当决策树深度达到18时预测准确率达到最高,随机种子数改变的同时,预测准确率也不断发生变化,但当随机种子数为76、77时,预测准确率达到100%,训练模型出现过拟合,同样验证与网格搜索得到的最佳参数匹配。模型最佳参数的深度为18、随机种子数为75时,预测结果如图11(b)所示。



(a) CRI; (b) CSR

1—试验值; 2—预测值。

图11 试验值和预测值对比

Fig. 11 Comparison of experimental and predicted values

图12显示的是51组测试数据点预测值与试验值之间的误差。由图12(a)可见:在CRI指标预测中, $>2\%$ 误差的数据有0组, $1\% \sim 2\%$ 误差的数据有5组,其余数据误差均小于 1% ;在CSR指标预测中, $>2\%$ 误差的数据有1组, $1\% \sim 2\%$ 误差的数据有4组,其余数据误差均小于 1% 。结果表明,参数调整良好的决策树算法对焦炭CRI和CSR性能预测精度达到 96% 以上,绝对误差基本控制在 2% 以内。由图12(b)可见:在CRI指标预测中, $>4\%$ 误差的数据有5组, $2\% \sim 4\%$ 误差的数据有6组, $<2\%$ 误差的数据有40组;在CSR指标预测中, $>4\%$ 误差的数据有1组, $2\% \sim 4\%$ 误差的数据有2组, $<2\%$ 误差的数据有48组。综合考虑试验值与预测值以及误差值的对比,决策树算法对焦炭CRI和CSR性能预测具有一定

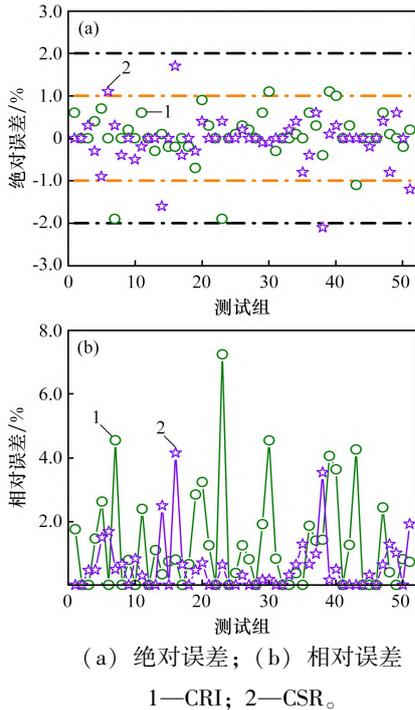


图 12 试验值与预测值误差

Fig. 12 Error of experimental and predicted values

的适应性。通过开发测试软件对工厂现场实际生
产应用有着一定的指导意义。

3.2 在线软件开发

基于 Python 3.9 运行环境，利用 Tkinter、Cv2 库函数将算法代码打包，开发形成线上预测软件，并通过工厂测试后投入应用。图 13(a) 为软件主要界面，包括数据读入、算法参数、数据训练和数据预测四大功能模块。图 13(b) 是实现决策树参数调节的功能，参数选择方式分为自动获取和手动输入两种。图 13(c) 是数据预测模块，输入灰分、挥发分、水分、硫分等基础属性数值，实现 CSR 和 CRI 数值的预测。预测结果如图 13(d) 所示。通过实时预测和工厂测试，软件预测精度分别达到 98.543% 和 96.825%，预测软件的运行时长为 0.1 ~ 0.3 s，实现在工业生产过程中对 CSR 和 CRI 两种属性的实时监控，以便后续优化调整基础属性值，避免原料浪费，促进高效、绿色生产。



(a) 主界面; (b) 调参系统; (c) 预测系统; (d) 结果显示

图 13 在线预测软件

Fig. 13 Online prediction software

4 结论

(1) 通过对比模型准确率且利用网格搜索和交叉验证方法确定模型超参数的最优组合。确定决策树模型超参数组合:深度为19、随机种子数为44,预测焦炭的CRI指标的准确率达98.543%;超参数组合:深度为18、随机种子数为75,预测焦炭的CSR指标的准确率达96.825%。

(2) 通过相关性分析,CRI数值影响的特征重要性为固定碳>M40/M25>M10>硫分>焦末含量>水分>灰分>挥发分;影响CSR数值的特征重要性为M10>硫分>M40/M25>焦末含量>水分>灰分>挥发分>固定碳。

(3) M40/M25和M10是影响CRI和CSR指标的主要因素,而灰分和挥发分对两者的影响都较为有限。固定碳对CRI有显著影响,但对CSR的作用则微乎其微。

(4) 利用Tkinter库和Cv2库将决策树程序源代码进行封装打包生成在线预测软件,实现对焦炭CRI和CSR指标实时监控,利于及时调整焦炭的成分达到生产需求,同时降低了碳排放和企业生产成本,促进企业高效和智能化生产。

参考文献:

- [1] 高冰,肖洪,张文强,等.高硫焦炭特性及对高炉冶炼的影响[J].炼铁,2017,36(1):56-58.
GAO Bing, XIAO Hong, ZHANG Wenqiang, et al. Characteristics of high-sulfur coke and its influence on blast furnace smelting [J]. Ironmaking, 2017, 36(1): 56-58.
- [2] 赵贵清.焦炭质量对高炉冶炼过程的影响[J].中国冶金,2007,17(11):14-17.
ZHAO Guiqing. Influence of coke quality on BF production [J]. China Metallurgy, 2007, 17(11): 14-17.
- [3] 刘颖义.王家岭瘦焦煤在宣钢配煤生产中的应用[J].中国冶金,2021,31(5):83-85.
LIU Yingyi. Application of Wangjialing lean coking coal in coal blending production of Xuanhua Steel [J]. China Metallurgy, 2021, 31(5): 83-85.
- [4] 侯启雄,孟佳文,曾泽泉,等.无烟煤与焦煤复配制备型焦的炭化过程研究[J].煤炭转化,2020,43(2):

58-64.

HOU Qixiong, MENG Jiawen, ZENG Zequan, et al. Study on carbonization process of formed coke prepared from anthracite and coking coal [J]. Coal Conversion, 2020, 43(2): 58-64.

- [5] 王鹏,齐洪涛,申岩峰,等.次烟煤与高硫焦煤共热解相互影响及动力学研究[J].洁净煤技术,2016,22(6):28-33.

WANG Peng, QI Hongtao, SHEN Yanfeng, et al. Research on interactions and kinetic during co-pyrolysis of subbituminous coal and high sulfur coking coal [J]. Clean Coal Technology, 2016, 22(6): 28-33.

- [6] 李杰,程欢,梁英华,等.用于高炉炼铁的低CSR焦炭机理探讨[J].铸造技术,2019,40(8):778-782.

LI Jie, CHENG Huan, LIANG Yinghua, et al. Study on the mechanism of low CSR coke used in blast furnace iron smelting [J]. Foundry Technology, 2019, 40(8): 778-782.

- [7] 孔德文.配煤结构优化在7m焦炉上的实践[J].河北冶金,2020(5):41-44;57.

KONG Dewen. Practice of coal blending structure optimization in 7 m coke oven [J]. Hebei Metallurgy, 2020(5): 41-44; 57.

- [8] 潘开灵,陈锡通,李德平.活惰组分比例对焦炭CRI和CSR指标的影响[J].燃料与化工,2017,48(6):25-27.

PAN Kailing, CHEN Xitong, LI Deping. Influence of active inert composition ratio on CRI and CSR of coke [J]. Fuel & Chemical Processes, 2017, 48(6): 25-27.

- [9] 庞克亮,刘冬杰,王明国,等.煤岩学在炼焦生产中的应用[J].钢铁,2015,50(10):26-29.

PANG Keliang, LIU Dongjie, WANG Mingguo, et al. Application of coal petrography analysis on coking production [J]. Iron and Steel, 2015, 50(10): 26-29.

- [10] 张代林,余亮,郑明东.炼焦煤的煤岩特征对其结焦性质的影响[J].钢铁,2009,44(1):15-18;30.

ZHANG Dailin, YU Liang, ZHENG Mingdong. Effect of petrographic properties of coking coal on quality of metallurgical coke [J]. Iron and Steel, 2009, 44(1): 15-18; 30.

- [11] 张代林,曾涛,李伟锋,等.焦煤显微结构特征与焦炭性质的关系[J].钢铁,2011,46(1):14-18.

ZHANG Dailin, ZENG Tao, LI Weifeng, et al. Relation between microstructure characteristics of coal and coke

- and quality of coke [J]. *Iron and Steel*, 2011, 46(1): 14 - 18.
- [12] 郭瑞,汪琦. 焦炭 CRI 和 CSR 指标的产生、发展和应用[J]. *炼铁*, 2012, 31(1): 45 - 48.
GUO Rui, WANG Qi. Generation, development and application of coke CRI and CSR indicators[J]. *Ironmaking*, 2012, 31(1): 45 - 48.
- [13] 陆泓彬,朱红春,姜周华,等. 基于 e-FCNN 的电弧炉终点温度预报[J]. *钢铁*, 2024, 59(1): 49 - 57.
LU Hongbin, ZHU Hongchun, JIANG Zhouhua, et al. Prediction of end-point temperature in electric arc furnace based on e-FCNN [J]. *Iron and Steel*, 2024, 59(1): 49 - 57.
- [14] 董择上,薛庆国,左海滨,等. 氧气高炉喷吹焦炉煤气数学模型[J]. *钢铁*, 2017, 52(4): 18 - 24.
DONG Zeshang, XUE Qingguo, ZUO Haibin, et al. Mathematical model analysis on coke oven gas injection into oxygen blast furnace[J]. *Iron and Steel*, 2017, 52(4): 18 - 24.
- [15] 苗泽凯,胡弘毅,程实,等. 焦炭反应性(CRI)和焦炭反应后强度(CSR)的相互关系[J]. *内蒙古石油化工*, 2015, 41(16): 1 - 3.
MIAO Zekai, HU Hongyi, CHENG Shi, et al. Interrelation between coke reaction index (CRI) and coke post-reaction strength (CSR)[J]. *Inner Mongolia Petrochemical Industry*, 2015, 41(16): 1 - 3.
- [16] 么秋香,杜美利,郎群. 焦炭强度预测回归分析及其 SPSS 的实现[J]. *中国煤炭*, 2012, 38(10): 74 - 77; 86.
YAO Qiuxiang, DU Meili, LANG Qun. Coke strength prediction by SPSS regression analysis[J]. *China Coal*, 2012, 38(10): 74 - 77; 86.
- [17] 程榆. 基于多元线性回归分析的焦炭质量预测研究[J]. *煤炭加工与综合利用*, 2017(7): 81 - 86.
CHENG Yu. Research on coke quality prediction based on multiple linear regression analysis[J]. *Coal Processing & Comprehensive Utilization*, 2017(7): 81 - 86.
- [18] 苏鑫,裴华健,吴迎亚,等. 应用经遗传算法优化的 BP 神经网络预测催化裂化装置焦炭产率[J]. *化工进展*, 2016, 35(2): 389 - 396.
SU Xin, PEI Huajian, WU Yingya, et al. Predicting coke yield of FCC unit using genetic algorithm optimized BP neural network [J]. *Chemical Industry and Engineering Progress*, 2016, 35(2): 389 - 396.
- [19] 阎少宏,吴宇航. 基于交叉验证的级联 BP 神经网络的焦炭质量预测模型[J]. *信息记录材料*, 2018, 19(10): 223 - 224.
YAN Shaohong, WU Yuhang. Coke quality prediction model based on cross-validation and cascade BP neural network[J]. *Information Recording Materials*, 2018, 19(10): 223 - 224.
- [20] 李爱莲,龚小帅. 量子遗传算法优化 BP 网络的焦炭质量模型[J]. *自动化应用*, 2017(8): 74 - 77.
LI Ailian, GONG Xiaoshuai. Coke quality model optimized by quantum genetic algorithm based on BP network[J]. *Automation Application*, 2017(8): 74 - 77.
- [21] 周洋,余文武,董相娟,等. 基于聚类分析和神经网络的高炉焦比预测模型[J]. *辽宁科技大学学报*, 2010, 33(3): 245 - 247; 257.
ZHOU Yang, YU Wenwu, DONG Xiangjuan, et al. A prediction model for blast furnace coke ratio with clustering analysis and neural network[J]. *Journal of University of Science and Technology Liaoning*, 2010, 33(3): 245 - 247; 257.
- [22] 詹艳艳,姜静,顾国利. 基于 GA 优化 RBF 网络的焦炭质量模型[J]. *电子技术*, 2015, 44(4): 16 - 18; 5.
ZHAN Yanyan, JIANG Jing, GU Guoli. The coke quality model based on GA optimization RBF network[J]. *Electronic Technology*, 2015, 44(4): 16 - 18; 5.
- [23] 刘剑,张凌波,王潇凌. 基于混沌菌群算法改进 BP 神经网络的焦炭塔生焦高度软测量建模[J]. *化工自动化及仪表*, 2019, 46(10): 811 - 815; 833.
LIU Jian, ZHANG Lingbo, WANG Xiaoling. Soft measurement modeling of coke height in coke tower based on chaos bacterial algorithm optimizing BP neural network [J]. *Control and Instruments in Chemical Industry*, 2019, 46(10): 811 - 815; 833.