

转炉吹炼终点碳锰预测数据驱动模型

洪 科^{1,3} 于学渊² 钟良才^{1,3*} 史秀全² 高 威² 赵 阳²

（1. 东北大学冶金学院，沈阳 110004； 2. 建龙集团抚顺新钢铁炼钢厂，辽宁 抚顺 113001； 3. 东北大学多金属共生矿生态化冶金教育部重点实验室，沈阳 110004）

摘 要：针对转炉吹炼终点碳和锰，采用大数据与机器学习算法 LGBM (Light Gradient Boosting Machine, 轻量梯度提升机)，建立了数据驱动的转炉吹炼终点碳、锰含量预测模型，模型输入变量包括铁水成分、铁水重量、废钢重量、耗氧量和造渣剂加入量，采用学习曲线调参方法确定模型的超参数最优值。研究表明，在大数据的前提下，数据分布均衡有利于提高机器学习算法建立的数据驱动模型的预测效果。在 $(-0.025\%, 0.025\%)$ 、 $(-0.020\%, 0.020\%)$ 和 $(-0.015\%, 0.015\%)$ 的误差范围，终点碳预测模型的命中率分别为 95.43%、94.06% 和 89.73%；终点锰预测模型的命中率分别为 97.00%、95.69% 和 90.61%。终点锰预测模型的均方根误差 (RMSE) 与终点碳预测模型的相近，但其决定系数 (R^2) 比终点碳预测模型的高，更接近 1。

关键词：转炉吹炼 终点碳 终点锰 数据驱动模型 预测

Data Driven Models of Carbon and Manganese Content Prediction at Blowing End Point in Converters

HONG Ke^{1,3}, YU Xueyuan², ZHONG Liangcai^{1,2}, SHI Xiuquang², GAO Wei², ZHAO Yang²

(1. School of Metallurgy, Northeastern University, Shenyang 110004, China; 2. Steelmaking Plant, Fushun New Iron & Steel Company, Jianlong Group, Fushun 113001, Liaoning, China; 3. Key Laboratory of Ecological Metallurgy of Polymetallic Ore of Ministry of Education, Northeastern University, Shenyang, 110004, China)

Abstract: For the carbon and manganese content at the end point of converter converting, data driven prediction models for carbon and manganese content at the end point of converter blowing were established by using big data and machine learning algorithm LGBM (Light Gradient Boosting Machine). The input variables of the model include the composition of hot metal, the weight of hot metal, the weight of scrap, oxygen consumption and the addition amount of slagging formers. The learning curve parameter adjustment method was used to determine the optimal hyperparameter values of the two models. The research results indicate that under the premise of big data, balanced data distribution is beneficial for improving the predictive performance of data-driven models established by machine learning algorithms. Within the error range of $(-0.025\%, 0.025\%)$, $(-0.020\%, 0.020\%)$, and $(-0.015\%, 0.015\%)$, the hit rates of the endpoint carbon prediction model are 95.43%, 94.06%, and 89.73%, respectively; The hit rates of the endpoint manganese prediction model were 97.00%, 95.69%, and 90.61%, respectively. The root mean square error (RMSE) of the endpoint manganese prediction model is near to that of the endpoint carbon prediction model, but its coefficient of determination (R^2) is higher and closer to 1 than that of the endpoint carbon prediction model.

Key words: converter blowing, endpoint carbon, endpoint manganese, data-driven model, prediction

1. 前言

目前转炉炼钢是我国的主要炼钢方法，转炉吹炼终点控制十分重要的，精确控制转炉吹炼终点，可以提高钢水质量，防止终点过氧化；减少或避免补吹，缩短冶炼实际，提高转炉产量；降低吹损和原材料消耗，节约能量，降低生产成本。因此，冶金工作者对转炉吹炼终点预报模型开展了许多的研究[1-12]。

本研究从转炉生产获得的实际数据，经数据处理，利用机器学习算法，建立转炉吹炼终点碳、锰含量的数据驱动模型，为转炉吹炼终点控制系统提供预测模型。

2. 研究方法

通过尝试不同的智能算法建立的数据驱动模型的预测效果对比，最后决定采用 LGBM 算法建立的数据驱动模型作为转炉终点碳锰预报模型。LGBM 是一个由微软提供、实现 GBDT(Gradient Boosting Decision Tree)算法的轻量级框架，解决了 GBDT 在处理大样本和高维度中遇到的问题，可应对多重共线性特征，提高模型的鲁棒性。GBDT 是以 Boosting 思想为基础的迭代决策树算法，通过学习之前回归树的残差，来对当前回归树进行训练，然后将所有回归树的结果累加得到最终的预测结果。

转炉实际生产数据集 $D=\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ ，转炉终点预测模型的最大迭代次数为 T ，损失函数为 L ，输出预测模型 $f(x)$ ，模型算法具体流程如下：

给出初始化弱学习器

$$f_0(x) = \arg \min_c \sum_{i=1}^m L(y_i, c) \quad i=1, 2, 3, \dots, m \quad (1)$$

式中 c 为使损失函数极小化的常数值，即使损失函数极小化的最优预测值，一般采用平方误差损失函数时 c 为训练集中炉次实际值的均值。本研究的终点预测值的平方误差损失函数为：

$$L(y_i, f(x_i)) = (y_i - f(x_i))^2 \quad (2)$$

计算损失函数的负梯度，迭代第 t 次第 i 炉次的终点预测值的损失函数负梯度为：

$$r_{ti} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{t-1}(x)} \quad (3)$$

其中，下标 t 为迭代次数， $t=1, 2, 3, \dots, T$ 。利用 (x_i, r_{ti}) 作为新的训练样本，拟合得到第 t 棵回归树，其叶子节点区域为 R_{tj} ， $j=1, 2, 3, \dots, J$ ， J 为回归树叶子节点个数。在叶子区域内求得最佳拟合值：

$$c_{tj} = \arg \min_c \sum_{x_i \in R_{tj}} L(y_i, f_{t-1}(x_i) + c) \quad (4)$$

更新强学习器：

$$f_t(x) = f_{t-1}(x) + \sum_{j=1}^J c_{tj} I(x \in R_{tj}) \quad (5)$$

式中 $I(x \in R_{tj})$ 用于判断集合中元素的指示函数。迭代 T 次后得到转炉吹炼终点预测模型为：

$$f(x) = f_T(x) = f_0(x) + \sum_{t=1}^T \sum_{j=1}^J c_{tj} I(x \in R_{tj}) \quad (6)$$

LGBM 在功能上的优化主要包括：直方图算法(Histogram)，基于深度限制的 Leaf-wise 的叶子垂直生长算法。

（1）基于 Leaf-wise 的决策树生长策略

通常决策树算法的生长策略是 **Level-wise**，如图 1 所示，主要思想是位于同一层的叶子节点每次都同时分裂，导致某些叶子节点的分裂增益减小，增加模型的计算量。**LGBM** 则采用 **Leaf-wise** 生长策略，如图 2 所示，每当进行分裂时，在当前所有叶子节点中，找到分裂增益最大的叶子节点进行分裂，并不是所有叶子节点都进行分裂，减少了模型的计算量，降低误差，提高模型的预测精度。

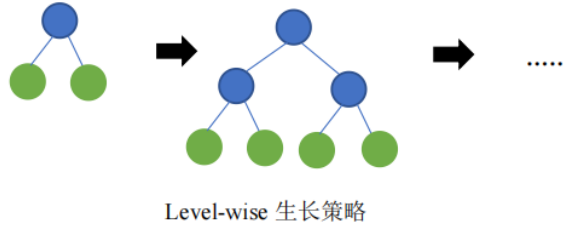


图 1 Level-wise 生长策略
Fig. 1 Level-wise growth strategy

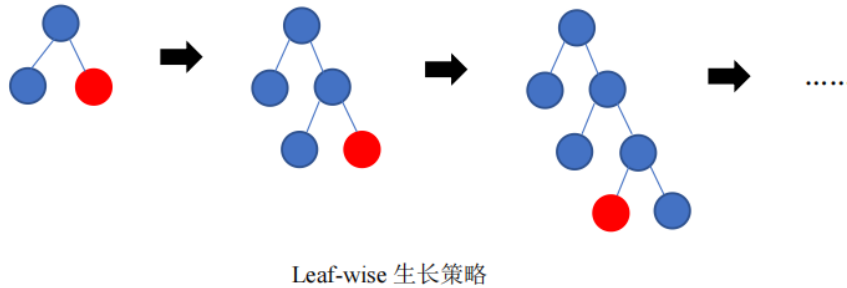


图 2 Leaf-wise 生长策略
Fig. 2 Leaf-wise growth strategy

（2）直方图算法

直方图的主要思想是首先对特征值进行装箱处理，将连续的浮点特征值离散成 n 个整数，形成多个单独的箱体（bins），与此同时构成一个宽度为 n 的直方图，如图 3 所示。

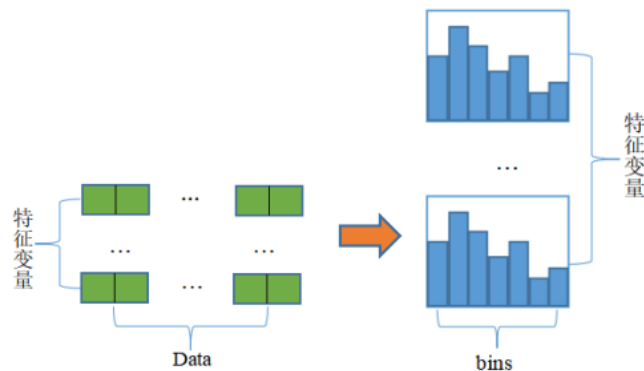


图 3 直方图算法
Fig. 3 Histogram algorithm

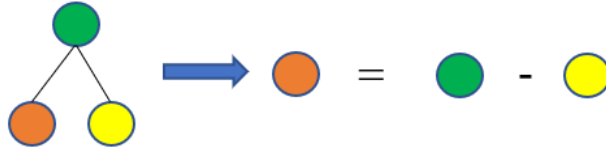


图 4 直方图做差加速算法

Fig. 4 Histogram subtraction acceleration algorithm

在遍历数据的时候，直方图中累积统计量以离散化后的值作为索引，当已完成一次遍历数据后，直方图已经累积了需要的统计量，之后依据直方图离散化后的值，寻找最优切分点，很大程度上减少了模型的计算量。直方图算法还可以做差加速，当前叶子节点的直方图可以由其父节点和兄弟节点做差得到，如图 4 所示，进一步提升模型训练速度。

采用均方根误差（RMSE）和决定系数（ R^2 ）评价与检测模型预测效果，计算公式如下：

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2} \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (f_i - y_i)^2}{\sum_{i=1}^N (\bar{y}_i - y_i)^2} \quad (8)$$

式中 N 为样本总量， f_i 为模型预测值， y_i 为样本实际值， $\bar{y}_i$ 为样本均值。

R^2 的取值范围在 0~1 之间， R^2 越接近 1，表明拟合效果越好。RMSE 可以衡量预测值与实际值的误差，其值越接近 0，模型的预测精度越高。

从冶金原理可知，影响转炉吹炼终点碳、锰含量的因素主要有铁水成分和重量、废钢重量、耗氧量、造渣剂消耗、吹炼枪位。本研究采用的模型输入参数包括：铁水 C、Si、Mn、P、S、Ti，铁水重量，废钢重量、耗氧量、石灰量、镁球量、石灰石量、生白云石量和矿石量。

在建立数据驱动模型时，机器学习算法的超参数是否重要，需要根据数据特点，确定算法的最优超参数。本研究采用学习曲线调参的方法来确定转炉吹炼终点碳、锰含量预测模型的 LGBM 算法的超参数最优值。

3 结果与分析

从转炉实际生产的经数据处理后的 1458 炉数据按 70% 和 30% 的比例随机划分为 1020 组训练数据、438 组测试数据；模型的超参数依据人工经验确定搜索范围，通过学习曲线方法寻找最优参数组合，以 R^2 为目标函数，确定模型的最佳超参数，表 1 和表 2 分别为最佳的超参数组合。

表 1 终点碳预测模型的最优参数组合

Table 2 Optimal parameters for the end-point carbon prediction model		
参数	寻优参数范围	寻优值
learning_rate	(0.01, 1.0)	0.018
max_depth	(1, 30)	2
n_estimators	(100, 1000)	830
num_leaves	(1, 50)	4
max_bin	(10, 500)	100
bagging_fraction	(0.1, 1.0)	0.6
bagging_freq	(1, 20)	7
feature_fraction	(0.01, 1.0)	0.7
feature_fraction_seed	(1, 20)	9
bagging_seed	(1, 20)	9
min_data_in_leaf	(1,10)	3
min_sum_hessian_in_leaf	(1,10)	3

表 2 终点锰预测模型的最优参数组合

Table 2 Optimal parameters for the end-point carbon prediction model		
参数	寻优参数范围	寻优值
learning_rate	(0.01, 1.0)	0.02
max_depth	(1, 30)	6
n_estimators	(100, 1000)	230
num_leaves	(1, 50)	4
max_bin	(10, 500)	220
bagging_fraction	(0.1, 1.0)	0.96
bagging_freq	(1, 20)	10
feature_fraction	(0.01, 1.0)	0.9
feature_fraction_seed	(1, 20)	9
bagging_seed	(1, 20)	9
min_data_in_leaf	(1,10)	2
min_sum_hessian_in_leaf	(1,10)	3

在上述的模型算法最优超参数下，得到的转炉吹炼终点碳含量预测结果如图 1 和图 2 所示，由图可知，在 0.04%~0.10%的碳含量范围，在误差（-0.015%，0.015%）范围外的数据点较少，在此终点碳范围，终点碳预测模型的预测效果较好，这是由于在此终点碳含量范围数据相对较多，模型训练时，可以学到在此终点碳含量范围的知识，因此，预测效果较好。而在终点碳含量大于 0.10%时，有较多的炉次在误差（-0.015%，0.015%）范围外，同时看出，在此终点碳含量范围，数据较少，模型在此范围学到的知识较少，因此终点碳的预报效果没有在低碳范围好。

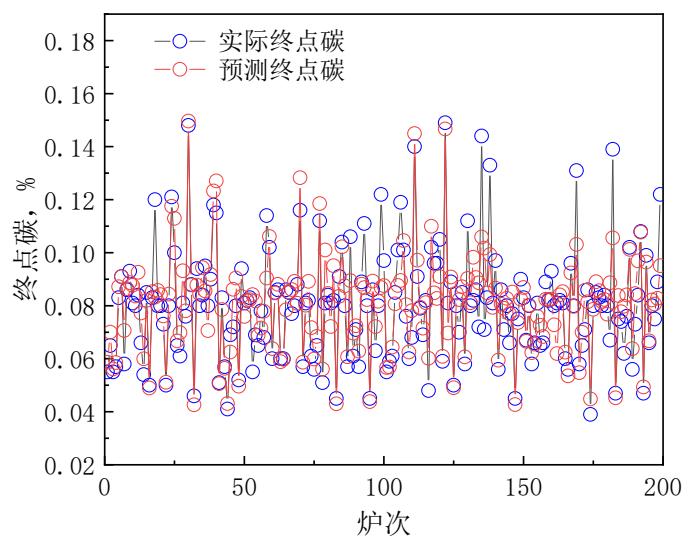


图 1 测试数据前 200 炉终点碳模型预测结果

Fig. 1 The prediction results of end-point carbon model from the first 200 heats

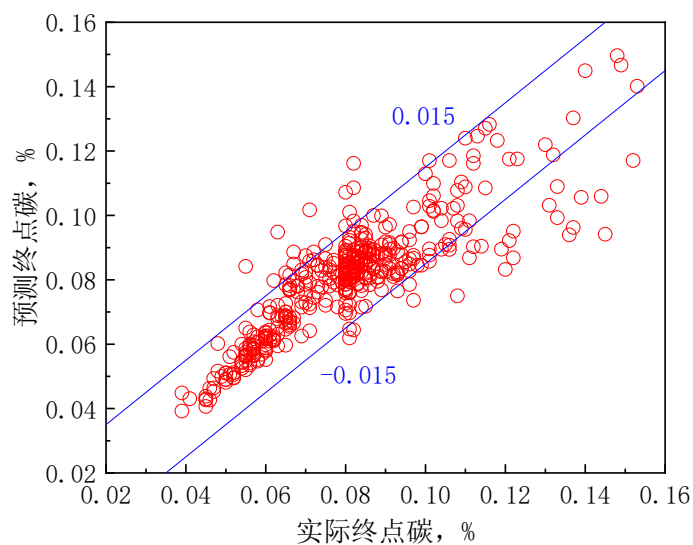


图 2 终点碳模型预测值与实际值对比

Fig. 2 Comparison between predicting and actual values from end point carbon model

图 3 和图 4 为转炉终点锰预测模型的预测效果, 由图可知, 由于终点锰含量在 0.08%~0.19%之间数据点发布较为均衡, 在该范围内终点锰预测模型的预测效果较好, 只是在此含量范围之外的两侧有少量炉次的预测的终点锰落在 $(-0.015\%, 0.015\%)$ 的误差范围之外。

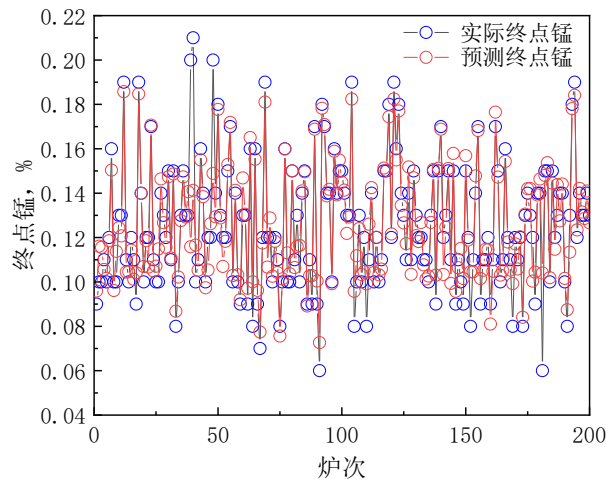


图 3 测试数据前 200 炉终点锰模型预测结果

Fig. 3 The prediction results of end-point manganese model from the first 200 heats

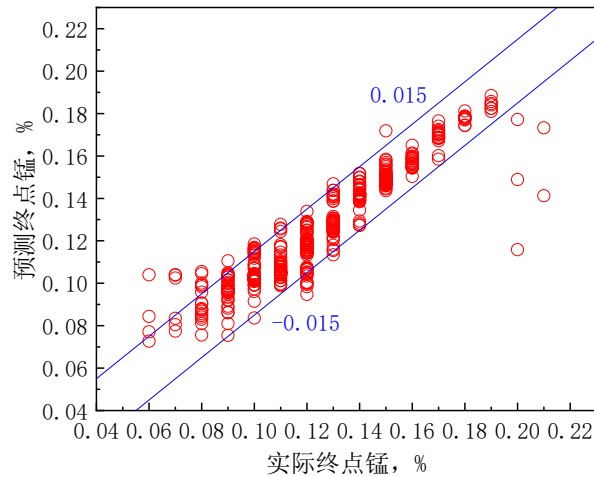


图 4 终点碳模型预测值与实际值对比

Fig. 4 Comparison between predicting and actual values from end point manganese model

表 3 为转炉终点碳、锰预测模型在不同误差范围的命中率，由表可知，在 $(-0.025\%, 0.025\%)$ 、 $(-0.020\%, 0.020\%)$ 和 $(-0.015\%, 0.015\%)$ 的误差范围，终点碳预测模型的命中率分别为 95.43%、94.06% 和 89.73%；终点锰预测模型的命中率分别为 97.00%、95.69% 和 90.61%。表 4 为终点碳、锰预测模型的均方根误差 (RMSE) 和决定系数 (R^2)，可以看出，终点锰预测模型的均方根误差 (RMSE) 与终点碳预测模型的相近，但其决定系数 (R^2) 比终点碳预测模型的高，更接近 1。

表 3 转炉终点碳锰预测模型的命中率

误差范围	终点碳命中率, %	终点锰命中率, %
$(-0.025\%, 0.025\%)$	95.43	97.00
$(-0.020\%, 0.020\%)$	94.06	95.69
$(-0.015\%, 0.015\%)$	89.73	90.61

表 4 转炉终点碳锰预测模型的评价指标

评价指标	终点碳	终点锰
RMSE	0.0099	0.0104
R^2	0.76	0.87

4 结论

（1）在大数据的前提下，数据分布均衡有利于提高机器学习算法建立的数据驱动模型的预测效果。

（2）在（-0.025%，0.025%）、（-0.020%，0.020%）和（-0.015%，0.015%）的误差范围，终点碳预测模型的命中率分别为 95.43%、94.06%和 89.73%；终点锰预测模型的命中率分别为 97.00%、95.69%和 90.61%。

（3）终点锰预测模型的均方根误差（RMSE）与终点碳预测模型的相近，但其决定系数（ R^2 ）比终点碳预测模型的高，更接近 1。

致谢

本研究得到了国家自然科学基金项目（U21A20117，51574069）；中央高校基本科研业务专项资金资助（N2125018）；科技部国家重点研发计划资助项目（2017YFB0304100）资助，在此表示衷心感谢。

参考文献

- [1] Bae Juhee, LI Y, Ståhl N, et al. Using machine learning for robust target prediction in a basic oxygen furnace system [J], Metall Mater Trans B, Process, 2020, 51(4): 1632.
- [2] HAN M, LIU C. Endpoint prediction model for basic oxygen furnace steel-making based on membrane algorithm evolving extreme learning machine [J], Applied Soft Computing, 2014, 19: 430.
- [3] Jo H, Hwang H J, Phan D, et al. Endpoint temperature prediction model for LD converters using machine-learning techniques [C]. 2019 IEEE 6th International Conference on Industrial Engineering and Applications, ICIEA 2019, 2019: 22.
- [4] Sala D A, Jalalvand A, Van Yperen-De Deyne A, et al. Multivariate time series for data-driven endpoint prediction in the basic oxygen furnace [C], Proceedings of 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018: 1419.
- [5] Rahnama A, LI Z, Sridhar S, Machine learning-based prediction of a BOS reactor performance from operating parameters [J], Processes, 2020, 8(3): 371.
- [6] Nenchev B, Panwisawas C, YANG X, et al. Metallurgical data science for steel industry: A case study on basic oxygen furnace [J], Steel Research International, 2022, in Press.
- [7] QI L, LIU H, XIONG Q, et al. Just-in-time-learning based prediction model of BOF endpoint carbon content and temperature via vMF mixture model and weighted extreme learning machine [J], Computers and Chemical Engineering, 2021, 154: 1.
- [8] 彭开香, 张学艺, 胡鑫裕. 一种基于即时学习的局部模型集成转炉炼钢终点预测方法[J], 冶金自动化, 2022, 46(2): 110.
- [9] 谷茂强, 徐安军, 刘旋, 等. 基于数据驱动的转炉二吹阶段钢水温度动态预测模型[J], 工程科学学报, 2022, <https://doi.org/10.13374/j.issn2095-9389.2022.01.05.002>.
- [10] 徐 钢, 黎 敏, 徐金梧, 等. 基于函数型数字孪生模型的转炉炼钢终点碳控制技术[J], 工程科学学报, 2019, 41(4): 5217.
- [11] 曾鹏飞, 刘辉. 基于二次相似性度量的即时学习转炉炼钢 终点碳温软测量方法[J], 计算机集成制造系统, 2021, 21(5): 1429.
- [12] 杨晓猛, 赵 阳, 钟良才, 等. 基于 XGBoost 算法的转炉吹炼终点预报[J], 炼钢, 2021, 37 (6) : 1.