

基于 XGBoost 算法的 LF 精炼钢水成分预报模型

李世健^{1,3,4}, 田大成¹, 刘震², 谢天², 秦振¹, 张刘瑜^{1,4}

1.苏州思萃熔接技术研究所有限公司, 江苏 张家港 215628; 2. 南京强思数字科技有限公司, 江苏 南京 210000; 3.北京科技大学钢铁冶金新技术国家重点实验室, 北京, 100083; 4. 江苏永钢集团, 江苏, 张家港 215628

A forecasting Model of Steel Composition of LF Refining Based on XGBoost

Algorithm

LI Shi-jian^{1,3,4}, TIAN Da-cheng¹, LIU Zhen², XIE Tian², QIN Zhen¹, ZHANG Liu-yu^{1,4}

1. Suzhou Sitri Welding Technology Research Institute Co., Ltd., Zhangjiagang, 215628, China; 2. Chance Digital Science & Technology Co., Ltd., Nanjing 210000, China; 3. State Key Laboratory of Advanced Metallurgy, University of Science and Technology Beijing, Beijing 100083, China; 4. Jiangsu Yonggang Group Co., Ltd., Zhangjiagang, 215628, China

1. 前言

钢铁工业是国民经济的重要支柱产业，是建设现代化强国的重要支撑。2021 年我国钢铁产量达到 10 亿吨，占世界总产量 53%。在以电弧炉或转炉-LF 精炼-真空脱气-连铸为主流的现代化炼钢流程中，LF 工序起着承上启下的作用，被誉为炼钢流程中的缓冲器^[1]。初炼后的钢水温度和成分通常具有一定波动性，通过 LF 精炼可以对其进行控制，定时为脱气或连铸工序提供温度和成分均合格的钢水。鉴于其优势，几乎所有优特钢的生产都需经过 LF 精炼工序，但企业目前在 LF 工序也面临着一些共性问题，例如，初炼炉终点往往存在一定的波动性，钢包到站后具体加多少脱氧剂、渣料，往往根据操作人员的经验而定，导致 LF 初期钢水成分和成渣性波动较大，这种波动性通常具有一定遗传性，最终导致产品质量波动大。如何实现 LF 精炼初期钢水成分的稳定控制是当前冶金工作者面临的一大难题。

当前，通过在物联网、大数据以及云计算基础上建立起来的智能制造技术，是实现钢铁工业生产过程智能化，助推其转型升级的有效方法^[2]。XGBoost (Extreme Gradient Boosting) 是由陈天奇在 GBDT 算法的基础上提出的集成算法模型，具有较高的运算速率和较好的收敛能力，且对特征数据的输入要求不高，从而被广泛应用在数据挖掘、统计、机器学习等领域^[3]。李红利、Wang Yuanyuan 等^[4,5]运用 XGBoost 集成算法开发了 LF 钢液温度预报模型；杨晓猛等^[6]运用 XGBoost 集成算法建立了转炉吹炼终点碳含量和温度预报模型，均取得了较好的预报效果。基于此，本文以江苏永钢集团炼钢二分厂实际生产数据为依据，运用 XGBoost 算法对 LF 精炼初期钢水成分 Al、Si、Mn 进行预报，旨在指导转炉出钢及 LF 精炼初期工艺，提高钢水成分的稳定性的。

2. 特征数据提取和预处理

永钢二炼钢厂生产流程为：50t 转炉-氩站-LF 精炼-VD 脱气-连铸。结合冶金原理分析和现场操作经验，初步筛选出的影响 LF 初期钢水成分的特征参数包括：铁水成分、转炉过程工艺参数、终点钢水成分和温度、出钢过程合金和渣料的加入量、氩站钢水成分、钢包温度、LF 初期原辅料的加入量等。

在转炉和 LF 工序中，系统的复杂性、仪器仪表本身故障以及工作操作的不当，导致极易出现异常数据和缺失数据。数据的准确性是精准预报的前提，因此很有必要对数据进行预处理。调取 python 中 duplicate 函数剔除重复项，使用 dropna 函数剔除数据库中的空值；利用 z-score 方法对数据进行标准化处理，消除量纲和变量自身变异大小以及数值大小的影响，数据标准化不会使原始数据信息发生变动^[7]。

3. 特征工程

特征工程在机器学习中是至关重要的一环。其主要目的是改变预测建模的特征以便更好的适应算法的训练，通过生成判别性高的特征来避免包含不相关参数，故又称之为变量消除。

XGBoost 筛选特征。在 Python 环境下运用 XGBoost 中的 plot-importance 函数建立输入特征与输出特征之间的关系，并根据各输入特征对输出特征的贡献度排序。排序原理为：特征重要性是通过对数据集中的每个特征进行计算，并进行排序得到。在单个决策树中通过每个特征分裂点改进性能度量的量来计算属性重要性，由节点负责加权和记录次数，记录次数即为该特征的分数量^[8]。特征变量分别与 LF 精炼钢水成分中 Al、Si、Mn 元素含量重要度相关性排序如图 1-3 所示。

特征重要性排序显现出特征对目标指数的贡献度，经过反复迭代，发现剔除分数较小的特征有利于模型训练。对 LF 精炼钢水成分中 Al 元素含量贡献度最大的有铝豆添加量、炉盖炉龄、进站温度、氩后 S 含量及钢包温度等因素；对 Si 元素含量贡献度最大的有氩后 Si 含量、炉盖炉龄、氩前 Si 含量及硅锰合金用量等因素；对 Mn 元素含量贡献度最大的有氩后 Mn 含量、铁水-氩后 Mn 元素含量差、氩前 Mn 含量等因素。

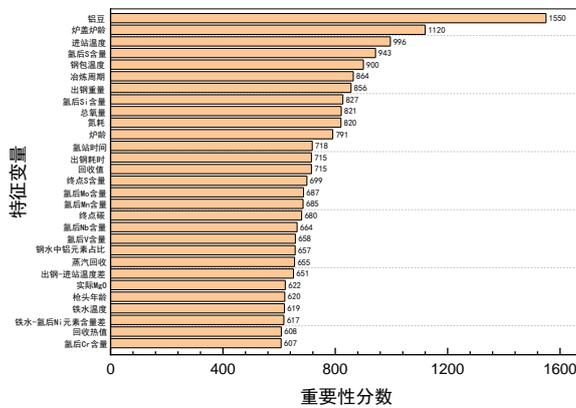


图 1 LF 精炼处理钢水成分 Al 的特征变量重要性

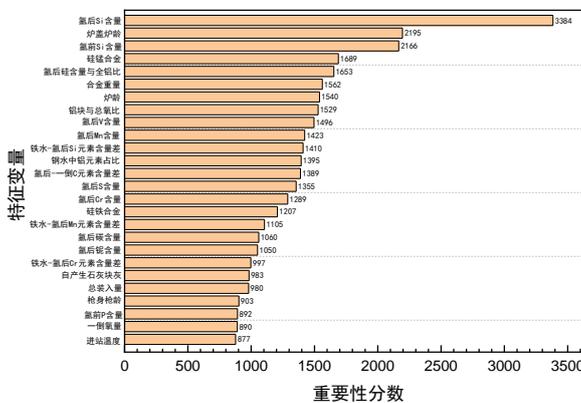


图 2 LF 精炼处理钢水成分 Si 的特征变量重要性

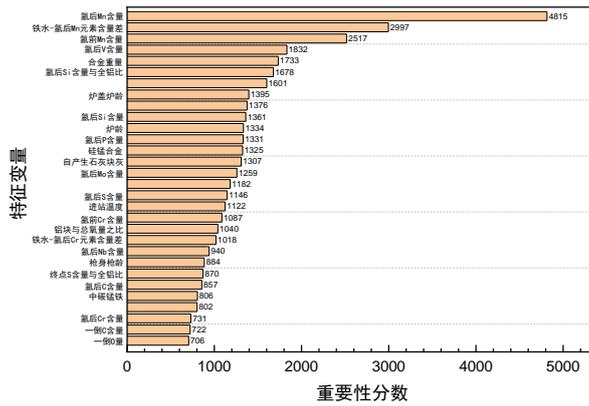


图3 LF 精炼处理钢水成分 Mn 的特征变量重要性

4. LF 精炼初期钢水成分预测模型

4.1. XGBoost 算法

XGBoost 是一种基于梯度 Boosting 的集成学习算法，又称极限梯度提升算法。其采用 CPU 多线程进行并行计算，具有极高的准确度和运算速度^[9]。其本质是一种树模型，不断进行特征分裂来增加一颗树去拟合上次预测的残差。目的是对原有目标函数进行改写和优化，同时进行泰勒展开，使算法收敛得更快，最终获得最优解。设立目标函数为：

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (1)$$

式中： Obj 为目标函数； l 为损失函数，即预测值与目标值间差距； Ω 为模型复杂度的正则化项，该项可降低模型的复杂度避免过拟合^[10]。 $\hat{y}_i^{(t-1)}$ 为第 i 个样本第 $t-1$ 轮迭代； f_t 为当前迭代误差。

4.2. 评价标准

评价标准采用平均绝对误差 (MAE) 和 R 平方 (R^2) 两种方法。其中 MAE 结果越小、 R^2 结果越接近 1，说明模型拟合效果越好，预测值和实际值差异越小。评价指标如下：

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

$$R^2 = 1 - \frac{\sum_{i=0}^m (y_i - \hat{y}_i)^2}{\sum_{i=0}^m (y_i - \bar{y}_i)^2} = 1 - \frac{RSS}{\sum_{i=1}^m (y_i - \bar{y}_i)^2} \quad (3)$$

式中： n 、 m 为样本预测个数； \hat{y}_i 为模型预测值； y_i 为实际值； \bar{y}_i 为平均值； RSS 为残差平方和。

4.3. XGBoost 回归模型训练

LF 精炼钢水成分预测模型所用数据均来自永钢二钢厂 2019 年 1 月 1 日至 2023 年 1 月 1 日的实际生产数据，其中 Si、Mn 元素收得率的预测模型将 42870 组数据随机分为训练数据 34296 组，测试数据 8757 组；Al 元素收得率的预测模型将 2000 组数据随机分为训练数据 1600 组，测试数据 400 组。模型经过反

复迭代，模型最主要的训练参数设置如下：booster 设为 gtree（树模型）；max-depth（树的最大深度）均设置为 6；eta（学习率）在 Si、Mn 模型中设置为 0.03，Al 模型中设置为 0.01；gbtree 分类器数目在 Si、Mn 模型中设置为 900，Al 模型中设置为 1000，其中学习率与分类器个数有着密切联系。

图 4-6 为 LF 精炼初期钢水成分中 Al、Si、Mn 三种元素含量的 XGBoost 模型预测值与实际值的对比图。Si、Mn 元素含量的 XGBoost 模型基分类器设置为 900 时，MAE 分别为 0.014151 和 0.022906， R^2 均达到了 0.988921；Al 元素含量的 XGBoost 模型基分类器设置为 1000 时，MAE 为 0.002345， R^2 达到了 0.683041。由此可见，LF 精炼钢水成分预报模型整体预报精度较高，对产线实际生产具有良好的指导意义，并且随着后续数据量的逐渐增多，排除多余的异常点和噪音，预测模型精度会随之升高。

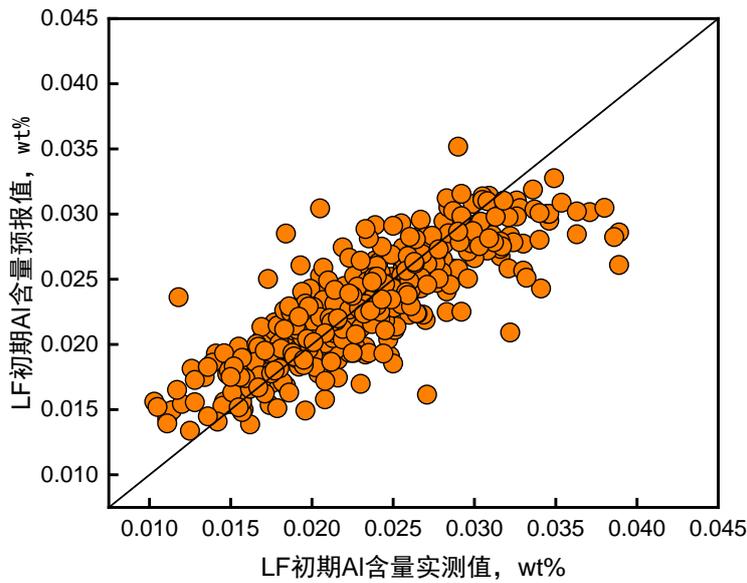


图 4 LF 精炼初期 Al 含量实测值与预报值对比

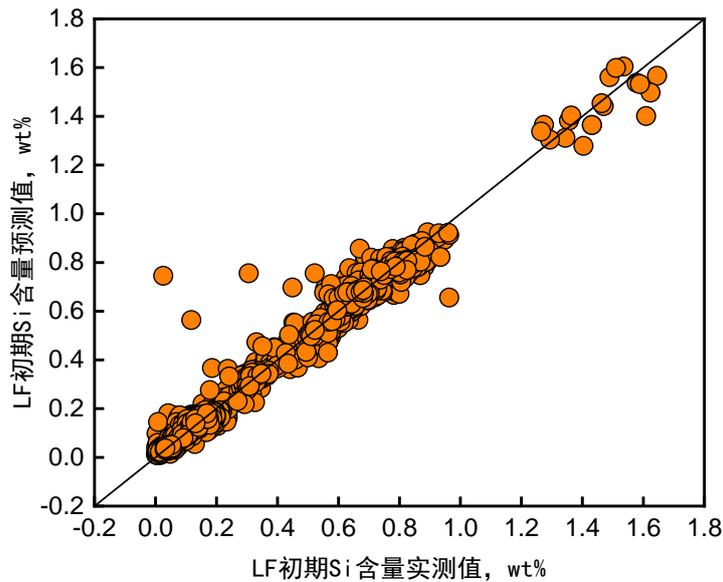


图 5 LF 精炼初期 Si 含量实测值与预报值对比

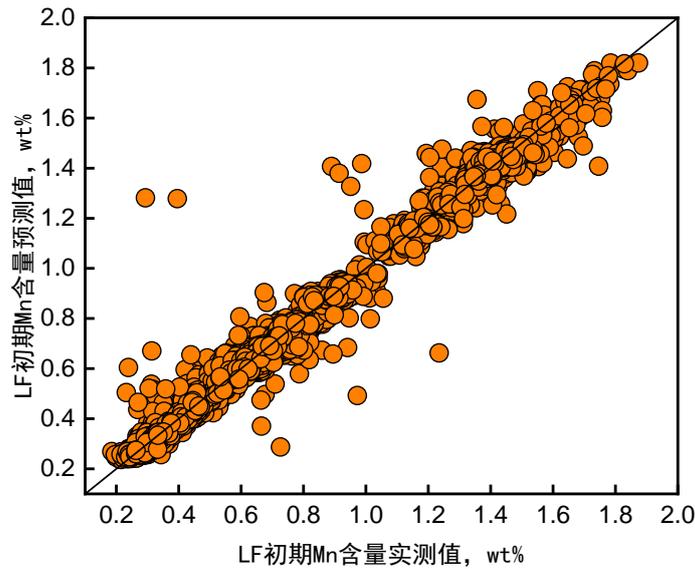


图 6 LF 精炼初期 Mn 含量实测值与预报值对比

5. 结论

以江苏永钢集团炼钢二分厂实际生产数据为依据, 运用 XGBoost 算法对 LF 精炼初期钢水成分 Al、Si、Mn 进行预报, 结果表明: LF 初期钢水中 Al、Si、Mn 预报值与实测值的 MAE (平均绝对误差) 分别达到 0.0023%、0.014%、0.023%, R^2 (相关系数) 分别为 0.68、0.99、0.99, 预报精度较高, 为 LF 初期钢水成分的精确控制奠定了基础。

参考文献

- [1] 李广帮,赵成林,赵素华,等.LF 精炼成分预报模型的开发[J].鞍钢技术,2009(04):26-29+34.
- [2] 王龙,冀秀梅,刘玠.人工智能在钢铁工业智能制造中的应用[J].钢铁,2021,56(04):1-8.
- [3] 蒋晋文,刘伟光.XGBoost 算法在制造业质量预测中的应用[J].智能计算机与应用,2017,7(06):58-60.
- [4] 李红利. 基于 XGBoost 混合模型的 LF 钢液温度预估研究[D].西安理工大学,2019.
- [5] Yw A, Ss A, Xc B, et al. Short-term load forecasting of industrial customers based on SVM and XGBoost-ScienceDirect[J]. International Journal of Electrical Power & Energy Systems, 129.
- [6] 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(06):1-8.
- [7] 陈恒志,杨建平,卢新春,等.基于极限学习机(ELM)的连铸坯质量预测[J].工程科学学报,2018,40(07):815-821.
- [8] 吕庆,刘月明,张振峰,等.基于承钢生产数据预测烧结矿 FeO 含量[J].钢铁研究学报,2018,30(12):957-962.
- [9] 柯凯, 彭其春, 彭霞林,等. 基于集成学习的转炉炼钢供氧量预测模型[J]. 炼钢, 2023, 39(1):6.
- [10] Friedman J H . Greedy Function Approximation: A Gradient Boosting Machine[J]. Annals of Statistics, 2001, 29(5):1189-1232.