

基于数据挖掘与清洗的高炉操作参数优化

刘 馨, 张卫军, 石 泉, 周 乐

(东北大学 冶金学院, 辽宁 沈阳 110819)

摘 要: 为了提高企业生产力, 实现“智慧钢厂”, 对企业的海量生产数据信息进行有效挖掘, 收集了某钢厂最近4年的高炉生产数据, 利用箱型图进行数据清洗, 提高数据质量。采取工艺理论和专家经验结合随机森林算法筛选出23个影响铁水质量和产量的特征参数。以铁水产量和铁水[Si+Ti]质量分数为目标参数, 通过 k -means聚类分析法将其分为3类。将分类结果与特征参数整合后进行分析, 得到造成铁水产量和质量大范围波动的13个参数, 同时提供了相应参数的合理控制范围。研究可对高炉稳定顺行以及数据挖掘在钢铁行业的应用提供指导。

关 键 词: 智慧钢厂; 数据挖掘; 特征工程; k -means 聚类; 随机森林; 高炉

中图分类号: TF 572 文献标志码: A 文章编号: 1005-3026(2020)08-1153-08

Operation Parameters Optimization of Blast Furnaces Based on Data Mining and Cleaning

LIU Xin, ZHANG Wei-jun, SHI Quan, ZHOU Le

(School of Metallurgy, Northeastern University, Shenyang 110819, China. Corresponding author: LIU Xin, E-mail: 1810574@stu.neu.edu.cn)

Abstract: Effective mining of mass production data helps to improve the productivity and the level of management to realize “smart steel works”. The production data for the last 4 years of one steel works were collected, and the box plot was used to clean the data to improve their quality. Twenty-three characteristic parameters affecting hot metal quality and yield were selected by technological theory and expertise combined with the random forest algorithm. Hot metal yield and Si + Ti content were taken as the objective parameters, which could be divided into three categories by k -means cluster analysis. Thirteen parameters contributing to the wide range fluctuation of hot metal yield and quality were obtained after the comprehensive analysis of classification results and characteristic parameters, and the reasonable variable ranges of corresponding parameters were provided. The research can guide the stable operation of blast furnaces and the application of data mining in the iron and steel industry.

Key words: smart steel works; data mining; feature engineering; k -means clustering; random forest; blast furnace

钢铁制造的许多生产环节周期较长, 目前部分高炉在正常运行(无中修)的情况下寿命可达10~15年^[1-2]。在此生产期间积累了庞大的数据, 诸如炉顶压力、炉顶温度、料重等数据, 但是由于数据处理技术平台的局限性, 海量的生产数据成为信息孤岛, 无法被整合并加以分析利用, 造成大量生产信息的浪费^[3-4]。通过数据挖掘可以充分发挥数据的信息价值, 解决“信息孤岛”的问

题, 加速钢铁生产走上信息化道路, 对于建立完整的管控平台, 实现产品质量管控与分析、能源优化、设备运行维护等功能具有重要的意义^[5-7]。

我国一些技术先进的钢铁企业已经开始了相关研究, 中冶京诚将大数据全流程分析系统应用在炼钢工序, 实现了钢的质量分析、在线预测与运行效果的判断^[8]。河钢承钢集团经过三年组建了信息流设备, 在2017年通过对热轧生产数据的管

收稿日期: 2019-09-23

基金项目: 国家“十三五”重点研发计划项目(2017YFA0700300); 国家自然科学基金资助项目(U1760115)。

作者简介: 刘 馨(1992-), 女, 河北张家口人, 东北大学博士研究生; 张卫军(1977-), 男, 辽宁沈阳人, 东北大学教授, 博士生导师。

理与应用,实现了产品质量的在线控制、诊断、评价与预报的功能^[9]. Brandenburger 等提出了一种多尺度的数据可视化表示方法,将大量生产数据与扁钢表面形态建立对应关系,建立了三层架构的质量监控系统并改进了扁钢的生产,将获得高分辨率可视数据报表的时间从几星期提高到几秒内^[10]. Han 等分析了唐山某钢厂1号高炉一年的生产数据,利用结构化次梯度投影的方法实现铁水中硅含量的动态预测,相比传统的串行算法,预测速度提高了54倍^[11]. Hua 等为解决高炉大数据中异常值对预测准确度的影响,提出了FCM-SVRs算法,对经过聚类剔除异常值的数据进行二次清洗,提高预测结果的准确性和工业数据的使用效率^[12].

由于高炉生产的复杂性,目前数据挖掘应用于炼铁方向的研究还比较少.吕庆等^[13-14]将数据挖掘技术应用于烧结生产,建立了烧结终点预报模型和FeO预报模型,但在使用大数据解决工厂实际问题时,注重数学算法分析的结果,没有结合专家经验,不能充分发挥模型的应用性.马富涛等^[15]结合钢铁行业需求,探究了大数据技术在高炉生产领域的应用前景.应用大数据技术的同时继承专家系统的优点,以特征学习为目的,能够更有效地挖掘数据内在的信息和规律.本文在运用数据挖掘技术的基础上融合了专家经验,对高炉生产参数和产品进行了全面分析和优化,对现场生产具有指导意义.研究所涉及的数据均来自某钢厂数据库中1#高炉的日均数据,由于生产环境不同,高炉数据的分析具有个体差异性.提供数据的炼铁厂拥有3座相同立式的高炉,并且拥有相同的烧结矿供应生产线.本文针对其中生产最稳定的1#高炉进行历史数据的挖掘分析,研究其稳定生产的内在规律,然后推广到另外两座相同立式的高炉指导生产.

1 高炉生产数据预处理

钢铁企业生产过程中积累了大量历史数据,采集过程中并未对数据进行合理的分类与规划.本文在考虑生产工艺要求的前提下,将高炉的生产数据进行分类,并厘清生产数据间的关系,对研究有着指导作用.高炉结构(几何形状、尺寸等)和操作参数(炉料配比、富氧量及风量等)的变化会影响过程参数(高炉内部传热、还原反应等),进一步影响高炉的生产指标(生铁产量、高炉寿命等).对于重点关注的指标参数,能直接接触并调控的是操作参数与结构参数,结构参数在生产

初期是已知参数,所以,研究重点是操作参数与指标参数两个变量间的关系.经过初步分类的数据中仍存在无效数据,即原始数据在存储过程中由于技术和管理问题,存在数据缺失、重复、异常等问题.数据存在质量问题不但无助于数据挖掘生成正确的决策,还可能产生负面影响^[16].因此原始数据库中的数据必须经过数据清洗、过滤处理,转换为高质量数据.

1.1 缺失数据处理

处理数据缺失问题时,若数据缺失比例过大,则该组数据不再适用数学算法进行补充,可以结合实际情况删除缺失率大于30%的参数.2015年3月—2016年5月期间,由于管理原因丢失了铁水温度数据,缺失率超过了1/3,考虑到铁水温度与生铁中硅含量呈负相关,决定不考虑铁水温度.

若参数缺失比例比较小,根据参数之间的相关关系填补缺失值,本文采用随机森林算法补充缺失值.如果参数缺失比例很小,则将缺失部分直接删除,不会对模型造成太大影响.

1.2 异常数据处理

异常数据是指明显偏离大多数样本数据的离群数值.箱型图法具有较好的耐抗性,不会破坏数据原始分布,在甄别异常数据方面表现出良好的性能,因此选用箱型图法识别与剔除异常数据.箱型图如图1所示,由五部分组成,分别是最小值、下四分位数(Q_1)、中位数(MD)、上四分位数(Q_3)和最大值.IQR表示四分位距,是 Q_1 与 Q_3 的间距.箱型图中以 $Q_1 - 1.5IQR$ 与 $Q_3 + 1.5IQR$ 作为温和异常点分界限(内限),以 $Q_1 - 3IQR$ 和 $Q_3 + 3IQR$ 作为极端异常点分界限(外限).外限以外的点成为异常数据的概率更大.由于高炉数据中温和异常点的数量过多,为了在保证数据质

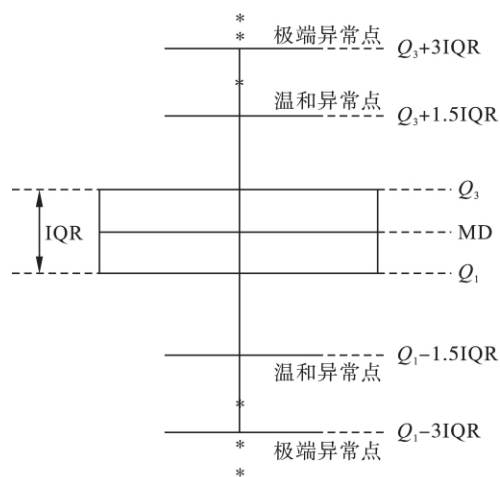


图1 箱线图
Fig. 1 Box plot

量的同时获得足够的样本数据,选择把极端异常点作为异常数据进行剔除。

数据预处理过程均在 Python 环境中进行,对某钢厂近 4 年的高炉生产数据进行清洗后,由 1 350 组数据得到 1 189 组数据。

2 高炉生产数据特征工程

在对高炉生产数据进行挖掘分析时,无关变量不仅无助于数据分析模型的辨识,还会干扰相关变量的作用,近而导致分析结果偏离实际情况;而遗弃一些相关性较大的变量,同样会导致模型很难反映系统特征,分析效果也会很差。因此,有效的特征参数选取非常重要,但单纯依靠数据挖掘算法有时也会将重要参数剔除。例如,在数据样本中,当某个重要参数一直被控制在合理范围时,由于波动很小,算法训练过后很有可能将其判定为无关参数而剔除,此类模型在应用时存在很大的潜在风险。对此,采取生产工艺理论经验与数据挖掘算法相结合,共同完成特征参数的选取。

2.1 基于机理经验筛选特征

高炉生产中影响铁水产量和质量的因素众多,汇总如下。

1) 风量对高炉生产的影响:在高炉全风操作以及漏风率稳定的情况下,送入高炉内部的风量与炉况行为相互作用。风量的波动,一定程度上预示着炉况波动。风量波动变化大,如风量下降较多时,炉况表现异常,且较难处理。在高炉日常生产中,风量降低 2%~4%,说明高炉生产已有潜在问题。

2) 风温对高炉生产的影响:高风温有助于提高炉缸温度,降低焦比。提高风温有助于增加热流比,降低吨铁热耗以及缩短软熔带位置高度,有利于减少 SiO_2 与焦炭和铁滴的接触时间和接触面积,近而抑制 $[\text{Si}]$ 的生成量。

3) 富氧对高炉生产的影响:富氧可以增加 CO 的生成量,不仅有助于强化高炉冶炼,增加高炉利用系数,而且缩短了 SiO_2 还原时间,有利于降低生铁中的硅含量。此外,富氧有利于改善高炉内部热量分布,有助于缓和低硅冶炼出现的上热下凉情况。

4) 炉顶温度和炉顶压力对高炉生产的影响:高炉热状态的变化影响下料状态。高炉采温点众多,其中炉顶温度波动能够间接反映炉顶煤气流分布,是高炉炉况进程的表现指标,直接反映了高炉内部的煤气利用率。高炉加压冶炼能够增加风量,延长煤气流在高炉内部的滞留时间,促进间接还原反应,有助于高炉顺行、降低焦比。除上述高

炉参数外,结合某钢厂生产经验收集到 20 个相关参数,但其中存在部分重复性的参数,包括富氧流量、富氧率、炉腹煤气量、炉腹煤气指数、透气性指数、阻力系数 K 值。其数据分布如图 2 所示。

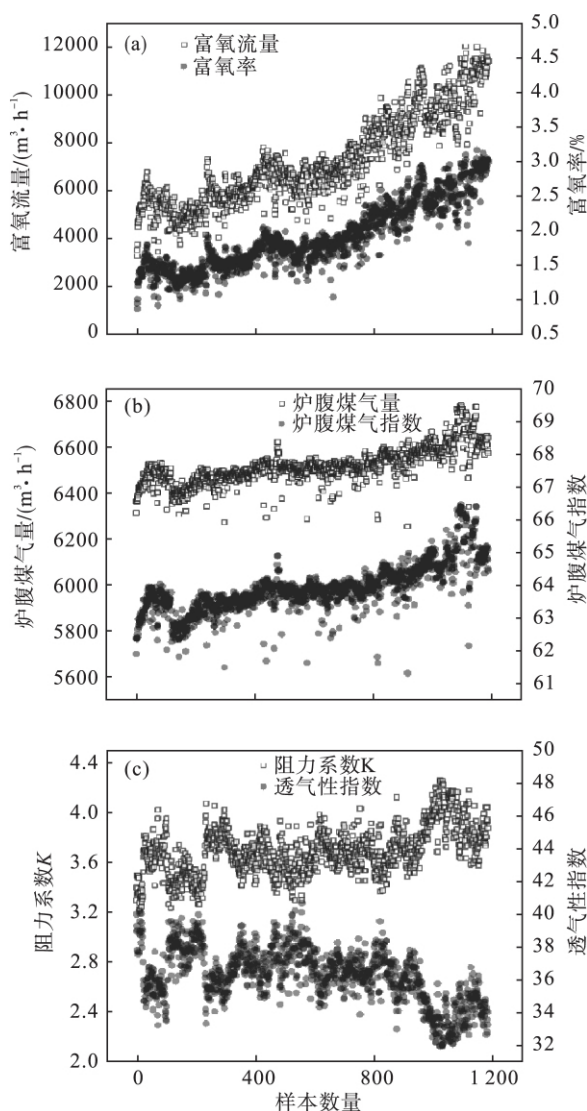


图2 共线性参数数据分布及相关性
Fig. 2 Data distribution and correlation analysis of redundant parameters

由图 2 可知,富氧流量与富氧率的分布规律基本一致(图 2a),且两者皮尔森相关性系数为 1,属于高度共线性参数,需要剔除其中一个;炉腹煤气量、炉腹煤气指数数据分布(图 2b)与富氧流量、富氧率情况相同;透气性指数与阻力系数 K 值数据分布相反(图 2c),皮尔森相关性系数为 -0.96 ,属于高度负相关,两者亦可相互替代。因此将富氧流量、炉腹煤气量、阻力系数 K 值剔除。最终保留风量、风压、风温、实际风速、透气性指数、鼓风动能、富氧率、全炉热负荷、炉顶压力、炉顶温度、上部压差、下部压差、焦比、煤比、综合燃比、炉腹煤气指数、高炉煤气利用率,共计 17 个参数。

2.2 基于随机森林算法筛选特征

生产现场有大量的数据采集点,但在生产中仅关注了其中一部分.针对这部分被忽略的数据,基于随机森林算法对其进行挖掘.

2.2.1 基于随机森林的特征选择方法介绍

特征选择的目标是找出与因变量相关度较高的特征变量.其一般步骤如下:

- ①对特征变量进行数据标准化;
- ②将随机森林算法中的特征变量按照贡献度 (feature importance) 进行降序排序;
- ③拟定删除比例,在当前的特征变量集中剔除贡献度不高的参数,得到新的特征集^[17].

2.2.2 特征参数筛选

1) 特征参数与目标参数的确定:除去上述 17 个参数,又汇总了与高炉生产可能相关的 19 个参数作为初始特征参数进行筛选,分别是:矿批量、焦批量、焦炭水分、入炉品位(铁矿石)、吨铁耗氧、机烧比、球比、杂矿比,以及高炉在 34、32、29、26、24、21、19、17、10、8、6 m 处的冷却壁 11 个测点温度.

选取产量指标实际铁水产量和质量指标铁水 [Si + Ti] 质量分数作为目标参数.质量指标选择 [Si + Ti] 质量分数有两个原因:一是由于某钢厂生产冶炼的含铁矿石以钕钛磁铁矿为主;二是因为选取的近 4 年数据中铁水温度数据缺失超过 30%,同时硅含量的变化与铁水温度呈负相关,能够反映铁水温度的变化,并且对清洗后的历史数据统计发现生铁中 [Si + Ti] 质量分数与 [Si] 质量分数的变化趋势基本一致. [Si + Ti] 质量分数与 [Si] 质量分数的变化趋势如图 3 所示.

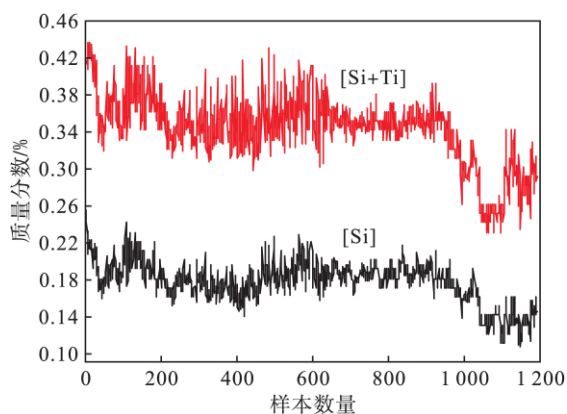


图3 铁水 [Si + Ti] 质量分数与 [Si] 质量分数的变化趋势比较

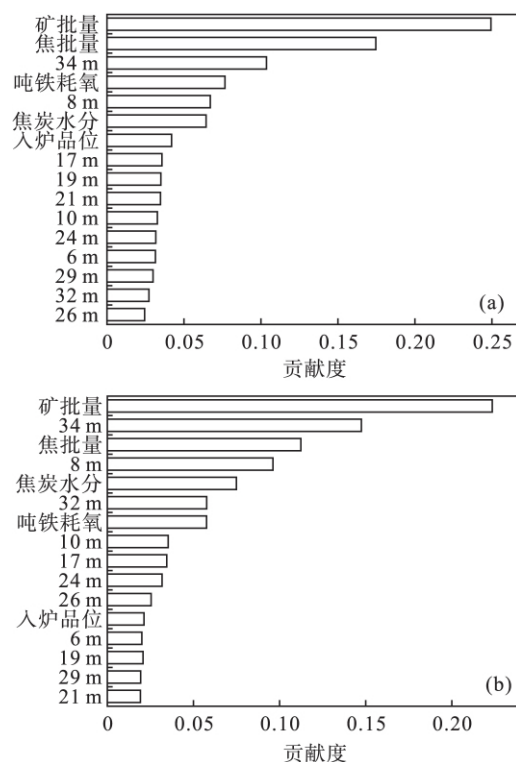
Fig. 3 Comparison of changes in [Si + Ti] content and [Si] content

2) 数据标准化处理:由于各参数的数据量级不同,量级大的参数对输出指标的贡献会掩盖量级小的参数,因此需要对数据进行标准化处理.本

文在 Python 环境下运用 sklearn 模块中的 preprocessing.scale() 函数对各参量进行数据标准化处理.

3) 贡献度排序:调用 Python 中 sklearn 模块中的 ExtraTreesRegressor() 函数建立特征参数与目标参数之间的关系,根据特征参数的贡献度进行排序,结果如图 4 所示.

由图 4 可以看出矿批量、焦批量、吨铁耗氧、焦炭水分,以及 34 m 和 8 m 处冷却壁测点温度对铁水产量和质量均有较明显的影响,因此将这 6 个参数选入特征参数的范畴.



图中纵坐标数字(如 34 m)表示此处冷却壁测点温度参数

图4 相关性排序

Fig. 4 Relevance ranking

- (a) 一特征参数与铁水产量的相关性排序;
(b) 一特征参数与 [Si + Ti] 质量分数的相关性排序.

将 2.1 节筛选的 17 个参数和 2.2.2 节筛选的 6 个参数,总计 23 个参数分类为操作参数、过程参数以及指标参数,各类参数与权重系数见表 1.各参数权重是由 xgboost 计算结果结合实际生产情况调整得到的综合结果.

3 高炉数据挖掘分析

通过工艺理论和专家经验并结合随机森林算法综合筛选出了 23 个影响高炉生产的核心参数.运用数据挖掘技术进一步分析这些参数对铁水产量和质量的影响形式.

表 1 参数分类与权重系数
Table 1 Parameter classification and weight coefficients

操作参数	权重	过程参数	权重	指标参数	权重
富氧率	0.07	炉顶温度	0.06	焦比	0.07
风量	0.07	炉顶压力	0.04	煤比	0.06
鼓风动能	0.05	上部压差	0.03	全炉热负荷	0.05
透气性指数	0.04	下部压差	0.03	燃料比	0.05
焦批量	0.04	34 m 冷却壁	0.03	高炉煤气利用率	0.05
矿批量	0.04	8 m 冷却壁	0.03	炉腹煤气指数	0.04
风温	0.04			吨铁耗氧	0.04
风压	0.03				
风速	0.02				
焦炭水分	0.02				

3.1 k-means 聚类分析

为了更准确地分析影响铁水产量和质量参数,首先对铁水产量和质量定义一个评价标准,即对其进行分类,然后结合 *k-means* 聚类函数进一步验证分类的合理性。*k-means* 函数是一种成熟的聚类算法,通过选取合理的 *k* 值(聚类中心),将数据进行分类。在分类之前需要先分析铁水产量

和 [Si + Ti] 质量分数的数据分布,确定合理的 *k* 值,然后通过 *k-means* 聚类函数对其进行分类,观察聚类效果并验证分类的合理性。

在分类之前对铁水产量和质量数据分布进行分析。以某钢厂 2 500 m³ 高炉近 4 年铁水产量和 [Si + Ti] 质量分数的日均数据(1 200 组数据)为样本,数据标准化后,二者数据分布如图 5 所示。

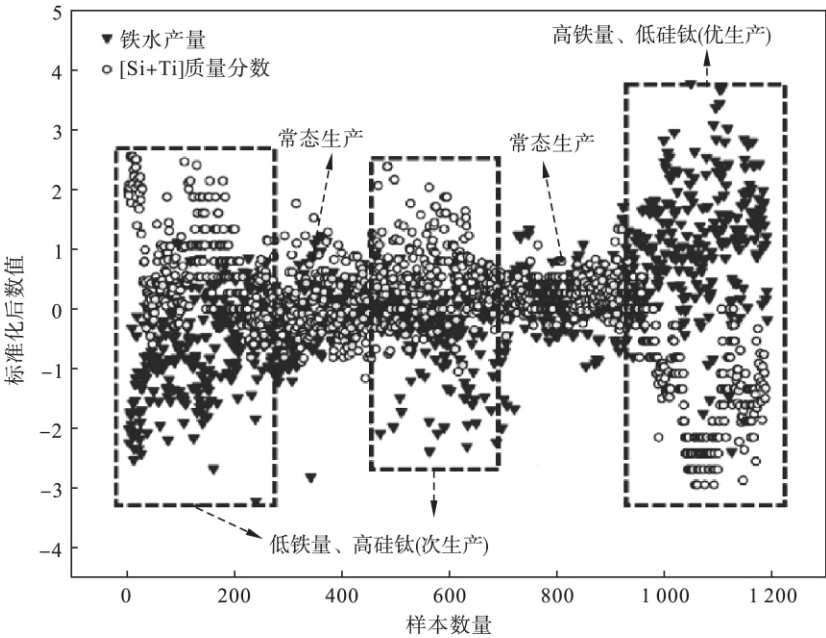


图 5 铁水产量和 [Si + Ti] 质量分数的变化趋势
Fig. 5 Change trend of hot metal yield and [Si + Ti] content

由图 5 可以看出,铁水产量与 [Si + Ti] 质量分数呈现一定的负相关性,即铁水产量高,则硅含量低,表现为前 200 组数据和第 600 组数据附近高 [Si + Ti] 质量分数对应低铁水产量;第 1 000 组数据以后低 [Si + Ti] 质量分数对应高铁水产量。数据分布规律与钢厂控制 [Si + Ti] 质量分数的同时提高铁水产量的要求相符。从数据分布结果确

定 *k* = 3,并将铁水产量高、[Si + Ti] 质量分数低的情况定义为优生产;将铁水产量低、[Si + Ti] 质量分数高的情况定义为次生产;铁水产量、[Si + Ti] 质量分数稳定的情况定义为常态生产。因此根据数据分析结果将铁水产量和质量初步划分为优生产、常态生产和次生产三个等级。

选取 3 个中心点,对标准化后的铁水产量和

[Si + Ti]质量分数数据进行聚类分析. 聚类结果如图6所示, 聚类的效果显著, 各类数据之间分界明显, 未出现数据相互掺杂的情况, 其中 class-1 数据代表优生产, 符合铁水产量高、[Si + Ti]质量分数低的规律, 该组样本量占总样本量的 17%; class-3 数据代表次生产, 符合铁水产量低、[Si + Ti]质量分数高的规律, 该组样本量占总样本量的 18%; class-2 数据代表常态生产, 该组样本量占总样本量的 65%. 聚类结果进一步验证了将铁水产量和 [Si + Ti]质量分数划分为三类的合理性. class-1 与 class-2 数据占比之和在 80% 以上, 代表整体生产水平良好; 但次生产数据占比超过了优生产, 接近 20%, 说明生产中炉况波动影响较大, 需要对核心参数与聚类结果间的关系作进一步研究.

3.2 聚类结果分析

根据聚类结果, 将铁水产量、[Si + Ti]质量分数综合为一个目标参数, 用类别 class-1, class-2, class-3 替换. 通过时间戳整合核心参数数据与各类别中的数据. 为了更清楚地观察核心参数对聚类结果的影响, 选择 class-1 和 class-3 条件下的数据(随机 200 组)进行对比, 观察核心参数波动与数据分类结果之间的关系. 利用反推法分析不同类别条件下各核心参数的分布规律.

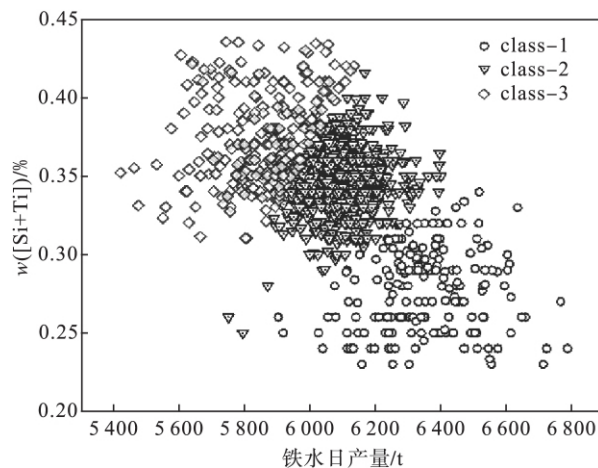
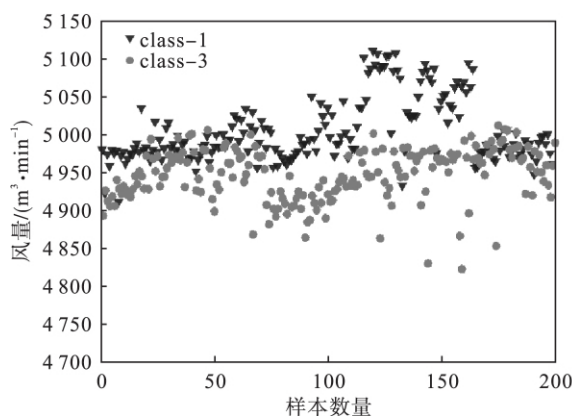


图6 铁水产量和 [Si + Ti]含量聚类结果

Fig. 6 Clustering results of hot metal yield and [Si + Ti] content

3.2.1 风量和风压的变化

如图7所示: class-1 风量和风压数据分布均明显高于 class-3, 其中 class-1 风量主要分布在 $4980 \sim 5100 \text{ m}^3/\text{min}$; class-3 风量主要分布在 $4900 \sim 5000 \text{ m}^3/\text{min}$. class-1 热风压力在 $370 \sim 400 \text{ kPa}$ 间波动, 主要集中在 380 kPa 以上; class-3 热风压力基本上分布在 380 kPa 以下. 由分析结果可知, 适当提高冷风流量和热风压力有利于提高铁水的产量和质量, 建议将风量调控在 $5000 \sim 5100 \text{ m}^3/\text{min}$, 风压调控在 $380 \sim 400 \text{ kPa}$.

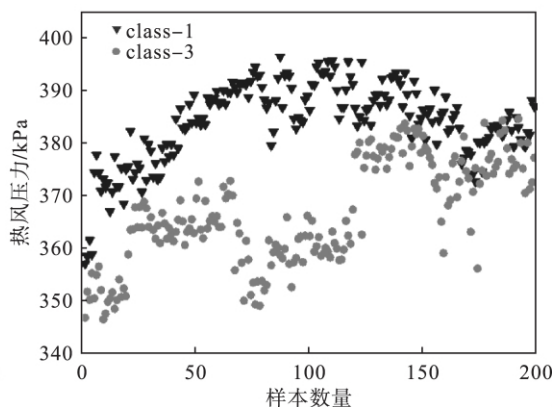


图7 冷风流量和热风压力统计分析

Fig. 7 Statistical analysis of cold air flow and hot air pressure

3.2.2 焦比和煤比的变化

由图8可知: class-1 焦比数据分布普遍低于 class-3, 主要分布在 $320 \sim 360 \text{ kg/t}$, 同时煤比数据分布普遍高于 class-3, 主要分布在 $140 \sim 170 \text{ kg/t}$; class-3 焦比和煤比分别集中在 $360 \sim 380 \text{ kg/t}$ 和 $120 \sim 140 \text{ kg/t}$. 统计结果表明, 适当提高煤比不仅能够降低焦比还有利于提高铁水的产量和质量, 因此建议将焦比控制在 $320 \sim 360 \text{ kg/t}$ 的同时煤比控制在 $140 \sim 170 \text{ kg/t}$.

3.2.3 焦批量和矿批量的变化

如图9所示: class-1 焦批量明显低于 class-3, 主要分布在 $12.5 \sim 14.5 \text{ t}$, 但矿批量主要分布在 $71 \sim 75 \text{ t}$; class-3 焦批量集中在 $14 \sim 16.7 \text{ t}$, 矿批量主要分布在 $69 \sim 71 \text{ t}$. 统计结果表明, 适当增加矿批量、减少焦批量, 有利于提高铁水的产量和质量, 建议焦批量调控范围为 $12.5 \sim 14.5 \text{ t}$, 矿批量调控范围为 $71 \sim 75 \text{ t}$.

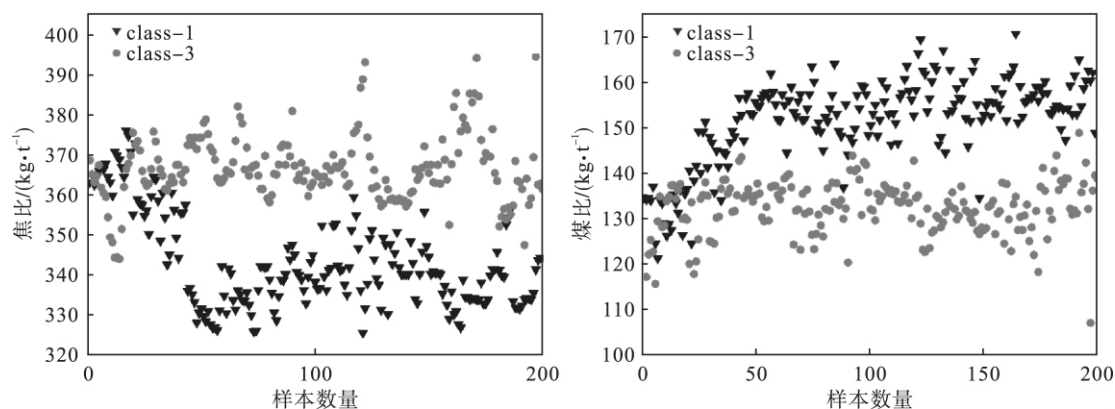


图 8 焦比和煤比数据统计分析

Fig. 8 Statistical analysis of coke ratio and coal ratio data

3.2.4 富氧率和炉腹煤气指数的变化

由图 10 可知: class-1 情况下富氧率和炉腹煤气指数均高于 class-3, 富氧率集中在 2.2% ~ 3.2%, 炉腹煤气指数集中在 63.5 ~ 66.5; class-3 中富氧率集中在 1.1% ~ 2%, 炉腹煤气指数集中在 62 ~ 64. 统计结果表明, 适当提高富氧率和炉腹煤气指数有利于提高铁水的产量和质量, 建议将富氧率控制在 2.2% ~ 3.2%, 炉腹煤气指数控制在 64 ~ 66.5.

除了上述 8 个参数外, 上部压差、吨铁耗氧、透气性指数、炉顶压力、34 m 处测点温度这 5 个参数在优生产和次生产条件下也存在比较明显的差异. 结合数据统计分析结果建议上部压差控制在 19 ~ 23 kPa, 吨铁耗氧控制在 280 ~ 292 m³/t, 透气性指数控制在 32 ~ 35.5, 炉顶压力控制在 238 ~ 241 kPa, 34 m 处测点温度控制在 75 ~ 105 °C. 其余 10 个核心参数差异并不明显, 可以维持现状, 代表日常生产中已经对其实现了合理控制.

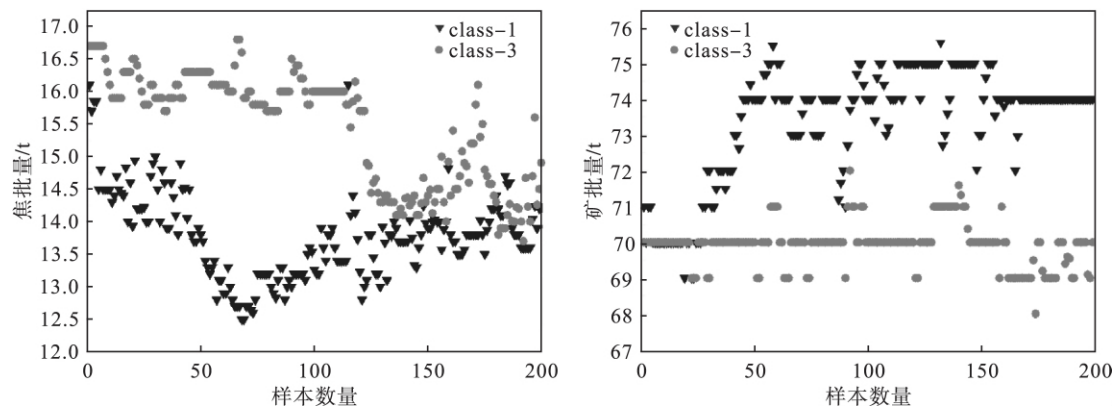


图 9 焦批量和矿批量数据统计分析

Fig. 9 Statistical analysis of the amount of coke batch and ore batch

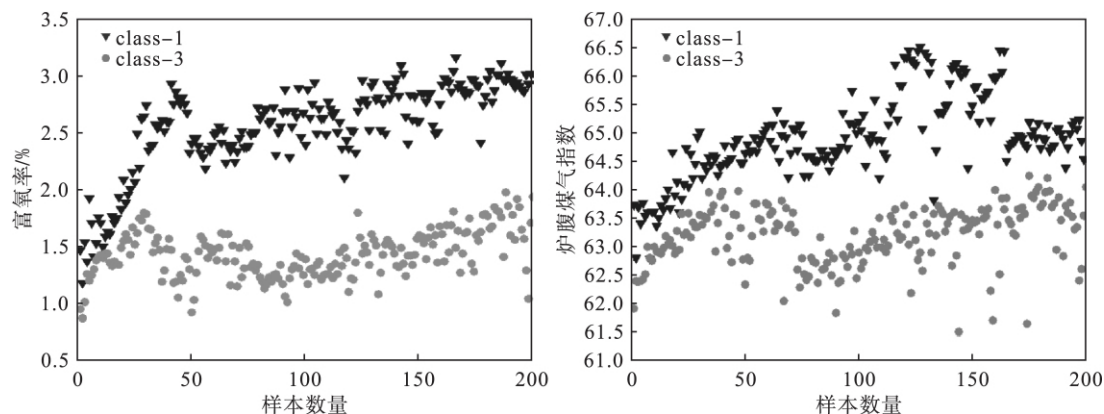


图 10 富氧率和炉腹煤气指数数据统计分析

Fig. 10 Statistical analysis of oxygen enrichment rate and belly gas index data

4 结 论

1) 采用工艺理论和专家经验并结合随机森林算法,全面地筛选出影响铁水产量和质量的23个核心参数,并将高炉生产参数划分为操作参数、过程参数和指标参数,为高炉操作参数分析优化建立了数据基础。

2) 通过 k -means 聚类算法和统计分析将目标参数(铁水产量和[Si+Ti]质量分数)划分为优生产、常态生产、次生产3个等级。运用统计分析针对不同分类结果,筛选出了13个波动差异明显的高炉参数,并得出最优类 class-1 炉况下对应的核心参数的参考取值范围,对高炉稳定顺行具有指导意义,为调整和优化高炉操作参数提供了新方向。

3) 建议风量控制在 $5\,000 \sim 5\,100 \text{ m}^3/\text{min}$,风压控制在 $380 \sim 400 \text{ kPa}$,焦比控制在 $320 \sim 360 \text{ kg/t}$,同时,将煤比控制在 $140 \sim 170 \text{ kg/t}$,焦批质量控制在 $12.5 \sim 14.5 \text{ t}$,矿批质量控制在 $71 \sim 75 \text{ t}$,富氧率控制在 $2.2\% \sim 3.2\%$,炉腹煤气指数控制在 $64 \sim 66.5$,上部压差控制在 $19 \sim 23 \text{ kPa}$,吨铁耗氧控制在 $280 \sim 292 \text{ m}^3/\text{t}$,透气性指数控制在 $32 \sim 35.5$,炉顶压力控制在 $238 \sim 241 \text{ kPa}$,34 m 处壁面温度控制在 $75 \sim 105 \text{ }^\circ\text{C}$ 。

参考文献:

- [1] 张寿荣,于仲洁. 中国炼铁技术60年的发展[J]. 钢铁, 2014, 49(7): 8-14.
(Zhang Shou-rong, Yu Zhong-jie. Development of ironmaking technology in the past 60 years [J]. *Iron and Steel*, 2014, 49(7): 8-14.)
- [2] 周渝生,曹传根,甘菲芳. 高炉长寿技术的最新进展[J]. 钢铁, 2003, 38(11): 70-74.
(Zhou Yu-sheng, Cao Chuan-gen, Gan Fei-fang. Recent development of BF long-campaign technology [J]. *Iron and Steel*, 2003, 38(11): 70-74.)
- [3] 徐雪松,杨胜杰. 大数据背景下中国钢铁生产能源管控路径优化研究[J]. 工业技术经济, 2017, 36(1): 32-40.
(Xu Xue-song, Yang Sheng-jie. The optimization of iron and steel enterprises energy management under the background of big data [J]. *Journal of Industrial Technological Economics*, 2017, 36(1): 32-40.)
- [4] 王钧超. 大数据时代产业经济信息分析及在宏观决策中的应用[D]. 北京: 中国地质大学, 2016.
(Wang Jun-chao. Analysis of industrial economic information and its application in macro decision in the era of big data. [D]. Beijing: China University of Geosciences, 2016.)
- [5] Krumeich J, Werth D, Loos P, et al. Advanced planning and control of manufacturing processes in steel industry through big data analytics: case study and architecture proposal[C/OL]. // IEEE International Conference on Big Data. Washington DC, 2014 [2019-08-25]. <https://ieeexplore.ieee.org/document/7004408>.
- [6] Guo S, Yu J X, Liu X J, et al. A predicting model for properties of steel using the industrial big data based on machine learning [J]. *Computational Materials Science*, 2019, 160: 95-104.
- [7] Kang L, Du H L, Zhang H, et al. Systematic research on the application of steel slag resources under the background of big data [J/OL]. *Complexity*, 2018 [2019-08-05]. <https://www.hindawi.com/journals/complexity/2018/6703908/>.
- [8] 中冶京诚全流程大数据分析系统探索智慧钢铁[N]. 世界金属导报, 2018-11-27.
(The whole process of big data analysis system to explore the wisdom of steel—CERI [N]. *World Metal Report*, 2018-11-27.)
- [9] 白瑞国,徐立山,包阔,等. 大数据过程质量控制系统在钢铁生产中的应用[J]. 中国冶金, 2018, 28(8): 76-80.
(Bai Rui-guo, Xu Li-shan, Bao Kuo, et al. Application of big data process quality control system in iron and steel production [J]. *China Metallurgy*, 2018, 28(8): 76-80.)
- [10] Brandenburger J, Colla V, Nastasi G, et al. Big data solution for quality monitoring and improvement on flat steel production [J]. *IFAC: Papers On Line*, 2016, 49(20): 55-60.
- [11] Han Y, Li J, Yang X L, et al. Dynamic prediction research of silicon content in hot metal driven by big data in blast furnace smelting process under Hadoop cloud platform [J/OL]. *Complexity*, 2018 [2019-08-25]. <https://doi.org/10.1155/2018/8079697>.
- [12] Hua C, Wu J, Li J, et al. Silicon content prediction and industrial analysis on blast furnace using support vector regression combined with clustering algorithms [J]. *Neural Computing and Applications*, 2017, 28(12): 4111-4121.
- [13] 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.
(Lyu qing, Liu Yue-ming, Zhang Zhen-feng, et al. Prediction of FeO content in sinter based on production data of Chengde Steel Mill [J]. *Journal of Iron and Steel Research*, 2018, 30(12): 957-962.)
- [14] 吕庆,刘颂,刘小杰,等. 基于大数据技术的烧结全产线质量智能控制系统[J]. 钢铁, 2018, 53(7): 1-9.
(Lyu qing, Liu song, Liu Xiao-jie, et al. Intelligent control system based on big data technology for whole production line of sintering quality [J]. *Iron and Steel*, 2018, 53(7): 1-9.)
- [15] 马富涛,张建良,张磊,等. 铁前数模技术进展与大数据应用探讨[J]. 钢铁, 2018, 53(12): 1-9.
(Ma Fu-tao, Zhang Jian-liang, Zhang lei, et al. Introduction of development and progress of mathematical modeling technology in iron-making area and discussion on application prospects of big data technology [J]. *Iron and Steel*, 2018, 53(12): 1-9.)
- [16] Dixon B E, Duke J, Grannis S. Measuring and improving the quality of data used for syndromic surveillance [J]. *Online Journal of Public Health Informatics*, 2017, 9(1): 182-189.
- [17] Genuer R, Poggi J M, Malot C T. Variable selection using random forests [J]. *Pattern Recognition Letters*, 2010, 31(14): 2225-2236.