

基于粒度聚类的转炉炼钢氧气消耗量预测

阳青锋^{1,2,3} 赖旭芝^{1,2,3} 杜胜^{1,2,3} 胡杰^{1,2,3} 陈略峰^{1,2,3} 吴敏^{1,2,3}

摘要 转炉炼钢是钢铁企业的主要耗氧工序, 预测转炉炼钢的氧气消耗量对氧气系统合理调度、保证生产安全具有重要意义. 考虑到转炉冶炼工况多、钢种数据粒度不统一, 提出一种基于粒度聚类的转炉炼钢氧气消耗量预测方法. 首先, 利用孤立森林异常检测方法剔除历史数据库中的异常数据; 接着, 采用皮尔逊相关性分析和互信息系数选取相关影响因子, 对不同钢种数据进行信息粒化, 实现数据特征提取和维度统一, 使用模糊 C 均值 (Fuzzy C-means, FCM) 划分工况并建立不同工况下的氧气消耗量预测子模型; 最后, 利用企业的实际生产数据进行实验, 验证所提方法的准确性和有效性.

关键词 转炉炼钢, 氧气消耗预测, 信息粒化, 工况识别

引用格式 阳青锋, 赖旭芝, 杜胜, 胡杰, 陈略峰, 吴敏. 基于粒度聚类的转炉炼钢氧气消耗量预测. 自动化学报, 2024, 50(1): 132-142

DOI 10.16383/j.aas.c230333

Converter Steelmaking Oxygen Consumption Prediction Based on Granularity Clustering

YANG Qing-Feng^{1,2,3} LAI Xu-Zhi^{1,2,3} DU Sheng^{1,2,3} HU Jie^{1,2,3} CHEN Lue-Feng^{1,2,3} WU Min^{1,2,3}

Abstract Oxygen consumption prediction in converter steelmaking is of great significance for the rational scheduling of the oxygen system and ensuring production safety in steel enterprise. Considering the diverse operating conditions of converter smelting and the inconsistent granularity of steel grade data, this paper proposes a prediction method for oxygen consumption in converter steelmaking based on granularity clustering. Firstly, the isolation forest anomaly detection method is used to remove abnormal data from the historical database. Then, Pearson correlation analysis and mutual information correlation coefficient are employed to select relevant influencing factors and achieve information granulation for different steel grade data, thereby extracting data features and unifying dimensions. Fuzzy C-means (FCM) clustering is utilized to divide the operating conditions and establish oxygen consumption prediction sub-models for different conditions. Finally, the accuracy and effectiveness of the proposed method are validated through experiments using actual production data from the steel enterprise.

Key words Converter steelmaking, oxygen consumption prediction, information granulation, operating condition recognition

Citation Yang Qing-Feng, Lai Xu-Zhi, Du Sheng, Hu Jie, Chen Lue-Feng, Wu Min. Converter steelmaking oxygen consumption prediction based on granularity clustering. *Acta Automatica Sinica*, 2024, 50(1): 132-142

转炉炼钢是钢铁企业的主要耗氧工序, 炼钢过程具有典型的周期性和间歇性特点, 炼钢在短时间内氧气需求激增, 与制氧机只能稳定制氧之间存在

生产矛盾, 难以保证氧气供需平衡. 预测转炉炼钢过程的氧气消耗, 可给予调度人员充足时间去调度决策, 保证氧气系统动态平衡.

针对转炉炼钢过程氧气消耗预测问题, 有学者从机理角度进行了研究. 文献 [1] 分析转炉炼钢过程的主要化学反应, 建立了转炉炼钢过程的氧气消耗预测模型, 该模型对认识转炉炼钢过程、实现氧气消耗在线预测具有积极作用. 但缺少对转炉炼钢过程中不同因子间的耦合性分析, 铁水中的化学物质种类繁多, 冶炼过程伴随众多的物理化学反应、复杂的相变过程以及各种不确定因素, 机理模型无法准确计算炼钢过程氧气的消耗量^[2]. 受现场炉体参数和检测精度的影响, 该方法不具备良好的普适性.

工业过程海量数据存储与计算能力的提升为基于数据驱动的预测方法提供了重要基础. 这类方法

收稿日期 2023-06-08 录用日期 2023-09-08

Manuscript received June 8, 2023; accepted September 8, 2023
高等学校学科创新引智计划资助项目 (B17040), 国家自然科学基金 (62303431), 湖北省自然科学基金 (2015CFA010, 2021CFB145, 2022CFB582), 中国博士后科学基金 (2023M733306) 资助

Supported by the 111 Project (B17040), National Natural Science Foundation of China (62303431), Natural Science Foundation of Hubei Province (2015CFA010, 2021CFB145, 2022CFB582), and China Postdoctoral Science Foundation (2023M733306)

本文责任编辑 杨涛

Recommended by Associate Editor YANG Tao

1. 中国地质大学 (武汉) 自动化学院 武汉 430074 2. 复杂系统先进控制与智能化湖北省重点实验室 武汉 430074 3. 地球探测智能化技术教育部工程研究中心 武汉 430074

1. School of Automation, China University of Geosciences, Wuhan 430074 2. Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074 3. Engineering Research Center of Intelligent Technology for Geoproduction, Ministry of Education, Wuhan 430074

利用数理统计知识和人工智能方法来挖掘数据特征, 不需要建立完全刻画机理过程的数学模型, 如多元线性回归^[3]、自回归滑动平均^[4]、长短记忆网络^[5]、支持向量机^[6]、T-S 模糊模型^[7]、回声状态网络^[8] 等方法. 上述研究可分为时序预测和回归预测两大类. 时序预测往往通过频谱分析、时序分解等技术提取历史数据特征, 利用时序网络挖掘数据变化规律, 达到预测的目的, 值得注意的是, 时序预测非常依赖历史数据的质量, 预测之前往往要对原始数据进行去噪处理. 回归预测将机理知识或专家经验知识与数据相融合, 采用主成分分析、皮尔逊相关性分析、互信息等方法提取主要特征因子, 采用回归网络挖掘工业过程的非线性关系进行预测.

上述方法在工业过程取得了良好应用, 但是这些方法直接用于转炉炼钢的氧气消耗量预测还存在一些问题. 首先, 在转炉高温、高压、高噪音的环境下, 传感器难以高精度在线测量甚至会采集一些异常数据, 数据驱动模型非常依赖数据的质量, 这些模型缺少数据的异常检测; 其次, 各类钢种的化学物质含量存在差异, 上述模型是针对单一工况的建模, 从钢种的角度来说, 转炉炼钢存在多工况的特点.

单一工况无法有效反映工业过程多工况的特性, 有许多学者对工业过程多工况预测进行了研究并取得了一定成果. 文献 [9] 针对烧结过程多工况特性, 提出了一种基于差分进化算法的多级预测模型, 预测烧结过程的综合焦比. 文献 [10] 提出一种基于时间序列数据分析的波动区间工况识别, 判断烧结过程波动工况. 文献 [11] 针对淬火过程建立了多工况板形预报模型. 上述研究侧重于单个样本数据特征的聚类, 而本文需从钢种角度进行聚类, 即不同钢种聚类成多个工况. 在实际生产中, 不同钢种的冶炼炉数与公司的订单需求有关, 导致不同钢种采集到的历史数据样本长度不同. 因此, 本文在识别转炉工况前需要解决不同钢种粒度不统一的问题, 本文提到的钢种数据粒度不统一是指不同钢种的样本数据长度不同.

针对上述问题, 本文提出一种基于粒度聚类的转炉炼钢氧气消耗量预测方法, 利用孤立森林对实际生产过程数据进行预处理, 剔除异常数据; 接着, 采用信息粒化处理不同钢种数据量不平衡的问题, 在此基础上进行模糊 C 均值 (Fuzzy C-means, FCM) 聚类; 然后针对不同的工况, 分别建立基于支持向量回归 (Support vector regression, SVR) 的工况子模型; 最后, 利用实际生产数据对本文所提方法进行实验, 验证本文所提方法的准确性和有效性.

本文的创新性主要包括以下两个方面: 首先是设计了转炉炼钢氧气消耗量预测方案, 该方案从工况的角度, 采用孤立森林、相关性分析、粒度聚类、支持向量回归进行转炉炼钢氧气消耗量预测; 其次是采用信息粒化解决了钢种在不同数据粒度下进行工况识别的问题.

1 转炉过程描述与方案设计

本节从转炉炼钢生产工艺过程出发, 对转炉炼钢氧气消耗机理进行详细分析, 指出转炉炼钢氧气消耗预测的难点, 并设计转炉炼钢氧气消耗量的预测方案.

1.1 转炉炼钢过程描述与分析

转炉炼钢的基本原理是在转炉中加入铁水、废钢、石灰石等原料, 在高温环境下通入高纯度氧气, 氧化去除铁水中的碳、磷、硫等物质, 同时利用氧化反应释放的热量加热炉体, 达到炼钢所需的温度. 转炉炼钢生产主要包括倒废钢、兑铁水、加热、脱硫、脱碳、调和、出钢等过程, 其工艺流程如图 1 所示.

高炉生产的铁水, 用铁牛罐车运输到转炉, 铁水装入转炉炉体后倒入废钢. 铁水在运输过程中会损耗大量热量, 在冶炼前需吹入氧气进行预加热, 提高铁水温度; 然后加入石灰石去除铁水中的含硫物质, 通过氧气枪向转炉内吹入高纯度氧气, 氧化铁水中的碳、磷、硫等化学物质, 分解出 CO 与 CO₂ 等气体; 接着加入合金元素, 调节钢水的化学成分, 达到不同钢种化学物质含量要求; 最后将钢水倒出并清理炉渣, 钢水运输到连铸、热轧、冷轧等工序进行后续加工.

高炉出铁口和转炉现场的环境十分恶劣, 铁水温度高达 1600 °C, 碳、磷、硫等参数实时在线监测困难. 本文研究的钢铁企业转炉过程参数通过以下方式获取: 1) 铁水经过出铁口时会进行取样, 检测铁水中各类物质含量; 2) 铁水与废钢在倒入转炉前进行称重. 本文的检测参数是以炉次为单位进行检测的, 即每一炉钢水在冶炼前检测. 每一炉次的时间跨度是 30 分钟左右, 其中冶炼时长在 18 分钟左右, 等待铁水 12 分钟左右. 深冷法制氧机的调节速度缓慢, 无法实现分钟级快速响应, 因此以炉次进行氧气消耗预测对氧气调度具有参考意义.

在整个吹炼过程中, 转炉炼钢的氧气消耗量与众多变量密切相关. 从物理化学角度来说, 首先, 冶炼的铁水、废钢以及加入的石灰石越多, 需要去除的杂质越多, 对氧气的需求量也越大, 因此铁水重量、废钢重量、石灰重量与氧气消耗可能存在较强

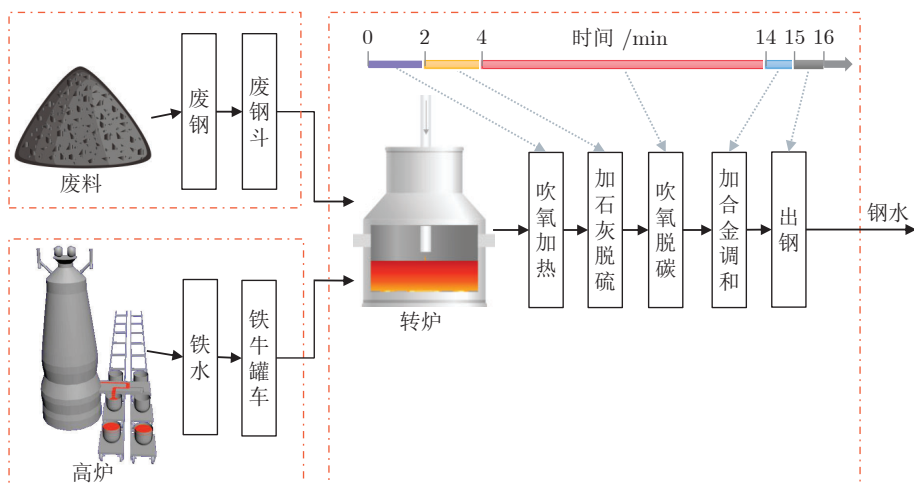


图 1 转炉冶炼过程

Fig.1 Converter steelmaking process

的线性相关性. 其次, 氧气与铁水接触时会与铁水中的化学物质发生各种化学反应, 包括氧化、还原、熔化、渗碳等, 在不同的阶段, 化学反应的程度不相同, 反应机理也存在差异, 在初期主要与碳反应, 在后期则主要与硅、锰等化学物质反应, 因此这类变量与氧气的消耗量存在严重的非线性关系. 此外, 上述变量之间并不是相互独立的, 还存在耦合作用, 例如, 不同类型废钢的残余元素成分不同. 铁水与废钢的比例对氧气吹入量有影响, 该因素也会直接影响化学反应进程.

除上述因素外, 还存在许多变量难以检测的问题. 不同类别的钢种其最终的化学成分不同, 也影响着转炉炼钢过程的氧气消耗量. 上述各类原因导致转炉炼钢的氧气消耗量难以预测, 有必要对转炉炼钢从冶炼钢种角度来划分工况, 对氧气消耗量进行分工况预测.

综合上述分析并结合转炉炼钢氧气消耗过程特点, 本文需考虑以下问题:

- 1) 现场环境恶劣, 采集到的数据存在异常值, 数据质量难以保证;
- 2) 转炉冶炼过程工况复杂, 参数变量多, 参数之间存在严重非线性;
- 3) 不同钢种间的数据粒度不统一.

本文的难点首先是工业过程数据维度过高, 难以对高维数据进行无监督异常检测. 其次是钢种粒度不统一时, 无法进行钢种聚类, 当前的工况识别算法和以往的研究都是对多条数据进行聚类, 每条数据维度为 $1 \times N$ 维, 而本文不同类型钢种的数据维度为 $m_i \times N$, 钢种不同, m_i 的值也不同, 因此无法以钢种为单位进行工况识别. 最后就是进行工况识别后, 转炉过程的数据被划分为多个工况, 模型

训练的数据量减小, 如何利用小样本数据进行高精度预测.

1.2 方案设计

考虑到转炉炼钢的氧气消耗量与转炉工况息息相关, 且不同钢种的数据粒度不统一. 结合转炉炼钢过程特点, 本文提出了一种基于粒度聚类的转炉炼钢氧气消耗预测方法, 其结构图如图 2 所示.

本文方法首先从转炉过程的历史数据库中获取转炉炼钢过程的相关数据, 工业过程的异常数据难以发现, 采用孤立森林对异常数据进行剔除. 然后, 针对检测参数众多的问题, 利用皮尔逊相关性分析和互信息法计算检测参数和计划参数与氧气消耗量的相关性, 选取与转炉炼钢耗氧具有强相关性的参数, 降低数据冗余对工况识别的影响. 接着, 按钢种对样本数据进行分类, 考虑到不同数据粒度的钢种无法聚类且转炉炼钢过程无统一标准划分转炉工况, 采用信息粒化提取不同钢种样本的数据特征并统一数据的维度, 对粒化后的数据进行模糊 C 均值聚类. 最后, 对不同类别的钢种数据, 分别建立不同工况的子模型. 在测试阶段, 将采集到的过程数据, 根据钢种编号分别输入训练好的工况子模型, 达到预测转炉炼钢过程氧气消耗量的目的.

2 转炉炼钢耗氧预测模型

本节主要介绍基于粒度聚类的转炉炼钢氧气消耗量预测模型. 为减小异常数据对预测模型精度的影响, 利用孤立森林去除转炉炼钢生产过程的异常数据; 针对工况复杂、数据粒度不统一的问题, 采用信息粒化和模糊 C 均值聚类划分不同工况类型, 建立不同工况下的转炉氧气消耗量预测子模型.

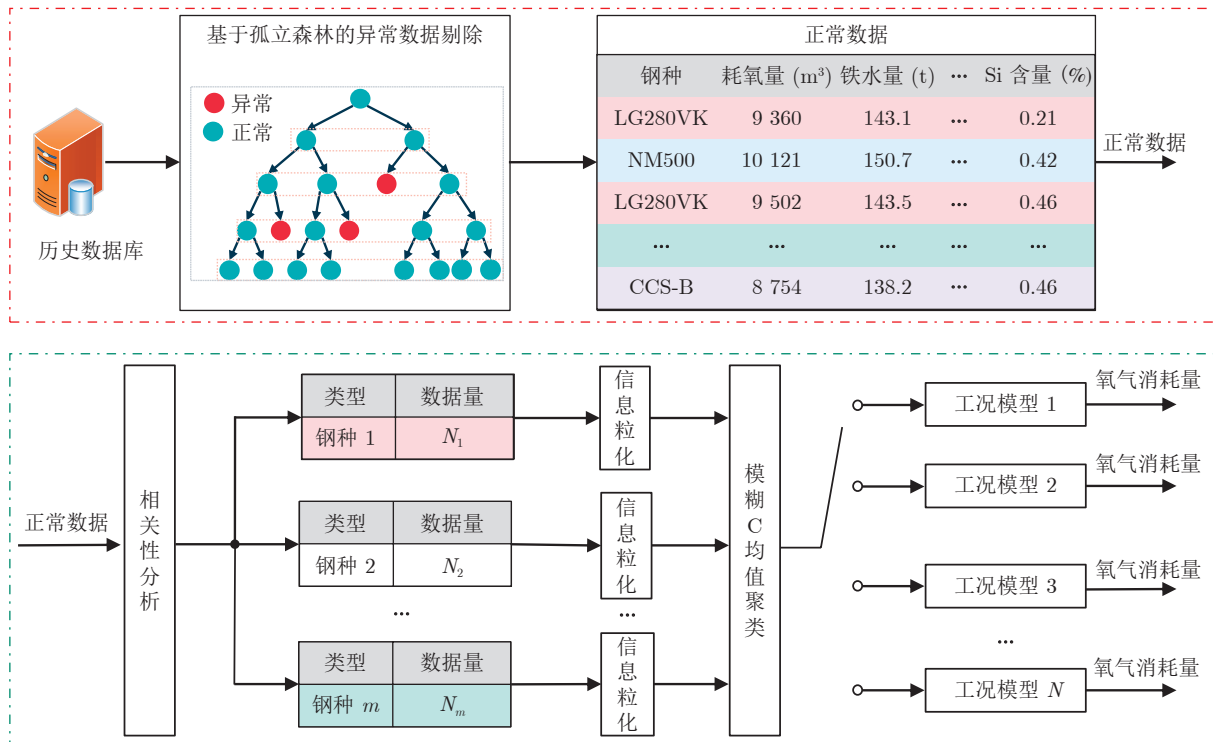


图 2 基于粒度聚类的氧气消耗量预测方案

Fig.2 Oxygen consumption prediction scheme based on granularity clustering

2.1 基于孤立森林的异常数据剔除

钢铁冶炼现场环境恶劣, 历史生产数据存在一些异常值. 这些异常数据虽然较少, 且与正常样本之间存在明显的差别, 若不加以剔除的话会直接影响相关参数的选取, 不利于神经网络对转炉炼钢氧气消耗量非线性模型的拟合, 因此有必要对异常数据进行剔除.

孤立森林采用一套高效的孤立策略来构造二叉树, 使异常样本少且与正常样本属性差异明显的样本能率先被分离出来^[12]. 该方法是一种无监督异常检测方法, 构建树的过程不涉及密度、距离等复杂运算, 计算复杂度低. 孤立森林不依赖数据的分布假设, 而基于概率分布的方法如高斯混合模型^[13]等方法, 需要正常数据满足正态分布, 这对于转炉炼钢过程的数据要求过于苛刻. 基于统计的 3σ 、Grubbs 等方法简单高效, 但其鲁棒性较差^[14]. 本文对多个转炉炼钢过程特征参数进行异常检测, 在高维度中, 数据点的密度分布会变稀疏, 基于密度的 DBSCAN 方法^[15] 处理高维数据存在维度灾难, 并不适用本文的异常检测. 因此, 本文选择孤立森林进行异常数据检测.

孤立森林的关键在于孤立树的构建, 图 3 为转炉过程数据样本的孤立树, 树节点的子节点采用随

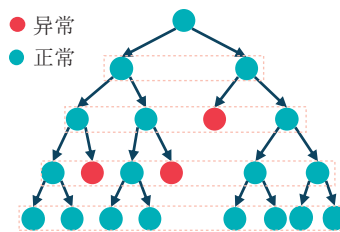


图 3 孤立树结构

Fig.3 Isolation tree structure

机采样的方式生成, 构建孤立树的步骤如下:

- 步骤 1. 选择转炉炼钢过程数据中的一个特征向量 X_i .
- 步骤 2. 随机选择该数据集中的一个值 x_n .
- 步骤 3. 把特征中小于 x_n 的数据放入左节点, 大于 x_n 的数据放入右节点.
- 步骤 4. 若二叉树的高度达到上限或所有子节点只有一条数据, 树的构建结束. 否则判断所有子节点, 若子节点只有一条数据则该节点结束, 否则该节点跳转步骤 2 继续构建新节点.

数据样本距离根节点越近, 该样本为异常样本的概率越高. 但是在构建孤立树的过程中, 转炉数据的特征向量和节点分割值随机选取, 仅使用一棵孤立树来确定样本的异常不具备鲁棒性, 因此需要

不断构造新的孤立树, 将多个孤立树集成为孤立森林, 来增加判别的可靠性以及对局部相对稀疏点的覆盖率. 构建多棵孤立树会增加计算负担, 考虑到数据密度较小的样本大概率会被率先孤立出来, 当孤立树的高度达到一定程度时, 剩下的数据所处位置密度较高, 这些数据是异常值的概率较低. 本文设置孤立树的最大高度为 100 来降低计算负担, 且孤立树的构建是相互独立的, 可以采用并行计算技术来加快孤立森林的构建速度. 为定量说明样本异常程度, 定义样本 x 异常得分函数 $S(x)$

$$S(x) = -\frac{1}{m} \sum_{i=1}^m (d_i(x) + c(n)) \quad (1)$$

其中, m 为孤立树的数量, $m \geq 100$ 时异常得分趋于稳定, n 为样本数量, $d_i(x)$ 为样本 x 距离第 i 棵树的根节点的路径长度, $c(n)$ 为孤立树的平均高度

$$c(n) = \begin{cases} 2(\ln(n-1) + \gamma) - \frac{2(n-1)}{n}, & n > 2 \\ 1, & n = 2 \\ 0, & n < 2 \end{cases} \quad (2)$$

其中, γ 为欧拉常数, $\gamma = 0.5772156649$.

通过上述方式, 可以得到样本的异常得分, 异常得分越高, 表示样本越可能为异常值.

转炉过程涉及的检测参数多达 20 个, 主要包括操作参数和成分参数, 操作参数受人为控制可直接调节, 如铁水重量 C_1 、废钢重量 C_2 、石灰重量 C_3 . 成分参数为铁水中各种化学元素的含量, 如 Si

含量 C_{Si} 、Mn 含量 C_{Mn} 、P 含量 C_P 等. 转炉过程的检测参数属于典型的高维数据, 随着维数增加, 数据空间大小会呈指数级增长, 导致转炉过程数据分布稀疏. 考虑到所有成分参数都是同时进行化验检测的, 在进行异常数据检测时可选择一个成分参数作为代表, 本文选择铁水硅含量作为铁水中化学元素含量的代表. 综上所述, 本文选择氧气消耗量、铁水重量、废钢重量、石灰重量以及铁水硅含量这 5 个特征参数进行异常值检测. 本文的符号含义如表 1 所示.

2.2 基于信息粒化的钢种聚类

根据转炉炼钢过程描述和特性分析可知, 转炉炼钢的氧气消耗量与检测参数之间存在一定的相关性. 其中检测参数主要指铁水中各种化学物质的含量, 计划参数则主要指冶炼的铁水量、废钢量、钢种等生产计划, 这些参数可能存在大量的冗余数据, 具体参数如表 2 所示. 为定量描述各类检测参数与氧气消耗量之间的相关性, 本文采用皮尔逊相关系数 ρ_{pec} ^[11] 和互信息相关系数 ρ_{NMI} ^[16] 来计算它们之间的相关性, 并选取强相关参数因子减小数据的冗余. 其中, 皮尔逊相关系数可描述过程参数与转炉炼钢氧气消耗量之间的线性关系, 归一化互信息相关系数则可描述它们之间的非线性关系.

不同钢种的化学物质含量不同, 从钢种角度来说, 冶炼不同钢种也可看作不同的生产工况. 对转炉炼钢过程的历史数据按冶炼的钢种进行归纳, 钢

表 1 符号说明
Table 1 Symbol description

符号	含义	符号	含义
C_1	铁水重量	C_2	废钢重量
C_3	石灰重量	C_{Si}	铁水中化学元素 Si 的含量
C_{Mn}	铁水中化学元素 Mn 的含量	C_P	铁水中化学元素 P 的含量
C_S	铁水中化学元素 S 的含量	C_{Cu}	铁水中化学元素 Cu 的含量
C_{As}	铁水中化学元素 As 的含量	C_{Sn}	铁水中化学元素 Sn 的含量
C_{Ti}	铁水中化学元素 Ti 的含量	C_V	铁水中化学元素 V 的含量
C_{Pb}	铁水中化学元素 Pb 的含量	C_{Zn}	铁水中化学元素 Zn 的含量
C_{Cr}	铁水中化学元素 Cr 的含量	C_{Ni}	铁水中化学元素 Ni 的含量
C_{Nb}	铁水中化学元素 Nb 的含量	C_{Mo}	铁水中化学元素 Mo 的含量
C_{Sb}	铁水中化学元素 Sb 的含量	C_W	铁水中化学元素 W 的含量
Y	氧气消耗量	$S(X)$	样本 X 的异常得分
ρ_{pec}	皮尔逊相关系数值	ρ_{NMI}	归一化互信息相关系数值
X_A	A 钢种	\tilde{X}_A	A 钢种信息粒
r_1	均方根误差 (Root mean square error, RMSE)	r_2	平均绝对误差 (Mean absolute error, MAE)
r_3	平均绝对百分比误差 (Mean absolute percentage error, MAPE)	r_4	最大绝对百分比误差 (Maximum absolute percentage error, MaxAPE)
D_b	Davies-Bouldin 指数		

表 2 相关参数与吹氧量的 ρ_{pcc} 和 ρ_{NMI} 值
Table 2 ρ_{pcc} and ρ_{NMI} values between relevant parameters and oxygen blowing amount

过程参数	C_1	C_2	C_3	C_{Si}	C_{Mn}
ρ_{pcc}	0.281	0.510	0.340	0.054	0.180
ρ_{NMI}	0.776	0.820	0.831	0.864	0.856
过程参数	C_{P}	C_{S}	C_{Cu}	C_{As}	C_{Sn}
ρ_{pcc}	0.016	0.117	0.120	0.102	0.012
ρ_{NMI}	0.758	0.812	0.766	0.757	0.710
过程参数	C_{Ti}	C_{V}	C_{Pb}	C_{Zn}	C_{Cr}
ρ_{pcc}	0.025	0.102	0.053	0.015	0.022
ρ_{NMI}	0.851	0.841	0.542	0.554	0.770
过程参数	C_{Ni}	C_{Nb}	C_{Mo}	C_{Sb}	C_{W}
ρ_{pcc}	0.019	0.019	0.046	0.036	0.030
ρ_{NMI}	0.718	0.535	0.472	0.314	0.609

种的类型多达几十种. 若对每一种钢种单独建立转炉炼钢的氧气消耗预测模型, 会使整个模型非常复杂, 并且小样本的钢种无法有效描述钢种冶炼过程的非线性关系. 为了解决上述问题, 需要对转炉炼钢过程的历史数据以钢种为单位进行聚类. 不同钢种的数据量存在差别, 在聚类之前要统一不同钢种的数据粒度. 如图 4 所示, 采用信息粒化与模糊 C 均值聚类相结合的方法来解决上述问题, 不同颜色代表不同钢种, 每个钢种的过程参数维度相同, 但数据长度不同. 先对每个钢种的过程参数进行信息粒化, 提取该钢种的主要数据特征并统一所有钢种

的数据维度, 接着采用模糊 C 均值算法对钢种进行分类, 将相似的钢种进行聚类.

假设 A 钢种数据为 $X_A = [x_A^1, x_A^2, \dots, x_A^k]$, 其中, k 为特征维数, $x_A^t = [x_1^t, x_2^t, \dots, x_n^t]^T$, 其中 n 为数据长度, $1 \leq t \leq k$. 对每个特征进行区间信息粒化得到 \tilde{x}_A^t , \tilde{x}_A^t 表示 A 钢种第 t 个特征的信息粒, 即 A 钢种数据信息粒化后为 $\tilde{X}_A = [\tilde{x}_A^1, \tilde{x}_A^2, \dots, \tilde{x}_A^k]$, 其中, $\tilde{x}_A^t = [x_l^t, x_m^t, x_u^t]^T$. 为有效保留原始数据的数据特征, 需构建合理的信息粒 \tilde{x}_A^t . 合理的信息粒应满足合理性和特殊性两个指标, 合理性指信息粒应尽可能多地覆盖原始数据点的个数, 特殊性

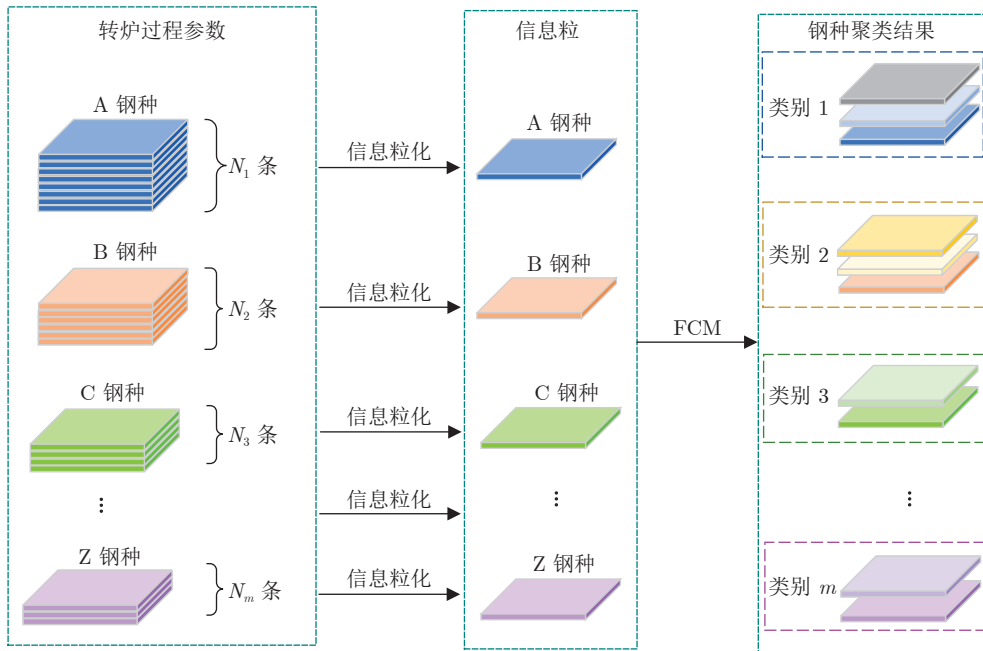


图 4 基于信息粒化的钢种模糊聚类

Fig. 4 Fuzzy clustering of steel types based on information granulation

则要求信息粒尽可能的精细^[17]. 合理性与特殊性之间相互制约, 定义乘积函数 $f = f_1 \times f_2$, 同时考虑合理性和特殊性两个指标^[18]

$$f_1 = \frac{1}{n} \text{card}(x_A^t | x_l^t \leq x_A^t \leq x_u^t) \quad (3)$$

$$f_2 = 1 - \frac{|x_u^t - x_l^t|}{\max(x_A^t) - \min(x_A^t)} \quad (4)$$

其中, $\text{card}(x_A^t | x_l^t \leq x_A^t \leq x_u^t)$ 表示 x_A^t 在 \hat{x}_A^t 范围内的个数, n 为总数, x_l^t 为信息粒的下限值, x_u^t 为信息粒的上限值, x_m^t 一般取 x_A^t 的中位数. f_1 代表信息粒的合理性, f_2 代表信息粒的特殊性, 信息粒的优化目标为函数 f 取最大值, 本文采用粒子群算法进行求解.

采用信息粒化, 可将转炉过程数据按钢种转化为 $\tilde{X} = [\tilde{X}_A, \tilde{X}_B, \dots, \tilde{X}_Z]^T$, 实现不同钢种数据维度的统一. 使用聚类算法将相似钢种看作同一类别, 可缓解小样本钢种无法拟合非线性模型的不足, 提高后续转炉炼钢氧气消耗预测模型的精度. 本文采用模糊 C 均值算法对粒化后的钢种数据进行聚类, 该算法通过求各个钢种到不同类别的欧氏距离和的最小值, 得到每种钢对所有类别中心的隶属度, 实现钢种聚类的目标^[19]. FCM 聚类算法的目标函数为

$$\min J = \sum_{i=1}^C \sum_{j=1}^n u_{ij}^m \|X_j - c_i\|^2 \quad (5)$$

其中, C 为钢种类别数, n 为样本数, m 为加权系数. c_i 为不同钢种类别的聚类中心, X_j 表示信息粒, u_{ij} 表示信息粒 X_j 属于 c_i 类的隶属度, u_{ij} 需满足约束条件

$$\sum_{i=1}^C u_{ij} = 1, u_{ij} \in [0, 1] \quad (6)$$

引入拉格朗日系数法对上述约束方程求解, 使目标函数 J 取极小值的必要条件为

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|X_j - c_i\|}{\|X_j - c_k\|} \right)^{\frac{2}{m-1}}} \quad (7)$$

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m X_j}{\sum_{j=1}^n u_{ij}^m} \quad (8)$$

模糊 C 均值算法通过迭代计算 c_i , u_{ij} 与 J , 直到目标函数值小于给定阈值或 u_{ij} 趋于稳定, 算法停止. 聚类算法完成后, 可计算钢种属于不同类别的隶属度值, 进而识别钢种所属的类别. 对粒化后

的钢种进行聚类, 得到的类别就是以钢种进行分类, 聚类中心是钢种的中心. 在线应用时, 根据需要冶炼的钢种名称即可判断所属的工况模型.

模糊 C 均值算法是无监督聚类, 需指定聚类数目以完成对钢种的聚类. 本文引入 Davies-Bouldin 指数 (Davies-Bouldin index, DBI) 来评价聚类效果^[20], 指数 D_b 值越小, 类别之间的相似度越低, 聚类效果越好. D_b 值的定义如下

$$D_b = \frac{1}{C} \sum_{i,j=1}^C \max_{i \neq j} \left(\frac{S_i + S_j}{d_{ij}} \right) \quad (9)$$

其中, S_i 表示属于 i 类的钢种到类中心距离的平均值, d_{ij} 表示 i 类与 j 类聚类中心间的距离.

2.3 基于支持向量回归的预测模型

综合考虑转炉炼钢过程氧气消耗量影响因子, 使用历史数据来建立氧气消耗量的预测模型. 转炉炼钢过程的数据按钢种聚类后, 按工况建立多个预测模型, 相较于建立单个预测模型, 历史数据被划分为多个工况, 训练模型的数据量减小. 支持向量回归模型通过高维映射方法来解决数据在低维空间难以非线性拟合的问题, 该模型在小样本建模上具有优秀的表现^[21], 本文采用支持向量回归来建立转炉炼钢的氧气消耗预测模型.

假设转炉炼钢氧气消耗量的相关影响因子为 $X = [X_1, X_2, \dots, X_k]^T$, 转炉炼钢氧气消耗量为 $Y = [y_1, y_2, \dots, y_n]$, 其中, k 为数据维度, $X_t = [x_1, x_2, \dots, x_n]$, n 为数据长度. 构建支持向量回归模型 $Y = wX + b$, 寻找超平面 w 和偏置 b 使模型的输入参数与模型输出之间具有最小拟合误差, 让离超平面最远的样本点距离最小. 转化为如下数学关系式

$$\begin{aligned} \min & \left(\frac{1}{2} \|w\|^2 + g \sum_{i=1}^n \lambda_i \right) \\ \text{s.t.} & \begin{cases} \|Y - wX - b\| \leq \lambda_i \\ \lambda_i \geq 0 \end{cases} \end{aligned} \quad (10)$$

其中, λ_i 为松弛变量, 在 λ_i 范围内的样本不计算损失, g 为 λ_i 的惩罚因子. 在此之前, 需使用核函数将输入数据映射到高维空间, 本文选用高斯核函数

$$K(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \quad (11)$$

其中, σ 为高斯核函数参数, x_i 和 x_j 代表转炉炼钢氧气消耗量数据的特征向量. 引入拉格朗日乘子 α 和 α' 对其对偶问题进行求解, 得到氧气消耗量预测函数 $f(x)$

$$f(x) = \sum_{i=1}^n (\alpha - \alpha') K(x_i, x_j) + b \quad (12)$$

3 仿真与结果分析

为验证本文所提方法的有效性, 本文使用我国某大型联合钢铁企业的 210 t 转炉生产过程 2023 年 1 月至 3 月的数据, 共收集和使用 708 组样本数据进行仿真实验. 其中 588 组数据作为训练集, 120 组数据用于测试.

对 588 组训练数据进行孤立森林异常检测, 检测到异常数据 45 组并将异常值剔除. 在图 5 中, 列举了其中 4 个检测参数使用孤立森林剔除异常值前后的箱线图, 其中绿色箱体为原始数据, 紫色箱体为去除异常值后的数据, 红色点为离群数据. 通过对比可以看出, 使用孤立森林后, 4 种参数的离群值数量大大减少, 剔除后的箱体上下限和箱线图的上、下须长度与原始数据差异很小, 说明孤立森林可以有效剔除转炉生产过程的异常数据并且不会改变转炉过程的训练数据分布.

利用剔除异常数据后的 543 组数据计算皮尔逊相关系数值和互信息相关系数值, 计算结果如表 2 所示. 根据表中的 ρ_{pcc} 值可知, 相较于其他过程参数而言, 转炉炼钢的氧气消耗量与铁水重量、废钢重量、石灰重量具有更强的线性相关性, 这符合转炉炼钢氧气消耗过程分析的物理化学特点. 废钢重量和石灰重量的 ρ_{NMI} 值超过了 0.8, 证明这两个变量与氧气消耗量具有强相关性, 铁水重量的 ρ_{NMI} 值为 0.776, 也与氧气消耗量具有较强相关性. 由于在进行相关性分析之前, 剔除了训练数据中的异常数据, 绝大部分参数与氧气消耗量的 ρ_{NMI} 值超过了 0.5, 为降低数据的冗余, 本文选择 ρ_{NMI} 超过 0.8 的过程参数作为氧气消耗的主要影响因子. 综合上述分析, 本文选择 C_1 (铁水重量)、 C_2 (废钢重量)、 C_3 (石灰重量)、 C_{Si} (Si 含量)、 C_{Mn} (Mn 含量)、 C_{S} (S 含量)、 C_{Ti} (Ti 含量)、 C_{V} (V 含量) 这 8 个变量作为转炉炼

钢过程氧气消耗量的影响因子.

剩余的 543 组训练数据中, 共有 31 类钢种数据, 首先对数据按钢种进行信息粒化, 根据粒化后的数据进行模糊 C 均值聚类, 计算不同聚类数目时的 D_b 值, 如图 6 所示. 聚类数目为 4 时, 取得最小的 D_b 值 0.471, 将钢种数据聚类为 4 类, 本文需建立 4 个不同工况的转炉炼钢氧气消耗预测子模型.

为对比验证本文所提基于粒度聚类的转炉炼钢氧气消耗预测方法的有效性, 分别与 Elman 网络、极限学习机 (Extreme learning machine, ELM) 以及支持向量回归方法等进行对比实验. 为比较各模型预测效果的优劣, 采用均方根误差 r_1 、平均绝对误差 r_2 、平均绝对百分比误差 r_3 、最大绝对百分比误差 r_4 作为评价指标, 评价不同模型的预测结果, 其计算公式如下所示

$$r_1 = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (13)$$

$$r_2 = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (14)$$

$$r_3 = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \quad (15)$$

$$r_4 = \max_{1 \leq i \leq N} \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \quad (16)$$

其中, N 为预测样本数, y_i 为预测值, \hat{y}_i 为实际值. r_1, r_2, r_3, r_4 的值越小, 表示模型的预测精度越高, 性能越好.

不同模型的预测误差如表 3 所示, 从表中可以看出, SVR 模型的性能优于 Elman 网络、ELM 网络和随机森林^[22] (Random forest, RF). 相较于单一的 Elman、ELM、SVR、RF 预测模型, 对应的基于粒度聚类的组合预测模型 IG-FCM-Elman、IG-FCM-ELM、IG-FCM-SVR (本文提出的方法)、IG-

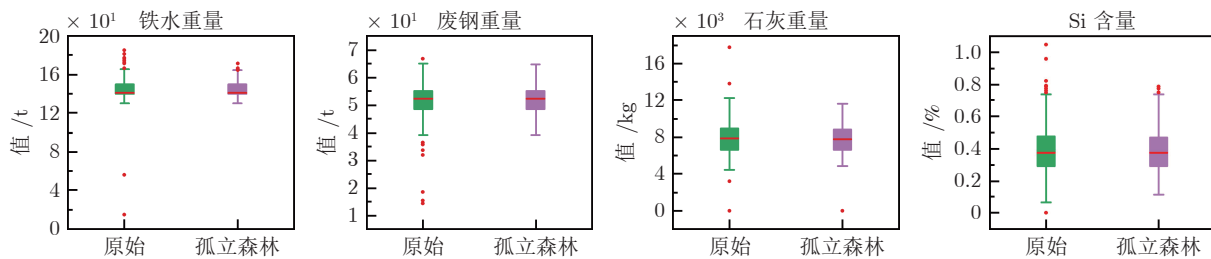
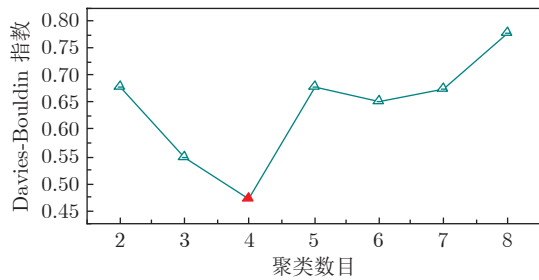


图 5 剔除异常值前后箱线图对比

Fig.5 Comparison of box plots before and after removing outliers

图 6 不同聚类数目的 D_b 值Fig. 6 D_b values for different numbers of clusters表 3 不同模型预测误差
Table 3 Prediction errors of different models

预测模型	r_1	r_2	r_3	r_4
Elman	575.60	465.41	4.85	0.19
IG-FCM-Elman	565.11	415.35	4.35	0.24
ELM	480.12	359.38	3.74	0.17
IG-FCM-ELM	471.48	352.71	3.68	0.24
RF	417.17	324.69	3.36	0.13
IG-FCM-RF	400.04	313.11	3.24	0.12
SVR	415.40	321.54	3.33	0.16
IG-FCM-SVR	407.64	304.74	3.16	0.14

FCM-RF 的 RMSE、MAE、MAPE 更小, 组合模型预测精度更具优势, 说明基于信息粒化的聚类能有效划分转炉冶炼的工况. IG-FCM-Elman 和 IG-FCM-ELM 的 MaxAPE 有所升高, 可能是模型分工况后样本数量减少、模型出现过拟合且模型预测远离历史数据分布的边界数据产生不稳定结果导致的; IG-FCM-SVR 的 MaxAPE 相较于单一的 SVR 有所降低, 说明了 SVR 模型的高斯函数内核以及在小样本建模上的良好性能. 本文所提 IG-FCM-SVR 方法的均方根误差、平均绝对误差、平均绝对百分比误差、最大绝对百分比误差分别为 407.64、304.74、3.16、0.14, 表明该模型能有效挖掘转炉炼钢过程氧气消耗的非线性特征, 本文所提方法能基

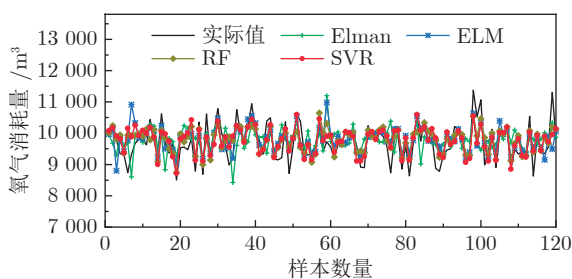


图 7 基于单一模型的氧气消耗量预测结果

Fig. 7 Oxygen consumption prediction results based on a single model

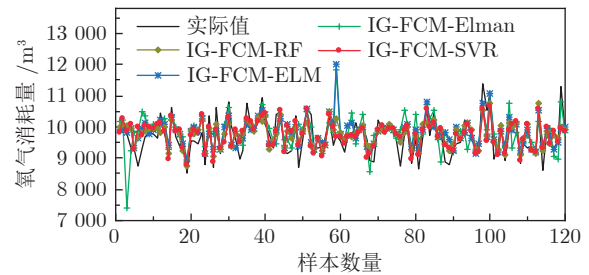


图 8 基于不同组合模型的氧气消耗量预测结果

Fig. 8 Oxygen consumption prediction results based on different combination models

本满足工业现场应用需求.

为直观展示不同模型的预测效果, 绘制不同方法的转炉炼钢氧气消耗量预测结果与实际检测值的曲线图. 图 7 为基于单一模型的氧气消耗量预测结果图, 在单一预测模型中, 支持向量回归预测模型的预测值受转炉炼钢氧气消耗相关因子的数值波动影响较小, 预测结果不会大幅度波动, 能更好地跟踪实际的氧气消耗量. 图 8 为基于不同组合模型的氧气消耗量预测结果图, 在组合模型中, ELM 和 Elman 的组合模型分别在第 59 个样本和第 3 个样本预测误差明显增大, 这是由于 ELM 的输出权重采用线性回归的方式进行求解, 可能存在奇异解, 无法有效应对生产过程的异常数值. Elman 网络是具有短期记忆功能的动态网络, 更适合时序预测, 而本文的预测是回归预测, 因此 Elman 网络在预测氧气消耗总量的表现上逊色于 SVR 网络.

结合图 7 和图 8 可知, 支持向量回归模型预测的可靠性更强. 基于信息粒化后的模糊 C 均值聚类结合支持向量回归的组合方法相对于其他方法拟合效果更好. 本文的数据异常值处理是针对训练集数据进行的, 而在实际检测过程中难免会出现少量无法识别的异常数据, 支持向量回归引入了松弛变量, 增加了对异常数据的敏感性, 因此, 采用支持向量回归模型更适合工业现场的应用.

4 总结

本文针对转炉炼钢氧气消耗量难以预测、不同钢种冶炼数据量不统一的问题, 提出了一种基于粒度聚类的转炉炼钢氧气消耗预测方法. 该方法采用信息粒化的方法统一不同钢种的数据维度, 在此基础上针对不同工况进行转炉炼钢氧气消耗量预测. 仿真结果表明, 本文所提方法对转炉炼钢氧气消耗预测的精度更高, 可为钢铁企业调节氧气系统平衡提供科学的参考信息. 下一步将参考转炉炼钢氧气消耗预测的相关成果, 继续研究钢铁企业氧气系统

平衡调度策略, 并将其用于钢铁企业能源管控系统中, 保证氧气系统动态平衡。

References

- Li Hua, Lei Hao. Analysis and prediction model of related affecting factors of oxygen consumption in converter steel-making. *Southern Metals*, 2022, **246**(3): 37–42
(李桦, 雷豪. 转炉炼钢冶炼氧气消耗相关影响因素分析及预测模型. 南方金属, 2022, **246**(3): 37–42)
- Zhang Jia-Yan, Ma Zhong-Hai, Qian Xiao-Bin, Li Shao-Ming, Lang Jia-Hong. Application of optimal control strategy to converter gas recovery system. *Acta Automatica Sinica*, 2012, **38**(6): 1017–1024
(章家岩, 马中海, 钱晓斌, 李绍铭, 郎佳红. 转炉煤气回收系统优化控制策略应用. 自动化学报, 2012, **38**(6): 1017–1024)
- Wang Z, Liu Q, Xie F M, Wang B, Lu X C, Fu G Q. Model for prediction of oxygen required in BOF steelmaking. *Ironmaking & Steelmaking*, 2012, **39**(3): 228–233
- Sen P, Roy M, Pal P. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. *Energy*, 2016, **116**(Part 1): 1031–1038
- Chen Zhong-Lin, Yang Cui-Li, Qiao Jun-Fei. The prediction of incomplete time series via TG-LSTM neural network. *Control Theory and Applications*, 2022, **39**(5): 867–878
(陈中林, 杨翠丽, 乔俊飞. 基于 TG-LSTM 神经网络的非完整时间序列预测. 控制理论与应用, 2022, **39**(5): 867–878)
- Wu M, Zhang K X, An J Q, She J H, Liu K Z. An energy efficient decision-making strategy of burden distribution for blast furnace. *Control Engineering Practice*, 2018, **78**: 186–195
- Li J P, Hua C C, Yang Y N, Guan X P. Bayesian block structure sparse based T-S fuzzy modeling for dynamic prediction of hot metal silicon content in the blast furnace. *IEEE Transactions on Industrial Electronics*, 2018, **65**(6): 4933–4942
- Zhao J, Wang W, Liu Y, Pedrycz W. A two-stage online prediction method for a blast furnace gas system and its application. *IEEE Transactions on Control Systems Technology*, 2011, **19**(3): 507–520
- Hu J, Wu M, Chen X, Du S, Zhang P, Cao W H, et al. A multi-level prediction model of carbon efficiency based on the differential evolution algorithm for the iron ore sintering process. *IEEE Transactions on Industrial Electronics*, 2018, **65**(11): 8778–8787
- Du S, Wu M, Chen L F, Cao W H, Pedrycz W. Operating mode recognition of iron ore sintering process based on the clustering of time series data. *Control Engineering Practice*, 2020, **96**: Article No. 104297
- Chen Cong, Wu Min, Chen Lue-Feng, Zhang Wen, Du Sheng. Flatness prediction method based on operating mode recognition for roller quenching process. *Control Theory and Applications*, 2021, **38**(9): 1407–1413
(陈聪, 吴敏, 陈略峰, 章文, 杜胜. 基于工况识别的辊式淬火过程板形预报方法. 控制理论与应用, 2021, **38**(9): 1407–1413)
- Liu F T, Ting K M, Zhou Z H. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery From Data*, 2012, **6**(1): 1–39
- Li L, Li W, Du Q, Tao R. Low-rank and sparse decomposition with mixture of Gaussian for hyperspectral anomaly detection. *IEEE Transactions on Cybernetics*, 2021, **51**(9): 4363–4372
- Zou Tong-Hua, Gao Yun-Peng, Yi Hui-Juan, Xu Chang-Bao, Xia Rui, Wu Cong. Processing of wind power a bnormal data based on Thompson tau-quartile and multi-point interpolation. *Automation of Electric Power Systems*, 2020, **44**(15): 156–162
(邹同华, 高云鹏, 伊慧娟, 徐长宝, 夏睿, 吴聪. 基于 Thompson
- tau-四分位和多点插值的风电功率异常数据处理. 电力系统自动化, 2020, **44**(15): 156–162)
- Wang Yue-Fei, Yu Jiong, Su Guo-Ping, Qian Yu-Rong, Liao Bin, Liu Su. ODIC-DBSCAN: A new analytical algorithm for inliers. *Acta Automatica Sinica*, 2019, **45**(11): 2107–2127
(王跃飞, 于炯, 苏国平, 钱育蓉, 廖彬, 刘粟. ODIC-DBSCAN: 一种新的簇内孤立点分析算法. 自动化学报, 2019, **45**(11): 2107–2127)
- Reshef D N, Reshef Y A, Finucane H K, Grossman S R, Mcvean G, Turnbaugh P J, et al. Detecting novel associations in large data sets. *Science*, 2011, **334**(6062): 1518–1524
- Pedrycz W. Granular computing for data analytics: A manifesto of human-centric computing. *IEEE/CAA Journal of Automatica Sinica*, 2018, **5**(6): 1025–1034
- Du Sheng, Wu Min, Chen Lue-Feng, Pedrycz Witold. Operating performance assessment based on granular clustering for iron ore sintering process. *Acta Automatica Sinica*, 2023, **49**(6): 1272–1282
(杜胜, 吴敏, 陈略峰, Pedrycz Witold. 基于粒度聚类的铁矿石烧结过程运行性能评价. 自动化学报, 2023, **49**(6): 1272–1282)
- Hu J, Wu M, Chen L F, Zhou K L, Zhang P, Pedrycz W. Weighted kernel fuzzy C-means-based broad learning model for time-series prediction of carbon efficiency in iron ore sintering process. *IEEE Transactions on Cybernetics*, 2022, **52**(6): 4751–4763
- Davies D L, Bouldin D W. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1979, **PAMI-1**(2): 224–227
- Zhou P, Guo D W, Wang H, Chai T Y. Data-driven robust M-LS-SVR-based NARX modeling for estimation and control of molten iron quality indices in blast furnace ironmaking. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(9): 4007–4021
- Wang T T, Wang X P, Ma R Z, Li X Y, Hu X P, Chan F T S, et al. Random forest-Bayesian optimization for product quality prediction with large-scale dimensions in process industrial cyber-physical systems. *IEEE Internet of Things Journal*, 2020, **7**(9): 8641–8653



阳青锋 中国地质大学(武汉)自动化学院硕士研究生. 主要研究方向为复杂工业过程控制与状态监测.

E-mail: yangqingfeng@cug.edu.cn

(YANG Qing-Feng Master student at the School of Automation, China University of Geosciences.

His research interest covers control and condition monitoring of complex industrial processes.)



赖旭芝 中国地质大学(武汉)自动化学院教授. 主要研究方向为非线性系统控制, 智能系统和机器人控制. 本文通信作者.

E-mail: laixz@cug.edu.cn

(LAI Xu-Zhi Professor at the School of Automation, China University of Geosciences. Her research interest covers nonlinear system control, intelligent systems, and robot control. Corresponding author of this paper.)



杜 胜 中国地质大学(武汉) 自动化学院教授. 主要研究方向为过程控制, 智能控制和计算智能.

E-mail: dusheng@cug.edu.cn

(**DU Sheng** Professor at the School of Automation, China University of Geosciences. His research interest covers process control, intelligent control, and computational intelligence.)



陈略峰 中国地质大学(武汉) 自动化学院教授. 主要研究方向为智能系统, 模式识别和计算智能.

E-mail: chenluefeng@cug.edu.cn

(**CHEN Lue-Feng** Professor at the School of Automation, China University of Geosciences. His research interest covers intelligent systems, pattern recognition, and computational intelligence.)



胡 杰 中国地质大学(武汉) 自动化学院教授. 主要研究方向为过程控制, 计算智能和人工智能.

E-mail: hujie@cug.edu.cn

(**HU Jie** Professor at the School of Automation, China University of Geosciences. His research interest covers process control, computational intelligence, and artificial intelligence.)



吴 敏 中国地质大学(武汉) 自动化学院教授. 主要研究方向为过程控制, 鲁棒控制和智能系统.

E-mail: wumin@cug.edu.cn

(**WU Min** Professor at the School of Automation, China University of Geosciences. His research interest covers process control, robust control, and intelligent systems.)