

# 控制与决策

Control and Decision

## 基于偏态深度分类的高炉硅含量及波动预测

罗世华, 陈坤

引用本文:

罗世华, 陈坤. 基于偏态深度分类的高炉硅含量及波动预测[J]. *控制与决策*, 2021, 36(2): 491–497.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.1116>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### 基于Stacking的高炉铁水质量区间预测

Prediction of blast furnace hot metal quality interval based on stacking

*控制与决策*. 2021, 36(2): 335–344 <https://doi.org/10.13195/j.kzyjc.2019.0786>

### 基于多维泰勒网的超前d步预测模型

d-step-ahead predictive model based on multi-dimensional Taylor network

*控制与决策*. 2021, 36(2): 345–354 <https://doi.org/10.13195/j.kzyjc.2019.0722>

### 基于改进堆叠自动编码器的循环冷却水系统工艺介质温度预测控制方法

Predictive control method of process medium temperature in circulating cooling water system based on improved stacked auto encoders

*控制与决策*. 2020, 35(12): 2835–2844 <https://doi.org/10.13195/j.kzyjc.2019.0694>

### 基于无标签、不均衡、初值不确定数据的设备健康评估方法

Equipment health risk assessment based on unlabeled, unbalanced data under uncertain initial condition

*控制与决策*. 2020, 35(11): 2687–2695 <https://doi.org/10.13195/j.kzyjc.2018.1493>

### 基于高炉炉况评价和反馈补偿的喷煤量决策优化

Decision-making optimization of coal injection volume based on evaluation of blast furnace condition

*控制与决策*. 2020, 35(11): 2803–2809 <https://doi.org/10.13195/j.kzyjc.2018.1553>

# 基于偏态深度分类的高炉硅含量及波动预测

罗世华<sup>†</sup>, 陈 坤

(江西财经大学 统计学院, 南昌 330013)

**摘 要:** 高炉冶炼是个具有高度复杂性、混沌性、时滞性的动态过程, 工业上常常用铁水硅含量反馈高炉炉温热状态波动变化, 而偏态投影深度在数据有偏时可以较好地反映出数据的离群情况, 在高维数据分类计算中十分稳健. 首先, 通过差分处理及相关性分析确定 11 个影响因素作为输入变量, 用于研究各变量变化对硅含量变化的关系; 然后, 将偏态投影深度值在 90% 的置信区间外的数据视作离群值, 分为稳定类和离群类; 最后, 对稳定数据利用 Elman 神经网络预测模型进行预测, 对于离群类利用 Logistic 模型在炉温不同波动方向下的规律进行归类预测. 实例仿真研究表明, 稳定类 157 炉的预测精度高达 85.3%, 离群类的预测精度达到 82.6%.

**关键词:** 硅含量; 偏态投影深度; Elman 神经网络; Logistic 模型; 预测

中图分类号: TF4 文献标志码: A

DOI: 10.13195/j.kzyjc.2019.1116

开放科学(资源服务)标识码(OSID):



引用格式: 罗世华, 陈坤. 基于偏态深度分类的高炉硅含量及波动预测[J]. 控制与决策, 2021, 36(2): 491-497.

## Prediction of blast furnace silicon content and fluctuation based on skewness depth classification

LUO Shi-hua<sup>†</sup>, CHEN Kun

(College of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China)

**Abstract:** Blast furnace smelting is a dynamic process with high complexity, chaos and time delay. In industry, molten iron silicon content is often used to feed back the fluctuation of blast furnace temperature and thermal state. The skew projection depth can reflect the outliers of the data well when the data is biased, and it is very robust in the classification calculation of high-dimensional data. Firstly, 11 influencing factors are determined as input variables through differential processing and correlation analysis in this paper, which are used to study the relationship between changes of various variables and changes of silicon content. Then, the data whose projection depth value is outside 90% confidence interval are regarded as outliers and classified into stable and outliers. Finally, the stable data are predicted by Elman neural network prediction model, and the outliers are classified and predicted by Logistic model under different fluctuation directions of furnace temperature. The simulation results show that the prediction accuracy of stable class 157 furnace is up to 85.3%, and that of outlier class is up to 82.6%.

**Keywords:** silicon content; skew projection depth; Elman neural network; Logistic model; prediction

## 0 引 言

高炉炼铁是个极其复杂, 同时具备滞后性、混沌性、非线性、时变性的动态过程<sup>[1]</sup>. 在传统工艺上, 铁水硅含量常常用来反馈炉内热反应状况, 因此, 硅含量的准确预测对炉温监控、高炉工序的稳定顺行都有着一定的指导意义.

基于炉内高温高压、强腐蚀性、热反应剧烈多变等特点, 炉温的波动不仅与炼铁的复杂机理相关, 同时受到众多不确定性因素的干扰, 使得炉温波动以及铁水硅含量的预测问题受到了种种限制. 机理模型

对专家经验知识依赖性较强<sup>[2]</sup>, 考虑了较多的主观因素, 预测效果时常有悖于实际. 随着大数据时代热潮涌进, 基于数据驱动的智能算法广受青睐, 目前主流的铁水硅含量预报模型主要有自回归模型<sup>[3]</sup>、神经网络预报模型<sup>[4-5]</sup>、贝叶斯网络模型<sup>[6-7]</sup>、支持向量机预测模型等<sup>[8-9]</sup>. 这些模型在炉况较为稳定的条件下都取得了较为满意的结果, 然而在炉况波动较大时预测精度不高, 难以真正满足在线闭环预测控制的需要.

鉴于神经网络较强的容错性、自适应性、动态记忆性、能够并行处理信息等优点, 在多元非线性回归

收稿日期: 2019-08-03; 修回日期: 2019-10-13.

基金项目: 国家自然科学基金项目(61973145); 江西省教育厅重点项目(GJJ180247).

责任编辑: 刘向杰.

<sup>†</sup>通讯作者. E-mail: luoshihua@aliyun.com.

中表现十分优异<sup>[10]</sup>. 然而, 由于影响硅含量的因素紊乱多变, 炉内热反应极端复杂, 往往使得采样数据波动明显, 这对网络的学习是十分麻烦的, 因此, 数据的有效分类对模型的训练和学习效率都至关重要.

统计深度函数作为近年来发展迅猛的稳健统计学上的分支, 最早由 Tukey<sup>[11]</sup> 在研究一维中位数向高维中位数推广中提出, 在处理高维数据排序、分类问题中应用十分广泛<sup>[12-14]</sup>. 偏态投影深度作为众多深度函数中的一员<sup>[15]</sup>, 具备仿射不变性、稳健性、较高离群崩溃点的优良特性, 能够很好地反映高维数据的分布特征和离群情况. 另外, 数据噪声往往会给模型的预测精度造成不必要的影响<sup>[16]</sup>, 直接忽略离群值又会付出信息缺失的代价, 针对 Logistic 模型易于实现与在分类上优异的性能<sup>[17-18]</sup>, 从概率的角度充分挖掘离群值对炉温波动(上升与下降)的潜在联系. 数据的有效分类能够让模型的学习与构建更加高效.

鉴于此, 本文提出了偏态投影深度分类算法, 依据输入样本点的深度值大小将数据分为稳定类与离群类. 对于稳定类, 建立 Elman 神经网络硅含量预报模型; 对于离群类, 建立 Logistic 炉温波动预测模型, 计算各离群点在炉温不同波动状态下的概率值, 并以概率大的方向作为波动预测方向, 在影响因素波动较大时, 及时给出对应的炉温控制操作指导.

## 1 预测模型及其算法

### 1.1 偏态投影深度

给定一元分布函数, 记  $Q_1(F_u)$ 、 $Q_3(F_u)$ 、 $\text{med}(F_u)$  分别为  $F_u$  的下四分位数、上四分位数和中位数, 则点  $x$  相对于  $R^p$  上随机向量  $X$  的分布函数  $F$  的偏态投影深度值定义为

$$\text{APD}(x, F) = (1 + \sup_{u \in S^{p-1}} \text{MO}(u, x, F))^{-1}. \quad (1)$$

其中:  $S^{p-1} = \{x \in R^p, \|x\| = 1\}$ , 且

$$\text{MO}(u, x, F) = \begin{cases} \frac{u^T x - \text{med}(F_u)}{Q_1(F_u) - \text{med}(F_u)}, & u^T x < \text{med}(F_u); \\ \frac{u^T x - \text{med}(F_u)}{Q_3(F_u) - \text{med}(F_u)}, & u^T x \geq \text{med}(F_u). \end{cases} \quad (2)$$

$F_u$  为  $X$  在方向  $u$  上一元投影  $u^T x$  所对应的函数, 与投影深度函数类似, 若满足下列之一:

$$\begin{cases} u^T x - \text{med}(F_u) = Q_1(F_u) - \text{med}(F_u) = 0, \\ u^T x - \text{med}(F_u) = Q_3(F_u) - \text{med}(F_u) = 0. \end{cases} \quad (3)$$

则强定义  $\text{MO}(u, x, F) = 0$ . 偏态投影深度在计算上比投影深度更快, 这是因为其定义中没有对称标度因子, 比投影深度计算更便捷, 在数据中有偏离中心的离群值存在的情况下, 偏态投影深度的计算值比投影深度更为精确, 这是因为总存在一些方向向量, 使得数据在有偏时  $|Q_1(F_u) - \text{med}(F_u)| \neq |Q_3(F_u) - \text{med}(F_u)|$ . 因为其深度值由内至外单调递减, 根据偏态投影深度值大小排序, 得到一个次序统计量序列

$$M = (x_{i1}, x_{i2}, \dots, x_{in}), x_{i1} \leq x_{i2} \leq \dots \leq x_{in}. \quad (4)$$

其中:  $x_i$  为方向  $i$  上的投影深度值,  $i$  为高维单位球内随机投影方向. 深度值越大越靠近数据中心, 根据深度值大小, 偏态投影深度算法能够很好地区分离群值.

### 1.2 Elman 网络

Elman 网络是递归神经网络的一种, 除了一般网络具备的前馈连接结构以外, 还包括由一组特殊结构单元构成的反馈连接, 具有动态记忆性, 称为关联层, 每个隐含层节点都连接着一个与之对应的关联层节点. 关联层能够从隐含层接收反馈信号, 记录并储存隐层单元前一时刻的输出值, 再自联到隐层的输出, 这种自联方式加强了网络对历史数据的敏感性, 通过这种内部反馈的方式增强了网络处理动态信息的能力, 从而实现动态反馈的过程. 其结构特征如图 1 所示.

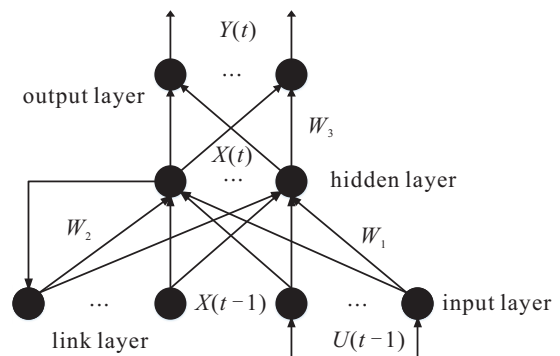


图 1 Elman 网络功能结构

网络的内部计算公式为

$$\begin{cases} X(t) = f_1(W_2 X(t-1) + W_1(U(t-1))), \\ Y(t) = f_2(W_3 X(t)). \end{cases} \quad (5)$$

其中:  $W$  为各层之间的连接权重矩阵,  $f_1$  为输入层到隐含层的传递函数,  $f_2$  为隐含层到输出层的传递函数, 分别采用  $\text{tansig}(\cdot)$  函数和  $\text{purelin}(\cdot)$  函数.

### 1.3 Logistic 模型

Logistic 回归模型, 作为广义线性模型的一个特例, 在非线性回归中应用广泛, 最常用于 0-1 分类模

型. 与其他多元线性模型不同的是, 为确保函数的连续性, 它将函数值通过单调可微  $\text{sigmoid}(\cdot)$  函数映射到  $[0, 1]$  之间, 最后与阈值大小进行比较确定分类结果. 假设 0 表示事件不发生, 1 表示事件发生, 则事件发生的概率可表示为

$$P(Y = 1|x_1, x_2, \dots, x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)} \quad (6)$$

其中  $P \in [0, 1]$ , 代表事件发生的概率, 则事件不发生的概率为  $1 - P$ , 两者比值  $\frac{P}{1 - P}$  称作事件发生的几率. 假设事件发生的几率与自变量之间存在某种线性关系, 有

$$\ln \frac{P}{1 - P} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k. \quad (7)$$

基于极大似然思想, 采用梯度下降方法计算模型系数  $\beta$ , Logistic 回归即是根据事件发生的概率作为归类判别依据. 通常做法是事先设定一个阈值, 即先验概率, 阈值大小依据先验知识决定, 然后与阈值大小对比归类.

### 1.4 模型预测及流程

在实际生产过程中, 经常会因不可测因素影响而产生异常数据——离群值. 本文采用偏态投影深度函数, 依据各输入样本点的深度值大小区分出离群值, 针对离群类和稳定类分别建立 Elman 网络模型和 Logistic 分类模型. 主要仿真流程如下.

step 1: 对原始数据进行一阶差分和相关分析处理, 选取一些随机投影方向向量, 参照式 (1), 计算出各输入变量在最优投影方向上的偏态投影深度值. 根据式 (4) 将所有点的深度值按大小排序, 根据深度 90% 置信区间将数据分为稳定和离群两大类.

step 2: 数据分类识别后, 采用 Elman 递归神经网络对稳定类数据仿真学习. 网络学习中, 若隐层节点数太少, 则容易造成网络误差太差的“欠拟合”现象; 若隐层节点数太多则会延长网络训练时间, 网络训练容易陷入局部极小点而得不到最优点, 出现“过拟合”现象. 因此, 合理隐层节点数的确定应该综合考虑网络结构复杂程度和误差大小. 根据神经网络隐含层节点的经验公式, 有

$$I = \sqrt{m + n} + c, \quad c \in [0, 10]. \quad (8)$$

其中:  $m$ 、 $n$  为输入层与输出层节点个数,  $I$  为隐含层节点个数. 由于目前确定神经网络参数没有统一的有效方法, 本文采用“试凑法”, 即在  $I$  所在的区间内通过设置最大迭代次数和学习率, 不断训练调整参数, 取均方误差最小的模型为最优模型, 最后在测试集上验证模型命中率.

step 3: 给离群值作出可信的调控指导对维持炉温波动的稳定极其重要, 具体作法是: 将稳定数据分为炉温上升和炉温下降两类, 分别计算两类的输入变量数据中心, 即深度最大样本点, 以该点作为炉温调控参考值.

step 4: 对于离群类, 建立 Logistic 分类回归模型, 对每个离群点计算炉温上升与炉温下降的发生概率, 探究不同离群状态下对炉温波动的影响, 以概率较大值作为归类判别, 并与实际值比较以验证模型准确率. 预测流程如图 2 所示.

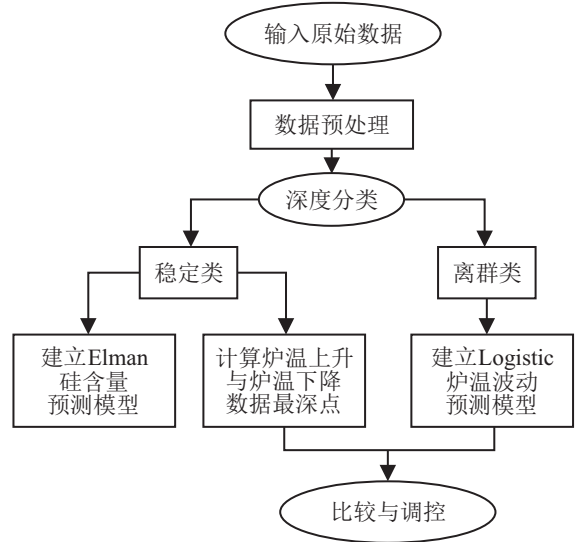


图 2 预测流程

## 2 仿真分析

### 2.1 参数选取及数据预处理

本文数据来源于中国包钢集团某高炉连续 840 炉次在线生产数据, 源数据有部分缺失, 考虑到数据采集条件及各参数与铁水硅含量之间的关联性, 最终选取 11 个皮尔逊系数最大的参数作为输入变量, 分别为风量、风温、风压、顶压、料速、富氧率、顶温、CO、CO<sub>2</sub>、配料碱度、冶炼强度; 输出变量为硅含量, 相关系数见表 1.

表 1 变量相关系数表

变量	风量	风温	风压	顶压	料速	富氧率	顶温	CO	CO <sub>2</sub>	配料碱度	冶炼强度
相关系数	-0.26	-0.13	-0.11	-0.11	-0.19	-0.17	0.11	-0.12	-0.13	0.53	-0.19

表2  $\alpha$ -th分位深度值

$\alpha$ -th	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	$\alpha = 0.95$	min	max
depth value	0.0823823	0.1201488	0.2262909	0.3536613	0.4012231	0.0084173	0.6046083

依据炼钢机理的知识,影响硅含量变化的因素繁多复杂,炉内涉及的反应主要为热反应与还原反应.风量、风温、顶温能够直接控制硅还原的热反应平衡,料速代表原料的投入批数;风压、顶压是反映炉况气化状态的测量指标,能够间接反映高炉煤气的分布状况;富氧率、CO、CO<sub>2</sub>对硅的氧化还原至关重要;配料碱度与冶炼强度作为高炉炼铁的客观因素也关乎着出铁炉温.这些参数中有些是人为可控参数,有些为不可控参数,考虑各因素前一期的时滞性影响,对原始数据作一阶差分处理,用于研究各工艺参数变化与硅含量变化之间的关系.模型输入变量为各参数的变化量,输出变量为硅含量变化量.

## 2.2 深度分类

数据总样本容量为  $N = 839$ , 偏态投影深度值能够较好地地区分数据集中的离群值. 对离群值的划分采用置信区间(深度域)的方式,若置信区间取值太小则会使离群值数量增多,在炉况不稳定的情况下难以反映离群数据对炉温波动趋势的影响;若置信区间取值太大,则会使离群值数量减少,会因数据量太小影响模型预测结果. 为了保证计算的准确性<sup>[17]</sup>,选取100000个随机投影方向,选取90%置信区间以外的点作为离群值,由式(1)计算各样本点的深度值后根据式(4)排序,建立深度值  $\alpha$ -th分位表如表2所示. 在深度值90%置信区间上 ( $APD > 0.1201488$ ) 将数据分为稳定类和离群类,稳定类样本容量  $N_1 = 756$ ,其中训练样本容量为600,预测样本容量为156,进而得到离群类样本容量  $N_2 = 83$ . 此外,本文所有程序算法均在 Matlab 2017Rb 版中完成.

## 2.3 仿真结果分析

Elman网络能够内部反馈、存储和利用过去时刻的输出信息,在实现非线性动态映射中表现优越,同时也存在一些缺陷. 在选择隐含层节点数方面目前没有切实高效的具体方法<sup>[5]</sup>,节点数太多或太少都会增加网络误差,不同问题背景下权衡选择尤为重要. 模型输入变量个数为11,输出变量个数为1,参照经验公式(8)先确定隐含节点大致范围[3,13],再通过遍历试凑的方式比较不同隐含层对应的模型误差,设置最大迭代次数10000并选择训练误差最小的模型,最终确定为Elman(11,5,1). 用该模型对剩余156炉测试样本预测,为了更直观地看到预测效果,还原

差分数数据后得到预测图如图3所示,模型残差图如图4所示.

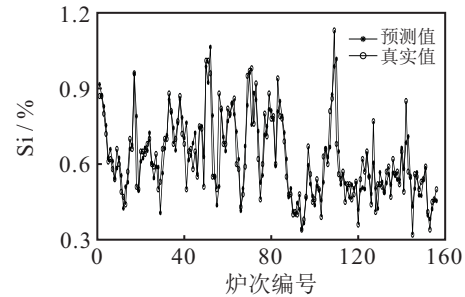


图3 硅含量预测图

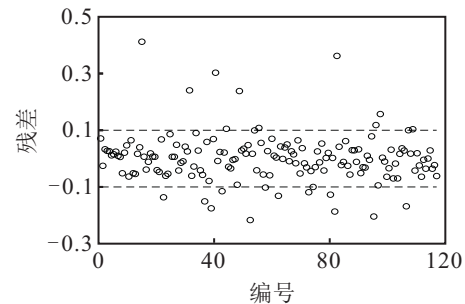


图4 模型残差图

学习率是梯度下降寻优的一个重要超参数,设置过大会影响网络收敛跳过最优解,设置过小会影响训练速度且易陷入局部最优. 本文采用自适应学习率梯度下降算法,经过多次实验后设置初始学习率为0.02. 此外,为避免网络过学习、泛化能力低的问题,采用早期中止法,若网络训练的验证误差超过一定次数仍不下降,则表明训练误差已达到最小,这里默认最大次数为6,超过最大次数网络训练就会中止,能够有效避免网络过学习. 为了评价模型的优越性,与未分类网络模型作对比,即在其他网络参数设置不变的情况下,将所有样本数据按照4:1比例划分训练网络,最终确定未分类模型为Elman(11,6,1). 对比主要体现在均方误差和命中率两项指标上,不失一般性,验证结果重复计算10次取平均值,通过比较,未分类模型命中率只有80.8%,分类后的Elman网络较分类前均方误差更小、命中率更高,两类模型参数对比结果见表3,表明分类后的模型优于未分类模型,对于波动较大的离群类分类研究具有理论意义.

由预测结果可以看出,预测值与真实值走向趋势基本保持一致,高炉炼铁是个复杂多变的动态过程,运作过程中炉况往往波动剧烈. 就目前国内高炉生

表3 两类模型参数比较

		隐含层					命中率/%
		4	5	6	7	8	
均方误差	已分类	0.0399	0.0375	0.0379	0.0403	0.0389	85.3
	未分类	0.0402	0.0398	0.0394	0.0398	0.0401	80.8

产条件而言,数千立方以上的现代化高炉运作较为平稳,一般采取0.05%的误差测算标准,而大多数高炉炉况波动较大,一般采用0.1%的误差测算标准,本文采用0.1%的测算误差. 预测156炉中仅有23炉在允许误差范围之外,预测精度高达85.3%,预测效果良好. 命中率计算公式如下:

$$\rho = \frac{1}{n} \sum_{i=1}^n \eta_i \times 100\% \quad (9)$$

其中

$$\eta = \begin{cases} 1, & |y - \hat{y}| \leq 0.1; \\ 0, & |y - \hat{y}| > 0.1. \end{cases} \quad (10)$$

$\rho$ 为命中率, $\eta = 1$ 表示命中,反之表示不命中; $\hat{y}$ 为预测值, $y$ 为实际值. 此外,为了避免离群值对炉温波动造成不必要的影响,及时调控离群参数对高炉的稳定顺行至关重要. 由于偏态投影深度具有较高崩溃点的优良性质,聚类中心不受少数离群值的干扰,将稳定数据集按照炉温的波动方向划分为炉温上升和炉温下降两类,依照式(1)分别计算两类最深点,炉温上升最深点为

$$R = (-33.51, 0.88, -1.26, -0.02, 7.94, 0, -3.06, -0.07, -0.23, 0.06, 0.09), \quad (11)$$

炉温下降最深点为

$$D = (36.61, -3, 0.45, -0.11, -2.65, -0.07, 2.46, 0, -0.1, -0.03, 0.06). \quad (12)$$

由于炉温稳定状态下各参数变化也相对稳定,最深点可为离群参数的及时调控提供参考范围,如图5所示. 通常离群参数对应的炉况极不稳定,炉温波动振幅较大的同时,调控操作也十分受限,工艺上仅凭工长经验难以及时调控,而最深点表征着炉况最佳

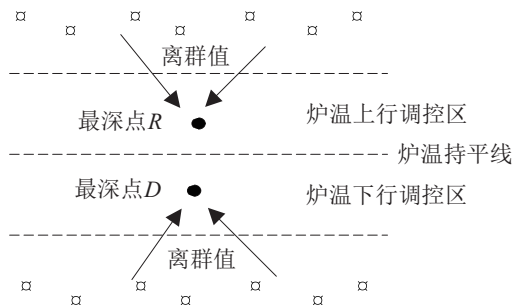


图5 炉温调控示意图

稳定状态的时态值,其附近小领域构成可操作缓冲空间,在异常炉况的监测与调控上更加高效. 对此,以数据最深点作为离群参数的参考指标具有更为科学的指导意义.

### 2.4 离群类预测

考虑到离群值的波动变化比较剧烈,直接用网络预测效果往往不尽如人意,直接忽略又会造成信息缺失. 而对于影响硅(Si)含量的离群波动值,预测其对硅含量的波动方向对高炉工长的操作选择、高炉作业中的故障防患都有指导意义. 本文拟用Logistic模型,分别计算不同离群状态下炉温上升与炉温下降的发生概率,以概率大小客观表示事件发生的可能性. 设定Logistic(0 - 1)变量模型,1表示炉温上升,0表示炉温下降. 从样本量上看,炉温上升和炉温下降数据量持平,表明两个状态发生的概率是几乎相等的. 因此,取先验概率为 $P = 0.5$ ,概率大于0.5表示炉温上升, $y_i = 1$ ;概率小于等于0.5表示炉温下降, $y_i = 0$ ,即

$$\begin{cases} P(y_i = 1|x_1, x_2, \dots, x_n) = 0.5. \\ y_i = \begin{cases} 1, & \hat{P} > P; \\ 0, & \hat{P} \leq P. \end{cases} \end{cases} \quad (13)$$

其中 $\hat{P}$ 为模型预测值. 由于离群类样本容量太小 $N_2 = 83$ ,为不失模型的泛化性,随机选取60个数据建立模型,剩余23个样本用于模型测试,得到模型参数如表4所示.

表4 Logistic模型参数表

variable	coefficients	variable	coefficients
(intercept)	0.467	$x_6$	-0.298
$x_1$	0.007	$x_7$	0.037
$x_2$	0.019	$x_8$	1.005
$x_3$	-0.145	$x_9$	1.079
$x_4$	0.144	$x_{10}$	-42.83
$x_5$	0.076	$x_{11}$	-22.926

chi<sup>2</sup>-statistic vs. constant model: 42, p-value = 1.59e-05

模型的p-value值为1.59e-05,模型通过检验,得到的Logistic模型可表示为

$$\ln \frac{P}{1-P} = 0.467 + 0.007x_1 + 0.019x_2 - 0.145x_3 + 0.144x_4 + 0.076x_5 - 0.298x_6 + 0.037x_7 + 1.005x_8 + 1.079x_9 - 42.83x_{10} - 22.926x_{11}. \quad (14)$$

然而,仅依靠对数值不能反映出事件发生概率的大小,需经过逆运算得到每个点的概率值,参照式(13)判别法则得到相应的分类结果如图6所示.其中:星型点代表真实类别,圆形点代表预测类别,重合的点表示命中.模型分类结果见表5和表6.

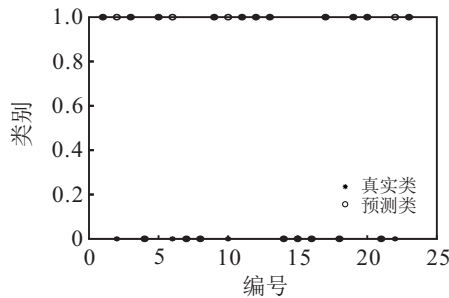


图6 模型分类预测图

表5 Logistic 模型预测表

编号	概率	实际波动方向	预测波动方向
1	0.997 837 882	上升	上升
2	0.635 937 320	下降	上升
3	0.782 180 261	上升	上升
4	0.028 904 813	下降	下降
5	0.529 388 748	上升	上升
6	0.991 265 477	下降	上升
7	0.212 452 012	下降	下降
8	0.474 313 965	下降	下降
9	0.987 103 198	上升	上升
10	0.892 590 119	下降	上升
11	0.985 863 800	上升	上升
12	0.767 471 744	上升	上升
13	0.984331164	上升	上升
14	0.125 715 675	下降	下降
15	0.227 512 360	下降	下降
16	0.013 391 621	下降	下降
17	0.875 948 175	上升	上升
18	0.058 434 585	下降	下降
19	0.917 767 698	上升	上升
20	0.959 960 412	上升	上升
21	0.004 273 553	下降	下降
22	0.590 727 278	下降	上升
23	0.985 000 337	上升	上升

表6 Logistic 模型命中率

sample size	hit number	error number	accuracy rate / %
23	19	4	82.6

由表6可见,模型预测精度高达82.6%,在炉温波动剧烈的状态下,能够较大程度上反映各参数变化对炉温波动的关系.如何应对离群值带来炉温动荡的潜在风险,及时对炉温不稳定性波动作出有效调控,方法可以参照前述R、D两个向量,在可调控范围内给予一定的调控缓冲空间,为实际生产提供科学的指导策略.

### 3 结论

炉温波动关乎着出铁质量和冶炼精度,异常值对炉温的影响不容小觑.基于数据驱动的数理模型对偏态数据的敏感性较强,模型的预测精度和计算复杂度往往受到很大影响.本文考虑离群值对炉温波动的影响,提出了基于偏态投影深度分类下的分类组合预测模型——Elman-Logistic 预测模型,偏态投影深度函数将多指标因素转换成单一指标区分离群类,建立Elman硅含量变化预测模型和Logistic 炉温波动预测模型.此外,以炉温不同波动方向下最深点为离群值的调控作为参考,两类模型预测效果良好,Elman模型命中率高达85.3%,Logistic模型命中率高达82.6%.

### 参考文献(References)

- [1] Shi L, Li Z L, Yu T, et al. Model of hot metal silicon content in blast furnace based on principal component analysis application and partial least square[J]. Journal of Iron and Steel Research (International), 2011, 18(10): 13-16.
- [2] Chu M S, Yang X F, Shen F M. Numerical simulation of innovative operation of blast furnace based on multi-fluid model [J]. Journal of Iron and Steel Research, 2006, 13(6): 8-15.
- [3] Saxen H. Short-term prediction of silicon content in pig iron[J]. Canadian Metallurgical Quarterly, 1994, 33(4): 319-326.
- [4] 文冰洁, 吴胜利, 周恒, 等. 基于BP神经网络的COREX铁水硅含量预测模型[J]. 钢铁研究学报, 2018, 30(10): 776-781. (Wen B J, Wu S L, Zhou H, et al. Prediction model of silicon content in COREX molten iron based on BP neural network[J]. Journal of Iron and Steel Research, 2018, 30(10): 776-781.)
- [5] 李泽龙, 杨春节, 刘文辉, 等. 基于LSTM-RNN模型的

- 铁水硅含量预测[J]. 化工学报, 2018, 69(3): 992-997.  
(Li Z L, Yang C J, Liu W H, et al. Prediction of silicon content in molten iron based on lstm-rnn model[J]. Journal of Chemical Engineering, 2018, 69(3): 992-997.)
- [6] 龚淑华, 刘祥官. 模糊贝叶斯网络应用于预测高炉铁水含硅量变化趋势[J]. 冶金自动化, 2005, 29(5): 30-32.  
(Gong S H, Liu X G. Application of fuzzy bayesian network to predict the change trend of silicon content in molten iron of blast furnace[J]. Metallurgical Automation, 2005, 29(5): 30-32.)
- [7] 刘学艺, 刘祥官, 王文慧. 贝叶斯网络在高炉铁水硅含量预测中的应用[J]. 钢铁, 2005, 40(3): 17-20.  
(Liu X Y, Liu X G, Wang W H. Application of bayesian network in prediction of silicon content in molten iron of blast furnace[J]. Iron and Steel, 2005, 40(3): 17-20.)
- [8] Xu X, Hua C C, Tang Y G, et al. Modeling of the hot metal silicon content in blast furnace using support vector machine optimized by an improved particle swarm optimizer[J]. Neural Computing and Applications, 2016, 27(6): 1451-1461.
- [9] Zhao W, Ting Z S, Tao T, et al. Adynamic particle filter-support vector regression method for reliability prediction[J]. Reliability Engineering and System Safety, 2013, 119(1): 109-116.
- [10] 樊兆峰, 马小平, 邵晓根. 非线性系统RBF神经网络多步预测控制[J]. 控制与决策, 2014, 29(7): 1274-1278.  
(Fan Z F, Ma X P, Shao X G. Multistep predictive control of nonlinear RBF neural network[J]. Control and Decision, 2014, 29(7): 1274-1278.)
- [11] Tukey J W. Mathematics and the picturing of data[C]. Proceedings of the International Congress of Mathematicians. Vancouver, 1975: 523-531.
- [12] Ghosh A K, Chaudhuri P. On maximum depth and related classifiers[J]. Scandinavian Journal of Statistics, 2005, 32(2): 327-350.
- [13] Cui X, Lin L, Yang G R. An extended projection data depth and its applications to discrimination[J]. Communications in Statistics—Theory and Methods, 2008, 37(4): 2276-2290.
- [14] Cuevas A, Febrero M, Fraiman R. Robust estimation and classification for functional data via projection-based depth notions[J]. Computational Statistics, 2007, 22(3): 481-496.
- [15] Liu X H, Zuo Y J. A Matlab package for computing projection depth[J]. Journal of Statistical Software, 2015, 65(2): 1-21.
- [16] Wang J J, Zhang W, Li Y N, et al. Forecasting wind speed using empirical mode decomposition and Elman neural network[J]. Applied Soft Computing, 2014, 23: 452-459.
- [17] Dang Y, Guo L, Lv D, et al. Classification of breast lesions based on a dual S-shaped logistic model in dynamic contrast enhanced magnetic resonance imaging[J]. Science China Life Sciences, 2011, 54(10): 889-896.
- [18] 李玉莹, 张景肖. 成分数据的 Logistic 回归模型研究[J]. 数理统计与管理, 2019, 38(3): 442-449.  
(Li Y Y, Zhang J X. Logistic regression model of component data[J]. Mathematical Statistics and Management, 2019, 38(3): 442-449.)

### 作者简介

罗世华(1975—), 男, 教授, 博士生导师, 从事复杂系统建模与优化等研究, E-mail: luoshihua@aliyun.com;

陈坤(1995—), 男, 硕士生, 从事数理建模与优化的研究, E-mail: 543787289@qq.com.

(责任编辑: 郑晓蕾)