

基于多源数据的铝土矿浮选生产指标集成建模方法

曹斌芳^{1,2}, 谢永芳^{1†}, 阳春华¹, 桂卫华¹, 王晓丽¹

(1. 中南大学 信息科学与工程学院, 湖南 长沙 410083; 2. 湖南文理学院 物理与电子科学学院, 湖南 常德 415000)

摘要: 在长流程浮选过程中, 生产指标难以在线检测, 造成操作不及时, 影响系统的稳定运行. 本文提出了一种基于多源数据的铝土矿浮选过程生产指标集成建模方法. 首先结合浮选机理和现场工人经验, 分析影响和反映生产指标的多源数据(生产数据和泡沫图像特征数据); 然后分别建立各生产指标预测子模型和同步误差补偿子模型; 最后采用信息熵和智能协调策略分别构建精矿品位和尾矿品位的集成预测模型. 工业验证和工况分析表明, 本文集成建模方法具有良好的预测性能和较强的泛化性, 为基于生产指标的浮选过程操作参数控制和全流程优化奠定基础.

关键词: 泡沫特征; 生产指标; 集成建模; 偏最小二乘; 正则极限学习机

中图分类号: TP29 **文献标识码:** A

Integrated modeling for production index of bauxite flotation based on multi-source data

CAO Bin-fang^{1,2}, XIE Yong-fang^{1†}, YANG Chun-hua¹, GUI Wei-hua¹, WANG Xiao-li¹

(1. School of Information Science and Engineering, Central South University, Changsha Hunan 410083, China;

2. College of Physics and Electronics Science, Hunan University of Arts and Sciences, Changde Hunan 415000, China)

Abstract: Because of the difficulty in online detecting the production index in a long-time flotation process, we are unable to take the counterpart operation timely to keep the production process stable. To deal with this problem, we propose an integrated prediction model for the production index of bauxite flotation based on multi-source data. Firstly, we make use of the flotation mechanism and the experiences of operators to analyze the multi-source data (production data and froth image features) that affects the production index. Then, the prediction sub-model and the corresponding compensation sub-model for each production index are built, and their parameters are optimized. Then, the integrated prediction models of concentrate grade and tailing grade are established by using the information entropy and coordination strategy. Industrial validation results show that the proposed method improves the prediction accuracy and adaptability, which lays a foundation for controlling the operating parameters of the flotation and the total process optimization.

Key words: froth features; production index; integrated model; partial least squares; regular extreme learning machine

1 引言(Introduction)

铝作为一种重要的有色金属原材料, 在国民经济中有着广泛的应用. 为了充分利用低品位铝土矿资源, 我国建立了首条选矿拜耳法生产线, 其方法是在拜耳法生产流程中增设一个浮选流程. 浮选是利用矿物表面物理化学性质差异进行矿物分离的有效方法, 其目的是提高矿物的富集程度, 最终提高拜耳法生产氧化铝的效率^[1]. 因此, 如何有效地提高浮选效果是氧化铝生产中的一个重要环节. 精矿品位和尾矿品位是铝土矿浮选过程的重要生产指标, 可通过粒度分析仪在线获得, 然而粒度分析仪价格贵, 且容易受铝土矿浆浓度过而导致堵塞损坏, 所以目前主要依靠人工化

验获得, 该方式具有很大的滞后性, 使得操作工人不能及时获得矿物品位的变化信息, 难以有效指导生产操作和过程的优化运行. 因此研究浮选过程生产指标的实时在线检测方法, 对稳定浮选过程、优化浮选操作以及提高我国铝土矿资源的综合利用水平具有重要的意义.

针对浮选过程生产指标的检测方法, 国内外学者进行了大量研究^[2-5]. 考虑到浮选过程机理复杂, 传统上基于机理的建模方法都存在较多的简化和假设^[4], 使得难以准确描述真实浮选过程. 目前浮选生产指标的检测主要是结合现场操作工人经验, 领域专家知识以及统计建模规则等完成. 如文献[5]用铜浮选过程中

收稿日期: 2014-01-07; 录用日期: 2014-05-16.

†通信作者. E-mail: yfxiemail.csu.edu.cn; Tel.: +86 13786163035.

基金项目: 国家自然科学基金重点资助项目(61134006); 国家创新研究群体科学基金资助项目(61321003); 国家自然科学基金资助项目(61473318, 61304126); 高等学校博士学科点专项基金博导类资助课题(20120162110076); 湖南省研究生科研创新项目(CX2014B077).

获得的原矿品位、进料速率、矿浆浓度以及矿浆液位作为模型输入数据, 比较了自回归滑动平均模型 (autoregressive moving average model, ARMAX)、神经网络、模糊组合模型以及偏最小二乘 (partial least squares, PLS) 几种预测模型在铜浮选精矿品位预测中的应用, 得出PLS能够得到相对准确的预测结果. 但由于PLS模型是线性结构, 采用经验风险最小的建模策略, 易导致模型的过拟合和泛化性差. 针对上述问题, 文献[6]以铜浮选过程中的进料速率、空气流速以及矿浆pH值作为输入变量, 采用递归偏最小二乘建立铅铜的精矿品位和回收率预测模型, 仿真结果表明该模型要优于PLS方法. 针对入矿数据的高维非线性特性, 文献[7]提出了采用核心主元分析 (kernel principal component analysis, KPCA) 提取主元特征, 然后建立磁铁矿品位预测模型. 虽然KPCA或主元分析 (principal component analysis, PCA) 能够完成对数据的降维和消除数据之间的冗余特性, 然而降维以后的主元并不一定与输出样本具有最大的相似度, 容易导致模型预测性能的不稳定.

近年来, 基于浮选泡沫图像特征的浮选过程生产指标检测方法受到广泛关注. 如文献[8]指出浮选泡沫的表面特征包含有大量与操作参数和生产指标相关的信息, 是判断浮选效果的重要依据. 文献[9]利用泡沫的颜色、表面纹理等视觉信息进行煤泥浮选过程生产指标的预测. 文献[10]研究了锌浮选过程泡沫大小、颜色、速度与精矿品位、回收率的关系, 进而构建了专家控制系统. 文献[11]研究了铂浮选过程, 泡沫纹理特征的提取以及与精矿品位的预估模型. 这些研究说明浮选泡沫表面的视觉特征能够反映浮选性能, 是完成浮选过程生产指标预测的一种有效方法.

上述文献[5-7, 9, 11]说明目前浮选过程的生产指标模型主要集中在构造某浮选流程 (如粗选流程) 的单一预测模型, 这种建模方法没有充分考虑到实际浮选过程是由多个浮选流程构成, 不同流程之间存在相互关联和耦合特性, 容易造成模型精度低和泛化性差. 研究表明集成多个子模型的方法可以有效的提高模型的精度, 泛化性和可信度^[12-14], 该方法已经成功的应用在磨机负荷参数测量^[15], 铅锌烧结过程参数预测^[16]和铜吹炼过程^[17]等复杂有色冶金过程. 为此, 本文将集成建模方法引入浮选过程, 结合浮选工艺和浮选对象的长流程, 时变特性以及不同输入源对模型生产指标的影响, 构建了一种全流程集成协调建模方法, 以期提高浮选过程生产指标模型的综合性能, 为后续研究浮选全流程的优化控制奠定基础.

针对以上分析, 考虑浮选过程采集的生产数据和多传感器图像特征数据间的互补性和冗余性, 本文提出一种基于多源数据的浮选过程生产指标集成建模方法. 首先结合工人经验分析影响和表征生产指标的

生产数据和多传感器视觉特征参数, 然后针对不同数据源的特点, 分别建立各生产指标预测子模型和误差补偿子模型. 最后采用信息熵和协调策略分别构建精矿品位和尾矿品位的集成预测模型. 对所建立的集成模型进行工业验证和工况分析, 以说明方法的有效性.

2 铝土矿浮选过程多源数据分析 (Multi-source data analysis in bauxite flotation process)

浮选是利用矿物表面物理化学性质的差异, 在气体、固体和液体三相体系中完成的复杂过程, 其目的是提高磨矿之后矿物的富集程度. 本文以中国某铝土矿浮选流程为例, 进行浮选工艺的分析. 该铝土矿为典型的一水硬铝矿, 具有高硅, 高铝和低铝硅比特点, 通常铝硅比 (Al_2O_3/SiO_2 , A/S) 在5~6. 该浮选流程采用浮选泡沫为精矿, 底流矿浆为尾矿的正浮选工艺, 共包括5种选别作业, 分别是6个粗选槽、2个粗扫槽、4个精II槽、4个精I槽和2个精扫槽, 该铝土矿浮选工艺流程如图1所示.

在矿物浮选流程中, 首先对上游磨矿产生的矿粒进行分级, 分级后的细粒进入浮选槽, 加入水和各种药剂, 在叶轮的搅拌下形成矿浆和气泡, 有用矿粒粘附于气泡上浮并溢流出粗选槽, 经管道进入精选I作业进行一次精选, 同时粗选槽的底流矿浆则进入粗扫槽; 然后精I槽的泡沫溢流至精选II形成最终精选泡沫, 而精选II的底流矿浆则进入精选I, 精选I的底流矿浆进入精扫槽; 精扫槽和粗扫槽的泡沫分别重新返回精I槽和粗选槽, 而它们的底流矿浆汇合流至总尾槽, 形成尾矿.

该浮选过程包括多槽泡沫的溢流和矿浆的底流, 是一个非常复杂的过程. 许多过程参数都会对最终生产指标产生影响, 各槽泡沫状态也能反映最终矿物品位的变化. 下面从影响生产指标的多源数据出发, 分析影响生产指标的生产数据和反映矿物品位的泡沫特征.

2.1 生产数据分析 (Production data analysis)

根据铝土矿浮选工艺学理论^[1, 18], 矿石性质对浮选脱硅有显著影响, 当给矿铝硅比较高时, 矿石可浮性较好, 有用矿物富集程度高, 精矿品位质量好, 反之, 当铝硅比降低至4.0以下时, 矿物分选困难, 富集度低. 磨矿粒度是反映矿浆颗粒大小分布的情况, 当磨矿细度增加时, 对提高精矿质量有利, 但是由于泥化也会影响浮选过程, 当粒度较粗时浮选速度变慢, 矿物单体解离不够充分, 容易脱落进入尾矿, 导致精矿和尾矿指标均不好. 因此对于正浮选脱硅工艺, 通常磨矿粒度应该满足小于0.074 mm的占90%以上. 矿浆浓度对于浮选机的充气量, 药剂消耗以及浮选时间都有显著影响, 浓度过大或过小都会影响生产指标. 通常

铝土矿正浮选一般控制入选浓度在30%左右. 矿浆酸碱度是影响浮选性能的重要因素, 在pH值较大的情况下, 由于矿浆中的 $\text{OH}^{[-]}$ 离子较多, 矿粒表面吸附大量的 $\text{H}^{[+]}$, 导致矿粒表面亲水性增大并阻碍捕收剂阴离子的吸附; 较小时取得相反的效果, 因此正浮选工艺一般控制pH在8~10. 影响工艺指标的操作变量有加药量、矿浆液位以及进气流量, 考虑到矿浆液位和进气流量变化时容易导致生产过程的工况波动, 出现“冒槽”或“跑槽”等生产事故, 所以通常是在保

证矿浆液位和进气流量稳定的情况下, 通过调整浮选药剂来达到目的. 浮选药剂主要是捕收剂, 在铝土矿浮选过程, 当捕收剂用量过多时, 泡沫层致密、粘度较高、泡沫带矿物多, 但也有较多的泥质物夹杂进入精矿, 导致精矿铝硅比降低; 当捕收剂用量不足, 矿物疏水性不够, 使得回收率下降, 所以药剂用量须合适. 综上分析, 影响浮选性能的主要生产数据有给矿品位 G_g 、给矿浓度 G_c 、给矿粒度 G_s 、加药量 D_o 和矿浆pH值 G_{ph} .

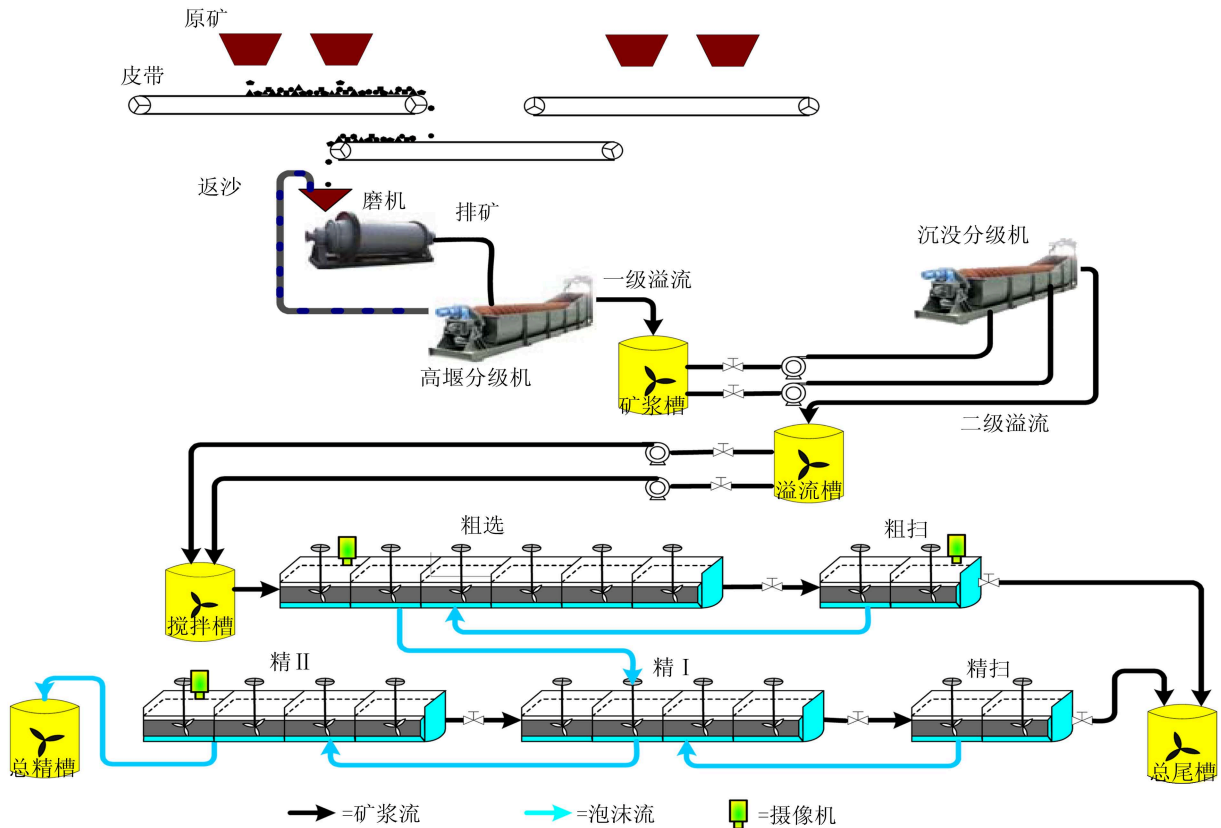


图1 浮选工艺流程图

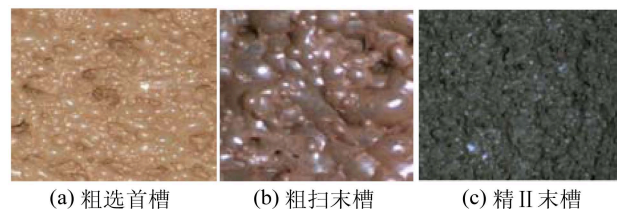
Fig. 1 Flow-diagram of the flotation

2.2 多传感器视觉特征分析(Multi-sensor image features analysis)

浮选泡沫图像的表面视觉特征是反映浮选性能的重要指标^[10, 19-21]. 传统上采用单一浮选槽获取的泡沫特征进行生产指标的表征, 但是由于单一浮选槽的视觉特征不能表征整个浮选流程的泡沫状态, 所以存在描述生产指标误差偏大的问题. 为此本文构建了铝土矿分布式机器视觉系统, 经过现场分析摄像机分别安装在粗选首槽、粗扫末槽和精II末槽上, 具体见图1中. 考虑到工业现场不同浮选槽的视觉特征在一定程度上都可以作为工人判断浮选工况的依据, 因此下面研究表征不同浮选槽的泡沫关键特征以及它们之间的关联性.

2.2.1 不同浮选工序泡沫特征选取(Froth features selection of different flotation stages)

在图1平台上采集不同浮选槽的泡沫图像, 如图2所示, 其中: 图2(a)为粗选首槽图像, 图2(b)为粗扫末槽图像, 图2(c)为精II末槽图像.



(a) 粗选首槽 (b) 粗扫末槽 (c) 精II末槽

图2 不同浮选槽获取的泡沫图像

Fig. 2 Froth images of different flotation stages

从采集的泡沫图像和浮选操作人员提供的信息可知: 粗选槽矿浆中矿物含量低, 泡沫层具有明显的气泡尺寸, 相对稳定, 颜色偏浅, 如图2(a)所示; 粗扫槽中, 矿浆中只含少量的有用矿粒, 泡沫层的气泡承载量低, 泡沫比较虚, 含矿量少, 颜色较浅, 如图2(b)所示; 精Ⅱ槽的气泡矿化程度比较高, 泡沫表现为粘稠状, 尺寸细小, 难以分辨单个泡沫, 颜色较深, 如图2(c)所示.

结合上述定性分析和运用相关系数法计算获得不同浮选槽表征生产指标的关键特征如表1所示.

表 1 不同浮选槽获取的关键图像特征

Table 1 Key image features of different cells

粗选首槽	精Ⅱ末槽	粗扫末槽
红色分量	红色分量	红色分量
流速	流速	流速
稳定性	稳定性	稳定性
大小	承载率	大小
承载率	纹理特征(熵)	承载率

2.2.2 时空配准及变换趋势(Variation trend and time-space registration)

铝土矿浮选过程是一个长流程, 大时滞系统, 不同浮选槽之间具有很大的延迟. 通过对现场实验跟踪可知, 泡沫从粗选首槽溢流到精Ⅱ末槽的时间大约为30 min, 从粗选首槽溢流到粗扫末槽的时间大约为95 min, 从而可以为不同工序的同步性和图像特征的对应性研究提供基础, 进而完成多传感器数据的时空配准. 下面对不同浮选槽获取的泡沫特征进行关联性分析, 其中泡沫特征来自于2011年4月-6月之间连续45天在图1平台上采集的样本数据. 精矿品位和图像特征之间的关系见文献[10, 21], 这里主要分析3个浮选槽图像特征的动态变换关系.

3个浮选槽红色分量均值之间的关系如图3所示, 可知同一泡沫状态在各槽的延迟时间以及颜色表征随着时间的变化关系. 说明各槽的色调值有大致相同的变化趋势, 也反映同一泡沫状态, 粗扫槽(scavenger)的色调平均值最小, 精选(cleaner)槽最大, 粗选(rougher)其次. 同样的方法分析3个不同浮选槽泡沫的流动速度如图4, 可知同一泡沫状态在粗选槽的流动速度最快, 精选其次, 扫选最慢, 随着浮选流程的变化, 各槽的流速有大致相同的变化趋势. 同样也可以分析泡沫的稳定性, 承载率的动态关系. 从图3和图4可知粗选流程采集的特征参数波动比较严重, 而精选流程相对稳定.

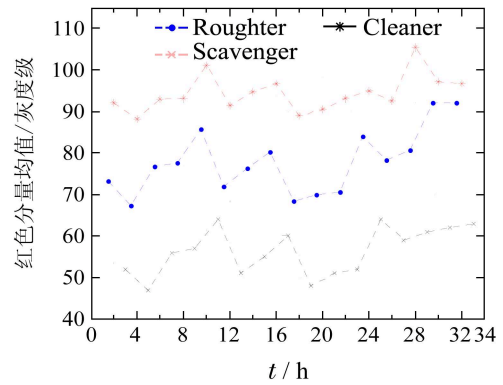


图 3 3槽样本图像色调分布趋势

Fig. 3 Distribution trends in the froth colour of different flotation stages

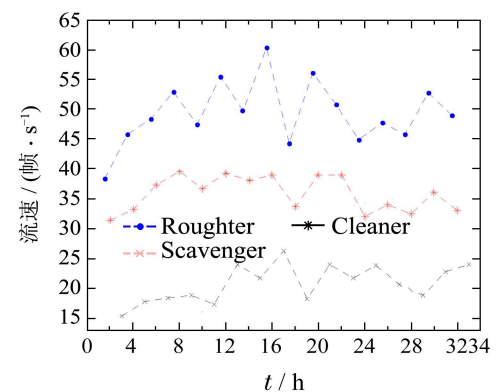


图 4 3槽样本图像速度特征分布趋势

Fig. 4 Distribution trends in the froth speed of different flotation stages

上述分析说明不同浮选流程之间具有相似的变化趋势, 但同时也存在局部点的不确定性, 这些数据分析为后续模型的建立提供依据.

3 生产指标集成建模(Production index integrated modeling)

影响生产指标的多源数据从不同角度表征和反映矿物品位, 并且各数据源内部之间存在超多维、大量冗余、非线性和互补性等问题. 为此本文提出对异源数据集分别构建生产指标预测子模型和同步误差补偿模型, 最后通过信息熵和智能补偿策略两种集成方法完成生产指标的预测.

3.1 建模策略(Modeling strategy)

结合铝土矿浮选过程, 提出了由数据预处理和时空配准模块, 预测子模型和误差补偿子模型模块, 以及加权集成模块共3部分组成的生产指标建模策略, 如图5所示.

数据预处理模块主要是针对工业现场整理的生产数据, 采集计算的图像特征数据, 以及工业化验的生产指标进行时空配准, 具体见第2.2节.

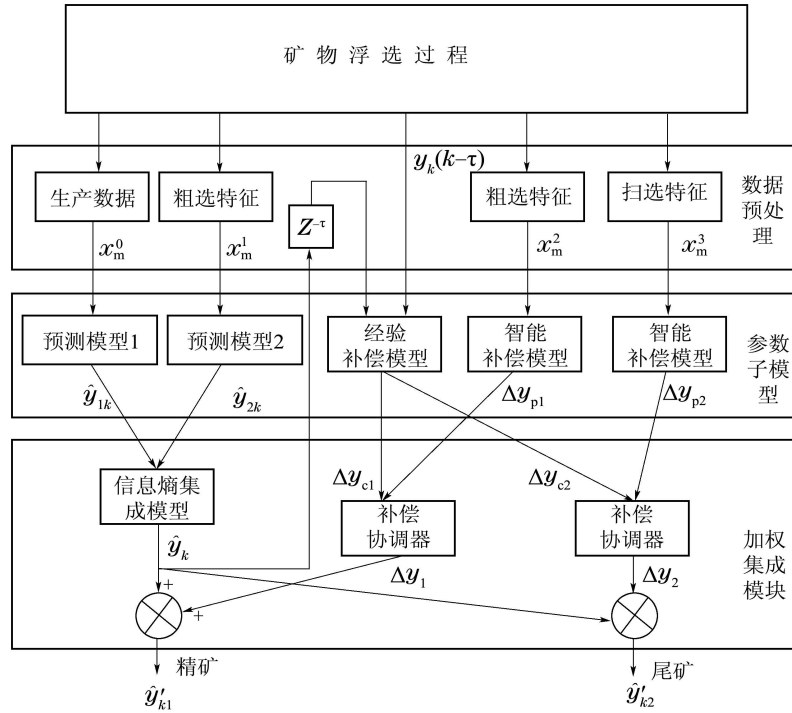


图5 浮选生产指标集成建模策略

Fig. 5 Strategy of production index integrated modeling

预测子模型模块建立生产数据, 粗选特征和浮选生产指标之间的模型; 误差补偿子模型模块建立精选泡沫特征和精矿品位误差之间的关系模型, 扫选特征和尾矿品位误差之间的关系模型。

加权集成模块则实现模型的信息熵集成和智能补偿集成, 最终获得生产指标的输出。

图5中: 用 x_m^0 表示现场生产数据, x_m^1 表示粗选槽采集的泡沫特征, x_m^2 表示精选槽采集的泡沫特征, x_m^3 表示扫选槽采集的泡沫特征, 其中 m 为样本个数, 取值范围为 $m = 1, 2, \dots, 5$. \hat{y}_{1k} 和 \hat{y}_{2k} 分别为两个预测子模型的输出, \hat{y}_k 为集成预测模型的预估值, y_k 为实际品位化验值; Δy_{p1} 和 Δy_{p2} 分别为精矿品位和尾矿品位采用智能补偿模型的输出量; Δy_{c1} 和 Δy_{c2} 分别为精矿品位和尾矿品位采用经验补偿模型的输出量; $\Delta \hat{y}_1$ 和 $\Delta \hat{y}_2$ 分别表示精矿品位和尾矿品位的补偿值; \hat{y}'_{k1} 表示精矿品位集成预测模型的输出, \hat{y}'_{k2} 表示尾矿品位集成预测模型的输出。

3.2 预测子模型(Predictive sub-model)

前面分析的多源数据中, 共包含4种不同的数据源, 用 x_m^0 表示现场生产数据, 具体包括给矿品位、给矿浓度、给矿粒度、加药量以及矿浆pH值; 用 x_m^1 表示粗选槽获取的泡沫图像特征, 包括泡沫红色分量、流速、稳定性、泡沫大小和承载率。这些数据之间存在超高维、大量冗余、非线性和互补性等特点, 常规的PLS方法能够提取数据中与生产指标

相关的潜变量实现数据的降维, 但不能解决过程数据的非线性问题, B样条函数具有较好的非线性拟合性能, 能将非线性问题用拟线性的方法转化为线性化问题^[22], 故本文建立基于B样条的PLS方法, 具体步骤如下:

Step 1 对输入矩阵的每列 $j(1 \leq j \leq p)$ 做如下变换 $x_j \rightarrow z_j$.

1) 定义 $\varepsilon_{j,l-1}$, h_j , M_j 分别为变量 x_j 上划分的区间分点、分段长度和分段个数, 确定 M_j , 计算区间分点 $\varepsilon_{j,l-1}$ 为

$$\varepsilon_{j,l-1} = \min(x_j) + (l-1)h_j, \quad (1)$$

式中: $h_j = \frac{\max(x_j) - \min(x_j)}{M_j}$, $l = 0, 1, \dots, M_j + 2$.

2) 对 $[x_j]_{n \times 1}$ 作3次B样条变换, 有

$$Z_j = \{z_{j,0}, z_{j,1}, \dots, z_{j,M_j+2}\} = \{\Omega_3(\frac{x_j - \varepsilon_{j,l-1}}{h_j}), l = 0, 1, \dots, M_j + 2\}, \quad (2)$$

其中

$$\Omega_3(\frac{x_j - \varepsilon_{j,l-1}}{h_j}) = \frac{1}{3!h_j^3} \sum_{p=0}^4 (-1)^p \binom{4}{p} (x_j - \varepsilon_{j,l-3+p})_+^3.$$

Step 2 对B样条变换后的矩阵进行中心标准化处理, 即

$$\tilde{z}_{j,l}^i = \frac{z_{j,l}^i - \bar{z}_{j,l}}{s_{j,l}}, \quad \tilde{y}_i = \frac{y_i - \bar{y}}{s_y}, \quad (3)$$

其中: $l = 0, 1, \dots, M_j + 2$; $j = 1, 2, \dots, p$; $i = 1, 2, \dots, q$, $\bar{z}_{j,l}$, \bar{y} 分别为 $z_{j,l}$, y 的样本均值, $s_{j,l}$, s_y 分别是 $z_{j,l}$, y 的样本方差.

记上述处理后的输入矩阵为 \tilde{Z} , 输出矩阵为 \tilde{Y} , 则可知经过标准化后输入输出矩阵组成的联合矩阵为

$$\begin{aligned} [X, Y] &= [x_1, x_2, \dots, x_p, y]_{n \times (p+1)} \rightarrow \\ [\tilde{Z}, \tilde{Y}] &= [(\tilde{Z}_1)_{n \times (M_1+3)}, \dots, (\tilde{Z}_p)_{n \times (M_p+3)}, \tilde{Y}] = \\ &[\tilde{z}_{1,0}, \dots, \tilde{z}_{1, M_1+2}, \dots, \\ &\tilde{z}_{p, M_1+2}, \tilde{y}]_{n \times \{(\sum_{j=1}^p (M_j + 3)) + 1\}}, \end{aligned} \quad (4)$$

从而得到新的各列的线性系数表达式为

$$\tilde{y} = \sum_{j=1}^p \sum_{l=0}^{M_j+2} \alpha_{j,l} \tilde{z}_{j,l}. \quad (5)$$

Step 3 对公式(5)在满足交叉有效性原理条件下, 提取最多的PLS成分数, 求得回归系数 $\alpha_{j,l}$ ($j = 1, 2, \dots, p$; $l = 0, 1, \dots, M_j + 2$).

Step 4 将式(3)代入式(5), 即

$$\frac{y - \bar{y}}{s_y} = \sum_{j=1}^p \sum_{l=0}^{M_j+2} \alpha_{j,l} \frac{z_{j,l} - \bar{z}_{j,l}}{s_{j,l}}, \quad (6)$$

得到

$$y = \beta_0 + \sum_{j=1}^p \sum_{l=0}^{M_j+2} \beta_{j,l} z_{j,l}, \quad (7)$$

其中: $\beta_{j,l} = s_y \frac{\alpha_{j,l}}{s_{j,l}}$, $\beta_0 = \bar{y} - \sum_{j=1}^p \sum_{l=0}^{M_j+2} \beta_{j,l} \bar{z}_{j,l}$.

Step 5 将回归系数和式(2)代入式(7), 得到输出与输入之间的非线性模型为

$$\begin{aligned} \hat{y} &= \beta_0 + \sum_{j=1}^p \hat{f}_j(x_j) = \\ &\beta_0 + \sum_{j=1}^p \sum_{l=0}^{M_j+2} \beta_{j,l} \Omega_3\left(\frac{x_j - \varepsilon_{j,l-1}}{h_j}\right). \end{aligned} \quad (8)$$

通过设定分段个数 M_j 和相关参数, 可以采用上述方法完成对生产数据 x_m^0 和粗选特征 x_m^1 的非线性预测模型的建立, 获得预测模型输出 \hat{y}_{1k} 和 \hat{y}_{2k} .

3.3 误差补偿子模型(Error compensation sub-model)

从浮选过程来看, 由于长流程预测模型带入的误差以及未知外界和生产扰动带来的模型误差需要进行补偿. 在泡沫浮选过程中, 粗选槽泡沫波动明显, 特征参数变化大, 而精选槽和扫选槽的特征变化相对比较平缓, 趋于稳定, 为此利用生产中积累

的大量与粗选流程同步的精选特征数据或扫选特征数据为验前知识和指导经验建立同步误差补偿模型. 该补偿模型主要是完成对预测模型的同步误差补偿和提高模型精度. 文中采用正则极限学习机(regularized extreme learning machine, RELM)神经网络^[7]构造生产指标补偿模型, 主要是因为RELM是一种速度快和泛化性能好的单隐层前馈神经网络学习算法, 另外采用小波函数作为激励函数, 可在一定程度上提高网络的收敛速度和降低误差.

将精选槽或扫选槽的泡沫特征作为模型输入, 采用正则极限学习机神经网络构造生产指标补偿模型, 即

$$\Delta y_{pr} = f_{\text{RELM}}(\mathbf{x}_r), \quad r = 1, 2, \quad (9)$$

其中: \mathbf{x}_r 为模型的输入, Δy_{pr} 表示由该模型获得的输出补偿量, $f_{\text{RELM}}(\cdot)$ 表示神经网络拟合的 Δy_{pr} 与 \mathbf{x}_r 之间的关系函数, 当 r 为1时表示精选槽的补偿模型, r 为2时表示扫选槽的补偿模型. 具体包括以下步骤:

Step 1 模型参数选择:

- 1) 以图像特征作为输入变量, 见表1精选特征 x_m^2 或扫选特征 x_m^3 ;
- 2) 以精矿品位或尾矿品位的补偿量作为模型输出, 即 Δy_{p1} 和 Δy_{p2} ;
- 3) 对上述变量进行归一化处理, 以消除不同量纲所带来的影响.

Step 2 构建小波极限学习机神经网络补偿模型, 选取模型参数:

- 1) 选Morlet小波作为隐含层激励函数;
- 针对激励函数的特性, 容易证明Morlet小波符合无限可微的条件, 可以作为激励函数构造RELM. Morlet小波函数表述为

$$h(t) = \cos(1.75t)e^{-2t^2/2}; \quad (10)$$

- 2) 确定参数 γ 和隐含层节点数 L , 随即设置权值向量 a_s 和隐含层阈值 b_s , $s = 1, 2, \dots, L$;
- 3) 计算隐层输出矩阵 H 和权值向量 $\hat{\beta}$;
- 4) 计算网络输出 $Y = H\hat{\beta}$, 更多参数详见文献[23].

3.4 加权集成策略(Integrated weighted method)

浮选流程长, 表层泡沫状态变化大, 仅仅通过粗选槽的泡沫特征不能充分反映精矿品位和尾矿品位的变换, 尤其工况波动时更加难以判断. 为了进一步解决浮选过程工况波动引起模型误差的问题, 系统提取粗选流程的历史同步特征参数, 通过智能残差补偿模型完成对误差的补偿, 同时针对粗选环节提出基于信息熵的补偿策略, 以达期提高模型预测

精度和泛化性.

3.4.1 基于信息熵的集成策略(Integrated strategy using comentropy)

集成模型中的权重因子通常根据操作工人经验或领域专家知识来确定. 由于浮选过程工况参数具有不确定性, 容易出现波动, 为了提高权重因子的可靠性, 本文采用的熵值法^[15]来确定权重因子的取值.

计算图5中基于生产数据的预测模型1和基于粗选特征的预测模型2在 k 时刻矿物品位的预估值 \hat{y}_{1k} 和 \hat{y}_{2k} , 集成预测模型的预估值 \hat{y}_k . 设 e_{nk} 为第 n 个预测模型在 k 时刻的相对误差, 按下式计算得到

$$e_{nk} = \begin{cases} 1 - \hat{y}_{nk}/y_k, & 0 \leq |1 - \hat{y}_{nk}/y_k| < 1, \\ 1, & |1 - \hat{y}_{nk}/y_k| \geq 1, \end{cases} \quad (11)$$

式中: y_k 表示 k 时刻品位实际检测值; $n = 1, 2, \dots, N$, N 为子模型个数, 文中取2; $k = 1, 2, \dots, M$, M 为训练样本个数. 利用熵值法确定加权系数的步骤如下:

Step 1 第 n 个单一预测模型在 k 时刻的预测相对误差的比重 p_{nk} :

$$p_{nk} = e_{nk} / \sum_{n=1}^N e_{nk}. \quad (12)$$

Step 2 第 n 个单一预测模型的熵值 E_n :

$$E_n = -\frac{1}{\ln M} \sum_{k=1}^M p_{nk} \ln p_{nk}. \quad (13)$$

Step 3 第 n 个单一预测模型的相对误差序列的变异程度 d_n :

$$d_n = 1 - E_n. \quad (14)$$

Step 4 各单一预测模型加权系数 w_n :

$$w_n = 1/(N-1)(1 - d_n / \sum_{n=1}^N d_n). \quad (15)$$

Step 5 计算信息熵集成模型的输出

$$\hat{y}_k = \sum_{n=1}^N w_n \hat{y}_{nk}. \quad (16)$$

通过上述步骤可以分别计算得到信息熵集成后的精矿品位和尾矿品位.

3.4.2 智能协调补偿策略(Intelligent coordinated compensation strategy)

选取样本空间分布的图像特征数据进行仿真说明, 智能补偿模型在样本区间内具有较高预测精度, 但对训练样本外的数据, 其预测能力有限. 为了解决样本空间外数据模型误差的问题, 提取历史检测数据通过经验补偿模型对误差进行补偿. 见图5集成模型, 设当前时刻为 k , 对于极限学习机神经网络

模型边界外的输入数据, 该时刻的误差补偿量为

$$\Delta y_{cv}(k) = y_{kv}(k - \tau) - \hat{y}_{kv}(k - \tau), \quad (17)$$

其中: $y_{kv}(k - \tau)$ 为 $k - \tau$ 时刻的品位化验值; $\hat{y}_{kv}(k - \tau)$ 为 $k - \tau$ 时刻集成模型的预测值; $v = 1, 2$ 分别表示精矿品位和尾矿品位; τ 为浮选流程生产过程的滞后时间, 与泡沫流速、取样检测时间等有关.

设 V 为小波极限学习机网络输入变量的取值区间, X 为输入向量. 则智能协调器的工作方式如下:

$$\text{IF } X \in V \text{ THEN } \Delta \hat{y} = \Delta y_{pr};$$

$$\text{IF } X \notin V \text{ THEN } \Delta \hat{y} = \Delta y_{cv}.$$

故可获得集成模型的输出为

$$\hat{y}'_k = \hat{y}_k + \Delta \hat{y}, \quad (18)$$

式中: \hat{y}_k 为集成预测模型的输出, $\Delta \hat{y}$ 为补偿模型的输出值, 可分别计算获得精矿指标和尾矿指标的输出值.

3.5 模型的更新(Model updating)

上述建模策略考虑了浮选过程的多源数据和长流程特性, 但是该模型是纯粹的数据模型, 容易受浮选过程的各种扰动影响, 使获取的数据具有很大的时变特性, 从而导致模型存在一定偏差, 不能很好的反映工况波动情况下的状态. 当工况发生较大波动, 如“跑槽”或“沉槽”时, 该模型将出现严重偏差. 为此, 本文结合上述特性, 采用滑窗方式更新模型参数, 具体步骤如下:

1) 计算模型预测误差. 计算新样本时刻模型的预测误差, 比较其和模型精度阈值之间的关系, 判断是否需要重新训练模型. 这里根据实际浮选现场需求, 精度阈值在 $\pm 5\%$ 以内较好.

2) 样本的取舍处理. 若模型需重新训练, 对新采集的数据样本与原训练样本做相关度检测, 若相关度大, 则可认为新样本未带入新的信息, 则舍去最新样本, 否则添加新样本. 当新样本加入到下一阶段的训练中时, 需要删除最老的一个训练样本. 实验中采用后台训练方式, 使模型随着系统的运行也能保持较高的预测精度.

4 模型的验证及工况分析(Model validation and industrial analysis)

4.1 模型的验证(Validation of the model)

为了保证多源数据和生产指标的时空一致性, 对同一时刻的矿浆取样、制样等流程及方法测定泡沫化学参数, 并制作成份化验样本, 获取矿浆浓度、矿物粒度、矿物品位和矿浆pH等. 采集泡沫图像的各个特征参数与此时的工艺指标参数的组合构成标准输入输出样本.

为了验证本文所提出的集成预测模型的性能, 选取2011年4-10月的实验数据共计580组数据样本, 该数据每8小时轮班采集化验, 分别进行模型的参数辨识和验证模型的性能. 实验中选取分段个数 $M_j = 3$, 对模型1和模型2的数据进行样条函数处理, 完成非线性问题转换成线性化问题, 然后用PLS的线性映射建立预测模型. 经实验测试选取模型更新时滑窗大小 $N = 100$ 进行模型的训练, 精度阈值为0.04. 基于RELM算法的单隐层前馈神经网络模型的生产指标补偿模型, 其参数设置如下: 输入层节点为5, 输出节点为2, 训练次数1000, $\gamma = 0.1$. 隐含层神经元的个数由经验公式确定, 采用RELM算法获得隐含层输出矩阵和最终输出.

通过基于信息熵的权值计算, 精矿品位的最优加权系数分别为 $w_{11} = 0.537$, $w_{12} = 0.463$; 尾矿品位的最优加权系数分别为 $w_{21} = 0.598$, $w_{22} = 0.402$; 进而可以获得当前 k 时刻信息熵集成模型的输出 \hat{y}_{k1} 和 \hat{y}_{k2} . 通过同步补偿模型, 最终获得本文集成预测模型的输出精矿品位 \hat{y}'_{k1} 和尾矿品位 \hat{y}'_{k2} , 其预测效果分别如图6-9所示.

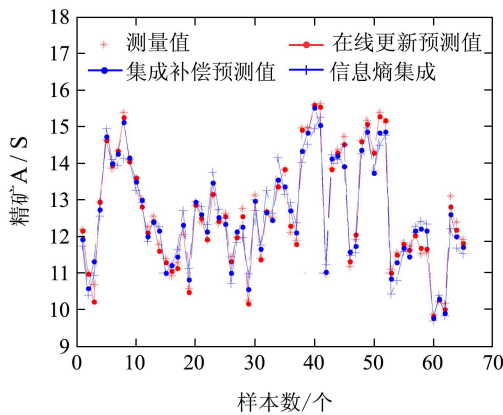


图6 精矿品位集成模型预测结果比较
Fig. 6 Comparison of integrated predicted concentrate grade vs. actual value

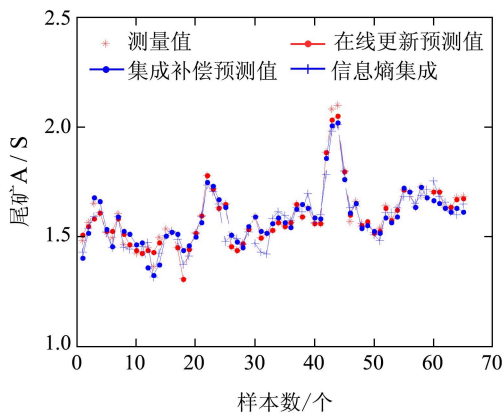


图7 尾矿品位集成模型预测结果比较
Fig. 7 Comparison of integrated predicted tailing grade vs. actual value

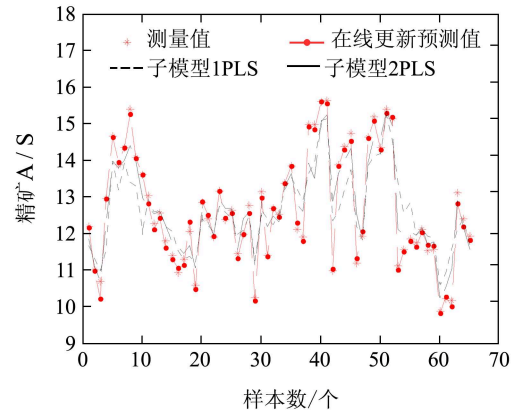


图8 精矿品位在线集成补偿预测模型与PLS模型比较
Fig. 8 Comparison of on-line integrated predicted concentrate grade vs. PLS model

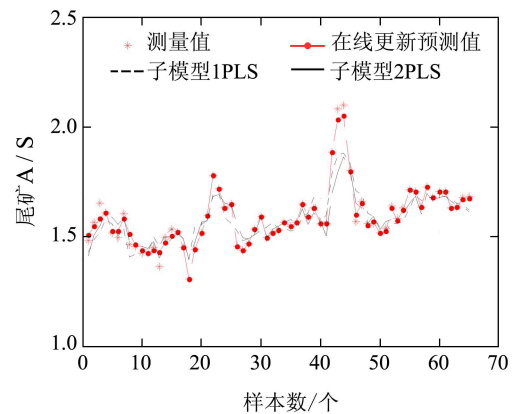


图9 尾矿品位在线集成补偿预测模型与PLS的比较
Fig. 9 Comparison of on-line integrated predicted tailing grade vs. PLS model

图6和图7中给出的曲线是本文提出的方法(信息熵集成模型, 信息熵集成补偿预测模型以及在线更新集成补偿预测模型)与实际值的比较, 这几种方法是结合浮选工艺, 在传统预测模型的基础上进行了一定的集成, 补偿以及在线修正后的结果, 整体上比较逼近实际值; 但是当浮选工况变化明显时, 在线更新集成补偿模型的跟踪预测性能能够适应工况变化, 逼近程度好于其他两种方法. 图8和图9给出了本文的在线集成补偿模型与传统PLS方法的比较(其中子模型1是基于生产数据的PLS模型, 子模型2是基于粗选特征的PLS模型), 由图8-9可知, 本文方法的逼近性能明显高于传统预测方法.

为了与本文提出的方法进行比较, 建立了基于生产数据的BPLS模型, 基于粗选特征的BPLS模型, 基于生产数据的PLS传统模型以及基于粗选特征的PLS模型. 计算其最大相对误差 $E_{R\max}$ 和均方根误差 (root mean square error, RMSE) 及相对误差 E 在5%之内的数据点比例, 结果如表2-3所示.

表2 精矿品位预测误差分析

Table 2 Predictive error analysis of concentrate grade

网络模型	相对误差在5%以内	$E_{R\max}/\%$	RMSE
在线更新集成补偿模型	94.87%	5.31	0.6032
集成补偿模型	94.10%	6.15	0.6741
信息熵集成	92.97%	8.27	1.5481
模型1BPLS	76.62%	11.74	3.7243
模型2BPLS	82.26%	15.62	2.9507
模型1 PLS	76.02%	19.73	4.6212
模型2 PLS	78.81%	16.40	3.2929

表3 尾矿品位预测误差分析

Table 3 Predictive error analysis of tailing grade

网络模型	相对误差在5%以内	$E_{R\max}/\%$	RMSE
在线更新集成补偿模型	92.05%	7.63	1.0204
集成补偿模型	90.63%	8.46	1.2604
信息熵集成	87.07%	16.72	3.7729
模型1BPLS	78.32%	19.21	4.5192
模型2BPLS	79.96%	16.63	3.1834
模型1 PLS	72.02%	21.64	4.3281
模型2 PLS	79.18%	17.37	3.9765

由表2和表3可以看出: 1) 本文的在线更新集成补偿模型的RMSE和 $E_{R\max}$ 最小, 说明该模型的精度较高, 预测值能很好的跟踪实际值. 2) 本文比较了信息熵集成模型与子模型1、子模型2的预测性能, 说明集成策略融合了不同输入源的特点, 使模型精度获得明显提高, 精矿品位相对误差控制在5%以内的比例由76.62%, 82.26%提高到92.97%. 3) 表中还比较了BPLS和PLS方法预测性能, 表明生产数据间或泡沫特征间存在非线性关系, BPLS方法使得模型的预测效果改善.

本文集成模型从影响矿物品位的不同输入源考虑, 分别构建了预测子模型和补偿模型, 可以很好的兼顾不同工况下生产指标的波动, 提高模型预测精度, 满足生产需求.

4.2 浮选工况状态判别(Discrimination of flotation state)

在给矿条件范围内, 铝土矿选厂要求达到的浮选精矿/尾矿指标分别为 $y_1^* \geq 10$ 和 $y_2^* \leq 1.8$, 即选厂期望值. 综合上述指标建立浮选过程工况评价指标如下:

$$Q = k_1 \left(\frac{y_1 - y_1^*}{y_1^*} \right)^2 + k_2 \left(\frac{y_2 - y_2^*}{y_2^*} \right)^2, \quad (19)$$

其中: Q 表示铝土矿浮选过程综合工况评价指标, y_1, y_1^* 分别表示精矿品位的预测值与选厂期望值; y_2, y_2^* 表示尾矿品位的预测值与选厂期望值; k_1 和 k_2 分别表示精矿品位和尾矿品位对综合工况评价指标的影响因子, 大小可通过专家经验确定, 且

$$\sum_{i=1}^2 k_i = 1. \quad (20)$$

工况判断模型根据计算获得的 Q 值将铝土矿浮选过程的综合工况分为: 优($0 \leq Q \leq 0.05$)、一般($0.05 \leq Q \leq 0.2$)、恶劣($Q \geq 0.2$)3个等级. 利用上述方法计算模型测试样本(图6和图7)中数据样本的工况分析如表4, 这里取 $k_1 = 0.55$ 和 $k_2 = 0.45$, 说明与现场实验时工人判断基本一致.

表4 工况状态判别

Table 4 Discrimination of flotation state

工况	优	一般	恶劣
Q	48	14	3
工人判断	51	11	3

然后根据工况判别情况, 进行后续操作指导. 如果工况为优, 则继续使用当前的操作参数控制策略; 如果工况一般, 则对各浮选流程进行判断, 完成分级协调控制; 如果工况恶劣, 则采用现场工人手动控制.

5 结论(Conclusions)

本文依据浮选矿物品位与入矿参数之间的相关性, 以及工业现场领域专家判别矿物品位的经验, 提出了基于多源数据的浮选生产指标集成建模方法, 主要工作和结论如下: 1) 分析了浮选过程的工艺机理和影响浮选性能的多源数据; 通过全局关联性分析说明不同浮选槽的泡沫特征具有很强的关联性和随机性. 2) 提出了基于多源数据的生产指标在线集成建模策略, 其模型的最大相对误差 $E_{R\max}$ 和RMSE最小, 精矿/尾矿品位相对误差控制在5%以内的比例为分别为94.87%和92.05%, 明显高于其他方法, 说明本文的在线集成建模策略的预测精度高、泛化性好. 3) 构建浮选工况判别模块, 为后续研究浮选过程的优化控制建立基础.

通过集成模型的建立, 为浮选矿物品位的预测提供了一种新的思路, 工况状态判别模型的构建, 为研究浮选流程的优化控制提供基础. 但在模型的通用性、在线更新性能, 以及模型的优化控制上还有待更进一步的研究.

参考文献(References):

- [1] 周长春. 铝土矿及其浮选技术 [M]. 徐州: 中国矿业大学出版社, 2011.

- (ZHOU Changchun. *Bauxite and Flotation Technology* [M]. Xuzhou: China University of Mining Technology Press, 2011.)
- [2] LIU J J, MACGREGOR J F. Froth-based modeling and control of flotation processes [J]. *Minerals Engineering*, 2008, 21(9): 642 – 651.
- [3] XU C, GUI W, YANG C, et al. Flotation process fault detection using output PDF of bubble size distribution [J]. *Minerals Engineering*, 2012, 26: 5 – 12.
- [4] NEETHLING S J. Simple approximations for estimating froth recovery [J]. *International Journal of Mineral Processing*, 2008, 89(1): 44 – 52.
- [5] GONZALEZ G D, ORCHARD M, CERDA J L, et al. Local models for soft-sensors in a rougher flotation bank [J]. *Minerals Engineering*, 2003, 16(5): 441 – 453.
- [6] DAYAL B S, MACGREGOR J F. Recursive exponentially weighted PLS and its applications to adaptive control and prediction [J]. *Journal of Process Control*, 1997, 7(3): 169 – 179.
- [7] 李海波, 柴天佑, 岳恒. 浮选工艺指标KPCA-ELM软测量模型及应用 [J]. 化工学报, 2012, 63(9): 2892 – 2898.
(LI Haibo, CHAI Tianyou, YUE Heng. Soft sensor of technical indices based on KPCA-ELM and application for flotation process [J]. *CIESC Journal*, 2012, 63(9): 2892 – 2898.)
- [8] COLE K E, WATERS K E, FAN X, et al. Combining positron emission particle tracking and image analysis to interpret particle motion in froths [J]. *Minerals Engineering*, 2010, 23(11): 1036 – 1044.
- [9] HARGRAVE J M, MILES N J, HALL S T. The use of grey level measurement in predicting coal flotation performance [J]. *Minerals Engineering*, 1996, 9(6): 667 – 674.
- [10] KAARTINEN J, H?T?NEN J, HY?TYNIEMI H, et al. Machine vision based control of zinc flotation-a case study [J]. *Control Engineering Practice*, 2006, 14(12): 1455 – 1466.
- [11] MARAIS C, ALDRICH C. Estimation of platinum flotation grades from froth image data [J]. *Minerals Engineering*, 2011, 24(5): 433 – 441.
- [12] DIETTERIEG T. Machine-learning research: four current directions [J]. *The Artificial Intelligence Magazine*, 1997, 18(4): 97 – 136.
- [13] HANSEN L K, SALAMON P. Neural network ensembles [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(10): 993 – 1001.
- [14] GUI W H, YANG C H, CHEN X F, et al. Modeling and Optimization Problems and Challenges Arising in Nonferrous Metallurgical Processes [J]. *Acta Automatica Sinica*, 2013, 39(3): 198 – 207.
- [15] 汤健, 柴天佑, 赵立杰, 等. 基于振动频谱的磨矿过程球磨机负荷参数集成建模方法 [J]. 控制理论与应用, 2012, 29(2): 183 – 191.
(TANG Jian, CHAI Tianyou, ZHAO Lijie, et al. Ensemble modeling for parameters of ball-mill load in grinding process based on frequency spectrum of shell vibration [J]. *Control Theory & Applications*. 2012, 29(2): 183 – 191.)
- [16] 陈晓方, 桂卫华, 王雅琳, 等. 基于智能集成策略的烧结块残硫软测量模型 [J]. 控制理论与应用, 2004, 21(1): 75 – 80.
(CHEN Xiaofang, GUI Weihua, WANG Yalin, et al. Soft-sensing model of sulfur content in agglomerate based on intelligent integrated strategy [J]. *Control Theory & Applications*, 2004, 21(1): 75 – 80.)
- [17] 王雅琳, 桂卫华, 阳春华, 等. 基于有限信息的铜吹炼动态过程智能集成建模 [J]. 控制理论与应用, 2009, 26(8): 860 – 866.
(WANG Yalin, GUI Weihua, YANG Chunhua, et al. Intelligent integrated modeling for the dynamic copper-converting process based on limited data information [J]. *Control Theory & Applications*, 2009, 26(8): 860 – 866.)
- [18] 耿增显, 柴天佑. 基于LS-SVM的浮选过程工艺技术指标软测量 [J]. 系统仿真学报, 2008, 20(23): 6321 – 6324.
(GENG Zengxian, CHAI Tianyou. Soft sensor of technical indices based on LS-SVM for flotation process [J]. *Journal of System Simulation*, 2008, 20(23): 6321 – 6324.)
- [19] REDDICK J F, HESKETH A H, MORAR S H, et al. An evaluation of factors affecting the robustness of colour measurement and its potential to predict the grade of flotation concentrate [J]. *Minerals Engineering*, 2009, 22(1): 64 – 69.
- [20] YANG C, XU C, MU X, et al. Bubble size estimation using interfacial morphological information for mineral flotation process monitoring [J]. *Transactions of Nonferrous Metals Society of China*, 2009, 19(3): 694 – 699.
- [21] 周开军, 阳春华, 牟学民, 等. 基于图像特征提取的浮选关键参数智能预测算法 [J]. 控制与决策, 2009, 24(9): 1300 – 1305.
(ZHOU Kaijun, YANG Chunhua, MOU Xuemin, et al. Intelligent prediction algorithm for floatation key parameters based on image features extraction [J]. *Control and Decision*, 2009, 24(9): 1300 – 1305.)
- [22] UNSER M, ALDROUBI A, EDEN M. B-spline signal processing: part I-theory [J]. *IEEE Transactions on Signal Processing*, 1993, 41(2): 821 – 833.
- [23] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications [J]. *Neurocomputing*, 2006, 70(1): 489 – 501.

作者简介:

曹斌芳 (1979–), 女, 博士研究生, 目前研究方向为复杂工业过程建模与优化、图像信息处理, E-mail: cao.bf163.com;

谢永芳 (1972–), 男, 教授, 博士生导师, 研究方向为复杂工业过程建模与优化, E-mail: yfxiemail.csu.edu.cn;

阳春华 (1965–), 女, 教授, 博士生导师, 主要研究方向为复杂工业过程建模与优化研究、图像处理等, E-mail: ychhmail.csu.edu.cn;

桂卫华 (1950–), 男, 教授, 博士生导师, 中国工程院院士, 主要研究方向为复杂工业过程建模、控制与优化等, E-mail: gwhmail.csu.edu.cn;

王晓丽 (1981–), 女, 副教授, 主要研究方向为复杂工业过程建模与优化控制, E-mail: xlwangmail.csu.edu.