

# 基于两层分块 GMM-PRS 的流程工业过程运行状态评价

邹筱瑜<sup>1</sup> 王福利<sup>1</sup> 常玉清<sup>1</sup> 郑伟<sup>2</sup>

**摘要** 过程运行状态评价旨在实时判断运行性能优劣程度,并追溯导致非优运行状态的原因,指导操作人员进行生产调整,保证企业经济效益。因此,对过程运行性能优劣评价的研究具有重要的理论和应用价值。本文针对定量、定性变量共存的流程工业过程运行状态评价问题,提出基于两层分块混合模型的评价方法。将流程工业过程根据其物理特性和管理方向划分子块,产生子块层和全流程层。在定量信息占主导地位的子块内,建立定量的高斯混合模型(Gaussian mixture model, GMM)。在定性信息占主导地位的子块内,建立定性概率粗糙集(Probabilistic rough set, PRS)模型。综合各子块运行状态信息,进一步判定全流程运行状态等级。针对非优运行状态等级,本文提出基于贡献率的非优原因追溯方法,在非优子块内进行原因追溯。最后,将所提方法应用于某黄金湿法冶炼生产过程,说明所提方法的可行性和有效性。

**关键词** 过程运行状态评价, 流程工业过程, 高斯混合模型, 概率粗糙集, 黄金湿法冶炼

**引用格式** 邹筱瑜, 王福利, 常玉清, 郑伟. 基于两层分块 GMM-PRS 的流程工业过程运行状态评价. 自动化学报, 2019, 45(11): 2071–2081

**DOI** 10.16383/j.aas.2018.c170412

## Plant-wide Process Operating Performance Assessment Based on Two-level Multi-block GMM-PRS

ZOU Xiao-Yu<sup>1</sup> WANG Fu-Li<sup>1</sup> CHANG Yu-Qing<sup>1</sup> ZHENG Wei<sup>2</sup>

**Abstract** Process operating performance assessment judges operating performance optimal degree online, and identifies the causes for non-optimal performance to guide the production adjustment for operators. Therefore, the research on process operating performance assessment is of great significance in both theory and practical applications. To solve the plant-wide process operating performance assessment problem with coexistence of quantitative and qualitative variables, a two-level multi-block hybrid model based assessment approach is proposed in this article. According to the physical property and management direction, a plant-wide process is classified into multiple sub-blocks. Hence, there are sub-block level and global level. In a sub-block that is dominated by quantitative information, a Gaussian mixture model (GMM) is established. Accordingly, in a qualitative information dominated sub-block, a probabilistic rough set (PRS) is built. Based on the sub-block-level performance grade information, the global-level performance grade can be further judged. For the non-optimal performance grade, cause identification is implemented in the non-optimal sub-block. A contribution rate based non-optimal cause identification method is developed in this research. In the end, to illustrate the feasibility and validity, the proposed technique is applied to a gold hydrometallurgy process.

**Key words** Process operating performance assessment, plant-wide process, Gaussian mixture model (GMM), probabilistic rough set (PRS), gold hydrometallurgy process

**Citation** Zou Xiao-Yu, Wang Fu-Li, Chang Yu-Qing, Zheng Wei. Plant-wide process operating performance assessment based on two-level multi-block GMM-PRS. *Acta Automatica Sinica*, 2019, 45(11): 2071–2081

过程运行状态评价在过程安全运行的前提下,综合考虑了产品质量、物耗、能耗、经济收益等因

素,对过程运行性能优劣进行进一步评价,包括过程运行状态优性在线评价和非优运行状态原因追溯两部分<sup>[1–5]</sup>。优性在线评价实时判断运行性能优劣程度,非优原因追溯诊断导致非优运行状态的原因,指导操作人员进行生产调整。理想的运行状态有助于提高综合经济效益和生产效率、降低生产成本。因此,对过程运行性能优劣评价的研究具有重要的理论和应用价值。

传统的过程运行性能评价方法可分为两类:基于定量信息的评价方法和基于定性信息的评价方法,其中,定量信息指用数值大小描述的变量信息,定性信息指定性描述的变量信息,主要通过语义变

收稿日期 2017-07-24 录用日期 2018-02-07  
Manuscript received July 24, 2017; accepted February 7, 2018  
国家自然科学基金(61533007, 61374146, 61374147, 61673092)资助  
Supported by National Natural Science Foundation of China (61533007, 61374146, 61374147, 61673092)  
本文责任编辑 邓方  
Recommended by Associate Editor DENG Fang  
1. 东北大学信息科学与工程学院 沈阳 110819 2. 国家电网辽宁省电力有限公司 沈阳 110819  
1. College of Information Science and Engineering, Northeastern University, Shenyang 110819 2. Liaoning Electric Power Co. LTD., State Grid, Shenyang 110819

量来描述. 基于定量信息的评价方法处理以定量信息为主的过程性能评价问题. 多元统计方法是一种应用最广泛的定量评价方法, 适用于过程先验知识较少的过程<sup>[6-8]</sup>. Liu 等提出了基于主成分分析法 (Principal component analysis, PCA) 和多集合主成分分析法 (Multi-set PCA, MsPCA) 的运行状态优性在线评价方法<sup>[9-10]</sup>, 但此类方法并没有考虑过程变量与评价指标之间的关系. 基于指标预测的评价方法, 虽避免了此问题, 但所需数据量非常庞大<sup>[1-2]</sup>. 概率框架下的性能评价方法, 如: 基于高斯混合模型 (Gaussian mixture model, GMM)<sup>[3, 10]</sup> 和贝叶斯理论 (Bayesian theory)<sup>[11-12]</sup> 的评价方法, 已广泛应用于性能评价中. 基于概率理论的评价方法需要先验知识辅助确定概率密度函数. 不同于经典方法过于严苛的要求, 智能评价方法, 如基于人工神经网络 (Artificial neural network, ANN) 的评价方法, 由于其学习能力和非线性处理能力强, 受到研究者的青睐<sup>[13-14]</sup>. 但是, 此类方法容易陷入局部最优值, 可能出现过拟合现象. 基于定性信息的评价方法处理以定性信息为主的过程性能评价问题. 最常用的处理定性信息的方法有贝叶斯网 (Bayesian network, BN)、模糊理论 (Fuzzy theory) 和粗糙集理论 (Rough set, RS) 等. BN 通过建立表示因果关系的网络和概率表来进行性能评价, BN 的构建通常需要大量过程因果知识<sup>[15-16]</sup>. 模糊理论通过隶属度函数来进行评价, 但隶属度函数和判定阈值的选取尚无严格的理论指导<sup>[17-18]</sup>. RS 在保持分类能力不变的前提下, 对数据表进行约简, 去除冗余信息, 提取启发式规则, 进行评价<sup>[19]</sup>. 但经典 RS 并未考虑数据与目标概念之间的覆盖关系, 因此, 概率粗糙集 (Probabilistic rough set, PRS) 应运而生<sup>[20-22]</sup>. PRS 定义了等价类与目标概念的隶属程度, 以后验概率的形式量化数据与目标概念之间的覆盖关系.

定量方法的优势在于: 精度高, 能够建立变量之间的相关性, 预测性能较好, 是提取过程内部特性的方法, 适用于变量测量准确的过程. 但是, 传统定量方法解释性差, 在样本数目少时, 可能会出现病态的模型. 定性方法的优势在于: 解释性强, 可以处理不精确的信息, 模型建立容易, 适用于过程存在定性信息的过程. 但是, 传统定性方法精度低, 要求数据类型覆盖所有可能的运行情况, 是提取过程外部特性的方法, 预测性能较差. 若采用定性方法处理定量变量, 需要将定量信息离散化, 在信息离散化过程中, 会损失有效信息, 降低评价精度. 综上所述, 定量与定性方法各有优、劣势, 相辅相成. 在实际流程工业生产中, 既有定量变量, 又有定性变量. 由于定性、定量变量共存的问题, 传统评价方法难以直接应用.

实际工业过程还可能面临一个巨大的挑战, 即

流程工业特性. 流程工业过程生产流程长, 规模庞大, 变量数目巨大, 变量相关性复杂. 一个流程工业生产过程, 通常包含若干生产单元. 同一个生产单元内, 变量强耦合, 不同生产单元间, 变量弱耦合. 生产过程从前至后, 依序进行, 每一个生产单元的生产时间不尽相同. 因此, 将传统的评价方法直接应用于流程工业过程, 常常难以得到令人满意的准确率. 流程工业过程生产周期长、变量众多、机理复杂, 难以建立准确的全局模型. 最常用的处理流程工业特性的方法就是将过程根据物理特性划分层次和子块, 这种措施已广泛应用于安全性能评价中<sup>[23-24]</sup>. Macgregor 等<sup>[25]</sup> 和 Jiang 等<sup>[26]</sup> 分别提出了分块的多元统计方法和分块的概率论方法, 来处理流程工业过程性能评价问题. 相比于分块方法, 分层的性能评价方法更注重子块之间的相关性<sup>[27]</sup>. 在分层或分块性能评价思想的基础上, 研究者在质量预测<sup>[28]</sup>、自适应<sup>[29]</sup> 等方向进行了进一步探索. 但目前对流程工业过程优性评价的研究还较少. 传统的分层分块性能评价方法难以直接应用于实际流程工业过程运行状态评价中, 主要原因如下: 1) 全流程的评价问题难以分解为子块的评价问题; 2) 子块的优性难以定义; 3) 未考虑定量、定性变量共存问题.

本文提出一种基于两层分块混合模型的流程工业过程运行状态评价方法. 横向上, 将流程工业过程, 根据其物理特性划分子块, 将联系紧密的设备或生产环节划分至同一子块内, 将联系相对较弱的设备或生产环节划分至不同子块; 纵向上, 形成两个评价层次即子块层和全流程层. 本文所提两层分块方法与传统方法不同之处在于, 所提方法能够评价子块的优劣程度, 不需要显式的全流程模型即可评价全流程的运行状态, 并快速定位非优的子块. 这种灵活的分层分块评价方式, 为混合模型的建立提供了便利. 在一个子块内, 综合考虑评价精度需求、定量和定性变量的比例、模型建立的复杂度, 来选择定量或者定性方法进行建模和评价. 不失一般性的, 本文假设: 以定量信息为主的子块, 采用 GMM 进行建模, 获取子块内各运行状态等级数据分布的概率密度函数; 以定性信息为主的子块, 采用 PRS 进行建模, 得到子块内各运行状态等级的推理规则; 于是, 可以建立两层分块 GMM-PRS (Gaussian mixture model-probabilistic rough set) 模型. 该混合模型的优势在于, 根据子块的数据特性, 灵活地选用恰当的评价方法, 可减少有效信息的损失, 保证方法的有效性. 最后, 本文将所提基于两层分块 GMM-PRS 模型的评价方法应用于国内某黄金湿法冶炼过程中, 验证其有效性. 此外, 综合经济效益是目前应用最广泛的过程运行状态评价指标之一, 本文采用综合经济效益为全流程运行状态评价指标.

## 1 基本方法简介

### 1.1 GMM 简介

高斯分布是一种常见的数据分布, 若高维空间点的分布近似为椭球体, 则可用单一高斯密度函数来描述这些数据的分布特性.

令  $\mathbf{x} \in \mathbf{R}^{1 \times J}$  是服从高斯分布的  $J$  维过程数据, 该类数据的概率密度函数可以用高斯函数表示:

$$g(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{J}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \times \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})(\boldsymbol{\Sigma})^{-1}(\mathbf{x} - \boldsymbol{\mu})^T\right] \quad (1)$$

其中, 参数  $\boldsymbol{\theta} = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ ,  $\boldsymbol{\mu}$  为该类数据的均值向量,  $\boldsymbol{\Sigma}$  为协方差矩阵<sup>[11]</sup>. 这些参数的取值决定了概率密度函数的特性, 如函数的中心点、宽窄和形状等.

一些过程数据不服从高斯分布, 但可以用高斯混合模型描述其分布特性. 假设该过程数据分布包含  $N$  个高斯分量, 第  $n$  个高斯分量的概率密度函数表示为  $g(\mathbf{x}|\boldsymbol{\theta}^n)$ , 其先验概率为  $\omega^n$ ,  $n = 1, 2, \dots, N$ . 则此过程概率密度函数为:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{n=1}^N \omega^n g(\mathbf{x}|\boldsymbol{\theta}^n) \quad (2)$$

数据  $\mathbf{x}$  属于各高斯分量的概率可用贝叶斯理论求得:

$$P(C^n|\mathbf{x}) = \frac{\omega^n g(\mathbf{x}|\boldsymbol{\theta}^n)}{\sum_{n'=1}^N \omega^{n'} g(\mathbf{x}|\boldsymbol{\theta}^{n'})} \quad (3)$$

其中,  $C^n$  表示第  $n$  个高斯成分.

### 1.2 PRS 简介

RS 是一种在不需过程先验知识的情况下进行推理的方法, 针对定性数据, 可进行高效、准确的推理. 但是, RS 存在没考虑子集间相关性和定义过于严格的问题. 因此, PRS 应运而生.

令  $U$  为目标的非空有限集合,  $U$  称作论域,  $A$  为一个有限的属性集合,  $R$  是  $A$  的一个子集. 对于任意  $\mathbf{x} \in U$ , 定义  $\mathbf{x}$  在  $R$  上的等价类  $[\mathbf{x}]_R$  为  $[\mathbf{x}]_R = \{\mathbf{y} \in U | \forall a \in R, f(\mathbf{x}, a) = f(\mathbf{y}, a)\}$ , 其中,  $f(\mathbf{x}, a)$  为  $\mathbf{x}$  在属性  $a$  上的取值. 给定一个非空子集  $X \subseteq U$  和一个等价类  $[\mathbf{x}]_R$ , 可以计算如下概率:

$$P(X|[\mathbf{x}]_R) = \frac{|[\mathbf{x}]_R \cap X|}{|[\mathbf{x}]_R|} \quad (4)$$

其中,  $|S|$  表示集合  $S$  的基, 即  $S$  中的元素个数.  $P(X|[\mathbf{x}]_R)$  表示  $[\mathbf{x}]_R$  中,  $X$  的覆盖程度.

给定阈值  $\alpha$  和  $\beta$ , 针对  $0 \leq \beta < \alpha \leq 1$  的情况,  $X$  的下近似、上近似、 $R$  边界域定义为:

$$\begin{cases} R_\alpha(X) = \{\mathbf{x} \in U | P(X|[\mathbf{x}]_R) \geq \alpha\} \\ \bar{R}_\beta(X) = \{\mathbf{x} \in U | P(X|[\mathbf{x}]_R) > \beta\} \\ BN_R(X) = \{\mathbf{x} \in U | \alpha > P(X|[\mathbf{x}]_R) > \beta\} \end{cases} \quad (5)$$

针对  $\alpha = \beta \neq 0$  的情况,  $X$  的下近似、上近似、 $R$  边界域定义为:

$$\begin{cases} R_\alpha(X) = \{\mathbf{x} \in U | P(X|[\mathbf{x}]_R) > \alpha\} \\ \bar{R}_\alpha(X) = \{\mathbf{x} \in U | P(X|[\mathbf{x}]_R) \geq \alpha\} \\ BN_R(X) = \{\mathbf{x} \in U | P(X|[\mathbf{x}]_R) = \alpha\} \end{cases} \quad (6)$$

$X$  的下近似中, 包含所有一定属于  $X$  的元素;  $X$  的上近似中, 包含所有可能属于  $X$  的元素. 如果边界域  $BN_R(X)$  为空, 那么  $X$  称为精确集; 否则,  $X$  称为粗糙集. 如果取  $\alpha = 1$ 、 $\beta = 0$ , PRS 退化为传统 RS.

## 2 两层分块 GMM-PRS 模型的建立

根据流程工业特性, 本文提出两层分块评价结构, 并根据每个子块的数据特性, 分别用定量或定性方法, 建立子块评价模型.

### 2.1 两层分块 GMM-PRS 结构

过程运行状态的优劣通常可反映在综合经济指标 (Comprehensive economic index, CEI) 上, CEI 越高, 运行状态越好, CEI 也成为了广泛接受的运行状态优性评价指标<sup>[23]</sup>. 传统评价方法对过程变量  $\mathbf{x}$  和评价指标 CEI 可建立一个单模型:

$$CEI = f(\mathbf{x}) \quad (7)$$

为了降低流程工业过程运行状态评价问题的规模、提高模型解释性, 本文提出如图 1 所示的两层分块结构; 并且, 对以定量和定性信息为主的子块, 分别进行定量和定性建模.

在子块层, 一个流程工业过程根据其过程特性, 划分为多个有物理意义的子块. 子块内, 变量相关性强; 子块间, 变量相关性弱. 在全流程层, 提取各子块间的相关性. 因此, 两层分块结构增强了模型解释性, 减少了问题规模, 降低了建模难度, 削弱了对子块性能无关变量的影响, 放大了对子块性能有关变量的影响.

将过程进行横向分块、纵向分层后, 得到了两层分块结构, 再根据每个子块的特性建立相应评价模型, 为全流程的评价提供基础. 按照子块的数据特性选择适当的建模方法, 保证了模型的有效性和精度. 高斯分布是一种常见的数据分布, 若高维空间点的

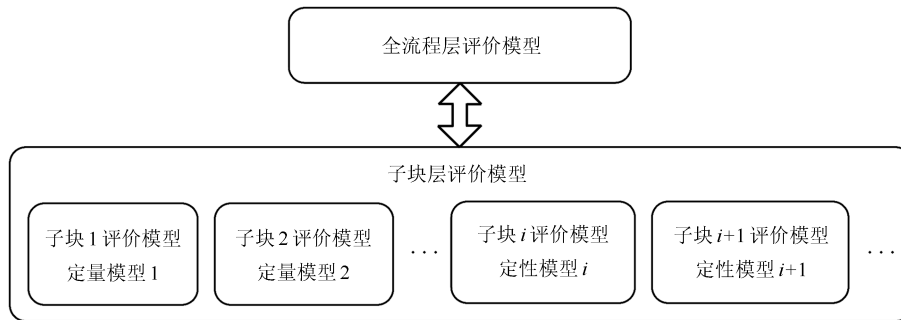


图 1 两层分块混合模型结构示意图

Fig.1 The illustration of the two-level multi-block hybrid model structure

分布近似为椭球体, 则可用单一高斯密度函数来描述这些数据的概率密度函数. 针对以定量信息为主的单模态过程, 同一运行状态等级的定量数据分布特性相似, 近似服从单高斯分布, 可视为所有定量数据分布的一个高斯成分. 过程中, 定性变量的数目和状态种类都较少, 因此, 定性变量可能出现状态组合种类不会很多, 其分布可以用历史数据进行学习. 而对于定性信息占主导地位的过程, 定性变量数目多, 相应定性状态的组合数目也会大幅增大. 那么, 基于 GMM 的方法, 会面临组合爆炸、建模数量庞大的问题. 也就是说, 以定性变量为主的过程, 不再适合用基于概率分布的方法来进行评价. Pawlak 教授提出的 RS 理论是一种在不确定性存在的前提下, 进行推理的方法, 现已广泛应用于安全性评价和风险评价等领域. 为改进 RS 无法处理不一致规则的问题, 概率粗糙集 PRS 方法应运而生. PRS 是一种具有严格理论支撑并且应用广泛的定性信息处理方法, 因此本文采用 PRS 对以定性信息为主的子块进行建模. 值得注意的是, 可选的子块建模方法并不局限于 GMM 和 PRS 方法.

### 2.1.1 运行状态等级确定

本文所述定性信息指用语义变量对变量状态进行描述的信息, 定量信息指用数值大小描述的变量信息. 建模数据中, 定量变量以变量取值的形式表示, 定性变量以变量状态等级序号的形式表示, 如温度的高、中、低三种状态, 分别对应状态等级 1、2、3. 其中, 定性变量状态等级只与变量幅值大小趋势相关, 与性能优劣无关. 本文所使用的定量数据是经过平滑处理后的数据, 定性数据根据其物理意义划分为了不同状态, 并用一系列连续的正整数对状态等级进行区分. 其中, 平滑处理的原理是用一定长度的滑动窗口内数据的平均水平来代表该滑动窗口的信息, 这种预处理方法在一定程度上克服了噪声的影响, 使此均值更能反映滑动窗口内的主要信息. 针对变化快速或噪声水平较低的过程, 为避免过程动态

特性被淹没, 滑动窗口长度不宜过长. 针对变化缓慢或噪声水平较高的过程, 为减少系统正常波动导致的误评价, 滑动窗口长度不宜过短.

### 2.2 基于两层分块 GMM-PRS 的离线建模

根据过程知识, 建立两层分块结构之后, 基于两层分块 GMM-PRS 的评价模型离线建立包括运行状态等级确定和模型建立, 其中, 模型建立分定量 GMM 和定性 PRS 模型.

针对一个复杂的流程工业过程, 在划分单元子块之后, 如果各子块存在独立的评价指标, 那么可以对每个子块建立以子块生产指标为评价标准的模型, 再在子块生产指标的基础上, 进行全流程的运行状态评价. 但是, 本文旨在解决无子块评价指标的流程工业过程运行状态优性评价问题. 此时, 子块的优性定义变得十分困难. 本文利用唯一的评价指标—全流程综合经济指标 CEI, 作为子块运行状态等级划分标准. 企业通常会有一定周期内对 CEI 进行估算并作为生产考核的标准, 但是估算周期比较长, 无法直接根据该估算结果实时指导生产. 所以, 我们可以将 CEI 视作定性变量, 只需确定其在一定周期内的定性状态. 根据全流程 CEI 的定性状态, 过程运行状态被划分为若干等级, 等级数目通常由过程评价的精度需求、过程检测情况等因素共同决定. 本文假设全流程综合经济指标包含  $N$  个状态, 对应的全流程运行状态等级由 1 至  $N$ , 优性依次降低. 在划分子块后, 就一个子块的一类运行情况而言, 定义此类运行情况下所能达到的最好全流程运行状态等级为这种运行情况下该子块的运行状态等级. 从另一个角度看, 该运行情况下, 当其他子块都处于最好匹配状态时, 该子块使全流程所能达到的最好等级代表了该子块所处运行情况的极限最好情况, 是子块所处运行情况固有特性的一种体现. 如图 2 所示, 建模数据的运行状态等级离线确定方法包括以下三个步骤: 1) 数据块划分, 2) 全流程层等级确定,

3) 子块层等级确定.

1) 数据块划分

令建模数据为  $\mathbf{X} \in \mathbf{R}^{H \times J}$ ,  $H$  表示样本个数,  $J$  表示变量个数. 根据变量和子块之间的关系, 将建模数据  $\mathbf{X}$  划分为  $I$  个子块, 用  $\mathbf{X}_i \in \mathbf{R}^{H \times J_i}$  ( $i = 1, 2, \dots, I$ ) 表示第  $i$  个子块的建模数据,  $J_i$  为第  $i$  个子块的变量数目.

2) 全流程层等级确定

根据全流程评价指标 CEI, 过程运行状态被划分为若干等级, 如: 优/良/中/差等. 那么, 每一个子块数据  $\mathbf{X}_i$ , 可以根据全流程评价指标 CEI, 划分为不同等级, 记为  $\tilde{\mathbf{X}}_i^1, \tilde{\mathbf{X}}_i^2, \dots, \tilde{\mathbf{X}}_i^N$ , 其中,  $\tilde{\mathbf{X}}_i^n$  表示子块  $i$  中全流程等级为  $n$  的数据,  $i = 1, 2, \dots, I$ ,  $n = 1, 2, \dots, N$ ,  $I$  为子块数目,  $N$  为全流程等级数目.

无论其他子块是否运行于最优匹配状态, 全流程层运行状态都不可能达到最优等级. 因此, 一个子块数据的子块层运行状态等级定义为: 该子块内相同数据所能达到的最好全流程层运行状态等级. 所以, 全流程层的等级数目和子块层的等级数目相等. 假设运行状态, 等级 1 到  $N$  的优性依次降低. 确定子块层运行状态等级  $n$  中数据的具体做法为: 以全流程层等级  $n, n+1, \dots, N$  中数据  $\tilde{\mathbf{X}}_i^n$  为基础, 将等级  $n+1, n+2, \dots, N$  中与等级  $n$  中数据相似度大于阈值  $\varepsilon$  的数据, 从原来的等级中转移至等级  $n$  的数据集中, 更新后的等级  $n$  中的数据为子块层运行状态等级为  $n$  的数据, 记为  $\mathbf{X}_i^n$ , 更新后的等级  $n+1, n+2, \dots, N$  中的数据为确定下一等级数据的基础. 两条数据的相似程度定义如下:

$$\text{sim}(\mathbf{x}_1, \mathbf{x}_2) = 1 - \frac{1}{J'} \sum_{j=1}^{J'} d(x_{1,j}, x_{2,j}) \quad (8)$$

其中,

$$d(x_{1,j}, x_{2,j}) = \begin{cases} \left| \frac{x_{1,j} - x_{2,j}}{x_j^{\max} - x_j^{\min}} \right| & (\text{定量变量}) \\ \frac{|x_{1,j} - x_{2,j}|}{A_j - 1} & (\text{定性变量}) \end{cases} \quad (9)$$

$x_{1,j}$  ( $x_{2,j}$ ) 是  $\mathbf{x}_1$  ( $\mathbf{x}_2$ ) 的第  $j$  个变量; 若第  $j$  个变量为定量变量,  $x_j^{\max}$  ( $x_j^{\min}$ ) 是该变量的工艺最大值 (最小值); 若第  $j$  个变量为定性变量,  $|x_{1,j} - x_{2,j}|$  表示  $x_{1,j}$  和  $x_{2,j}$  对应定性状态等级的等级差值的绝对值;  $A_j$  是第  $j$  个变量的状态等级数目,  $J'$  为变量数目.

根据上述三个步骤, 子块数据的全流程层等级和子块层等级能相应确定. 由于上述等级划分规则, 并不需要建立显式的全流程层模型. 全流程层运行状态等级由子块层中最劣的子块运行状态等级决定, 原因将在第 4 节的全流程运行状态在线评价方法中阐述.

2.2.1 定量建模

针对以定量信息为主的子块, 将每一个运行状态等级的数据作为一个高斯分量, 分别建立单高斯模型, 拟合各等级数据的概率密度函数. 但由于少数定性变量的存在, 无法直接建立高斯模型.

假设  $\mathbf{x}$  来自于第  $i$  个子块的第  $n$  个等级, 即  $\mathbf{x} \in \mathbf{X}_i^n$ . 令  $\mathbf{x} = [\tilde{\mathbf{x}}, b \pm \tilde{\mathbf{x}}]$ , 其中,  $\tilde{\mathbf{x}}$  表示定性变量,  $\tilde{\mathbf{x}}$  表示定量变量.  $\tilde{\mathbf{x}}$  根据定性变量状态的不同, 存在多种组合形式. 以定量信息为主的子块所含定性变量数目少, 一个定性变量的状态种类一般不会很多, 不是每一种理论上存在的定性组合都会在实际应用中出现. 所以,  $\tilde{\mathbf{x}}$  中可能出现的定性状态组

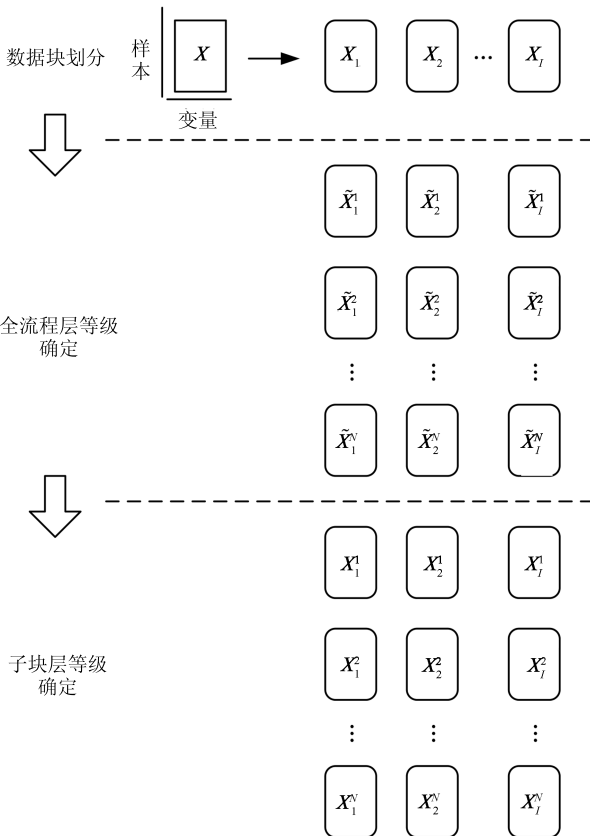


图 2 离线数据划分示意图

Fig. 2 The diagram of data processing

3) 子块层等级确定

由于全流程层运行状态等级不能单独取决于一个子块的运行状态, 所以, 相似的子块数据可能被标记了不同的全流程层运行状态等级. 对于一个子块: 若该子块运行于最优运行状态, 并且其他子块运行于最优匹配状态时, 全流程层运行状态可能达到最优等级; 若该子块运行于一个非优运行状态中,

合种类一般不会很多. 用  $(\bar{\mathbf{X}}_i^n)_k$  表示  $\mathbf{X}_i^n$  中, 定性变量对应为  $(\tilde{\mathbf{x}}_i^n)_k$  的样本的定量变量部分, 其中,  $k = 1, 2, \dots, K$ ,  $K$  为定性变量组合的数目. 假设  $\bar{\mathbf{x}} \in (\bar{\mathbf{X}}_i^n)_k$ , 则针对定性变量为  $(\tilde{\mathbf{x}}_i^n)_k$  的情况, 第  $i$  个子块第  $n$  个等级的概率密度函数可用高斯函数表示为:

$$g[\bar{\mathbf{x}} | (\boldsymbol{\theta}_i^n)_k, (\tilde{\mathbf{x}}_i^n)_k] = \frac{1}{(2\pi)^{\frac{J_i}{2}} |(\bar{\boldsymbol{\Sigma}}_i^n)_k|^{\frac{1}{2}}} \times \exp\left[-\frac{1}{2}(\bar{\mathbf{x}} - (\bar{\boldsymbol{\mu}}_i^n)_k)(\bar{\boldsymbol{\Sigma}}_i^n)_k^{-1}(\bar{\mathbf{x}} - (\bar{\boldsymbol{\mu}}_i^n)_k)^T\right] \quad (10)$$

参数  $(\bar{\boldsymbol{\theta}}_i^n)_k = \{(\bar{\boldsymbol{\mu}}_i^n)_k, (\bar{\boldsymbol{\Sigma}}_i^n)_k\}$ ,  $(\bar{\boldsymbol{\mu}}_i^n)_k$  和  $(\bar{\boldsymbol{\Sigma}}_i^n)_k$  分别是均值和协方差,  $i = 1, 2, \dots, I$ ,  $n = 1, 2, \dots, N$ .

此外, 还需要通过历史数据统计以下概率:

$$\begin{cases} P[(\boldsymbol{\theta}_i^n)_k] = \frac{Num[(\bar{\mathbf{X}}_i^n)_k]}{\sum_{n=1}^N Num[(\bar{\mathbf{X}}_i^n)_k]} \\ P[(\tilde{\mathbf{x}}_i^n)_k] = \frac{\sum_{n=1}^N Num[(\bar{\mathbf{X}}_i^n)_k]}{\sum_{k=1}^K \sum_{n=1}^N Num[(\bar{\mathbf{X}}_i^n)_k]} \\ P[(\tilde{\mathbf{x}}_i^n)_k | (\boldsymbol{\theta}_i^n)_k] = \frac{Num[(\bar{\mathbf{X}}_i^n)_k]}{\sum_{k=1}^K Num[(\bar{\mathbf{X}}_i^n)_k]} \end{cases} \quad (11)$$

其中,  $Num[\boldsymbol{\psi}]$  表示矩阵  $\boldsymbol{\psi}$  中的样本个数.

一旦确定上述参数, 给定一个样本  $\mathbf{x} = [\tilde{\mathbf{x}}, \bar{\mathbf{x}}] \in \mathbf{X}_i^n$ , 可以计算如下概率:

$$\begin{aligned} P[\mathbf{X}_i^n | \mathbf{x}] &= P[(\boldsymbol{\theta}_i^n)_k | (\tilde{\mathbf{x}}_i^n)_k, \bar{\mathbf{x}}] = \\ &= \frac{Pr[(\boldsymbol{\theta}_i^n)_k, \bar{\mathbf{x}} | (\tilde{\mathbf{x}}_i^n)_k] Pr[(\tilde{\mathbf{x}}_i^n)_k]}{Pr[\bar{\mathbf{x}} | (\tilde{\mathbf{x}}_i^n)_k] Pr[(\tilde{\mathbf{x}}_i^n)_k]} = \\ &= \frac{Pr[(\boldsymbol{\theta}_i^n)_k | (\tilde{\mathbf{x}}_i^n)_k] Pr[\bar{\mathbf{x}} | (\tilde{\mathbf{x}}_i^n)_k, (\boldsymbol{\theta}_i^n)_k]}{Pr[\bar{\mathbf{x}} | (\tilde{\mathbf{x}}_i^n)_k]} = \\ &= \frac{P[(\tilde{\mathbf{x}}_i^n)_k | (\boldsymbol{\theta}_i^n)_k] \times P[(\boldsymbol{\theta}_i^n)_k] \times g[\bar{\mathbf{x}} | (\tilde{\mathbf{x}}_i^n)_k, (\boldsymbol{\theta}_i^n)_k]}{P[(\tilde{\mathbf{x}}_i^n)_k] \sum_{n=1}^N P[(\boldsymbol{\theta}_i^n)_k] g[\bar{\mathbf{x}} | (\tilde{\mathbf{x}}_i^n)_k, (\boldsymbol{\theta}_i^n)_k]} \end{aligned} \quad (12)$$

### 2.2.2 定性建模

针对以定性信息为主的子块, 采用 PRS 进行建模. 基于 PRS 的离线建模包含以下三个主要步骤:

1) 数据预处理; 2) 决策表组织; 3) 属性约简.

1) 数据预处理

PRS 是一种以定性或离散数据为基础的推理方法, 因此, 需要将数据进行相应预处理. 针对定性数据, 为后文计算方便, 将变量的各定性状态用一系列整数表示. 针对以定性信息为主的子块中少数的定量数据, 需要将定量数据进行离散化, 得到一系列离散数值. 离散化处理方法很多, 如等距离划分、等频率划分、Naive Scaler 算法、基于断点重要性的离散化算法、基于属性重要性的离散化算法等<sup>[30]</sup>.

2) 决策表组织

决策表的每一列表示一个属性, 每个属性的取值被划分为若干离散状态. 通常, 属性可分为条件属性和决策属性. 决策表每一行代表论域中的一个元素和一种推理规则. 以子块内过程变量为条件属性, 以子块层运行状态等级为决策属性, 分别建立各子块决策表.

3) 属性约简

属性约简目的在于简化决策表, 在保持分类能力不变的前提下, 删除对决策没有影响的条件属性. 常用的属性约简方法有: 一般约简算法、基于差别矩阵和逻辑运算的属性约简算法、归纳属性约简算法等<sup>[30]</sup>.

## 3 基于两层分块 GMM-PRS 的过程运行状态在线评价和非优原因追溯

基于两层分块 GMM-PRS 的过程运行状态在线评价方法, 先在子块层, 对各个子块分别进行评价, 再在全流程层, 综合各子块信息得到最终评价结果. 针对非优运行状态, 在非优的子块内进行原因追溯.

### 3.1 子块层的运行状态在线评价

用  $\mathbf{x}_t$  表示  $t$  时刻子块  $i$  的数据. 若子块  $i$  为以定量变量为主的子块,  $\mathbf{x}_t$  可分解为  $[\tilde{\mathbf{x}}_t, \bar{\mathbf{x}}_t]$ .  $\mathbf{x}_t$  中的定性变量取值  $\tilde{\mathbf{x}}_t$  与  $(\tilde{\mathbf{x}}_i^n)_k$  的相似度为  $\text{sim}(\tilde{\mathbf{x}}_t, (\tilde{\mathbf{x}}_i^n)_k)$ . 针对等级  $n$ , 如果  $\max\{\text{sim}[\tilde{\mathbf{x}}_t, (\tilde{\mathbf{x}}_i^n)_k], k = 1, 2, \dots, K\}$  小于一个事先定义的判定阈值  $\delta$  ( $0 < \delta \leq 1$ ), 那么认为  $\mathbf{x}_t$  不可能处于此等级, 令  $P[\mathbf{X}_i^n | \mathbf{x}_t] = 0$ . 否则,  $P[\mathbf{X}_i^n | \mathbf{x}_t]$  可以根据式 (12) 获得.

若子块  $i$  为以定性变量为主的子块, 在子块内, 首先, 对定量变量进行离散化. 然后, 从历史数据中得到  $\mathbf{x}_t$  的等价类  $[\mathbf{x}_t]_R$ , 其中,  $R$  为条件属性集合. 最后, 根据式 (4) 计算  $[\mathbf{x}_t]$  属于子块层第  $n$  等级的概率:

$$P(\mathbf{X}_i^n | \mathbf{x}_t) = \frac{|[\mathbf{x}_t]_R \cap \mathbf{X}_i^n|}{|[\mathbf{x}_t]_R|} \quad (13)$$

其中,  $n = 1, 2, \dots, N$ ,  $i = 1, 2, \dots, I$ .

$t$  时刻, 子块  $i$  所处运行状态等级  $\hat{L}_{t,i}$  为:

$$\hat{L}_{t,i} = \arg \max_n \{P(\mathbf{X}_i^n | \mathbf{x}_t), n = 1, 2, \dots, N\} \quad (14)$$

### 3.2 全流程运行状态在线评价

在获得所有子块的子块层运行状态等级后, 全流程层运行状态等级  $\hat{L}_t$  与子块层最劣的子块运行状态等级相等. 假设全流程层等级 1 至  $N$ , 优性依次递减. 那么全流程层运行状态等级  $\hat{L}_t$  表示为:

$$\hat{L}_t = \max \{ \hat{L}_{t,i}, i = 1, 2, \dots, I \} \quad (15)$$

显然, 根据子块等级定义, 由于各子块运行状态等级被定义为相似度大于阈值  $\varepsilon$  的同类数据所能达到的历史最好全流程层等级, 所以全流程层运行状态等级不可能比任何一个子块的运行状态等级更优. 也就是说, 全流程层运行状态等级不会比子块层最劣的子块运行状态等级更优. 从另一个角度看, 如果全流程层运行状态等级比子块层最劣的子块运行状态等级更劣, 说明全流程层运行状态等级比所有子块运行状态等级都更差. 这种情况在实际生产中较少出现, 大部分子块运行状态应与全流程运行状态相一致. 所以, 定义全流程层运行状态等级为子块层最劣运行状态等级, 如式 (15) 所示.

### 3.3 非优原因追溯

如果全流程层运行状态等级  $\hat{L}_t$  为非优运行状态, 那么需要在非优的子块内进行原因追溯. 若子块  $i$  所处运行状态等级  $\hat{L}_{t,i}$  为非优等级, 那么在该子块内追溯非优原因变量. 当前非优数据与最优等级数据进行差异度的比较, 差异大的变量为非优原因变量.

针对以定量变量为主的子块, 用  $(\mathbf{x}_i^{\text{opt}})_k = [(\bar{\mathbf{x}}_i^{\text{opt}})_k, (\bar{\boldsymbol{\mu}}_i^{\text{opt}})_k]$  代表最优运行状态的特性, 其中,  $k = 1, 2, \dots, K$ . 计算非优数据  $\mathbf{x}_t$  与  $(\mathbf{x}_i^{\text{opt}})_k$  的相似度  $\text{sim}[\mathbf{x}_t, (\mathbf{x}_i^{\text{opt}})_k]$ . 于是, 可以得到与  $\mathbf{x}_t$  相似度最高的数据, 记为  $(\mathbf{x}_i^{\text{opt}})_*$ . 定义  $\mathbf{x}_t$  和  $(\mathbf{x}_i^{\text{opt}})_*$  中第  $j$  个变量的差异度为  $d[x_{t,j}, (x_i^{\text{opt}})_{*,j}]$ , 计算方法与式 (9) 相同. 定义贡献率

$$CR_{t,j} = \frac{d[x_{t,j}, (x_i^{\text{opt}})_{*,j}]}{\sum_{j'=1}^{J_i} d[x_{t,j'}, (x_i^{\text{opt}})_{*,j'}]} \quad (16)$$

其中,  $\sum_{j=1}^{J_i} CR_{t,j} = 1$ . 贡献率  $CR_{t,j}$  较大的变量为非优原因变量.

针对以定性变量为主的子块, 直接在最优运行状态等级历史数据中, 查找与当前非优数据  $\mathbf{x}_t$  相似

度最高的数据, 记为  $(\mathbf{x}_i^{\text{opt}})_*$ . 与上述追溯方法类似, 用式 (16) 中的公式, 计算变量贡献率, 贡献率较大的属性为非优属性.

## 4 基于两层分块 PRS-GMM 的流程工业过程运行状态评价方法在黄金湿法冶炼过程中的应用

湿法冶金过程是现代工业生产中金属富集、分离与提取的重要手段和技术. 湿法冶金, 又称之为化学冶金, 是相对于火法冶金和电解法冶金而言, 一种利用液相环境的特点, 通过一定的化学反应, 进行目标金属的提炼和萃取的技术. 黄金湿法冶炼通过液相环境, 将矿石中固相的金, 浸出至矿浆中, 形成液相的金氰络合物离子. 在浸出子块中, 氰化钠是一种重要的添加药剂, 并通过影响浸出率来影响综合经济效益. 然后, 通过洗涤进行固液分离, 得到矿渣和富含金氰络合物离子的贵液. 其中, 贵液经过锌粉, 发生置换反应, 得到金泥. 在置换环节中, 锌粉的添加量和质量对运行状态影响较大.

本文将所提评价方法应用于国内某黄金湿法冶炼过程中, 该过程可划分为五个子块: 第一次浸出、第一次洗涤、第二次浸出、第二次洗涤和第二次置换, 分别对应子块层的五个子块. 两浸两洗的工艺设置是为了提高浸出率. 黄金湿法冶炼过程是一个复杂的流程工业过程, 同时包含定量和定性变量. 第一次浸出和第二次浸出子块以定量信息为主, 因此, 用 GMM 对这两个子块进行建模. 第一次洗涤和第二次洗涤子块以定性信息为主, 因此, 用 PRS 对这两个子块进行建模. 至于置换子块, 定性和定量信息大量共存, 没有某种信息占主导地位的现象. 但是, 其影响优性的关键变量均为定性变量, 故采用 PRS 对置换子块进行建模.

选取 36 个过程变量, 列于表 1. 根据综合经济效益先将黄金湿法冶炼全流程层运行状态划分为优、中、差 3 个等级, 分别对应等级 1、2、3. 从湿法冶金仿真平台, 选取 3000 组数据进行离线建模, 其中, 每个等级的数据各 1000 组, 建模数据量充分, 可以建立准确的离线模型. 然后, 确定每个子块各个等级所包含的数据, 即确定数据子块层等级. 根据 5 个子块的特性, 分别建立 GMM 或 PRS 模型. 设置相似度判定阈值  $\varepsilon = 0.9$ .

重新选取 400 组数据进行在线测试, 实验设计如表 2 所示. 在实验 1 中, 前 100 组数据运行状态等级为优 (等级 1), 后 100 组数据由于氰化钠添加量 2 (子块 3, 定量) 不足, 导致运行状态等级变为差 (等级 3). 在实验 2 中, 前 100 组数据运行状态等级为优, 后 100 组数据由于锌粉添加量 (子块 5, 定性)

表 1 过程变量列表  
Table 1 The process variable list

序号	指标名称	数据类型	位置
1	第一次浸出前矿石固金品位	定性变量	一次浸出
2	第一次浸出前矿浆浓度	定性变量	一次浸出
3	第一次浸出前调浆水量	定量变量	一次浸出
4	第一次浸出调浆后矿浆流量	定性变量	一次浸出
5	第一次浸出氰化钠添加量	定量变量	一次浸出
6	第一次浸出后氰根离子浓度	定量变量	一次浸出
7	第一次浸出充气量	定量变量	一次浸出
8	第一次浸出溶氧浓度	定量变量	一次浸出
9	第一次浸出后金氰络合物离子浓度	定量变量	一次浸出
10	第一次洗涤前矿浆浓度	定性变量	一次洗涤
11	第一次洗涤前矿浆流量	定性变量	一次洗涤
12	第一次洗涤后贵液流量	定性变量	一次洗涤
13	第一次洗涤后滤饼流量	定性变量	一次洗涤
14	第一次洗涤后金氰络合物离子浓度	定性变量	一次洗涤
15	第二次浸出前矿石固金品位	定性变量	二次浸出
16	第二次浸出前矿浆浓度	定性变量	二次浸出
17	第二次浸出前调浆水量	定性变量	二次浸出
18	第二次浸出调浆后矿浆流量	定量变量	二次浸出
19	第二次浸出氰化钠添加量	定性变量	二次浸出
20	第二次浸出后氰根离子浓度	定量变量	二次浸出
21	第二次浸出充气量	定量变量	二次浸出
22	第二次浸出溶氧浓度	定量变量	二次浸出
23	第二次浸出后金氰络合物离子浓度	定量变量	二次浸出
24	第二次洗涤前矿浆浓度	定性变量	二次洗涤
25	第二次洗涤前矿浆流量	定性变量	二次洗涤
26	第二次洗涤后贵液流量	定性变量	二次洗涤
27	第二次洗涤后滤饼流量	定性变量	二次洗涤
28	第二次洗涤后金氰络合物离子浓度	定量变量	二次洗涤
29	置换前贵液金氰络合物离子浓度	定量变量	置换
30	脱氧塔压力 1	定量变量	置换
31	脱氧塔压力 2	定量变量	置换
32	脱氧塔压力 3	定量变量	置换
33	置换前贵液流量	定性变量	置换
34	锌粉添加量	定性变量	置换
35	锌粉平均粒径	定性变量	置换
36	金泥品位	定性变量	置换

表 2 实验设计  
Table 2 The experiment design

实验	描述
1	前 100 组数据运行状态等级为优 (等级 1), 后 100 组数据由于第二次浸出氰化钠添加量 (子块 3, 定量) 不足, 导致运行状态等级变为差 (等级 3).
2	前 100 组数据运行状态等级为优 (等级 1), 后 100 组数据由于锌粉添加量 (子块 5, 定性) 过量, 导致运行状态等级变为中 (等级 2).

过量, 导致运行状态等级变为中. 其中, 在实际生产中, 锌粉添加量只能获得 8 小时的累积量, 因此, 在本实验中, 将该变量作为定性变量进行处理. 用所提方法进行运行状态在线评价和非优原因追溯, 以验证其有效性.

实验 1 的评价结果如图 3 所示, 在评价时间内: 子块 1、2、4、5 都处于等级 1; 子块 3 前 100 个评价点处于等级 1, 后 100 个评价点处于等级 3; 全流程运行状态评价结果为前 100 个评价点处于等级 1, 后 100 个评价点处于等级 3. 显然, 评价结果与实际运行状态等级设置一致. 从第 101 个评价点起, 全流程运行状态等级为非优等级, 导致全流程非优的是子块 3, 第二次浸出子块. 因此, 在子块 3 中, 进行非优原因追溯. 第 101 个评价点的非优原因追溯结果如图 4 所示, 显示非优原因变量有第二次浸出氰化钠添加量、第二次浸出后氰根离子浓度和第二次浸出后金氰络合物离子浓度. 事实上, 由于第二次氰化钠添加量不足, 会导致第二次浸出后氰根离子和金氰络合物离子浓度下降. 因此, 非优原因追溯结果与实际情况一致.

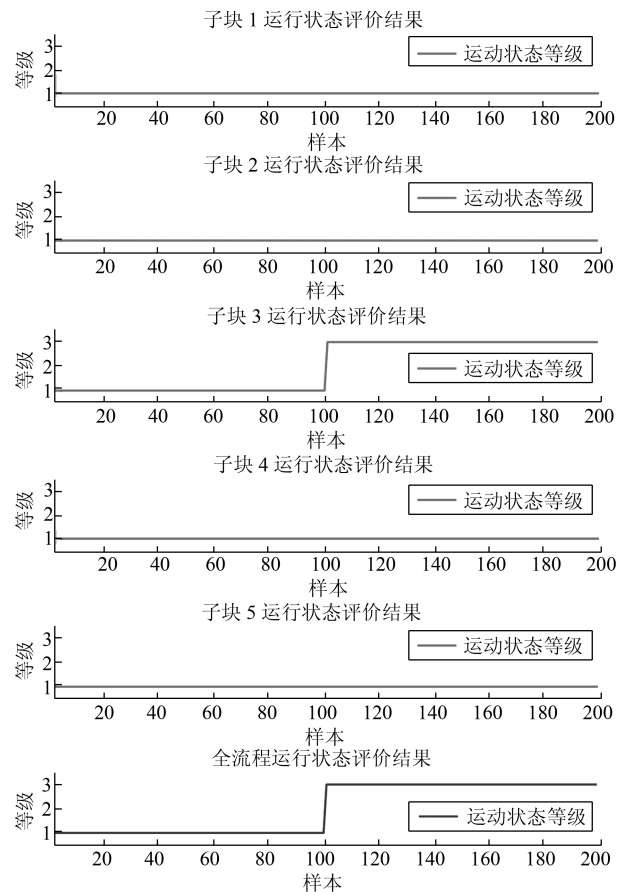


图 3 实验 1 运行状态等级在线评价结果  
Fig. 3 The assessment result in Case 1

实验 2 的评价结果如图 5 所示, 在评价时间内:



子块 1、2、3、4 都处于等级 1; 子块 5 前 100 个评价点处于等级 1, 后 100 个评价点处于等级 2; 相应的, 全流程评价结果为前 100 个评价点处于运行状态等级 1, 后 100 个评价点处于等级 2. 评价结果与实际运行状态等级设置一致. 然后, 在非优的子块中, 进行原因追溯. 非优的子块为置换子块, 是一个定性建模的子块, 其追溯结果展示于图 6. 实际上, 锌粉添加量过多, 置换率已无法再提升, 反而会增加置换物耗, 降低总经济效益, 导致运行状态等级变为非优.

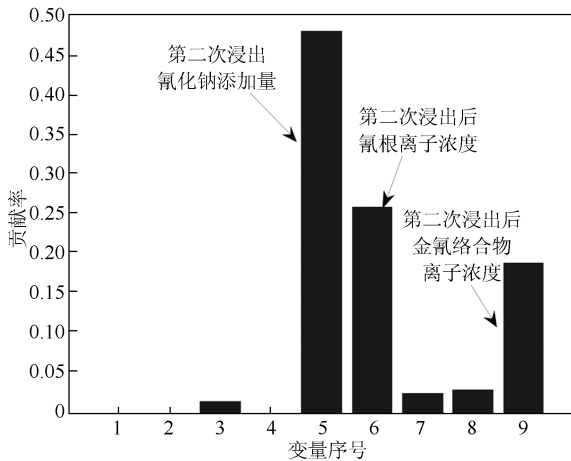


图 4 实验 1 子块 3 中非优原因追溯结果

Fig. 4 The cause identification result within Sub-block 3 in Case 1

与传统评价方法相比, 两层分块 GMM-PRS 主要在分层分块和混合模型两方面做了改进. 分层分块使得评价难度降低, 解释性强, 可以直接在非优子块中进行原因追溯, 快速定位原因变量. 混合模型充分利用不同变量提供的信息, 提高了评价精度. 将 PRS、两层分块 PRS、GMM、两层分块 GMM 和两层分块 GMM-PRS 方法分别应用于黄金湿法冶炼过程运行状态评价中, 经过多次试验, 评价准确率均值如表 3 所示. 其中, 基于 PRS 的评价, 对全流程建立一个 PRS 模型; 基于两层分块 PRS 的评价, 将全流程进行层次和子块的划分, 对每个子块分别建立一个 PRS 模型; 基于 GMM 的评价, 针对全流程不同定性变量的组合, 分别建立全流程的 GMM; 基于两层分块 GMM 的评价, 将全流程进行层次和子块的划分, 在子块内, 针对不同定性变量的组合, 分别建立 GMM. 从准确率的对比可以看出, 相比于 PRS、两层分块 PRS、GMM 方法, 本文所提两层分块 GMM-PRS 具有明显优势. 两层分块 GMM 方法, 由于应用了本文所提的分层分块和针对不同定性变量组合分别建立 GMM 的思想, 具有与两层分块 GMM-PRS 相当的正确率. 但是, 相比于两层分块 GMM 方法, 所提两层分块 GMM-PRS 具有模

型数量少、计算量小、计算时间短的优势.

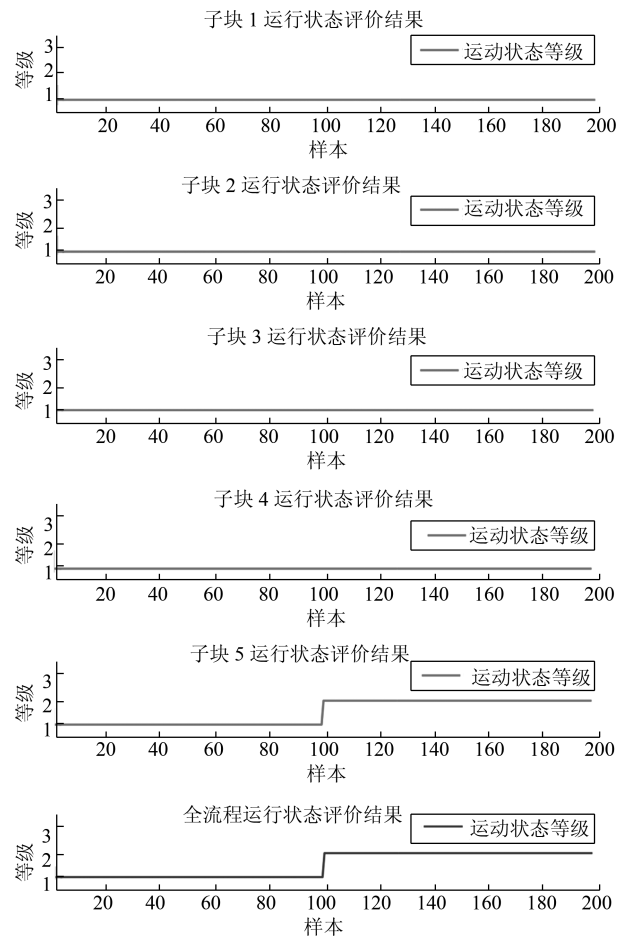


图 5 实验 2 运行状态等级在线评价结果

Fig. 5 The assessment result in Case 2

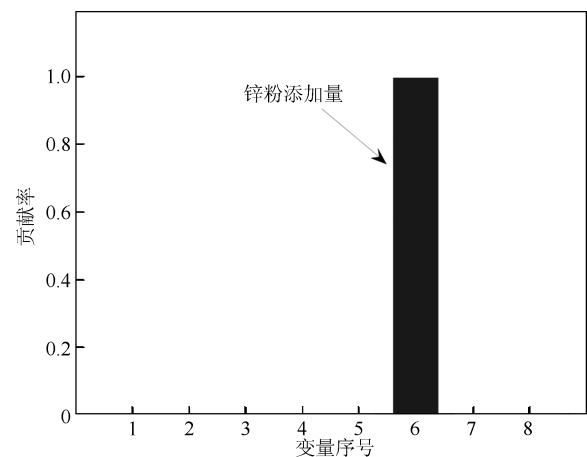


图 6 实验 2 子块 5 中非优原因追溯结果

Fig. 6 The cause identification result within Sub-block 5 in Case 2

表 3 不同方法评价准确率对比

Table 3 The assessment accuracy rate comparison of different methods

	PRS	两 层 分 块 PRS	GMM	两 层 分 块 GMM	两 层 分 块 GMM- PRS
评价准确率	75.3%	81.0%	86.2%	97.6%	97.9%

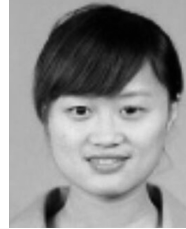
## 5 结论

针对定量、定性变量共存的流程工业过程运行状态评价问题, 本文提出了基于两层分块混合模型的评价方法. 将流程工业过程, 根据其物理特性划分为运行子块, 同时, 形成了子块层和全流程层, 两个评价层次. 在子块层, 对于以定量信息为主的子块, 根据不同的定性变量状态组合, 分别建立 GMM. 对于以定性信息为主的子块, 将定量变量进行离散化, 建立 PRS 模型. 全流程层运行状态等级由子块层最劣运行状态等级决定. 当过程运行于非优运行状态等级, 非优的子块可根据子块运行状态评价结果进行确定. 在非优的子块内, 本文提出了基于贡献率的原因追溯方法. 最后, 本文将所提方法应用于湿法冶金过程运行状态评价中, 并与传统方法进行了比较. 仿真结果证明了所提方法的有效性和优势.

## References

- Liu Y, Chang Y Q, Wang F L. Online process operating performance assessment and nonoptimal cause identification for industrial processes. *Journal of Process Control*, 2014, **24**(10): 1548–1555
- Liu Y, Wang F L, Chang Y Q, Ma R C. Comprehensive economic index prediction based operating optimality assessment and nonoptimal cause identification for multimode processes. *Chemical Engineering Research and Design*, 2015, **97**(1): 77–90
- Liu Y, Wang F L, Chang Y Q, Ma R C. Operating optimality assessment and nonoptimal cause identification for non-Gaussian multimode processes with transitions. *Chemical Engineering Science*, 2015, **137**: 106–118
- Zou X Y, Wang F L, Chang Y Q, Zhang B. Process operating performance optimality assessment and non-optimal cause identification under uncertainties. *Chemical Engineering Research and Design*, 2017, **120**: 348–359
- Zou X Y, Chang Y Q, Wang F L, Zhao L P. Process operating performance optimality assessment with coexistence of quantitative and qualitative information. *Canadian Journal of Chemical Engineering*, 2018, **96**(1): 179–188
- Qin S J. Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, 2003, **17**(8–9): 480–502
- Zhao C H, Gao F R. Critical-to-fault-degradation variable analysis and direction extraction for online fault prognostic. *IEEE Transactions on Control Systems Technology*, 2017, **25**(3): 842–854
- Li W Q, Zhao C H, Gao F R. Linearity evaluation and variable subset partition based hierarchical process modeling and monitoring. *IEEE Transactions on Industrial Electronics*, 2018, **65**(3): 2683–2692
- Liu Y, Wang F L, Chang Y Q. Online fuzzy assessment of operating performance and cause identification of nonoptimal grades for industrial processes. *Industrial & Engineering Chemistry Research*, 2013, **52**(50): 18022–18030
- Liu Y, Wang F L, Chang Y Q. Operating optimality assessment based on optimality related variations and nonoptimal cause identification for industrial processes. *Journal of Process Control*, 2016, **39**: 11–20
- Yu J, Qin S J. Multimode process monitoring with bayesian inference-based finite gaussian mixture models. *AIChE Journal*, 2008, **54**(7): 1811–1829
- Wolbrecht E, D'ambrosio B, Paasch R, Kirby D. Monitoring and diagnosis of a multi-stage manufacturing process using Bayesian networks. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 2000, **14**(1): 53–67
- Kumar S, Kaur T. Development of ANN based model for solar potential assessment using various meteorological parameters. *Energy Procedia*, 2016, **90**: 587–592
- Rojas-Moraleda R, Valous N A, Gowen A, Esquerre C, Härtel S, Salinas L, et al. A frame-based ANN for classification of hyperspectral images: assessment of mechanical damage in mushrooms. *Neural Computing and Applications*, 2017, **28**(1): 969–981
- Ren J, Wang J, Jenkinson I, Xu D L, Yang J B. A Bayesian network approach for offshore risk analysis through linguistic variables. *China Ocean Engineering*, 2007, **21**(3): 371–388
- Hosack G R, Hayes K R, Dambacher J M. Assessing model structure uncertainty through an analysis of system feedback and Bayesian networks. *Ecological Applications*, 2008, **18**(4): 1070–1082
- Biglarfadafan M, Danehkar A, Pourebrahim S, Shabani A A, Moeinaddini M. Application of strategic fuzzy assessment for environmental planning; case of bird watch zoning in wetlands. *Open Journal of Geology*, 2016, **6**(11): 1380–1400
- Wang Zheng-Shuai, Liu Bing-Jing, Deng Ka-Zhong. Fuzzy extension assessment model of old goaf stability. *Chinese Journal of Underground Space and Engineering*, 2016, **12**(2): 553–559  
(王正帅, 刘冰晶, 邓喀中. 老采空区稳定性的模糊可拓评价模型. 地下空间与工程学报, 2016, **12**(2): 553–559)
- Kusiak A. Rough set theory: a data mining tool for semiconductor manufacturing. *IEEE Transactions on Electronics Packaging Manufacturing*, 2001, **24**(1): 44–50

- 20 Ziarko W. Probabilistic rough sets. In: Proceeding of the 10th International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing. Regina, Canada: Springer, 2005. 283–293
- 21 Yao Y Y. Probabilistic rough set approximations. *International Journal of Approximate Reasoning*, 2008, **49**(2): 255–271
- 22 Yao Y Y. Probabilistic approaches to rough sets. *Expert Systems*, 2003, **20**(5): 287–297
- 23 Liu Q, Qin S J, Chai T. Multiblock concurrent PLS for decentralized monitoring of continuous annealing processes. *IEEE Transactions on Industrial Electronics*, 2014, **61**(11): 6429–6437
- 24 Deng X G, Wang L. Multimode process fault detection method using local neighborhood standardization based multi-block principal component analysis. In: Proceeding of the 29th Chinese Control and Decision Conference. Chongqing, China: IEEE, 2017. 5615–5621
- 25 Macgregor J F, Jaeckle C, Kiparissides C, Koutoudi M. Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal*, 1994, **40**(5): 826–838
- 26 Jiang Q C, Yan X F. Monitoring multi-mode plant-wide processes by using mutual information-based multi-block PCA, joint probability, and Bayesian inference. *Chemometrics and Intelligent Laboratory Systems*, 2014, **136**: 121–137
- 27 Rännar S, MacGregor J F, Wold S. Adaptive batch monitoring using hierarchical PCA. *Chemometrics and Intelligent Laboratory Systems*, 1998, **41**(1): 73–81
- 28 Chen G, McAvoy T J. Multi-block predictive monitoring of continuous processes. *IFAC Proceedings Volumes*, 1997, **30**(9): 73–77
- 29 Katsaros G, Kousiouris G, Gogouvitis S V, Kyriazis D, Menychtas A, Varvarigou T. A Self-adaptive hierarchical monitoring mechanism for Clouds. *Journal of Systems and Software*, 2012, **85**(5): 1029–1041
- 30 Wang Guo-Yin. *Rough Set Theory and Knowledge Acquisition*. Xian: Xian Jiaotong University Press, 2001.  
(王国胤. Rough 集理论与知识获取. 西安: 西安交通大学出版社, 2001.)



**邹筱瑜** 东北大学信息科学与工程学院博士研究生. 主要研究方向为复杂工业过程运行状态评价与非优原因追溯.

E-mail: xiaoyuzou\_neu@hotmail.com

(**ZOU Xiao-Yu** Ph.D. candidate at the College of Information Science and Engineering, Northeastern University.

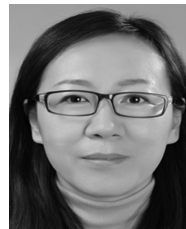
Her research interest covers complex process operating performance optimality assessment and non-optimal cause identification.)



**王福利** 东北大学信息科学与工程学院教授. 主要研究方向为复杂工业过程运行状态评价、优化、故障诊断. 本文通信作者. E-mail: wangfuli@ise.neu.edu.cn

(**WANG Fu-Li** Professor at the College of Information Science and Engineering, Northeastern University. His research interest covers operating performance optimality assessment, optimization, and fault diagnosis of complex system. Corresponding author of this paper.)

performance optimality assessment, optimization, and fault diagnosis of complex system. Corresponding author of this paper.)



**常玉清** 东北大学信息科学与工程学院教授. 主要研究方向为复杂工业过程运行状态评价和故障诊断.

E-mail: changyuqing@ise.neu.edu.cn

(**CHANG Yu-Qing** Professor at the College of Information Science and Engineering, Northeastern University.

Her research interest covers operating performance optimality assessment and fault diagnosis of complex system.)



**郑伟** 国家电网辽宁省电力有限公司研究员. 主要研究方向为复杂工业过程运行状态评价与非优原因追溯.

E-mail: zhengwei\_stg@hotmail.com

(**ZHENG Wei** Researcher in Liaoning Electric Power Co. LTD., State Grid. His research interest covers operating performance optimality assessment and non-optimal cause identification.)

ment and non-optimal cause identification.)