

# 控制与决策

Control and Decision

## 部分子块通讯的分布式PCA厂级工业过程监测方法

曹跃, 陈志文, 袁小锋, 王雅琳, 桂卫华

引用本文:

曹跃, 陈志文, 袁小锋, 等. 部分子块通讯的分布式PCA厂级工业过程监测方法[J]. *控制与决策*, 2020, 35(6): 1281–1290.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2018.1485>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### 基于Block-RPLS模型自适应更新的质量预测方法

Quality prediction method based on adaptive updating of Block-RPLS model

*控制与决策*. 2018, 33(3): 455–462 <https://doi.org/10.13195/j.kzyjc.2017.0070>

### 基于多块MICA-PCA的全流程过程监控方法

Plant-wide process monitoring based on multiblock MICA-PCA

*控制与决策*. 2018, 33(2): 269–274 <https://doi.org/10.13195/j.kzyjc.2016.1222>

### 基于多块相对变换独立主元分析的故障诊断方法

Fault diagnosis approach based on relative transformation ICA of multiblock

*控制与决策*. 2018, 33(11): 2009–2014 <https://doi.org/10.13195/j.kzyjc.2017.0869>

### 结合KL距离与图像域分块的SAR图像分割

Segmentation of SAR image based on Kullback-Leibler distance and regular tessellation

*控制与决策*. 2018, 33(10): 1767–1774 <https://doi.org/10.13195/j.kzyjc.2017.0758>

### 基于自适应分块表观模型的视觉目标跟踪

Visual tracking based on adaptive patches appearance model

*控制与决策*. 2016(3): 448–452 <https://doi.org/10.13195/j.kzyjc.2015.0035>

# 部分子块通讯的分布式PCA厂级工业过程监测方法

曹 跃, 陈志文, 袁小锋, 王雅琳<sup>†</sup>, 桂卫华

(中南大学 信息科学与工程学院, 长沙 410083)

**摘 要:** 针对现代工业过程具有检测、控制变量多,且变量关系复杂、耦合严重等特点,厂级工业过程监测通常采用分块的方式建立模型.然而,实际过程中每个子块间存在物流、能流的交换和信息的传递,这种建模方式通常只利用子块本身的信息建立模型,恰恰忽略了子块之间内在的联系,可能导致子块监测模型一定程度上丢失其他子块所提供的过程信息.针对现有方法的局限性,提出一种部分子块通讯的分布式主元分析(PCA)过程监测方法.首先利用先验知识对厂级过程进行分块,增加用于描述子块间连接方式的拓扑矩阵;以迭代更新的思路,利用其他节点提供的压缩信息,按照拓扑矩阵不断更新系数矩阵,进而得到最终的监测模型;代入待监测数据后,综合考虑其余子块携带的异常信息得出监测结果.在 tennessee eastman 过程和加氢裂化过程中,将所提出的方法与传统分块PCA、全部子块交互的分布式PCA方法进行对比,所得结果表明了所提出方法的合理性和有效性.

**关键词:** 分布式监测方法; 厂级过程监测; 部分子块通讯; 拓扑矩阵; 迭代更新

中图分类号: TP277; TE62

文献标志码: A

## Distributed PCA for plant-wide processes monitoring with partial block communication

CAO Yue, CHEN Zhi-wen, YUAN Xiao-feng, WANG Ya-lin<sup>†</sup>, GUI Wei-hua

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

**Abstract:** In modern industrial processes, there are numerous measure and control variables with serious coupling and complex relationships. Alternatively, distributed (multi-block) process monitoring methods can handle this problem for plant-wide process monitoring. However, the exchange of energy, materials and information exists among blocks in industrial processes. The distributed methods mentioned above always constructs a monitoring model in each block without communication from other blocks. Those modelling methods ignore the inherent connection between blocks and the monitoring models may be affected without the information from others. In order to overcome this limitation of the distributed methods, a distributed principal component analysis (PCA) process monitoring method based on the industrial process connection is proposed in this paper. Firstly, this method divides the plant-wide processes into blocks by prior knowledge. The topology matrix is proposed to describe the connection between blocks. In each iteration, the parameters of the monitoring model are updated according to the topology matrix using compressed information from other blocks and the final monitoring model is obtained. After substituting the tested data, the abnormal information carried by the residual blocks is taken into consideration in each blocks, and the final monitoring results are obtained. In the tennessee eastman process and the hydrocracking process, the proposed method is compared with the traditional multi-block PCA and fully connected distributed PCA method, which demonstrates the feasibility and effectiveness of the proposed method.

**Keywords:** distributed process monitoring; plant-wide process monitoring; partial block communication; topology matrix; iteration updating

## 0 引 言

传统的过程监测<sup>[1-3]</sup>(见图1(a))往往通过构建过程模型来判断该过程是否处于正常工作状态,能否及时反映过程的异常状态并提醒现场工人作出相应的调整,因此,过程监测是工业过程正常有序运行的

重要保证<sup>[4-5]</sup>.然而,现代工业过程以大型集散控制系统(DCS)为核心,配备了大量的用于直接检测数字仪表、间接检测的软测量<sup>[6-7]</sup>设备以及结构复杂的控制回路,导致传统过程监测方法难以对现代工业过程实现全局的监测.因此,近些年来,国内外学者针对此

收稿日期: 2018-10-30; 修回日期: 2019-01-11.

基金项目: 国家自然科学基金重大项目(61590921); 中央高校基本科研业务费专项资金项目(2017zzts135).

责任编辑: 姜斌.

<sup>†</sup>通讯作者. E-mail: ylwang@csu.edu.cn.

问题提出了多种分布式监测方法<sup>[8-12]</sup>(见图1(b)). 这些分布式方法通过对过程进行分块,再在每个子块中独立地建立监测模型,达到分布式监测的目的. 这些方法主要可分为两类:基于机理的分块和基于数据的分块. Ge等<sup>[13]</sup>基于机理分块提出了一个适用于厂级过程的分布式监测框架,并针对 tennessee eastman (TE)过程对该过程分为3块,最后通过贝叶斯推理对各个子块的监测结果进行了汇总. Jiang等<sup>[14]</sup>利用互信息谱(MI-spectral)评价两个变量之间的相关度,再利用传统的聚类方法对其进行聚类,将聚类作为分块的结果,最后也通过贝叶斯推理对各个子块的监测结果进行汇总. 此外,Zhang等<sup>[15]</sup>基于核偏最小二乘提出了厂级分布式监测方法,并成功地应用于连续退火过程. 同样在该过程,Liu等<sup>[16]</sup>基于多级的主元分析(principal component analysis, PCA)也得到了较好的监测效果.

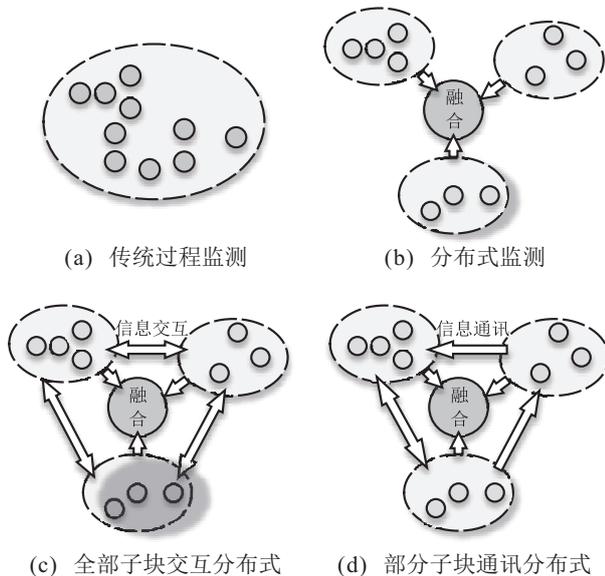


图1 4种过程监测方法结构示意图

尽管上述监测方法均得到了较好的效果,但是,这些分布式监测方法均将过程分块后分别独立地建立监测模型,忽略了子块之间的内在联系,一定程度上丢失了其他子块所提供的过程信息,而这部分信息也是分布式过程监测中重要的参考依据. 针对现有方法的局限性,本文将全部子块信息交互<sup>[17]</sup>的分布式典型相关分析(canonical correlation analysis, CCA)引入工业过程监测中,见图1(c). 该方法首先将工业过程进行机理分块,再按照子块特点确定CCA的输入和输出变量. 经过不断迭代更新,对分布式CCA监测模型进行训练,并在TE过程中验证了该方法的有效性.

需要强调的是,在该全部子块信息交互的分布式

过程监测方法中,如果将其他节点提供的全部数据与当前节点数据直接全部整合建模是不可行的,需要对其他节点的数据进行压缩并传输后,再与当前节点融合后进行建模. 原因有以下几点: 1) 利用全部的数据进行建模与传统全局建模方法无异,违背了分布式思想的初衷; 2) 在大量数据的现代工业现场,如果将全部数据进行建模,则会带来大量的计算负荷; 3) 传输其他节点的压缩数据,一方面提供了其他节点重要的异常信息,另一方面,由于主要利用当前节点未压缩的数据进行建模,使其在模型中起主导作用,如果某个子块发生异常,则能初步确定异常发生在该区域,进而为后续诊断提供重要依据.

如上文所述,全部子块交互的分布式监测方法默认每个子块之间都是可以交互信息的,在迭代更新时,每个子块可从其他所有子块中获取信息. 然而,结合实际工业过程,并不是所有子块都可以交互信息,且不同的工业应用背景中子块间的连接方式也不同,从而导致迭代更新方式的改变以及模型结构的变化. 此外,CCA算法中输入、输出变量的选取可能具有一定的随意性. 综上,这些因素给部分子块通讯的分布式监测方法带来了较大的挑战. 因此,本文为了解决实际工业过程中子块的连接方式难以通用描述的问题,增加拓扑矩阵,并将PCA算法引入该分布式过程监测框架中,提出一种部分子块通讯的分布式PCA厂级过程监测方法(见图1(d)). 该方法首先通过机理知识对工业过程进行分块,增加用于通用描述实际工业过程的拓扑矩阵,使监测模型更符合实际工业过程的特征,进一步提高监测方法的精度. 经过迭代更新系数矩阵,使分布式过程监测模型不断完善,直到模型的相关参数收敛或达到最大迭代次数后,保存监测模型的各个参数,获得最终的监测模型. 代入待监测数据后,经过系数矩阵的转换,将当前子块与其余子块的信息进行融合,再将融合后的数据分别代入每个子块对应的模型中,分别计算得到各个子块的监测模型. 最后通过每个子块监测结果的融合,得到最终的监测结果.

## 1 PCA算法及其过程监测方法

本节主要回顾经典的PCA算法,并简要介绍基于PCA的过程监测方法.

### 1.1 PCA算法

由Pearson<sup>[18]</sup>提出的PCA算法是一种广泛应用于数据降维的经典方法. 通常,该方法将原始数据映射到另一个由主元变量形成的坐标系中,从而达到降

维的目的. 假设, 原始数据  $\mathbf{X}^{n \times m}$  包含  $n$  个变量,  $m$  个采样点, 通过 PCA, 该数据集可被分解为两部分, 即

$$\mathbf{X} = \mathbf{T}_A \mathbf{P}_A^T + \mathbf{E}. \quad (1)$$

其中:  $\mathbf{T}_A \in \mathbf{R}^{n \times A}$  为得分矩阵,  $\mathbf{P}_A \in \mathbf{R}^{m \times A}$  为负载矩阵,  $\mathbf{E} \in \mathbf{R}^{n \times m}$  为残差矩阵,  $A$  为主元变量的个数. 关于主元变量的确定方法有很多, 常见的方法有累计方差贡献率等.

## 1.2 基于PCA的过程监测方法

过程监测是保证过程安全正常生产的重要方法. 目前, 在大数据的背景下, 基于数据的多元统计方法成为过程监测最常用的方法之一<sup>[19-20]</sup>. 基于PCA的过程监测主要分为离线建模和在线监测两部分.

### 1.2.1 离线建模

基于PCA算法, 利用正常的过程数据建立PCA过程监测模型. 如式(1)所示, 原始的正常数据  $\mathbf{X}^{n \times m}$  可被分解为主元部分和残差部分, 即

$$\mathbf{T}_A = \mathbf{X} \mathbf{P}_A, \hat{\mathbf{X}} = \mathbf{T}_A \mathbf{P}_A^T, \mathbf{E} = \mathbf{X} - \hat{\mathbf{X}}, \quad (2)$$

其中  $\mathbf{X}$  为原始正常数据通过得分矩阵  $\mathbf{T}_A$  和负载矩阵  $\mathbf{P}_A$  重构后得到的矩阵.

为了判断待监测数据是否处于异常状态, 通常构建  $T^2$  和  $Q$  统计量. 在离线建模部分, 两个统计量对应的控制限  $T_{lim}$  和  $Q_{lim}$  可由式(2)中  $\mathbf{T}_A$ 、 $\mathbf{P}_A$  和  $\mathbf{E}$  计算得到. 限于篇幅不再赘述, 详细推导参见文献[21].

### 1.2.2 在线监测

基于PCA离线建模, 利用待监测数据  $\mathbf{x}_{new} \in \mathbf{R}^{1 \times m}$  计算相应的  $T^2$  和  $Q$  统计量, 并与对应的控制限  $T_{lim}$  和  $Q_{lim}$  比较, 进而评价该数据被采样时的过程是否处于异常状态.

将  $\mathbf{x}_{new}$  代入已建立的如下监测模型中, 计算对应的两个统计量  $T^2$  和  $Q$ :

$$T^2 = \mathbf{x}_{new} \mathbf{P}_A (\Lambda)^{-1} \mathbf{P}_A^T \mathbf{x}_{new}^T, \quad (3)$$

$$Q = \mathbf{e}^T \mathbf{e}, \mathbf{e} = (\mathbf{I}_m - \mathbf{P}_A \mathbf{P}_A^T) \mathbf{x}_{new}, \quad (4)$$

其中  $\mathbf{I}_m$  为  $m$  维的单位矩阵.

将计算得到的  $T^2$  和  $Q$  统计量与离线建模所得对应的控制限  $T_{lim}$  和  $Q_{lim}$  进行比较, 如果计算得到的统计量超过对应的控制限, 则表明当待监测数据被采样时过程很可能处于异常状态.

## 2 部分子块通讯的分布式厂级过程监测方法

本节主要介绍本文所提出的部分子块通讯的分布式厂级监测方法.

## 2.1 厂级过程变量分块与数据预处理

### 2.1.1 厂级过程变量分块

随着工业过程的不断发展, 现代工业过程配备了大量的数字检测仪表和控制回路, 并利用DCS将全厂的检测和操作变量进行采集、储存以及整理. 变量繁多、耦合严重、非线性等特点给厂级过程监测带来了巨大的挑战. 为实现厂级过程的全局监测, 通常的方法是将全厂的变量进行分块, 形成若干子块, 再在每个子块中分别建立监测模型. 本文所提出的分布式监测框架也遵循该思路, 通过对厂级变量的划分而构成. 考虑到实际工业过程中装置、设备以及仪表的物理连接情况, 本文采用基于机理(先验知识和专家经验)的分块方式, 对全厂的变量进行划分. 全厂的变量  $\mathbf{X}^{n \times m}$  经过机理划分为  $B$  块

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_B]. \quad (5)$$

本文采取基于机理的分块方式, 主要根据实际过程的物理连接和专家经验. 该方法通常以设备的功能和所属区域进行划分, 具有一定的合理性和可行性; 同时, 如果监测到异常, 则可极大地方便全局监测后异常变量的诊断与定位.

### 2.1.2 数据预处理

在工业过程中噪声是不可避免的, 在使用数据前, 需要对数据的离群噪声点进行去除, 本文利用  $3\sigma$  准则进行第一步预处理.

考虑到物料的流量、气体的压力、反应器的温度等具有不同的量纲, 在建立分布式监测模型之前需要对数据进行标准化. 原始数据  $\mathbf{X} = \{x_{i,j}\} (i = 1, 2, \dots, n, j = 1, 2, \dots, m)$  按列进行标准化, 如下式所示:

$$(x_{i,j} - \bar{x}_j) / \sigma_j, \quad (6)$$

其中  $\bar{x}_j$  和  $\sigma_j$  分别为  $\mathbf{X}$  第  $j$  列各元素的平均值和标准差.

## 2.2 部分子块通讯的分布式监测模型离线构建

考虑现代工业过程数据量大的特点, 本文采用有数据更新的迭代方式不断更新模型参数, 最终建立分布式监测模型. 此外, 考虑到实际工业过程中稳定工作状态下仍可能存在的较小偏移或波动, 为使模型能更好地实时反映过程的特征, 且跟踪过程的动态信息, 在实际应用中应引入最新的稳态数据不断更新监测模型.

为建立部分子块通讯的分布式过程监测模型, 本文采用迭代建模的思路, 其中每个子块对应一个

系数矩阵,分别记为 $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_i, \dots, \mathbf{V}_B (i = 1, 2, \dots, B)$ . 该思路以当前子块为主要更新的子块,其余子块根据上次的系数矩阵随之更新. 在当前子块训练后,保存并更新系数矩阵,跳转至下一个子块进行训练,直到达到预期的精度或最大迭代次数后停止迭代,获取最终的监测模型.

首先,将原始数据集 $\mathbf{X}$ 进行划分,并分别对变量进行标准化,且保存每个变量的均值和标准差,用于在线监测部分中数据的初始化. 同时,鉴于随机化方法的思路,利用0-1的随机数对每个系数矩阵 $\mathbf{V}_i \in \mathbf{R}^{m_i \times A_i}$ 进行初始化. 其中: $m_i$ 表示第 $i$ 个子块中包含的变量个数, $A_i$ 表示第 $i$ 个子块对应的主元个数. 注意:为保持迭代算法的总体性和简约性,本方法将每个子块的主元个数保持相等,即

$$A_1 = A_2 = \dots = A_B.$$

另外,考虑子块间数据传输通信代价,主元个数一般由人工经验方式确定,当每个子块具有10~50个变量时,每块主元个数2~4个为宜.

其次,过程经上述数据分块、数据预处理以及参数初始化后,进入循环迭代. 每次迭代取长度为 $L$ 的数据子块 $\mathbf{X}^s \in \mathbf{R}^{L \times m}$ ( $s$ 为当前迭代次数)进行模型训练. 当迭代次数为 $s$ 时,以当前子块 $i$ 为中心,再获取其余子块的信息,构建数据集 $\mathbf{Y}_i^s \in \mathbf{R}^{(m_i + \sum A_{-i}) \times L}$ ,如下式所示:

$$\mathbf{Y}_i^s = [\mathbf{V}_1 \mathbf{X}_1^{sT}; \dots; \mathbf{X}_i^{sT}; \mathbf{V}_{i+1} \mathbf{X}_{i+1}^{sT}; \dots; \mathbf{V}_B \mathbf{X}_B^{sT}]. \quad (7)$$

如式(1),将数据集 $\mathbf{Y}_i^s$ 利用PCA进行分解,可得

$$\mathbf{Y}_i^s = \mathbf{T}_i^s \mathbf{P}_i^s + \mathbf{E}_i^s. \quad (8)$$

经PCA分解后,保存当前的负载矩阵 $\mathbf{P}_i^s$ 和得分矩阵 $\mathbf{T}_i^s$ 为子块 $i$ 在第 $s$ 迭代中最优的负载矩阵和得分矩阵.

为使子块间通讯更符合实际工业过程,本文增加描述子块间连接方式的拓扑矩阵 $\mathbf{C} \in \mathbf{R}^{B \times B}$ ,即矩阵中的每个元素表示子块间是否有直接联系,如果有则为1,否则为0. 该拓扑矩阵能反映任何工业过程子块之间的连接方式. 以工业过程最为常见的流程工业串联方式为例,假设过程被分为3个子块(如图2所示),其对应的拓扑结构如下式所示:

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \quad (9)$$

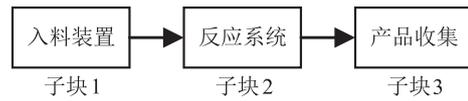


图2 常见的流程工业串联方式

入料装置与反应系统有直接连接,则该拓扑矩阵的元素 $c_{12}$ 为1;同理,入料装置与产品收集无直接连接,则 $c_{13}$ 为0.

再次,根据PCA算法所得的得分矩阵 $\mathbf{T}_i^s$ 、负载矩阵 $\mathbf{P}_i^s$ 以及子块的拓扑矩阵 $\mathbf{C}$ ,并以第 $i$ 个子块在第 $s$ 次迭代为例,介绍该算法迭代更新的方式,如下式所示:

$$\mathbf{V}^{s+1} = \mathbf{V}^s; \quad (10)$$

$$\mathbf{V}_j^{s+1} = \begin{cases} \mathbf{G}_i^s, & c_{i,i} = 1; \\ \mathbf{V}_j^s, & c_{i,j} = 0; \\ \mathbf{V}_j^s \mathbf{G}_j^s, & c_{i,j} = 1; \end{cases} \quad (11)$$

$$\mathbf{P}_i^s = [\mathbf{G}_1^s, \dots, \mathbf{G}_i^s, \dots, \mathbf{G}_B^s]^T. \quad (12)$$

其中: $\mathbf{V}_i^s$ 是子块 $i$ 在第 $s$ 次迭代时的系数矩阵;同理, $\mathbf{G}_i^s$ 是负载矩阵 $\mathbf{P}_i^s$ 的子矩阵, $\mathbf{G}_i^s \in \mathbf{R}^{m_i \times A_i}$ , $\mathbf{G}_j^s \in \mathbf{R}^{A_i \times A_i}$ . 式(11)表示了算法的核心思路,其中 $c_{i,i} = 1$ 始终成立,因为当前迭代的子块必与该子块本身存在信息的传递. 式(12)为负载矩阵 $\mathbf{P}_i^s$ 的分块形式.

最后,判断迭代是否收敛 $|\lambda_1^{s+1} - \lambda_1^s| < \delta$ ( $\delta$ 为设定的参数)或达到最大迭代次数. 如果满足,则跳出迭代,保存最终迭代的各项参数,完成分布式监测模型的建立;否则,返回迭代的入口继续迭代.

### 2.3 部分子块通讯的分布式监测模型在线监测

基于已经建立的分布式监测模型,代入待监测的数据 $\mathbf{X}_{\text{new}}$ 进行监测,判断该组数据采样时过程的工作状态是否处于异常.

与经典PCA方法类似,将待监测的数据利用建模数据所得的均值和方差进行标准化,再按照建模的分块方法对标准化后的待监测数据进行分块,有

$$\mathbf{X}_{\text{new}} = [\mathbf{X}_1^*, \mathbf{X}_2^*, \dots, \mathbf{X}_B^*]. \quad (13)$$

利用建模中最后更新的系数矩阵 $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_B$ 对待监测数据进行变换,即

$$\mathbf{Z}_i = [\mathbf{V}_1 \mathbf{X}_1^*, \dots, \mathbf{X}_i^*, \mathbf{V}_{i+1} \mathbf{X}_{i+1}^*, \dots, \mathbf{V}_B \mathbf{X}_B^*]^T, \quad (14)$$

其中 $\mathbf{Z}_i$ 为第 $i$ 个子块的待监测数据结合了其余节点信息后所得的综合数据.

参照用于过程监测的经典PCA,分别利用每个子块的综合数据  $Z_i$ , 分别代入对应的监测模型, 计算每个子块的统计量  $T_i^2$  和  $Q_i$ , 再与每个子块所得的控制限  $T_{lim,i}^2$  和  $Q_{lim,i}$  进行比较. 如果计算所得统计量超过了对应的控制限, 则认为该子块出现异常.

为保证所提出分布式工业过程监测算法的敏感度, 对每个子块的监测结果按照以下策略进行融合, 得出最终的监测结果:

$$\begin{cases} \exists T_i^2 > T_{lim,i}^2 \vee \exists Q_i > Q_{lim,i}, \mathcal{A}, \\ \forall T_i^2 < T_{lim,i}^2 \ \& \ \forall Q_i < Q_{lim,i}, \mathcal{N}. \end{cases} \quad (15)$$

其中:  $i = 1, 2, \dots, B$ ;  $\mathcal{A}$  和  $\mathcal{N}$  分布表示系统处于异常和正常状态.

### 2.4 过程监测方法小结

上述数据分块与预处理、监测模型离线建立以及过程数据在线监测, 详细介绍了本文所提出的部分子块通讯的分布式工业过程监测方法. 为了更清楚地说明本文所提出监测方法的结构, 下面简要地给出该方法的步骤.

分布式监测方法主要步骤如下:

step 1: 对原始数据集  $X$  按照机理进行分块, 每个变量分别进行标准化. 将系数矩阵  $V_i$  用  $0 \sim 1$  的随机数初始化.

step 2: 进入迭代. 取长度为  $L$  的数据块进行迭代. 当前为第  $s$  次迭代, 以第  $i$  个子块为中心将该子块数据与其余子块数据进行融合, 得到综合的数据块  $Y_i^s$ .

step 3: 将数据块  $Y_i^s$  利用PCA分解, 得出负载矩阵和得分矩阵.

step 4: 考虑实际背景的子块连接方式, 构建拓扑矩阵  $C$ , 再根据式(10)~(12), 计算并迭代更新系数矩阵  $V_i$ .

step 5: 保存当前的负载矩阵和得分矩阵, 并判断迭代算法是否收敛或是否达到最大迭代次数. 如果是, 则跳出迭代循环, 并进入在线监测部分; 如果否, 则返回 step 2.

step 6: 获取待监测的数据  $X_{new}$ , 根据原始数据的分块方式进行分块, 再根据原始数据的均值和标准差进行标准化.

step 7: 将标准化后的分块待监测数据, 分别利用建模时得到的系数矩阵  $V_i$  进行转换, 得到每个子块对应的综合数据集  $Z_i$ .

step 8: 将每个子块的数据集  $Z_i$  分别代入子块对

应的PCA模型中, 得到各自的监测结果.

step 9: 根据监测结果的汇总策略, 如式(15)所示, 对每个子块的监测结果进行汇总, 得到最终的监测结果.

在上述步骤中, step 1~step 5 为分布式过程监测方法的离线建模部分, step 6~step 9 为在线监测部分.

图3给出了本文所提出的过程监测方法的示意框图.

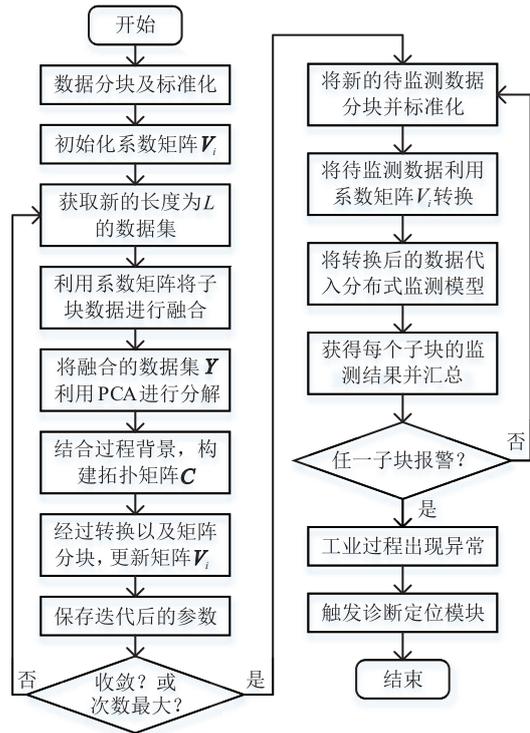


图3 本文所提出的分布式过程监测方法示意

## 3 工业案例验证

为了验证本文方法的有效性和实用性, 将本文方法分别应用至TE过程Benchmark和实际的加氢裂化过程中, 并与传统分块PCA、全部子块交互的分布式PCA进行对比.

### 3.1 TE过程

TE过程<sup>[22-23]</sup>是一个经典的Benchmark仿真模型, 广泛用于验证和测试过程监测以及故障诊断方法. 如图4所示, TE过程包括一个反应器, 一个压缩机, 一个产品汽提塔, 一个产品冷凝器和一个气液分离器. 该过程包含12个操作变量, 22个连续检测变量和19个成分检测变量. 通常在验证监测和诊断方法时, 选取其中的52个变量作为原始数据.

该过程模拟了21种标准的故障, 其中前15个以及第21个故障为已知故障, 其余的第16~第20个为5个未知的故障. 在本文中, 利用TE过程的正常数据

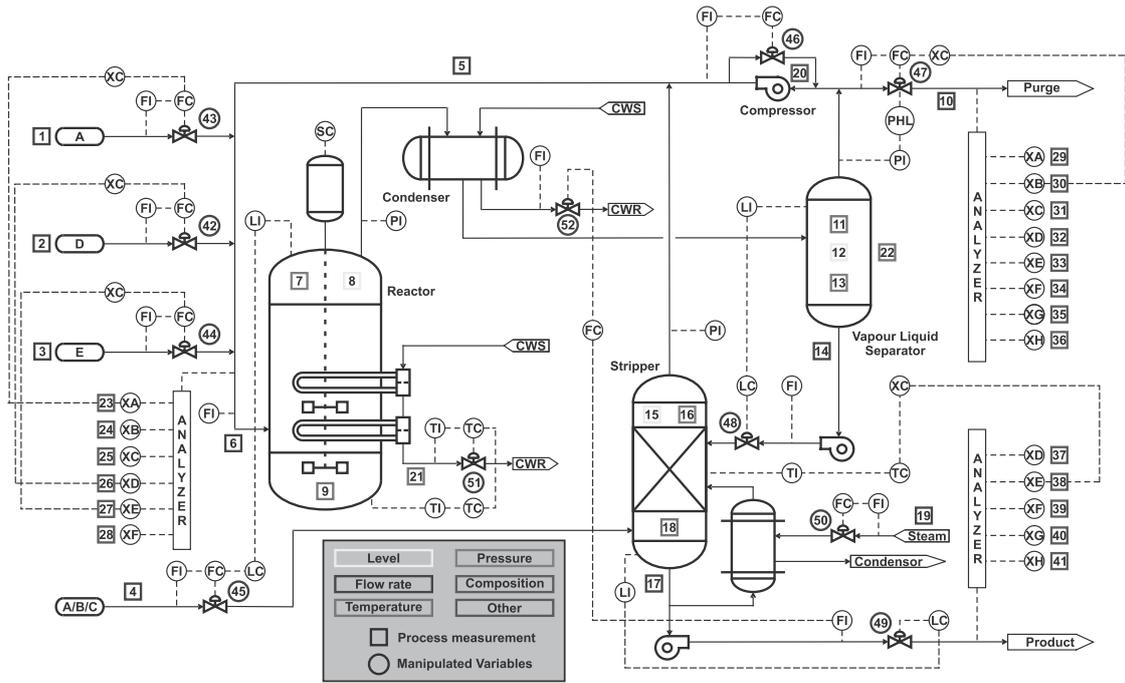


图4 TE过程流程

建立分布式监测模型,并分别将21组故障数据代入建立的监测模型中以验证所提出监测方法.其中:每组故障数据各有960个采样点,前160个为正常数据,从第161个采样点后分别注入21种不同的故障.由于故障3、9和15通常很难监测,本文略去这3种故障,对其余故障进行验证.需要指出的是,本文方法需要大量数据进行迭代更新,因此,基于960个正常数据,将用于建模的数据扩展至20万个采样点.由于TE过程的正常数据处于同一个稳态,可通过打乱原有采样点顺序的方法进行扩展.

为了验证本文方法的有效性,将该方法与传统分块PCA、全部子块交互的分布式PCA方法进行对比,即将利用21种故障数据分别代入这3种方法所建立的监测模型中,获得各自的检测率.传统分块PCA是将全部的数据进行分块后,分别建立PCA监测模型;全部子块交互的分布式PCA与本文方法分块方式相同,但不含描述工业过程的拓扑矩阵.表1给出了这3种方法在TE过程21种故障的监测对比结果(故障检测率).与传统PCA和全部子块交互分布式PCA相比,本文所提出的基于拓扑矩阵的部分连通分布式PCA在大多数典型故障的情况下都能进行较好的监测.此外,从故障的检测率来看,本文所提出的分布式算法,虽然利用了大量的压缩信息,有效降低了由于信息通信带来的较大计算负荷,但仍能保持与其他两种方法检测率接近,具有一定的可行性与工程实践意义.

表1 TE过程典型故障监测结果

编号	传统分块PCA		全部交互PCA		部分通讯PCA	
	$T^2$	$Q$	$T^2$	$Q$	$T^2$	$Q$
1	0.9950	0.9975	0.9875	0.9987	0.9900	0.9987
2	0.9825	0.9913	0.9749	0.9925	0.9737	0.9925
4	0.4100	1.0000	0.0940	1.0000	0.0589	1.0000
5	0.2625	0.4313	0.3158	0.5163	0.3083	0.5865
6	0.9900	1.0000	0.9950	1.0000	0.9950	1.0000
7	1.0000	1.0000	0.5150	1.0000	0.5075	1.0000
8	0.9750	0.9813	0.9286	0.9887	0.9110	0.9937
10	0.4163	0.7125	0.5401	0.6479	0.5226	0.6967
11	0.5213	0.7438	0.2168	0.8596	0.1717	0.8860
12	0.9875	0.9863	0.9424	0.9950	0.9311	0.9950
13	0.9513	0.9575	0.9373	0.9624	0.9373	0.9787
14	0.9988	0.9988	0.6704	1.0000	0.4887	1.0000
16	0.2325	0.6488	0.4511	0.5589	0.4273	0.6767
17	0.8013	0.9738	0.7130	0.9599	0.6391	0.9599
18	0.8913	0.9300	0.8860	0.9411	0.8910	0.9323
20	0.3738	0.6950	0.4737	0.7268	0.4637	0.7707
21	0.3338	0.5988	0.3496	0.5514	0.3371	0.5376

综上所述,本文所提出的部分连通分布式PCA监测方法,首先对TE过程根据专家经验和先验知识进行了分块;然后利用拓扑矩阵客观地反映了TE过程各子块间的实际连接方式;最后根据拓扑矩阵进行压缩通信,降低了传输成本,得到了最终的分布式监测模型.因此,本文方法从全局角度分析了各子块之间的连接关系,且在保持原有监测性能的前提下降低了通信代价,在大部分故障案例中展示了较好的监测性能.为了更好地说明本文方法的有效性,下面以故障1为例进行对比和讨论.

TE过程中,故障1为管道4中进料B含量恒定,而A/C进料比例发生变化.该故障直接影响管道4中A、C进料的比例,随后被输送至汽提塔底部,而汽提塔顶部输入的是经过气液分离器分离后的液体部分(产品部分).因此,该故障直接影响汽提塔的汽提效果,进而影响到产品的成分以及汽提塔顶部气体返回反应系统的成分.从机理分析可知,该故障发生首先会引起管道4中A/C进料比例变化,且在全部分子块交互和部分子块通讯的分布式PCA监测算法中,与该故障的相关变量被划分至子块1中.因此,子块1对该故障敏感度更高.对比图5(b)与图6(a),因本文方法考虑了描述实际过程的拓扑矩阵,故能更快速地反映子块1中故障的特征,进而能相对较好地监测TE过程的异常状态.如图5(a)所示,该监测结果是利用传统分块PCA通过等权值融合后的监测结果.

综上所述,3种方法的监测效果都较为理想,其中本文方法在利用大量压缩数据后,仍能保持较好的监

测效果,且敏感度较高.因此,本文方法具有一定的有效性.

### 3.2 加氢裂化过程

加氢裂化过程是一个典型的石化工业过程.如图7所示,该过程包含两个主反应器,加氢精制反应器和加氢裂化反应器,4个高低压分离器以及一个主分馏塔.

作为油品的二次加工过程,加氢裂化过程的原料油主要有常减压过程中的减压蜡油,以及催化柴油、焦化蜡油等组分相对较重的油.原料油经过混合后,通过加氢裂化反应器出口热油的换热预热后,与加热后的氢气一同进入加氢精制反应器.加氢精制反应器的主要功能是去除原料油中的S、N以及O等化学元素杂质.净化后的原料油送至加氢裂化反应器,将大分子原料油在催化剂的作用下裂解为小分子,即常用的产品油,如石脑油、航空煤油和柴油等.因为加氢精制和加氢裂化两个反应器中的大部分化学反

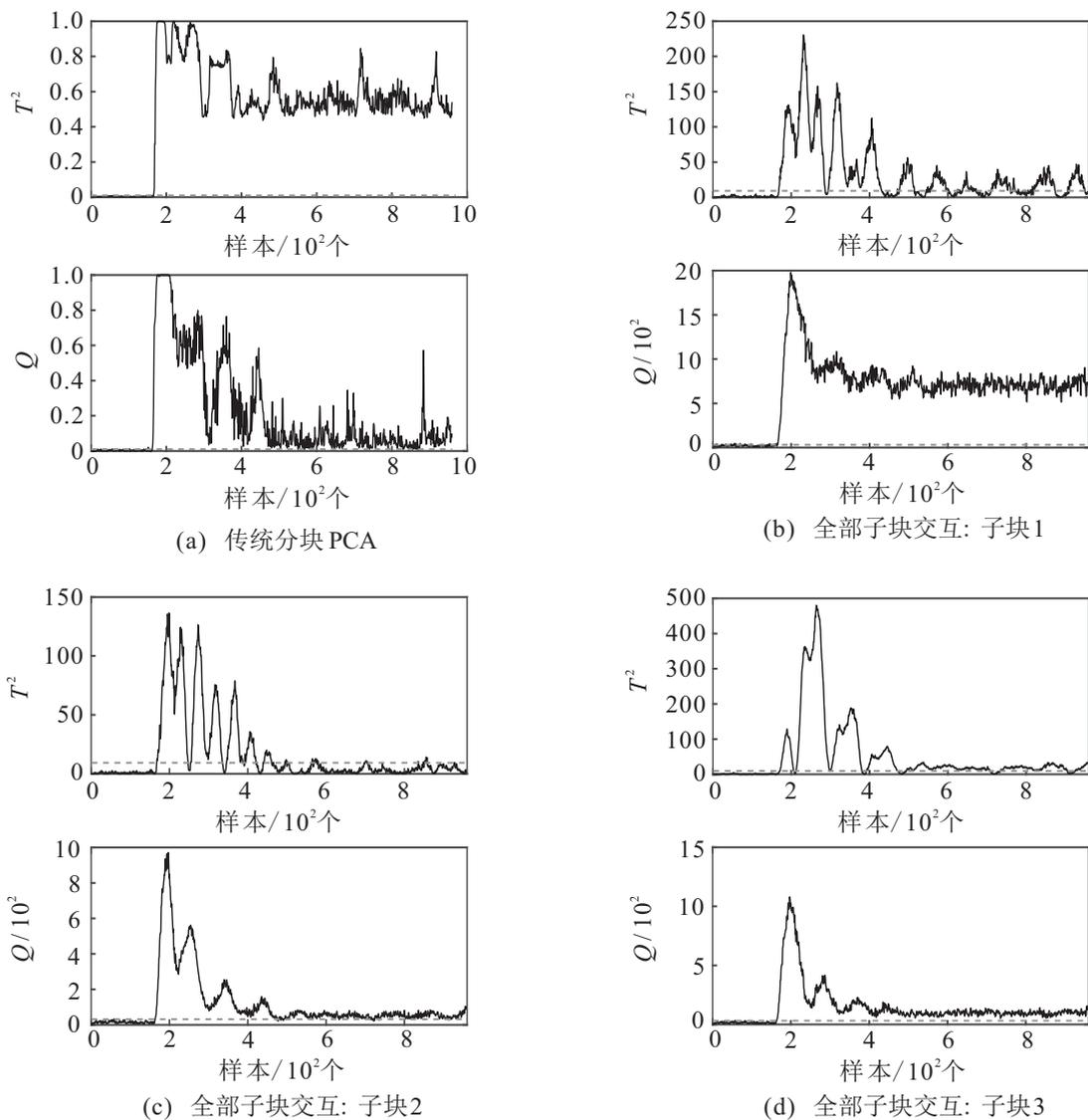


图5 分块PCA及全部子块交互PCA故障1监测结果

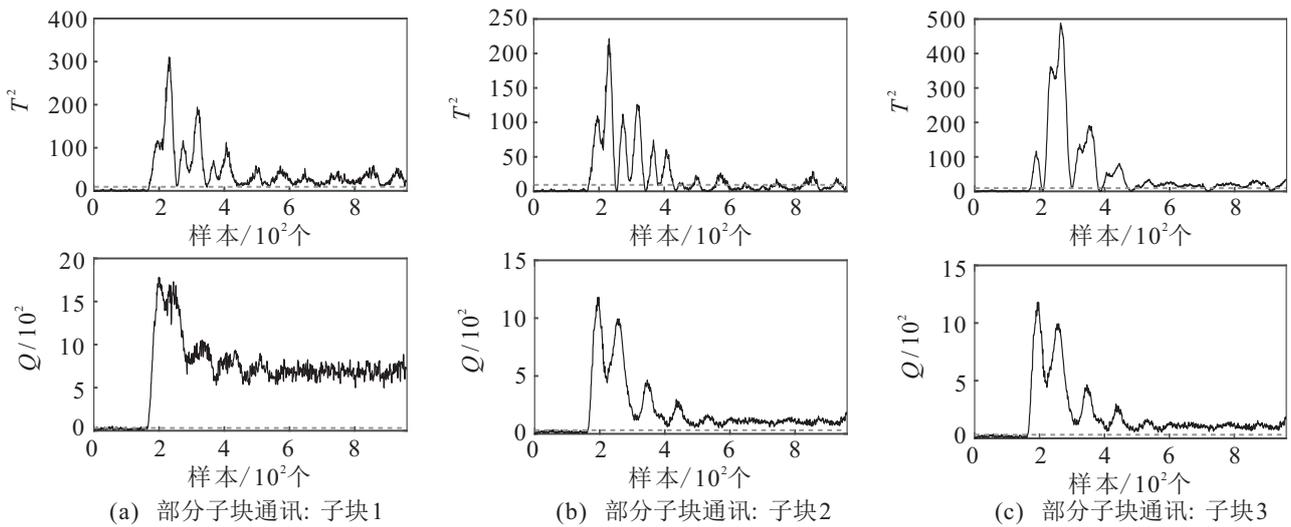


图6 部分子块通讯PCA故障1监测结果

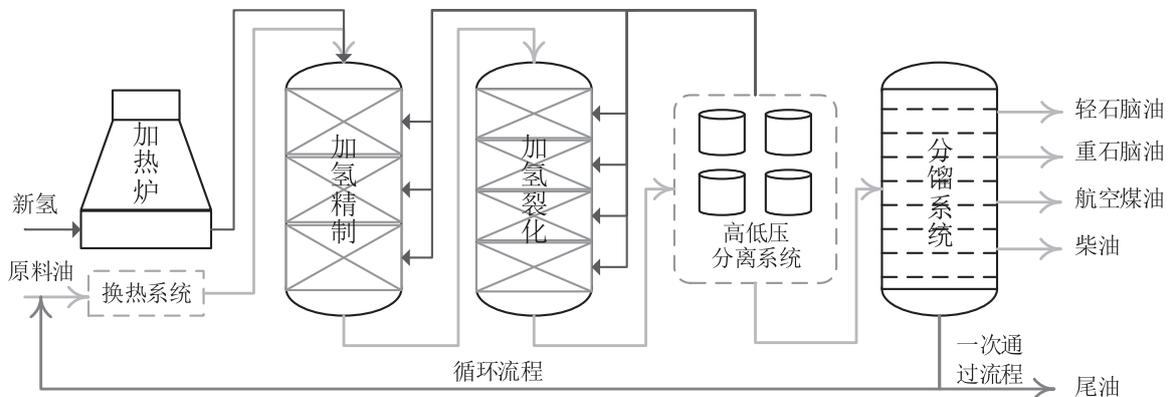


图7 加氢裂化流程简图

应发出热量,所以两个反应器的每个床层间均有过量的冷氢注入,一方面可以降低反应带来的温升,另一方面增加氢分压. 4个高低压分离器用于分离从加氢裂化反应器塔底生成的油气混合物,使其进行气液分离,回收一部分过来的氢气,并将混合油送至分馏系统. 主分馏塔按照不同产品油的物理馏程将混合油通过各个测线抽出,达到分离的目的,同时,塔底有尾油(一部分未反应的原料油和重质油)排出,按照生产计划不同,尾油可能重新返回入料系统再次加工,或送至下一道工序进行加工.

氢气是加氢裂化过程的必要原料,而且为了保持加氢裂化氢气管网中氢气的分压,需要从其他装置中注入大量的新氢. 通常加氢裂化使用的氢气由煤制氢炉供给,然而煤制氢炉的工作状态不稳定,时常会发生故障,导致新氢供给量不足,加氢裂化氢气管网压力下降,这使加氢裂化过程不得不减负荷生产. 如果煤制氢炉的故障没有被及时发现,则加氢精制与加氢裂化反应将无法按照期望进行,甚至会引起严重的事故. 在本文中,选取某次煤制氢炉故障导致氢气管网压力异常的案例作为验证本文方法的实例.

在这个实例中,前80个采样点为加氢裂化正常运行的数据,从第81个开始出现煤制氢炉故障. 需要注意的是,随着煤制氢炉故障的发生,现场工人不断进行相应的调整,使加氢裂化过程逐渐返回正常工作状态. 在第350个采样点时出现人工调整的波动情况,使加氢裂化过程出现相应的波动.

如图8和图9所示,传统分块PCA、全部子块交互的分布式PCA以及本文所提出的部分子块通讯的分布式PCA都得到了较好的监测结果. 本文所提出方法,在利用压缩数据的前提下,监测效果与传统分块PCA方法接近,表明本文方法具有一定的可行性. 对比图8(b)和图9(a),尤其在第130个采样点左右,部分子块连通监测方法中的子块1相比于全部子块交互方法的子块1监测效果要好. 此外,由图8和图9的监测结果可见,子块1监测出较为明显的故障信息,表明故障很可能发生在子块1上,这与实际加氢裂化过程中煤制氢炉故障相符,同时,也表明监测方法可以对异常现象发生的原因进行初步定位,从而验证了本文方法的合理性和有效性.

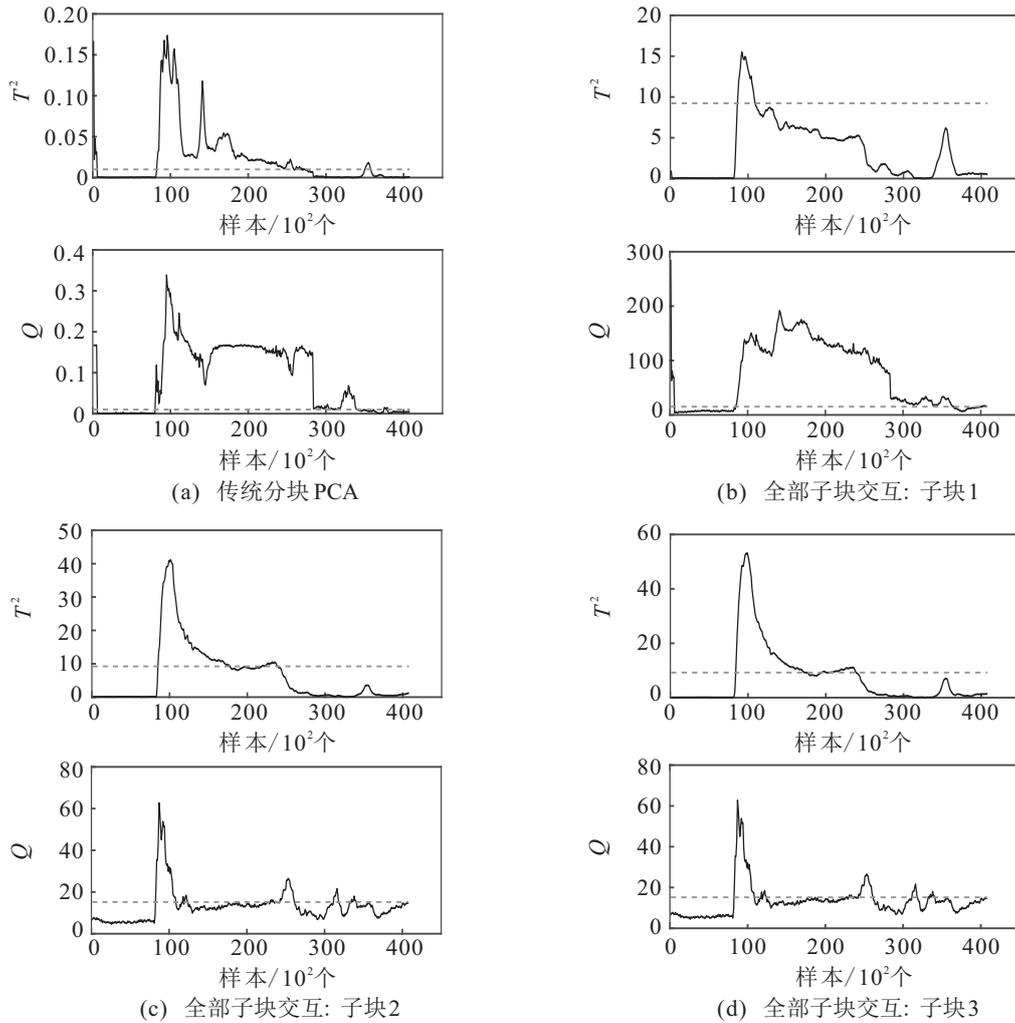


图8 分块PCA及全部子块交互PCA循环氢异常监测结果

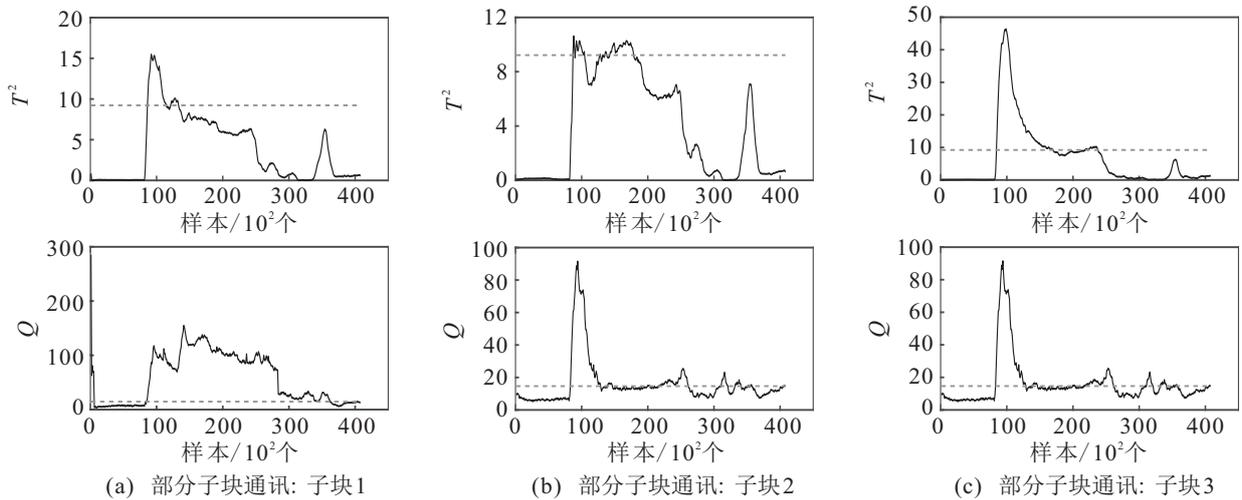


图9 部分子块通讯PCA循环氢异常监测结果

### 4 结论

本文提出了一种部分子块通讯的分布式PCA工业过程监测方法. 该方法首先通过机理对工业过程进行分块, 再结合实际工业过程, 通过所提出的拓扑矩阵描述子块间的内在联系, 使监测模型按照该矩阵利用大量压缩数据进行迭代更新, 进而得到更准确的

分布式监测模型. 本文所提出的方法考虑了子块间的连接方式及异常信息压缩通讯, 增加了方法的监测敏感度, 且能初步确定故障发生的大致范围. 最后, 将本文方法应用于TE过程和加氢裂化过程, 并与传统分块PCA、全部子块交互的分布式PCA方法进行对比, 验证了该方法的合理性和有效性.

## 参考文献(References)

- [1] Ding S X. Model-based fault diagnosis techniques[M]. Berlin, Heidelberg: Springer, 2008: 21-50.
- [2] Hoskins J C, Himmelblau D M. Artificial neural network models of knowledge representation in chemical engineering[J]. Computers & Chemical Engineering, 1988, 12(9): 881-890.
- [3] Lee H J, Ahn B S, Park Y M. A fault diagnosis expert system for distribution substations[J]. IEEE Transactions on Power Delivery, 2000, 15(1): 92-97.
- [4] Chen Z, Ding S X, Peng T, et al. Fault detection for non-gaussian processes using generalized canonical correlation analysis and randomized algorithms[J]. IEEE Transactions on Industrial Electronics, 2018, 65(2): 1559-1567.
- [5] Chen Z, Zhang K, Ding S X, et al. Improved canonical correlation analysis-based fault detection methods for industrial processes[J]. Journal of Process Control, 2016, 41: 26-34.
- [6] Yuan X, Ge Z, Song Z, et al. Soft sensor modeling of nonlinear industrial processes based on weighted probabilistic projection regression [J]. IEEE Transactions on Instrumentation & Measurement, 2017, 66(4): 837-845.
- [7] Yuan X, Wang Y, Yang C, et al. Probabilistic density-based regression model for soft sensing of nonlinear industrial processes[J]. Journal of Process Control, 2017, 57: 15-25.
- [8] Zhu J, Ge Z, Song Z. Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data[J]. IEEE Transactions on Industrial Informatics, 2017, 13(4): 1877-1885.
- [9] Jiang Q, Yan X. Plant-wide process monitoring based on mutual information-multiblock principal component analysis[J]. ISA Transactions, 2014, 53(5): 1516-1527.
- [10] Xu C, Zhao S, Liu F. Distributed plant-wide process monitoring based on PCA with minimal redundancy maximal relevance[J]. Chemometrics and Intelligent Laboratory Systems, 2017, 169: 53-63.
- [11] Ge Z, Song Z. Two-level multiblock statistical monitoring for plant-wide processes[J]. Korean Journal of Chemical Engineering, 2009, 26(6): 1467-1475.
- [12] Liu Q, Qin S J, Chai T. Multiblock concurrent PLS for decentralized monitoring of continuous annealing processes[J]. IEEE Transactions on Industrial Electronics, 2014, 61(11): 6429-6437.
- [13] Ge Z, Chen J. Plant-wide industrial process monitoring: A distributed modeling framework[J]. IEEE Transactions on Industrial Informatics, 2016, 12(1): 310-321.
- [14] Jiang Q, Yan X. Nonlinear plant-wide process monitoring using MI-spectral clustering and Bayesian inference-based multiblock KPCA[J]. Journal of Process Control, 2015, 32: 38-50.
- [15] Zhang Y, Zhou H, Qin S J, et al. Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares[J]. IEEE Transactions on Industrial Informatics, 2010, 6(1): 3-10.
- [16] Liu Q, Qin S J, Chai T. Decentralized fault diagnosis of continuous annealing processes based on multilevel PCA[J]. IEEE Transactions on Automation Science & Engineering, 2013, 10(3): 687-698.
- [17] Bertrand A, Moonen M. Distributed canonical correlation analysis in wireless sensor networks with application to distributed blind source separation [J]. IEEE Transactions on Signal Processing, 2015, 63(18): 4800-4813.
- [18] Pearson K. LIII. On lines and planes of closest fit to systems of points in space[J]. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901, 2(11): 559-572.
- [19] Ge Z, Song Z, Gao F. Review of recent research on data-based process monitoring[J]. Industrial & Engineering Chemistry Research, 2013, 52(10): 3543-3562.
- [20] Yin S, Ding S X, Xie X, et al. A review on basic data-driven approaches for industrial process monitoring[J]. IEEE Transactions on Industrial Electronics, 2014, 61(11): 6418-6428.
- [21] Kruger U, Xie L. Statistical monitoring of complex multivariate processes: With applications in industrial process control[M]. West Sussex: John Wiley & Sons, 2012: 83-92.
- [22] Downs J J, Vogel E F. A plant-wide industrial process control problem[J]. Computers & Chemical Engineering, 1993, 17(3): 245-255.
- [23] Shen Y, Ding S X, Haghani A, et al. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process[J]. Journal of Process Control, 2012, 22(9): 1567-1581.

## 作者简介

曹跃(1992-), 男, 博士生, 从事工业过程监测与诊断的研究, E-mail: csucaoyue@csu.edu.cn;

陈志文(1986-), 男, 讲师, 博士, 从事机器学习、数据挖掘与分析、轨道交通安全监测、系统健康管理及故障诊断等研究, E-mail: zhiwen.chen@csu.edu.cn;

袁小锋(1988-), 男, 副教授, 博士, 从事工业大数据和智能制造、机器学习与模式识别、人工智能与深度学习、过程监测与软测量建模等研究, E-mail: yuanxf@csu.edu.cn;

王雅琳(1973-), 女, 教授, 博士生导师, 从事复杂过程建模与控制、智能控制与智能优化、过程模拟与计算机仿真、计算机工程与应用等研究, E-mail: ylwang@csu.edu.cn;

桂卫华(1950-), 男, 教授, 博士生导师, 中国工程院院士, 从事智能优化制造、复杂工业过程建模与优化、复杂工业过程故障诊断等研究, E-mail: gwh@csu.edu.cn.

(责任编辑: 李君玲)