

基于缺失数据的误差生成策略及其在故障检测中的应用

蓝 艇, 朱 莹, 俞海珍, 童楚东[†]

(宁波大学 信息科学与工程学院, 宁波 315211)

摘 要: 误差生成是基于机理模型故障检测方法的核心本质, 但鲜有应用于统计过程监测方法中. 为此, 提出一种基于缺失数据的误差生成策略, 将能反映出采样数据对统计模型拟合程度的误差作为新的被监测对象实施故障检测. 所提出的基于缺失数据的主元分析 (MD-PCA) 方法通过逐一假设各变量测量数据缺失后, 利用缺失数据处理方法推测出相应缺失数据的估计值, 并对缺失数据的实际值与估计值之间的误差实施基于 PCA 模型的故障检测. 利用误差实施故障检测的优势在于, 生成的误差能在一定程度上降低原测量变量的非高斯性程度, 而且误差体现的是对应缺失变量中与其他测量变量不相关的成分信息, 更能揭示各测量变量的本质. 通过在 TE 过程上的实验充分验证了所提出方法的优势, 以及 MD-PCA 方法用于故障检测的可行性与优越性.

关键词: 缺失数据; 主成分分析; 统计过程监测; 误差生成; 故障检测; TE 过程

中图分类号: TP277

文献标志码: A

Missing data based method for residual generation and its application for fault detection

LAN Ting, ZHU Ying, YU Hai-zhen, TONG Chu-dong[†]

(Faculty of Electrical Engineering & Computer Science, Ningbo University, Ningbo 315211, China)

Abstract: Residual generation is the essential step in the model-based fault detection methods, but it has not been applied in the statistical process monitoring approaches. Therefore, a missing data based residual generation strategy is proposed, the generated residual which can indicate the fitness of sampled data to the developed statistical model is utilized for fault detection. The proposed missing data based principal component analysis (MD-PCA) method first assumes the measured data of individual variables is missing one by one, and the technique handling missing data is then employed for calculating the estimation of the corresponding missing variable. Ultimately, the residual between the actual and estimated data is modeled and monitored using the PCA-based fault detection approach. The advantages of utilizing residual for fault detection lie in that the generated residual can reduce the non-Gaussianity of the original measured variable to some extent, and that the residual reflects the uncorrelated information from other measured variables in the corresponding missing variable, and more essential characteristic of individual variables can be recovered. The case study in the TE process sufficiently demonstrates these advantages of the proposed method, and the feasibility and superiority of the MD-PCA method are validated as well.

Keywords: missing data; principal component analysis; statistical process monitoring; residual generation; fault detection; TE process

0 引 言

随着市场竞争的日趋激烈化, 保证生产过程的安全性及产品质量的稳定性是企业稳健发展的根本途径. 近几十年来, 以故障检测为核心要素的过程监测方法与技术一直都是工业界和学术界重点关注的对象^[1-4]. 在故障检测研究的起始阶段, 基于机理模型设计观测器是主流的实施方案. 通常来讲, 误差生成是基于机理模型方法的核心^[2-3]. 这类方法大多利用过

程对象中某些参数实际值与估计值之间的误差来反应过程运行状态, 以达到实时故障检测的目的^[5-6]. 这些被监测的误差信息在正常运行状态下一般都服从高斯分布, 只有在异常状态下才会出现较大波动. 然而, 随着现代化工业对象规模的不断复杂化与大型化, 获取符合一定精度要求的机理模型变得越来越不切实际, 取而代之的是, 基于数据的过程建模与故障检测方法^[7-9]. 这主要得益于先进仪表与计算机技术

收稿日期: 2018-04-23; 修回日期: 2018-08-06.

基金项目: 国家自然科学基金项目(61503204, 61773225); 浙江省自然科学基金项目(LY16F030001).

责任编辑: 方华京.

[†]通讯作者. E-mail: tongchudong@nbu.edu.cn.

在工业过程中的广泛应用,海量的生产过程数据可以很容易地被测量与储存. 与基于机理模型的故障检测方法生成误差不同,基于数据的故障检测方法旨在挖掘过程数据中的潜藏的有用信息,并对这些有用信息进行监测.

在诸多的基于数据的方法中,多变量统计过程监测(Multivariate statistical process monitoring, MSPM)是最常见的故障检测方法. 其中,有以主元分析(Principal component analysis, PCA)与独立元分析(Independent component analysis, ICA)为最主流的数据分析与建模方法,已被广泛用于工业过程监测中,各种延伸拓展形式层出不穷^[10-15]. 一般来讲,PCA算法的宗旨在于捕捉原始高维数据变量间的相关性特征,使提取的低维主元子空间最大化保留原始数据的方差信息. 然而,PCA模型用于监测这些主元时,需要假设过程数据服从或近似服从高斯分布以方便地确定出监测统计量的控制上限. 若不满足高斯分布假设,则PCA故障检测模型无法较好地描述正常数据的波动范围,相应地故障检测能力将会大幅下降. 相比较之下,ICA方法不存在这个假设,能更好地适应于非高斯过程对象的建模与监测问题^[16-18].

值得一提的是,无论是PCA还是ICA,对应的故障检测模型通常都是采用基于距离的监测统计量来定义正常数据范围. 例如,Hotelling's T^2 与 Q 统计监测指标分别定义的是马氏距离与欧氏距离. 从几何空间角度看,PCA与ICA在将原始数据进行转换后,定义了一个超椭球(或球体)来限定其波动范围. 换言之,所有位于超椭球之外的数据点都是故障数据,而处于其内的点则为正常数据样本. 可想而知,若是这些经PCA或ICA转换过后的特征成分不服从高斯分布,则定义的超椭球内部就存在较多的“空洞”,处于这些“空洞”中的点应该是故障数据,但是PCA或ICA却无法将其有效识别出来.

受基于机理模型误差生成思路的启发,本文在传统PCA模型的基础上提出一种基于缺失数据的误差生成方法,并将其用于故障检测中. PCA模型中的缺失数据处理方法一般都在真实存在缺失数据的情况下才会使用,若不考虑缺失数据的影响,则这类方法就不会得到关注. 本文借鉴缺失数据方法旨在生成过程变量实际值与估计值之间的误差,并利用误差实施故障检测. 与利用机理模型描述过程对象输入-输出关系的不同之处在于,本文所提出的基于缺失变量的PCA(Missing data based PCA, MD-PCA)方法利用的是从数据角度出发建立的统计模型,所生成的误差

主要体现了采样数据对该统计模型的拟合程度. 通过在Tennessee Eastmann(TE)^[19-21]仿真实验平台上的实验中可以发现,MD-PCA方法可在一定程度上降低原测量变量的非高斯性特征,而且生成的误差与其他测量变量之间不存在显著的相关性,更能揭示对应缺失变量的本质. 此外,在TE过程上的对比分析还表明了该方法的优越性与可行性.

1 经典过程监测模型

1.1 基于PCA的过程监测方法

作为一种被广泛使用的多变量数据分析工具,PCA算法旨在按如下所示形式分解经标准化后的训练数据 $\mathbf{X} \in \mathbf{R}^{n \times m}$ (n 为样本数, m 为变量数)^[22]:

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

其中: $\mathbf{T} \in \mathbf{R}^{n \times m}$ 和 $\mathbf{P} \in \mathbf{R}^{m \times m}$ 分别是得分矩阵和载荷矩阵; $\mathbf{E} = \mathbf{T}_r\mathbf{P}_r^T = \mathbf{X}(\mathbf{I} - \mathbf{P}_s\mathbf{P}_s^T) \in \mathbf{R}^{n \times m}$ 为残差矩阵; $\mathbf{T}_s \in \mathbf{R}^{n \times d}$ 和 $\mathbf{P}_s \in \mathbf{R}^{m \times d}$ 分别是PCA模型中主元子空间的得分矩阵与载荷矩阵,分别对应于 \mathbf{T} 和 \mathbf{P} 中的前 d 列. 参数 $d < m$ 为选取的占主导地位的主元个数,其取值通常采用累计方差贡献率(Cumulative percentage variance, CPV)或交叉验证(Cross validation, CV)法来确定. 主元子空间中的载荷矩阵 \mathbf{P} 其实是由相关性矩阵 $\mathbf{C} = \mathbf{X}^T\mathbf{X}/(n-1)$ 前 d 个最大特征值 $\lambda_1, \lambda_2, \dots, \lambda_d$ 所对应的特征向量组成. 因此,PCA算法旨在挖掘训练数据中各变量间的相关性.

实施在线故障检测时,传统PCA故障检测模型主要依赖如下两个统计量监测新样本 $\mathbf{x} \in \mathbf{R}^{m \times 1}$ 是否出现故障,即

$$T^2 = \mathbf{x}^T \mathbf{P}_s \mathbf{A}_s \mathbf{P}_s^T \mathbf{x} \leq \frac{d(n-1)}{n-d} F_\alpha(d, n-d), \quad (2)$$

$$Q = \mathbf{x}^T (\mathbf{I} - \mathbf{P}\mathbf{P}^T) \mathbf{x} \leq g \chi_{h,\alpha}^2. \quad (3)$$

其中: $\mathbf{A}_s = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_d\}$, α 为置信限, $F_\alpha(d, n-d)$ 表示自由度为 d 与 $n-d$ 的 F 分布, $g \chi_{h,\alpha}^2$ 表示权重为 $g = v/2b$ 、自由度为 $h = 2a^2/b$ 的 χ^2 分布, a 与 b 分别是训练数据集对应的 Q 统计量的估计均值和估计方差^[22].

近年来,分布式的PCA方法得到了众多学者的青睐,这主要得益于多模型的效果通常优于单个模型^[23]. Ge等^[14]曾利用PCA模型中各载荷向量对过程变量实施分解,累计得到 $d+1$ 个子变量块,并对各个子块分别建立基于PCA的过程监测模型. Ge等^[14]提出的这种Distributed PCA方法不需要过程的机理知识,而完全是从数据统计角度出发实施过程分解的. 通过贝叶斯概率融合的方式,可将 $d+1$ 个PCA

模型的 T^2 与 Q 监测指标分别融合成两个概率型指标 BIC_{T^2} 与 BIC_Q ,以方便最终故障决策.

1.2 基于ICA的过程监测方法

与PCA算法类似,ICA旨在从训练数据集 \mathbf{X} 中提取 $k < m$ 个独立元,即

$$\mathbf{X}^T = \mathbf{A}\mathbf{s} + \mathbf{F}. \quad (4)$$

其中: \mathbf{F} 为模型残差矩阵, $\mathbf{A} \in \mathbf{R}^{m \times k}$ 与 $\mathbf{s} \in \mathbf{R}^{k \times n}$ 分别是混合矩阵与独立元矩阵.求解ICA模型通常依赖于传统的FastICA迭代算法,但是该FastICA迭代过程对初始值很敏感,迭代结果不稳定.为此, Lee等^[24]提出一种改进形式的Modified ICA算法.该算法首先利用PCA将原始数据白化成互不相关的主元,然后迭代求解混合矩阵 $\mathbf{W} \in \mathbf{R}^{k \times n}$,从而从训练数据 \mathbf{X} 中分离出独立元 \mathbf{s} ,即

$$\mathbf{s} = \mathbf{W}\mathbf{X}^T. \quad (5)$$

值得注意的是,PCA白化过程需尽可能包含最多个数的主元,但可以删除特征值几乎等于0所对应的主元^[24]. Modified ICA(MICA)算法的具体实施细节可参考文献[24].

与PCA不同的是,ICA模型中潜藏成分 \mathbf{s} 是按照非高斯最大化原则提取的,一般不服从高斯分布.但是,ICA模型用于在线故障检测时,同样采取了与PCA模型相似的统计量

$$I^2 = \mathbf{y}D^{-1}\mathbf{y}^T, \quad Q = (\mathbf{x} - \mathbf{x}\mathbf{W}^T\mathbf{A}^T)(\mathbf{x} - \mathbf{x}\mathbf{W}^T\mathbf{A}^T)^T. \quad (6)$$

其中: $\mathbf{y} = \mathbf{x}\mathbf{W}^T \in \mathbf{R}^{l \times r}$ 为当前监测数据 \mathbf{x} 所对应的独立元, $D = \mathbf{s}\mathbf{s}^T / (n - 1)$ 为训练数据独立元的协方差矩阵.由此可看出,传统PCA与ICA故障检测模型都依赖于基于距离的统计指标,即 I^2 为平方马氏距离, Q 为平方欧氏距离.无论监测指标的上限如何确定,其定义的正常数据变化范围都是超椭球体或超球体.因此,降低非高斯性对于距离型监测指标具有重要意义.

MICA算法提取的是非高斯成分信息,模型残差 \mathbf{F} 则主要保留了训练数据中的高斯变化信息.因此,可进一步利用PCA算法对残差 \mathbf{F} 进行分析,相应的故障检测方法称之为MICA-PCA,采用了 I^2 、 T^2 以及 Q 统计量实施监测^[25].

2 基于缺失数据的故障检测方法

2.1 PCA模型中的缺失数据处理方法^[26]

在生产过程实际采样中,有时会出现某个或某几个测量变量的采样数据缺失.针对这些缺失数据,传

统PCA模型无法直接计算出其对应的主元,但可以采取如下方式估计出相应的主元信息.不失一般性,设缺失数据向量可以描述成

$$\mathbf{x} = [\mathbf{x}^\#, \mathbf{x}^{*\top}]^T \in \mathbf{R}^{m \times 1}, \quad (7)$$

其中 $\mathbf{x}^\#$ 和 \mathbf{x}^* 分别表示缺失数据和已测量到的数据.由于PCA模型已知,载荷矩阵也可以做类似划分,即 $\mathbf{P} = [\mathbf{P}^\#, \mathbf{P}^*]$.相关性矩阵可以写成如下形式:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{P}^\# \mathbf{\Lambda} \mathbf{P}^{\#\top} & \mathbf{P}^\# \mathbf{\Lambda} \mathbf{P}^{*\top} \\ \mathbf{P}^* \mathbf{\Lambda} \mathbf{P}^{\#\top} & \mathbf{P}^* \mathbf{\Lambda} \mathbf{P}^{*\top} \end{bmatrix}, \quad (8)$$

相应的缺失数据估计值为

$$\hat{\mathbf{x}}^\# = \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \mathbf{x}^* = \mathbf{P}^\# \mathbf{\Lambda} \mathbf{P}^{*\top} (\mathbf{P}^* \mathbf{\Lambda} \mathbf{P}^{*\top})^{-1} \mathbf{x}^*. \quad (9)$$

2.2 基于误差生成的故障检测方法

在前述缺失数据处理方法的基础上,提出如图1所示的MD-PCA故障模型.首先,依次假设新采样数据 $\mathbf{x} \in \mathbf{R}^{m \times 1}$ 中第 i 个变量出现数据缺失,并利用已建立的PCA模型推测出相应的估计值 $\hat{\mathbf{x}}_i^\#$,即

$$\hat{\mathbf{x}}_i^\# = \mathbf{P}^\# \mathbf{\Lambda} \mathbf{P}^{*\top} (\mathbf{P}^* \mathbf{\Lambda} \mathbf{P}^{*\top})^{-1} \mathbf{x}_i^*. \quad (10)$$

其中 $\mathbf{x}_i^* \in \mathbf{R}^{(m-1) \times 1}$ 为向量 \mathbf{x} 中的可观测数据, $i = 1, 2, \dots, m$.然后,计算缺失数据的实际值 $\mathbf{x}_i^\# \in \mathbf{R}^{1 \times 1}$ 与估计值 $\hat{\mathbf{x}}_i^\#$ 之间的误差

$$e_i = \mathbf{x}_i^\# - \hat{\mathbf{x}}_i^\#. \quad (11)$$

最后,对误差向量 $\mathbf{e} = [e_1, e_2, \dots, e_m]^T$ 实施基于传统PCA模型的故障检测.值得注意的是,逐一假设各变量测量数据缺失时不需要按照一定的顺序执行,但需要保证有且仅有一次假设各个变量的测量数据缺失.

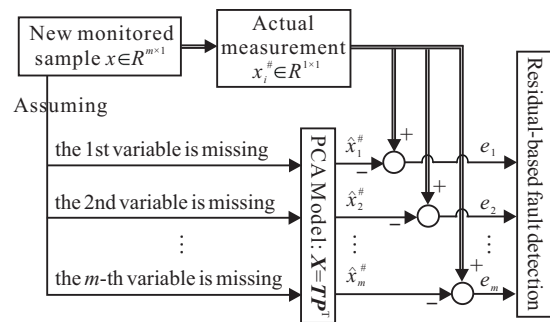


图1 基于缺失数据的PCA故障检测模型

2.3 基于MD-PCA的故障检测流程

如图2所示,基于MD-PCA的故障检测实施流程分为离线建模与在线故障检测两个部分,详细的实施步骤如下所示.

离线建模流程:

1) 对正常工况下的训练数据 $\mathbf{X} \in \mathbf{R}^{n \times m}$ 实施标准化处理,以消除变量间量纲差异的影响.

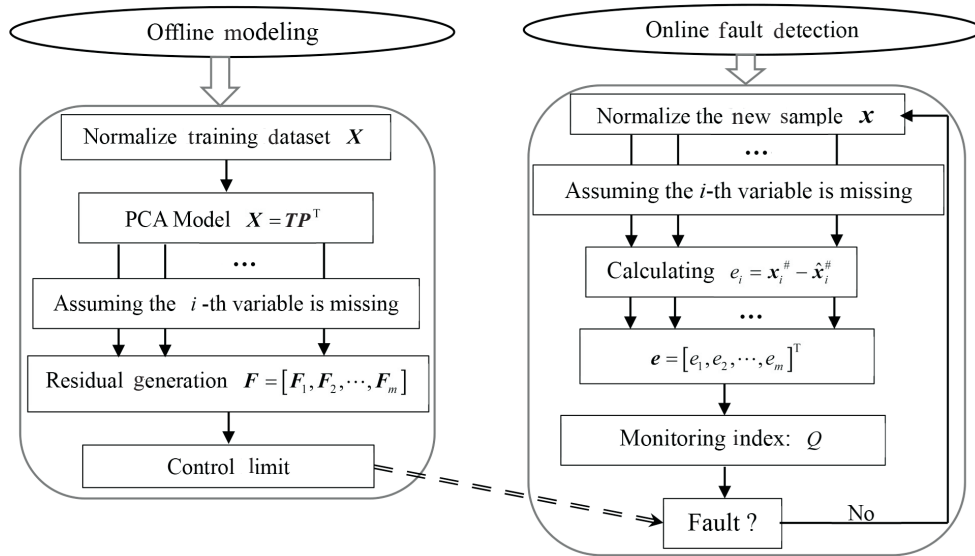


图2 MD-PCA的实施流程

2) 利用PCA算法对标准化后的数据 X 建立PCA模型 $X = TP^T$.

3) 假设 X 中第 i 个变量的测量数据缺失, 并依据式(10)和(11)计算相应的估计误差 $F_i \in R^{n \times 1}$.

4) 确定对误差实施监测的统计量的控制上限

$$Q_c = g\chi_{h,\alpha}^2 \quad (12)$$

其中: $g = \text{tr}(C_F(C_F + \varepsilon I)^{-1})^2 / \text{tr}(C_F(C_F + \varepsilon I)^{-1})$, 自由度 $h = [\text{tr}(C_F(C_F + \varepsilon I)^{-1})]^2 / \text{tr}(C_F(C_F + \varepsilon I)^{-1})^2$, 协方差矩阵 $C_F = F^T F / (n - 1)$, $F = [F_1, F_2, \dots, F_m]$. 由于误差 F 的共线性问题, 无法直接求取 C_F 的逆矩阵, 因此需增加一个矫正系数 ε , ε 的选取只需要显著小于误差 F 中的方差即可^[27].

在线故障检测实施流程:

5) 对新采样数据 x 实施与步骤1)中一样的标准化处理.

6) 依次假设 x 中第 i 个数据缺失, 并利用式(10)和(11)计算相应的估计误差 e_i .

7) 按照式(2)和(3)对得到的误差向量 $e = [e_1, e_2, \dots, e_m]^T$ 计算监测统计量 $Q = e^T(C_F + \varepsilon I)^{-1}e$, 并判断是否超限.

3 仿真案例研究

TE 仿真模型因其结构的复杂性, 已成为不同控制方法和过程监测策略的标准实验平台^[28-29]. TE 过程主要由连续搅拌反应器、产品冷凝器、气液分离塔、汽提塔和离心式压缩机等5个生产单元组成, 可连续测量22个过程变量和12个操作变量, 还可以仿真模拟如表1所列的21种不同的故障. 在本文的研究中, 选取如参考文献[24]中所列的33个变量作为监测变

量. 离线建模阶段, 利用正常工况下的960个样本分别建立基于MD-PCA、PCA、MICA、Distributed PCA、和MICA-PCA的故障检测模型. 其中, 置信限统一取值99%, PCA模型中采用CPV > 85%确定保留的主元个数, MICA模型中各参数与参考文献[24]一致, 白化过程时剔除特征值小于 10^{-3} 所对应的主元.

首先, 利用TE过程的训练数据验证MD-PCA方法具备降低原测量变量非高斯程度的能力. 选取TE过程33个连续测量变量中非高斯特征较为明显的3个变量, 即第10、第19和第31号变量, 将其各自的高斯分布检验图及其对应的误差高斯分布检验图呈现于图3中. 从图3(a)中可以明显地发现, TE过程中这3个测量变量明显不满足高斯分布假设. 而经过本文所提出的误差生成方法转换成相应误差 θ_{10} 、 θ_{19} 、 θ_{31} 后, 图3(b)所示的误差非高斯程度得到明显降低, 这对于利用距离型监测指标实施故障检测的PCA方法而言具有重要性意义. 因为, 监测对象越接近于高斯分布, 距离型监测指标所定义的超椭球(或球体)中“空洞”现象出现的几率就会越少, 所定义的正常范围也就越准确.

其次, 利用TE过程另一组正常工况下的测试数据集以及18种故障工况下测试数据集, 对比验证MD-PCA方法相比于其他4种故障检测方法的优越性与实用性, 相应的误报率与故障漏报率详情列于表1中. 由于TE过程第3、第9和第15种故障对采样数据的影响甚微, 已被很多文献证实它们很难被可靠地检测出来^[30-33], 在本文的研究中不予考虑. 针对另一组包含500个正常样本的训练数据集的误报率情况

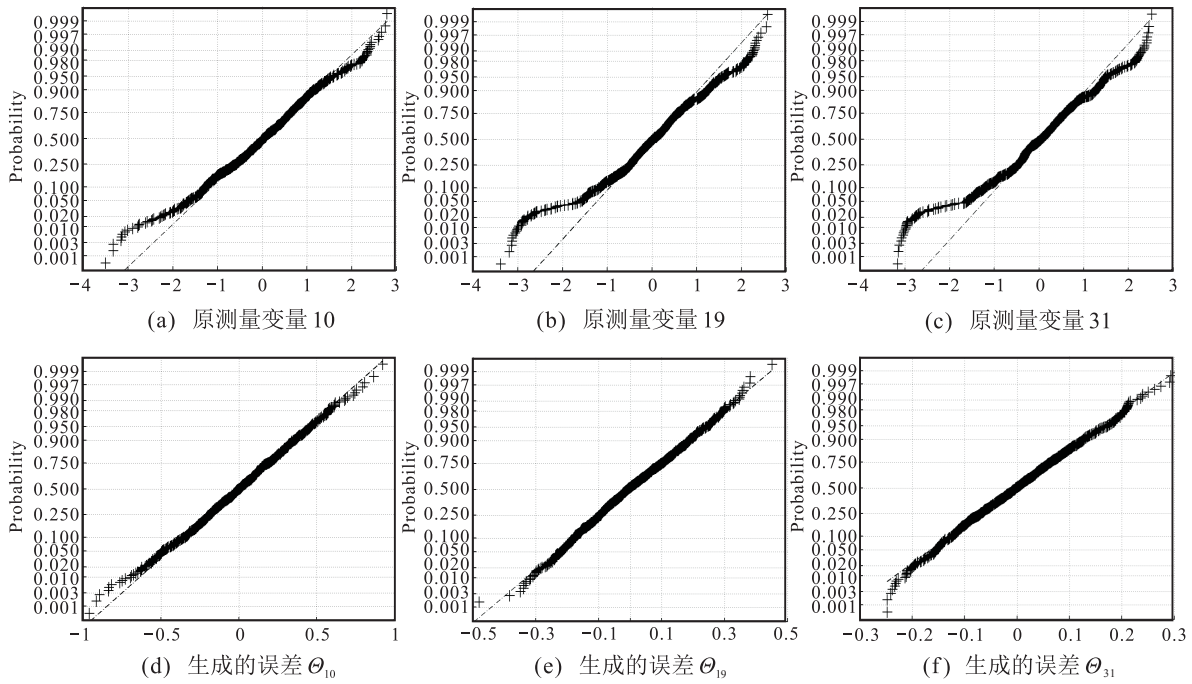


图3 测量变量及其对应误差的高斯分布检验

表1 TE过程各工况的监测结果对比

No.	Description	MD-PCA	PCA		Modified ICA		Distributed	PCA	MICA-PCA		
		<i>Q</i>	<i>T</i> ²	<i>Q</i>	<i>I</i> ²	<i>Q</i>	<i>BIC</i> _{<i>T</i>₂}	<i>BIC</i> _{<i>Q</i>}	<i>I</i> ²	<i>T</i> ²	<i>Q</i>
0	Normal operating condition	1.60	0.40	2.20	0.80	0.60	0.00	1.80	0.80	0.40	1.00
1	A/C feed ratio, B composition constant.	0.00	0.88	0.13	0.25	0.50	0.88	0.25	0.25	0.25	0.75
2	B composition, A/C ratio constant.	2.75	1.63	4.00	2.50	1.38	1.63	1.50	2.50	1.75	3.13
4	Reactor cooling water inlet temperature.	0.00	79.13	0.00	32.63	5.38	11.13	0.00	32.63	44.25	0.00
5	Condenser cooling water inlet temperature.	0.00	75.75	75.88	76.13	74.25	75.00	66.63	76.13	74.38	81.75
6	A feed loss.	0.00	0.88	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00
7	C header pressure loss-reduced availability.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50
8	A,B, C feed composition.	2.25	3.13	13.88	2.25	98.88	2.7	52.25	2.25	2.25	16.00
10	C feed temperature.	13.88	70.13	70.88	24.38	29.13	60.88	39.38	24.38	28.63	44.13
11	Reactor cooling water inlet temperature.	16.88	59.38	23.88	62.13	31.13	27.50	9.50	62.13	37.25	46.50
12	Condenser cooling water inlet temperature.	0.25	1.63	9.25	2.50	1.38	0.88	0.63	2.50	2.13	15.38
13	Reaction kinetics.	4.75	6.38	4.75	4.88	5.50	5.75	5.00	4.88	5.13	9.00
14	Reactor cooling water valve.	0.00	0.75	0.00	0.13	0.00	0.00	0.00	0.13	0.00	1.38
16	Unknown	10.63	86.50	67.75	24.00	24.75	80.00	43.13	24.00	30.13	26.00
17	Unknown	2.50	23.63	4.13	12.25	6.38	10.13	2.25	12.25	14.00	4.50
18	Unknown	9.88	10.75	9.75	10.50	9.63	10.75	9.38	10.50	9.75	10.25
19	Unknown	7.63	89.00	82.25	50.00	39.13	77.00	51.13	50.00	61.63	25.13
20	Unknown	10.00	68.25	48.38	33.88	33.38	59.63	41.13	33.88	34.25	36.63
21	Fixed position of valve in stream 4.	56.38	60.75	51.13	46.00	76.38	56.38	44.63	46.00	68.50	91.63

来看, MD-PCA 方法的误报率虽不是最小,但也不是最大,误报率都在可接受范围内. 针对 TE 过程 18 种故障工况的监测中,取得最小故障漏报率的数值已用粗体标出. 由表 1 可以发现,MD-PCA 方法在绝大多数的故障监测上都能取得最小的故障漏报率,而且 MD-PCA 方法显著地改善了针对 TE 过程故障 5、故

障 10、故障 16、故障 19 和故障 20 的故障检测效果,相应的故障漏报率得到了大幅度的降低. 此外,采用多模型实施监测的 Distributed PCA 方法效果整体优越于传统 PCA 方法. 虽然,在故障 2、故障 11、故障 17、故障 18 和故障 21 的监测上,MD-PCA 方法未取得最佳的监测效果,但是,故障漏报率的差别是很微小的,不

足以引起重视。

最后,将故障5的过程监测详情展示于图4中。从图4的对比中可以发现,MD-PCA方法的 Q 指标能持续不断地触发故障警报,而PCA与MICA都存在较明显的漏报情况。

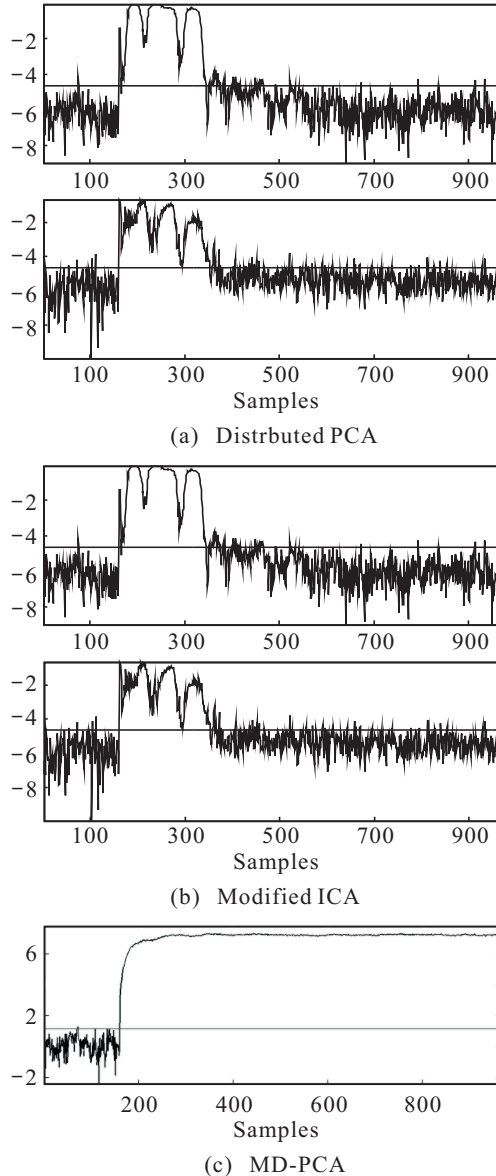


图4 故障5的过程监测结果

4 结论

本文提出了一种基于缺失数据误差生成方法的故障检测模型,并通过对比实验验证了该方法的优越性。由于MD-PCA方法监测的对象为缺失数据实际值与估计值之间的误差,而误差一般都服从或近似服从高斯分布,在原始数据不服从高斯分布的情况下,PCA故障检测模型的高斯分布假设依旧得以满足。这在很大程度上扩大了传统PCA故障检测方法的适用范围。另外,从TE过程上的对比实验也可以发现,MD-PCA方法丝毫不逊色于能应对非高斯过程监测问题的Modified ICA方法,而且在某几个故

障上还能体现出较大优势。然而,本文的研究只限于将MD-PCA应用于故障检测,还未涉及后续的故障诊断,未来需开展相应的研究工作。

参考文献(References)

- [1] Yin S, Ding S X, Xie X, et al. A review on basic data-driven approaches for industrial process monitoring[J]. *IEEE Trans Industrial Electronics*, 2014, 61(11): 6418-6428.
- [2] Tong C, El-farra N H, Palazoglu A, et al. Fault detection and isolation in hybrid process systems using a combined data-driven and observer-design methodology[J]. *AIChE Journal*, 2014, 60(8): 2805-2814.
- [3] Tidiri K, Chatti N, Verron S, et al. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges[J]. *Annual Reviews in Control*, 2016, 42: 63-81.
- [4] Ge Z. Review on data-driven modeling and monitoring for plant-wide industrial processes[J]. *Chemometrics & Intelligent Laboratory Systems*, 2017, 171: 16-25.
- [5] Pierri F, Paviglianiti G, Caccavale F, et al. Observer-based sensor fault detection and isolation for chemical batch reactors[J]. *Engineering Applications of Artificial Intelligence*, 2008, 21: 1204-1216.
- [6] Sotomayor O A Z, Odloak D. Observer-based fault diagnosis in chemical plants[J]. *Chemical Engineering J*, 2005, 112(1): 93-108.
- [7] Ge Z, Song Z, Ding S, et al. Data mining and analytics in process industry: The role of machine learning[J]. *IEEE Access*, 2017, 5: 20590-20616.
- [8] Severson K, Chaiwatanodom P, Braatz R D. Perspectives on process monitoring of industrial systems[J]. *Annual Reviews in Control*, 2016, 42: 190-200.
- [9] 蒋栋年, 李炜. 基于数据驱动残差评价策略的故障检测方法[J]. *控制与决策*, 2017, 32(7): 1181-1188. (Jiang D N, Li W. Fault detection method based on data-driven residual evaluation strategy[J]. *Control and Decision*, 2017, 32(7): 1181-1188.)
- [10] 刘强, 柴天佑, 赵立杰. 基于数据和知识的工业过程监视及故障诊断综述[J]. *控制与决策*, 2010, 25(6): 801-807. (Liu Q, Chai T, Zhao L J. Progress of data-driven and knowledge-driven process monitoring and fault diagnosis for industry process[J]. *Control and Decision*, 2010, 25(6): 801-807.)
- [11] 韩敏, 张占奎. 基于加权核独立成分分析的故障检测方法[J]. *控制与决策*, 2016, 31(2): 242-248. (Han M, Zhang Z K. Fault detection method based on weighted kernel independent component analysis[J]. *Control and Decision*, 2016, 31(2): 242-248.)
- [12] 童楚东, 蓝艇, 史旭华. 基于互信息的分散式动态PCA故障检测方法[J]. *化工学报*, 2016, 67(10): 4317-4323. (Tong C D, Lan T, Shi X H. Fault detection by decentralized dynamic PCA algorithm on mutual

- information[J]. *CIESC Journal*, 2016, 67(10): 4317-4323.)
- [13] Liu Y, Zhang G, Xu B. Compressive sparse principal component analysis for process supervisory monitoring and fault detection[J]. *J of Process Control*, 2017, 50: 1-10.
- [14] Ge Z, Song Z. Distributed PCA model for plant-wide process monitoring[J]. *Industrial & Engineering Chemical Research*, 2013, 52(5): 1947-1957.
- [15] Tong C, Lan T, Shi X H. Fault detection and diagnosis of dynamic processes using weighted dynamic decentralized PCA approach[J]. *Chemometrics & Intelligent Laboratory Systems*, 2017, 161: 34-42.
- [16] Cai L, Tian X. A new fault detection method for non-Gaussian process based on robust independent component analysis[J]. *Process Safety and Environmental Protection*, 2014, 92(6): 645-658.
- [17] Tong C, Lan T, Shi X H. Double-layer ensemble monitoring of non-Gaussian processes using modified independent component analysis[J]. *ISA Transactions*, 2017, 68: 181-188.
- [18] 江伟, 王振雷, 王昕. 基于混合分块DMICA-PCA的全流程过程监控方法[J]. *化工学报*, 2017, 68(2): 759-766.
(Jiang W, Wang Z L, Wang X. Plant-wide process monitoring based on mixed multiblock DMICA-PCA[J]. *CIESC Journal*, 2017, 68(2): 759-766.)
- [19] Downs J J, Vogel E F. A plant-wide industrial process control problem[J]. *Computers & Chemical Engineering* 1993, 17(3): 245-255.
- [20] Yin S, 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]. *J of Process Control*, 2012, 22: 1567-1581.
- [21] Rato T J, Reis M. Fault detection in the Tennessee Eastman benchmark process using dynamic principal components analysis based on decorrelated residuals (DPCA-DR)[J]. *Chemometrics & Intelligent Laboratory Systems*, 2013, 125: 101-108.
- [22] Qin S J. Statistical process monitoring: Basics and beyond[J]. *J of Chemometrics*, 2003, 17(7/8): 480-502.
- [23] Tong C D, Shi X H. Decentralized monitoring of dynamic processes based on dynamic feature selection and informative fault pattern dissimilarity[J]. *IEEE Trans on Industrial Electronics*, 2016, 63(6): 3804-3814.
- [24] Lee J M, Qin S J, Lee I B. Fault detection and diagnosis based on modified independent component analysis[J]. *AIChE Journal*, 2006, 52(10): 3501-3514.
- [25] Ge Z, Song Z. Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors[J]. *Industrial & Engineering Chemical Research*, 2007, 46(7): 2054-2063.
- [26] Nelson P R C, Taylor P A, Macgregor J F. Missing data methods in PCA and PLS: Score calculations with incomplete observations[J]. *Chemometrics & Intelligent Laboratory Systems*, 1996, 35: 45-65.
- [27] Yu J, Qin S J. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models[J]. *AIChE Journal*, 2008, 54(7): 1811-1829.
- [28] 侯平智, 张明, 徐晓滨, 等. 基于K近邻证据融合的故障诊断方法[J]. *控制与决策*, 2017, 32(10): 1767-1774.
(Hou P Z, Zhang M, Xu X B, et al. Fault diagnosis based on KNN evidence fusion[J]. *Control and Decision*, 2017, 32(10): 1767-1774.)
- [29] 童楚东, 史旭华. 基于互信息的PCA方法及其在过程监测中的应用[J]. *化工学报*, 2015, 66(10): 4101-4106.
(Tong C D, Shi X H. Mutual information based PCA algorithm with application in process monitoring[J]. *CIESC Journal*, 2015, 66(10): 4101-4106.)
- [30] Zhang H, Qi Y, Wang L, et al. Fault detection and diagnosis of chemical process using enhanced KECA[J]. *Chemometrics & Intelligent Laboratory Systems*, 2017, 161: 61-69.
- [31] Bernal-de-lazaro J M, Llanes S O, Prieto M A, et al. Enhanced dynamic approach to improve the detection of small-magnitude faults[J]. *Chemical Engineering Science*, 2016, 14: 166-179.
- [32] 熊伟丽, 郭校根. 一种基于多工况识别的过程在线监测方法[J]. *控制与决策*, 2018, 33(3): 403-412.
(Xiong W L, Guo X G. A process on-line monitoring method based on multi-mode identification[J]. *Control and Decision*, 2018, 33(3): 403-412.)
- [33] 刘洋, 张国山. 基于敏感稀疏主元分析的化工过程监测与故障诊断[J]. *控制与决策*, 2016, 31(7): 1213-1218.
(Liu Y, Zhuang G S. Chemical process monitoring and fault diagnosis based on sensitive sparse principal component analysis[J]. *Control and Decision*, 2016, 31(7): 1213-1218.)

作者简介

蓝艇(1976—), 男, 讲师, 博士, 从事计算智能、数据驱动建模等研究, E-mail: lanting@nbu.edu.cn;

朱莹(1976—), 女, 讲师, 博士, 从事数据驱动的工业过程监测与软测量等研究, E-mail: zhuying@nbu.edu.cn;

俞海珍(1975—), 女, 高级实验师, 博士, 从事智能控制算法与优化计算等研究, E-mail: yuhaizhen@nbu.edu.cn;

童楚东(1988—), 男, 副教授, 博士, 从事数据驱动的工业过程监测与软测量等研究, E-mail: tongchudong@nbu.edu.cn.

(责任编辑: 孙艺红)