

基于 DPCA 残差互异度的故障检测与诊断方法

张成¹ 戴絮年¹ 李元¹

摘要 针对动态主元分析方法中残差自相关性降低过程故障检测率问题, 提出基于动态主元分析残差互异度的故障检测与诊断方法. 首先, 应用动态主元分析 (Dynamic principal component analysis, DPCA) 计算动态过程数据的残差得分; 接下来, 应用滑动窗口技术并结合互异度指标 (Dissimilarity) 来监控过程残差得分状态; 最后, 利用基于变量贡献图的方法进行过程故障诊断分析. 本文方法通过 DPCA 捕获过程的动态特征, 同时互异度指标区别于传统的平方预测误差 (Square prediction error, SPE), 它可以有效地对具有自相关性的残差得分进行过程状态监控. 通过一个数值例子和 Tennessee Eastman (TE) 过程的仿真实验并与传统方法对比分析, 仿真结果进一步证实了本文方法的有效性.

关键词 动态主元分析, 互异度, 滑动窗口, 故障诊断

引用格式 张成, 戴絮年, 李元. 基于 DPCA 残差互异度的故障检测与诊断方法. 自动化学报, 2022, 48(1): 292–301

DOI 10.16383/j.aas.c190884

Fault Detection and Diagnosis Based on Residual Dissimilarity in Dynamic Principal Component Analysis

ZHANG Cheng¹ DAI Xu-Nian¹ LI Yuan¹

Abstract Aiming at the problem of reducing process fault detection rate because of residual autocorrelation in dynamic principal component analysis, a novel fault detection and diagnosis based on residual dissimilarity in dynamic principal component analysis is proposed in this paper. Firstly, Dynamic principal component analysis (DPCA) is used to calculate residual score of a dynamic process. Next, moving window technology and dissimilarity index are utilized to monitor the status of this process in residual score. Finally, a fault diagnosis method based on contribution chart of monitored variables is used for discovering the reason causing abnormal change of this process. The proposed method to capture dynamic characteristics of a process through DPCA, meanwhile, the proposed dissimilarity index, which is different from the conventional squared prediction error (SPE), can effectively monitor the process in which the residual scores contain significant autocorrelation. The effectiveness of DPCA-Diss is tested in a numerical case and the Tennessee Eastman (TE) process.

Key words Dynamic principal component analysis (DPCA), dissimilarity, moving window, fault diagnosis

Citation Zhang Cheng, Dai Xu-Nian, Li Yuan. Fault detection and diagnosis based on residual dissimilarity in dynamic principal component analysis. *Acta Automatica Sinica*, 2022, 48(1): 292–301

在现代工艺中, 企业对系统更安全更可靠的需要推动了过程故障检测技术的快速发展. 该类技术通过及时检测过程扰动、设备故障等特殊事件, 不仅在保证工艺和人员安全方面起到了重要作用, 而

且提高了工艺效率和产品质量. 随着分布式控制系统 (Distributed control systems, DCS) 对大量变量的自动测量和存储, 多元统计过程监控 (Multivariate statistical process monitoring, MSPM) 方法被成功地应用于各种工业过程的在线监控领域^[1-3].

针对相关过程变量引起的共线性问题, 主元分析 (Principal component analysis, PCA)^[4-5] 和偏最小二乘 (Partial least squares, PLS)^[6-7] 等方法相继被提出并在化工过程检测领域得到了广泛的应用. PCA 在过程检测时通常将原始空间分为主元子空间 (Principal component subspace, PCS) 和残差子空间 (Residual subspace, RS), 然后分别使用 Hotelling's T^2 和平方预测误差 (Square prediction error, SPE) 作为统计量来监控样本状态^[8]. 其

收稿日期 2019-12-24 录用日期 2020-03-11

Manuscript received December 24, 2019; accepted March 11, 2020

国家自然科学基金项目 (61490701, 61673279), 辽宁省自然科学基金项目 (2019-MS-262), 辽宁省教育厅基金项目 (LJ2019013)

Supported by National Natural Science Foundation of China (61490701, 61673279), Liaoning Natural Science Fund Project (2019-MS-262), Liaoning Provincial Department of Education Fund Project (LJ2019013)

本文责任编辑 杨浩

Recommended by Associate Editor YANG Hao

1. 沈阳化工大学技术过程故障诊断与安全性研究中心 沈阳 110142

1. Research Center for Technical Process Fault Diagnosis and Safety, Shenyang University of Chemical Technology, Shenyang 110142

中, T^2 是归一化主元得分平方和, 即对 PCA 模型内变化的度量, 而 SPE 是平方误差之和, 即对 PCA 模型未捕获变化量的度量. 为了检测非线性过程, 基于核理论的主元分析方法被提出, 即核主元分析 (Kernel PCA, KPCA)^[9]. KPCA 的基本思想是首先通过非线性映射将低维样本数据映射到高维特征空间, 然后在特征空间中执行 PCA 完成故障检测^[10-11]. 值得注意的是, PCA 和 KPCA 方法均应用 T^2 和 SPE 两个统计量进行过程状态监控. T^2 和 SPE 能够较好完成过程监控的前提条件是变量服从多元高斯分布且样本间相互独立^[12]. 事实上, 众多生产过程, 如石油化工、半导体蚀刻工艺等, 变量间存在较强的非高斯特征且样本间存在自相关特征^[13], 这些显著特征制约了上述方法的故障检测性能. 针对上述不足, Ku 等提出了动态主元分析 (Dynamic PCA, DPCA)^[14]. DPCA 通过“时滞移位”方法将过程的静态信息和动态信息实现同步提取. 在进行故障检测时, DPCA 同样采用 T^2 和 SPE 统计量监控样本状态. SPE 统计量主要衡量样本在空间上的变化信息, 而没有考虑样本存在时间信息上的变化. 因此当样本具有时刻相关性时, SPE 统计量通常无法及时检测出故障. Kano 等定义了一个互异度指标来监测过程时间序列数据的分布^[15]. 在故障检测时, 互异度方法和 DPCA 使用不同的统计量监控样本状态. 互异度方法通过对具有不同特征的过程数据分布进行定量分析来检测典型连续过程中操作条件的变化. 近年来, 许多学者对互异度方法进行了深入研究并取得了丰硕的成果, 如 Wang 等提出了基于互异度方法的渐进式故障检测与诊断方法, Zhao 等将互异度方法应用于批次过程故障检测中. 一系列成功的理论研究和应用表明, 互异度方法能够快速、有效地检测过程变量间相关性的变化^[16-17].

DPCA 方法能够提取过程的动态变化信息, 其在化工过程故障检测方面的有效性已经得到了验证^[18-20]. 然而, 文献 [21] 指出, 由 DPCA 获得的主元和残差中仍然保留了较强的自相关性. 这种特征降低了 T^2 和 SPE 控制图的故障检测性能. 针对动态主元分析方法中残差自相关性降低过程故障检测率问题, 本文提出基于 DPCA 残差互异度的故障检测与诊断方法. 首先, 通过 DPCA 方法将输入数据空间划分成主元子空间和残差子空间. 接下来, 对于残差得分应用滑动窗口技术并结合互异度指标完成对样本状态的监控. 该方法能够降低过程动态特征对故障检测的影响, 通过仿真实验进一步证明了本文方法的有效性.

1 动态主元分析和互异度指标

1.1 动态主元分析

假设 $X = [\mathbf{x}(1) \ \mathbf{x}(2) \ \cdots \ \mathbf{x}(m)]$ 为包含 m 个样本的数据集, 其中 $\mathbf{x}(i) \in \mathbf{R}^n$ ($i = 1, 2, \dots, m$). 若过程的时滞参数 $\text{lag} = k$, 则可通过式 (1) 生成包含过程静态特征和动态特征的增广矩阵,

$$Y = [\mathbf{y}(1) \ \mathbf{y}(2) \ \cdots \ \mathbf{y}(m)] \quad (1)$$

其中, $\mathbf{y}(i) = [\mathbf{x}^T(i) \ \mathbf{x}^T(i-1) \ \cdots \ \mathbf{x}^T(i-k)]^T$.

接下来, 对 Y 通过式 (2) 进行标准化处理,

$$\bar{Y} = \Sigma^{-1}(\bar{Y} - \mathbf{m}\boldsymbol{\xi}^T) \quad (2)$$

其中, \mathbf{m} 和 Σ 分别为 Y 的均值向量和标准差对角阵, $\boldsymbol{\xi}$ 为元素全为 1 的列向量. 标准化后的增广矩阵 $\bar{Y} = [\bar{\mathbf{y}}_1 \ \bar{\mathbf{y}}_2 \ \cdots \ \bar{\mathbf{y}}_m]$ 的协方差矩阵可由式 (3) 计算得到,

$$C = \frac{1}{m-1} \bar{Y} \bar{Y}^T \quad (3)$$

根据式 (4) 可计算 C 的特征值 λ 和特征向量 \mathbf{p} ,

$$C\mathbf{p} = \lambda\mathbf{p} \quad (4)$$

记 C 按降序排列的特征值与相应的特征向量依次为 $\lambda_1, \lambda_2, \dots, \lambda_{n \times (k+1)}$ 和 $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{n \times (k+1)}$, 依据累计方差贡献率 (Cumulative percent variance, CPV) 可将特征向量分为主元负载矩阵 $\bar{P} = [\mathbf{p}_1 \ \mathbf{p}_2 \ \cdots \ \mathbf{p}_r]$ 和残差负载矩阵 $\tilde{P} = [\mathbf{p}_{r+1} \ \mathbf{p}_{r+2} \ \cdots \ \mathbf{p}_{n \times (k+1)}]$, 其中, r 为被保留的主元数. 于是, 可通过式 (5) 和 (6) 计算样本 $\bar{\mathbf{y}}_i$ 的主元得分和残差得分,

$$\bar{t}_i = \bar{P}^T \bar{\mathbf{y}}_i \quad (5)$$

$$\tilde{t}_i = \tilde{P}^T \bar{\mathbf{y}}_i \quad (6)$$

同时, $\bar{\mathbf{y}}_i$ 可以被分解为如下形式:

$$\bar{\mathbf{y}}_i = \bar{P} \bar{P}^T \bar{\mathbf{y}}_i + \mathbf{e}_i \quad (7)$$

其中, \mathbf{e}_i 称为 $\bar{\mathbf{y}}_i$ 的残差向量. 通常应用 T^2 和 SPE 完成过程监控, 如式 (8) 和 (9) 所示.

$$T^2 = \bar{\mathbf{t}}^T \Lambda^{-1} \bar{\mathbf{t}} = \bar{\mathbf{y}}^T \bar{P} \Lambda^{-1} \bar{P}^T \bar{\mathbf{y}} \quad (8)$$

$$SPE = \mathbf{e}^T \mathbf{e} \quad (9)$$

1.2 互异度指标

假设 $X_1 = [\mathbf{x}_1^{(1)} \ \mathbf{x}_2^{(1)} \ \cdots \ \mathbf{x}_{N_1}^{(1)}]$ 和 $X_2 = [\mathbf{x}_1^{(2)} \ \mathbf{x}_2^{(2)} \ \cdots \ \mathbf{x}_{N_2}^{(2)}]$ 为已被中心化的两个数据集, 其中 $\mathbf{x}_j^{(i)} \in \mathbf{R}^n$. X_i ($i = 1, 2$) 的协方差矩阵可表示为 $R_i = \frac{1}{N_i - 1} X_i X_i^T$.

记 $X = [X_1, X_2]$, 于是 X 的协方差矩阵可表示为:

$$R = \frac{N_1 - 1}{N - 1} R_1 + \frac{N_2 - 1}{N - 1} R_2 \quad (10)$$

其中, $N = N_1 + N_2$. 将 R 通过式 (11) 进行特征分解:

$$RP_0 = P_0 \Xi \quad (11)$$

其中, Ξ 为 R 的特征值构成的对角矩阵, P_0 为相应的特征向量矩阵. 构造变换矩阵 P_1 :

$$P_1 = P_0 \Xi^{-\frac{1}{2}} \quad (12)$$

同时, P_1 与 R 满足:

$$P_1^T R P_1 = I \quad (13)$$

将数据集 X_i 做如下变换:

$$\Psi_i = \sqrt{\frac{N_i - 1}{N - 1}} X_i^T P_1 \quad (14)$$

则矩阵 Ψ_i 的协方差矩阵可表示为:

$$S_i = \frac{N_i - 1}{N - 1} P_1^T R_i P_1 \quad (15)$$

通过式 (10)、(12) 和 (15) 可得:

$$I = S_1 + S_2 \quad (16)$$

记 S_i 的特征值与特征向量分别为 $\lambda_j^{(i)}$ 和 $\mathbf{v}_j^{(i)} \in \mathbf{R}^n$, 满足如下关系:

$$S_i \cdot \mathbf{v}_j^{(i)} = \lambda_j^{(i)} \cdot \mathbf{v}_j^{(i)} \quad (17)$$

由式 (16) 和 (17) 可得:

$$S_2 \cdot \mathbf{v}_j^{(1)} = (1 - \lambda_j^{(1)}) \cdot \mathbf{v}_j^{(1)} \quad (18)$$

式 (17) 和 (18) 说明, S_2 与 S_1 具有相同的特征向量, 而其特征值满足如下关系:

$$\lambda_j^{(2)} = 1 - \lambda_j^{(1)} \quad (19)$$

综上, 如果 X_1 和 X_2 相似, $\lambda_j^{(i)}$ ($i = 1, 2$) 的值近似相等, 即 $\lambda_j^{(i)}$ ($i = 1, 2$) 的值在 0.5 附近波动. 在故障检测过程中, 可以建立如下统计量完成过程监控.

$$D = \frac{1}{n} \sum_{i=1}^n (\lambda_i^{(1)} - 0.5)^2 \quad (20)$$

2 基于 DPCA 残差互异度的故障检测与诊断方法

在 DPCA 方法中, 通常采用 SPE 统计量对过程数据残差得分进行监控. 该统计量可以进一步整理为如下形式:

$$SPE = \mathbf{e}^T \mathbf{e} = \tilde{\mathbf{t}}^T \tilde{\mathbf{t}} \quad (21)$$

可以看出, SPE 统计量本质上在监控样本的残

差得分到坐标原点的距离. 为了保证 SPE 统计量的有效性, 进一步要求过程数据残差得分应具有较低的自相关性. 事实上, 由 DPCA 所捕获的残差得分通常存在较强的自相关性. 接下来, 通过一个简单的数值例子进行分析验证上述结论, 主要模型如式 (22) 所示:

$$z(i) = 0.8z(i-1) + u(i-1) \quad (22)$$

其中, $u(i) = 0.7u(i-1) + w(i-1)$, $u(0) = 0$, w 是均值为 0、方差为 1 的高斯白噪声. 由式 (22) 生成 1000 个样本, 在时滞参数 $\text{lag}=1$ 的前提下应用 DPCA 方法对上述数据集 $X = [\mathbf{u} \ \mathbf{z}]$ 进行分析. 记 $\bar{Y} = [\bar{u}(i) \ \bar{z}(i) \ \bar{u}(i-1) \ \bar{z}(i-1)]$ 为 X 被标准化的增广矩阵, \bar{Y} 的协方差矩阵特征值如图 1 所示. 由图 1 可以看出, 最后一个特征值较小且接近于 0, 这说明通过分析 \bar{Y} 使得过程的动态特征被捕获. 在本例中, 由于 3 个主元可以包含原始数据 99.7% 的变化信息, 因此被保留的主元数设置为 3.

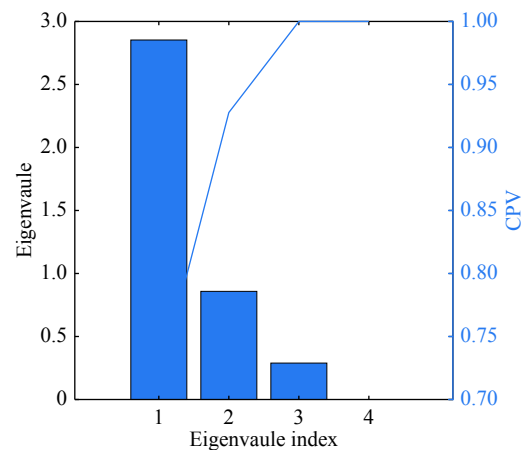


图 1 主元累计方差贡献率

Fig. 1 Cumulative percent variance of principal component

接下来, 通过式 (23) 分析本例中残差得分的自相关性.

$$r_k = \frac{c_k}{c_0} \quad (23)$$

其中, $c_k = \frac{1}{N} \sum_{j=1}^{N-k} (\tilde{t}(j) - \bar{\tilde{t}})(\tilde{t}(j+k) - \bar{\tilde{t}})$, $k = 1, 2, \dots, K$. 图 2 给出了残差得分的自相关性统计图, 可以看出, 本例中的残差得分仍然具有较强的自相关性. 文献 [21] 指出统计指标的自相关性也是影响故障检测结果的一个重要因素. 对比式 (21) 和 (23) 可以看出, 式 (21) 中的 SPE 统计量只考虑到了数据的空间分布情况, 而没有考虑数据的时间分布情况. 式 (23) 实质上在衡量相邻时刻数据间的自相关

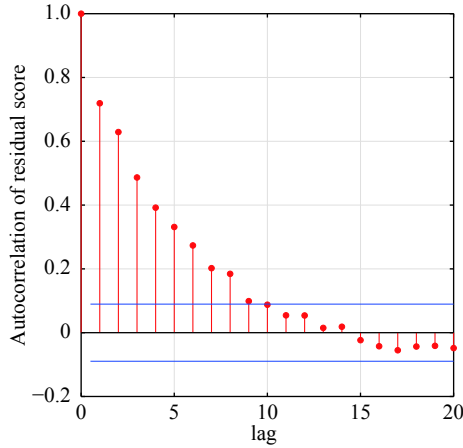


图 2 DPCA 残差得分自相关性

Fig.2 Autocorrelation of residual score in DPCA

性. 若数据相邻时刻间的相关性较弱, 说明数据的空间信息对故障检测起主要作用, 则 SPE 统计量可以有效地检测出过程故障. 若数据间存在较强的时刻相关性, 说明时间信息对故障检测也起着至关重要的作用, 则 SPE 统计量的故障检测能力有所下降. 基于上述分析及式 (21) 可知, 残差得分的自相关性会影响 SPE 统计量的故障检测性能.

本文结合 DPCA 方法能够有效捕获过程动态特征和互异度指标能有效监控具有自相关性数据的优势, 提出基于 DPCA 残差互异度的故障检测与诊断方法 (Fault detection and diagnosis based on residual dissimilarity in dynamic principal component analysis, DPCA-Diss). DPCA-Diss 故障诊断过程主要由以下三部分构成: 离线建模、在线检测和故障诊断.

2.1 离线建模

- 1) 采集两组正常情况下时间序列数据集 X_1 和 X_2 ;
- 2) 构造 X_i ($i = 1, 2$) 的增广矩阵 Y_i ($i = 1, 2$);
- 3) 应用 Y_1 的均值和方差对 Y_i 进行标准化处理, 记为 \bar{Y}_i ($i = 1, 2$);
- 4) 计算 \bar{Y}_1 的协方差矩阵的特征值与特征向量并确定残差负载矩阵 \tilde{P} ;
- 5) 计算残差得分矩阵 \tilde{T}_i ($i = 1, 2$);
- 6) 根据滑动步长 s 和窗宽 w , 依次在 \tilde{T}_1 和 \tilde{T}_2 上选取窗口数据 $\tilde{T}_1^{(k)}$ 和 $\tilde{T}_2^{(k)}$ 并计算 $\tilde{T}_1^{(k)}$ 和 $\tilde{T}_2^{(k)}$ 的互异度指标 $D^{(k)}$;
- 7) 确定 $D^{(k)}$ 的控制限 D^α .

值得注意的是, 本文采用经验法确定统计量的控制限. 经验法的一般性描述如下: 99% 置信上限确定为位于 99% 检验样本所在位置的统计值, 即在

该统计值以下可以找到 99% 的校验样本. 经验法确定控制限的意义在于可以为不同方法提供公平的比较^[22].

2.2 在线检测

- 1) 采集第 k 窗口数据并计算其增广矩阵 $\Theta^{(k)}$;
- 2) 应用第 2.1 节中的均值方差对 $\Theta^{(k)}$ 进行标准化处理, 记为 $\bar{\Theta}^{(k)}$;
- 3) 应用第 2.1 节中 \tilde{P} 计算 $\bar{\Theta}^{(k)}$ 的残差得分矩阵;
- 4) 计算 $\tilde{T}_1^{(k)}$ 和 $\tilde{T}_2^{(k)}$ 的互异度指标 $D^{(k)}$;
- 5) 判断: 若 $D^{(k)} > D^\alpha$, $\Theta^{(k)}$ 为故障, 否则, $\Theta^{(k)}$ 正常.

2.3 故障诊断

当 $\Theta^{(k)}$ 被识别为过程故障时, 其相应的故障诊断方案如下.

首先根据式 (20), 确定 $D^{(k)}$ 中贡献较大的特征值 $\lambda_j^{(1)}$. 本质上, $\lambda_j^{(1)}$ 是 $\tau_j = \sqrt{\frac{w-1}{2w-1}} \Theta^{(k)} \tilde{P} P_1^{(k)} V_1^{(k)} \xi_j$ 的方差, 其中 ξ_j 为单位阵的第 j 个列向量. 若记 $\eta_j = \sqrt{\frac{w-1}{2w-1}} \tilde{P} P_1^{(k)} V_1^{(k)} \xi_j$, 则 $\tau_j = \sum_{i=1}^n \theta_i^{(k)} \eta_{ij}$, 其中 $\theta_i^{(k)}$ 为 $\Theta^{(k)}$ 的第 i 列. 于是, $\lambda_j^{(1)} = \frac{1}{w-1} \tau_j^T \tau_j = \frac{1}{w-1} \sum_{i=1}^n \eta_{ij} (\theta_i^{(k)})^T \Theta^{(k)} \eta_j$, 因此, $\theta_i^{(k)}$ 对 $(\lambda_j^{(1)} - 0.5)^2$ 的贡献可表示为

$$con_j = \frac{\left[\eta_{ij} (\theta_i^{(k)})^T \Theta^{(k)} \eta_j - (w-1) \frac{0.5}{n} \right] (\lambda_j^{(1)} - 0.5)}{w-1} \quad (24)$$

通常情况下, 具有较大贡献的 $\theta_i^{(k)}$ 被确认为过程的异常变量.

注 1. DPCA-Diss 方法结合了 DPCA 和互异度指标两种方法的优点, 既可以获取过程数据的动态信息, 又能消除数据的自相关性对故障检测的影响. 但是 DPCA-Diss 方法并不是两种方法的简单结合. DPCA 方法既可以对数据进行维数约减, 又能获取数据的动态信息, 但是 DPCA 无法有效消除数据自相关性对过程状态监控的影响. 互异度方法虽然可以解决数据自相关性的问题, 但是当工业数据具有较多变量时, 计算复杂度会增加. 同时, 基于互异度方法的故障检测策略无法直接通过控制图观测出故障产生原因. DPCA 和互异度两种方法可以优势互补, 因此, 本文结合两种算法的优势, 提出了 DPCA-Diss 方法, 并且给出了基于贡献图的故障诊断策略, 该策略解决了互异度方法无法通过控制图查找出故障原因的问题.

3 数值例子

在本节中文献 [14] 中的一个动态过程例子被用于测试本文方法的有效性. 具体模型如式 (25) 和 (26) 所示:

$$\mathbf{x}(t) = \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix} \mathbf{x}(t-1) + \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} \mathbf{u}(t-1) \quad (25)$$

$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t) \quad (26)$$

其中, $\mathbf{v} = [v_1 \ v_2]^T$ 是均值为 $[0 \ 0]^T$, 方差为 $[0.1^2 \ 0.1^2]^T$ 的高斯白噪声; \mathbf{u} 是过程相关输入且满足:

$$\mathbf{u}(t) = \begin{bmatrix} 0.811 & -0.266 \\ 0.477 & 0.415 \end{bmatrix} \mathbf{u}(t-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix} \mathbf{w}(t-1) \quad (27)$$

其中, $\mathbf{w} = [w_1 \ w_2]^T$ 是均值为 $[0 \ 0]^T$, 方差为 $[1 \ 1]^T$ 的高斯噪声, $\mathbf{u}(0) = [0 \ 0]^T$. 依文献 [22] 所述, 在本例中监控数据集由过程相关输入 \mathbf{u} 和过程输出 \mathbf{y} 构成. 由以上模型随机生成两组各包含 2000 个样本的数据集用于确定 DPCA-Diss 方法的控制限. 为了验证 DPCA-Diss 的有效性, 由该模型生成包含 2000 个样本的测试数据集 F . 在 F 中, 从第 1001 s 起 w_1 的均值由 0 变化为 2 并持续到过程结束.

在本节中, 传统的 PCA 和 DPCA 方法对本例也进行了测试. 各种方法具体参数设置如下: 1) 在 PCA 中, $PCs = 3$; 2) 在 DPCA 中, 时滞参数 $lag = 2$, $PCs = 5$; 3) 在 DPCA-Diss 中, 经交叉验证确定窗宽为 100, 滑动步长为 1. 各种方法的故障检测率 (Fault detection rate, FDR) 如表 1 所示.

表 1 各种方法故障检测率

Table 1 Fault detection rates using different methods

检测方法	FDR (%)
PCA-SPE	4.4
DPCA-SPE	16.6
DPCA-Diss	95.1

图 3 为 PCA-SPE 的故障检测结果, 可以看出该方法并不能有效识别动态过程的异常变化, 因此故障检测率仅为 4.4%. DPCA 可以捕获过程的动态特征, 因此 DPCA-SPE 的故障检测率相比 PCA-SPE 有所提高, 但是仅达到了 16.6%, 如图 4 所示. DPCA-SPE 故障检测率仍然较低的主要原因是其残差中存在较强的自相关性, 如图 5 所示. DPCA-Diss 的故障检测结果如图 6 所示, 可以看出 DPCA-

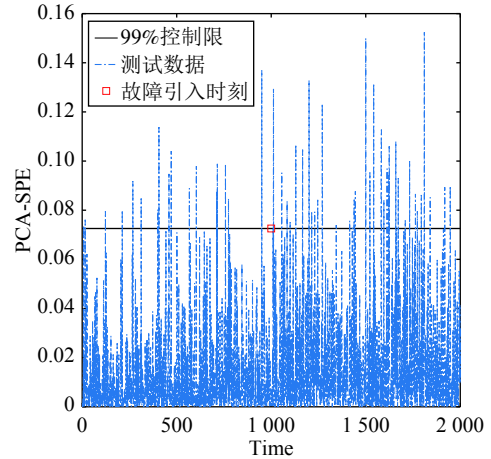


图 3 PCA-SPE 故障检测结果

Fig. 3 Fault detection results using PCA-SPE

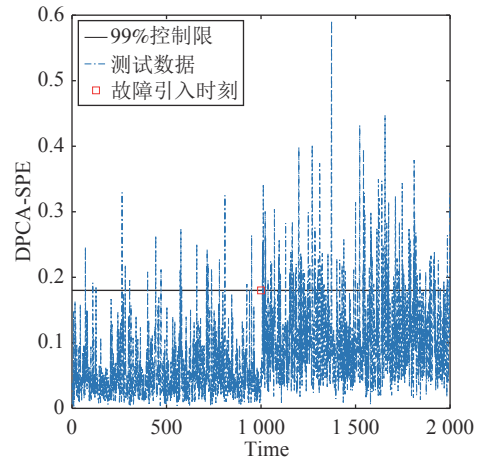


图 4 DPCA-SPE 故障检测结果

Fig. 4 Fault detection results using DPCA-SPE

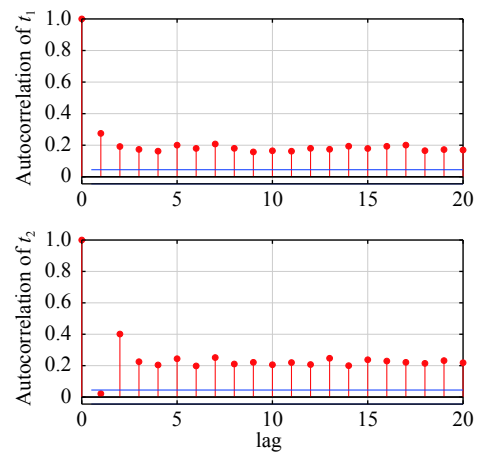


图 5 DPCA 残差得分自相关性

Fig. 5 Autocorrelation of residual score in DPCA

DPCA-Diss 方法在本例中具有最高的故障检测率, 达到

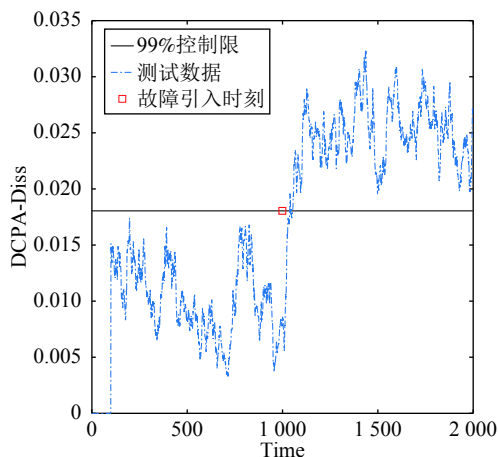


图 6 DPCA-Diss 故障检测结果

Fig.6 Fault detection results using DPCA-Diss

95.1%. 因为 Diss 可以有效提取数据在时间上的变化信息即自相关性. 以上故障检测结果说明本文方法在降低残差得分自相关性影响的同时能够提高动态过程故障检测率. 图 7 给出 DPCA-Diss 方法的故障诊断结果. 可以看出, 在第 1051 s ~ 1055 s 时间内, 变量 u_1 的贡献显著大于其他变量的贡献, 这一结果说明, 过程的异常变化主要是由于变量 u_1 发生变化引起的. 事实上, 由于 u_1 是 $w = [w_1 \ w_2]^T$ 的线性组合, 当 $w = [w_1 \ w_2]^T$ 中心发生漂移时必将引起 u_1 的异常发生. 由图 8 可以看出, 1001 s 后数据的变量中心发生了明显的偏移, 这与 DPCA-Diss 诊断结果一致.

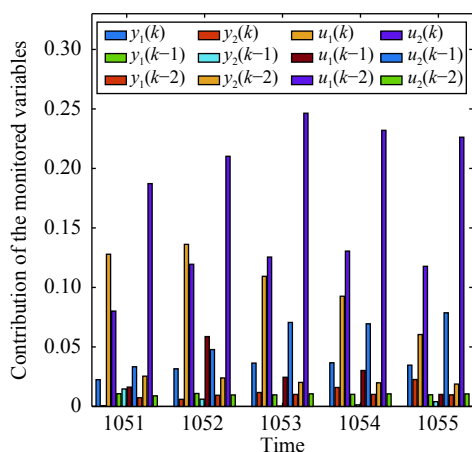
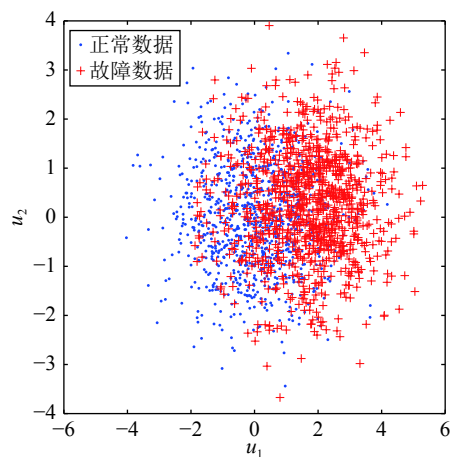


图 7 监控变量贡献图

Fig.7 Contribution charts of the monitored variables

4 Tennessee Eastman 过程

1993 年, Downs 和 Vogel 设计了 TE 过程仿真器, 截止目前, 该仿真器已经被广泛应用于过程监

图 8 输入变量 u Fig.8 Input variable u

控领域并取得了许多先进的研究成果^[23-24]. TE 过程包含反应器、冷凝器、汽液分离器、循环压缩机和产品汽提器 5 个基本操作单元, 如图 9 所示. 由该过程采集的数据包含 12 个操纵变量、22 个连续过程测量变量和 19 个成分测量变量. 本节采用已经被广泛使用的 TE 数据集进行仿真测试, 该数据集在 TE 仿真器中通过 3 分钟采样间隔并持续运行 48 小时获得^[25]. 训练集中包含 960 个正常状态下采集到的样本, 同时, 过程故障在 8 小时后引入并持续到过程结束. 如文献^[26]所述, 在本节中对上述数据集的 33 个变量进行监控.

在 DPCA-Diss 方法中, 采用包含 480 个正常样本的校验集合进行过程故障检测控制限的确定. 本例中 $\text{lag} = 1$, 55 个主元被用于建立模型, 同时, 窗宽设置为 60, 滑动步长为 1. 为了对比分析, 本节对 PCA、KCPA、DPCA 和 Diss 方法也进行了实验. 表 2 给出上述方法的故障检测率. 综合对比上述方法对于 TE 故障的检测效果, PCA 和 KPCA 方法的检测率相对较低, 其原因是这两种方法均无法获取数据的动态信息. DPCA 相较于前两种方法检测率有所提高, 主要原因是该方法可以有效捕获数据的动态信息. 但对于部分故障 DPCA 检测率仍旧较低, 原因是这些故障具有较强的自相关性. 在 TE 原始数据集上使用 Diss 方法进行故障检测时, 其对于大部分故障检测率依然较低, 其根本原因是 TE 过程数据变量较多, 数据自相关特征不显著. DPCA-Diss 方法的总体检测效果最优, 主要原因是该方法既可以提取数据的主要动态信息, 又可以成功消除数据自相关性对故障检测的影响.

接下来, 通过两个具体的故障对 DPCA-Diss 方法进行分析. 图 10 和 11 分别给出 PCA, DPCA

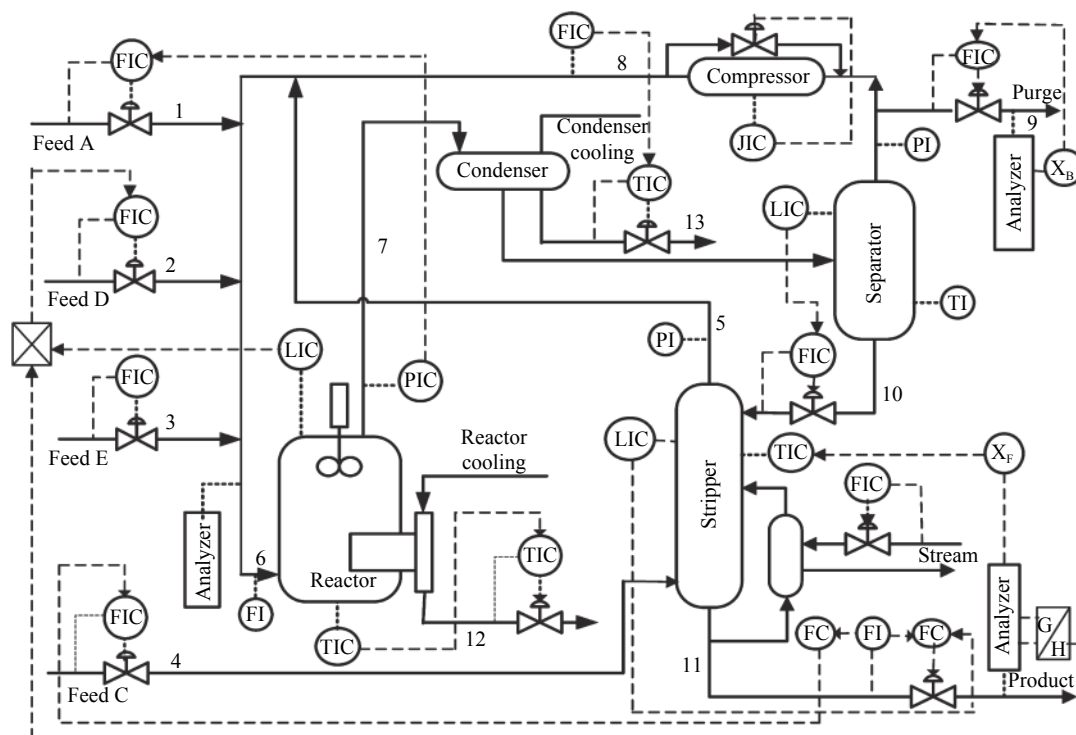


图 9 TE 过程

Fig.9 Layout of TE process

表 2 各种方法故障检测率 (%)

Table 2 Fault detection rates using different methods (%)

故障号	PCA-SPE	KPCA-SPE	DPCA-SPE	Diss	DPCA-Diss
1	99.75	93.75	98.88	34.25	96.13
2	91.88	94.50	91.75	8.88	93.63
4	99.88	53.88	4.25	9.75	21.63
5	64.25	7.50	100.00	14.13	99.38
6	100.00	63.38	100.00	96.38	99.88
7	34.25	98.63	20.00	30.25	48.75
8	79.38	40.25	65.25	70.25	96.75
10	56.50	3.63	90.13	30.88	96.63
11	67.38	53.00	7.75	76.75	91.00
12	87.00	57.75	97.75	98.88	99.75
13	94.50	62.00	93.13	65.75	93.50
14	89.13	88.13	7.75	62.00	94.38
15	3.25	3.13	2.50	2.50	77.50
16	56.63	2.88	89.63	54.25	98.38
17	94.88	74.00	71.88	88.50	96.63
18	90.38	85.75	89.50	87.00	89.00
19	51.88	6.00	46.25	73.13	97.13
20	60.13	21.63	87.38	71.00	90.00
21	37.00	5.25	16.75	22.50	33.88

和 DPCA-Diss 方法对故障 5 和 19 的检测结果. 对

于故障 5 而言, 由于 DPCA 成功获取了过程数据的动态信息, 因此本文方法与 DPCA-SPE 具有较高的故障检测率, 均达到 99% 以上. 而在故障 19 的检测故障中, 虽然 DPCA 方法获得了过程数据动态信息, 但是残差得分仍存在自相关性, 而本文方法有效解决了该问题, 因此本文方法故障检测率显著高于其他两种方法.

为了进一步确定故障产生的原因, 图 12 和 13 分别给出了故障 5 和故障 10 的诊断结果. 故障 5 是在仿真 8 小时后由冷凝器冷却水入口温度发生阶跃变化引起. 由于没有测量冷却水入口温度, 所以操纵变量冷凝器冷却水流速 (x_{33}) 是最相关的变量, 如图 14 所示. 故障 10 是由 C 进料温度发生随机变化引起的, 基于 TE 过程的机理, 这种变化必将引起汽提塔温度 (x_{18}) 的异常变化, 如图 15 所示. 故障 5 和故障 10 的诊断结果与实际一致, 结果验证了本文诊断策略的有效性.

5 结论

针对动态主元分析方法中残差自相关性降低过程故障检测率问题, 提出了基于动态主元分析残差互异度的故障检测与诊断方法. 本文方法通过滑动窗口技术与互异度指标相结合降低了过程动态特征和数据自相关性对故障检测的影响; 同时, 本文给

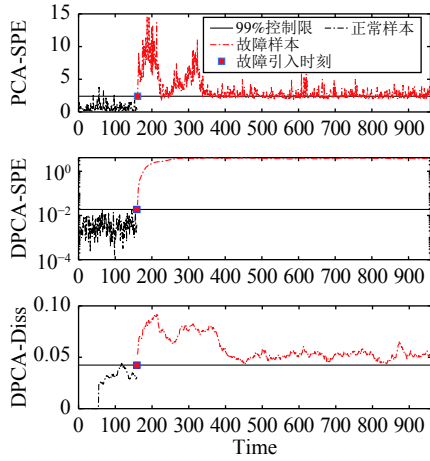


图 10 故障 5 检测结果
Fig.10 Detection results of Fault5

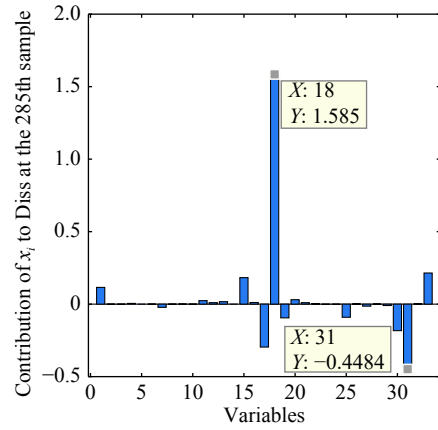


图 13 故障 10 贡献图
Fig.13 Contribution chart of Fault10

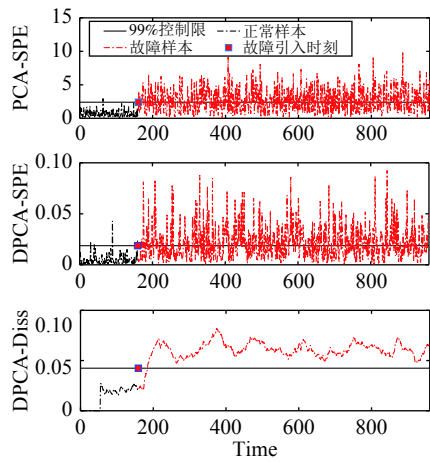


图 11 故障 19 检测结果
Fig.11 Detection results of Fault19

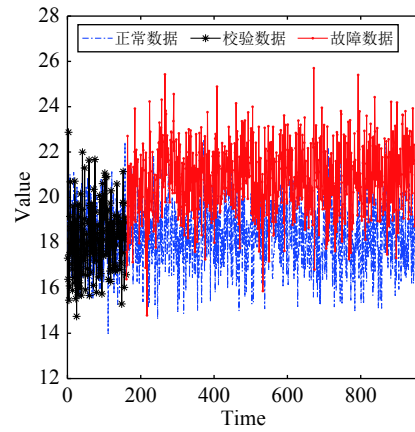


图 14 变量 33
Fig.14 Variable 33

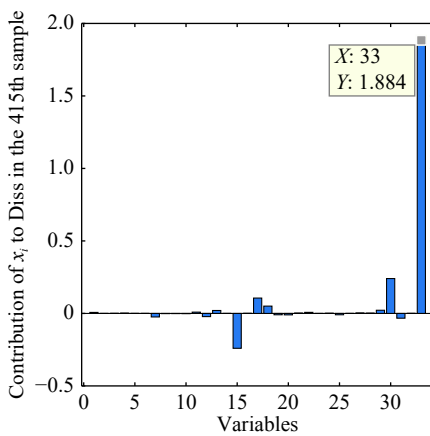


图 12 故障 5 贡献图
Fig.12 Contribution chart of Fault5

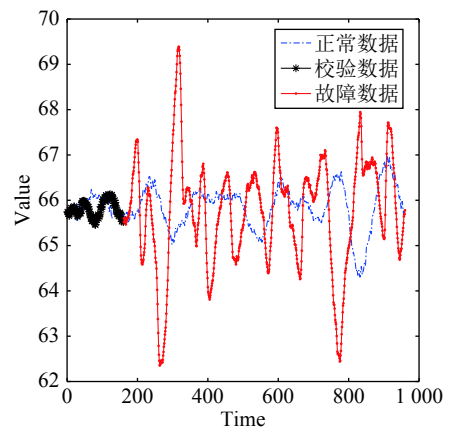


图 15 变量 18
Fig.15 Variable 18

出了基于贡献图的故障诊断策略,解决了无法通过控制图获取故障产生原因的问题。对比传统的故障方法,本文提出的方法具有更优异的故障检测性能。

本文采用的滑动窗口技术对发生于初始时刻的故障检测通常较为迟缓,因此接下来的研究方向为对初始时刻发生的故障进行及时有效地监控。

References

- 1 Qin S J. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 2012, **36**(2): 220–234
- 2 Liu Qiang, Zhuo Jie, Lang Zi-Qiang, Qin S. Joe. Perspectives on data-driven operation monitoring and self-optimization of industrial processes. *Acta Automatica Sinica*, 2018, **44**(11): 1944–1956 (刘强, 卓洁, 郎自强, 秦泗钊. 数据驱动的工业过程运行监控与自优化研究展望. 自动化学报, 2018, **44**(11): 1944–1956)
- 3 Ge Z Q, Song Z H, Gao F R. Review of recent research on data-based process monitoring. *Industrial & Engineering Chemistry Research*, 2013, **52**(10): 3543–3562
- 4 Zhang Cheng, Gao Xian-Wen, Li Yuan. Fault detection strategy based on principal component score difference of k nearest neighbors. *Acta Automatica Sinica*, 2020, **46**(10): 2229–2238 (张成, 高宪文, 李元. 基于 k 近邻主元得分差分的故障检测策略. 自动化学报, 2020, **46**(10): 2229–2238)
- 5 Feng Li-Wei, Zhang Cheng, Li Yuan, Xie Yan-Hong. DLNS-PCA-based fault detection for multimode batch process. *CIESC Journal*, 2018, **69**(07): 3159–3166 (冯立伟, 张成, 李元, 谢彦红. 基于双近邻标准化和 PCA 的多阶段过程故障检测. 化工学报, 2018, **69**(07): 3159–3166)
- 6 Hu Yi, Wang Li, Ma He-He, Shi Hong-Bo. Online nonlinear process monitoring using kernel partial least squares. *CIESC Journal*, 2011, **62**(09): 2555–2561 (胡益, 王丽, 马贺贺, 侍洪波. 基于核 PLS 方法的非线性过程在线监控. 化工学报, 2011, **62**(09): 2555–2561)
- 7 Wang L, Shi H B. Improved kernel PLS-based fault detection approach for nonlinear chemical processes. *Chinese Journal of Chemical Engineering*, 2014, **22**(06): 657–663
- 8 Zhang Cheng, Guo Qing-Xiu, Feng Li-Wei, Li Yuan. Fault detection strategy based on locality preserving projections-weighted k nearest neighbors in multimodal batch processes. *Control Engineering of China*, 2019, **36**(10): 1682–1689 (张成, 郭青秀, 冯立伟, 李元. 基于局部保持投影-加权 k 近邻规则的多模态间歇过程故障检测策略. 控制理论与应用, 2019, **36**(10): 1682–1689)
- 9 Lee J M, Yoo C K, Choi S W, Vanrolleghem P A, Lee I B. Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 2004, **59**(1): 223–234
- 10 Zhao Xiao-Li, Zhao Rong-Zhen. A method of dimension reduction of rotor faults data set based on fusion of global and local discriminant information. *Acta Automatica Sinica*, 2017, **43**(4): 560–567 (赵孝礼, 赵荣珍. 全局与局部判别信息融合的转子故障数据集降维方法研究. 自动化学报, 2017, **43**(4): 560–567)
- 11 Zhang Y W, Ma C. Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS. *Chemical Engineering Science*, 2011, **66**(1): 64–72
- 12 Zhang Cheng, Guo Qing-Xiu, Li Yuan, Gao Xian-Wen. Fault detection strategy based on difference of score reconstruction associated with principal component analysis. *Control Theory & Application*, 2019, **36**(05): 774–782 (张成, 郭青秀, 李元, 高宪文. 基于主元分析得分重构差分的故障检测策略. 控制理论与应用, 2019, **36**(05): 774–782)
- 13 Peng Kai-Xiang, Ma Liang, Zhang Kai. Review of quality-related fault detection and diagnosis techniques for complex industrial processes. *Acta Automatica Sinica*, 2017, **43**(3): 349–365 (彭开香, 马亮, 张凯. 复杂工业过程质量相关的故障检测与诊断技术综述. 自动化学报, 2017, **43**(3): 349–365)
- 14 Ku W, Storer R H, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemom. Intell. Lab. Syst.*, 1995, **30**(1): 179–196
- 15 Kano M. Statistical process monitoring based on dissimilarity of process data. *AIChE Journal*, 2002, **48**(6): 1231–1240
- 16 Wang G Z, Liu J C, Li Y. A progressive fault detection and diagnosis method based on dissimilarity of process data. In: Proceedings of the 2014 IEEE International Conference on Information and Automation. Hailar, China: IEEE, 2014. 1211–1216
- 17 Zhao C H, Wang F L, Jia M X. Dissimilarity analysis based batch process monitoring using moving windows. *AIChE Journal*, 2007, **53**(5): 1267–1277
- 18 Zhang C, Guo Q X, Li Y. Fault detection method based on principal component difference associated with DPCA. *Journal of Chemometrics*, 2018, **33**(4): e3082
- 19 Rato T J, Reis M S. Fault detection in the Tennessee Eastman benchmark process using dynamic principal components analysis based on decorrelated residuals. *Chemom Intel Lab Syst*, 2013, **125**(7): 101–108
- 20 Shi Huai-Tao, Liu Jian-Chang, Ding Xiao-Di, Tan Shuai, Wang Xue-Mei. Fault detection based on hybrid dynamic principal component analysis. *Control Engineering of China*, 2012, **19**(1): 152–155 (石怀涛, 刘建昌, 丁晓迪, 谭帅, 王雪梅. 基于混合动态主元分析的故障检测方法. 控制工程, 2012, **19**(1): 152–155)
- 21 Rato T J, Reis M S. Advantage of using decorrelated residuals in dynamic principal component analysis for monitoring large-scale systems. *Industrial & Engineering Chemistry Research*, 2013, **52**(38): 13685–13698
- 22 Wang J, He Q P. Multivariate statistical process monitoring based on statistics pattern analysis. *Industrial & Engineering Chemistry Research*, 2010, **49**(17): 7858–7869
- 23 Li G, Alcalá C F, Qin S J, Zhou D H. Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the Tennessee Eastman process. *IEEE Transactions on Control Systems Technology*, 2011, **19**(5): 1114–1127
- 24 Yu H, Khan F. Improved latent variable models for nonlinear and dynamic process monitoring. *Chemical Engineering Science*, 2017, **168**: 325–338
- 25 Jiang Q C, Yan X F. Non-Gaussian chemical process monitor-

ing with adaptively weighted independent component analysis and its applications. *Journal of Process Control*, 2013, **23**(9): 1320–1331

- 26 Zhang M G, Ge Z Q, Song Z H, Fu R W. Global-Local structure analysis model and its application for fault detection and identification. *Industrial & Engineering Chemistry Research*, 2011, **50**(11): 6837–6848



张 成 沈阳化工大学副教授, 东北大学博士研究生. 主要研究方向为复杂工业过程故障诊断.

E-mail: zhangcheng@syuct.edu.cn

(ZHANG Cheng Associate professor at Shenyang University of Chemical Technology, Ph.D. candidate at

North-eastern University. His research interest covers fault diagnosis of processes.)



戴絮年 沈阳化工大学信息工程学院硕士研究生. 主要研究方向为基于数据驱动的多工况过程故障检测.

E-mail: daixunian1996@163.com

(DAI Xu-Nian Master student at the Institute of Information Engineering, Shenyang University of

Chemical Engineering. His research interest covers data-driven based techniques for multiple operating conditions fault detection.)



李 元 沈阳化工大学教授. 2004 年获得东北大学博士学位. 主要研究方向为系统识别, 故障检测, 复杂过程故障诊断. 本文通信作者.

E-mail: li-yuan@mail.tsinghua.edu.cn

(LI Yuan Professor at Shenyang University of Chemical Technology.

She received her Ph.D. degree from Northeastern University in 2004. Her research interest covers system identification, fault detection, and complex process fault diagnosis. Corresponding author of this paper.)