

基于稀疏残差距离的多工况过程故障检测方法研究

郭小萍¹ 刘诗洋¹ 李元¹

摘要 针对多工况过程, 本文提出一种新的基于稀疏残差距离 (Sparse residual distance, SRD) 统计指标的故障检测方法. 首先对正常的多工况标准化后数据直接进行稀疏分解, 提取多个工况数据间相关关系, 得到字典和对应的稀疏编码, 以便构建全局检测模型, 避免分工况且突出数据特征. 然后计算正常多工况数据的近似值, 构建稀疏残差空间, 提出计算稀疏残差 k 近邻距离构建故障检测统计量, 利用 k 近邻捕捉过程具有的非线性、多工况特征. 最后通过数值案例和 TE (Tennessee Eastman) 生产过程进行仿真实验, 验证了所提方法的有效性.

关键词 稀疏残差距离, 稀疏分解, k 近邻距离, 多工况过程, 故障检测

引用格式 郭小萍, 刘诗洋, 李元. 基于稀疏残差距离的多工况过程故障检测方法研究. 自动化学报, 2019, 45(3): 617–625

DOI 10.16383/j.aas.c170389

Fault Detection of Multi-mode Processes Employing Sparse Residual Distance

GUO Xiao-Ping¹ LIU Shi-Yang¹ LI Yuan¹

Abstract For multi-mode processes, a new fault detection method employing sparse residual distance (SRD) is proposed in this paper. Firstly, standardized normal multi-mode process data is directly used for sparse decomposition to extract correlation between multi-mode data, and a global detection model is established using the obtained dictionary and corresponding sparse coding, so as to avoid distinguishing modes and highlight data characteristics. Then calculating the approximate value of the normal multi-mode process data to construct the sparse residual space, in which sparse residual k -nearest neighbor distance is proposed, thus the nonlinear and multi-mode features of the process can be captured further by using k -nearest neighbor. Finally, the effectiveness of the proposed method is verified by a numerical example and the Tennessee Eastman (TE) production process.

Key words Sparse residual distance, sparse decomposition, k -nearest neighbor distance, multi-mode processes, fault detection

Citation Guo Xiao-Ping, Liu Shi-Yang, Li Yuan. Fault detection of multi-mode processes employing sparse residual distance. *Acta Automatica Sinica*, 2019, 45(3): 617–625

随着计算机技术、自动化技术、网络技术的迅猛发展, 现代生产过程的自动控制水平不断提高, 使得进行过程监视与故障诊断, 提升过程的安全性与可靠性, 保证过程稳定与产品的质量成为可能. 生产过程发生故障轻则造成产品质量不合格、资源浪费等后果, 重则引起火灾、爆炸等危害人员安全和影响社会安定的恶性事件^[1]. 实际工业生产过程中, 生产

负荷、产品特性、原料组分等因素的改变会导致生产工况发生变化, 产生多工况过程. 多工况过程的数据具有非高斯、非线性、多模态、中心漂移等特征, 各工况中心不同, 其变量至少有一个是多峰分布, 因此多工况过程的故障检测技术更加复杂且对于保证过程稳定与安全具有重要意义.

针对多工况过程, 常用的两种研究方法^[2]为: 多模型法和全局模型法. 多模型法利用各工况数据建立多个局部模型, 通过多次建模将局部单工况过程扩展成为多工况过程. 该方法首要问题即为如何将数据分离成不同子集对应的多个工况. 如许仙珍等^[3]对传统主成分分析 (Principle component analysis, PCA) 不能有效解决多工况问题的缺点进行了改进, 提出了一种基于 PCA 混合模型的多工况过程故障监测方法, 该方法的优点在于能自动获取工况数目, 无需过程先验知识. 熊伟丽等^[4]针对不同稳态工况之

收稿日期 2017-07-13 录用日期 2018-02-07
Manuscript received July 13, 2017; accepted February 7, 2018
国家自然科学基金 (61490701, 61673279), 辽宁省教育厅重点实验室项目 (LZ2015059), 辽宁省自然科学基金 (201602584), 辽宁省教育厅项目 (L2016007, L2015432) 资助
Supported by National Natural Science Foundation of China (61490701, 61673279), Department of Education Key Laboratory Project of Liaoning Province (LZ2015059), Natural Science Foundation of Liaoning Province (201602584), and Foundation of Liaoning Educational Committee (L2016007, L2015432)
本文责任编辑 王伟
Recommended by Associate Editor WANG Wei
1. 沈阳化工大学信息工程学院 沈阳 110142
1. College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142

间的过渡过程提出了一种基于多工况识别的过程监测方法, 利用独立主成分分析 (Independent component analysis, ICA) 和 PCA 提取各阶段数据的信息. Ge 等^[5] 针对多工况间歇过程提出了一种新的基于贝叶斯推理的过程监测方法, 该方法采用 ICA-PCA 提取数据特征, 利用支持向量数据描述 (Support vector data description, SVDD) 方法进行故障检测, 构造超球体, 以超球体的半径作为控制限. Zhao 等^[6] 提出了一种能够同时解决多工况多阶段问题的故障监视方法. 孙贤昌等^[7] 提出了一种基于高斯混合模型 (Gaussian mixture models, GMM) 的多工况过程故障诊断方法, 该方法使用 GMM 方法对数据进行聚类分析, 得到工况数和不同工况的分布参数, 而后采用 PCA 方法建模进行故障检测. 多模型法的优点在于步骤分明、监控结果易于解释, 故障识别和诊断相对容易, 但在实际应用中存在依赖过程知识、在线识别工况、模型集优化和模型切换策略等问题.

全局模型法直接从具有多工况特性的历史数据中学习全局模型, 构建表征各工况特性的统一检测指标, 可以避免多模型法存在的问题. 郭红杰等^[8] 提出了一种基于局部近邻标准化策略的故障检测方法, 该方法运用局部近邻标准化策略对数据进行预处理, 消除数据的多工况特性. 王国柱等^[9] 提出了一种加权 KNN (K-nearest neighbor, KNN) 重构的故障诊断方法, 文中采用 KNN 规则进行故障检测, 基于 KNN 的故障检测方法是以前距离为衡量, 将数据的非高斯、非线性、多工况的特点随着建模数据直接包含在所建的模型中. Ge 等^[10] 将 SVDD 方法应用于多工况过程, 通过在特征空间构造一个最小超球体将故障数据与正常数据区分. 钟娜等^[11] 提出了一种基于局部熵成分分析 (Local entropy component analysis, LECA) 的故障检测方法, 通过 KNN-Parzen 窗方法估计变量的局部概率密度, 利用信息熵理论挖掘局部信息熵, 最后采用 ICA 方法建模进行故障检测. 针对多模态问题, Ning 等^[12] 提出了基于标签一致性字典学习算法 (Label consistent dictionary learning, LCDL) 和稀疏贡献图的过程故障诊断方法, 该方法通过标签一致性 K-SVD 方法从正常历史数据中得到字典, 加入一个单位矩阵扩展字典, 运用扩展字典对检测样本进行稀疏编码, 最后以重构误差作为统计量进行在线监控, 该方法对字典进行扩展, 增加了数据的复杂度和计算量, 而且以重构误差作为统计量的误报率偏高.

本文针对多工况过程, 提出一种在稀疏残差空间构建 k 近邻距离统计量的故障检测方法. 将多工况正常数据混合在一起标准化用于建立全局模型, 避免工况识别的工作; 采用稀疏编码方法提取分

布散、特征不够集中的多工况数据的字典和稀疏编码, 保持工况特征的同时突出数据特征; 采用多工况正常数据的近似值和原始值的偏差构建新的样本空间-稀疏残差空间, 提出在稀疏残差空间引入距离统计量作为故障检测指标, 通过稀疏残差空间数据间 k 近邻距离大小衡量数据的相似性, 兼得 k 近邻算法解决非高斯、非线性、多工况问题的优点, 改善多工况过程故障检测效果.

1 基于稀疏残差距离的故障检测方法

1.1 稀疏残差距离 (SRD) 的构建

本文提出在稀疏残差空间引入距离统计量, 简称 SRD (Sparse residual distance), 具体构建内容如下:

定义多工况正常数据矩阵 $X \in \mathbf{R}^{n \times m}$, 其中 n 是变量个数, m 是样本个数, 使用式 (1) 所示 Z -score 方法对矩阵 X 进行标准化处理, 得到均值为 0, 方差为 1 的矩阵 $X' \in \mathbf{R}^{n \times m}$;

$$x'_{ij} = \frac{x_{ij} - \text{mean}(\mathbf{x}_j)}{\text{std}(\mathbf{x}_j)} \quad (1)$$

其中, x_{ij} 是矩阵 X 第 i 行、第 j 列的数值, $\text{mean}(\mathbf{x}_j)$ 是矩阵 X 第 j 列的均值, $\text{std}(\mathbf{x}_j)$ 为矩阵 X 第 j 列的方差, x'_{ij} 为标准化后矩阵 X' 第 i 行、第 j 列的数值.

设置初始字典 $Dic_0 \in X'_{n \times m_1}$ ($m_1 < m$) 和初始训练数据 $X_{tr} \in X'_{n \times m_2}$ ($m_2 < m$), 使用正交匹配追踪 (Orthogonal matching pursuit, OMP) 算法^[13] 计算初始训练数据对应的稀疏编码 $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_j \ \cdots \ \mathbf{a}_{m_2}]$ ^[14]:

$$\begin{aligned} \min_{\mathbf{a}_j} \|\mathbf{x}_{tr}^j - Dic_0 \mathbf{a}_j\|_2^2 \\ \text{s. t. } \|\mathbf{a}_j\|_0 \leq s, \quad j = 1, 2, \dots, m_2 \end{aligned} \quad (2)$$

其中, $\mathbf{x}_{tr}^j \in \mathbf{R}^{m_2}$ 表示 X_{tr} 的第 j 个样本, $\mathbf{a}_j \in \mathbf{R}^{m_2}$ 表示与第 j 个样本对应的稀疏编码, s 表示稀疏编码的稀疏级别上限. 计算更新字典原子 \mathbf{d}_k , $k = 1, 2, \dots, m_2$ 的误差矩阵 E_k :

$$\begin{aligned} \|X_{tr} - Dic_0 A\|_F^2 &= \|X_{tr} - \sum_{j=1}^{m_2} \mathbf{d}_j \mathbf{A}_j\|_F^2 = \\ \|(X_{tr} - \sum_{j \neq k} \mathbf{d}_j \mathbf{A}_j) - \mathbf{d}_k \mathbf{A}_k\|_F^2 &= \\ \|E_k - \mathbf{d}_k \mathbf{a}_k\|_F^2 & \end{aligned} \quad (3)$$

为了确保得到稀疏编码的准确性, 只更新 \mathbf{d}_k 起作用的稀疏编码, 因此定义非零项索引集 $\omega_k = \{j | 1 \leq j \leq m_2, X_k(j) \neq 0\}$ 和矩阵 $\Omega_k \in \mathbf{R}^{m_2 \times |\omega_k|}$,

其中 Ω_k 在 $(\omega_k(j), j)$ 处为 1, 其余为 0. 重新计算 $E_k^R = E_k \Omega_k$, 对 E_k^R 进行 SVD 分解, 即 $E_k^R = U \Delta V^T$, 更新字典原子 \tilde{d}_k 为 U 的第一列, 更新相应的稀疏编码 A_k^R 为 V 的第一列乘以 $\Delta(1, 1)$. 重复上述更新计算至达到稀疏级别上限. 以上字典 Dic 和稀疏编码 A 全部更新完毕.

根据得到的字典 Dic 和稀疏编码 A 计算 X_{tr} 的近似值 \hat{X}_{tr} 及两者之间的残差矩阵 R (Residual):

$$\hat{X}_{tr} = Dic \cdot A \quad (4)$$

$$R = X_{tr} - \hat{X}_{tr} \quad (5)$$

将 R 矩阵的每一列看作一个样本, 构建新的样本空间, 计算新样本间的欧氏距离^[15].

$$D_t^2 = \sum_{i=1}^k d_{ij}^2$$

$$d_{ij} = R^i - R^j \quad (6)$$

其中, d_{ij} 表示第 i 个样本与其第 j 个样本之间的距离. 将得到的距离从小到大排列, 取前 l 个, 即为第 i 个新样本的 l 个近邻样本. 计算 l 个近邻样本距离平方和, 构建距离统计量进行故障检测.

1.2 基于 SRD 的故障检测步骤

SRD 方法分为离线建模和在线检测两部分.

1.2.1 离线建模

1) 收集多工况正常数据矩阵 X , 使用 Z-score 方法对 X 进行标准化处理, 得到矩阵 X' ;

2) 确定初始字典 D 和训练数据 X_{tr} , 根据式 (2) 获得与 D 对应的稀疏编码 A , 使用训练数据更新 D 和 A , 得到更准确表达数据特征的字典 \hat{D} 和稀疏编码 \hat{A} ;

3) 由式 (4) 计算 X' 的近似值 \tilde{X}' , 由式 (5) 得到残差, 构造残差空间 R ;

4) 将残差空间 R 的每一列看作一个样本, 根据式 (6) 计算各样本间 l 个近邻的距离平方和, 作为计算控制限的统计指标;

5) 使用核密度估计法求出置信水平为 α 的控制限 D_α^2 .

1.2.2 在线检测

1) 收集待检测数据, 使用建模数据的均值及方差进行标准化处理, 得到标准化后的数据矩阵为 X_{new} ;

2) 保持建模时的字典不变, 根据式 (2) 计算矩阵 X_{new} 的稀疏编码 A_{new} ;

3) 由式 (4) 计算近似值 \hat{X}_{new} , 由式 (5) 得近似值与待检测数据之间的残差矩阵 R' , 产生新的样本空间;

4) 将 R' 矩阵的每一列看作一个样本, 由式 (6) 计算各样本间 l 个近邻的距离平方和 D_t^2 ;

5) 比较 D_t^2 与控制限 D_α^2 的大小, 若 $D_t^2 < D_\alpha^2$, 则数据为正常数据, 反之, 则为故障数据.

具体流程图如图 1 所示.

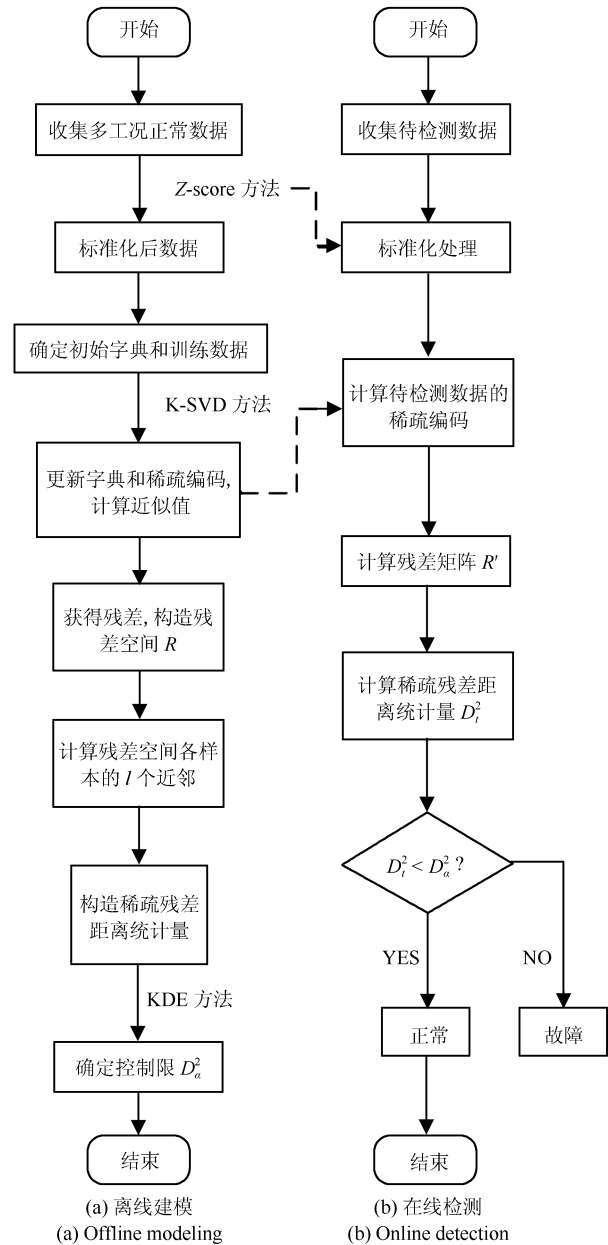


图 1 过程检测模型建立流程图
Fig.1 Flow chart of establishing process detection model

2 仿真研究

2.1 数值仿真

使用的模型如下^[16]:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0.5768 & 0.3766 \\ 0.7382 & 0.0566 \\ 0.8291 & 0.4009 \\ 0.6519 & 0.2070 \\ 0.3972 & 0.8045 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} \quad (7)$$

其中, x_1 、 x_2 、 x_3 、 x_4 、 x_5 为模型的 5 个变量, $[s_1 \ s_2]^T$ 是潜在变量, $[e_1 \ e_2 \ e_3 \ e_4 \ e_5]^T$ 是服从均值为 0, 标准差为 0.01 的高斯分布的 5 个独立的白噪声. 给出两个模态的数据代表不同工况:

$$\text{模态 1: } \begin{cases} s_1: \text{Uniform}(-10, -7) \\ s_2: \text{N}(-5, 1) \end{cases} \quad (8)$$

$$\text{模态 2: } \begin{cases} s_1: \text{Uniform}(2, 5) \\ s_2: \text{N}(7, 1) \end{cases} \quad (9)$$

所用建模数据由两种工况分别产生 400 个样本组成正常训练样本集, 因此训练数据共有 800 个样本. 随机选取训练样本中 300 个正常样本作为初始字典, 剩余的 500 个样本作为初始训练数据, 通过 OMP 算法得到的稀疏编码为 300×500 维. 在工况 1 中, $T = 401$ 时刻起在变量 x_5 上加一个幅值为 4 的阶跃信号作为故障, 即测试数据的前 400 个样本为正常样本, 剩余 400 个样本为故障样本, 共 800 个样本. 通过使用建模过程获得的字典计算测试数据的稀疏编码为 300×800 维. 本数值案例的稀疏距离统计量参数的选取采用交叉检验法, 选取适合的近邻个数为 3. 选取变量 x_1 和 x_2 两个变量方向, 分别绘制建模数据投影图、标准化处理后数据投影图和 SRD 处理后的数据投影图, 分别如图 2(a)、图 2(b) 和图 2(c) 所示.

由图 2(a) 可以看出, 建模数据分为两个工况, 工况 1 的数据分布在 30 至 80 之间, 数据中心大约为 $(-10, 50)$, 工况 2 的数据分布在 0 至 20 之间, 数据中心大约为 $(5, 10)$, 数据跨度较大, 分布较散, 中心不够明确; 由图 2(b) 可以看出, 标准化后, 数据依然分为两个工况, 此时工况 1 的中心为 $(-1, 1)$, 工况 2 的中心为 $(1, -1)$, 两个工况数据的中心呈中心对称; 在图 2(c) 中, 数据的中心与图 2(b) 一样, 与图 2(a) 相比, 数据的分布形态基本一致, 且提取数据特征后的数据跨度明显减小, 数据分布相对更集中, 数据特征得到强化, 更加明显突出, 由此说明本文方法是针对具有多工况特征过程的方法.

使用本文所提 SRD 算法的故障检测结果如图 3 所示, 由图可以看出, 前 400 个样本基本全部位于控制限以下, 误报率为 0.25%; 剩余 400 个样本全部

位于控制限上方, 故障全部被检测出来, 漏报率为 0, 说明该方法能够有效解决多工况问题.

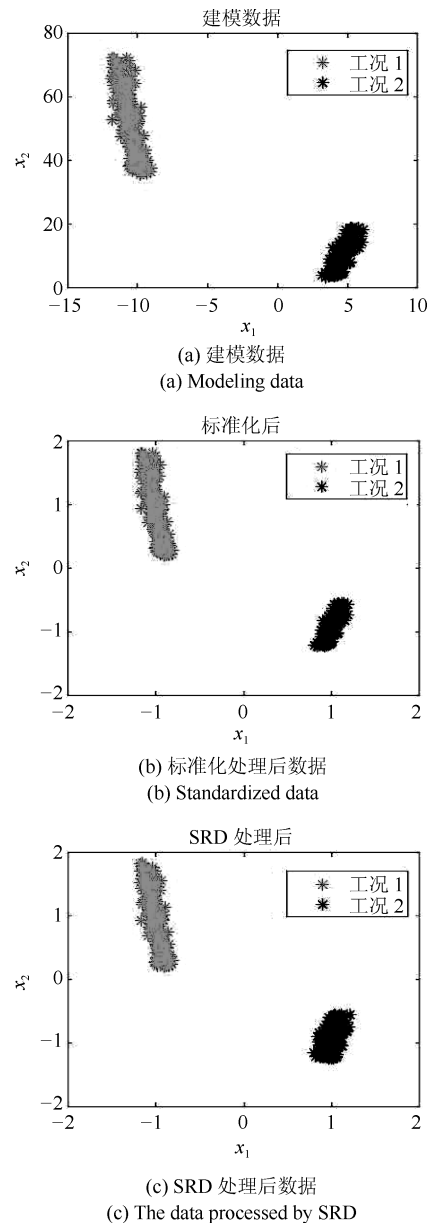


图 2 数值投影图

Fig. 2 Numerical data projection

2.2 TE 过程仿真

TE 过程是 1993 年由 Downs 和 Vogel 以伊曼曼化学公司的实际工艺流程为基础, 改进后提出的一个复杂非线性过程^[17], TE 过程适合于研究过程控制技术、过程监控、故障检测与诊断等方向. 该过程由反应器、冷凝器、压缩机、汽提塔和汽/液分离器 5 个主要的操作单元组成, 包括 8 种成分, 有 6 种生产模式, 如表 1 所示. 过程中有 41 个测量变量和 12 个控制变量, 详细的工艺流程如图 4 所示.

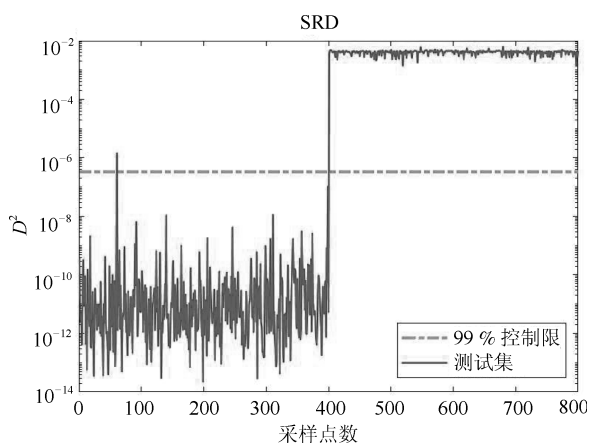


图 3 SRD 故障检测结果

Fig. 3 Fault detection results by SRD

TE 过程各种工况的采样频率均是 3 分钟/次, 每次仿真都运行 48 个小时. 仿真从正常工况开始进行, 在仿真开始 8 小时后引入故障, 共有 21 种故障可以引入, 本文验证算法有效性采用的故障如表 1 所示. 本文仿真实验采用 TE 过程生产模式 1 和 3, 如表 2 所示, 分别代表工况 1 和工况 2, 将两个工况的数据混合使用, 用于建模的两个工况的正常样本数为 1 200, 用于检测的样本数为 1 200, 检测的样本中包含正常数据 200 和故障数据 1 000. 采用的变量为 41 个测量变量和 12 个控制变量, 共 53 个变量.

在基于 SRD 的故障检测方法中, 为避免不同工况过渡过程对检测效果产生影响, 本文首先对多工

况数据进行合理选取, 通过交叉检验法选取建模数据集 ($53 \times 1\,200$). 随机选取建模数据集中 400 个样本作为初始字典, 其维数为 53×400 ; 剩余 800 个样本为训练数据, 其维数为 53×800 ; 二者通过 OMP 计算后获得的稀疏编码为 400×800 维. 对于检测的数据集, 使用建模数据集的字典 (53×400) 计算得到的稀疏编码为 $400 \times 1\,200$ 维. 为验证 SRD 方

表 1 TE 过程故障

Table 1 Failures of TE process

故障编号	性质描述	变化类型
IDV 1	物料 A/C 进料比改变, 物料 B 含量不变	阶跃
IDV 2	物料 A/C 进料比不变, 物料 B 含量改变	阶跃
IDV 4	反应器冷却入口温度改变	阶跃
IDV 6	物料 A 进料损失	阶跃
IDV 7	物料 C 压力损失	阶跃
IDV 13	反应动力学参数改变	慢偏移
IDV 16	未知	未知

表 2 本文采用的 TE 过程生产模式

Table 2 TE process production model used in this paper

生产模式	G/H 比率	产品生产率
1	50/50	7 038 kgh-1 G 和 7 038 kgh-1 H
3	90/10	1 000 kgh-1 G 和 1 111 kgh-1 H

“kgh-1 G” 表示 “每小时生产多少千克的 G 产品”, “kgh-1 H” 表示 “每小时生产多少千克的 H 产品”.

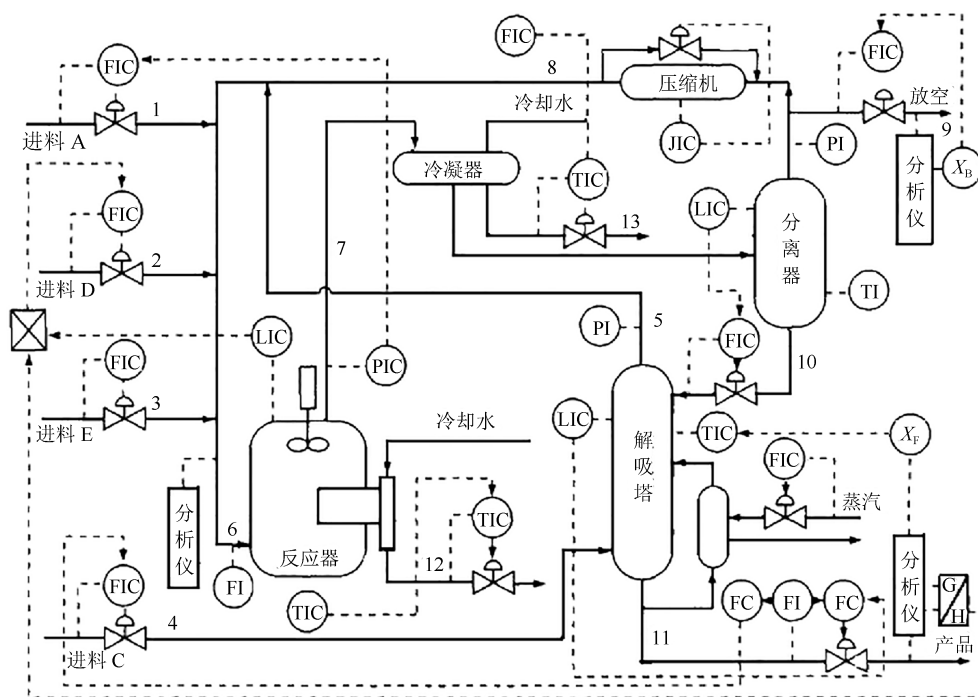


图 4 TE 过程工艺流程图

Fig. 4 TE process flow chart

法能够准确提取数据特征,任取两个变量方向分别绘制建模数据投影图、标准化处理后数据投影图和 SRD 处理后的数据投影图,分别如图 5(a)、图 5(b)和图 5(c)所示。

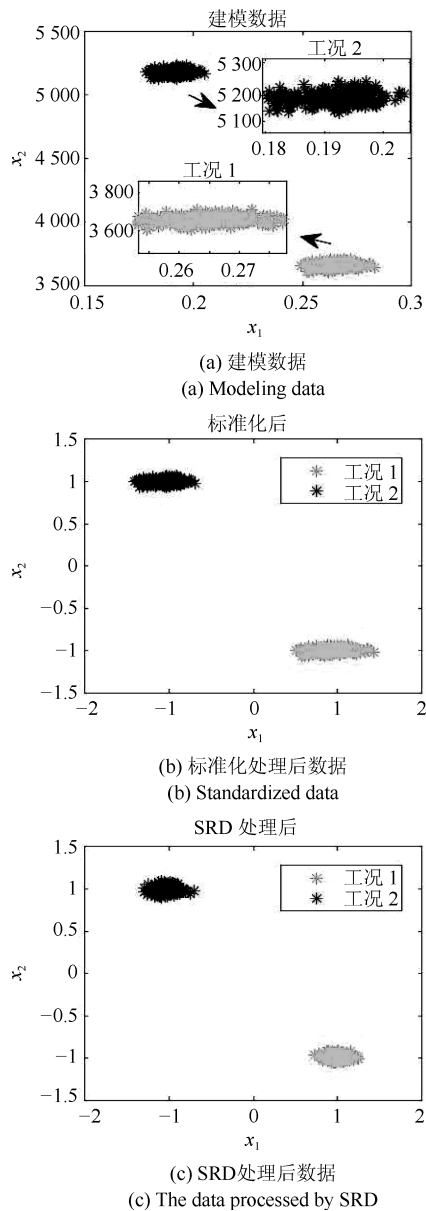


图 5 数据投影图

Fig. 5 Data projection

由图 5(a)可以看出,建模数据分为两个工况,工况 1 的数据分布在 3 600 至 3 750 之间,工况 2 的数据分布在 5 100 至 5 250 之间,数据跨度较大,分布较散,中心不够明确,特征不够突出;由图 5(b)可以看出,标准化后,数据依然分为两个工况,工况 1 的中心为 (1, -1),工况 2 的中心为 (-1, 1);在图 5(c)中,数据依然分为两个工况,中心也保持不变,工况 1 的数据分布在 -1.3 至 -0.7 之间,工况 2 的

数据分布在 0.7 至 1.3 之间,数据跨度明显减小。图 5(c)和图 5(a)中数据的分布形态基本一致,而且提取数据特征后的数据分布相对更集中,数据特征得到强化,更加明显突出,便于捕捉,在此基础上构建多工况数据的故障检测统计量会更准确,由此也说明本文方法是针对具有多工况特征过程的方法。

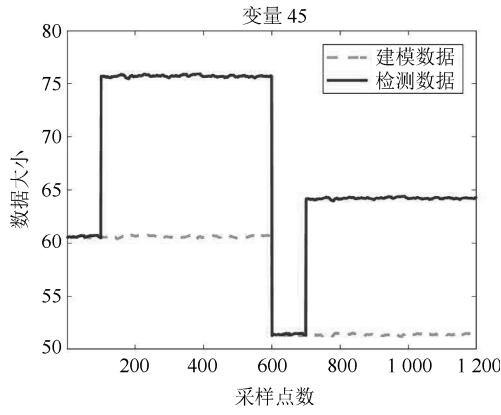
实验过程中,构建稀疏距离统计量的参数 l 的选取采用交叉检验法,通过多次试验选取合适的 l 值,即 $l=3$,故后续实验涉及到的参数 l 均选取 3。

本文验证所提 SRD 方法有效性的同时,还与以 SPE 为统计量的 KSVD-R (K-singular value decomposition residual) 方法的检测结果进行了比较。以故障 6 和故障 13 为例对故障检测结果进行说明。故障 6 和故障 13 的故障检测结果如图 6、图 7 所示,从图中可以看到数据具有多工况特征,分为两个工况,前 600 个样本为工况 1,后 600 个样本为工况 2;每一工况的前 100 个样本为正常样本,后 500 个为故障样本。

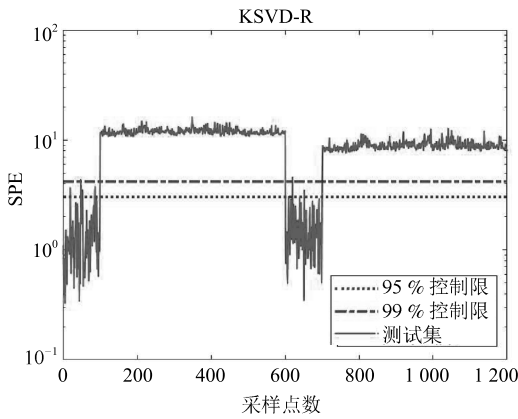
由表 1 知,故障 6 (A 进料损失)为阶跃故障,经过多次试验发现变量 45 (总进料量)体现阶跃最明显,具有明显的故障特征,视其为影响故障的主要变量,具体趋势见图 6(a)。图 6(b)为以 SPE 为统计量的 KSVD-R 故障检测方法的故障检测结果图,由图可知故障全部位于置信度为 99% 的控制限上方,正常样本基本位于该控制限下方,仅有少数几个样本位于其上方,显然该方法存在误报样本。图 6(c)为本文所提 SRD 方法的故障检测结果图,由图可知故障样本全部位于置信度为 99% 的控制限上方,正常样本全部位于该控制限下方,显然 SRD 方法误报率极低,趋近于 0。将图 6(a)和图 6(c)进行比较,可以看出,数据趋势基本一致,说明 SRD 算法可以在保证数据结构、工况特征不变的基础上进行故障检测。将图 6(b)和图 6(c)进行比较,可以看出 SRD 方法的误报样本少于 KSVD-R 方法,且 SRD 方法正常样本和故障样本数值差距稍大,说明在残差空间引入的距离统计量确实能够使故障样本和正常样本区分更加明显。

由表 1 知,故障 13 (反应动态)为慢偏移故障,经过多次试验发现变量 21 (反应器冷却水出口温度)和变量 51 (反应器冷却水流量)具有明显的的数据特征,视其为影响故障的主要变量,具体趋势见图 7(a)。图 7(b)为 KSVD-R 方法的故障检测结果图,由图可见大部分故障位于置信度为 99% 的控制限上方,正常样本基本位于该控制限下方,仅有少数几个样本位于其上方,显然该方法存在误报样本和漏报样本。

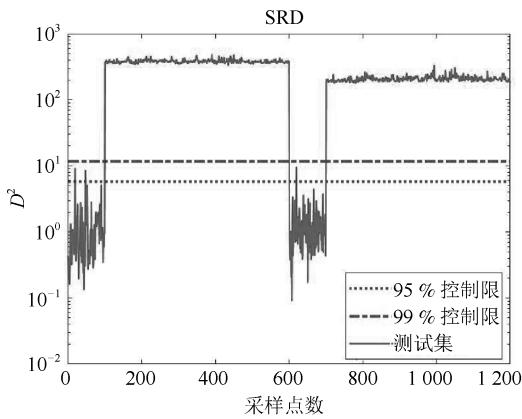
图 7(c)为 SRD 方法的故障检测结果图,由图可知绝大部分故障样本位于置信度为 99% 的控制



(a) 变量趋势图
(a) Tendency chart of variate



(b) 基于 KSVD-R 的故障检测结果
(b) Fault detection results based on KSVD-R

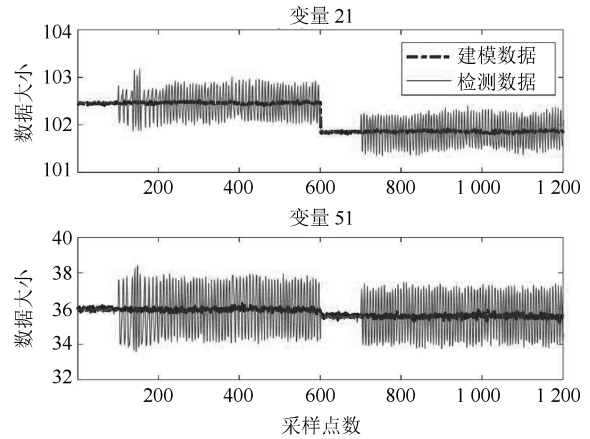


(c) 基于 SRD 的故障检测结果
(c) Fault detection results based on SRD

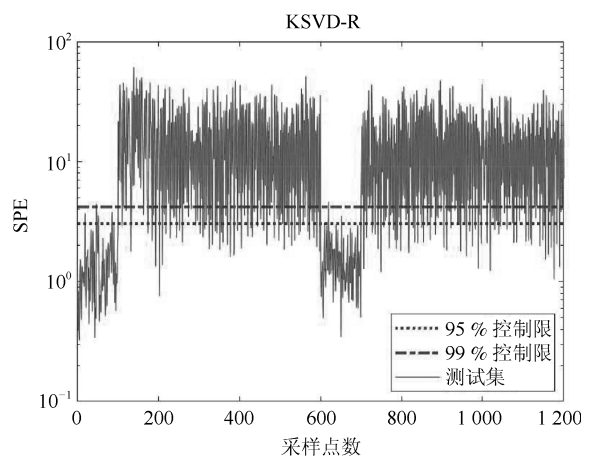
图 6 故障 6 故障检测结果

Fig.6 Fault 6 fault detection results

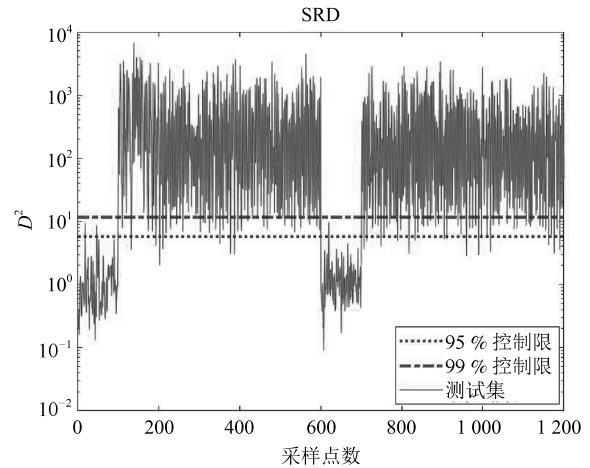
限上方, 正常样本全部位于该控制限下方, 显然 SRD 方法检测率更高, 误报率更低. 将图 7(a) 和图 7(c) 进行比较, 可以看出, 数据趋势基本一致, 再次验证 SRD 方法可以保持原始数据特征. 将图 7(b) 和图 7(c) 进行比较, 可以看出 SRD 方法的误报样本明显少于 KSVD-R 方法, 图 7(b) 中工况 1 的故障样



(a) 变量趋势图
(a) Tendency chart of variate



(b) 基于 KSVD-R 的故障检测结果
(b) Fault detection results based on KSVD-R



(c) 基于 SRD 的故障检测结果
(c) Fault detection results based on SRD

图 7 故障 13 故障检测结果

Fig.7 Fault 13 fault detection results

本和工况 2 的正常样本界限模糊, 存在交叉现象, 而图 7(c) 中正常样本和故障样本数值差距更大, 区分更加明显, 更加说明在残差空间引入的距离统计量

确实能够使故障样本和正常样本区分更加明显。

由图 6(c) 和图 7(c) 可以看出, 各工况故障检测效果均良好, 说明 SRD 方法无需对待检测样本进行工况识别仍能够检测到不同工况的故障, 由此说明 SRD 方法使用的多工况数据混合建模方式是有效的, 适用于解决多工况问题。

选取几个对比较明显的故障, 将其误报率及检测率汇总成表, 如表 3 所示。由表 3 可以看出:

1) 对于误报率, 依据置信度为 95% 的控制限, 与 KSVD-R 方法相比, SRD 方法的误报率降低了 2.5%; 依据置信度为 99% 的控制限, 则降低了 0.1%; 由此说明距离统计量对误报率的改进起一定作用;

2) 对于故障 1、故障 4、故障 6、故障 7, 无论使用哪种方法, 故障检测率均为 100%;

3) 依据置信度为 95% 的控制限, 故障 2、故障 13、故障 16 使用 SRD 方法的故障检测率比 KSVD-R 方法分别提高了 1.4%、8.1%、4.5%; 依据置信度为 99% 的控制限, 故障 13、故障 16 使用 SRD 方法的故障检测率比 KSVD-R 方法分别提高了 8.2% 和 8.4%; 进一步说明残差空间引入距离统计量确实可以改进故障检测性能。

表 3 误报率及检测率汇总表
Table 3 False alarm rate and detection rate summary table

故障号	KSVD-R		SRD		KSVD-R		SRD	
	误报率 (%)	检测率 (%)	误报率 (%)	检测率 (%)	误报率 (%)	检测率 (%)	误报率 (%)	检测率 (%)
1	4.50	0.10	100	100	2.00	0	100	100
2	4.50	0.10	100	98.6	2.00	0	100	100
4	4.50	0.10	100	100	2.00	0	100	100
6	4.50	0.10	100	100	2.00	0	100	100
7	4.50	0.10	100	100	2.00	0	100	100
13	4.50	0.10	89.6	82.1	2.00	0	97.8	90.2
16	4.50	0.10	94.6	90.2	2.00	0	98.6	94.7

注 1: 表 3 中误报率和漏报率下属两列数据, 靠前的一列为控制限为 95% 的数值, 靠后的一列为控制限为 99% 的数值

3 结束语

本文提出了一种新的基于稀疏分解残差距离的多工况过程故障检测方法。该方法将多工况数据混合建模, 提取数据特征, 依据获得的数据特征计算近似值及其与测量值的残差, 在残差空间首次引入 k 最近邻距离统计量进行故障检测。SRD 方法与 KSVD-R 方法相比, 故障检测效果有明显提高。同时该方法可以避免进行线性化、高斯化、分工况等数据预处理, 提高了多工况过程故障检测的有效性。

将 SRD 方法运用于多工况过程的故障定位、故障分类将是进一步研究的内容。

References

- Zhou Dong-Hua, Li Gang, Li Yuan. *Data-driven Based Process Fault Detection and Diagnosis Technology*. Beijing: Science Press, 2011.
(周东华, 李钢, 李元. 数据驱动的工业过程故障诊断技术. 北京: 科学出版社, 2011.)
- Lu Chun-Hong. *Research on Data-driven Fault Detection and Diagnosis Techniques and Their Applications* [Ph.D. dissertation], Jiangnan University, China, 2015
(卢春红. 基于数据驱动的故障检测与诊断技术及其应用研究 [博士学位论文], 江南大学, 中国, 2015)
- Xu Xian-Zhen, Xie Lie, Wang Shu-Qing. Multi-mode process monitoring method based on PCA mixture model. *CI-ESC Journal*, 2011, **62**(3): 743–752
(许仙珍, 谢磊, 王树青. 基于 PCA 混合模型的多工况过程监控. 化工学报, 2011, **62**(3): 743–752)
- Xiong Wei-Li, Guo Xiao-Gen. A process on-line monitoring method based on multi-mode identification. *Control and Decision*, 2018, **33**(3): 403–412
(熊伟丽, 郭校根. 一种基于多工况识别的过程在线监测方法研究. 控制与决策, 2018, **33**(3): 403–412)
- Ge Z Q, Song Z H. Bayesian inference and joint probability analysis for batch process monitoring. *AIChE Journal*, 2013, **59**(10): 3702–3713
- Zhao C H. Concurrent phase partition and between-mode statistical analysis for multimode and multiphase batch process monitoring. *AIChE Journal*, 2014, **60**(2): 559–573
- Sun Xian-Chang, Tian Xue-Min, Zhang Ni. Multi-mode process fault diagnosis method based on GMM. *Computers and Applied Chemistry*, 2014, **31**(1): 33–39
(孙贤昌, 田学民, 张妮. 一种基于 GMM 的多工况过程故障诊断方法. 计算机与应用化学, 2014, **31**(1): 33–39)
- Guo Hong-Jie, Xu Chun-Ling, Shi Hong-Bo. Multimode process monitoring based on local neighborhood standardization strategy. *Journal of Shanghai Jiao Tong University*, 2015, **49**(6): 868–875, 883
(郭红杰, 徐春玲, 侍洪波. 基于局部邻域标准化策略的多工况过程故障检测. 上海交通大学学报, 2015, **49**(6): 868–875, 883)
- Wang Guo-Zhu, Liu Jian-Chang, Li Yuan, Shang Liang-Liang. Fault diagnosis of industrial processes based on weighted k -nearest neighbor reconstruction analysis. *Control Theory & Applications*, 2015, **32**(7): 873–880
(王国柱, 刘建昌, 李元, 商亮亮. 加权 k 最近邻重构分析的工业过程故障诊断. 控制理论与应用, 2015, **32**(7): 873–880)
- Ge Z Q, Song Z H. Bagging support vector data description model for batch process monitoring. *Journal of Process Control*, 2013, **23**(8): 1090–1096
- Zhong Na, Deng Xiao-Gang, Xu Ying. Fault detection method based on LECA for multimode process. *CIESC Journal*, 2015, **66**(12): 4929–4940
(钟娜, 邓晓刚, 徐莹. 基于 LECA 的多工况过程故障检测方法. 化工学报, 2015, **66**(12): 4929–4940)
- Ning C, Chen M Y, Zhou D H. Sparse contribution plot for fault diagnosis of multimodal chemical processes. *IFAC-PapersOnLine*, 2015, **48**(21): 619–626

- 13 Lian Qiu-Sheng, Shi Bao-Shun, Chen Shu-Zhen. Research advances on dictionary learning models, algorithms and applications. *Acta Automatica Sinica*, 2015, **41**(2): 240–260 (练秋生, 石保顺, 陈书贞. 字典学习模型、算法及其应用研究进展. *自动化学报*, 2015, **41**(2): 240–260)
- 14 Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 2007, **53**(12): 4655–4666
- 15 Chen Chuan. Fault Detection Algorithm for Batch Process Based on Nearest Neighbor Rule [Master thesis], University of Electronic Science and Technology of China, China, 2015 (陈川. 基于近邻规则的间歇过程故障检测算法研究 [硕士学位论文], 电子科技大学, 中国, 2015)
- 16 Ge Z Q, Song Z H. Multimode process monitoring based on Bayesian method. *Journal of Chemometrics*, 2009, **23**(12): 636–650
- 17 Downs J J, Vogel E F. A plant-wide industrial process control problem. *Computer & Chemical Engineering*, 1993, **17**(3): 245–255



郭小萍 沈阳化工大学信息工程学院副教授. 2006 年获得东北大学博士学位. 主要研究方向为基于数据驱动技术的复杂过程故障检测与诊断.
E-mail: gxp2001@sina.com
(**GUO Xiao-Ping** Associate professor at the Institute of information engineering, Shenyang University of Chem-

ical Engineering. She received her Ph.D. degree from Northeastern University in 2006. Her research interest covers data-driven based techniques for complex process fault detection and diagnosis.)



刘诗洋 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: liusy2018@sina.com

(**LIU Shi-Yang** Master student at the Institute of Information Engineering, Shenyang University of Chemical Engineering. Her research interest covers

data-driven based techniques for multiple operating conditions fault detection.)



李元 沈阳化工大学信息工程学院教授. 2005 年获得东北大学博士学位. 主要研究方向为基于数据驱动技术的复杂过程故障检测与诊断. 本文通信作者.

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

(**LI Yuan** Professor at the Institute of information engineering, Shenyang University of Chemical Engineering.

She received her Ph.D. degree from Northeastern University in 2005. Her research interest covers data-driven based techniques for complex process fault detection and diagnosis. Corresponding author of this paper.)