

基于 SFFS-RBPCA 的高维复杂工业过程 故障诊断方法研究

金寅峰, 翁琪航, 任少君, 司风琪

(东南大学 能源热转换及其过程测控教育部重点实验室, 江苏 南京 210096)

摘要: 基于重构的故障分离方法能够抑制残差污染影响, 有效降低误诊率, 但该类方法计算量会随系统维度和故障变量数量呈指数级上升, 难以直接应用于高维复杂工业过程的在线故障诊断。因此, 提出的主成分分析法是一种基于序列特征选择算法的重构主成分分析故障诊断方法, 该方法基于历史数据建立主成分分析监测模型, 利用综合指标对实时数据进行故障检测, 在故障分离过程中引入序列特征选择方法来定位故障变量, 并采用数学仿真算例和实际工程算例对该方法的诊断性能进行验证。结果表明: 所提方法可以在较小计算量的情况下保证高诊出率和低误诊率, 在诊断精度和诊断效率之间达到良好平衡, 能够有效处理高维系统复杂故障, 满足了在线诊断需求。

关键词: 故障诊断; 主成分分析; 复杂工业过程; 特征选择

中图分类号: TK221

文献标识码: A

DOI: 10.16146/j.cnki.rndlge.2024.02.022

[引用本文格式] 金寅峰, 翁琪航, 任少君, 等. 基于 SFFS-RBPCA 的高维复杂工业过程故障诊断方法研究[J]. 热能动力工程, 2024, 39(2): 175-183. JIN Yin-feng, WENG Qi-hang, REN Shao-jun, et al. Research on fault diagnosis method of high-dimensional complex industrial process based on SFFS-RBPCA[J]. Journal of Engineering for Thermal Energy and Power, 2024, 39(2): 175-183.

Research on Fault Diagnosis Method of High-dimensional Complex Industrial Process based on SFFS-RBPCA

JIN Yin-feng, WENG Qi-hang, REN Shao-jun, SI Feng-qi

(Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education, Southeast University,
Nanjing, China, Post Code: 210096)

Abstract: The reconstruction-based fault isolation method can suppress the influence of the smear effect and effectively reduce the false alarm rate. However, the computational cost of these methods will increase exponentially with the system dimension as well as the number of fault variables, making it challenging to directly apply in the real-time fault diagnosis of high-dimensional complex industrial process. Therefore, a new fault diagnosis method integrating the sequential floating forward selection aided reconstruction-based principal component analysis (SFFS-RBPCA) was proposed, the PCA monitoring model was established based on historical samples, and the combined index was used to detect the faults of real-time data. Then, the sequence feature selection method was introduced to locate fault variables in the fault isolation process. Furthermore, a simulation example and a practical industrial case were employed to verify the diagnostic performance of the proposed method. The results show that the proposed method can ensure a high fault detection rate and a low false alarm rate by consuming a small amount of computa-

收稿日期: 2023-03-10; 修订日期: 2023-04-12

基金项目: “十四五”国家重点研发项目(2022YFB4100702)

Fund-supported Project: “the 14th Five-Year Plan” National Key R&D Program of China (2022YFB4100702)

作者简介: 金寅峰(1998-), 男, 东南大学硕士研究生.

tion, achieving a good balance between diagnostic accuracy and diagnostic efficiency. The proposed method can effectively deal with the complex faults of high-dimensional systems and meet the online diagnosis requirements.

Key words: fault diagnosis, principal component analysis (PCA), complex industrial process, feature selection

引 言

故障诊断技术是保证工业过程安全稳定运行的重要手段。现有的故障诊断方法主要分为机理分析方法以及数据驱动方法两大类^[1]。相比于机理分析方法,数据驱动方法不需要复杂的理论推导,可以从历史数据中获取系统参数之间的关联性,在缺少系统结构参数和系统内部原理的情况下也能够建立较为准确的监测模型^[2]。主成分分析(Principal Component Analysis, PCA)是一种典型的数据驱动方法,能够有效处理高维复杂数据,已经在化工^[3-5]、发电和钢铁等领域得到了广泛的研究和应用。该方法首先通过 SPE 和 T^2 等统计量来检测系统运行状态。当统计量高于控制限值阈值时,则认为系统出现故障状态。当系统被检测为故障状态后,下一步就是快速分离出根源故障变量,从而定位出故障位置。贡献图法(Contribution Plot, CP)是最为常用的故障分离方法,该方法计算简单,但容易出现残差污染问题^[6],面对大幅度或多变量复杂故障时会造成大量的误报情况。

针对这一问题,研究者们提出了多种改进方法。Rahoma 等人^[7]提出了一种基于稀疏主成分分析的故障诊断方法,通过对重载矩阵进行稀疏处理,降低故障变量对正常变量的影响,从而达到降低残差污染的目的。Dunia 等人^[8]提出了一种重构方法(Reconstruction-based, RB),通过数学算例和工程算例验证了该方法具有较强的残差污染抑制能力。Alcala 等人^[9]用数学推导从理论上证明了重构方法的优越性,即在处理较大幅度单参数故障时,重构方法能够完全克服残差污染影响。Mnassri 等人^[10]研究了重构方法在多参数故障诊断中的适用性,认为重构方法能够在较大幅度多参数故障时也能克服残差污染的影响。在此基础上,研究者们进一步将重构方法拓展到非线性动态系统的故障诊断中去^[11-15]。

然而,重构方法需要先假设故障方向再进行数据重构,但在实际过程中无法提前获知故障方向。因此,重构方法常采用穷举搜索来选择故障变量组合,需要消耗大量计算资源,且计算量随系统维度和故障变量数量呈指数级上升。为提升计算效率, Kariwala 等人^[16]和 He 等人^[17]通过分支边界算法(Branch and Bound, BAB)对故障分离方法进行优化。然而,该方法虽然相较于穷举搜索节省了许多计算成本,但面对高维多参数系统时,其计算量依然十分庞大。Yan 等人^[18]通过采用 LASSO 方法搜索故障变量组合从而大幅缩短诊断时间。但这类方法在计算时需要进行简化处理,搜索结果存在一定误差。

本文提出一种新的基于序列特征选择算法的重构主成分分析方法(Sequential Floating Forward Selection Aided Reconstruction-based Principal Component Analysis, SFFS-RBPCA),该方法可以在保证低漏报和低误报的同时大幅降低计算量和诊断时间。采用数学仿真算例和实际工程过程算例对 SFFS-RBPCA、BAB-RBPCA、LASSO-RBPCA 3 种方法进行性能比较。结果表明,本文所提方法能够在诊断精度和诊断效率之间达到良好平衡,总体表现优于其他两种效率优化算法,并且对工业过程具有良好的适应性,能够快速准确地诊断出系统实际运行过程中发生的故障。

1 基于重构的 PCA 方法

1.1 PCA 基本算法

对 n 组 m 维数据集 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ 进行归一化处理,将数据集缩放到零均值、单位方差,从而消除变量之间量纲不同的影响。PCA 的目的是寻找最优投影矩阵 $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_l] \in \mathbb{R}^{m \times l}$, $l < m$, 使其满足如式(1)所示目标函数:

$$\operatorname{argmin}_{\mathbf{P}^T \mathbf{P} = \mathbf{I}_l} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{P} \mathbf{P}^T \mathbf{x}_i\|_2^2 \quad (1)$$

式中: $\|\cdot\|_2^2$ — L_2 范数的平方; \mathbf{I}_l — l 维的单位矩阵;

x_i —第 i 个样本; n —样本个数; \mathbf{P} —待求矩阵; \mathbf{P}^T — \mathbf{P} 的转置。由式(2)可以看出,式(1)和式(3)目标函数等价。

$$\sum_{i=1}^n \|x_i - \mathbf{P}\mathbf{P}^T x_i\|_2^2 + \sum_{i=1}^n \|\mathbf{P}\mathbf{P}^T x_i\|_2^2 = \sum_{i=1}^n x_i^T x_i$$

$$= \sum_{i=1}^n \|x_i\|_2^2 \quad (2)$$

$$\operatorname{argmax}_{\mathbf{P}^T \mathbf{P} = \mathbf{I}_l} \sum_{i=1}^n \|\mathbf{P}\mathbf{P}^T x_i\|_2^2 = \operatorname{argmax}_{\mathbf{P}^T \mathbf{P} = \mathbf{I}_l} \sum_{i=1}^n \|\mathbf{P}^T x_i\|_2^2 =$$

$$\operatorname{argmax}_{\mathbf{P}^T \mathbf{P} = \mathbf{I}_l} \operatorname{tr}(\mathbf{P}^T \sum_x \mathbf{P}) \quad (3)$$

式中: $\sum_x = \sum_{i=1}^n x_i x_i^T$ —测量样本的协方差矩阵; $\operatorname{tr}(\cdot)$ —矩阵的迹运算操作符。将 \sum_x 特征值分解,则目标函数(1)对应的最优解 $\mathbf{P}^* = [\mathbf{p}_1^*, \mathbf{p}_2^*, \dots, \mathbf{p}_l^*]$ 为 \sum_x 的前 l 个特征向量。超参数 l 可以采用重构误差方差准则 (Variance Reconstruction Error, VRE) 进行选取, VRE 准则如式(4)所示:

$$\min_l \operatorname{VRE}(l) = \min_l \sum_{i=1}^n \frac{u_i}{\operatorname{var}\{\xi_i^T \sum_x \xi_i\}} \quad (4)$$

式中: $u_i \equiv \operatorname{var}\{\xi_i^T(x - x_i)\}$ —重构误差方差; ξ_i —重构方向。当 $\operatorname{VRE}(l)$ 取到最小值时,该 l 即为最佳主元个数。

1.2 故障检测

PCA 通常采用 SPE, T^2 和综合指标 φ 3 类统计指标作为检测系统是否异常的依据^[19]。用 Q 来表示 3 种故障检测指标的统一形式,如式(5)所示:

$$Q(x) = \|\mathbf{M}^{\frac{1}{2}} x\|^2 = x^T \mathbf{M} x \quad (5)$$

式中: \mathbf{M} —二次型统计指标 Q 对应的半正定的矩阵,可以由式(6)和式(7)计算得到:

$$\mathbf{M} = \mathbf{S} \mathbf{A} \mathbf{S}^T \quad (6)$$

$$\mathbf{A} = \begin{cases} \operatorname{diag}(0 \cdots 0 \underbrace{1 \cdots 1}_l) & \text{SPE} \\ \operatorname{diag}(\lambda_1 \cdots \lambda_l 0 \cdots 0) & T^2 \\ \operatorname{diag}(\lambda_1 \Gamma_{T^2}^2 \cdots \lambda_l \Gamma_{T^2}^2 \Gamma_{\text{SPE}}^2 \cdots \Gamma_{\text{SPE}}^2) & \varphi \end{cases} \quad (7)$$

式中: \mathbf{S} — \sum_x 的特征向量构成的矩阵; $\lambda_1 \cdots \lambda_l$ — \sum_x 的前 l 个特征值, l 为主元个数; Γ_Q^2 —统计指标 Q 的相应阈值,其统一表达式如式(8)所示:

$$\Gamma_Q^2 = g^Q \chi_\alpha^2(h^Q) \quad (8)$$

式中: α —卡方分布的置信阈值;参数 g^Q 和 h^Q 可以由式(9)和式(10)所示:

$$g^Q = \frac{\operatorname{tr}[(\sum_x \mathbf{M})^2]}{\operatorname{tr}[(\sum_x \mathbf{M})]} \quad (9)$$

$$h^Q = \frac{\operatorname{tr}[(\sum_x \mathbf{M})]^2}{\operatorname{tr}[(\sum_x \mathbf{M})]^2} \quad (10)$$

当故障检测指标满足 $Q(x) > \Gamma_Q^2$, 则认为过程数据存在故障。

1.3 基于重构的故障分离方法

假设 x 为一个故障测量样本,可以将 x 分为无故障部分 x^* 和故障部分 Ξf ,如式(11)所示。

$$x = x^* + \Xi f \quad (11)$$

式中: $\Xi \in \mathbb{R}^{m \times v}$ —故障方向; m —样本的维数; v —故障变量个数; $f \in \mathbb{R}^{v \times 1}$ —故障幅值。假设故障发生在第 p 个变量 v_p 和第 q 个变量 v_q 上。沿变量组合 $\{v_p, v_q\}$ 重构后的样本 x_i 和相应的指标 Q_i^{rb} 如式(12)和式(13)所示:

$$x_i = x - \Xi_i f_i \quad (12)$$

$$Q_i^{\text{rb}} = \|\mathbf{M}^{\frac{1}{2}} x_i\|^2 = \|\mathbf{M}^{\frac{1}{2}} x - \mathbf{M}^{\frac{1}{2}} \Xi_i f_i\|^2 \quad (13)$$

式中: $\Xi_i = [\xi_p \quad \xi_q] \in \mathbb{R}^{m \times 2}$ 为沿故障的重构方向; ξ_p 和 ξ_q —第 p 个和第 q 个标准正交基; f_i —沿 Ξ_i 的重构幅值估计值, f_i 的最佳估计值通过式(14)计算得到:

$$f_i = \operatorname{argmin}_f \{Q_i^{\text{rb}}\} = (\Xi_i^T \mathbf{M} \Xi_i)^{-1} \Xi_i^T \mathbf{M} x_i \quad (14)$$

当重构后的故障检测指标 Q_i^{rb} 能满足 $Q_i^{\text{rb}} < \Gamma_Q^2$, 则可以认为此时重构方向 Ξ_i 可以代表实际故障方向。

2 基于 SFFS-RBPCA 的故障诊断方法

2.1 序列浮动前向搜索算法

在实际工业过程中故障方向通常难以直接获得,因此如何准确而高效地定位故障相关变量是重构方法的另一个难题。当预设故障方向包含了额外的变量,可以通过剔除多余的变量得到更小的重构指标^[9]。由于故障所涉及的变量通常是稀疏的^[11],为避免预设故障方向包含了非故障变量,需要对其进行约束。最佳故障方向应满足两个条件:(1) 重构指标小于控制阈值;(2) 去除任意一个故障变量后,重构指标将大于控制阈值,如式(15)所示:

$$\begin{cases} Q_g^{\text{rb}} < \Gamma_Q^2 \\ \forall v_k \in g, Q_{g - \{v_k\}}^{\text{rb}} > \Gamma_Q^2 \end{cases} \quad (15)$$

式中:下标 g —目标集,包含所有故障变量; v_k —第 k 个变量; Q_g —沿 g 对应重构方向的重构指标; $Q_{g - \{v_k\}}^{\text{rb}}$ —沿 g 去掉变量 v_k 后对应重构方向的重构

指标。

序列浮动前向选择 (Sequential Floating Forward Selection, SFFS) 是一种贪心搜索算法^[20], 通过插入和剔除来寻找最优变量集合, 考虑一个从 10 维系统中搜索 3 个异常变量的变量选择问题。同样地搜索问题, 采用穷举搜索 (Exhaustive Search, ES) 需要 $C_{10}^1 + C_{10}^2 + C_{10}^3 = 175$ 次计算, 而 SFFS-RBPCA 仅需 38 次计算, 相较于 ES 算法可以大幅减少计算量。随着系统维度和故障变量个数的增加, 采用 ES 算法进行故障变量搜索所需的计算量将会呈指数级上升, 而采用 SFFS-RBPCA 算法的计算量增长率将会处于可控范围内。

2.2 SFFS-RBPCA 算法流程

SFFS-RBPCA 的算法流程如下所示:

(1) 初始化设置故障个数 $f = 0$, 目标集 $g = \emptyset$, 候选集 $F = \{v_1, v_2, \dots, v_m\}$ 包括所有过程变量。

(2) 计算重构统计量集 $Q_f = \{Q_1, Q_2, \dots, Q_{n-f}\}$, 其中 Q_i 为沿重构方向 Ξ_i 计算的重构统计量, Ξ_i 的选择基于 g_i , 可以表示为如式(16)所示关系:

$$\begin{cases} g_i = \{v_i\} \cup g \\ F = F - \{v_i\}, i \in \{1, \dots, |F|\} \end{cases} \quad (16)$$

找到最佳重构统计量 $Q_j = \max\{Q_i\}$, 更新最优目标集 $g = g \cup \{v_j\}$ 。

(3) 令 $f=f+1$, 返回第二步, 直到最佳重构统计量 Q_j 小于相关控制阈值 Γ_Q^2 。

(4) 计算重构统计量集 $Q_f = \{Q_1, Q_2, \dots, Q_{n-f}\}$, 此时 Ξ_i 的选择基于 $g_i, g_i = g - \{x_i\}, i \in \{1, \dots, |g|\}$ 。找到最佳重构统计量 $Q_j = \min_i\{Q_i\}$, 并更新最优目标集 $g = g - \{v_j\}$ 。

(5) 当最优重构统计量 Q_j 超出对应控制阈值 Γ_Q^2 , 迭代停止, 否则令 $f=f-1$, 更新最优目标集 $g = g - \{v_j\}$, 并返回步骤(4)。

(6) 目标集 g 中的变量即为最终的故障变量。

3 仿真实验与分析

3.1 数学仿真算例

为验证本文所提方法的故障诊断效果, 采用式(17)构建仿真样本矩阵 X :

$$X = AT + e \quad (17)$$

式中: $A \in \mathbb{R}^{m \times l}$ — $-1 \sim 1$ 之间服从随机分布的核

心矩阵, m —变量个数; l —主元个数; $T = [t_1, t_2, \dots, t_l]^T \in \mathbb{R}^{l \times n}$ —内部状态矩阵, $t_i \in \mathbb{R}^{n \times 1}, i \in \{1, 2, \dots, l\}$ 满足均值为零, 标准差为特定值的高斯分布, n —样本数; $e \in \mathbb{R}^{m \times n}$ —过程中的噪声, 满足高斯分布。

通过式(17)分别构建两组样本集, 用两组训练集分别训练 PCA 模型。第 1 组训练集为 1 000 个 10 维数据样本, 测试集为 2 000 个 10 维数据样本, 第 2 组训练集为 1 000 个 20 维数据样本, 测试集为 2 000 个 20 维数据样本。采用蒙特卡洛 (Monte Carlo) 方法, 如式(18)所示, 将故障随机添加到测试集样本中, 共生成 11 组故障数据集, 如表 1 所示。

表 1 数学仿真算例对照故障数据集

Tab. 1 Comparative fault data sets of mathematical simulation examples

编号	实验组	系统维度	样本数	故障变量数	故障幅度
1	1	10	2 000	1	(2,3)
2	1	10	2 000	2	(2,3)
3	1	10	2 000	3	(2,3)
4	1	10	2 000	4	(2,3)
5	1	10	2 000	5	(2,3)
6	2	20	2 000	1	(2,3)
7	2	20	2 000	2	(2,3)
8	2	20	2 000	3	(2,3)
9	2	20	2 000	4	(2,3)
10	2	20	2 000	5	(2,3)
11	2	20	2 000	6	(2,3)

$$x_{\text{fault}} = x^* + \xi_i f \quad (18)$$

式中: x^* —模拟生成的测试集样本; ξ_i —故障方向, 按照故障参数个数的取值设定每个组合包含的故障变量个数; f —故障幅度, 设定各测试集样本的故障幅度服从(2,3)区间的均匀分布。

每个故障数据集分别采用 SFFS-RBPCA、LASSO-RBPCA 和 BAB-RBPCA 3 种方法进行故障分离。为评估各方法的故障诊断效果, 采用故障诊出率 (Fault Detection Rate, FDR)、误诊率 (False Alarm Rate, FAR)、样本平均诊断时间 (Time Consumption, TC) 和单样本平均重构次数 (Mean Count, MC) 来进行判断, 各指标计算公式如下:

$$FDR = \frac{1}{m} \sum_{i=1}^m \frac{r_i}{g_i} \times 100\% \quad (19)$$

$$FAR = \frac{1}{m} \sum_{i=1}^m \frac{e_i}{n - g_i} \times 100\% \quad (20)$$

$$TC = \frac{T_2 - T_1}{n} \quad (21)$$

$$MC = \frac{\text{count}_{\text{total}}}{n} \quad (22)$$

式中: r_i —第 i 个测试样本中被诊出的参数个数; e_i —第 i 个测试样本中被误诊的参数个数; g_i —第 i 个测试样本中设定的故障参数个数; T_1 和 T_2 —测试集故障诊断的起始和结束时间; $\text{count}_{\text{total}}$ —样本在故障诊断计算过程中进行重构计算总次数, 单次重构计算过程如式(13)和式(14)所示。

两组实验结果如表 2 和表 3 所示。

表 2 10 维仿真数据各算法故障诊断结果

Tab. 2 Fault diagnosis results of each algorithm of 10-dimensional simulation data

故障数	BAB-RBPCA				LASSO-RBPCA				SFFS-RBPCA			
	FDR/%	FAR/%	TC/ms	MC	FDR/%	FAR/%	TC/ms	MC	FDR/%	FAR/%	TC/ms	MC
1	99.95	1.68	0.48	11.90	100.00	1.56	0.04	1.14	99.95	0.54	0.12	10.44
2	99.93	1.16	0.81	22.11	100.00	3.74	0.11	2.31	99.95	0.87	0.40	21.52
3	99.77	1.29	1.41	41.36	100.00	12.51	0.16	3.93	99.83	1.16	0.68	30.97
4	99.99	0.20	2.39	77.09	98.15	31.66	0.27	6.18	99.69	0.81	1.03	43.08
5	99.98	0.30	3.97	128.13	93.47	51.39	0.35	7.94	98.79	3.97	1.60	58.30

表 3 20 维仿真数据各算法故障诊断结果

Tab. 3 Fault diagnosis results of each algorithm of 20-dimensional simulation data

故障数	BAB-RBPCA				LASSO-RBPCA				SFFS-RBPCA			
	FDR/%	FAR/%	TC/ms	MC	FDR/%	FAR/%	TC/ms	MC	FDR/%	FAR/%	TC/ms	MC
1	97.40	0.50	9.07	202.30	98.25	0.81	0.08	1.14	98.25	0.22	0.25	20.44
2	97.63	0.41	16.61	340.00	99.98	1.82	0.17	2.33	98.30	0.30	0.66	41.26
3	97.53	0.44	24.54	498.36	100.00	2.73	0.24	3.46	98.45	0.26	1.14	59.87
4	96.93	0.42	30.89	682.53	100.00	3.57	0.41	4.57	97.95	0.16	1.81	76.94
5	95.89	0.61	89.15	2 271.98	100.00	5.33	0.38	5.80	97.23	0.16	2.04	93.11
6	94.87	0.60	211.56	5 509.57	100.00	6.09	0.53	6.85	96.16	0.04	2.77	107.72

从诊断精度上看, 针对不同维度不同故障个数的测试集, BAB-RBPCA 始终保持较高诊出率和较低误诊率, 而 LASSO-RBPCA 在两组实验中表现出不同的诊断效果。在同样故障变量个数的情况下, 在第 1 组实验的误诊率明显高于第 2 组实验。主要原因在于 LASSO 在搜寻最优变量组合时将 L_0 范数简化为 L_1 范数, 导致计算结果存在误差, 随着故障变量个数与系统维度之间的比例不断增高, 计算误差将进一步加大, 导致误诊率变高。由于两组实验对应数据集的维度分别为 10 维和 20 维, 因此在同样故障变量个数的情况下, 第 1 组实验中故障变量个数所占系统维度比例是第 2 组的两倍, 使得两组实验结果存在明显差异。SFFS-RBPCA 的诊断效果与 BAB-RBPCA 相似, 具有较高的诊出率和较低的误诊率, 明显优于 LASSO-RBPCA, 如图 1(a) 和

1(b) 所示。

从诊断速率上看, BAB-RBPCA 的诊断耗时随故障变量个数和系统维度不断快速上升。对于第 1 组实验, 当故障变量个数从 1 增加到 5 后, 其计算时间和重构次数分别从 0.48 ms 和 11.9 次上升到 3.97 ms 和 128.13 次。对于第 2 组实验, 当故障变量个数从 1 增加到 6 后, 其计算时间和重构次数分别从 9.07 ms 和 202.3 次上升到 211.56 ms 和 5 509.57 次。与 BAB-RBPCA 相比, LASSO-RBPCA 和 SFFS-RBPCA 的计算时间和重构次数随故障变量个数和系统维度变化而上升但并不明显, 始终保持较快的计算速率。虽然 LASSO-RBPCA 的重构次数明显低于 SFFS-RBPCA, 如图 1(c) 和 1(d) 所示, 但计算时间并没有太大优势。其原因在于 LASSO-RBPCA 每次重构都需要先找出各维度最优变量组合, 再进

行重构指标计算, SFFS-RBPCA 则相反。因此, LASSO-RBPCA 的平均单次重构时间要高于 SFFS-RBPCA 方法, 如图 1(e) 和(f) 所示。

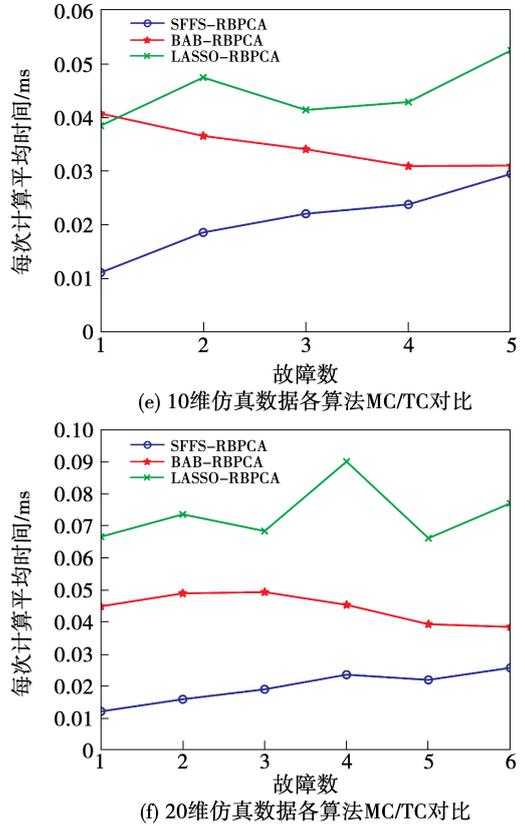
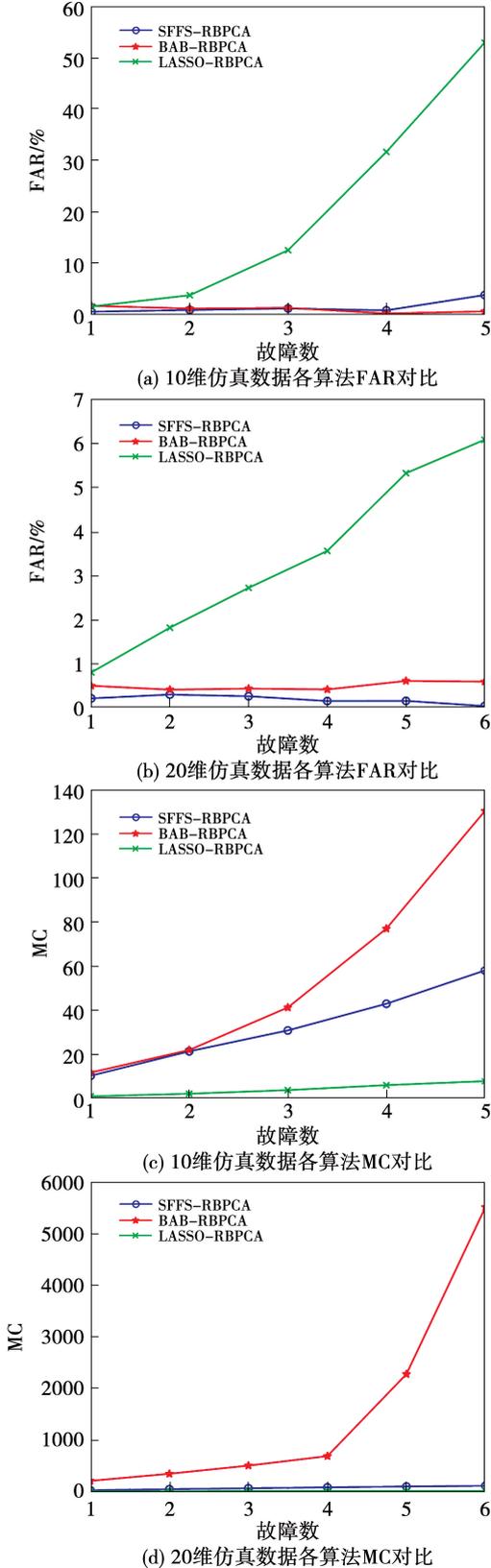


图 1 仿真数据各算法故障诊断结果
Fig. 1 Fault diagnosis results of each algorithm of simulation data

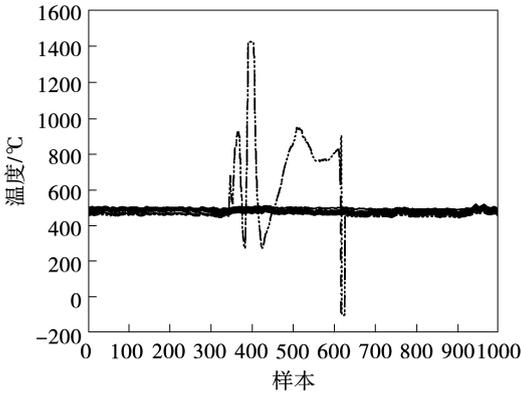
综上所述, SFFS-RBPCA 在保持较高诊断精度的同时, 诊断耗时不会随系统维度和故障数的上升而急剧增加, 计算成本总体呈可控状态, 相较于 BAB-RBPCA 和 LASSO-RBPCA 两种方法具有明显优势, 达成了诊断精度和诊断速率的良好平衡, 能够有效处理高维系统复杂故障。

3.2 实际工程算例

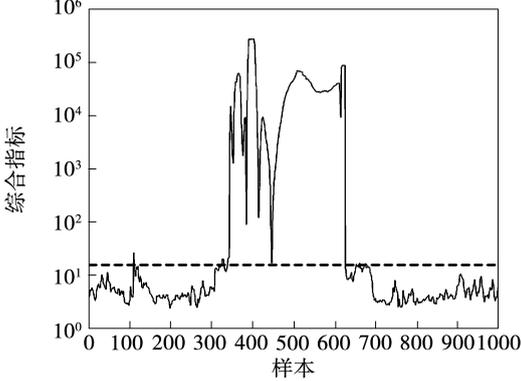
以某 1 000 MW 机组锅炉屏式过热器为研究对象, 采集 PS18Z 管组系统 32 个壁温测点共 8 595 组历史运行数据, 作为 PCA 模型训练样本集。该管组 2020 年 10 月 13 日发生了双温度传感器并发故障, 采集故障发生期间的运行数据作为测试样本集。

温度测点测量数据如图 2(a) 所示。可以看出, 第 4 号和第 17 号温度测点在第 350 至 620 样本区间出现数据异常。采用训练得到的 PCA 模型对故障样本进行故障检测, 检测结果如图 2(b) 所示。

其中,代表各样本的综合指标,虚线代表阈值。从图 2(b)可以看出,PCA 方法能够准确地识别出故障样本。



(a) 屏式过热器 PS18Z 管路温度测点信息



(b) PCA 故障检测结果

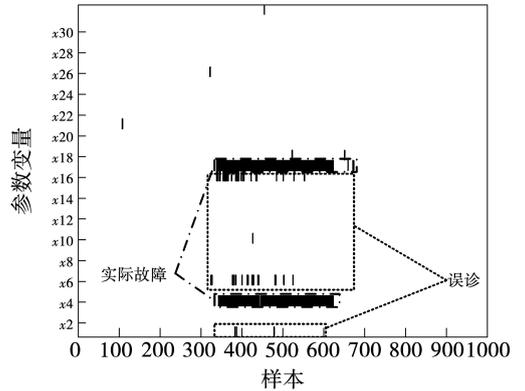
图 2 屏式过热器 PS18Z 管路温度测量数据和诊断结果

Fig. 2 Temperature measurement data and diagnosis results of PS18Z pipeline of platen superheater

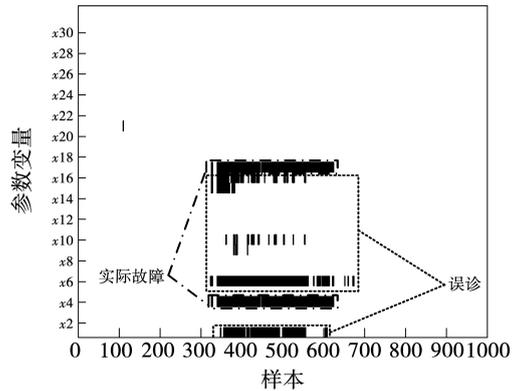
分别采用 SFFS-RBPCA、BAB-RBPCA 和 LASSO-RBPCA 进行故障分离测试。各算法实际工程过程故障诊断结果如图 3 所示。诊断结果中的黑色方块代表被分离出故障变量,其中,框内的黑色方块表示被正确识别的故障样本,虚线框内的黑色方块表示被错误诊断的正常样本。可见,3 种方法都具有较好的诊出效果,而 SFFS-RBPCA 和 BAB-RBPCA 的误诊样本数量明显少于 LASSO-RBPCA。

各方法故障诊断速率如图 4 所示。由图 4 可知,SFFS-RBPCA 的平均样本诊断时间为 382 ms,略

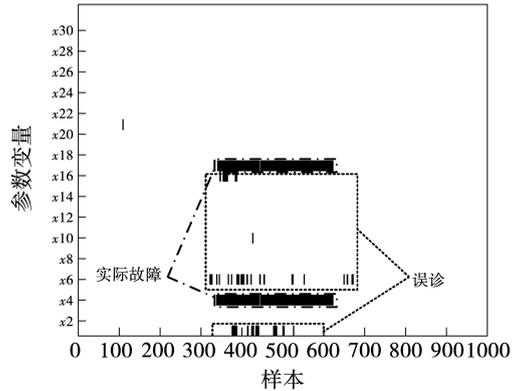
高于 LASSO-RBPCA 的 230 ms,但明显低于 BAB-RBPCA 的 3 923 ms,满足在线诊断要求。



(a) BAB-RBPCA 故障诊断结果



(b) LASSO-RBPCA 故障诊断结果



(c) SFFS-RBPCA 故障诊断结果

图 3 各算法实际工程过程故障诊断结果

Fig. 3 Fault diagnosis results of each algorithm in actual engineering process

综上,SFFS-RBPCA 相较于 BAB-RBPCA 和 LASSO-RBPCA 两种方法具有明显优势,能够同时满足低漏报、少误报和快速诊断的要求。

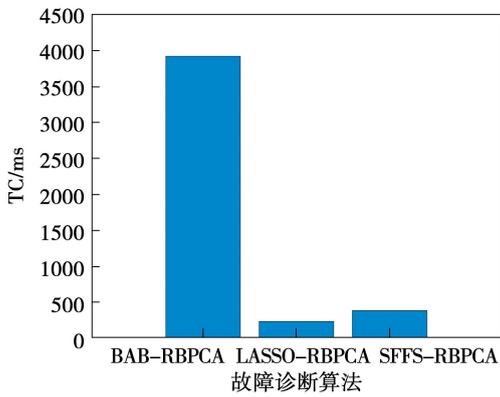


图4 各方法故障诊断速率对比

Fig.4 Comparison of fault diagnosis rates of various methods

4 结论

(1) 采用数学仿真算例对 SFFS-RBPCA 进行验证, 研究表明, SFFS-RBPCA 性能明显优于 BAB-RBPCA 和 LASSO-RBPCA 两种方法, 达成了诊断精度和诊断速率的良好平衡, 能够有效处理高维系统复杂故障。

(2) 针对 32 维工业系统的双传感器并发故障, SFFS-RBPCA 能够在低漏报和少误报的情况下实现快速诊断, 单个样本的平均诊断时间约为 382 ms, 满足了在线诊断需求。

参考文献:

[1] 叶林, 葛鸥翔, 郭永红, 等. 基于堆叠自动编码器的汽轮发电机多指标故障诊断模型[J]. 中国电机工程学报, 2022, 42(10): 3656-3670.
YE Lin, GE Ou-xiang, GUO Yong-hong, et al. Multi-index fault diagnosis model of steam turbo generator based on stacked autoencoder model [J]. Proceedings of the CSEE, 2022, 42(10): 3656-3670.

[2] WILHELM Y, REIMANN P, GAUCHEL W, et al. Overview on hybrid approaches to fault detection and diagnosis: Combining data-driven, physics-based and knowledge-based models[C]//Procedia CIRP, Amsterdam, USA; Elsevier, 2021.

[3] 林扬, 何亚东, 袁壮, 等. 基于 PCA-SVDD 的化工过程异常工况检测[J]. 过程工程学报, 2022, 22(7): 970-978.
LIN Yang, HE Ya-dong, YUAN Zhuang, et al. Abnormal condition

detection in chemical process based on PCA-SVDD[J]. The Chinese Journal of Process Engineering, 2022, 22(7): 970-978.

- [4] 王依宁, 解大, 王西田, 等. 基于 PCA-LSTM 模型的风电机组相互作用预测[J]. 中国电机工程学报, 2019, 39(14): 4070-4081.
WANG Yi-ning, XIE Da, WANG Xi-tian, et al. Prediction of interaction between grid and wind farms based on PCA-LSTM model [J]. Proceedings of the CSEE, 2019, 39(14): 4070-4081.
- [5] 武凯, 孙彦广, 张琳. 基于核主成分分析的热连轧断带故障诊断[J]. 中国冶金, 2020, 30(11): 60-65.
WU Kai, SUN Yan-guang, ZHANG Lin. Fault diagnosis of strip breaking in hot strip rolling based on kernel principal component analysis [J]. China Metallurgy, 2020, 30(11): 60-65.
- [6] WESTERHUIS J A, GURDEN S P, SMILDE A K. Generalized contribution plots in multivariate statistical process monitoring [J]. Chemometrics and Intelligent Laboratory Systems, 2000, 51(1): 95-114.
- [7] RAHOMA A, IMTIAZ S, AHMED S. A new criterion for selection of non-zero loadings for sparse principal component analysis (SPCA) [J]. Canadian Journal of Chemical Engineering, 2021, 99(1): 356-368.
- [8] DUNIA R, QIN S J. Subspace approach to multidimensional fault identification and reconstruction [J]. AIChE Journal, 1998, 44(8): 1813-1831.
- [9] ALCALA C F, QIN S J. Reconstruction-based contribution for process monitoring [J]. Automatica, 2009, 45(7): 1593-1600.
- [10] MNASSRI B, OULADSINE M. Reconstruction-based contribution approaches for improved fault diagnosis using principal component analysis [J]. Journal of Process Control, 2015, 33: 60-76.
- [11] 任少君, 肖晋飞, 司风琪, 等. 一种抑制残差污染的补偿型自联想神经网络 [J]. 东南大学学报(自然科学版), 2020, 50(4): 712-720.
REN Shao-jun, XIAO Jin-fei, SI Feng-qi, et al. A compensation auto-associative neural network for overcoming smearing effects [J]. Journal of Southeast University (Natural Science Edition), 2020, 50(4): 712-720.
- [12] ALCALA C F, QIN S J. Reconstruction-based contribution for process monitoring with kernel principal component analysis [C]// Proceedings of the 2010 American Control Conference (ACC), New York, USA; IEEE, 2010.
- [13] 仇韬, 张清峰, 丁艳军, 等. PCA 在非线性系统传感器故障检测和重构中的应用 [J]. 清华大学学报(自然科学版), 2006, 46(5): 708-711.

- QIU Tao, ZHANG Qing-feng, DING Yan-jun, et al. Nonlinear sensor fault detection and data rebuilding based on principle component analysis [J]. Journal of Tsinghua University (Science and Technology), 2006, 46(5): 708 - 711.
- [14] LI G, QIN S J, CHAI T W. Multi-directional reconstruction based contributions for root-cause diagnosis of dynamic processes [C] // 2014 American Control Conference American Control Conference (ACC), New York, USA: IEEE, 2014.
- [15] ELSHENAWY L M, MAHMOUD T A. Fault diagnosis of time-varying processes using modified reconstruction-based contributions [J]. Journal of Process Control, 2018, 70: 12 - 23.
- [16] KARIWALA V, ODIOWEI P E, CAO Y, et al. A branch and bound method for isolation of faulty variables through missing variable analysis [J]. Journal of Process Control, 2010, 20(10): 1198 - 1206.
- [17] HE B, YANG X, CHEN T, et al. Reconstruction-based multivariate contribution analysis for fault isolation: A branch and bound approach [J]. Journal of Process Control, 2012, 22(7): 1228 - 1236.
- [18] YAN Z, YAO Y. Variable selection method for fault isolation using least absolute shrinkage and selection operator (LASSO) [J]. Chemometrics and Intelligent Laboratory Systems, 2015, 146: 136 - 146.
- [19] YUE H, QIN S J. Reconstruction-based fault identification using a combined index [J]. Industrial and Engineering Chemistry Research, 2001, 40(20): 4403 - 4414.
- [20] 周 阳, 周 炎, 周 桃, 等. 基于标准序列浮动前向特征选择的改进算法研究 [J]. 计算机测量与控制, 2017, 25(7): 294 - 297.
- ZHOU Yang, ZHOU Yan, ZHOU Tao, et al. Research on improved algorithm based on the sequential floating forward selection [J]. Computer Measurement & Control, 2017, 25(7): 294 - 297.

(姜雪梅 编辑)

(上接第 108 页)

参考文献:

- [1] 刘效德, 陶乐仁, 郑志泉, 等. 垂直矩形窄通道换热特性实验研究 [J]. 制冷学报, 2014, 35(1): 14 - 19, 87.
- LIU Xiao-de, TAO Le-ren, ZHENG Zhi-gao, et al. Experimental study for heat transfer characteristics of vertical rectangular narrow channels [J]. Journal of Refrigeration, 2014, 35(1): 14 - 19, 87.
- [2] ISHIBASHI E, NISHIKAWA K. Saturated boiling heat transfer in narrow spaces [J]. International Journal of Heat and Mass Transfer, 1969, 12(8): 863 - 894.
- [3] 高立博, 郑志泉, 陶乐仁, 等. 板式升膜蒸发器蒸发换热特性的实验研究 [J]. 低温工程, 2011, 183(5): 42 - 45.
- GAO Li-bo, ZHENG Zhi-gao, TAO Le-ren, et al. Experiment study of evaporation heat transfer characteristics in plate climbing film evaporator [J]. Cryogenics, 2011, 183(5): 42 - 45.
- [4] 高晓凯, 陶乐仁, 高立博, 等. 升膜式板式换热器的换热性能研究 [J]. 低温与超导, 2013, 41(2): 71 - 74.
- GAO Xiao-kai, TAO Le-ren, GAO Li-bo, et al. Experiment study on heat transfer coefficient in climbing film plate heat exchanger [J]. Cryogenics & Superconductivity, 2013, 41(2): 71 - 74.
- [5] 张伟龙, 魏新利, 孟祥睿, 等. 热流密度及充灌率对板式蒸发器环路热管运行特性影响的实验研究 [J]. 工程热物理学报, 2020, 41(8): 2020 - 2024.
- ZHANG Wei-long, WEI Xin-li, MENG Xiang-rui, et al. Experimental study on the influence of heat flux and filling ratio on the operation characteristics of plate evaporator loop heat pipe [J]. Journal of Engineering Thermophysics, 2020, 41(8): 2020 - 2024.
- [6] SU G, GOU J, QIU S, et al. Theoretical calculation of annular upward flow in a narrow annuli with bilateral heating [J]. Nuclear Engineering and Design, 2003, 225: 219 - 247.
- [7] DU D X, TIAN W X, SU G H, et al. Theoretical study on the characteristics of critical heat flux in vertical narrow rectangular channels [J]. Applied Thermal Engineering, 2012, 36: 21 - 31.
- [8] 胡永攀, 陶乐仁, 黄嘉宇, 等. 单面加热竖直窄通道内环状流沸腾传热特性 [J]. 热能动力工程, 2019, 34(5): 105 - 111.
- HU Yong-pan, TAO Le-ren, HUANG Jia-yu, et al. Heat transfer characteristics of annular upward flow in vertical narrow channel heated from one side [J]. Journal of Engineering for Thermal Energy and Power, 2019, 34(5): 105 - 111.

(刘颖 编辑)