

DOI:10.19651/j.cnki.emt.2417427

基于迭代建模的滑窗主元分析故障检测方法^{*}

程为康 张家良

(西安工业大学电子信息工程学院 西安 710021)

摘要: 针对传统主元分析(PCA)方法在工业过程故障检测中的高虚警率和故障检测不及时的问题,本研究提出了一种基于迭代建模的滑窗主元分析故障检测方法。为了提高故障检测实时性,在建模过程中,采用迭代方法逐步剔除PCA建模数据中的异常样本,优化PCA模型。为了降低虚警率,在检测过程中,采用滑动观测窗口统计异常样本数量,通过构造第二置信限检测故障。为了提高故障检测准确性,采用一种合成统计量作为检测指标,能够同时考虑主成分方向和残差空间的异常。为了验证本研究方法的有效性,分别采用数值算例和田纳西-伊斯曼(TE)过程进行了仿真实验,其中,数值算例的故障检测准确率达到89.20%,虚警率为1.33%;TE过程的故障检测准确率达到99.39%,虚警率为3.12%。

关键词: 主元分析;故障检测;迭代建模;滑动窗口;TE过程

中图分类号: TP273;TN98 **文献标识码:** A **国家标准学科分类代码:** 510.8010

Fault detection based on iterative modeling and sliding window principal component analysis

Cheng Weikang Zhang Jialiang

(School of Electronic Information Engineering, Xi'an Technological University, Xi'an 710021, China)

Abstract: To address the high false alarm rate and delayed fault detection in traditional principal component analysis (PCA) methods for industrial process fault detection, this paper proposes an iterative modeling-based sliding window PCA fault detection method. To improve detection real-time performance, an iterative approach is used during the modeling process to progressively remove outlier samples from the PCA model data, optimizing the PCA model. To reduce the false alarm rate, a sliding observation window is employed to count the number of outlier samples, and a second confidence limit is constructed for fault detection. To enhance fault detection accuracy, a composite statistic is used as the detection index, which considers anomalies in both the principal component direction and the residual space. To validate the effectiveness of the proposed method, simulation experiments were conducted using numerical examples and the Tennessee-Eastman (TE) process. In the numerical examples, the fault detection accuracy reached 89.20% with a false alarm rate of 1.33%. For the TE process, the fault detection accuracy reached 99.39% with a false alarm rate of 3.12%.

Keywords: principal component analysis; fault detection; iterative modeling; sliding window; TE process

0 引言

随着科学技术的迅猛发展,工业系统的复杂性和现代化水平不断提高,过程监控与故障诊断在提升系统稳定性和保障其正常运行方面显得愈加重要。特别是随着工艺流程的复杂程度逐步增加,故障诊断已成为过程控制领域的研究热点之一^[1-3]。多元统计过程监控方法在过程监控领域中得到了广泛应用,其中主元分析(principal component

analysis, PCA)是应用最为普遍的一种。PCA是一种数据驱动方法,通过对过程变量进行线性变换,将原始数据空间分解为主元子空间和残差子空间,并在两个空间中构造统计量以反映空间变化。在实际应用中,PCA将观测向量投影到两个子空间中,计算相应的 Hotelling's T^2 统计量和平方预测误差(squared prediction error, SPE)用于故障检测和识别^[4-5]。

传统的主元分析方法对测试数据中的随机噪声具有高

收稿日期:2024-11-22

* 基金项目:陕西省自然科学基金基础研究计划项目(2023-JC-YB-579)资助

度敏感性,这可能导致虚警率上升。文献[6]提出了一种用于核电站传感器故障检测的鲁棒 PCA 方法,采用移动平均滤波方法减少噪声引发的虚警,提升了检测准确性和可靠性。文献[7]提出了一种基于 PCA 的广义似然比故障检测方法,通过在给定虚警率的条件下最大化故障检测概率,提供最优检测性能。文献[8]提出了一种基于数据重构的 PCA 故障诊断方法,将 SPE 统计量分解为主元相关的残差变量和一般变量残差用于故障判断并通过数据重构减少故障数据的影响。文献[9]将一种基于数据变化率的预处理方法用于 PCA 故障检测,能够有效检测系统变量中的微小故障。文献[10]提出一种基于故障敏感主元的多块 PCA 故障监测方法,通过定义故障敏感系数来优化主元选择,并利用贝叶斯信息准则融合各子模型监测结果。文献[11]提出了一种分散加权的 ReliefF-PCA 方法,用改进的 ReliefF 算法选择主成分,保留小方差信息,通过加权模型和贝叶斯信息准则整合不同故障。文献[12]提出了一种结合长短期记忆(long short term memory, LSTM)和主元分析的离心式压缩机动态监测方法,通过滑动窗口和 LSTM 预测未来数据点,有效降低误报率。文献[13]提出了一种增强的时间算法耦合自适应稀疏 PCA 方法,通过引入时间状态计算和动态稀疏性优化,提升了传统 PCA 模型的可解释性。文献[14]提出了一种结合 PCA 和支持向量机的故障诊断方法,通过三层分类模型和滑窗法实现在线诊断,提升了诊断的准确性和速度。文献[15]通过 PCA 得分重构差分并计算重构差分的统计值用于故障检测。文献[16]则结合 PCA 和 k 近邻,构建了新的统计量,有效检测非线性和多模态场景中的故障。文献[17]通过构建得分差分和残差矩阵,在各子空间中建立新统计量,有效降低了数据结构的干扰。文献[18]提出了一种基于加权自适应递归 PCA 的故障检测方法,显著减少计算复杂度,在保证原有检测延迟性能的基础上降低了低虚警率。文献[19]提出了一种结合数据预处理和虚警减少的故障检测方法,通过消除数据中的奇异点和波动,并基于统计量降低 T^2 和 SPE 统计量的虚警,提高了 PCA 模型的稳定性和可靠性。上述这些文献虽然都对降低故障虚警率或提高检测及时性问题进行了研究,但未对同时解决这两个问题进行深入探讨。文献[20]提出了一种基于统计和一种基于迭代的主元分析故障检测方法,能够降低虚警率和提高检测及时性。但该方法采用的是传统的 SPE 和 T^2 统计量,在提高检测准确性方面还可深入研究。

为了提高故障检测准确性和实时性,减少虚警率,本文提出了一种基于迭代建模的滑窗主元分析故障检测方法。该方法采用一种合成统计量,能够同时考虑主成分方向和残差空间的异常,首先利用原始数据建立 PCA 模型,确定原始置信限,再通过迭代方法逐步剔除失控样本,更新建模数据,得到优化的 PCA 模型并确定主元个数和最终置信限。然后,在检测过程中,采用一种滑动观测窗口统计异常

样本数量,构造第二置信限检测故障。最后,通过数值仿真和 TE 过程仿真实验验证了提出方法的有效性。

1 主元分析

假设一个包含了 m 个传感器的测量样本,每个传感器各有 n 个独立采样,构造为一个 n 行 m 列的数据阵 $\mathbf{X}_{n \times m}$,表示为:

$$\mathbf{X}_{n \times m} = \begin{bmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \cdots & \mathbf{x}_{1m} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \cdots & \mathbf{x}_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}_{n1} & \mathbf{x}_{n2} & \cdots & \mathbf{x}_{nm} \end{bmatrix} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_m] \quad (1)$$

对 $\mathbf{X}_{n \times m}$ 进行标准化处理:

$$\mathbf{Y}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (2)$$

式中: x_{ij} 表示数据集的元素, μ_j 表示 $\mathbf{X}_{n \times m}$ 的每一列元素的均值, σ_j 表示 $\mathbf{X}_{n \times m}$ 的每一列的标准方差。

标准化后的矩阵为 \mathbf{Y} , 计算 \mathbf{Y} 的协方差矩阵 \mathbf{S} 。

$$\mathbf{S} = \text{cov}(\mathbf{Y}) \approx \frac{1}{n-1} \mathbf{Y}^T \mathbf{Y} \quad (3)$$

对 \mathbf{S} 进行特征值分解,得到矩阵的特征值 λ_i 和特征向量,将特征值由大到小进行排列,得到 $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_n \geq 0$, 由累计方差贡献率 CPV 可以确定主元个数。

$$\text{CPV} = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^m \lambda_i} \quad (4)$$

式中: k 为主元个数。

根据 PCA 模型可得:

$$\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{E} = \mathbf{T} \mathbf{P}^T + \mathbf{E} \quad (5)$$

$$\mathbf{T} = \mathbf{Y} \mathbf{P} \quad (6)$$

式中: $\hat{\mathbf{Y}}$ 为数据在主元空间的投影,是被建模的部分, $\mathbf{P} \in \mathbf{R}^{m \times k}$ 为负载矩阵,由 \mathbf{S} 的前 k 个特征向量组成, $\mathbf{T} \in \mathbf{R}^{n \times k}$ 为得分矩阵, \mathbf{T} 的各列被称为为主元变量,也是得分矩阵的列数, \mathbf{E} 为在残差空间的投影,是未被建模的部分,主要反映非正常数据噪声变化的情况。

PCA 故障诊断方法通常用 SPE 统计量或 Hotelling's T^2 来检测过程是否发生故障异常。

SPE 统计量衡量样本在残差空间投影的变化。

$$\text{SPE} = \|(\mathbf{I} - \mathbf{P} \mathbf{P}^T) \mathbf{x}\|^2 \leq \delta_\alpha \quad (7)$$

式中: δ_α 表示 SPE 统计量在置信水平为 α 时的控制限。

控制限的计算公式为:

$$\delta_\alpha = \theta_1 \left(\frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{1/h_0} \quad (8)$$

式中: $\theta_i = \sum_{j=k+1}^m \lambda_j^i$ ($i = 1, 2, 3$), $h_0 = 1 - 2\theta_1 \theta_3 / 3\theta_2^2$, λ_j 为 \mathbf{X} 的协方差矩阵的特征值, c_α 为标准正太分布在置信水平 α

下的阈值,当 SPE 统计量位于控制限内时认为工作正常,相反之表明有故障发生。

Hotelling's T^2 统计量衡量样本在主元空间投影的变化。

$$T^2 = \mathbf{x}^T \mathbf{P} \mathbf{A}^{-1} \mathbf{P}^T \mathbf{x} \leq T_\alpha^2 \quad (9)$$

式中: $\mathbf{A} = \text{diag}\{\lambda_1, \dots, \lambda_k\}$, T_α^2 表示置信度为 α 的 T^2 控制限,控制限的计算公式为:

$$T_\alpha^2 = \frac{k(n^2 - 1)}{n(n - k)} F_{k, n-k; \alpha} \quad (10)$$

式中: $F_{k, n-k; \alpha}$ 是自由度为 $k, n - k$, 置信水平为 α 的 F 分布临界值,当 T^2 统计量位于控制限内时认为工作正常,相反之表明有故障发生。

2 迭代建模的滑窗主元分析故障检测方法

在多元统计过程监控中,常用的统计指标包括 SPE 和 T^2 统计量,但这两者各有局限性。SPE 统计量仅反映样本在残差空间的偏离,难以直接与系统物理变量相关联,故障如果仅影响主成分方向, SPE 统计量可能无法检测到异常。 T^2 统计量无法检测主成分空间外的异常(如噪声或未建模动态引起的故障),如果噪声影响主要出现在残差空间内, T^2 统计量的检测能力可能较差。

因此,本文采用一种基于 SPE 统计量和 T^2 统计量的混合指标进行故障检测。该合成统计量能够综合 SPE 和 T^2 统计量的优势,同时检测到主成分方向和残差方向的异常。

合成统计量计算公式为:

$$\varphi(\mathbf{x}) = \frac{\text{SPE}(\mathbf{x})}{\delta_\alpha} + \frac{T^2(\mathbf{x})}{T_\alpha^2} = \mathbf{x}^T \Phi \mathbf{x} \quad (11)$$

式中: $\Phi = (\mathbf{P} \mathbf{A}^{-1} \mathbf{P}^T) / T_\alpha^2 + (\mathbf{I} - \mathbf{P} \mathbf{P}^T) / \delta_\alpha$ 。

合成统计量的控制限为:

$$\zeta^2 = g \chi_h^2 \quad (12)$$

式中: $g = \text{tr}(\mathbf{S}\Phi)^2 / \text{tr}(\mathbf{S}\Phi)$, $h = [\text{tr}(\mathbf{S}\Phi)]^2 / \text{tr}(\mathbf{S}\Phi)^2$ 。

针对于在故障检测中虚警较高,并且故障检测响应不够及时的问题,提出了基于迭代建模的滑窗主元分析故障检测方法。

该方法首先利用迭代方法优化模型,通过逐步剔除训练数据中失控的样本(即合成统计量超出置信限的样本),从而获得最终的 PCA 模型。在此基础上,使用统计量方法定义一个滑动观测窗口,监测该窗口内的报警次数(即合成统计量超过置信限的次数)。当报警总数超过预设限值时,将窗口内的当前样本确认为故障,否则视为虚警。结合迭代和统计方法,能够显著减少由外部环境干扰和模型误差导致的合成统计量虚警,使 PCA 模型对传感器上发生的小故障更为敏感。基于迭代建模的滑窗主元分析方法建模流程图如图 1 所示,主元分析故障检测模型的建模步骤为:

1) 利用训练样本数据建立 PCA 模型,然后确定合成统计量的原始置信限(即 ζ^2)。

2) 利用建立的 PCA 模型对训练数据中的所有样本依次进行检验,若所有样本的合成统计量均在原置信限内后,转到步骤 4)。

3) 若某一样本合成统计量超出原始置信限,则从训练数据中删除此样本,更新训练样本数据,返回步骤 1)。

4) 建立最终的 PCA 模型,确定最终主元个数和置信限。

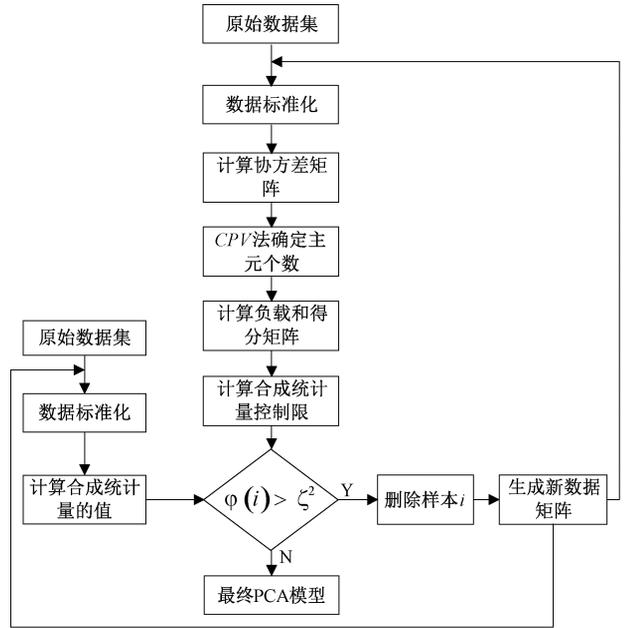


图 1 基于迭代建模的滑窗主元分析方法建模流程图

Fig. 1 Modeling flow chart of sliding window principal component analysis method based on iterative modeling

通过上述步骤,从训练数据中去除所有失控样本,并通过迭代过程建立优化的 PCA 模型。建立 PCA 模型后,采用一种基于第二置信限的滑窗主元分析方法进行故障检测。基本思想是定义一个滑动观测窗口,当前观测窗口内的报警总数(报警意味着合成统计量超过其置信限值)大于预设的限值时,则观测窗口内的当前样本就被确认为故障^[20]。

一般情况下,考虑一个为 $100\alpha\%$ 的控制极限,用于过程故障检测,当合成统计量的置信限为 ζ^2 ,则在正常运行状态下,合成统计量以 α 的概率超出置信限。即:

$$\begin{cases} p_{normal} = 1 - \alpha \\ p_{fault} = \alpha \end{cases} \quad (13)$$

α 可由下式得到:

$$\alpha = \frac{x_{alarm}}{x_{alarm} + x_{normal}} \times 100\% \quad (14)$$

式中: x_{alarm} 为检验过程中合成统计量超出第一置信限的个数, x_{normal} 为检验过程中合成统计量在第一置信限内的个数。

进一步考虑一个滑动观测单元(即观测窗口的长度),

有 n_1 个测试样本沿时间剖面的测量全长分布。如果每个观测单元中测试样本的合成统计量相互独立,则可以近似地认为合成统计量服从伯努利分布。然后其概率密度函数可以表示为:

$$b(k; n_1, \alpha) = \binom{n_1}{k} \alpha^k (1 - \alpha)^{n_1 - k} \quad (15)$$

式中: k 为一个观测单元中合成统计量的总报警数。

对此可以得到一个观测单元中合成统计量报警个数小于 s 的概率。对于在滑动窗口内识别到的故障点可以建立一个 $100\beta\%$ 的控制极限,该极限表示为在滑动窗口内允许的最大报警数,也就是前面描述中设置的限制值。可以描述为:

$$F(s; n_1, \alpha) = \sum_{i=0}^s C_{n_1}^i \alpha^i (1 - \alpha)^{n_1 - i} \leq \beta \quad (16)$$

通过式(16)可以推导出最大允许的 s ,当 $\beta = 1$ 时,最大允许 s 是观测窗口的长度。因此,在工业过程故障检测过程中,根据统计量给 β 分配一个合理的数值,就可以得到一个观测单元中合成统计量的最大允许报警数。 β 在实际工业过程中主要是由模型的精度决定,根据工业故障检测的多元统计分析方法的经验,通常设置在 $\beta \in [0.9, 0.99]$ 。

根据第二节内容分析可以得出 ζ^2 为第一置信限,则在上述中合成统计量允许的最大 s 称为第二置信限(即预设极限值 $m(Q)$),根据式(14)~(16)在给定的 α, β, n_1 的情况下,可以得到合成统计量的第二个置信极限值(即 $m(Q)$)。

为了确定第二置信限的值,滑动窗口大小 n_1 的选择也是非常重要的,滑动窗口的选择必须确保有一个失控的数据点允许在这个窗口,即 $s \geq 1$ 。另一方面窗口大小应该相对较小,以保持 s 的小值,从而获得对实际过程故障的足够灵敏度。选择一个合适的窗口大小对于故障检测的灵敏度非常重要,具体还是要根据实际情况确定观测窗口的大小。基于迭代建模的滑窗主元分析方法故障检测流程图如图2所示。

故障检测步骤为:

- 1) 采集第 i 个窗口内的第 j 个样本 x_{ij} , 对数据进行标准化处理,计算合成统计量控制限。
- 2) 若合成统计量的值大于控制限,转到步骤4);若合成统计量的值小于控制限,则进行下一步判断。
- 3) 若 j 等于 n 时,则令 $i = i + 1, j = 1$,转到步骤1);若 j 不等于 n 时,则令 $j = j + 1$,转到步骤1)。
- 4) 将报警数 $s(Q)$ 设置为0,令 $s(Q) = s(Q) + 1$,不断累积报警次数,直到窗口内的所有样本检查完毕。
- 5) 如果总报警数 $s(Q)$ 大于第2个置信限 $m(Q)$,则将当前 x_{ij} 视为真故障状态。否则将作为虚警处理,直接忽略。

3 实验仿真与结果分析

3.1 数值仿真

为了评估所提出的降低虚警的方法对 PCA 模型性能

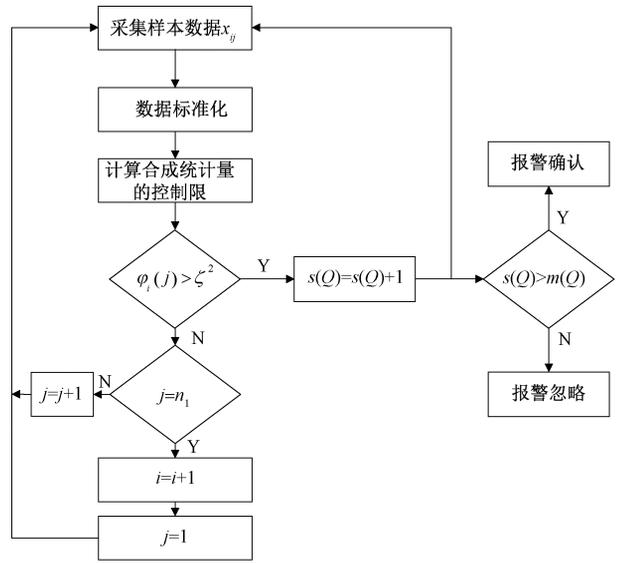


图2 基于迭代建模的滑窗主元分析方法故障检测流程图

Fig. 2 Fault detection flow chart of sliding window principal component analysis method based on iterative modeling

的影响,采用一个数值算例进行仿真实验^[21]。

$$\begin{cases} x_1 = 0.1 \times \text{randn}(1, n) \\ x_2 = 0.2 \times \text{randn}(1, n) \\ x_3 = 0.3 \times \text{randn}(1, n) \\ x_4 = -1.3x_1 + 0.2x_2 + 0.8x_3 + 0.1 \times \text{randn}(1, n) \\ x_5 = x_2 - 0.3x_3 + 0.1 \times \text{randn}(1, n) \\ x_6 = x_1 + x_4 + 0.1 \times \text{randn}(1, n) \end{cases} \quad (17)$$

式中: randn 是一个生成正态分布随机数的函数,产生均值为0、标准差为1的高斯噪声,模拟随机噪声或扰动。

采集1000个正常状态下的样本建立PCA故障检测模型。本实验共设置2种故障,一个属于漂移类别的,漂移幅度从0逐渐增加到1的线性序列,另一个故障模拟传感器测量精度较低的情况,模拟增加噪声(增加倍率为0.02)的方式。采集1000个样本用于测试所建立的PCA模型,对测试数据中的 x_1 的后400个样本点引入这两种故障。以合成统计量为指标衡量上述方法的降低虚警的情况和故障敏感性。仿真结果如图3、4所示。

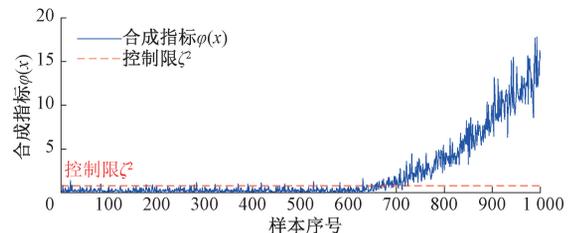


图3 在漂移情况下的检测图

Fig. 3 Detection diagram under drift condition

通过图3可以看出,在第600个样本之后引入漂移故

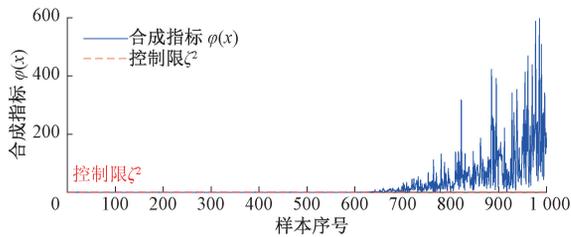


图 4 在精度下降情况下的检测图

Fig. 4 Detection diagram under the condition of precision decline

障,合成统计量逐渐上升,并且在后续样本中出现了越来越大的偏差。漂移故障使得模型能够逐渐检测到异常,且合成统计量随时间显著上升,最后大幅超出控制限。通过图 4 可以看出,在第 600 个样本后,合成统计量波动性显著增加,表现出频繁的高峰值,表明精度下降故障的影响。

将文中方法与基于 SPE 统计量和 T^2 统计量的主元分析故障诊断方法进行了比较,故障检测准确率如表 1 所示。

表 1 数值仿真故障检测准确率

Table 1 Numerical simulation fault detection accuracy

| 故障类型 | 准确率/% | | |
|------|---------|-----------|-------|
| | SPE 统计量 | T^2 统计量 | 合成指标 |
| 漂移 | 87.60 | 86.80 | 89.20 |
| 精度下降 | 82.00 | 81.00 | 83.70 |

通过表 1 可以得出,对漂移和精度下降这两种故障,合成指标在提高故障检测准确率方面优于单独使用的 SPE 和 T^2 统计量。合成指标结合了 SPE 和 T^2 的优势,提供了更全面、精准的故障检测能力。

将文中方法与传统 PCA 故障检测方法、基于统计方法的 PCA 故障检测方法^[20]进行了比较,故障检测结果如表 2 所示。

表 2 数值仿真故障检测结果

Table 2 Numerical simulation fault detection results

| 故障类型 | 检测指标 | 传统 | 统计 | 本文 |
|------|-------|------|------|------|
| | | PCA | PCA | 方法 |
| 漂移 | 虚警率/% | 1.83 | 1.50 | 1.33 |
| | 及时性 | 788 | 797 | 769 |
| 精度下降 | 虚警率/% | 2.50 | 2.17 | 2.08 |
| | 及时性 | 703 | 711 | 697 |

通过表 2 中结果分析可以得出,本文所提方法与传统 PCA 故障检测方法、基于统计方法的 PCA 故障检测方法相比,虚警率显著降低。在故障检测及时性方面,本文所提的方法分别在第 769 和 697 个测试样本检测出故障,而传统 PCA 故障检测方法分别在第 788 和 703 个测试样本检测出故障,基于统计方法的 PCA 故障检测方法分别在第

797 和 711 个测试样本检测出故障,故障检测及时性显著提高。

3.2 TE 过程仿真

TE 过程是由美国 Tennessee Eastman 化学公司创建的,为进行过程监控提供了一个实际化学工业生产过程的仿真实例。可以利用 TE 过程仿真中采集的数据来进行各种故障诊断方法的比较,验证各种方法的有效性。整个生产过程由反应器、冷凝器、气液分离器、循环压缩机和解吸塔 5 个操作单元组成。TE 过程包含 53 个变量,其中 22 个过程测量变量、19 个成分测量变量以及 12 个操作变量^[22]。TE 过程的数据集中每个样本都有 52 个观测变量,数据每 3 min 采样一次,不仅包含正常运行的数据集,还有 21 种故障环境下的数据集,21 种故障都是从第 161 个样本开始引入的,TE 过程流程如图 5 所示。

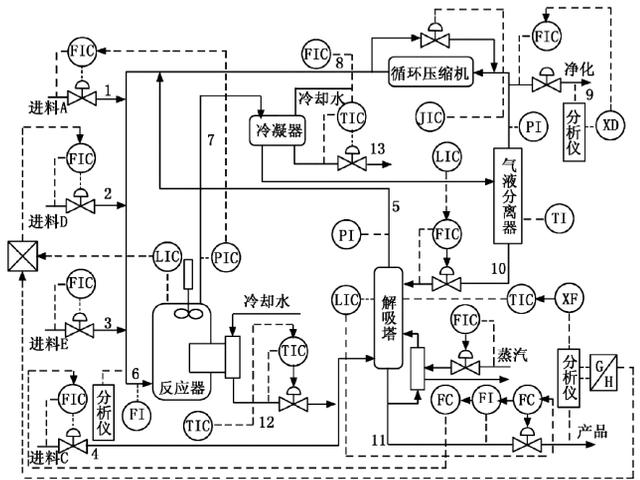
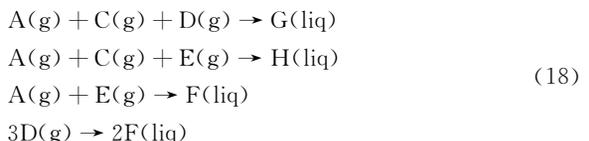


图 5 TE 过程流程图

Fig. 5 TE Process flow diagram

主要有 4 种气体物料参与反应,分别为 A、C、D 和 E,生产出两种产品 G、H,并伴有一种副产品 F,此外在产品的进料中含有少量的惰性气体 B。在催化剂的作用下,反应器中主要有 4 个同时进行的化学反应,其中的两个主化学反应生成的液态产物分别为产物 G 和产物 H,同时会生成副产物 F,化学反应方程式为:



在 TE 仿真实验过程中,选取过程的 3 种故障类型验证所提方法的可行性和有效性,分别为物料 C 压力损失,物料 A、B、C 的组成变化,动力学系数发生变化。故障原因及类型如表 3 所示。

采用本文所提出的方法对上述 3 种故障在 MATLAB 中进行仿真检测,从而验证所提方法的有效性,检测结果如图 6~8 所示。

表3 3种故障信息

Table 3 Three types of fault information

| 编号 | 故障现象 | 故障类型 |
|----|----------------|--------|
| 1 | 物料 C 压力头损失 | 阶跃故障 |
| 2 | 物料 A、B、C 的组成变化 | 随机变化故障 |
| 3 | 动力学常系数发生变化 | 慢漂移故障 |

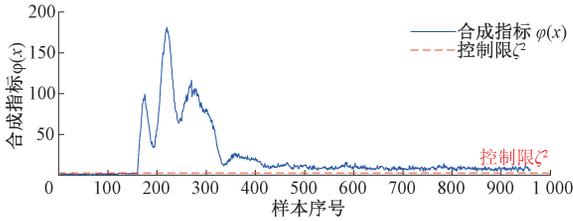


图6 故障1检测图

Fig. 6 Fault 1 detection diagram

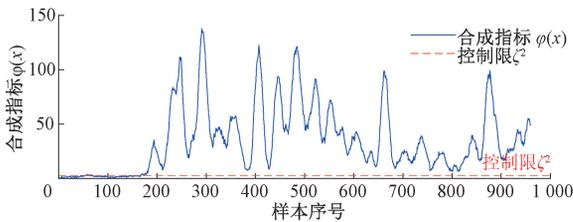


图7 故障2检测图

Fig. 7 Fault 2 detection diagram

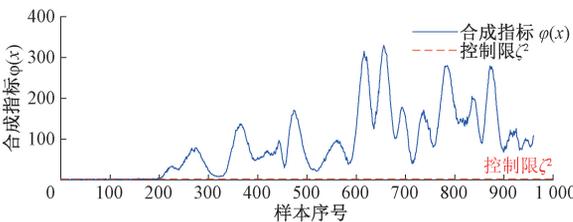


图8 故障3检测图

Fig. 8 Fault 3 detection diagram

从图6~8可以看出,从第161个样本开始引入3种不同故障后,合成统计量逐渐上升,并在后续样本中偏差逐步加大。不同类型的故障使模型能够逐步检测到异常,同时合成统计量随时间显著增加,最终超出控制限。

将文中方法与基于SPE统计量和 T^2 统计量的主元分析故障诊断方法进行了比较,TE过程故障检测准确率如表4所示。

表4 TE过程故障检测准确率

Table 4 TE process fault detection accuracy

| 故障类型 | 准确率/% | | |
|------|--------|-----------|-------|
| | SPE统计量 | T^2 统计量 | 合成指标 |
| 故障1 | 98.58 | 98.33 | 99.39 |
| 故障2 | 96.77 | 95.10 | 97.29 |

通过表4可以看出,在TE仿真实验中,合成指标的故障检测准确率优于SPE和 T^2 统计量的准确率,验证了合成指标提升检测准确性的有效性。

将文中提出的方法与传统PCA故障检测方法及基于统计的PCA故障检测方法进行了比较,故障检测结果如表5所示。

表5 TE过程故障检测结果

Table 5 TE process fault detection results

| 故障类型 | 检测指标 | 传统 | 统计 | 本文 |
|------|-------|-------|-------|-------|
| | | PCA | PCA | 方法 |
| 故障1 | 虚警率/% | 4.38 | 3.75 | 3.12 |
| | 及时性 | 188 | 200 | 177 |
| 故障2 | 虚警率/% | 16.88 | 15.62 | 15.00 |
| | 及时性 | 200 | 216 | 194 |
| 故障3 | 虚警率/% | 5.62 | 5.23 | 4.85 |
| | 及时性 | 227 | 242 | 218 |

根据表5中数据对比,传统PCA在3种故障下的虚警率最高,误报较多,而本文方法在3种故障情况下均表现出最低的虚警率,从而提升了故障检测的可靠性。在故障检测的及时性方面,本文方法比其他两种PCA模型更迅速地检测到故障(分别在第177、第194和第218个测试样本时),显著提高了检测的及时性。

4 结论

本文提出了一种基于迭代建模的滑窗主元分析故障检测方法,旨在改进传统PCA在工业过程故障检测中的性能,特别是在降低虚警率和提高检测及时性方面。该方法采用合成统计量作为检测指标,能够同时考虑主成分方向和残差空间的异常,从而提高了故障检测的准确性。通过在建模阶段迭代剔除训练数据中的异常样本,最终构建了更加稳健的PCA模型。在检测阶段,采用滑动观测窗口统计样本的异常次数,以进一步提升对随机噪声、环境扰动以及模型误差的抗干扰能力。仿真结果表明,本方法在多个故障场景下有效降低了虚警率,并在故障发生时实现了更为快速的响应,证明了其在复杂工业系统监测中的有效性和敏感性。

该方法的成功验证了将迭代优化与统计检测结合在复杂工业过程监测中的潜力,为智能化的工业故障检测提供了一种实用的解决思路。未来研究可以进一步扩展该方法,结合多元统计分析及机器学习技术,以提高对更复杂非线性、多模态工业过程的适应性,并开发更广泛的实时故障检测应用。

参考文献

- [1] 徐雄,林海军,刘悠勇,等.融合PCA与自适应K-Means聚类的水电机组故障检测在线方法[J].电子测

- 量与仪器学报, 2022, 36(3): 260-267.
- XU X, LIN H J, LIU Y Y, et al. On-line fault detection method of hydraulic turbine combining PCA and adaptive K-Means clustering [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 260-267.
- [2] 李榕, 申志, 李元. 基于核熵成分分析的工业过程多类型故障诊断[J]. 电子测量技术, 2023, 46(10): 40-45.
- LI R, SHEN ZH, LI Y. Multi-type fault diagnosis of industrial process based on KECA [J]. Electronic Measurement Technology, 2023, 46(10): 40-45.
- [3] 杨帅, 张岩, 梁永春, 等. 基于多维特征与优化 SVM 在高压断路器故障分类中的应用[J]. 国外电子测量技术, 2024, 43(8): 150-159.
- YANG SH, ZHANG Y, LIANG Y CH, et al. Application of multidimensional feature-based and optimized SVM in high-voltage circuit breaker fault classification [J]. Foreign Electronic Measurement Technology, 2024, 43(8): 150-159.
- [4] 江兵, 李响, 巢一帆, 等. 基于 KPCA-CGSSA-KELM 的变压器故障识别方法[J]. 电子测量与仪器学报, 2024, 38(5): 139-147.
- JIANG B, LI X, CHAO Y F, et al. Transformer fault recognition method based on KPCA-CGSSA-KELM [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(5): 139-147.
- [5] 谢彦红, 杨滕, 贾冬妮, 等. 基于主元分析的滑动窗口累积和的微小故障检测[J]. 计算机应用与软件, 2023, 40(5): 60-66, 96.
- XIE Y H, YANG T, JIA D N, et al. Minor fault detection based on PCA and moving window cumulative sum [J]. Computer Applications and Software, 2023, 40(5): 60-66, 96.
- [6] ZHU SH M, XIA H, ANNOR-NYARKO M, et al. A robust strategy for sensor fault detection in nuclear power plants based on principal component analysis[J]. Annals of Nuclear Energy, 2021, 164: 108621.
- [7] HARROU F, NOUNOU M N, NOUNOU H N, et al. Statistical fault detection using PCA-based GLR hypothesis testing[J]. Journal of Loss Prevention in the Process Industries, 2013, 26(1): 129-139.
- [8] 杜海莲, 苗诗瑜, 杜文霞, 等. 改进主元分析方法及数据重构在工业系统中的故障诊断研究[J]. 南京理工大学学报(自然科学版), 2019, 43(1): 72-77.
- DU H L, MIAO SH Y, DU W X, et al. Research on fault diagnosis of industrial process based on improved PCA method and data reconstruction [J]. Journal of Nanjing University of Science and Technology (Natural Science Edition), 2019, 43(1): 72-77.
- [9] 鲍中新, 文成林, 马雪. 一种基于数据变化率的预处理及主元分析故障诊断方法 [J]. 电子学报, 2021, 49(11): 2234-2240.
- BAO ZH X, WEN CH L, MA X. Data preprocessing and PCA fault diagnosis method based on rate of change transformation [J]. Acta Electronica Sinica, 2021, 49(11): 2234-2240.
- [10] 顾炳斌, 熊伟丽, 史旭东. 基于故障敏感主元的多块 PCA 故障监测方法 [J]. 高校化学工程学报, 2019, 33(6): 1499-1508.
- GU B B, XIONG W L, SHI X D. Multi-block PCA process monitoring based on fault sensitive principal components [J]. Journal of Chemical Engineering of Chinese Universities, 2019, 33(6): 1499-1508.
- [11] YANG Y H, CHEN X M, ZHANG Y, et al. A novel decentralized weighted ReliefF-PCA method for fault detection [J]. IEEE Access, 2019, 7: 140478-140487.
- [12] WANG Y, HU SH L. State monitoring and fault prediction of centrifugal compressors based on long short-term memory and principal component analysis [J]. PeerJ Computer Science, 2024, 10: e2433.
- [13] ZHANG J X, DAI Y Y, FENG Z M, et al. An enhanced temporal algorithm-coupled optimized adaptive sparse principal component analysis methodology for fault diagnosis of chemical processes [J]. Process Safety and Environmental Protection, 2023, 174: 663-680.
- [14] CAO H S, SUN P W, ZHAO L, et al. PCA-SVM method with sliding window for online fault diagnosis of a small pressurized water reactor [J]. Annals of Nuclear Energy, 2022, 171: 109036.
- [15] 张成, 郭青秀, 李元, 等. 基于主元分析得分重构差分的故障检测策略 [J]. 控制理论与应用, 2019, 36(5): 774-782.
- ZHANG CH, GUO Q X, LI Y, et al. Fault detection strategy based on difference of score reconstruction associated with principal component analysis [J]. Control Theory and Applications, 2019, 36(5): 774-782.
- [16] ZHANG CH, GUO Q X, LI Y, et al. Fault detection in the Tennessee Eastman benchmark process using principal component difference based on K-Nearest neighbors [J]. IEEE Access, 2020, 8: 49999-50009.
- [17] 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 [J]. 自动化学报, 2020, 46(10): 2229-2238.
- ZHANG CH, GAO X W, LI Y. Fault detection strategy based on principal component score difference of k nearest neighbors [J]. Acta Automatica Sinica,

- 2020, 46(10): 2229-2238.
- [18] PORTNOY I, MEIENDEZ K, PINZON H, et al. An improved weighted recursive PCA algorithm for adaptive fault detection [J]. Control Engineering Practice, 2016, 50: 69-83.
- [19] LI W, PENG M, WANG Q ZH. Improved PCA method for sensor fault detection and isolation in a nuclear power plant [J]. Nuclear Engineering and Technology, 2019, 51(1): 146-154.
- [20] LI W, PENG M, WANG Q ZH. False alarm reducing in PCA method for sensor fault detection in a nuclear power plant [J]. Annals of Nuclear Energy, 2018, 118: 131-139.
- [21] WANG Z H, WEN CH L, XU X M, et al. Fault diagnosis method based on gap metric data preprocessing and principal component analysis [J]. Journal of Control Science and Engineering, 2018, 2018(1): 1-9.
- [22] JIANG Q CH, YAN X F, HUANG B. Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference [J]. IEEE Transactions on Industrial Electronics, 2016, 63(1): 377-386.

作者简介

程为康, 硕士研究生, 主要研究方向为工业过程故障检测与诊断。

E-mail: 2437623073@qq.com

张家良(通信作者), 博士, 讲师, 主要研究方向为工业过程故障。

E-mail: zjl512@163.com