

去主元相关性 DKPCA 故障检测与诊断方法

张成¹, 韩宏宇², 李元³

(1. 沈阳化工大学理学院, 辽宁 沈阳 110142;

2. 沈阳化工大学计算机科学与技术学院, 辽宁 沈阳 110142;

3. 沈阳化工大学信息工程学院, 辽宁 沈阳 110142)

摘要:针对动态核主元分析(Dynamic Kernel Principal Component Analysis, DKPCA)在动态非线性过程监控中没有降低数据动态性的影响,导致统计量 T^2 具有显著自相关性的问题,提出一种基于去主元相关性的DKPCA(Dynamic Kernel Principal Component Analysis based on Removing Principal Component Correlation, DKPCA-RPCC)故障检测与诊断方法。首先,对原始数据 X 进行时滞扩展生成增广矩阵 Y 并使用KPCA计算主成分 M ;其次,利用已知数据重构增广矩阵 \hat{Y} ,再使用KPCA计算主成分 \hat{M} ;然后,通过主成分之间的差异来构建统计量进行故障检测;最后,利用基于变量贡献图的方法进行故障诊断。通过数值例子和田纳西-伊斯曼(Tennessee-Eastman, TE)过程进行仿真验证,并将仿真结果与KPCA、DPCA和DKPCA的结果进行对比。仿真结果说明,该方法在动态非线性过程监控中构建的统计量故障检测性能更高且具有较低的自相关性。

关键词:动态核主元分析;过程监控;自相关性;去主元相关性;故障诊断

中图分类号:TP 277

文献标识码:A

文章编号:1673-629X(2023)04-0161-07

doi:10.3969/j.issn.1673-629X.2023.04.024

Fault Detection and Diagnosis Method Based on Removing Principal Component Correlation DKPCA

ZHANG Cheng¹, HAN Hong-yu², LI Yuan³

(1. School of Science, Shenyang University of Chemical Technology, Shenyang 110142, China;

2. School of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110142, China;

3. School of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China)

Abstract: Aiming at the problem that Dynamic Kernel Principal Component Analysis (DKPCA) does not reduce the influence of data dynamics in dynamic nonlinear process monitoring, resulting in significant autocorrelation of statistic T^2 , a fault detection and diagnosis method for DKPCA based on removing principal component correlation (DKPCA-RPCC) was proposed. Firstly, the original data X was time-delayed to generate the augmented matrix Y , and the principal component M was calculated by KPCA. Secondly, the known data was used to reconstruct the augmented matrix \hat{Y} and then calculate the main component \hat{M} by KPCA. Then, the difference between principal components was used to construct statistics for fault detection. Finally, the method based on variable contribution graph was used for fault diagnosis. Numerical examples and Tennessee-Eastman (TE) process were used for simulation verification, and the simulation results were compared with those of KPCA, DPCA and DKPCA. The simulation results show that the proposed method has higher fault detection performance and lower autocorrelation in dynamic nonlinear process monitoring.

Key words: dynamic kernel principal component analysis; process monitoring; autocorrelation; remove principal component correlation; fault diagnosis

0 引言

随着现代工业对产品质量和生产安全要求的逐步

提高,基于数据驱动的多元统计过程监控(Multivariate Statistical Process Monitoring, MSPM)得到了广泛的

收稿日期:2022-06-10

修回日期:2022-10-12

基金项目:国家自然科学基金科学项目(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)

作者简介:张成(1979-),男,副教授,博士,研究方向为化工过程监视、故障检测与诊断;通讯作者:李元(1964-),女,教授,博士,研究方向为数据驱动的过程故障诊断。

应用^[1-2]。

针对多变量监控问题,传统的主元分析(Principal Component Analysis, PCA)监控方法可以很好地用于稳态线性过程,或者用于监控非线性过程的局部操作区域(近似线性)^[3-4]。但是将PCA方法应用于非线性或动态过程可能导致低效和不可靠的过程监控,主要原因是线性静态的PCA方法不适合描述过程变量中的非线性或动态特征。为了有效地监控非线性过程,基于核主元分析(Kernel PCA, KPCA)的多元统计过程监控方法被应用于工业过程^[5]。然而, KPCA并没有考虑变量的时间相关性,即数据具有动态性时, KPCA的监控性能就受到制约。为了描述数据的动态特征, Ku等提出了一种动态主元分析(Dynamic Principal Component Analysis, DPCA)方法^[6],即将PCA应用于一个时间滞后的数据矩阵^[7-8],从而将样本之间的相关性转化为变量之间的相关性。DPCA监控方法也只适用于线性过程,如果将DPCA方法应用于动态非线性过程监控,那么该方法的监控性能就受到制约。为了解决动态非线性过程的监控问题, Choi等提出了一种基于动态核主元分析(Dynamic KPCA, DKPCA)^[9]的过程监控方法。DKPCA的主要思想是在执行KPCA之前对数据矩阵进行了时滞扩展。虽然DKPCA能提取数据的非线性特征和动态特征,检测结果也优于KPCA和DPCA的结果,但是通过大量的实验证明,统计量 T^2 上仍含有显著的自相关性^[10],这说明通过时滞扩展并没有彻底解决数据含有动态性的问题。

针对上述问题,该文提出一种基于去主元相关性的DKPCA动态非线性过程故障检测与诊断方法(Dynamic Kernel Principal Component Analysis based on Removing Principal Component Correlation, DKPCA-RPCC),其基本思想如下:首先,将低维数据引入时滞参数构成增广矩阵,再通过KPCA方法计算增广矩阵的主成分,称为真实得分;其次,假设当前时刻样本缺失,利用前移时刻样本估计当前值,构成新的增广矩阵,再按照相同的方法计算新矩阵的主成分,称为预估得分;然后,使用真实得分与预估得分之间的差异来构建统计量进行故障检测;最后,利用基于变量贡献图的方法进行故障诊断。该方法在DKPCA的基础上去除主元之间的相关性,能够最大限度地降低样本之间动态性的影响。

1 动态核主元分析

动态核主元分析^[9]的核心思想是在应用KPCA之前对数据矩阵进行时滞扩展。该方法不仅可以描述变量之间的非线性特征,而且由于加入了额外的时移

变量,所以也能描述变量之间的自相关性和滞后交叉相关性。此外,通过选择适当时移数 l ,使得变量之间的非线性和动态关系都出现在方差较小的主元所对应的噪声子空间中^[11]。

假设 $\mathbf{X} = [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \cdots \ \mathbf{x}_n^T]^T$ 是包含 n 个样本的数据集,其中 $\mathbf{x}_i \in \mathbb{R}^{1 \times m}$, m 表示变量个数。通过引入时滞参数 l 构建增广向量,如式(1)所示。

$$\mathbf{x}_i^{(l)} = [\mathbf{x}_i \ \mathbf{x}_{i-1} \ \cdots \ \mathbf{x}_{i-l}]^T \quad (1)$$

其中, i 代表一个任意的采样瞬间, $\mathbf{x}_i^{(l)}$ 表示瞬间时刻 i 的增广向量。

记 $\mathbf{y}_i = \mathbf{x}_i^{(l)}$,该方法通过叠加增广向量 \mathbf{y}_i 得到增广矩阵 $\mathbf{Y} = [\mathbf{y}_1^T \ \mathbf{y}_2^T \ \cdots \ \mathbf{y}_n^T]^T$ 。将增广矩阵 \mathbf{Y} 进行中心化,得到 $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}_1^T \ \bar{\mathbf{y}}_2^T \ \cdots \ \bar{\mathbf{y}}_n^T]^T$ 。在动态核主元分析中,首先,将 $\bar{\mathbf{y}}_i$ 通过非线性函数 φ 映射到高维空间,高维空间的数据集为 $\varphi(\bar{\mathbf{Y}}) = [\varphi(\bar{\mathbf{y}}_1)^T \ \varphi(\bar{\mathbf{y}}_2)^T \ \cdots \ \varphi(\bar{\mathbf{y}}_n)^T]^T$ 。 $\varphi(\bar{\mathbf{Y}})$ 的协方差矩阵如式(2)所示。对协方差矩阵进行奇异值分解,得到特征向量 \mathbf{p} 和特征值 λ ,如式(3)所示。

$$\mathbf{C} = \frac{1}{n-1} \varphi(\bar{\mathbf{Y}})^T \varphi(\bar{\mathbf{Y}}) \quad (2)$$

$$\mathbf{C} \mathbf{p} = \lambda \mathbf{p} \quad (3)$$

通过 φ 映射之后,原始数据在高维空间的维数较大,无法直接计算特征值和特征向量,因此引入径向基核函数,通过核技巧间接计算高维空间数据集的特征值和特征向量,从而简化计算。

式(4)和式(5)分别表示核矩阵和中心化的核矩阵,通过奇异值分解计算 $\bar{\mathbf{K}}$ 的特征值和特征向量,如式(6)所示,其中 \mathbf{u} 为矩阵 $\bar{\mathbf{K}}$ 的特征向量, μ 为矩阵 $\bar{\mathbf{K}}$ 的特征值。

$$\mathbf{K} = \varphi(\bar{\mathbf{Y}}) \varphi(\bar{\mathbf{Y}})^T \quad (4)$$

$$\bar{\mathbf{K}} = \mathbf{K} - \frac{1}{n} \mathbf{K} \mathbf{E}^T - \frac{1}{n} \mathbf{E} \mathbf{K} + \frac{1}{n^2} \mathbf{E} \mathbf{K} \mathbf{E}^T \quad (5)$$

$$\bar{\mathbf{K}} \mathbf{u} = \mu \mathbf{u} \quad (6)$$

在高维空间中,样本协方差矩阵的特征向量和特征值与核矩阵的特征向量和特征值之间存在以下关系,如式(7)所示。为了简化计算,将协方差矩阵的特征向量单位化,得到单位向量 \mathbf{p} ,如式(8)所示。高维空间中数据集的得分 t_i 可通过式(9)计算得到,其中 \mathbf{P} 为协方差的特征向量矩阵, \mathbf{U} 为核矩阵的特征向量矩阵。

$$\begin{cases} p_i = \varphi(\bar{\mathbf{Y}})^T \mathbf{u}_i \\ \lambda_i = \frac{1}{n-1} \mu_i \end{cases} \quad (7)$$

$$\mathbf{p}_i = \frac{1}{\sqrt{\lambda}} \varphi(\bar{\mathbf{Y}})^T \mathbf{u}_i \quad (8)$$

$$t_i = \frac{1}{\sqrt{\lambda}} K_i U \quad (9)$$

DKPCA 的监控指标^[9]如式(10)所示:

$$T^2 = t \Lambda^{-1} t^T \quad (10)$$

其中, Λ 是由特征值 λ 组成的对角矩阵。

2 基于去主元相关性的 DKPCA 故障检测和诊断方法

DKPCA 方法通过时滞扩展并没有完全解决数据具有动态性的问题,这导致统计量仍含有较强的自相关性^[10]。为了解决这一问题,将 DKPCA 与缺失值处理技术^[11]相结合并利用主成分的差异构建统计量。

具体方法如下:在 DKPCA 中,建立一个包含当前与过去测量值的矩阵 Y 。对于每个新的观测向量 y_i 可以计算得分。假设当前的观测向量缺失,可以使用缺失值处理技术,从过去的测量值中估计出当前时刻的测量值并计算得分,这本质上是对得分和观察值进行了一步预测。因此,得分之间的差异几乎是连续非相关的^[8],这意味着可以使用得分差异更好地对样本进行监控。所选择的缺失值处理技术为条件均值替换法^[12],该方法对缺失数据的测量矢量进行重新排列,不失一般性。

$$Y = \begin{bmatrix} Y^\# \\ Y^* \end{bmatrix} \quad (11)$$

式中, $Y^\#$ 表示缺失的测量值, Y^* 表示已知的测量值。

相应的,把特征向量 P 进行划分 $P^T = [P^{\#T} \ P^{*T}]^T$,缺失的测量值 $Y^\#$ 可以通过期望最大化算法 (Expectation Maximization, EM)^[13-14] 进行估计。在每次迭代时,缺失值被已知数据在当前正态分布条件下的均值和协方差矩阵所代替^[13],如式(12)所示。

$$\hat{Y}^\# = E(Y^\# | Y^*, \bar{Y}, S) \quad (12)$$

假设增广矩阵已知。为了计算缺失向量的观测值,只需要 EM 算法的期望步骤,把 P 代入到 S 的表达式,可以得到式(13)。使用式(13)可以计算缺失观测值的条件期望,如式(14)所示。

$$S = \begin{bmatrix} P^\# \Lambda P^{\#T} & P^\# \Lambda P^{*T} \\ P^* \Lambda P^{\#T} & P^* \Lambda P^{*T} \end{bmatrix} \quad (13)$$

$$\hat{Y}^\# = P^\# \Lambda P^{*T} (P^* \Lambda P^{*T})^{-1} Y^* \quad (14)$$

最终可以得到一个新的矩阵,如式(15)所示。

$$\hat{Y} = \begin{bmatrix} \hat{Y}^\# \\ Y^* \end{bmatrix} \quad (15)$$

使用 KPCA 方法分别计算矩阵 Y 、 \hat{Y} 的得分,如式(16)和式(17)所示,其中 K_1 是由矩阵 Y 计算得到的核矩阵, U 表示 K_1 的特征向量矩阵, M 表示矩阵 Y 的

得分,称为真实得分。 K_2 是由矩阵 \hat{Y} 计算得到的核矩阵, \hat{M} 表示矩阵 \hat{Y} 的得分,称为预估得分。

$$M = K_1 U \quad (16)$$

$$\hat{M} = K_2 U \quad (17)$$

上述过程可以对真实得分估计出最符合实际的测量值,并定义一个统计量 W ,如式(18)所示,其中 m 、 \hat{m} 分别表示每个训练样本和观测样本的得分, $S_{m-\hat{m}}$ 表示真实得分与观测得分之间差异的样本协方差矩阵。

$$W = (m - \hat{m})^T S_{m-\hat{m}}^{-1} (m - \hat{m}) \quad (18)$$

(1) 离线建模。

第一步:对训练数据集 X 进行标准化。

第二步:根据时移 l 生成增广矩阵 Y_x ,然后通过 KPCA 计算真实得分 M_x 。

第三步:假设增广矩阵 $Y_x^\#$ 缺失,利用 EM 算法估计缺失值 $\hat{Y}_x^\#$,得到新的增广矩阵 \hat{Y}_x ,再通过 KPCA 计算预估得分 \hat{M}_x 。

第四步:根据式(18)计算统计量 W ,使用核密度估计方法^[15]确定控制限 W_{UCL} 。

(2) 在线检测。

第一步:对测试样本应用训练数据的均值和方差进行标准化。

第二步:根据时移 l 生成增广向量 Y_{f_i} ,使用训练数据的核矩阵进行中心化,求出测试数据的真实得分 M_{f_i} 。

第三步:假设 Y_{f_i} 缺失的部分为 $Y_{f_i}^\#$,利用 EM 算法估计缺失值 $\hat{Y}_{f_i}^\#$,得到新的增广向量 \hat{Y}_{f_i} ,使用训练数据的核矩阵进行中心化,求出测试数据的预估得分 \hat{M}_{f_i} 。

第四步:根据式(18)计算测试样本的 W_i ,若 $W_i > W_{UCL}$,则判定 Y_{f_i} 为故障样本,否则 Y_{f_i} 为正常样本。

(3) 故障诊断。

对文中方法的动态非线性过程进行分析可以发现,低维空间的增广矩阵与非线性主元构成的监控统计量之间存在一定的关系,即故障变量和非线性主元之间有较大的相关性,而与故障无关变量相关性较弱^[16]。贡献度计算方法如式(19)所示,其中 $Y_{f,i}^T$ 表示测试数据的第 j 个变量, $M_{f,i,d} - \hat{M}_{f,i,d}$ 表示测试数据的第 i 个非线性主元的差异, g_i 表示第 i 个非线性主元的方差, $\text{cont}_{j,i}$ 表示第 i 个样本的第 j 个变量的贡献度, CON_j 表示所有样本第 j 个变量贡献度的累加和。一般情况下, $\text{cont}_{j,i}$ 具有较大贡献的变量为引发故障的变量。

$$\begin{cases} \text{cont}_{j,i} = \frac{Y_{f,i,j}^T \cdot (M_{f,i,d} - \hat{M}_{f,i,d})}{g_i} \\ \text{CON}_j = \sum_{i=1}^q \text{cont}_{j,i} \end{cases} \quad (19)$$

3 仿真实验

3.1 数值例子

使用一个非线性动态数值例子^[9]测试文中方法的有效性,具体模型如式(20)所示。

$$\begin{cases} \mathbf{z}(t+1) = \mathbf{A}\mathbf{z}(t) + \mathbf{B}f(\mathbf{u}(t)) \\ \mathbf{y}(t) = \mathbf{z}(t) + \mathbf{o}(t) \\ \mathbf{u}(t) = \mathbf{C}\mathbf{u}(t-1) + \mathbf{D}\mathbf{w}(t-1) \end{cases} \quad (20)$$

其中, $\mathbf{u}(t)$ 、 $\mathbf{y}(t)$ 、 $\mathbf{z}(t)$ 分别是输入变量、输出变量以及状态变量, $\mathbf{o}(t)$ 是服从均匀分布 $U \sim (0, 0.5)$ 的随机噪声, $\mathbf{w}(t)$ 是服从均匀分布 $U \sim (0, 5)$ 的白噪声, 系统 \mathbf{A} 、 \mathbf{B} 、 \mathbf{C} 、 \mathbf{D} 、 $\mathbf{w}(t)$ 、 $\mathbf{o}(t)$ 分别为:

$$\mathbf{A} = \begin{bmatrix} 0.118 & -0.191 & 0 \\ 0.847 & 0.264 & 0.9 \\ 0.214 & -0.11 & 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0.05 & 0.1 \\ 0.05 & 0.05 \\ 0 & 0.05 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix}$$

$$\mathbf{w}(t) = [w_1(t) \quad w_2(t)]^T$$

$$\mathbf{o}(t) = [o_1(t) \quad o_2(t) \quad o_3(t)]^T$$

其中, 非线性映射函数 $f(\mathbf{u}(t))$ 如式(21)所示:

$$f(\mathbf{u}(t)) = [u_1^2(t) \quad u_2^2(t)]^T \quad (21)$$

由以上模型随机生成 1 000 个正常数据作为训练样本, 然后, 在变量 w_1 加上一个幅值为 -4.5 的阶跃变化作为故障数据。再由该模型生成 100 组测试数据集, 每组测试数据前 500 个为正常样本, 后 500 个为故障样本。输出变量 $u(t)$ 、 $y(t)$ 被用于过程监控。采用 DPCA、KPCA、DKPCA 以及文中方法对此动态非线性过程例子进行仿真实验, 并对上述的四种方法的结果进行分析。

采用 DPCA 对该数值例子进行故障检测时, 按照 95% 的累计方差贡献率选取主元, 因此主元数设置为 5。按照平行分析法^[17], 时滞参数的选取与特征值有关, 特征值为 0 或接近于 0 时表示变量之间线性无关, 因而时滞参数 l 的值等于特征值为 0 或者接近于 0 的个数^[6]。DPCA 提取的主元之间含有较强的相关性, 特别是前两个主元之间的相关性最大, 这将导致计算的统计量 T^2 也含有较强的自相关性, 严重影响故障的检测结果。

采用 KPCA 对此数值例子进行故障检测时, 通过经验法^[18]确定核宽参数的范围以及选取最优核宽参数。低维空间的数据集通过非线性函数映射, 使得数据在高维空间线性可分, 此时故障样本与正常样本之

间的差异最大, 检测结果达到最优。采用 DKPCA 对此数值例子进行故障检测时, 核宽参数 σ 和时滞参数 l 采用经典的网格搜索和交叉验证^[19]方式进行选取, 该方法选取不同的参数时, 训练数据的准确率如图 1 所示。

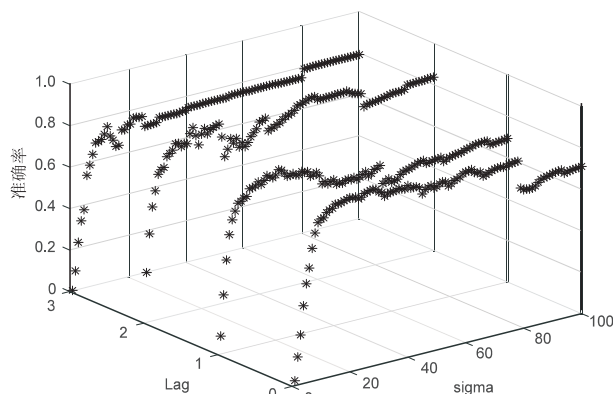


图1 校验数据的准确率

采用文中方法进行故障检测时, 核宽参数与时滞参数的选取均与 DKPCA 保持一致。图 2 给出了上述四种方法对于 100 组测试数据的检测结果, 由图 2 可以看出, 文中方法的检测结果明显高于其他三种方法且平均检测效率最高。图 3 表示四种方法统计值自相关性的对比。由图 3 可以看出, DPCA、KPCA 和 DKPCA 方法的统计值都含有较强的一步自相关性, 而

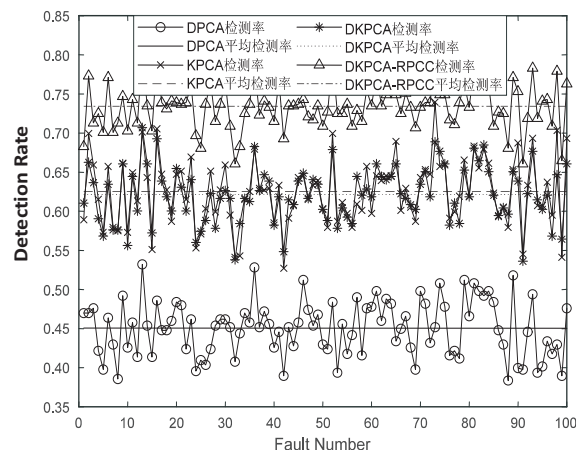


图2 四种方法故障检测结果

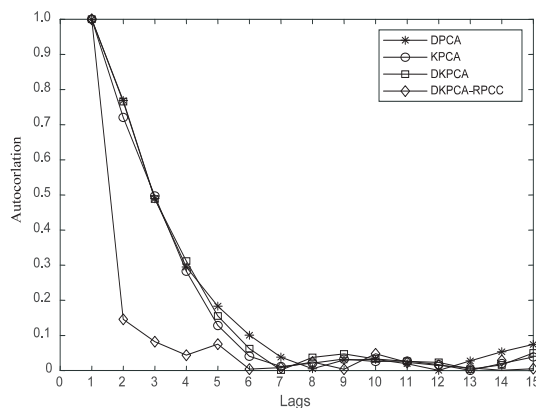


图3 统计量的自相关性

文中方法的统计量的自相关性明显降低,因此能够提高故障检测率。图4给出了文中方法对第一组数据的故障检测结果,由图4可知,故障样本与正常样本几乎完全分离,证明了文中方法的有效性。使用公式(19)对测试数据进行故障诊断,测试数据集10个变量对样本的贡献程度,如图5所示,从图5可以看出第4、9变量的贡献度较大,所以故障发生与这两个变量的相关性最大。

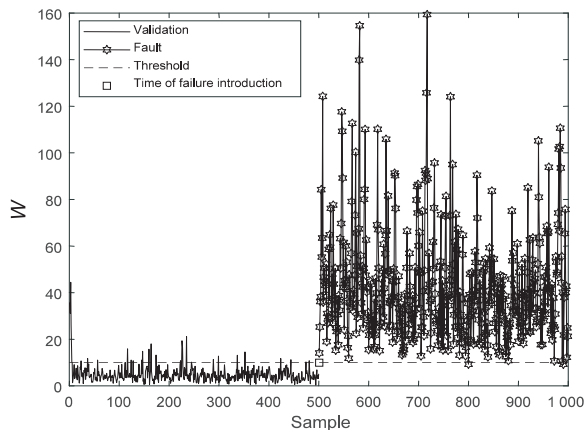


图4 DKPCA-RPCC 故障检测结果

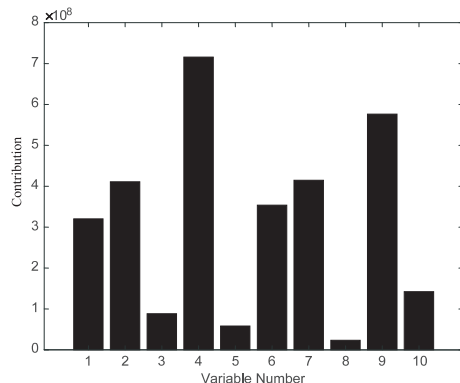


图5 样本变量的累计贡献

3.2 Tennessee Eastman

本节的仿真数据由最新的 TE 仿真器生成^[20],其基本结构如图6所示。数据集在 TE 仿真器中通过0.01小时采样间隔并持续运行10小时获得。正常状态下采集1000个样本作为训练集;同时,过程故障在3小时后引入并持续到过程结束。TE 过程共包含41个测量变量和12个控制变量,仿真器运行过程中有三个变量恒定不变,所以选取剩余的50个变量进行过程监控。

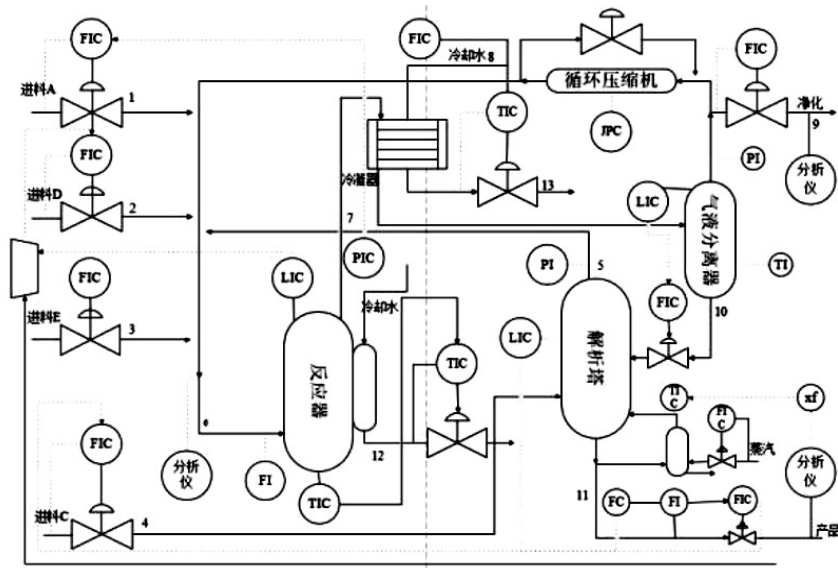


图6 TE 过程

仿真使用的计算机环境为 Intel(R) Core(TM) i7-7700HQ CPU @ 2.80 GHz 处理器的 64 位操作系统,且所有实验结果均在 MATLAB 中实现。

利用文中方法对 TE 数据进行仿真验证,并与 KPCA、DPCA、DKPCA 进行对比分析。四种方法均按照 95% 的累计贡献率选取主元,此外 DPCA 的时滞参数选取为 1, DKPCA 和文中方法的核宽参数和时滞参数均通过网格搜索和交叉验证的方式选取。表1给出了上述各种方法的检测结果,由于第六组数据缺失,因此不做测试。以第27组测试数据为例进行故障检测,图7分别给出 DKPCA-RPCC、KPCA、DPCA 和 DKPCA

对于故障27的检测结果。对比四种方法的检测结果可以看出,文中方法的检测率最高。为了进一步确定故障产生的原因,使用式(19)对第27组数据进行故障诊断,图8给出了故障27的诊断结果。从图中可以看出,第9个和第49个变量的贡献度最大,即故障27是由变量9和49引起的。变量9和49分别代表反应堆温度和反应堆冷却水流量,图9给出了反应堆温度和反应堆冷却水流量的变化趋势,由图可以看出与计算的结果恰好相互印证,与实际相符。

图10表示四种方法统计量的自相关性,明显可以看出,DKPCA-RPCC方法构建的统计量具有较低水平

的自相关性。根据故障检测率以及构建统计量的自相关性检验所提出方法的性能。统计量的自相关性越低,说明主成分之间含有的动态性越低,故障检测性能更好。

表 1 TE 过程中各种方法故障检测率

故障号	KPCA	DPCA	DKPCA	DKPCA-RPCC
	T^2	T^2	T^2	W
1	99.14	99.43	17.13	25.25
2	93.42	97.42	96.66	100
3	1	27.51	14.95	89.11
4	100	100	100	100
5	0.14	53.58	39.62	100
7	100	100	100	100
8	87.55	90.11	68.65	92.16
9	3.58	23.78	13.64	87.95
10	84.26	89.26	90.57	93.61
11	95.28	99.28	99.56	100
12	35.34	80.95	85.34	98.26
13	94.28	95.13	29.17	77.21
14	98.86	99	99.42	100
15	0	10.89	1.74	98.26
16	5.87	4.58	3.77	4.64
17	81.55	84.53	37.88	55.88
18	49.07	63.04	61.10	73.88
19	91.27	96.42	97.68	97.53
20	84.98	85.10	86.50	89.70
21	5.87	4.44	4.06	5.37
22	0.86	30.66	21.04	85.78
23	9.87	10.03	5.52	73.88
24	80.54	86.96	86.94	91.87
25	74.68	85.39	77.94	99.13
26	78.25	90.11	92.89	96.23
27	70.10	94.70	80.84	98.11
28	0.29	35.24	30.19	91.58

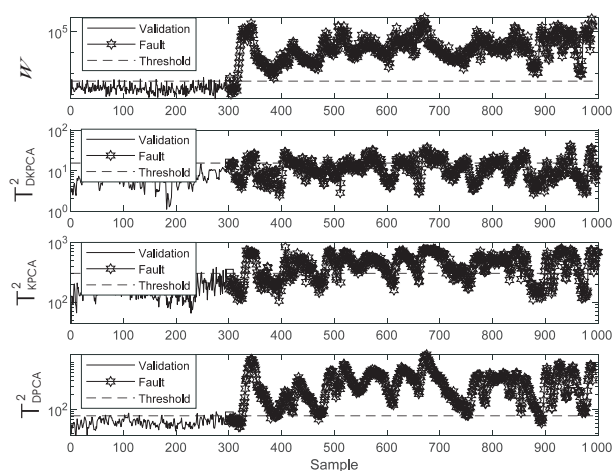


图 7 故障 27 的检测结果

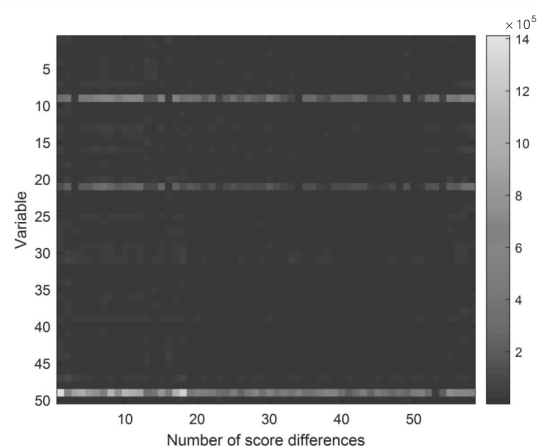


图 8 故障 27 的贡献图

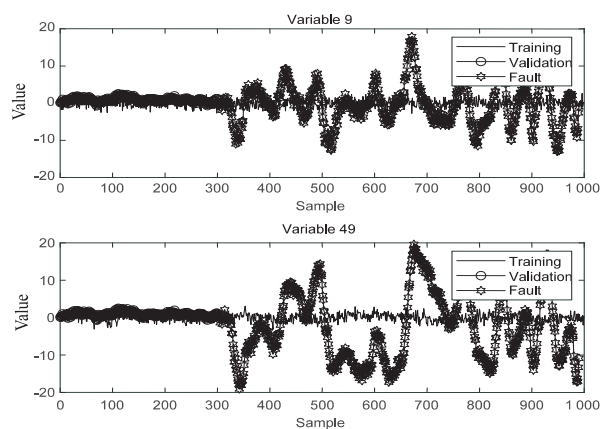


图 9 反应堆温度和反应堆冷却水流量

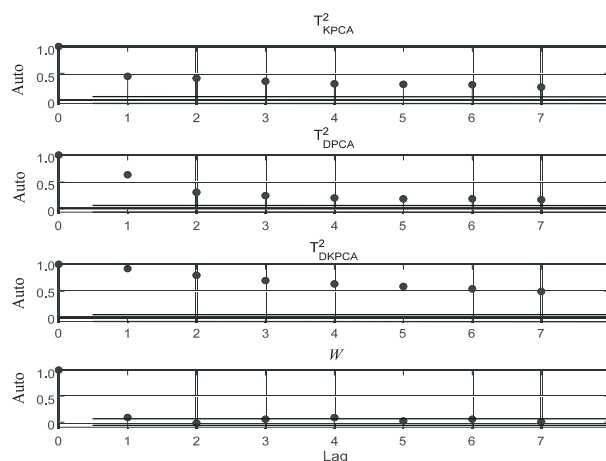


图 10 四种方法统计量的自相关

4 结束语

提出了一种基于去主元相关性的 DKPCA 动态非线性过程故障检测与诊断方法。该方法相对传统的 DKPCA 方法有两点改进,一是通过缺失值处理技术预测与原始矩阵保留相同特征的矩阵,二是通过真实得分与预估得分作差消除相同特征并在差异空间构造新的统计量。该方法进一步降低了动态性的影响,使得统计量的自相关性显著降低。另外,给出了基于变量累计贡献率的故障诊断方法。通过数值例子和 TE 过

程验证该方法在故障检测与诊断中具有更好效果,对于动态非线性监控过程具有一定的指导意义。

由于该方法在计算过程中多次使用核函数,因而出现检测过程运行时间长,占用内存空间大的问题,接下来将考虑如何优化算法的问题,提高算法的运行效率。

参考文献:

- [1] 牛玉广,王世林,林忠伟,等.基于多元统计过程监控的锅炉过程故障检测[J].动力工程学报,2017,37(10):829-836.
- [2] HUANG J, YAN X. Quality-driven principal component analysis combined with kernel least squares for multivariate statistical process monitoring[J]. IEEE Transactions on Control Systems Technology, 2019, 27(6): 2688-2695.
- [3] BASHA N, NOUNOU M, NOUNOU H. Multivariate fault detection and classification using interval principal component analysis[J]. Journal of Computational Science, 2018, 27(7): 1-9.
- [4] GUO J, WANG X, LI Y, et al. Fault detection based on weighted difference principal component analysis[J]. Journal of Chemometrics, 2017, 31(11): e2926.
- [5] 郭金玉,李涛,李元.基于核主成分空间支持向量机的过程监视方法[J].山东科技大学学报:自然科学版,2022,41(1):114-122.
- [6] KU W, STORER R H, GEORGAKIS C. Disturbance detection and isolation by dynamic principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1995, 30(1): 179-196.
- [7] RATO T J, REIS M S. Fault detection in the Tennessee Eastman benchmark process using dynamic principal components analysis based on decorrelated residuals (DPCA-DR)[J]. Chemometrics and Intelligent Laboratory Systems, 2013, 125(6): 101-108.
- [8] RATO T J, REIS M S. Advantage of using decorrelated residuals in dynamic principal component analysis for monitoring large-scale systems[J]. Industrial & Engineering Chemistry Research, 2013, 52(38): 13685-13698.
- [9] SANG W C, LEE I B. Nonlinear dynamic process monitoring based on dynamic kernel PCA[J]. Chemical Engineering Science, 2004, 59(24): 5897-5908.
- [10] RATO T J, REIS M S. Statistical process control of multivariate systems with autocorrelation[J]. Computer Aided Chemical Engineering, 2011, 29(2): 497-501.
- [11] ARTEAGA F, FERRER A. Dealing with missing data in MSPC: several methods, different interpretations, some examples[J]. Journal of Chemometrics, 2002, 16(8-10): 408-418.
- [12] LIM S, KIM S J, PARK Y, et al. A deep learning-based time series model with missing value handling techniques to predict various types of liquid cargo traffic[J]. Expert Systems with Applications, 2021, 184: 115532.
- [13] GAO K, LI D D. Estimating fluorescence lifetimes using the expectation-maximisation algorithm[J]. Electronics Letters, 2018, 54(1): 14-16.
- [14] 井沛良,段宇,韩超,等.基于高斯混合模型和期望最大化算法的非高斯分布圆概率误差估计方法研究[J].兵工学报,2019,40(2):369-376.
- [15] KANG Y J, NOH Y J, LIM K. Development of a kernel density estimation with hybrid estimated bounded data[J]. Journal of Mechanical Science and Technology, 2018, 32(12): 5807-5815.
- [16] DENG X G, TIAN X M. Nonlinear process fault diagnosis method using kernel principal component analysis[J]. Journal of Shandong University, 2005, 35(3): 103-106.
- [17] RATO T J, REIS M S. Defining the structure of DPCA models and its impact on process monitoring and prediction activities[J]. Chemometrics & Intelligent Laboratory Systems, 2013, 125(6): 74-86.
- [18] 陈伟.基于参数优化粒子群算法的核主元分析及其在青霉素发酵过程监控的应用[D].广州:广东技术师范大学,2021.
- [19] 韦哲,张宇刚,石栋栋,等.基于网格搜索与交叉验证的支持向量机糖尿病并发症诊断预测[J].中国医学装备,2020,17(8):8-11.
- [20] BATHELT A, N. RICKER L, JELALI M. Revision of the Tennessee Eastman process model[J]. IFAC - PapersOnLine, 2015, 48(8): 309-314.