

# 基于稀疏动态主元分析的故障检测方法

段怡雍, 吴平, 高金凤

(浙江理工大学 机械与自动控制学院, 杭州 310018)

**摘要:** 文章将动态主元分析 (Dynamic Principal Component Analysis, DPCA) 和稀疏主元分析 (Sparse Principal Component Analysis, SPCA) 两种方法结合起来, 提出一种新的稀疏动态主元分析方法, 并将其用于工业过程的故障检测; 所提出的稀疏动态主元分析方法通过对过程数据的动态增广矩阵进行稀疏主元的求解, 获取稀疏的负荷向量, 该方法既考虑到了过程数据的动态特性, 又降低了过程数据的冗余度, 同时降低了计算负荷, 非常适合工业过程的实时故障检测; 此外, 还提出了一种前向选择算法, 用于确定稀疏主元中的非零负荷数目; 最后, 将所提出方法应用于数值例子和田纳西-伊斯曼过程, 并将与主元分析、动态主元分析和稀疏主元分析等 3 种方法相比较, 表明所提方法可以获得更好的故障检测效果。

**关键词:** 主元分析; 动态主元分析; 稀疏动态主元分析; 非零负荷; 故障检测

## Fault Detection Based on Sparse Dynamic Principal Component Analysis

Duan Yiyong, Wu Ping, Gao Jinfeng

(Faculty of Mechanical Engineering & Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China)

**Abstract:** In this paper, a new sparse dynamic principal component analysis (SDPCA) technique is proposed, which combines two popular monitoring methods, dynamic principal component analysis (DPCA) and sparse principal component analysis (SPCA). The proposed SDPCA is used for fault detection for industrial process. In the proposed SDPCA method, the sparse loading vectors are derived by solving an optimization problem through the dynamic augmented matrix of process data. SDPCA technique not only considers the temporal correlation of process data, but also reduces redundancy of the process data, meantime reduces the computation load. Moreover, we will discuss a new forward selection algorithm for determining the number of non-zero loadings. The proposed SDPCA method is assessed through a numerical example and Tennessee Eastman benchmark process. Results show that the SDPCA based fault detection method could obtain a better performance compared with PCA, DPCA and SPCA based methods.

**Keywords:** principal component analysis; sparse principal component analysis; dynamic principal component analysis; non-zero loading; fault detection

## 0 引言

现代工业过程正朝着智能化的方向发展, 为了保证产品的质量、避免生产安全事故的发生, 实时高效的故障检测系统在工业过程中的作用变得越来越重要。主元分析 (Principal Component Analysis, PCA) 作为一种简单实用的多元统计过程监控方法, 能够有效地剔除冗余信息, 只需要正常工况下的历史数据即可建立模型, 在工业过程的故障检测和诊断中得了广泛的应用<sup>[1-3]</sup>。考虑到过程变量本身的时序自相关性, Ku 等<sup>[4]</sup>提出了动态主元分析方法 (Dynamic Principal Component Analysis, DPCA), 通过用带有时间滞后特性的变量构造出动态数据矩阵后, 利用 PCA 方法提取相关主元, 提高了故障检测的精度。

但是通过 PCA 方法得到的负荷向量中的元素通常是非零的, 不仅不利于特征的提取和解释, 计算量也非常大。为了获取稀疏负荷向量, 稀疏主元分析方法 (Sparse Principal Component Analysis, SPCA) 近年来得到了极大的发展, 并在生物学、医学等各个领域得到了广泛的应用<sup>[5-6]</sup>。Jolliffe<sup>[7]</sup>提出了 SCoTLASS 算法来获得稀疏主元。Zou 等<sup>[8]</sup>提出利用 LASSO 惩罚项来获得稀疏主元。Leng 等<sup>[9]</sup>提出了简单自适应主元分析, 用自适应的 LASSO 惩罚项取代传统的 Elastic Net。Kang 等<sup>[10]</sup>进一步的提出了自适应稀疏主元分析 (ASPCA)。彭必灿等<sup>[11]</sup>, 刘洋等<sup>[12]</sup>, Gajjar 等<sup>[13]</sup>采用 SPCA 实现了工业过程的故障检测及诊断。Gajjar 等<sup>[13]</sup>并提出一种前向选择方法来确定稀疏主元的非零负荷数目。

本文结合 DPCA 和 SPCA 两者的特点, 提出了一种基于稀疏动态主元分析法 (Sparse Dynamic Principal Component Analysis, SDPCA) 的故障检测方法。该方法先通过叠加时间滞后变量的方式建立测量数据的动态增广矩阵, 然后再通过 Lasso 约束函数获取增广矩阵的稀疏主元。此外,

收稿日期: 2018-10-09; 修回日期: 2018-10-04。

基金项目: 浙江理工大学科研启动基金 (14022086-Y)。

作者简介: 段怡雍 (1994-), 男, 浙江衢州人, 硕士生, 主要从事过程监测方向的研究。

高金凤 (1978-), 女, 安徽合肥人, 博士, 教授, 主要从事网络控制系统、多智能体系统方向的研究。

本文还提出了一种新的稀疏主元非零负荷数目的确定方法, 该方法考虑到了稀疏主元之间的关联性, 对文献 [13] 中提出的前向选择法进行了改进, 在保证累积贡献率的情况下, 进一步提高了稀疏度。最后, 通过数值仿真和田纳西伊斯曼过程来验证所提方法的性能, 并与 PCA、DPCA 和 SPCA 方法进行对比。

## 1 基于主元分析的故障检测

### 1.1 主元分析

给定一个具有  $n$  个观测值和  $m$  个过程变量的数据集  $\mathbf{X} \in \mathbf{R}^{n \times m}$ , 对  $\mathbf{X}$  进行特征值分解, 可得:

$$\text{cov}(\mathbf{X}) = \frac{\mathbf{X}^T \mathbf{X}}{n-1} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \quad (1)$$

式中,  $\mathbf{\Lambda} \in \mathbf{R}^{m \times m}$  为特征值的对角矩阵, 且对角数值沿对角线递减。  $\mathbf{V} \in \mathbf{R}^{m \times m}$  为酉矩阵, 其列向量为主元的负荷向量。

选取  $\mathbf{V}$  的前  $k$  列得到一个新的负荷矩阵  $\mathbf{P} \in \mathbf{R}^{m \times k}$ , 可得得分矩阵:

$$\mathbf{T} = \mathbf{X} \mathbf{P} \quad (2)$$

将  $\mathbf{T}$  投影返回  $m \times n$  维的观测空间可以得到一个对应主元所包含信息的新矩阵  $\hat{\mathbf{X}}$ , 做  $\mathbf{X}$  与  $\hat{\mathbf{X}}$  的差可以得到残差矩阵  $\mathbf{E}$ :

$$\hat{\mathbf{X}} = \mathbf{T} \mathbf{P}^T \quad (3)$$

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} \quad (4)$$

主元数目的求取可以通过设定贡献度百分比 (Cumulative Percent Variance, CPV) 来完成,

$$\text{CPV}(k) = \frac{\sum_{i=1}^k \Lambda_i}{\sum_{i=1}^m \Lambda_i} \times 100\% \geqslant CL. \quad (5)$$

式中,  $CL$  为设定值, 满足上式成立的  $k$  的最小值即为满足当前贡献率阈值的最优解。

### 1.2 主元分析的故障检测

Hotelling  $T^2$  和  $Q$  统计量常被用于基于多变量统计法的故障检测。  $T^2$  统计量由计算主元的得分向量在空间中的马氏距离获得, 而  $Q$  统计量则是残差向量在残差空间中的欧式距离,

$$T_i^2 = \mathbf{x}_i^T \mathbf{P} \mathbf{\Lambda}^{-1} \mathbf{P}^T \mathbf{x}_i, \quad (6)$$

$$Q_i = \mathbf{e}_i^T \mathbf{e}_i \quad (7)$$

式中,  $\mathbf{x}_i$  为数据集在  $i$  时刻的  $m$  维观测向量,  $\mathbf{e}_i$  为该观测向量的残差向量。

通常,  $T^2$  统计量的阈值用自由度为  $k$  和  $n-k$  的  $F$  分布计算,  $Q$  统计量的阈值则可用  $\chi^2$  分布计算出来:

$$T_\alpha^2 = \frac{k(n-1)(n+1)}{n(n-k)} F_\alpha(k, n-k). \quad (8)$$

$$Q_\alpha = g \chi_{h, \alpha}^2; \quad g = \frac{\bar{v}}{2\bar{\mu}}, \quad h = \frac{2\bar{\mu}}{\bar{v}}$$

式中,  $\alpha$  为给定的显著性水平,  $\bar{\mu}$  和  $\bar{v}$  分别为数据集  $Q$  统计量的估计均值和方差。一旦计算得到的  $T_i^2$  或  $Q_i$  统计量高

于相应的阈值, 则认为过程有故障发生。

## 2 基于稀疏动态主元分析的故障检测

### 2.1 动态主元分析

工业过程一般都具有较强的动态特性, 因此过程数据具有强烈的序列相关性。为了解决工业过程的动态特性, DPCA 通过扩展观测矩阵, 可以消除数据的自相关性, 从而提高故障检测的精度<sup>[14-15]</sup>。

通过在每个观测向量后叠加  $l$  个滞后的观测向量来获得增广矩阵, 构建方法如下:

$$\mathbf{X}(l) = \begin{bmatrix} \mathbf{x}_t^T & \mathbf{x}_{t-1}^T & \cdots & \mathbf{x}_{t-l}^T \\ \mathbf{x}_{t-1}^T & \mathbf{x}_{t-2}^T & \cdots & \mathbf{x}_{t-l-1}^T \\ \cdots & \cdots & \ddots & \cdots \\ \mathbf{x}_{t+1-n}^T & \mathbf{x}_{t+1-n+1}^T & \cdots & \mathbf{x}_{t-n}^T \end{bmatrix} \quad (10)$$

式中是数据集在  $t$  时刻的  $m$  维观测值,  $n$  是总的样本数目。然后将该增广矩阵代替原观测矩阵来进行主元分析。一般情况下, 在工业过程控制中序列的滞后参数选择 1 或 2<sup>[16]</sup>。

DPCA 通过增广矩阵可以产生更多的信息关联, 比静态 PCA 更适用于动态工业过程的故障检测。但通过 DPCA 得到的主元数目过多, 使得变量之间的关系更不容易解释, 并导致计算量过大。

### 2.2 基于稀疏动态主元分析的故障检测

本文在 DPCA 的基础上, 提出稀疏动态主元分析法 (SDPCA), 通过结合了 DPCA 和 SPCA 方法的优点, 以提高故障检测的精度, 并减少实时计算量, 其主要步骤包括:

- 1) 获得动态增广矩阵;
- 2) 确定主元中的非零负荷数目;
- 3) 对增广矩阵进行稀疏主元求解。

首先, 通过公式 (10) 获得设计矩阵, 这里参数可采用交叉验证进行选取。然后再确定非零负荷的数目, 本文提出了一种新的非零负荷数目的确定方法, 并在下一小节中给出具体的过程。最后, 在确定的非零负荷数目下, 将降维问题转化为回归最优化问题, 对其进行稀疏求解。这里, 我们采用类似于 Zou 等<sup>[8]</sup> 提出的 SPCA 中的优化算法来获取增广矩阵的稀疏主元, 即在 DPCA 模型上增加 LASSO 惩罚项:

$$\begin{aligned} (\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \arg \min_{\mathbf{A}, \mathbf{B}} \sum_{i=1}^n \|\mathbf{x}(l)_i - \mathbf{A} \mathbf{B}^T \mathbf{x}(l)_i\|^2 + \\ \lambda \sum_{j=1}^k \|\boldsymbol{\beta}_j\|^2 + \lambda_{1,j} \sum_{j=1}^k \|\boldsymbol{\beta}_j\|_1, \\ \text{Subject to } \mathbf{A}^T \mathbf{A} = \mathbf{I}_{k=k}. \end{aligned} \quad (11)$$

式中,  $\mathbf{A}$  和  $\mathbf{B}$  是  $(l+1) m \times k$  的矩阵,  $\hat{\mathbf{A}}$  和  $\hat{\mathbf{B}}$  是使右式取得最小值的  $\mathbf{A}$  和  $\mathbf{B}$  的解,  $\boldsymbol{\beta}_j$  是  $\mathbf{B}$  的第  $j$  列向量, 同时也是稀疏后的主元。  $\lambda$  以及  $\lambda_{1,j}$  是惩罚项的系数,  $\lambda > 0$ 。调整后的解释方差可以由以下公式取得:

$$\text{SA} = \left[ \frac{\text{diag}(qr(\mathbf{X}(l)\hat{\mathbf{P}}))}{n} \right]^{-2} \quad (12)$$

式中,  $\mathbf{SA}$  为解释方差的对角矩阵,  $\hat{\mathbf{P}}$  为稀疏后的负荷向量矩阵,  $qr(\cdot)$  表示对括号内的矩阵进行 QR 分解后得到上三角  $\mathbf{R}$  矩阵。此时, 该模型的贡献度可计算如下:

$$CPV = \frac{\sum_{i=1}^k SA_i}{\sum_{i=1}^m \Delta_i} \times 100\% \quad (13)$$

由于稀疏后的负荷向量不一定是正交的,  $T^2$  统计量及其控制限的计算如下:

$$T_i^2 = t_i \gamma^{-1} t_i^T \quad (14)$$

$$T_a^2 = \frac{k(n-1)}{n-k} F_{\alpha}(k, n-k) \quad (15)$$

其中,  $\gamma$  是稀疏得分向量的协方差矩阵。而  $Q$  统计量的计算方式则与 PCA 模型一致, 可采用公式 (7) 和公式 (9)。

### 2.3 非零负荷数目的确定

如何选择每个主元的非零负荷数目一直是稀疏主元分析的难点。过多的非零负荷会造成计算上的繁琐, 所以需要尽可能地减少数目, 但同时又要保证剩下的非零负荷拥有足够多的相关信息。

Jolliffe 指出当主元的解释方差相近时, 约束标准的选择对模型的最终效果影响不大<sup>[17]</sup>。在此基础上, Gajjar 等<sup>[18]</sup> 针对 SPCA 模型和对应的 PCA 模型的解释方差之间的联系, 提出了一种非零负荷数目的前向选择算法, 具体步骤如下:

- 1) 通过公式 (1) 求解得特征值矩阵。
- 2) 计算可获得稀疏主元的公式 (11), 其中主元个数为  $k$ , 最后一个主元的非零负荷数目为  $q$  ( $k$  和  $q$  的初始值为 1)。
- 3) 如果  $\eta_1 \geq 90\%$ , 则确定  $q$  为最后一个负荷向量的非零负荷数目; 否则  $q=q+1$ , 返回步骤 2)。
- 4) 当  $k$  值与传统 PCA (DPCA) 的主元个数相同时, 或是稀疏后的主元贡献度超过一定阈值时, 停止计算。否则  $k=k+1$ ,  $q=1$ , 返回步骤 2)。

其中,  $\eta_1$  为当前 SPCA 主元的解释方差与所对应特征值的比值, 计算公式为:

$$\eta_1 = \frac{SA_k}{\Delta_k} \quad (16)$$

但需要注意的是, 在 Gajjar 等的方法中, 每个稀疏主元的非零负荷数目的选择只和它所对应的一个特征值有密切关系, 而忽略了与它之前的特征值的关系, 可能导致解释方差过低。为此, 本文将此方法进行了改进, 提出了一种新的基于前向选择的非零负荷数目算法如下:

- 1) 通过公式 (1) 求解得特征值矩阵。
- 2) 计算可获得稀疏主元的公式 (11), 其中主元个数为  $k$ , 最后一个主元的非零负荷数目为  $q$  ( $k$  和  $q$  的初始值为 1)。
- 3) 如果  $\eta_1 \geq 80\%$  且  $\eta_2 \geq 90\%$ , 则确定  $q$  为最后一个负荷向量的非零负荷数目; 否则  $q=q+1$ , 返回步骤 2)。
- 4) 当稀疏后的主元贡献度超过 85% 时, 停止计算。否则  $k=k+1$ ,  $q=1$ , 返回步骤 2)。

其中,  $\eta_2$  为当前  $k$  个 SPCA (或 SDPCA) 的方差的和与

所对应的  $k$  特征值的和的比值, 计算公式为:

$$\eta_2 = \frac{\sum_{i=1}^k SA_i}{\sum_{i=1}^k \Delta_i} \times 100\% \quad (17)$$

在本文提出的算法中, 主要依靠来确定非零负荷的数目, 因此不仅包含了之前的主元的解释方差信息, 也将参数考虑进来, 以防止当前的解释方差过低, 加强了每个稀疏主元的非零向量数目与特征值之间的联系。此外, 还可直接地通过调整贡献度的大小来确定主元的数目。

## 3 仿真

### 3.1 数值仿真

为了表明方法的有效性, 采用文献 [4] 的数值例子如下:

$$z(k) = \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix} z(k-1) + \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} u(k-1) \quad (18)$$

$$y(k) = z(k) + v(k)$$

$$u(k) = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix} u(k-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix} w(k-1) \quad (19)$$

其中: 输入变量  $w$  是均值为零, 方差为 1 的白噪声,  $v$  是均值为零, 方差为 0.1 的白噪声。输入变量  $u$  和输出变量  $y$ , 均为可测变量。采集正常情况下的 100 个样本, 用于建模。为了模拟故障, 在非正常情况下, 在第 10 个样本上引入故障。其中, 故障 1 是,  $w_1$  变量添加一个单位阶跃。故障 2 是, 将阶跃的幅值增加至 3。

为了减少随机变量的影响, 本次实验一共生成了 1000 组数据集。在设置 PCA 和 SPCA 的主元个数为 3, DPCA 和 SDPCA 的主元个数为 4 时, DPCA 和 SDPCA 的动态迟滞设置为 1。此时 PCA 和 DPCA 模型的贡献率可达到 98% 以上, SPCA 的贡献率在 92% 至 96% 之间, SDPCA 的贡献率在 89% 至 94% 之间。4 种方法的置信度均设置为 99%。表 1 为分别用 PCA、SPCA、DPCA 和 SDPCA 4 种方法所得到的故障检测率 (Fault Detection Rate, FDR) 平均值。

表 1 各种故障检测方法的检测率 %

检测方法	故障 1( $T^2$ )	故障 1(Q)	故障 2( $T^2$ )	故障 2(Q)
PCA	3.80	2.18	24.45	4.41
SPCA	3.85	2.38	25.03	5.78
DPCA	3.64	2.90	3.72	3.41
SDPCA	3.72	3.41	25.17	11.47

从表 1 中可以看出, 4 种故障检测方法在  $T^2$  统计量下的检测结果差别并不大, 但 SDPCA 方法用 Q 统计量测得的检测率明显高于其它 3 种方法。并且从 SPCA 与 PCA、SDPCA 与 DPCA 的对比可以看出, 经过稀疏处理之后的检测效果会有小的提升。

### 3.2 TE 过程仿真

田纳西-伊斯曼过程是一种模仿真实化学工业现场的标准控制过程, 被广泛的用作控制、优化及故障诊断的仿真对象。该过程主要包括了 5 个操作单元, 分别是反应器、冷凝器、气液分离器、循环压缩机、汽提塔。过程数据的工况被人为设定 21 种故障情况, 每种故障情况以 3 分钟为采样间隔分别采集了 960 组采样数据, 每组数据含有 53 个特征变量, 每组数据的在第 161 个采样时刻引入不同的故障。本文选取了第 1 至 22 个测量变量及第 42 至第 52 个控制变量, 共 33 个特征变量用于故障检测。

表 2 TE 过程在  $T^2$  统计量下的检测率 %

故障号	PCA	SPCA	DPCA	SDPCA
1	<b>99.13</b>	<b>99.13</b>	<b>99.13</b>	<b>99.13</b>
2	98.38	98.38	<b>98.50</b>	<b>98.50</b>
3	0.88	1.50	1.25	<b>5.63</b>
4	<b>20.88</b>	7.25	4.50	4.88
5	24.25	23.75	24.00	28.13
6	<b>99.13</b>	<b>99.13</b>	99.00	99.00
7	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
8	96.88	96.38	<b>97.00</b>	96.88
9	1.75	1.88	0.50	<b>5.50</b>
10	29.88	26.63	25.88	<b>39.38</b>
11	<b>40.63</b>	19.88	19.25	17.88
12	98.38	97.50	99.00	<b>99.25</b>
13	93.63	93.00	<b>94.00</b>	93.25
14	99.25	81.00	<b>99.88</b>	<b>99.88</b>
15	1.38	1.38	0.88	<b>6.88</b>
16	13.50	10.63	10.25	<b>21.25</b>
17	76.38	72.63	75.75	<b>79.50</b>
18	89.25	89.25	89.13	<b>90.38</b>
19	11.00	13.63	17.50	<b>23.50</b>
20	31.75	34.50	32.63	<b>36.50</b>
21	39.25	26.88	43.25	<b>44.38</b>

这里, 将 SDPCA 方法的动态迟滞设置为 1。经过本文提出的非零负荷数目的选择方法, 数据的 SDPCA 模型一共产生了 30 个稀疏主元。每个主元的非零负荷的个数分别为 24, 23, 9, 4, 5, 4, 2, 2, 2, 2, 4, 2, 2, 3, 2, 2, 2, 1, 1, 2, 1, 1, 1, 2, 1, 4, 2, 2, 2 和 2, 贡献率为 85.48%。DPCA 方法同样选取 85% 的贡献度, 其模型含有 24 个主元。SPCA 与 PCA 方法的检测数据主元数目均为 14。在置信度设为 99% 情况下,  $T^2$  统计量检测的结果如表 2 所示, Q 统计量的检测结果如表 3 所示。表 2 和表 3 中, 每个故障的最优检测率用加粗字体表示。

通过表 2 可以看出, 4 种方法在  $T^2$  统计检测结果中, SDPCA 在其中的 16 个故障数据下的检测效果都是最优的。特别是故障 3、故障 9 和故障 15 这几个传统方法难以检测的故障, 在其它 3 种方法下的检测率均低于 2%, 而通过

SDPCA 方法在  $T^2$  统计量下可以得到高于 5% 的检测率。通过表 3 可以看出, 在 Q 统计检测结果中, SDPCA 方法在 11 个故障下的检测率都是最高的。尤其在故障 10 和故障 16 的检测中, SDPCA 在两种统计量下的检测效果均明显优于其它 3 种方法。图 1~4 为故障 10 在 4 种方法下的 Q 统计量检测结果。

表 3 TE 过程在 Q 统计量下的检测率 %

故障号	PCA	SPCA	DPCA	SDPCA
1	<b>99.88</b>	<b>99.88</b>	99.75	99.75
2	96.00	97.63	96.13	98.00
3	3.38	<b>4.25</b>	4.00	2.25
4	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
5	24.13	27.13	27.50	30.25
6	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
7	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
8	86.13	96.63	89.38	<b>96.88</b>
9	2.38	2.38	<b>3.88</b>	1.88
10	29.13	45.75	36.75	<b>51.63</b>
11	76.13	79.00	87.38	<b>87.75</b>
12	90.75	98.88	91.25	<b>99.00</b>
13	<b>95.25</b>	95.00	95.13	94.88
14	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	98.88
15	3.88	<b>4.88</b>	4.13	4.63
16	32.25	39.25	37.63	<b>47.13</b>
17	95.88	95.13	<b>97.00</b>	96.25
18	<b>90.25</b>	89.75	<b>90.25</b>	89.50
19	17.75	10.38	<b>30.13</b>	13.63
20	51.63	49.50	55.00	<b>55.00</b>
21	<b>48.88</b>	<b>55.38</b>	48.00	52.63

此外, 稀疏算法可以有效减少非零负荷的数目, TE 过程的数据经过稀疏动态处理之后, SDPCA 模型的所有主元中所包含的非零负荷数目为 116, 而 DPCA 模型下共包含了 792 个非零负荷。非零负荷的大量减少, 使得 SDPCA 相比 DPCA 具有更高的实时计算效率。表 4 列出了同一台服务器使用 MATLAB 软件分别用 4 种方法采用  $T^2$  统计量处理 TE 过程故障数据所用的时间。从表中可以看出经过数据动态处理后的 SDPCA 和 DPCA 两种方法所花费的时间更长一些, 但经过稀疏处理后的 SDPCA 方法比 DPCA 方法的检测时间更短一些。

表 4 TE 过程的检测时间

方法	PCA	SPCA	DPCA	SDPCA
时长/s	0.37	0.27	0.51	0.46

## 4 结束语

结合动态主元分析和稀疏主元分析的特点, 本文提出

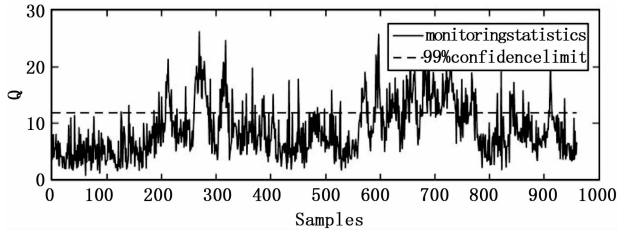


图 1 TE 过程故障 10 在 PCA 下的检测结果

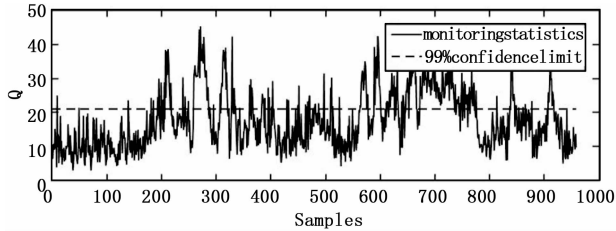


图 2 TE 过程故障 10 在 DPCA 下的检测结果

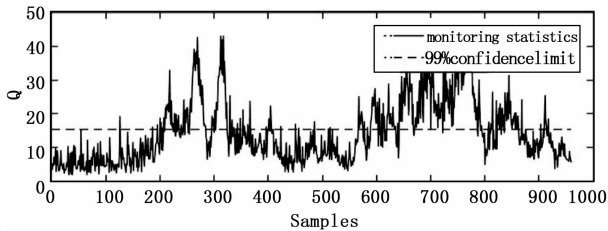


图 3 TE 过程故障 10 在 SPCA 下的检测结果

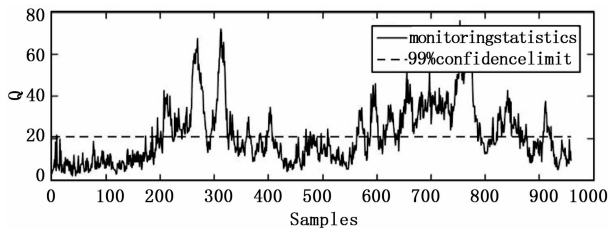


图 4 TE 过程故障 10 在 SDPCA 下的检测结果

了一种基于稀疏动态主元分析的故障检测方法,并进一步改进了非零负荷数目的前向选择算法。该方法不仅考虑到了实际工业数据的动态特性,同时也大量地减少了非零负荷的数量,从而提高了计算效率,并增加了特征的可解释性。通过数值和 TE 过程的仿真结果表明,所提方法能够获得良好的动态过程的故障检测性能。未来的研究将进一步研究非零负荷数目的优化选择,以及将动态稀疏方法推广到其它的多变量统计方法中。

#### 参考文献:

- [1] Jolliffe I T, Cadima J. Principal component analysis: a review and recent developments [J]. *Philos Trans A Math Phys Eng Sci*, 2016, 374 (2065): 20150202.
- [2] Ge Z, Song Z, Ding S X, et al. Data mining and analytics in the

process industry: the role of machine learning [J]. *IEEE Access*, 2017, 5 (99): 20590–20616.

- [3] Zeng J, Kruger U, Geluk J, et al. Detecting abnormal situations using the Kullback–Leibler divergence [J]. *Automatica*, 2014, 50 (11): 2777–2786.
- [4] 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.
- [5] Feng C M, Gao Y L, Liu J X, et al. A simple review of sparse principal components analysis [A]. *International Conference on Intelligent Computing [C]*. Springer International Publishing, 2016: 374–383.
- [6] Trendafilov N T. From simple structure to sparse components: a review [M]. Kluwer Academic Publishers, 2014, 29 (3–4): 431–454.
- [7] Jolliffe I T, Trendafilov N T, Uddin M. A Modified principal component technique based on the LASSO [J]. *Journal of Computational & Graphical Statistics*, 2003, 12 (3): 531–547.
- [8] Zou H, Hastie T, Tibshirani R. Sparse principal component analysis [J]. *Journal of Computational & Graphical Statistics*, 2006, 15 (2): 265–286.
- [9] Leng C, Wang H. On General Adaptive Sparse Principal Component Analysis [J]. *Journal of Computational & Graphical Statistics*, 2009, 18 (1): 201–215.
- [10] Liu K L, et al. Adaptive partitioning PCA model for improving fault detection and isolation [J]. *中国化学工程学报 (英文版)*, 2015, 23 (6): 981–991.
- [11] 彭必灿, 张正道. 基于稀疏主元分析的过程监控研究 [J]. *计算机工程与应用*, 2014, 50 (18): 240–245.
- [12] 刘洋, 张国山. 基于敏感稀疏主元分析的化工过程监测与故障诊断 [J]. *控制与决策*, 2016, 31 (7): 1213–1218.
- [13] Gajjar, S, Kulahci, M, Palazoglu, A. Real-time fault detection and diagnosis using sparse principal component analysis [J]. *Journal of Process Control*, 2017.
- [14] 石怀涛, 刘建昌, 丁晓迪, 等. 基于混合动态主元分析的故障检测方法 [J]. *控制工程*, 2012, 19 (1): 152–155.
- [15] Huang J, Yan X. Dynamic process fault detection and diagnosis based on dynamic principal component analysis, dynamic independent component analysis and Bayesian inference [J]. *Chemometrics & Intelligent Laboratory Systems*, 2015, 148: 115–127.
- [16] Ku W, Storer R H, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis [J]. *Chem. intell. lab. syst*, 1995, 30 (1): 179–196.
- [17] Jolliffe I T. Rotation of principal components: choice of normalization constraints [J]. *Journal of Applied Statistics*, 2012, 22 (1): 29–35.