

# 控制与决策

Control and Decision

## 基于递归稀疏主成分分析的工业过程在线故障监测和诊断

刘金平, 王杰, 刘先锋, 唐朝晖, 马天雨, 肖文辉

引用本文:

刘金平, 王杰, 刘先锋, 等. 基于递归稀疏主成分分析的工业过程在线故障监测和诊断[J]. *控制与决策*, 2020, 35(8): 2006–2012.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2018.1661>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### 基于正交信号修正与高效偏最小二乘的质量相关故障检测方法

Quality-related fault detection method based on orthogonal signal correction and efficient PLS  
*控制与决策*. 2020, 35(5): 1167–1174 <https://doi.org/10.13195/j.kzyjc.2018.0708>

### 基于多块相对变换独立主元分析的故障诊断方法

Fault diagnosis approach based on relative transformation ICA of multiblock  
*控制与决策*. 2018, 33(11): 2009–2014 <https://doi.org/10.13195/j.kzyjc.2017.0869>

### 正交信号校正的自回归模型及其在动态过程监测中的应用

Orthogonal signal correction based auto-regression model with application to dynamic process monitoring  
*控制与决策*. 2016, 31(8): 1505–1508 <https://doi.org/10.13195/j.kzyjc.2015.0750>

### 基于敏感稀疏主元分析的化工过程监测与故障诊断

Chemical process monitoring and fault diagnosis based on sensitive sparse principal component analysis  
*控制与决策*. 2016, 31(7): 1213–1218 <https://doi.org/10.13195/j.kzyjc.2015.0629>

### 基于主元和判别集成分析的模拟电路故障诊断

Fault diagnosis of analog circuit by integrating principal component analysis with linear discriminant analysis  
*控制与决策*. 2015(7): 1321–1324 <https://doi.org/10.13195/j.kzyjc.2014.0774>

# 基于递归稀疏主成分分析的工业过程在线故障监测和诊断

刘金平<sup>1†</sup>, 王 杰<sup>1</sup>, 刘先锋<sup>1</sup>, 唐朝晖<sup>3</sup>, 马天雨<sup>1,2</sup>, 肖文辉<sup>3</sup>

(1. 湖南师范大学 智能计算与语言信息处理湖南省重点实验室, 长沙 410081;  
2. 湖南师范大学 物理与电子学院, 长沙 410081; 3. 中南大学 自动化学院, 长沙 410083)

**摘要:** 提出一种基于递归稀疏主成分分析(recursive sparse principal component analysis, RSPCA)的工业过程故障监测与诊断方法,可用于时变工业过程的自适应故障监测与诊断. 通过引入弹性回归网,将主成分问题转化为Lasso与Ridge结合的凸优化问题,采用秩-1矩阵修正对协方差矩阵进行递归分解,递归更新稀疏载荷矩阵和监测统计量的过程控制限,以实现连续工业过程长时间自适应故障监测,对检测出来的故障通过贡献图法实现对故障的诊断. 在田纳西-伊斯曼(TE)过程进行实验验证,结果表明,与传统的故障监测方法相比,所提出的方法有效降低了故障漏检率和误报率,且时间复杂度低,确保了故障监测的灵敏度和实时性.

**关键词:** 递归稀疏主成分分析; 工业过程故障监测; 弹性回归网; 田纳西-伊斯曼过程

中图分类号: TP273 文献标志码: A

## Online fault monitoring and diagnosis using recursive sparse principal component analysis

LIU Jin-ping<sup>1†</sup>, WANG Jie<sup>1</sup>, LIU Xian-feng<sup>1</sup>, TANG Zhao-hui<sup>3</sup>, MA Tian-yu<sup>1,2</sup>, XIAO Wen-hui<sup>3</sup>

(1. Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha 410081, China; 2. School of Physics and Electronics, Hunan Normal University, Changsha 410081, China; 3. School of Automation, Central South University, Changsha 410083, China)

**Abstract:** A fault monitoring and diagnosis method for industrial processes based on recursive sparse principal component analysis (RSPCA) is proposed, which can be used for adaptive fault monitoring and diagnosis of time-varying industrial processes. By introducing the elastic regression net, the principal component analysis problem is changed to a convex optimization problem combining Lasso and Ridge regression. Successively, the covariance matrix is decomposed recursively by rank -1 matrix correction, which leads to the recursive updating of the sparse load matrix and the process control limit of monitoring statistics to realize the long-time adaptive fault monitoring of continuous industrial processes. With regard to the detected faults, the fault diagnosis is realized by using the contribution graph method. Experiments in Tennessee-Eastman (TE) process show that the proposed method effectively reduces the rate of failure miss detection and false alarm with low time complexity compared with traditional fault monitoring methods, which can ensure the sensitivity and real-time performance of the fault monitoring.

**Keywords:** recursive sparse principal component analysis; industrial process fault diagnosis; elastic regression network; Tennessee-Eastman process

## 0 引言

及时发现工业生产中可能出现的故障并对故障发生原因进行诊断,以便及时制定或调整过程操作策略,对于保证工业产品质量、降低能源消耗和实现安全生产均具有非常重要的现实意义<sup>[1]</sup>. 多变量统计过程监测(MSPM)方法在工业控制过程监测领域得到了广泛的应用<sup>[2]</sup>,最常见的MSPM方法包括主成分分

析(PCA)、偏最小二乘(PLS)<sup>[3]</sup>等.

PCA是一种代表性的MSPM方法,通过线性空间变换将高维数据空间投影到低维线性空间,得到少数几个不相关的主成分对原始空间进行描述<sup>[4]</sup>. 然而,基于PCA所获得的主元通常为非零值,当部分变量值相差不多时,容易造成共线问题,导致故障监测灵敏度差.

收稿日期: 2018-12-04; 修回日期: 2019-04-01.

基金项目: 国家自然科学基金项目(61971188, U1701261, 61771492); 湖南省自然科学基金项目(2018JJ3349); 图像信息处理与智能控制教育部重点实验室(华中科技大学)开放基金项目(IPIC2017-03); 湖南省研究生科研创新项目(CX2018B312).

†通讯作者. E-mail: ljp202518@163.com.

一些学者指出,PCA稀疏化是解决上述问题的一种有效途径。Zou等<sup>[5]</sup>结合Lasso算子和Ridge算子的弹性网(elastic net)惩罚项<sup>[6]</sup>提出了稀疏主成分分析算法(sparse PCA, SPCA)。SPCA算法可以使载荷向量稀疏化,减少主元和监控指标的计算量,从而更加有效地揭示数据内在结构<sup>[7]</sup>。

传统的MSPM方法基本上都假定过程工况是静态的。然而,因工业过程原料成分复杂多变、生产设备磨损老化、传感器漂移等多种因素的影响,工业过程状态随时会发生变化<sup>[8-10]</sup>。基于静态历史数据所建立的正常工况监测模型往往无法适应复杂工业过程的动态变化,特别是对于长时间连续运行的流程工业过程,传统的多元分析统计模型极易产生故障误报、漏报、反应不够迅速等问题<sup>[11]</sup>。

鉴于此,本文提出一种基于RSPCA的工业过程故障监测方法,可实现时变工况过程的自适应在线故障监测。在稀疏主元模型更新过程中,采用基于秩-1矩阵的修正<sup>[12]</sup>对协方差矩阵进行递归分解,从而递归更新稀疏主成分。TE过程故障监测与诊断实验验证了所提出方法的有效性和优越性。

## 1 RSPCA算法

本节首先简要回顾PCA和SPCA的基本原理,然后推导RSPCA算法,并分析算法的收敛性和时间复杂度。

### 1.1 PCA和SPCA

取一段观测数据集 $X \in R^{n \times m}$ ,标准化处理得到 $\bar{X}$ (下文将标准化后的 $\bar{X}$ 仍记为 $X$ ),有

$$X = D_\sigma^{-\frac{1}{2}} [X - E(X)]. \quad (1)$$

其中 $E(X)$ 和 $D_\sigma$ 分别为数据集 $X$ 对应的均值矩阵和方差矩阵。

假设经过PCA变换的新坐标系为 $W \in R^{n \times p}$ , $W$ 为酉矩阵,可以得到PCA的优化准则为

$$\begin{aligned} & \arg \max_W \text{tr}(W^T X^T X W); \\ & \text{s.t. } W^T W = I. \end{aligned} \quad (2)$$

由采用拉格朗日余项法求解上述优化问题可知, $W$ 实际上是相关系数矩阵 $X^T X$ 的 $p$ 个最大特征值对应特征向量所组成的矩阵,采用主成分贡献率<sup>[13]</sup>确定主元个数( $p$ 值)。

本文通过引入弹性回归网对主成分进行稀疏求解。首先通过Ridge惩罚项将主成分问题转换为回归问题,引入参数 $\alpha$ 和 $\beta$ ,假设 $\beta_j$ 为稀疏主成分空间的第 $j$ 个载荷向量, $X_i$ 为第 $i$ 个观测样本,其中 $\alpha, \beta \in$

$R^{n \times p}$ ,对于任意的 $\lambda_2 > 0$ ,有

$$\arg \max_{\alpha, \beta} \left\{ \|X_i \beta \alpha^T - X_i\|_2^2 + \lambda_2 \sum_{j=1}^p \|\beta_j\|_2^2 \right\}. \quad (3)$$

加入Lasso惩罚项将主成分的载荷矩阵稀疏化,最终得到SPCA的优化准则为

$$\arg \max_{\alpha, \beta} \left\{ \|X_i \beta \alpha^T - X_i\|_2^2 + \lambda_2 \sum_{j=1}^p \|\beta_j\|_2^2 + \lambda_{1,j} \sum_{j=1}^p \|\beta_j\|_1 \right\}. \quad (4)$$

### 1.2 RSPCA

考虑到新增样本数据时,历史样本矩阵及相关统计特征更新为

$$X_{k+1}^0 = \begin{bmatrix} X_k^0 \\ X_{n_{k+1}}^0 \end{bmatrix}, \quad (5)$$

$$m_{k+1} = \frac{N_k}{N_{k+1}} m_k + \frac{n_{k+1}}{N_{k+1}}, \quad (6)$$

$$\Sigma_{k+1}^2 = \frac{N_k - 1}{N_{k+1}} m_k + \frac{\|X_{n_{k+1}}^0 - l_{n_{k+1}} m_{k+1}^T\|_2^2}{N_{k+1}}. \quad (7)$$

其中: $X_k^0 \in R^{N_k \times m}$ 为 $k$ 次采样矩阵, $n_{k+1}$ 为第 $k+1$ 次采样数据块大小, $N_k$ 为前 $k$ 次采样数据块大小, $m_k \in R^{m \times 1}$ 为历史样本矩阵均值向量, $l_{n_{k+1}} = [1, 1, \dots, 1]^T \in R^{n_{k+1} \times 1}$ , $\Sigma_k \in R^{m \times m}$ 为标准差的对角阵。令 $\Delta m_{k+1} = m_{k+1} - m_k$ , $X_{k+1}^0$ 可以标准化为

$$X_{k+1} = \begin{bmatrix} X_k \Sigma_k \Sigma_{k+1}^{-1} - l_k \Delta m_{k+1}^T \Sigma_{k+1}^{-1} \\ X_{n_{k+1}} \end{bmatrix}. \quad (8)$$

格拉姆矩阵 $G_k = X_k^T X_k$ ,当更新为小样本事件时,认为 $\Sigma_k \Sigma_{k+1}^{-1} \approx 1$ ,相应的 $G_{k+1}$ 更新为

$$G_{k+1} = G_k + N_k \Sigma_k^T \Delta m_{k+1} \Delta m_{k+1}^T \Sigma_k^{-1} + G_{n_{k+1}}. \quad (9)$$

SPCA计算分为两个相互独立阶段进行。对于结合Ridge惩罚项的回归模型(称为 $L_2$ 模型),采用最小二乘法求解。为使 $L_2$ 模型取得最小值,固定参数 $\alpha$ ,可得

$$\beta_k^* = (G_k + \lambda I)^{-1} G_k \alpha_k. \quad (10)$$

令 $\zeta = N_k \Sigma_k^T \Delta m_{k+1} \Delta m_{k+1}^T \Sigma_k^{-1}$ ,当 $N_k \gg n_{k+1}$ 时, $\beta_{k+1}^*$ 递推更新为

$$\begin{aligned} \beta_{k+1}^* &= \beta_k^* + (G_{k+1}^{-1} + \lambda I) \times \\ & G_{n_{k+1}} \left[ \alpha_{k+1} - \left( \frac{\zeta}{G_{n_{k+1}}} + I \right) \beta_k^* \right]. \end{aligned} \quad (11)$$

对于参数 $\alpha$ 固定为主成分的载荷矩阵 $W_k$ ,当样本更新为 $X_{k+1}$ 时,本文采用秩-1修正算法递推更新载荷矩阵 $W_{k+1}$ ,引入如下定理<sup>[12]</sup>。

**定理1** 如果 $H = \text{diag}(h_1, h_2, \dots, h_n)$ ,则 $V^T (H + \gamma \eta \eta^T) V = \text{diag}(s_1, s_2, \dots, s_n)$ 为 $H + \gamma \eta \eta^T$ 的Schur

分解,其中  $H \in R^{n \times n}, \gamma \neq 0, \eta \in R^n, V$  为酉矩阵,特征值和特征向量分别如下:

1) 函数  $f(\varsigma) = 1 + \gamma\eta^T(H - \varsigma I)^{-1}$  的  $n$  个零点为  $\eta_1, \eta_2, \dots, \eta_n$ .

2) 向量  $V_i = \|(H - \varsigma_i I \eta)^{-1}\|_2^{-1} (H - \varsigma_i I \eta)^{-1}$ , 其中  $i = 1, 2, \dots, n$ .

相应的格拉姆矩阵  $G_k$  迭代分解为

$$G_{k+1} = W_{k+1}^T D_{k+1}^2 W_{k+1}. \quad (12)$$

最后,利用 Lasso 模型(称为  $L_1$  模型,下同)使回归系数缩减为 0,得到  $p$  个 Lasso 问题

$$\hat{\beta}_{k+1} = \min \sum_{j=1}^p \|X\beta_{k+1} - X\beta_{k+1}^*\|_2^2 + \lambda_{1,j} \|\beta_{k+1,j}^*\|_1. \quad (13)$$

采用坐标下降法<sup>[14-15]</sup> 稀疏载荷矩阵  $\beta_{k+1}^*$ . 针对  $\hat{\beta}_{k+1}$  的第  $j$  个分量,通过软阈值  $S_\lambda(\cdot)$  算子求解,引入如下定理.

**定理 2** 对于任意的  $\mu \in R$ , 当  $\lambda > 0$  时,  $\frac{S_\lambda(\mu)}{a}$  表示函数  $\frac{1}{2}ax^2 - bx + \lambda|x|$  的最小值点.

将该定理应用于稀疏化递归载荷矩阵,得到式(12)的解为

$$\beta_{k+1,j} = \frac{S_\lambda\left(\sum_{i=1}^{N_{k+1}} X_{ij} E_i^j\right)}{\sum_{i=1}^{N_{k+1}} X_{ij}^2}. \quad (14)$$

$$S_\lambda(\mu) = \begin{cases} \mu - \lambda, & \mu > \lambda; \\ 0, & |\mu| \leq 0; \\ \mu + \lambda, & \mu < -\lambda. \end{cases} \quad (15)$$

$$E_i^j = X\beta_{k+1}^* - \sum_{u \neq j} X_{iu} \beta_{k+1,u}. \quad (16)$$

通过固定其中一个分量进行迭代更新,最后得到稀疏解.

综上, RSPCA 算法流程如下.

step 1: 标准化历史样本矩阵  $X_k^0$ , 计算载荷矩阵  $W_k$ 、平均值向量  $m_k$ 、标准差  $\Sigma_k$ .

step 2: 由式(10)得到初始化参数  $\beta_k^*$ .

step 3: 新增样本数据时,由式(8)和(9)递归更新  $X_{k+1}, G_{k+1}$ .

step 4: 由式(12)更新载荷矩阵  $W_{k+1}$ .

step 5: 由式(11)递推更新参数  $\beta_{k+1}^*$ .

step 6: 由式(14)对参数  $\beta_{k+1}^*$  迭代稀疏化处理,最终得到稀疏载荷  $\beta_{k+1}$ .

step 7: 标准化  $\beta_{k+1}$ , 得到递归更新后的稀疏主成

分  $Y_{k+1} = X_{k+1}\beta_{k+1}$ .

RSPCA 的收敛性由如下定理说明.

**定理 3** RSPCA 产生的递推稀疏载荷矩阵是收敛的.

**证明** 更新后的样本矩阵  $X_{k+1}$  仍然为连续值,  $L_2$  模型在该定义域仍是可导的,故最优  $L_2$  模型有唯一最小二乘解  $\beta_{k+1}^*$ ,由式(11)递推  $\beta_{k+1}^*$  也是唯一解.而子问题(14)通过坐标下降法求解收敛于稳定点.  $\square$

采用递归更新时,递推计算总的时间复杂度为  $O(N_{k+1}n_{k+1}p + N_{k+1}n_{k+1}^2p + N_{k+1}p)$ ,远低于非递归所需时间复杂度  $O(N_{k+1}^2m + N_{k+1}p)$ ,节省了计算时间,更加适用于在线更新.

## 2 基于 RSPCA 的工业过程故障监测与诊断

当过程变量发生异常时,变量的统计量也将会发生变化,因此可以分析过程数据的统计量,从中抽取变化特征,从而对故障进行监测和诊断.

### 2.1 基于 RSPCA 模型的过程监测

多变量故障监测实际上是通过监视两个多元统计量——Hotelling- $T^2$  和 SPE(或  $Q$ ) 统计量<sup>[16]</sup>,以获取整个生产过程运行状况的实时信息,进而判断整个过程的运行状态是否异常.

在 RSPCA 模型中,第  $k+1$  个样本的  $T^2$  和  $Q$  统计量定义如下:

$$T_{k+1}^2 = X_{k+1}\beta_{k+1}R_{k+1}^{-1}\beta_{k+1}^T X_{k+1}^T, \quad (17)$$

$$Q_{k+1} = (I - \alpha_{k+1}\beta_{k+1}^T)G_{k+1}(I - \beta_{k+1}\alpha_{k+1}^T). \quad (18)$$

在 RSPCA 模型中,第  $k+1$  个样本的  $T^2$  和  $Q$  统计量的阈值为

$$T_{k+1}^2 = \frac{p(N_{k+1} - 1)(N_{k+1} + 1)}{N_{k+1}(N_{k+1} - p)} F_\alpha(p, N_{k+1} - p), \quad (19)$$

$$Q_{k+1} = \theta_1 \left[ \frac{C_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}}. \quad (20)$$

其中:  $F_\alpha(p, n-p)$  表示分位点为  $p$  和  $N_{k+1} - p$  的  $F$  分布;  $C_\alpha$  表示正态分布点,置信度均为  $(1 - \alpha)$ ;  $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}, \theta_i = \sum_{j=p+1}^j \lambda_j^i, \lambda_j$  为协方差矩阵  $R_{k+1}$  中几个较小的特征根.

### 2.2 基于 RSPCA 模型的过程诊断

多元统计量  $T^2$  和  $Q$  超出阈值时,表明过程中出现了异常,但无法判断出造成故障或异常的原因.贡献图法<sup>[17]</sup> 给出了各个观测变量对监测统计量的贡

献,能够从异常的 $T^2$ 和 $Q$ 统计量找出高贡献率的过程变量,从而实现简单的故障分离和诊断功能.

过程变量对 $T^2$ 和 $Q$ 的贡献定义<sup>[18]</sup>如下:

$$\text{Con}(T_{k+1}^2) = X_i \beta_{k+1} \Lambda^{-\frac{1}{2}} \beta_{k+1}, \quad (21)$$

$$\text{Con}(Q_{k+1}) = [X_i(I - \beta_{k+1} \beta_{k+1}^T)]^2. \quad (22)$$

其中: $X_i$ 为故障样本, $\Lambda$ 为稀疏主成分协方差矩阵.

### 3 实验验证

在TE过程<sup>[19]</sup>进行实验验证. TE过程较好地模拟了实际复杂工业过程的典型特征,常被作为标准模型来评价过程监测和故障诊断方法的可行性.

#### 3.1 过程介绍

TE过程共有4种气态反应物A、C、D和E,2种气态产物G和H. 整个过程主要包括5个操作单元:反应器、冷凝器、循环压缩机、分离器和汽提塔. 气态反应物进入反应器,生成产物G和H,产物和残余反应物通过冷凝器送入气液分离器进行分离,气态通过压缩机进入循环管道与新鲜进料混合送入反应器循环使用进入汽提塔进行精制,从汽提塔底得到的流股中主要包含TE过程的产物G和H,送至下游的过程见文献[20].

#### 3.2 实验设计

TE过程数据包含正常状态和21种不同的故障状态. 除反应器中搅拌器的搅拌速度外,每种状态有52个变量,其中12个过程操作变量(XMV(1)~XMV(12))、41个测量变量(XMEAS(1)~XMEAS(41))构成某一采样时刻的观测向量,按顺序构成监测矩阵,分为训练集和测试集.

训练集中正常样本数据在25h运行仿真下获得,每隔3min采样一次,得到500个样本;测试集样本中的故障数据集在48h运行仿真下获得,每隔3min采样一次,得到960个样本,每种故障引入点为第161个样本点,一直持续到大约第360个样本点,开始恢复到正常状态.

为测试RSPCA算法的有效性,首先对算法性能进行测试,实验将本文所提出RSPCA算法的载荷向量与PCA的载荷向量进行对比,以评估算法的性能;然后将RSPCA方法与PCA、PLS、KPCA、SPCA监测方法进行对比,验证基于RSPCA方法的故障监测能力. 为保证实验的公平性,所有方法的故障监测控制置信限为99%,在各模型方法中设定主成分的贡献率为90%,实验以故障5为例进行故障监测.

各方法参数设置如下:在KPCA算法中,采用核半径为20的高斯核函数;在PLS算法中,根据主成分

贡献率理论计算相应的主成分个数,选取变量35作为质量数据;在SPCA算法中,Lasso惩罚系数 $\lambda_1 = 0.5$ ,迭代次数和收敛条件分别为300和 $1 \times 10^{-6}$ ,RSPCA的参数设置与SPCA算法设置相同.

为排除实验的偶然性,对21种类型的故障数据进行在线监测,以验证本文所提出方法的有效性,采用误报率作为评价指标评估RSPCA方法的故障监测能力.

实验以故障5为例进行故障诊断,选取第180个样本点(故障点),通过贡献图法对检测出故障的样本进行诊断,识别出故障根源变量.

#### 3.3 结果分析与讨论

##### 3.3.1 验证性结果分析

图1为PCA两个主元PC1和PC2的载荷向量以及RSPCA的两个稀疏主元SPC1和SPC2的载荷向量,可以看出PCA的两个主元载荷向量分布在坐标轴以外的区域,而RSPCA的两个主元的载荷向量分布在坐标轴上. 产生这种现象的原因是由于RSPCA模型稀疏化了载荷向量. 相对于传统PCA模型,RSPCA模型减少了无关变量对主元的影响,可以精确描述数据的内在空间,当过程出现扰动、误操作和故障时,对故障监测和诊断的灵敏度更高.

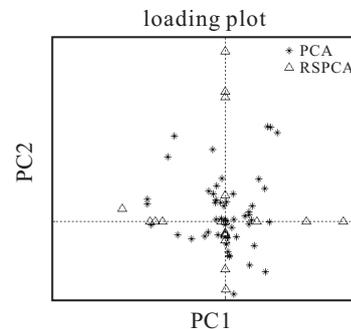


图1 载荷向量图

图2(a)~图2(e)分别为PCA、PLS、KPCA、SPCA和RSPCA模型对故障5监测的结果,图中直线代表置信度为99%的控制限,曲线为各样本点 $T^2$ 和 $Q$ 统计量的值,当统计量超过过程控制限时,认为是故障点.

由图2可见,各模型和RSPCA模型均能在故障引入的样本点(即第161个样本点)附近检测出故障,统计量 $T^2$ 和 $Q$ 都明显超出控制限,一直到大约第360个样本点,开始恢复到正常状态. PCA、PLS和SPCA模型的 $Q$ 统计图中,正常状态下的部分样本数400~960,仍有大量样本被检测为故障,造成误检. 在故障状态下的部分样本数200~300,有少量样本被检测为正常,造成漏检;在KPCA模型的 $Q$ 统计图

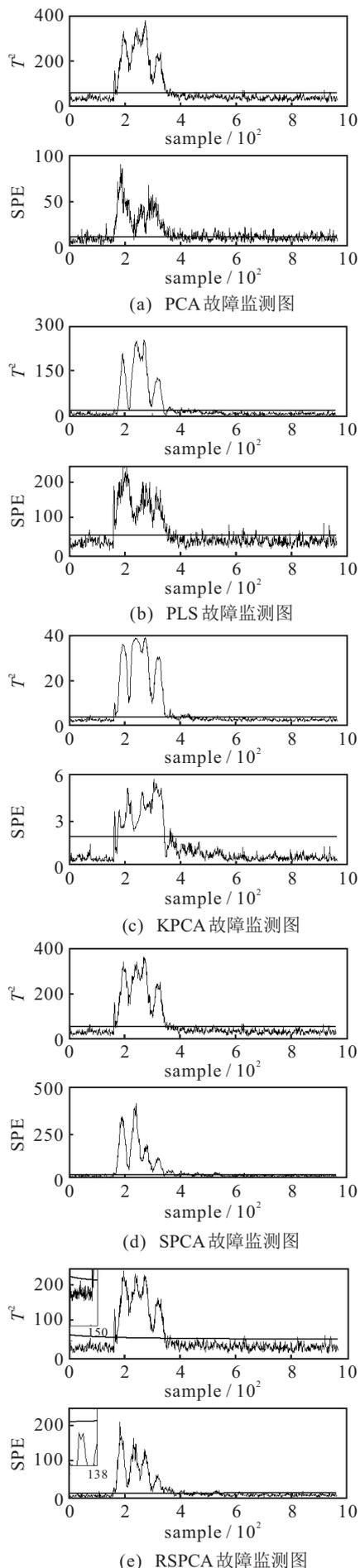


图2 故障5监测图

中,在正常状态下的部分样本数400~960,误报数明显减少,但在故障状态下的部分样本数160~200存在较高的故障漏报;各模型 $T^2$ 统计图中,在正常状态下的部分样本数也存在一定的误诊,而本文所提出的RSPCA方法通过将监测出来的正常样本重新加入训练集进行更新,同时更新相应的统计控制限,可极大地改善故障监测中的漏检和误检.由图2(e)可见,样本0~160和样本360~960的误报明显降低,样本200~300之间的漏报也优于其他监测方法.

实际工业过程的正常数据大多数是非稳态的,因此,基于静态历史数据所建立的监测模型往往无法适应复杂工业过程的动态变化,监测过程中极易产生误报和漏报.所提出的RSPCA模型具有递归更新能力,能够有效降低故障误报率和漏报率.

### 3.3.2 工业过程故障监测性能分析

表1为RSPCA方法和其他方法故障监测的误报率.可以看出,RSPCA方法对故障监测误报率低于PCA方法和SPCA方法.虽然该方法在故障15等少数故障类型进行监测时的误报率并不是最低,但从绝大多数故障监测的误报率和统计的平均值MFDR看,RSPCA方法占有较大优势.基于PCA方法的故障监测方法产生了大量误报,获得较差的监控表现;PLS方法结合了典型相关、主成分分析和多变量回归,监测能力比PCA方法稍强;KPCA方法针对过程数据的非线性特性,采用核函数将非线性数据转换为线性数据,其监测能力较PCA方法和PLS方法有所提高;RPCA方法也是一种递归更新的PCA模型,具有较强的自适应能力,但该方法的主成分监测模型的灵敏度不高,在监测过程中容易造成误检和漏检,基于RPCA的故障监测方法并没有达到较好的效果;SPCA方法通过对主成分稀疏化可以有效解决上述问题,监测能力比PCA方法高;RPCA方法和SPCA方法所构建的监测模型,都仅提高了过程监控某一方面的能力.RSPCA方法是SPCA方法与RPCA方法的有机结合,能够同时提高故障监测的自适应能力和灵敏度,有效降低了故障的误报率和漏检率,在时变过程的故障监测中具有较好的综合性能.

### 3.3.3 工业过程故障诊断结果分析

图3为利用贡献图法对故障5第180个样本点进行故障诊断的曲线图.可以看出,对于 $T^2$ 统计量主要贡献的变量号为16、25、31和35,对于 $Q$ 统计量主要贡献的变量号为11、16、25.其中第16个变量对故障5的贡献率最大,即汽提塔压力发生异常直接导致冷凝器冷却水入口温度发生异常.当汽提塔压力发生异

表1 各方法故障监测误报率对比

单位: %

Fault No.	PCA		PLS		KPCA		RPCA		SPCA		RSPCA	
	$T^2$	$Q$										
IDV1	62.50	64.69	62.60	63.12	62.60	62.50	62.50	63.84	62.60	62.92	62.50	62.81
IDV2	62.71	65.00	62.50	63.02	62.81	62.50	62.63	63.05	62.71	62.54	62.50	60.81
IDV3	1.35	20.62	4.06	3.16	6.04	0	0.46	6.38	1.08	18.23	0	5.94
IDV4	32.81	65.52	62.81	7.71	22.92	19.46	14.86	22.63	29.69	63.33	6.46	12.60
IDV5	3.23	25.31	4.90	4.90	6.04	1.52	2.89	10.74	2.23	15.94	2.10	5.00
IDV6	62.50	64.58	62.60	63.02	62.50	62.50	60.28	63.65	62.50	62.92	59.97	61.31
IDV7	62.50	65.42	19.90	62.60	62.50	60.94	61.91	65.34	62.50	62.81	60.50	62.60
IDV8	62.60	64.79	59.79	62.81	62.60	62.40	62.50	61.41	62.60	60.17	62.50	56.29
IDV9	93.23	16.56	6.25	3.44	8.23	0	1.10	12.43	2.96	8.44	0	6.67
IDV10	27.40	50.21	25.00	28.65	21.87	22.50	17.14	32.66	27.81	41.56	12.47	28.69
IDV11	34.58	5.71	54.17	15.83	28.13	18.02	29.31	16.12	34.27	14.24	23.02	8.54
IDV12	62.60	66.25	63.12	62.92	63.02	62.40	62.60	62.39	62.60	60.15	62.19	61.12
IDV13	62.50	65.00	62.50	62.81	62.71	62.50	62.50	63.81	62.50	62.23	62.50	61.65
IDV14	62.60	66.25	62.50	62.81	62.92	62.29	62.60	65.19	62.60	64.06	62.60	63.33
IDV15	21.46	20.42	8.65	6.67	13.54	1.56	14.37	16.36	15.94	19.90	9.88	13.67
IDV16	21.35	51.77	26.25	28.65	18.54	17.19	19.94	26.35	13.44	16.98	7.83	9.55
IDV17	52.08	67.08	55.94	57.71	51.56	58.75	49.57	55.00	55.10	51.65	45.10	51.42
IDV18	62.71	66.35	62.60	63.02	62.50	62.50	61.26	56.83	62.81	63.23	60.00	54.76
IDV19	6.67	37.71	3.13	12.19	9.48	27.14	4.65	16.72	6.88	21.0	1.95	12.42
IDV20	26.15	52.81	22.19	37.92	24.37	21.25	23.44	26.40	22.92	37.40	19.23	34.20
IDV21	32.71	60.21	25.74	39.79	20.63	24.17	22.56	26.19	26.46	48.02	20.31	35.96
MFDR	39.34	50.58	38.91	38.70	37.88	36.77	36.15	39.69	38.20	43.70	33.51	36.64

常时, 导致流6的成分C(变量25)、流9的成分C(变量31)和流9的成分G(变量35)的流速发生异常. 而冷凝器冷却水入口温度发生变化, 又间接导致产品分离器的温度发生异常(变量11).

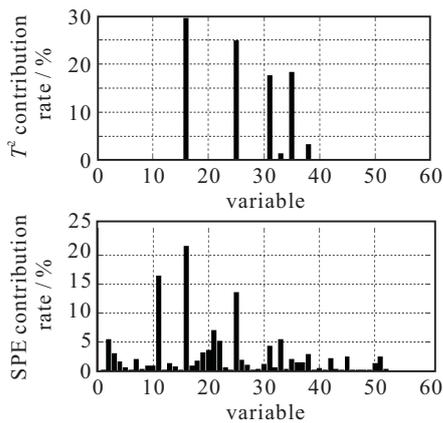


图3 故障前后对比

图4为变量16故障前后对比. 可以看出, 该变量在故障引入的样本点(即第161个样本点)附近发生阶跃, 偏离出正常状态, 随后又恢复至正常状态, 表明该变量为故障变量, 由此验证了本文所提出方法通过贡献图可以将故障根源变量从正常变量有效地隔离出来, 从而实现故障诊断.

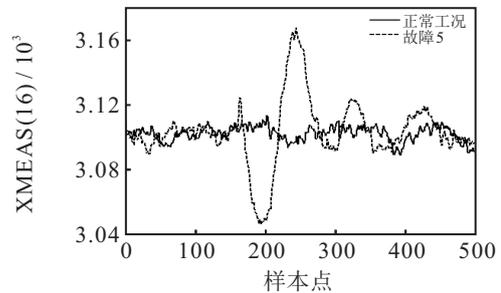


图4 故障诊断图

### 4 结论

本文提出了一种基于RSPCA的复杂工业过程监测与诊断方法. 基于所建立的RSPCA模型, 可实时递归更新过程故障监测统计量的过程控制限, 从而实现流程工业过程监测模型的自适应更新. 所提出方法有效克服了传统离线建模不适应于时变过程在线监测的缺点, 提高了过程监测的实时性和准确性. TE过程故障监测结果表明, 所提出方法能有效降低故障监测误报率和漏检率, 并且能够准确进行故障溯源, 为时变的复杂工业过程长时间监测提供了一种新的解决方案. 下一步工作是将所提出方法在长流程的复杂工业过程(如矿物浮选过程)中进行工业验证与应用, 以进一步改进所提出方法的性能.

## 参考文献(References)

- [1] 张坤鹏, 姜斌, 陈复扬, 等. 基于时变模型辨识的高速列车复合故障诊断[J]. 控制与决策, 2019, 34(2): 274-278.  
(Zhang K P, Jiang B, Chen F Y, et al. Composite fault diagnosis of high speed train based on time-varying model identification[J]. Control and Decision, 2019, 34(2): 274-278.)
- [2] Niu Y, Wang S, Lin Z, et al. Fault detection of industrial processes based on multivariate statistical process monitoring[J]. Journal of Chinese Society of Power Engineering, 2017, 37(10): 829-836.
- [3] Bing S, Shi H. Fault detection and classification using quality-supervised double-Layer method[J]. IEEE Transactions on Industrial Electronics, 2018, 65(10): 8163-8172.
- [4] Song B, Zhou X, Tan S, et al. Process monitoring via key principal components and local information based weights[J]. IEEE Access, 2019, 99(7): 15357-15366.
- [5] Zou H, Hastie T, Tibshirani R. Sparse principal component analysis[J]. Journal of Computational and Graphical Statistics, 2006, 15(2): 265-286.
- [6] Alameer A, Ghazaei G, Degenaar P, et al. Object recognition with an elastic net-regularized hierarchical MAX model of the visual cortex[J]. IEEE Signal Processing Letters, 2016, 23(8): 1062-1066.
- [7] Yu H, Khan F, Garaniya V. A sparse PCA for nonlinear fault diagnosis and robust feature discovery of industrial processes [J]. Aiche Journal, 2016, 62(5): 1494-1513.
- [8] Sapena-Bano A, Burriel-Valencia J, Pineda-Sanchez M, et al. The harmonic order tracking analysis method for the fault diagnosis in induction motors under time-varying conditions[J]. IEEE Transactions on Energy Conversion, 2017, 32(1): 244-256.
- [9] Du Y, Budman H, Duever T A, et al. Fault detection and classification for nonlinear chemical processes using lasso and gaussian process[J]. Industrial & Engineering Chemistry Research, 2018, 57(27): 8962-8977.
- [10] Yan R, Xiao H, Wang Z, et al. Detection, isolation and diagnosability analysis of intermittent faults in stochastic systems [J]. International Journal of Control, 2017, 91(8): 1-24.
- [11] Zhang L, Yong L, Chang Q. Intermittent connection fault diagnosis for CAN using data link layer information[J]. IEEE Transactions on Industrial Electronics, 2017, 64(3): 2286-2295.
- [12] Hastie T, Mazumder R, Zadeh R, et al. Matrix completion and low-rank SVD via fast alternating least squares[J]. Journal of Machine Learning Research, 2015, 16(1): 3367-3402.
- [13] Yue L, Ge Z. Weighted random forests for fault classification in industrial processes with hierarchical clustering model selection [J]. Journal of Process Control, 2018, 64(4): 62-70.
- [14] Jiang Y, He Y, Zhang H. Variable selection with prior information for generalized linear models via the prior LASSO method[J]. Journal of the American Statistical Association, 2016, 111(5): 355-376.
- [15] Bien J, Taylor J, Tibshirani R. A lasso for hierarchical interactions[J]. Annals of Statistics, 2018, 41(3): 1111-1141.
- [16] Vanhatalo E, Kulahci M. The effect of autocorrelation on the hotelling  $T_2$  control chart[J]. Quality & Reliability Engineering International, 2016, 31(8): 1779-1796.
- [17] Vouillamoz N, Mosar J, Deichmann N. Multi-scale imaging of a slow active fault zone: Contribution for improved seismic hazard assessment in the Swiss Alpine foreland[J]. Swiss Journal of Geosciences, 2017, 110(2): 547-563.
- [18] Gajjar S, Kulahci M, Palazoglu A. Real-time fault detection and diagnosis using sparse principal component analysis[J]. Journal of Process Control, 2017, 67: 112-128.
- [19] Yi J, Di H, He H, et al. A novel framework for fault diagnosis using kernel partial least squares based on an optimal preference matrix[J]. IEEE Transactions on Industrial Electronics, 2017, 64(5): 4315-4324.
- [20] Amin M T, Imtiaz S, Khan F. Process system fault detection and diagnosis using a hybrid technique[J]. Chemical Engineering Science, 2018, 189: 191-211.

## 作者简介

刘金平(1983—), 男, 副教授, 博士, 从事工业过程自动化等研究, E-mail: ljp202518@163.com;

王杰(1995—), 男, 硕士生, 从事复杂工业过程自动化监控的研究, E-mail: wangjie\_hunnu@163.com;

刘先锋(1964—), 男, 教授, 从事数据挖掘等研究, E-mail: ljp202518@163.com;

唐朝晖(1965—), 男, 教授, 博士, 从事复杂工业过程故障诊断等研究, E-mail: zhtang@csu.edu.cn;

马天雨(1978—), 男, 讲师, 博士, 从事复杂工业过程智能监控等研究, E-mail: mty@hunnu.edu.cn;

肖文辉(1981—), 女, 讲师, 博士生, 从事复杂工业过程智能监控的研究, E-mail: 53819362@qq.com.

(责任编辑: 郑晓蕾)