

# 基于改进高效偏最小二乘的质量相关故障诊断

孔祥玉, 解建<sup>†</sup>, 罗家宇, 李强  
(火箭军工程大学 导弹工程学院, 陕西 西安 710025)

**摘要:** 高效偏最小二乘(EPLS)作为偏最小二乘(PLS)的扩展算法之一, 在质量相关故障检测中取得了良好的应用效果。然而, 研究发现当系统中存在一些与产品质量无关的信息时会导致EPLS的检测率降低, 影响工业生产安全及效益。同时, 传统的基于贡献图的故障诊断方法在无故障时输入变量会对故障检测指标的贡献值不均等, 从而影响故障诊断效果。针对上述问题, 本文提出了一种改进高效偏最小二乘(IEPLS)的质量相关故障诊断方法。所提方法首先用正常数据建立IEPLS算法模型, 利用获得的模型参数对过程变量进行空间分解。然后在分解后的空间中定义局部信息增量均值和局部动态阈值, 结合故障判据进行故障检测。当故障发生后, 利用每个变量的新息矩阵计算对故障总体的新息贡献率, 根据各个变量新息贡献率大小实现对故障变量的定位。最后, 使用田纳西伊士曼过程(TEP)对算法性能进行了验证。

**关键词:** 过程监控; 质量相关; 故障诊断; 改进高效偏最小二乘; 新息贡献率; TE过程

**引用格式:** 孔祥玉, 解建, 罗家宇, 等. 基于改进高效偏最小二乘的质量相关故障诊断. 控制理论与应用, 2020, 37(12): 2645–2653

DOI: 10.7641/CTA.2020.00197

## Quality-related fault diagnosis based on improved efficient partial least squares

KONG Xiang-yu, XIE Jian<sup>†</sup>, LUO Jia-yu, LI Qiang

(Department of Missile Engineering, Rocket Force University of Engineering, Xi'an Shaanxi 710025, China)

**Abstract:** Efficient partial least squares (EPLS), as one of the extended algorithms of partial least squares (PLS), has achieved good application results in the quality-related fault detection. However, it is found that when there is information unrelated to the product quality in the system, the detection rate of EPLS is reduced, which can affect the safety and efficiency of the industrial production. Meanwhile, the traditional contribution graph-based fault diagnosis method has unequal contribution values to the fault detection index when there is no fault, thereby affecting the fault diagnosis effect. As such, an improved EPLS (IEPLS) quality-related fault diagnosis method is proposed in this paper. Firstly, the normal data are used to establish the IEPLS-based model, and the obtained model parameters are employed to spatially decompose the process variables. Secondly, the local mean value of the incremental information and the local dynamic threshold are defined to detect the fault in the decomposed space. When there is a fault, the new information matrix of each variable is applied to calculate the new information contribution rate to the total failure, and the fault variable is located based on the new information contribution rate of each variable. Finally, the performance of the proposed algorithm is verified by using the Tennessee Eastman process (TEP).

**Key words:** process monitoring; quality-related; fault diagnosis; improved efficient partial least squares; new information contribution rate; TE process

**Citation:** KONG Xiangyu, XIE Jian, LUO Jiayu, et al. Quality-related fault diagnosis based on improved efficient partial least squares. *Control Theory & Applications*, 2020, 37(12): 2645–2653

## 1 引言

近年来, 随着传感器技术的快速发展, 一些复杂工业设备的数据采集变得更加便捷。然而, 面对这些大

型复杂设备和系统, 对其进行故障诊断却变得十分困难。原因在于传统故障检测诊断方法大多是通过系统的物理模型来进行的, 面对高度集成化和复杂化的系

收稿日期: 2020-04-16; 录用日期: 2020-08-20.

<sup>†</sup>通信作者。E-mail: hdxiejian@163.com; Tel.: +86 17775219951.

本文责任编辑: 周东华。

国家自然科学基金项目(61673387, 61833016, 61903375), 陕西省自然科学基金项目(2020JM-356)资助。

Supported by the National Natural Science Foundation of China (61673387, 61833016, 61903375) and the Natural Science Foundation of Shaanxi Province (2020JM-356).

统对其建立物理模型严重影响了生产效率和成本。因此,基于数据驱动<sup>[1]</sup>的多元统计过程监控<sup>[2-3]</sup>(multivariate statistical process monitoring, MSPM)方法因其不需要建立系统精确的物理模型逐渐成为了该领域的研究热点,并在实际生产中取得了很好的应用效果。传统的多元统计过程监控方法主要有主成分分析(principal component analysis, PCA)<sup>[4]</sup>,独立成分分析(independent component analysis, ICA)<sup>[5]</sup>和偏最小二乘(partial least squares, PLS)<sup>[6-7]</sup>,其中以PLS算法的应用最为广泛。PLS算法在对过程变量执行空间分解时可以将过程变量分解为两部分:一部分与最终产品性能指标相关即主元空间(principal component space, PCS),另一部分与最终产品性能指标无关即残差空间(residual space, RS)。对于质量无关子空间,因为其包含的质量无关信息不会影响最终产品质量,其影响往往可以忽略不计。而通过对质量相关子空间的监控可以获得影响产品质量的过程变量信息,方便人们对系统进行故障检测和诊断,避免不必要的经济损失和安全风险。

PLS算法虽然在实际生产中取得了较好的应用效果,但是当系统中含有较多的质量无关信息时其性能明显下降。为了解决该问题,去除质量无关信息的影响,Trygg和Wold<sup>[8]</sup>首先提出了一种正交信号校正(orthogonal signal correction, OSC)的预处理方法,该方法在对PLS进行建模前剔除了过程变量中的质量无关信息,很好的解决了质量无关信息对质量相关故障监控的干扰。为了从根本上实现对过程变量的正交分解,Yin等人<sup>[9]</sup>提出了一种新的空间分解算法即改进的偏最小二乘算法(modified partial least squares, MPLS),该方法通过计算过程变量与质量变量的正交关系对过程变量进行正交分解,从而在质量相关子空间中去除了正交信息,该方法明显提高了质量相关故障检测率并且避免了大量的迭代过程。虽然MPLS从根本上解决了质量无关信息的干扰,但是MPLS在计算过程中存在广义逆计算,这可能会导致过程变量中与质量相关的信息丢失。针对这一问题,Peng等人<sup>[10]</sup>提出了高效偏最小二乘(efficient partial least squares, EPLS)算法,该算法在MPLS的基础上对过程变量的残差子空间进行PCA分解,分别对残差子空间中的主成分及噪声进行监控,明显提高了质量相关故障的可检测性。然而,在实际操作中EPLS算法是基于一个近似的假设对过程变量进行与质量相关的正交分解,这使得分解得到的质量相关子空间中依然含有少量与质量无关的信息,对连续流水生产形式的工业过程造成较大影响。为此,Kong等人<sup>[11]</sup>提出了基于正交信号校正的高效偏最小二乘(orthogonal signal correction efficient partial least squares, OSC-EPLS)的质量相关故障检测方法,该方法利用OSC预处理算法处理过程变量后再使用EPLS算法进行空间分解,获得了更能

反映产品关键性能指标的质量相关子空间,从而提高了质量相关故障的检测率。

为了能够在准确检测出系统的质量相关故障后及时排除故障,尽可能减少经济损失和人员伤亡,还必须对系统进行故障诊断,找出导致故障发生的原因并对故障变量进行定位。最常用的PLS故障诊断方法多为基于贡献图<sup>[12]</sup>的故障诊断方法,在实际应用中具有较好的性能,但是在无故障时的过程变量会造成故障检测指标的贡献值不均等,影响算法准确定位故障变量。为了解决上述问题,本文提出使用一种基于改进的高效偏最小二乘(improved efficient partial least squares, IEPLS)质量相关故障诊断方法,IEPLS算法旨在解决两方面问题:1)传统的EPLS和OSC-EPLS等故障检测算法对质量相关故障的检测率较低,且质量无关故障误报率较高;2)基于贡献图的故障诊断方法对故障指标定位时的贡献不均等,导致在故障诊断过程中对主要的故障变量贡献值低。

本文所提的IEPLS质量相关故障诊断方法的主要创新如下:1)针对EPLS和OSC-EPLS算法在对质量相关故障检测时检测率较低的问题,提出使用局部协方差矩阵与局部数据信息增量矩阵<sup>[13]</sup>的方法对EPLS的质量相关子空间设计新的统计量和控制限。IEPLS解决了质量相关故障检测率较低的问题,并在一定程度上保证了质量无关故障较低的误报率,提高了算法的整体性能;2)在新的统计量和控制限的基础上,提出在EPLS的质量相关子空间结合新息贡献率<sup>[13]</sup>技术对故障变量进行定位,并对比传统的基于贡献图的故障诊断方法。实验结果表明,本文所提的IEPLS质量相关故障诊断方法对质量相关故障的检测性能更好,且对故障变量的定位具有较好的准确性。

本文结构安排如下:第2节对标准PLS算法进行简要回顾;第3节介绍了EPLS及其扩展的质量相关故障检测方法;第4节提出了基于IEPLS的质量相关故障诊断方法;第5节结合田纳西伊士曼过程,验证了所提算法的性能;第6节对IEPLS算法进行了总结。

## 2 标准PLS

标准PLS算法<sup>[6]</sup>给出一对输入输出数据矩阵 $X$ 和 $Y$ , $X$ 代表自变量, $Y$ 是因变量,并分别具有如下形式:

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_N^T \end{bmatrix} \in \mathbb{R}^{N \times m}, Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix} \in \mathbb{R}^{N \times l}, \quad (1)$$

$$x_i \in \mathbb{R}^m, y_i \in \mathbb{R}^l, i = 1, \dots, N, \quad (2)$$

其中: $N$ 代表的是采样数, $m$ 和 $l$ 分别代表输入变量和输出变量的个数。将原始变量空间投影到一个不相关的潜因子空间

$$T = [t_1 \cdots t_A] \in \mathbb{R}^{N \times A}, \quad (3)$$

$T$ 是得分矩阵。由于得分矩阵 $T$ 无法从输入 $X$ 中直接

获得, 需要引入权重  $R = W(P^T W)^{-1}$ , 由此可得到

$$T = X R. \quad (4)$$

基于上述变换, 可以将原始变量  $X$  分为主元子空间  $\hat{X}$  和残差子空间  $\tilde{X}$ ,  $Y$  分为可预测部分  $\hat{Y}$  与不可预测部分  $\tilde{Y}$ , 其外部模型如下所示:

$$\begin{cases} X = \hat{X} + \tilde{X} = T P^T + \tilde{X}, \\ Y = \hat{Y} + \tilde{Y} = T Q^T + \tilde{Y}, \end{cases} \quad (5)$$

其中:  $A$  是主元个数, 由交叉验证得到;  $P \in \mathbb{R}^{m \times A}$  和  $Q \in \mathbb{R}^{l \times A}$  分别是  $X$  和  $Y$  的负载矩阵. PLS 建模的各个参数使用非线性迭代偏最小二乘算法 (nonlinear iterative partial least squares, NIPALS)<sup>[14]</sup> 求得, 在 NIPALS 迭代过程中,  $Y$  的特征向量被用来回归  $X$  的特征向量, 而  $X$  的特征向量被用来对  $Y$  进行回归. 最终, 算法通过旋转特征向量逐步提高相关性收敛获得模型参数. 标准PLS求解模型参数的方法虽然存在计算量较大的问题, 然而其最大缺点是对过程变量  $X$  使用了斜交分解导致了质量相关信息的丢失, 从而造成了质量相关故障检测率较低, 影响了其过程监控的性能.

### 3 EPLS及其扩展算法的过程监控技术

#### 3.1 EPLS算法

由于标准PLS的上述缺点, 人们相继提出了其他算法用来解决斜交分解问题. EPLS作为一种更为先进的故障检测算法, 其应用效果要明显优于标准PLS 及其扩展算法. EPLS算法首先给出了期望的输出变量  $Y$  的分解如下:

$$Y = XM + E_y = \hat{Y} + E_y, \quad (6)$$

式中: 第1部分  $\hat{Y}$  与  $X$  相关, 第2部分  $E_y$  与  $X$  无关. 则会有

$$\text{cov}(e_y, x^T) = 0. \quad (7)$$

基于式(6)和式(7), 推导出如下关系<sup>[10]</sup>:

$$\frac{1}{N} Y^T X = \frac{1}{N} M^T X^T X + \frac{1}{N} E_y^T X \approx M^T \frac{X^T X}{N}, \quad (8)$$

那么,  $X$  和  $Y$  之间的相关矩阵为

$$M = (X^T X)^\dagger X^T Y. \quad (9)$$

EPLS算法具体步骤如下:

1) 对建模数据中的输入  $X$  和输出  $Y$  进行数据标准化;

2) 计算相关系数矩阵  $M$ :

$$M = (X^T X)^\dagger X^T Y; \quad (10)$$

3) 利用式(11)对  $MM^T$  执行奇异值分解(singular value decomposition, SVD)<sup>[15]</sup>:

$$MM^T = [\hat{P}_M \tilde{P}_M] \begin{bmatrix} \Lambda_M & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{P}_M^T \\ \tilde{P}_M^T \end{bmatrix}; \quad (11)$$

4) 由上式获得参数求出两个正交投影算子如式(12)与式(13):

$$\Xi_M = \hat{P}_M \hat{P}_M^T, \quad (12)$$

$$\Xi_M^\perp = \tilde{P}_M \tilde{P}_M^T; \quad (13)$$

5) 当测试数据到来时, 将  $X$  沿着正交投影算子分别向  $\text{span}\{M\}$  和  $\text{span}\{M\}^\perp$  上投影后得到分解后的质量相关子空间  $\hat{X}$  和质量无关子空间  $\tilde{X}$ :

$$\begin{cases} \hat{X} = X \Xi_M = X \hat{P}_M \hat{P}_M^T, \\ \tilde{X} = X \Xi_M^\perp = X \tilde{P}_M \tilde{P}_M^T; \end{cases} \quad (14)$$

6) 使用式(15)对质量无关子空间  $\tilde{X}$  执行PCA分解:

$$\tilde{X} = \hat{X}_r + \tilde{X}_r = T_{X_r} P_{X_r}^T + \tilde{X}_r, \quad (15)$$

其中  $T_{X_r}$  和  $P_{X_r}$  分别是  $\tilde{X}$  的得分矩阵和负载矩阵;

7) 得到最终的EPLS模型如下:

$$\begin{cases} X = \hat{X} + T_{X_r} P_{X_r}^T + \tilde{X}_r, \\ Y = \hat{Y} + E_y; \end{cases} \quad (16)$$

8) 利用式(17)分别对空间  $\hat{X}$ ,  $\hat{X}_r$  构造  $T^2$  统计量, 对  $\tilde{X}_r$  使用式(18)构造  $Q^2$  统计量<sup>[16]</sup>如下:

$$\begin{cases} T_{\hat{X}}^2 = t_{\hat{X}}^T \left( \frac{\hat{P}_M^T X^T X \hat{P}_M}{N-1} \right)^{-1} t_{\hat{X}}, \\ T_{\hat{X}_r}^2 = t_{\hat{X}_r}^T \left( \frac{T_{\hat{X}_r}^T T_{\hat{X}_r}}{N-1} \right)^{-1} t_{\hat{X}_r}, \end{cases} \quad (17)$$

$$Q_X = \|\tilde{X}\|^2 = \tilde{X}^T (I - P_{X_r} P_{X_r}^T) \tilde{X}, \quad (18)$$

其中:  $t_{\hat{X}} = \hat{P}_M^T X$ ,  $t_{\hat{X}_r} = P_{X_r}^T \tilde{X}$ ;

9) 构造基于  $\chi^2$  分布的控制限:

$$\begin{cases} J_{\text{th}, T_{\hat{X}}^2} = \hat{g} \chi_{\alpha, \hat{h}}^2, \hat{g} = \frac{\hat{S}}{2\hat{\mu}}, \hat{h} = \frac{2\hat{\mu}^2}{\hat{S}}, \\ J_{\text{th}, T_{\hat{X}_r}^2} = \hat{g} \chi_{\alpha, \hat{h}_r}^2, \hat{g}_r = \frac{\hat{S}_r}{2\hat{\mu}_r}, \hat{h}_r = \frac{2\hat{\mu}_r^2}{\hat{S}_r}, \\ J_{\text{th}, T_{\tilde{X}}^2} = \tilde{g} \chi_{\alpha, \tilde{h}_r}^2, \tilde{g} = \frac{\tilde{S}}{2\tilde{\mu}}, \tilde{h}_r = \frac{2\tilde{\mu}^2}{\tilde{S}}, \end{cases} \quad (19)$$

其中  $\hat{\mu}$  和  $\hat{S}$ ,  $\hat{\mu}_r$  和  $\hat{S}_r$ ,  $\tilde{\mu}$  和  $\tilde{S}$  分别是正态过程中  $T_{\hat{X}}^2$ ,  $T_{\hat{X}_r}^2$  和  $Q_X$  对应的均值和方差;

10) 结合上述设计的统计量和控制限与故障判据在相应空间进行故障检测.

由于 Li 等人<sup>[17]</sup>提出的 MPLS 算法对  $MM^T$  使用 SVD 分解时可能会导致  $X$  中质量相关信息的丢失. 因此, Peng 等人<sup>[10]</sup>在分析了 MPLS 算法的这个不足之后, 提出对残差子空间  $\tilde{X}$  进行 PCA 分解, 提高质量相关故障的检测率. 如式(16)所示,  $\hat{X}$  作为主元子空间专门负责预测质量信息  $Y$ , 而  $T_{X_r}$  则代表输入中对预测输出无用的变量, 其中  $\tilde{X}_r$  是系统噪声. EPLS 故障检测算法是在 MPLS 的基础上改进而来, 不仅很好的解决对  $X$  斜交分解所产生的问题, 同时使用 SVD 分解避免了大

量的迭代过程,减少了算法的计算量.

### 3.2 OSC-EPLS算法

EPLS算法相比MPLS算法其质量相关故障检测精度得到了明显提升,然而EPLS算法的输入数据由于存在噪声干扰等情况导致其模型不能完全分解,其质量相关子空间仍然包含质量无关信息影响模型对质量信息的监测.因此Kong等将OSC算法引入到EPLS中,提出了基于4类OSC-EPLS模型过程监控技术,用来降低输入数据中与质量正交的信息. OSC预处理算法的详细步骤如下<sup>[11]</sup>:

- 1) 对正常数据中的输入 $X$ 和输出 $Y$ 进行标准化处理;
- 2) 对 $X$ 和 $Y$ 利用PLS迭代算法求出向量 $w$ 和 $p$ ;
- 3) 利用式(20)求出模型参数 $w_{\perp}$ :

$$w_{\perp} = p - \frac{w^T p}{w^T w} w; \quad (20)$$

- 4) 利用式(21)计算正交得分向量 $t_{\perp}$ :

$$t_{\perp} = X w_{\perp}; \quad (21)$$

- 5) 计算正交得分矩阵

$$P_{\perp} = \frac{X^T t_{\perp}}{t_{\perp}^T t_{\perp}}; \quad (22)$$

- 6) 使用式(23)得出去正交后的输入矩阵 $X_{osc}$ :

$$X_{osc} = X - t_{\perp} p_{\perp}^T; \quad (23)$$

- 7) 循环上述过程直至满足算法的收敛精度要求.

$$\begin{cases} P_{\perp} = [P_{\perp} \ p_{\perp}], \\ W_{\perp} = [W_{\perp} \ w_{\perp}]. \end{cases} \quad (24)$$

利用式(23)计算出的输入矩阵 $X_{osc}$ 使用式(16)的EPLS模型进行空间分解,然后对质量相关空间和质量无关空间分别进行检测.本节主要介绍了基于(orthogonal projections to latent structures, OPLS)的正交信号预处理算法,其他3种OSC预处理算法如SWosc<sup>[18]</sup>, TFosc<sup>[19]</sup>和OPLS(PCA)的详细步骤见参考文献.

## 4 基于改进的高效偏最小二乘故障检测与辨识技术

本节设计了基于IEPLS的故障诊断方法,通过结合局部信息增量技术对现有EPLS算法的统计量和控制限进行改进,提高了现有算法的适应性和检测性能.另外,由于局部信息增量矩阵反映了相邻采样时刻之间新的信息差异,因此利用该思想可以定义局部信息增量矩阵为新息矩阵,利用新息矩阵计算新息贡献率对故障变量进行定位,从而实现IEPLS对质量相关故障的诊断. IEPLS算法的具体步骤如下所示:

- Step 1** 对建模数据 $X$ 和 $Y$ 进行标准化处理;  
**Step 2** 利用式(25)求取建模数据 $X$ 和 $Y$ 的关系

矩阵 $M$ :

$$M = (X^T X)^{\dagger} X^T Y; \quad (25)$$

**Step 3** 利用式(26)对 $MM^T$ 进行SVD分解得到正交投影算子:

$$MM^T = [\hat{P}_M \ \tilde{P}_M] \begin{bmatrix} A_M & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{P}_M^T \\ \tilde{P}_M^T \end{bmatrix}, \quad (26)$$

$$\Xi_M = \hat{P}_M \hat{P}_M^T, \quad (27)$$

$$\Xi_M^{\perp} = \tilde{P}_M \tilde{P}_M^T; \quad (28)$$

**Step 4** 在观测数据前加入一段正常数据组成测试数据 $X_t$ 并对其进行标准化处理;

**Step 5** 将 $X_t$ 向 $\text{span}\{M\}$ 和 $\text{span}\{M\}^{\perp}$ 上正交分解得到式(29):

$$X_t = \hat{X}_t + \tilde{X}_t = X_t \Xi_M + X_t \Xi_M^{\perp}; \quad (29)$$

**Step 6** 假设当前时刻为 $n$ ,从正常数据选取长度为 $L$ 的数据组成局部数据矩阵 $X_{t,n}$ ;那么下一时刻的局部数据矩阵 $X_{t,n+1}$ 为<sup>[13]</sup>

$$X_{t,n} = [X_t(n' - L + 1) \cdots X_t(n')], \quad (30)$$

$$X_{t,n+1} = [X_t(n' - L + 2) \cdots X_t(n'), X_t(n)], \quad (31)$$

其中 $n'$ 为正常数据中的某一时刻;

**Step 7** 求两个时刻之间所形成的公共数据矩阵 $G_{n,n+1}$ ,并定义局部信息增量均值 $g_{n,n+1}$ 为

$$G_{n,n+1} = [X_t(n' - L + 2) \cdots X_t(n')], \quad (32)$$

$$g_{n,n+1} = \frac{G_{n,n+1} I_{L-1}}{L-1}, \quad (33)$$

其中 $I_{L-1} = [1 \ 1 \cdots 1] \in \mathbb{R}^{(L-1) \times 1}$ .为了计算方便定义 $K_n = X_{t,n}(X_{t,n})^T = X_t(n' - L + 1)X_t(n' - L + 1)^T + G_{n,n+1} G_{n,n+1}^T$ ;

**Step 8** 计算 $n$ 时刻的局部协方差矩阵 $R_n$ :

$$R_n = \frac{K_n}{L-1} - \frac{LB_n B_n^T}{L-1}, \quad (34)$$

$$\text{其中 } B_n = \frac{(L-1)g_{n,n+1} + X_t(n' - L + 1)}{L};$$

**Step 9** 类似Step 8,计算下一时刻的局部协方差矩阵 $R_{n+1}$ :

$$R_{n+1} = \frac{K_{n+1}}{L-1} - \frac{LB_{n+1} B_{n+1}^T}{L-1}; \quad (35)$$

**Step 10** 如式(36)计算局部信息增量矩阵 $D_{n+1}$ :

$$D_{n+1} = R_{n+1} - R_n; \quad (36)$$

**Step 11** 求取局部信息增量均值 $\gamma_{n+1}$ 如式(37):

$$\gamma_{n+1} = \frac{\sum_{i=1}^m \sum_{j=1}^m |D_{n+1}[i,j]|}{m^2}; \quad (37)$$

**Step 12** 计算局部动态控制限 $\sigma_{n+1}$ 如式(38):

$$\sigma_{n+1} = \frac{1.5}{L} \sum_{k=n'-L+1}^{n'} \gamma^*, \quad (38)$$

其中 $\gamma^*$ 为未发生故障的局部信息增量均值;

**Step 13** 设计故障检测判据:

- a) 当 $\gamma_{n+1} \geq \sigma_{n+1}$ 时, 检测有质量相关故障发生;
- b) 当 $\gamma_{n+1} < \sigma_{n+1}$ 时, 检测无质量相关故障发生, 继续故障检测;

**Step 14** 利用新息矩阵 $D_{n+1}$ 计算各变量的新息贡献率 $P(i)$ <sup>[13]</sup>如式(39)所示:

$$P(i) = \frac{|\sum_{j=1}^m D_{n+1}[i, j]|}{\sum_{i=1}^m \sum_{j=1}^m |D_{n+1}[i, j]|}; \quad (39)$$

**Step 15** 分析各个变量的新息贡献率得出故障诊断结果.

最后, 本文所提的IEPLS算法步骤总结如下:

离线建模:

**步骤1** 对建模数据进行预处理;

**步骤2** 利用建模数据训练IEPLS模型;

**步骤3** 获得IEPLS的模型参数.

在线检测:

**步骤4** 对测试数据进行预处理;

**步骤5** 利用IEPLS模型对测试数据进行空间分解;

**步骤6** 利用设计的统计量和控制限进行故障检测;

故障判断:

**步骤7** 利用新息贡献率诊断方法对故障空间进行故障诊断;

**步骤8** 获得各个变量对故障的贡献结果进行故障变量定位.

**注1** 需要进行说明的是式(38)设计的控制限中参数“1.5”为控制限的优化系数, 在离线建模阶段通过对大量已知的正常工况数据使用粒子群(particle swarm optimization, PSO)智能优化算法训练得到的一个最优参数. 因本文研究重点旨在提高算法的检测率和故障诊断性能, 所以对于PSO算法本文不展开论述, 具体计算方法见文献[14]和文献[20].

## 5 仿真实验与讨论

本节中使用田纳西伊士曼过程验证EPLS, OSC-EPLS和IEPLS算法在故障检测与诊断中的性能, 包括TEP实验参数初始化、基于IEPLS的故障检测方法和基于IEPLS的故障诊断方法3部分, 实验结果表明所提算法的性能相比EPLS和OSC-EPLS有了明显提升.

### 5.1 TEP实验参数初始化

TEP是由美国田纳西伊士曼化工公司开发的一个工业模拟器<sup>[21-22]</sup>, 该仿真被广泛用作连续过程的策略、监控和诊断的研究平台. 田纳西伊士曼过程共包含4种反应物( $A, C, D, E$ ), 生成两种产物( $G, H$ ). 另外, 该过程还包含一种惰性物质 $B$ 和副产物 $F$ , 系统中的化学反应如下:

$$\begin{cases} A(g) + C(g) + D(g) \rightarrow G(lig), \\ A(g) + C(g) + E(g) \rightarrow H(lig), \\ A(g) + E(g) \rightarrow F(lig), \\ 3D(g) \rightarrow 2F(lig). \end{cases} \quad (40)$$

其中:  $g$ 代表气体,  $lig$ 代表液体. 整个流程由5个操作单元组成, 包括化学反应器、冷凝器、压缩机、汽/液分离器和分离器. TEP共包含22个数据集被用于故障检测与诊断等研究, 该数据集由1个正常数据和21个特定的故障数据集组成. 其中正常数据集共包含480个样本, 而每个故障数据集共包含960个样本. 在15个已知的故障数据集中又分为两种类型的故障数据, 即质量相关故障IDV(3, 4, 9, 11, 14, 15)和质量无关故障IDV(1, 2, 5-8, 10, 12, 13). 在本实验中, 选取数据集中的22个过程变量XMEAS(1-22)和11个操纵变量XMV(1-11)作为输入 $X$ , 选取过程变量XMEAS(35)作为质量变量 $y$ ; 另外, 选取100个正常数据用来模拟系统在正常工作时所产生的采样数据, 选取800个故障数据用来模拟系统在故障情况下产生的采样数据, 二者共同组成测试数据集用来进行故障检测. 实验参数初始化: 选取主元个数 $A = 15$ , 由交叉验证所得; 选取固定窗口长度 $L = 5$ , 由粒子群智能优化(PSO)算法寻优所得. 实验共分为两部分, 实验1通过TEP仿真实验对3种算法故障检测性能进行对比分析, 并对所提算法的性能进行了验证总结; 实验2则验证了所提算法的故障诊断性能, 最后分析实验结果给出结论.

### 5.2 基于IEPLS的故障检测方法

为了验证故障5下质量变量 $y$ 是否受到影响, 采用 $Q_y$ 统计量<sup>[23]</sup>对输出 $y$ 的残差空间进行直接监测, 该 $Q_y$ 统计量定义如式(41)所示:

$$Q_y = \|y - \hat{y}\|^2 \sim g\chi^2_{\alpha, h}, \quad (41)$$

其中:  $\hat{y}$ 是质量变量的估计值;  $Q_y$ 仅用于故障分类, 不能实现故障检测. 图1是IDV(5)下质量变量 $y$ 在不同工况下的对比图, 从图中可以看出, 在系统工作正常时, 质量变量 $y$ 始终稳定在固定的区间内, 没有发生异常波动. 当故障数据到来后, 质量变量 $y$ 直接超出控制限, 在故障到来的一刻发生了明显的异常. 然后由于系统反馈控制的作用, 质量变量 $y$ 又回到固定的区间内, 实验结果表明质量变量 $y$ 确实受到了故障的影响.

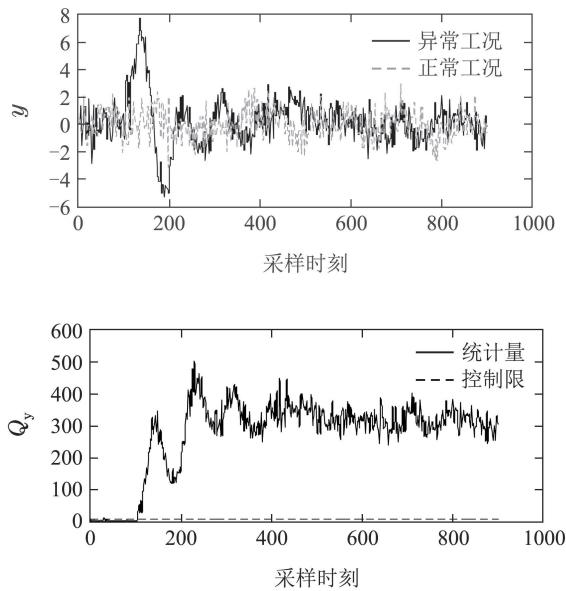
图1 IDV(5)下质量变量 $y$ 的变化对比图

Fig. 1 Comparison chart of changes in quality variable  $y$  under IDV (5)

图2~4分别为3种算法的IDV(1)故障检测图,其中:实线部分为算法的统计量,虚线表示算法的控制限。3组实验各选取了900组数据,0~100区段为正常数据,100~900区段为故障数据,900~1000空白区段表示算法继续进行故障检测的一个实时过程。图2为EPLS算法的IDV(1)故障检测图,从图中可以看出EPLS算法在对IDV(1)进行检测时存在多处漏报情况,其检测率较低。图3是加入OSC预处理算法后的IDV(1)检测图,从图中可以看出OSC-EPLS算法在100~900区段的漏报率相比于EPLS算法明显减少,证明OSC预处理算法提前去除了正交信息使得在进行质量相关信息检测时,其检测率明显提高,算法的整体性能得到了很大改善。

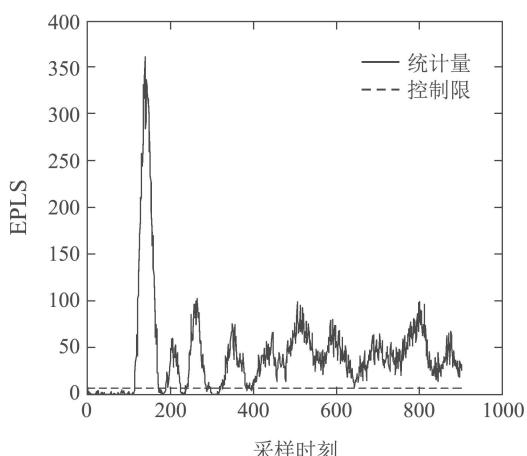


图2 EPLS的IDV(1)故障检测结果

Fig. 2 The IDV(1) fault detection result of EPLS

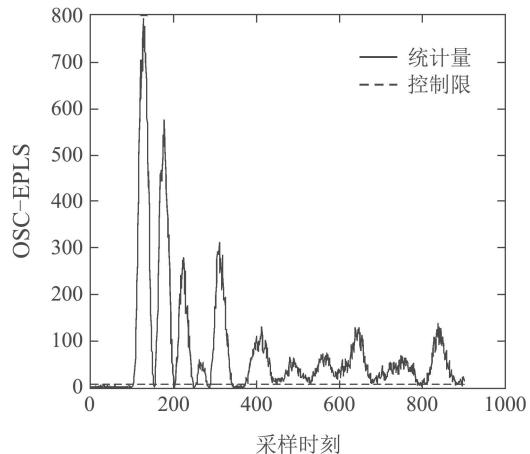


图3 OSC-EPLS的IDV(1)故障检测结果

Fig. 3 The IDV (1) fault detection result of OSC-EPLS

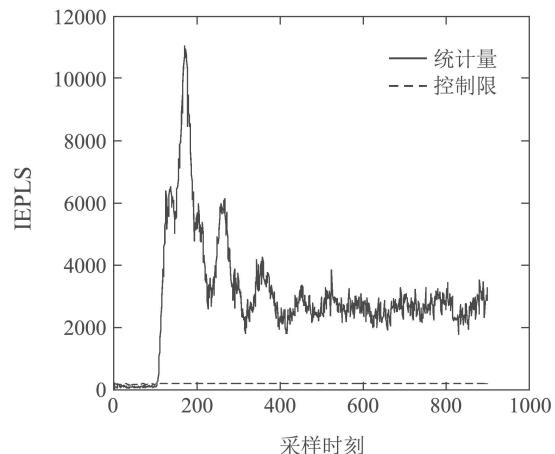
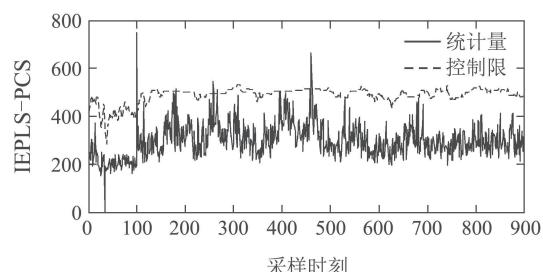


图4 IEPLS的IDV(1)故障检测结果

Fig. 4 The IDV(1) fault detection result of IEPLS

图4是所提的IEPLS算法的IDV(1)故障检测图,从图中可以看出,IEPLS设计的局部动态控制限在测试中当包含正常数据时可以准确根据正常数据特点而变化,其误报率大大降低;另一方面,在100~900故障数据区段IEPLS的漏报率很低,质量相关故障均被IEPLS算法有效检测报警,相比EPLS和OSC-EPLS算法而言其质量相关故障检测性能有了很大提高。

图5是IEPLS对质量无关故障IDV(4)的检测结果,从图中可以看出在故障发生后,IEPLS的质量相关子空间没有检测到质量无关故障的发生;而在质量无关空间中,IEPLS算法准确检测出了质量无关故障的发生,验证了IEPLS检测质量无关故障的有效性。



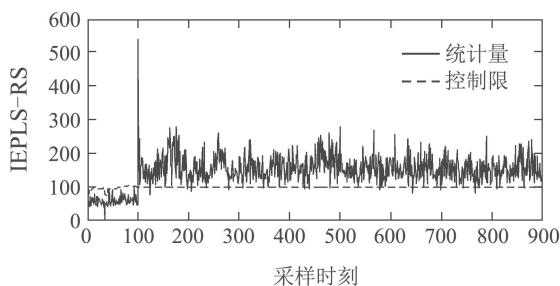


图 5 IEPLS 的 IDV(4) 故障检测结果  
Fig. 5 The IDV(4) fault detection result of IEPLS

表1为3种算法所有质量相关故障的检测率。从表1中可以看出IEPLS对质量相关故障的检测性能明显优于EPLS和OSC-EPLS。表1中加粗部分为3种算法中检测率最高的一组,可以看出IEPLS除了在IDV(5)算法的检测率低于OSC-EPLS算法以外,其余各组的

质量相关故障检测率均高于EPLS和OSC-EPLS。其中,IEPLS算法的IDV(1),IDV(6)对质量相关故障实现了100%监测报警。

表2中给出了3种算法的质量相关故障误报率,加粗部分为3种算法中误报率最高的一组,可以看出EPLS算法和OSC-EPLS算法均存在较高的误报率。3种算法中OSC-EPLS算法的误报率较高,其中IDV(6)的质量相关故障误报率高达47.3%,可以看出OSC-EPLS算法虽然提高了质量相关故障检测率,但是其质量相关故障误报率也明显提高。本文所提的IEPLS算法除了在IDV(5)算法的误报率高于EPLS和OSC-EPLS算法0.9%以外,其余各组的质量相关故障误报率均低于EPLS或OSC-EPLS。最后,根据多次实验得出的实验数据可以看出本文所提的IEPLS算法的质量相关故障检测性能明显提高。

表 1 TE 过程中 3 种算法的质量相关故障检测率

Table 1 Quality-related fault detection rate of three algorithms in TE process

故障编号	已知的过程故障		故障检测率/%		
	故障描述		EPLS	OSC-EPLS	IEPLS
IDV(1)	A/C供料比故障, B浓度恒定(管道4)	90.51	91.26	<b>100.00</b>	
IDV(2)	B浓度故障, A/C供料比恒定(管道4)	88.26	99.13	<b>99.33</b>	
IDV(5)	冷凝器冷却水入口温度变化	<b>99.87</b>	99.75	47.17	
IDV(6)	A供料损失(管道1)	99.00	98.00	<b>100.00</b>	
IDV(7)	物料C压力损失(流4)	41.07	64.79	<b>92.41</b>	
IDV(8)	A, B, C供料浓度变化(管道4)	67.80	91.26	<b>98.20</b>	
IDV(10)	C供料温度变化	45.94	49.31	<b>55.35</b>	
IDV(12)	压缩器冷却水内部温度变化	84.02	84.89	<b>98.85</b>	
IDV(13)	反应器动力学变化	90.26	94.01	<b>96.60</b>	

表 2 TE 过程中 3 种算法的质量相关故障误报率

Table 2 Quality-related fault false alarm rate of three algorithms in TE process

故障编号	已知的过程故障		故障误报率/%		
	故障描述		EPLS	OSC-EPLS	IEPLS
IDV(1)	A/C供料比故障, B浓度恒定(管道4)	0	0	0	
IDV(2)	B浓度故障, A/C供料比恒定(管道4)	1.00	<b>28.00</b>	4.00	
IDV(5)	冷凝器冷却水入口温度变化	0	0	<b>0.90</b>	
IDV(6)	A供料损失(管道1)	0	<b>47.30</b>	6.00	
IDV(7)	物料C压力损失(流4)	2.00	<b>25.00</b>	0	
IDV(8)	A, B, C供料浓度变化(管道4)	7.00	<b>32.00</b>	4.20	
IDV(10)	C供料温度变化	0	<b>5.00</b>	3.00	
IDV(12)	压缩器冷却水内部温度变化	<b>14.00</b>	1.00	4.60	
IDV(13)	反应器动力学变化	2.00	<b>12.00</b>	1.10	

表3中给出了3种算法的质量无关故障误报率,加粗部分为3种算法中误报率最高的一组。可以看出EPLS算法对质量无关故障的检测性能整体较差,而OSC-EPLS算法的质量无关故障检测性能明显优于

EPLS,其中IDV(9)的质量无关故障误报率仅0.87%,可以看出OSC-EPLS算法明显降低了质量无关故障的误报率。本文所提的IEPLS算法除了在IDV(3),IDV(9)和IDV(11)的误报率高于OSC-EPLS算法0.98%,

5.8%和7.34%以外,其余各组的质量无关故障误报率均低于OSC-EPLS。最后,根据多次实验得出的实验数据可以看出本文所提的IEPLS算法在保持质量无关

故障误报率较低的情况下,其质量相关故障检测性能明显提高,对于工业过程故障检测而言具有更好的实际应用效果。

表3 TE过程中3种算法的质量无关故障误报率

Table 3 Quality-unrelated fault false alarm rate of three algorithms in TE process

故障编号	已知的过程故障 故障描述	故障误报率/%		
		EPLS	OSC-EPLS	IEPLS
IDV(3)	D供料温度(管道2)	<b>13.61</b>	3.62	4.60
IDV(4)	反应冷却水内部温度	<b>10.99</b>	3.37	2.12
IDV(9)	D供料温度	<b>7.62</b>	0.87	6.67
IDV(11)	反应冷却水内部水温度	10.24	6.87	<b>14.21</b>
IDV(15)	压缩冷凝水温度	<b>10.49</b>	5.24	3.23

### 5.3 基于IEPLS的故障诊断方法

在本节中,分别使用基于传统的贡献图方法和基于新息贡献率方法对IDV(8)进行故障诊断。当故障8数据到来后,管道4中出现了ABC供料比变化的随机故障,直接影响的是过程变量4即物料A和C流量,这时由于系统的闭环反馈控制原因,其他的变量如质量变量和操作变量26也受到了该故障的影响。图6所示为使用传统贡献图方法对IDV(8)进行故障诊断的结果,可以看出当故障8数据到来后,传统贡献图的诊断结果表明变量16,18~20,31做出了主要贡献。然而,传统贡献图没有诊断出变量4的贡献,这与实际的故障情况并不符合,由此可以看出基于传统的贡献图故障诊断方法并不能够完全准确诊断出故障变量。

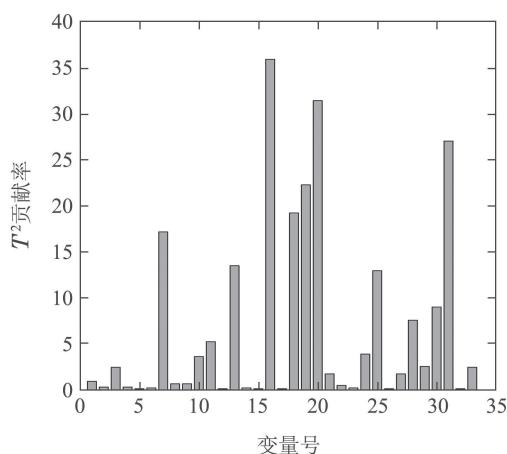


图6 基于贡献图的IDV(8)故障诊断结果

Fig. 6 The fault diagnosis result of IDV(8) based on contribution plot

图7所示为使用新息贡献率方法对故障8进行故障诊断的结果,从图中可以看出IEPLS对故障8的诊断结果显示出了变量4对故障8的主要贡献。另外,IEPLS同时也诊断出了变量8~11,20~22,26~28对故障8的贡献结果,此诊断结果与真实情况是吻合的。实验结果

表明,所提的IEPLS故障诊断算法明显优于传统贡献图方法的诊断性能。

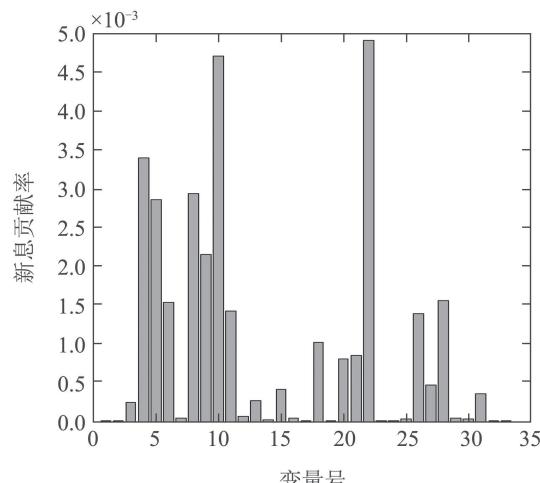


图7 基于IEPLS的IDV(8)故障诊断图结果

Fig. 7 The fault diagnosis result of IDV(8) based on IEPLS

## 6 结论

本文通过在EPLS的质量相关子空间设计新的统计量和控制限,并将新息贡献率应用在该空间提出了一种基于IEPLS的质量相关故障诊断方法。该方法首先使用正常数据获得模型参数,然后对输入数据进行空间分解。通过在分解后的质量相关子空间设计新的统计量和控制限检测质量相关故障,接着结合新息贡献率对已发生故障的各个变量进行定位,最后得出发生故障的变量信息。通过TEP仿真平台与EPLS和OSC-EPLS的实验结果对比分析,所提IEPLS算法不仅提高了对质量相关故障的检测性能,而且IEPLS对故障变量的定位性能也明显提高。值得一提的是,IEPLS算法的质量无关空间虽然对产品质量的影响可以忽略不计,但是提高其监控性能将是本课题继续研究的一个方向。

**参考文献:**

- [1] JIANG Y C, YIN S, KAYNAK O. Data-driven monitoring and safety control of industrial cyber-physical systems: basics and beyond. *IEEE Access*, 2018, 6(3): 1 – 2.
- [2] LUO L J, BAO S Y, MAO J F. Adaptive selection of latent variables for process monitoring. *Industrial & Engineering Chemistry Research*, 2019, 58(21): 9075 – 9086.
- [3] PRATS J M, FERRER A. Statistical process control based on multivariate image analysis: a new proposal for monitoring and defect detection. *Computers & Chemical Engineering*, 2014, 71(4): 501 – 511.
- [4] DING X H, HE L H, CARIN L. Bayesian robust principal component analysis. *IEEE Transactions on Image Processing*, 2011, 20(12): 3419 – 3430.
- [5] ARIFF M A, PAL B C. Coherency identification in interconnected power system—an independent component analysis approach. *IEEE Transactions on Power Systems*, 2013, 28(2): 1747 – 1755.
- [6] WONG K H, VALENTINA R N, LI K M, et al. The quality control of two pueraria species using raman spectroscopy coupled with partial least squares analysis. *Journal of Raman Spectroscopy*, 2015, 46(4): 361 – 368.
- [7] WANG Wei, CHAI Tianyou, ZHAO Lijie. Dynamic partial least squares modeling with recurrent neural networks of stable learning. *Control Theory & Applications*, 2012, 29(3): 337 – 341.  
(王魏, 柴天佑, 赵立杰. 带有稳定学习的递归神经网络动态偏最小二乘建模. 控制理论与应用, 2012, 29(3): 337 – 341.)
- [8] TRYGG J, WOLD S. Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics: A Journal of the Chemometrics Society*, 2002, 16(3): 119 – 128.
- [9] YIN S, DING S X, ZHANG P, et al. Study on modifications of PLS approach for process monitoring. *Proceedings of the 18th IFAC World Congress*. Milano: IFAC, 2011, 44(1): 12389 – 12394.
- [10] ZHANG K, PENG K X, YOU B, et al. Quality-relevant fault monitoring based on efficient projection to latent structures with application to hot strip mill process. *IET Control Theory and Applications*, 2015, 9(7): 1135 – 1145.
- [11] KONG Xiangyu, LUO Jiayu, ZHANG Qi, et al. Quality-related fault detection method based on orthogonal signal correction and efficient PLS. *Control and Decision*, 2020, 35(5): 1167 – 1174.  
(孔祥玉, 罗家宇, 张琪, 等. 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. 控制与决策, 2020, 35(5): 1167 – 1174.)
- [12] LI G, QIN S Z, JI Y D, et al. Total PLS based contribution plots for fault diagnosis. *Acta Automatica Sinica*, 2009, 35(6): 759 – 765.
- [13] WEN Chenglin, HU Yucheng. Fault diagnosis based on information incremental matrix. *Acta Automatica Sinica*, 2012, 38(5): 1 – 8.
- [14] BJORCK A, INDAHL U G. Fast and stable partial least squares modelling: a benchmark study with theoretical comments. *Journal of Chemometrics*, 2017, 31(8): 2883 – 2898.
- [15] XU G H, HU G R, SONG Y. Singular value decomposition(SVD)-based nonlinear filtering algorithm. *Journal of Shanghai Jiaotong University*, 2005, 31(1): 1 – 9.
- [16] QIN S J. Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, 2003, 17(9): 480 – 502.
- [17] LI G, QIN S J, ZHOU D H. Geometric properties of partial least squares for process monitoring. *Automatica*, 2010, 46(1): 204 – 210.
- [18] WOLD S, ANTTI H, LINDGREN F, et al. Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 1998, 44(1): 175 – 185.
- [19] FEARN T. On orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems*, 2000, 50(1): 47 – 52.
- [20] YANG H Y. *Advanced prognosis and health management of aircraft and spacecraft subsystems*. Massachusetts: Massachusetts Institute of Technology, 2000.
- [21] JIAO J F, HAN Y, WANG G. A quality-related fault detection approach based on dynamic least squares for process monitoring. *IEEE Transactions on Industrial Electronics*, 2016, 63(4): 2625 – 2632.
- [22] YAN Aijun, WANG Yingjie, WANG Dianhui. Fault diagnosis method using learning case-based reasoning for Tennessee-Eastman process. *Control Theory & Applications*, 2017, 34(9): 1179 – 1184.  
(严爱军, 王英杰, 王殿辉. Tennessee-Eastman过程的学习型案例推理故障诊断方法. 控制理论与应用, 2017, 34(9): 1179 – 1184.)
- [23] WANG G, YIN S. Enhanced quality-related fault detection approach based on orthogonal signal correction and modified PLS. *IEEE Transactions on Industrial Informatics*, 2015, 11(2): 398 – 405.

**作者简介:**

孔祥玉 教授, 博士, 博士生导师, 目前研究方向为随机系统分析、非线性系统建模及其应用、复杂系统故障诊断等, E-mail: xiangyu\_kong01@163.com;

解 建 硕士研究生, 目前研究方向为复杂系统特征提取和故障检测与诊断, E-mail: hdxiejian@163.com;

罗家宇 博士研究生, 目前研究方向为复杂系统下数据驱动的故障监测与诊断等, E-mail: 540629964@qq.com;

李 强 硕士研究生, 目前研究方向为特征提取和多元统计过程监控, E-mail: 1058097905@qq.com.