

# 基于改进动态主元分析在半实物 仿真系统中的研究

高 强, 常 勇

(天津理工大学自动化学院天津市复杂系统控制理论及应用重点实验室, 天津 300384)

**摘 要:** 为了实现数据驱动技术在工业中的实际应用, 开发了以蒸馏塔作为被控对象的半实物仿真系统, 将数据驱动方法应用到流程工业半实物仿真系统. 针对动态主元分析方法存在的计算负荷大, 计算效率低的问题, 提出了一种改进动态主元分析方法, 利用不可区分度和交叉程度去除众多变量中的不相关变量或相关度较小的变量, 减少数据量. 针对系统中的典型故障, 数据驱动方法能够检测出半实物仿真系统中的异常, 而且与传统动态主元分析比较, 改进算法降低漏报率和误报率, 提高诊断可靠性, 并且能及时检测出生产过程的微小故障.

**关键词:** 数据驱动; 蒸馏塔; 半实物仿真; 动态主元分析; 不可区分度

**中图分类号:** TP277      **文献标识码:** A      **文章编号:** 0372-2112 (2017)03-0565-05

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2017.03.009

## The Research of Hardware-in-the-Loop Simulation System Based on the Improved Dynamic Principal Component Analysis

GAO Qiang, CHANG Yong

(Tianjin Key Laboratory for Control Theory & Applications in Complicated Systems, Tianjin University of Technology, Tianjin 300384, China)

**Abstract:** In order to apply the data-driven technology in industry field, with the distillation column as one of the controlled objects, the hardware-in-the-loop simulation system is developed, and the data driven methods are applied to the process industry hardware-in-the-loop simulation system. Aiming at the huge computation and low efficiency of the dynamic principal component analysis method, an improved dynamic principal component analysis method is proposed, which remove the irrelevant variables or low relevant variables, reduce the amount of data, and improve the diagnostic efficiency by indiscernibility and the cross-degree. For the typical faults of the systems, the application result shows that the data-driven methods can detect the fault of hardware-in-the-loop simulation system, and compared with the traditional dynamic principal component, the simulation results show that the proposed method is more reliable with lower missing rate, and lower false rate, and in addition, it can detect the small process faults timely.

**Key words:** data driven; distillation column; hardware-in-the-loop simulation; dynamic principal component analysis (DPCA); indiscernibility

### 1 引言

随着工业 4.0 的提出和发展, 现代工业过程自动化程度和复杂性越来越高<sup>[1-4]</sup>, 工业生产过程中安全性和可靠性问题成为人们关注的焦点<sup>[5,6]</sup>. 过程监控在保证工业过程的安全生产和提高产品质量方面发挥重要作用, 因此过程监控技术成为众多学者研究的热门. 在监控技术领域, 数据驱动方法凭借: (1) 大数据时代的到

来<sup>[7]</sup>, 各种智能仪表被广泛使用, 海量的过程数据被收集和归档, 这些数据包含系统的状态信息; (2) 系统复杂性越来越高, 很难建立系统模型, 而数据驱动方法无需建立精确的物理模型, 这两点优势被广泛应用于流程工业的过程监控和故障诊断<sup>[8-10]</sup>.

主元分析法 (Principal Component Analysis, PCA) 是数据驱动方法中应用最广泛的一种多元统计方法, PCA 能有效监视生产过程, 但 PCA 是一种静态方法, 其应用

的前提条件是各变量不存在序列相关性. 实际流程工业过程存在很强的动态特性, 根本无法满足这一条件, 导致 PCA 在故障检测时, 误报和漏报的概率增加. 针对工业过程动态特性的问题, Ku 等人<sup>[11,12]</sup>通过引入变量观测值构造增广矩阵, 然后对增广矩阵进行 PCA 建模来解决这一问题, 该方法虽然考虑了工业过程的动态特性, 提取了变量之间的相关性, 但是存在明显的不足, 增广矩阵极大的增加了主元数, 增加了计算量, 降低了计算效率. 本文提出基于不可区分度的动态主元, 利用不可区分度去除不相关变量和相关度较小的变量, 保留相关度较大的变量, 减少数据维度, 降低误报率和漏报率, 提高诊断可靠性.

## 2 改进动态主元分析

### 2.1 不可区分度和交叉程度

在流程工业中, 一个复杂系统存在很多过程和操作变量, 这些变量的过程值被采集存储, 在传统多元统计过程监控过程中, 监控变量的选取往往考虑所有的过程变量, 给过程监控带来很多不便, 实际故障中, 对于某类特定故障来说, 只有很少部分变量受到影响, 很多变量没有太大变化, 正常数据跟故障数据区别不大, 完全可以不用考虑, 所以本文提出了基于不可区分度和交叉程度的动态主元分析方法, 该方法利用不可区分度和交叉程度两个参数, 去除样本集中的不相关变量, 降低数据维度, 简化算法的计算量, 提高故障诊断效率和精度.

首先获取一组正常样本数据  $\mathbf{Z}$ , 对正常数据的每一个变量求平均值

$$\hat{z}_i = \frac{1}{n} \sum_j z_{ji}, \quad i = 1, 2, \dots, m \quad (1)$$

其中,  $n$  是样本个数,  $m$  是变量个数.

然后对某个故障数据  $\mathbf{Z}'$  求平均值

$$\hat{z}'_i = \frac{1}{n} \sum_j z'_{ji}, \quad i = 1, 2, \dots, m \quad (2)$$

确定阈值

$$\delta_i = \frac{\hat{z}_i + \hat{z}'_i}{2} \quad (3)$$

根据阈值可以得到正常数据中错分的数据和故障数据中错分的数据, 这里统称为样本错分点集合

$$U_i = \{u_{i1}, u_{i2}, \dots, u_{ik}\}, \quad 0 \leq k \leq 2n, \quad i = 1, 2, \dots, m$$

此处  $m$  是指数据中变量数,  $k$  是正常数据与故障数据的错分点个数, 对于不同监控变量,  $k$  不相同.

不可区分度

$$\mu_i = \sum_j^k \left| u_{ij} - \frac{\hat{z}_i + \hat{z}'_i}{2} \right|, \quad u_{ij} \in U_i \quad (4)$$

交叉程度

$$\eta_i = \frac{k}{2n} \times 100\%, \quad i = 1, 2, \dots, m \quad (5)$$

对于每个故障, 根据  $\mu_i$  和  $\eta_i$  的值, 选择  $\mu_i$  和  $\eta_i$  值比较小的变量, 去除  $\mu_i$  和  $\eta_i$  比较大的变量使数据更精简, 诊断效果更好. 对于同类的已知故障, 分别采用不可区分度和交叉程度对数据变量进行筛选, 获取  $\mu_i$  和  $\eta_i$  比较小的变量, 组成新的训练样本数据矩阵  $\mathbf{X}$ , 利用筛选后的样本数据  $\mathbf{X}$  进行建模. 对于故障数据同样选取筛选过后的变量进行检测.

### 2.2 动态特性分析

经典主元分析方法是一种静态的方法, 该方法应用的前提条件是: 某一时刻的观测数据与过去时刻的观测数据是统计独立的, 即它们之间序列无关. 对于典型的流程工业过程, 该条件只对长时间采样间隔有效, 为了更好的监视生产过程, 降低损失, 需要更短的采样时间间隔, 因此序列相关在实际工业过程数据中是普遍存在的特点, 针对工业数据这一特点, Ku 等人提出了动态主元分析方法, 通过增广时间序列来扩大数据矩阵, 将数据自相关特性纳入到相应的增广矩阵中, 即: 对于 2.1 节中得到的新矩阵

$$\mathbf{X} = [x_1, x_2, \dots, x_m], \quad x_i \in \mathbb{R}^n$$

通过增加前  $h$  个时刻的观测数据<sup>[11,13]</sup>, 得到增广矩阵为:

$$\mathbf{X}(h) = \begin{bmatrix} \mathbf{x}_t^T & \mathbf{x}_{t-1}^T & \dots & \mathbf{x}_{t-h}^T \\ \mathbf{x}_{t-1}^T & \mathbf{x}_{t-2}^T & \dots & \mathbf{x}_{t-h-1}^T \\ \dots & \dots & \dots & \dots \\ \mathbf{x}_{t+h-n}^T & \mathbf{x}_{t+h-n-1}^T & \dots & \mathbf{x}_{t-n}^T \end{bmatrix} \quad (6)$$

其中,  $\mathbf{x}_t^T$  是  $t$  时刻得到的  $m$  维观测向量, 将式(6)进行 PCA 建模, 静态和动态的关系都呈现在残差子空间中, 残差子空间的  $Q$  统计量是真正相互独立的.

### 2.3 改进算法监控

本文提出的改进动态主元分析方法具有处理数据之间动态关系的能力, 根据正常训练样本和已知故障数据集, 利用不可区分度和交叉程度, 提取系统中相关的变量, 剔除不相关变量, 简化诊断样本数据, 提高算法诊断效率, 改进算法诊断流程如图 1 所示.

#### (1) 离线训练

**步骤 1** 从数据库中提取出训练数据和一些已知的故障数据集.

**步骤 2** 根据训练数据和故障数据集, 通过式(4)、式(5)计算不可区分度和交叉程度.

**步骤 3** 根据不可区分度和交叉程度去除相关程度较小或者不相关变量, 确定监控变量的个数.

**步骤 4** 获得新的训练矩阵  $\mathbf{X}$ .

**步骤 5** 确定滞后  $h$ , 通过观测值扩展对筛选后的

新矩阵构建增广矩阵  $X(s)$ .

步骤 6 对增广矩阵  $X(s)$  建立 PCA 模型.

步骤 7 计算控制限  $T_{\alpha}^2$  和  $SPE_{\alpha}$ .

(2) 在线监控

步骤 1 提取系统实时运行的过程数据, 获得检测样本.

步骤 2 根据离线训练所确定的变量, 去除相关程度较小或者不相关的变量构成新的测试样本数据.

步骤 3 对筛选后的新样本矩阵构建增广矩阵  $X_{new}$ .

步骤 4 计算测试样本数据的  $T_{new}^2$  和  $SPE_{new}$ .

步骤 5 将  $T_{new}^2$  统计量和  $SPE_{new}$  统计量与控制限  $T_{\alpha}^2$  和  $SPE_{\alpha}$  比较.

步骤 6 若计算的  $T_{new}^2$  统计量和  $SPE_{new}$  统计量超出控制限, 则有故障发生, 未超出控制限, 则断定无故障发生.

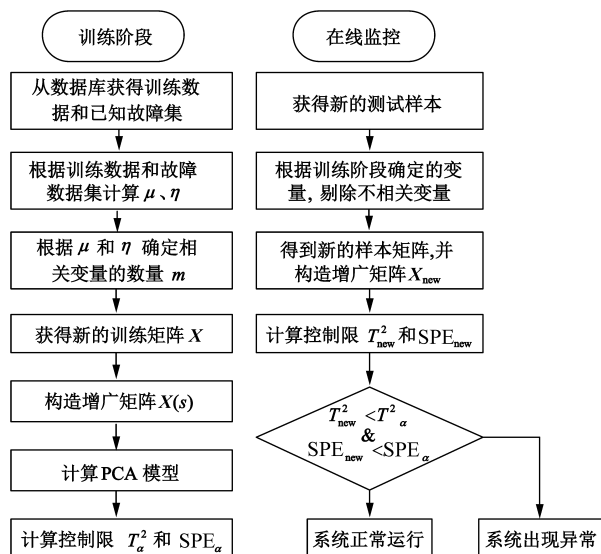


图1 基于改进算法的故障诊断流程图

### 3 半实物仿真系统

#### 3.1 半实物仿真介绍

集散控制系统(Distributed Control Systems, DCS)其“集中管理, 分散控制”的理念得到广泛的应用和发展, 已在工业过程控制领域得到基本普及, 在石油化工、电力等行业中更是起到极其重要的作用. 但在科研中, 由于被控对象规模庞大、生产原料复杂、设备维护困难、生产过程中存在高温和高压的危险等问题, 理论研究人员针对 DCS 系统故障诊断的实验研究难以开展. 本文提出了一种利用半实物仿真技术, 将硬件控制器与计算机仿真模型相结合的 DCS 系统实验平台, 用以研究数据驱动故障诊断.

系统采用真实的 DCS 系统构成, 以真实的硬件控制器与虚拟的计算机仿真模型相结合的方式构成故障诊断技术的研究平台与验证环境. 采用西门子 SIMATIC PCS7 软件作为开发环境, 以西门子公司开发的蒸馏塔单元模板作为 DCS 系统实验平台的被控对象. DCS 系统实验平台整体结构如图 2 所示. 该平台增加了故障诊断站, 实现系统的实时诊断和监控, 保证系统的正常运行.

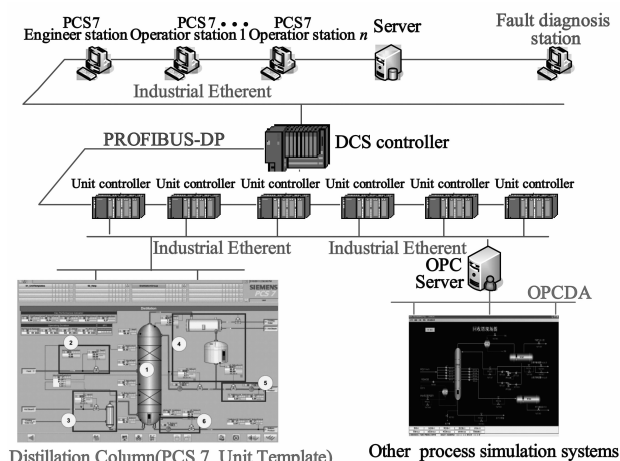


图2 流程工业半实物仿真系统

#### 3.2 半实物仿真在故障诊断研究中的优势

在故障诊断研究中, 国内外很多研究人员只能对纯仿真过程进行研究, 比如田纳西伊斯曼过程(Tennessee-Eastman, TE)<sup>[14]</sup>. 该过程是 Down 和 Vogel 两位工程师根据工厂实际生产开发的仿真模拟系统, 虽然 TE 仿真过程很接近实际生产, 但依然存在弊端: 该过程是离线的, 系统故障通过算法设定实现, 故障固定且无法改变. 而且该过程只设定了 21 个故障, 故障种类和个数偏少, 然而, 在实际流程工业生产中, 故障发生是随机的、多种多样的.

半实物仿真系统采用真实的控制器, 通过操作界面可以设定阀门、浓度、流量、温度等各种各样的故障, 对于同类故障, 可以设定不同的故障参数, 比较符合实际工厂中的随机故障. 而且研究者可以通过操作界面直观的观测故障及其影响.

### 4 仿真研究

蒸馏塔作为化工与炼油企业必不可少的重要装置, 一旦出现阀门、浓度等故障, 就会给企业生产带来很大的损失, 因此蒸馏塔的故障检测与诊断成为化工生产的重要环节. 笔者选取了 14 个监控变量, 采集了 960 组正常工况下的采样数据用于离线训练建立模型, 设定了 6 组故障如表 1 所示, 主要包括三类, 分别是阀门故障、浓度故障、流量故障, 并分别采集了 960 组故障数

据用于研究.

表2记录了DPCA与改进DPCA两种方法对于6种故障的诊断误报率,通过表2可以看出,改进算法比DPCA误报率明显降低.

故障1、2、3都是阀门故障,对于阀门类故障,DPCA算法能及时检测出故障,但是有较大的误报率,如表2所示,改进算法明显有较低的漏报率.

表1 蒸馏塔故障

序号	故障描述	位置
1	塔顶出口阀故障参数:50%~45%	蒸馏塔
2	冷水阀门故障参数:60%~55%	回流冷凝器
3	热蒸汽阀门故障参数:50%~45%	蒸发器
4	进料浓度故障参数:0.5~0.6	进料器
5	进料浓度故障参数:0.5~0.51	进料器
6	进料流量故障参数:20 L/min~20.1 L/min	进料器

表2 故障检测误报率(单位:%)

故障	DPCA		改进DPCA	
	$T^2$	SPE	$T^2$	SPE
1	18.68	12.45	5.49	8.79
2	12.09	12.45	1	7.64
3	15.46	12.71	1.77	4.24
4	6.36	9.54	2.47	0
5	8.36	18.91	1.45	2.55
6	7.64	9.82	1.45	2.91

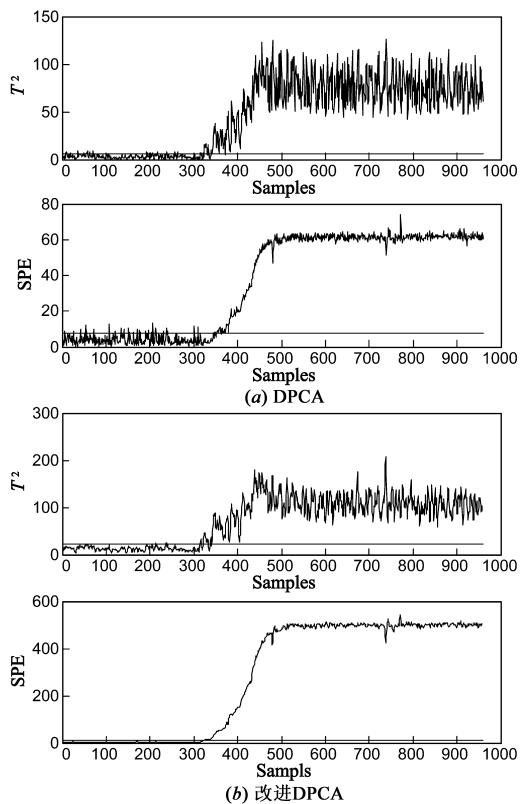


图3 故障4的检测结果

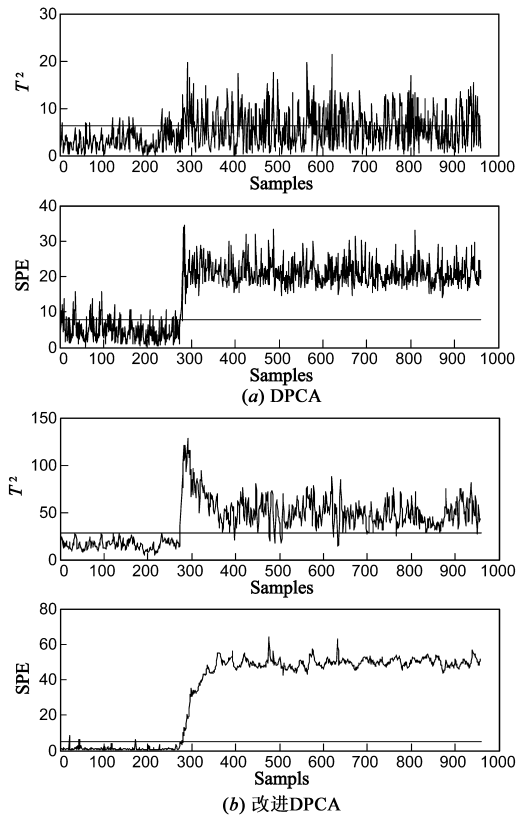


图4 故障5的检测结果

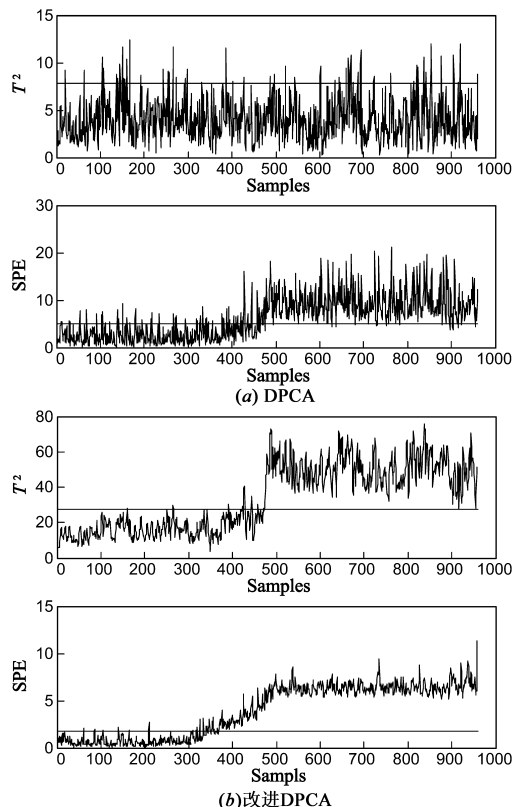


图5 故障6的检测结果

故障 4 是进料浓度发生较大变化,浓度是关乎工业生产的重要指标,一旦浓度受到较大扰动,产品质量就会大大降低,针对浓度故障,DPCA 算法能够很好检测出系统故障,并且有较低的误报率,但是出现一定的延时,而改进 DPCA 算法比 DPCA 具有更低的误报率和更快的诊断速度,相比 DPCA 具有更好的性能,如图 3 所示。

故障 5 也是进料浓度故障,故障 6 是流量故障。相比故障 4,故障 5 是小扰动故障浓度值变化 0.01,故障 6 也是小扰动故障,流量值变化 0.1,对于此类小扰动故障,DCS 系统本身的报警系统无法检测出,DPCA 算法的 SPE 统计量能够检测出故障,但是  $T^2$  统计量漏报率太高,无法检测出故障,而改进算法的 SPE 统计量和  $T^2$  统计量则都能及时的检测到故障,而且降低的误报率,表现出更好的诊断性能,如图 4、5 所示。

## 5 结论

集散控制系统(DCS)在工业中已广泛应用。为在未来实现基于数据驱动技术的故障诊断在工业生产中的全面应用,我们开发了流程工业半实物仿真系统,成功将数据驱动应用到半实物仿真系统中,实现了故障诊断从纯仿真对象到半实物的研究和应用,进一步推进了数据驱动技术在实际系统中的应用。提出了改进 DP-CA 算法,该方法不仅考虑了过程工业的动态特性,而且利用不可区分度和交叉程度能去除系统中的不相关变量和相关程度较小的变量,降低计算量,提高计算效率。将改进算法与传统方法应用到蒸馏塔故障诊断,实验结果表明改进算法降低了误报率和漏报率、诊断速率更快、并且能够检测出微小故障。

## 参考文献

- [1] Qin S J. Process data analytics in the era of big data[J]. *AIChE Journal*, 2014, 60(9): 3092 - 3100.
- [2] Dai X, Gao Z. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis[J]. *IEEE Transactions on Industrial Informatics*, 2013, 9(4): 2226 - 2238.
- [3] Wang G, Yin S. Quality-related fault detection approach based on orthogonal signal correction and modified PLS [J]. *IEEE Transactions on Industrial Informatics*, 2015, 11(2): 398 - 405.
- [4] Peng K, Zhang K, Li G. Online contribution rate based fault diagnosis for nonlinear industrial processes[J]. *Acta Automatica Sinica*, 2014, 40(3): 423 - 430.
- [5] 黄江涛,王明辉,李武劲,等. 基于动态权值的多分类器故障诊断系统[J]. *电子学报*, 2012, 40(4): 734 - 738. Huang Jiang-tao, Wang Ming-hui, Li Wu-jin, et al. Multiple classifier fault diagnosis system based on dynamic weight[J]. *Acta Electronica Sinica*, 2012, 40(4): 734 - 738. (in Chinese)
- [6] Zhao F, Koutsoukos X, Haussecker H, et al. Monitoring and fault diagnosis of hybrid systems[J]. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 2005, 35(6): 1225 - 1240.
- [7] 李玮,张大方,黄昆,等. 面向大数据处理的高精度多维计数布鲁姆过滤器[J]. *电子学报*, 2015, 43(04): 652 - 657. Li Wei, Zhang Da-fang, Huang Kun, et al. Accurate multi-dimension counting bloom filter for big data processing[J]. *Acta Electronica Sinica*, 2015, 43(04): 652 - 657. (in Chinese)
- [8] 周福娜,文成林,等. 基于指定元分析的多级相对微小故障诊断方法[J]. *电子学报*, 2010, 38(8): 1874 - 1879. Zhou Fu-na, Wen Cheng-lin, et al. DCA based multi-level small fault diagnosis method [J]. *Acta Electronica Sinica*, 2010, 38(8): 1874 - 1879. (in Chinese)
- [9] 汪楚娇,夏士雄,牛强. 免疫粒子群算法及其在矿井提升机故障诊断中的应用[J]. *电子学报*, 2010, 38(2A): 94 - 98. Wang Chu-jiao, Xia Shi-xiong, Niu Qiang. Artificial immune particle swarm optimization for fault diagnosis of mine hoist[J]. *Acta Electronica Sinica*, 2010, 38(2A): 94 - 98. (in Chinese)
- [10] Wang G, Yin S, Kaynak O. An LWPR-based data-driven fault detection approach for nonlinear process monitoring [J]. *IEEE Transactions on Industrial Informatics*, 2014, 10(4): 2016 - 2023.
- [11] 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.
- [12] Li G, Qin S J, Zhou D. A new method of dynamic latent-variable modeling for process monitoring[J]. *IEEE Transactions on Industrial Electronics*, 2014, 61(11): 6438 - 6445.
- [13] Choi S W, Lee I. Nonlinear dynamic process monitoring based on dynamic kernel PCA [J]. *Chemical Engineering Science*, 2004, 59(24): 5897 - 5908.
- [14] Downs J J, Vogel E F. A plant-wide industrial process control problem[J]. *Computers & Chemical Engineering*, 1993, 17(3): 245 - 255.

## 作者简介



高 强 男,1968 年 2 月出生于天津,现任天津理工大学自动化学院副院长。主要研究方向:系统辨识与优化控制;非线性系统建模、分析与控制;故障诊断与容错控制;过程控制系统技术及先进控制。

常 勇 男,1992 年 1 月出生于安徽临泉县,现为天津理工大学研究生,主要研究故障诊断,过程控制及先进系统。  
E-mail: chang133122332@ sina. com