

文章编号: 1002-0411(2001)04-359-06

# 非线性主元分析故障检测和诊断方法及应用

赵立杰 王 纲 李 元

(沈阳化工学院高级过程控制中心 沈阳 110021)

**摘 要:** 本文针对间歇生产过程的特点, 基于多方向主元分析方法(MPCA)和非线性理论, 提出了一种非线性多元统计分析方法——最小窗口方法, 该方法突破了 MPCA 方法单模型、线性化的建模方式, 创新性地构造了适合间歇生产过程特点的多模型结构非线性建模方法, 并侧重于在线间歇过程性能监视和故障诊断的实时性, 消除了预报未来测量值带来的误差, 提高了过程性能监视和故障诊断的准确率. 本文详细地讨论了最小窗口 PCA 建模方法、原理、应用实例. 基于该方法设计的聚氯乙烯生产过程性能监视和故障诊断系统充分验证了该方法的有效性.

**关键词:** 多元统计分析; 主元分析; 间歇过程; 故障检测和诊断

中图分类号: TP277 O212.4

文献标识码: B

## STUDY OF A NONLINEAR PCA FAULT DETECTION AND DIAGNOSIS METHOD

ZHAO Li-jie WANG Gang LI Yuan

(Advanced Process Control Center, Shenyang Institute of Chemical Technology, Shenyang 110021)

**Abstract:** In view of characteristics of batch process, this paper proposes a new real time and nonlinear minimum window principal component analysis (MWPCA) method based on multiway principal component analysis and nonlinear theory. MWPCA method breaks through linear MPCA modeling with single model structure and innovates in a nonlinear multimodel structure for batch process modeling. The method emphasizes particularly on real-time characteristic in on-line batch process performance monitoring and eliminates error caused by predicting future measurements of process variables, increases the accuracy of process performance monitoring and fault diagnosis. MWPCA modeling procedures, principle and its application are discussed in detail. PVC process performance and fault diagnosis system based on MWPCA method verify the validity of the method.

**Keywords:** multivariate statistical, PCA (Principal Component Analysis), batch processes, FDD (Fault Detection and Diagnosis)

### 1 引言(Introduction)

故障检测和诊断(FDD)技术是近年来国际控制理论研究和实际应用的热点问题之一<sup>[1,2]</sup>, 其中基于多元统计分析方法的故障检测和诊断技术更为国际过程控制领域重视和关注. 尤其是间歇过程, 由于起停频繁、动态特性变化快、时序操作严格等特点, 使实际过程监视和故障诊断更具有挑战性. MPCA<sup>[3]</sup>方法是一种适合间歇生产过程的线性化建模方法. 但是大多数间歇过程往往具有高度复杂的非线性, MPCA 线性化的建模方法往往难以保证诊断的准确性<sup>[4]</sup>. 另外在线间歇生产过程性能监视和故

障诊断的实时性也是 MPCA 方法实际应用时受到限制的重要原因. 本文针对间歇过程生产的特点, 提出了最小窗口 PCA 方法, 解决间歇过程非线性统计建模和故障诊断实时性问题.

### 2 最小窗口 PCA 方法 (MWPCA method)

每次间歇生产过程数据集合构成了一族测量变量的时间轨迹. 多个间歇生产过程的时间轨迹集合可以表示为三维数据矩阵的形式  $X(I \times J \times K)$ . 其中  $I$  代表批次个数,  $J$  代表测量变量,  $K$  代表时间序列. 对于三维数据矩阵  $X$ , MPCA 方法首先将三维

数据矩阵以某种方式展开成二维数据矩阵  $X$ , 然后对新展开的二维数据矩阵做 PCA 分解, 从中抽取间歇过程主要变化信息. 数据矩阵  $X$  或展开矩阵  $X$  被分解为得分向量 Scores 和载荷向量 Loadings 乘积的和, 再加上残差矩阵  $E$ .

$$X = \sum_{r=1}^R t_r \odot p_r + E \quad (1)$$

其中  $t_r$  是得分向量,  $p_r$  是载荷向量,  $R$  是系统中保留的主元个数,  $t_r$  和  $p_r$  的乘积和代表间歇过程主要变化信息, 是批量、过程变量和时间的函数, 残差部分  $E$  表征间歇过程数据中次要变化, 在最小二乘意义上尽可能小. 三维数据矩阵展开成二维数据矩阵后 MPCA 方法数据处理和分析过程等同于普通 PCA 方法<sup>[3]</sup>. 因此 MPCA 是一种线性建模方法. MPCA 方法通过抽取测量变量平均轨迹可以去掉部分非线性, 但是并没有从本质上解决多元统计建模方法的非线性问题.

非线性间歇过程要求非线性的建模方式. 本文提出的最小窗口 PCA 较好地解决了多元统计数学模型的非线性问题. 最小窗口 PCA 建模方法的主要设计思路如下: 首先将大数据空间划分成一系列子数据空间, 然后分别建立子数学统计模型. 最小窗口 PCA 方法具体建模过程如下:

(i) 三维数据矩阵  $X(I \times J \times K)$  按时间坐标轴的方向依次向右展开, 按照时间顺序将三维数据空间划分成一系列子数据空间, 每个子数据空间具有相同的数据结构  $x_k(I, J)$ , 其中  $x_k(I, J)$  表示  $k$  时刻,  $I$

个批次, 过程  $J$  个测量变量的数据集合.

$$x_k(I, J) = X(I, J, K) \quad (2)$$

$$X = [x_1, x_2, \dots, x_k \dots x_k]$$

(ii) 子数据空间量化. 分别计算子数据空间矩阵的均值、方差, 然后依次将子数据空间集合量化成均值为 0, 方差为 1 的数据矩阵.

$$x_k = \frac{x_k - E(x_k)}{\sqrt{\text{Var}(x_k)}} \quad (3)$$

其中,  $E(x_k)$  为第  $k$  个子数据空间的均值,  $\text{Var}(x_k)$  是第  $k$  个子数据空间数据矩阵的方差.

(iii) 子数据空间分解. 相对不同批次同一时刻的测量值阵列对应一个子数据空间. 在子数据空间应用 PCA 方法分别建立子数学统计模型.

$$x_k = \sum_{r=1}^R t_{k,r} \odot p_{k,r} + E_k$$

$$P = [p_1 \ p_2 \ \dots \ p_k \ \dots \ p_k]$$

$$T = [t_1 \ t_2 \ \dots \ t_k \ \dots \ t_k]$$

其中,  $P$  为所有子数据空间的载荷向量,  $p_k$  为第  $k$  个子数据空间的载荷矩阵;  $T$  是所有子数据空间的得分向量,  $t_k$  为第  $k$  个子数据空间的得分向量.  $E_k$  是模型  $k$  的残差矩阵.

MPCA 建模方法实质上是使用正常批次数据测量变量和平均轨迹之间的偏差建立统计模型. 然后, 比较新批次过程测量变量和平均轨迹之间的偏差, 如果偏差和模型偏差统计不一致表明新批次过程是异常批次过程, 产品质量下降.

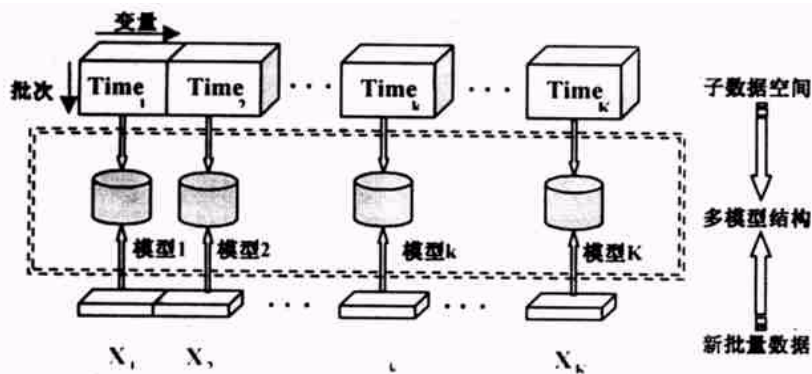


图 1 最小窗口 PCA 建模和监视原理图

Fig. 1 Principle of modeling and detection based on MWPCA

最小窗口 PCA 方法建模和监视原理如图 1 所示. 大型数据空间划分成小的子数据空间后分别建立一系列子数学统计模型, 这种动态建模方式轻松地解决了 MPCA 方法线性建模的局限性, 同时将单模型结构拓展到多模型结构, 所以说最小窗口 PCA 方法是一种非线性的动态多模型结构的建模方法.

### 3 故障检测和诊断(Fault detection and diagnosis)

间歇过程性能监视和故障诊断技术的关键问题之一是在线实时监视和故障诊断策略. MPCA 方法要实现在线实时监视功能的前提条件之一是必须具

备一个完整的测量变量时间轨迹集合. 令当前时刻为  $k$ ,  $k$  时刻测量变量集合  $x_k = [x_{k,1}, x_{k,2}, \dots, x_{k,J}]$ ; 批次过程开始到当前时刻时间序列集合  $X = [x_1, x_1, \dots, x_k]$ ;  $k$  时刻以后测量变量未来时间序列集合  $X_f = [x_{k+1}, x_{k+2}, \dots, x_k]$ , 在间歇反应过程中只能得到当前时刻及其以前测量变量的数据轨迹  $X = [x_1, x_2, \dots, x_k]$ , 未来时刻测量变量的时间轨迹  $X_f = [x_{k+1}, x_{k+2}, \dots, x_k]$  未知, 所以在间歇操作周期内, 只能得到部分数据集合. 只有整个间歇过程结束, 才能得到完整的批次变量时间轨迹. 由于间歇过程操作周期中测量数据的不完整性以及预报未来数据误差的不可避免性, 限制了 MPCA 方法实时在线应用时的精度.

最小窗口 PCA 方法在不需要间歇过程完整数据集合的情况下, 根据当前时刻的数据, 就可以实现间歇过程在线实时监视和故障诊断功能. 最小窗口 PCA 方法在监视过程中比较  $k$  时刻的数据轨迹和  $k$  时刻子模型, 根据当前数据和模型参数之间的接近程度实时检测和判断新批量过程数据统计特征和正常数据统计特征的一致性. 子模型的统计投影和过程性能监视原理与 PCA 方法类似. 最小窗口 PCA 方法在过程性能监视和故障诊断过程中不必预报未来时刻测量变量时间轨迹, 因此消除了预报未来测量值带来的误差, 提高了监视精度和故障诊断的正确率.

最小窗口 PCA 方法利用正常数据集合确定各个子数据空间的统计控制指标<sup>[1]</sup>, 预测误差平方和  $SPE_k$ , Hotelling  $T_k^2$ ,  $i = 1, 2, \dots, K$ , 和统计指标的计算如下所示:

#### (1) 预测误差平方和 SPE

监视过程性能变化的另一个统计指标是预测误差平方和 SPE, 它表示每次采样在变化趋势上与统计模型之间的误差, 是模型外部数据变化的一种测度. 最小窗口 PCA 方法中子预测误差平方和  $SPE_k$  的定义如下:

$$SPE_k = e_k^T e_k = X_k(I - P_R P_R^T) X_k^T \quad (5)$$

第  $k$  个子模型的预测平方误差和统计控制限  $SPE_{k,\alpha}$

$$SPE_{k,\alpha} = \theta_{k,1} \cdot$$

$$\cdot \left[ \frac{c_\alpha}{\theta_{k,1}} \frac{2\theta_{k,2} h_{k,0}}{\theta_{k,1}} + 1 + \frac{\theta_{k,2} h_{k,0} (h_{k,0} - 1)}{\theta_{k,1}} \right]^{\frac{1}{h_{k,0}}} \quad (6)$$

$$\theta_{k,i} = \sum_{r=R+1}^{m \wedge n(I,J)} \lambda_{k,r}^i \quad i = 1, 2, 3 \quad (7)$$

$$h_{k,0} = 1 - (2\theta_{k,1} \theta_{k,3}) / (3\theta_{k,2}^2)$$

$$SPE_\alpha = [SPE_{1,\alpha} \quad SPE_{2,\alpha} \quad \dots \quad SPE_{k,\alpha} \quad SPE_{K,\alpha}] \quad (8)$$

其中  $c_\alpha$  是置信度为  $\alpha$  的统计, 若  $SPE_k < SPE_{k,\alpha}$ , 说明该时刻  $SPE_k$  统计正常, 否则  $k$  时刻批量过程异常.

#### (2) Hotelling $T^2$

统计指标 Hotelling  $T^2$  是得分向量 Scores 的标准平方和, 指示每个采样在变化趋势和幅值上偏离模型的程度.  $T^2$  表征了 MPCA 模型内部变化的一种测度. 子数学模型 Hotelling  $T^2$  统计指标  $T_k^2$  定义:

$$T_k^2 = t_{k,i} \lambda_k^{-1} t_{k,i}^T = X_k P_{k,R} \lambda_k^{-1} \cdot P_{k,R}^T X_k^T \quad (9)$$

其中  $\lambda_k^{-1}$  为对角阵, 对角元素为模型中保留的  $R$  个特征向量所对应的特征值的逆. Hotelling  $T^2$  的统计控制限:

$$T_{k,\alpha}^2 = \frac{R \cdot (m - 1)}{m - R} F_\alpha(R, \omega) \quad (10)$$

$$\begin{cases} n > 300, \omega = 300; \\ n \leq 300, \omega = m - R; \end{cases}$$

$$T_\alpha^2 = [T_{1,\alpha}^2 \quad T_{2,\alpha}^2 \quad \dots \quad T_{k,\alpha}^2 \quad T_{K,\alpha}^2] \quad (11)$$

其中  $F_\alpha(n, n-p)$ , 是  $100\alpha\%$  临界点,  $p, n-p$  两个自由度的  $F$  分布的上限.  $m$  是采样次数,  $n$  是变量个数,  $R$  是模型主元个数. 如果  $T_k^2 < T_{k,\alpha}^2$  则监视过程在第  $k$  个时刻处于可控状态.

### 4 聚氯乙烯生产过程性能监视和故障诊断系统方案设计 (Design of PVC process performance and fault diagnosis system)

聚氯乙烯(PVC)生产过程工艺十分复杂, 而且易燃、易爆、易中毒, 生产安全性和产品质量指标要求高. 聚氯乙烯生产工艺复杂, 引起产品质量波动的因素多种多样, 譬如原材料配比误差、反应设备磨损、测量仪表失效以及操作员操作不熟练等都会引起产品质量波动, 甚至会引起重大经济损失和人员伤亡. 聚氯乙烯生产过程技术关键是严格控制聚合反应阶段聚合釜内温度和压力, 尤其是釜内温度波动严格控制在  $\pm 0.2^\circ\text{C}$ . 因此, 有必要对聚氯乙烯间歇生产过程进行性能监视和故障检测. 如能在聚合反应早期检测出过程的异常情况, 可以帮助操作人员采取各种补救措施来消除故障或停止该批次生产, 以减少原材料的浪费和设备的占用, 降低不安全

生产事故的发生率。

### (1) 优化设计系统整体结构

根据现场工艺工程师和控制工程师的先验知识,依据聚氯乙烯生产过程的工艺特点,确定聚合反应阶段关键过程变量和影响聚氯乙烯产品质量的主要因素。(以1#釜为例)

表1 聚合反应模型变量

Tab. 1 Model variables of PVC process

序号	变量(工位号)
1	温度设定点 TIC-P101-SP
2	釜内温度 TIC-P101
3	夹套入口温度 TIC-P102
4	进水入口温度 TI-P107
5	挡板出口温度 TI-P108
6	夹套出口温度 TI-P109
7	釜内压力 PIC-P102
8	挡板水流量 FIC-P101
9	夹套水流量 FIC-P102
10	釜内搅拌功率 JI-P101

### (2) 模型参考数据预处理

多元统计数学模型的前提和基础是建立一个代表聚合反应过程正常操作的历史数据库。模型数据库的好坏直接关系到过程性能监视和故障诊断的正确性和准确性。在原有DCS控制系统基础上利用TCP/IP协议实时采集的现场数据不可避免地受到各种随机扰动和测量噪声的影响,数据存在失真现象。数据预处理主要包括数据文件格式转换、数据量化与还原、大误差剔除、模型数据长度截取等。选择52个批次过程,其中50个批次过程为正常批次过程,2个批次过程为异常批次过程。每次批次过程历时4小时10分钟,采样周期为5s,所以每个聚合反应过程是3000个采样时刻数据轨迹,52个批次过程构成历史数据 $X(52 \times 10 \times 3000)$ 集合。在建立数学模型前分析历史数据库,选择正常批次过程变量时间轨迹集合,剔除异常批次数据,所以选择50个正常批次过程数据轨迹组成模型参考数据库 $X(50 \times 10 \times 300)$ 。

### (3) 模型数据空间重组

依据最小窗口PCA方法建立数学模型,模型数据矩阵结构和表达形式发生了变化。参考模型数据库是由50个批次、10个变量、3000个采样时刻构成的三维数据矩阵 $X(50 \times 10 \times 300)$ ,而最小窗口PCA方法要求每一个子数学模型具备独立的子数据空间,因此三维参考数据库重新组合成一系列子

数学空间。具体展开方式是将三维数据矩阵沿时间轴方向依次向右展开构成一系列二维数据集合 $X(50 \times 10)$ ;最小窗口PCA模型是由一系列子数据空间构成,每个子数据空间是由一个采样时刻所有批次过程测量变量构成的数据阵列。

### (4) 非线性多元统计数学模型

最小窗口PCA方法建模实质是将一个大型复杂的数据空间,沿时间轨迹分割成一系列子数据空间,在每一个子数据空间上分别建立子数学模型,多个子数学模型构成了非线性多元统计数学模型。首先将三维数据矩阵 $X(50 \times 10 \times 3000)$ 沿时间轴方向分割成竖直片,依次展开得到3000个子数据块 $X(50 \times 10)$ 。采用最小窗口PCA方法分别处理每个子数据块,共建立3000个子数学统计模型。依次分析每个子数据空间主元方差贡献率,应用方差贡献率方法确定子数据空间保留的主元个数。

### (5) 基于正常数据集合建立置信限或域

基于模型数据库,应用多元统计学基本原理,计算各种统计指标,例如,Hotelling  $T^2$ , SPE等。基于正常数据建立的置信限或域是衡量模型正常与异常的分界线。

### (6) 新批次过程性能监视和故障诊断

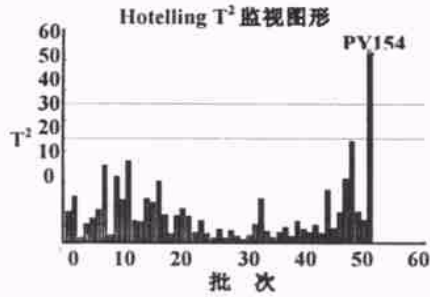
新批次数据投影到离线模型后,比较新批次统计控制量(SPE、Hotelling  $T^2$ )和统计控制指标,评价过程性能;故障条件下分析时间子SPE和变量子SPE贡献图找出故障发生的原因。为了清晰直观地表示离线监视结果,采用多种统计控制图如SPE、Hotelling  $T^2$ 和PC#监视图形。每种统计控制图分别给出95%和99%两种统计控制限或控制域。如果新批次过程的统计指标处于统治控制域或控制限内时,表明监视过程处于一种统计控制状态,否则表明系统处于故障失控状态。图形化表达信息丰富、表达方式直观、易于理解。如果检测到故障,则寻找故障发生的时间、种类、引发故障的主要变量。

## 5 聚氯乙烯生产过程性能监视和故障诊断实例(Example of PVC process performance and fault diagnosis)

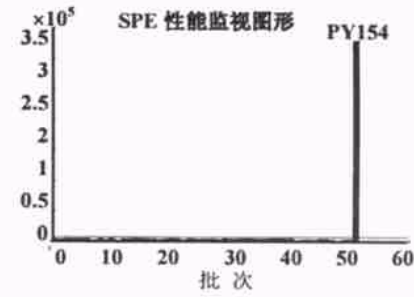
下面以聚合反应实例“PY154”说明非线性PCA故障检测和诊断过程,其中编号“PY1#”表示1号聚合釜第#次聚合反应批次过程。图2(A)(B)分别为聚氯乙烯生产过程离线监视SPE和Hotelling  $T^2$ 图,两种统计指标均表明“PY154”批次过程超出正常操作范围,故障类型为“动作报警”;图

(C) 表明引发批次过程“PY154”异常的变量是釜内温度(变量 2)和釜内压力(变量 7), 图(D)表明故障发生时间为聚合反应过程中期(800~ 1100). 图 3 (A) (B) 是聚合过程在线监视和诊断的 SPE 和 Hotelling  $T^2$  结果图, 从图中可以明显地看出故障

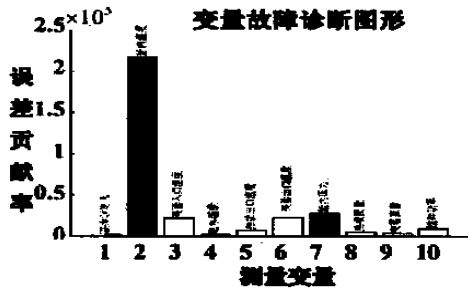
类型(“动作报警”), 故障发生时间(中期), 分析每个异常时刻, 还可以识别故障变量. 根据现场工程师的先验知识和图(C)(D) 给出变量 2 和变量 7 的轨迹曲线和平均轨迹的比较结果, 可以验证离线、在线过程监视和故障诊断的有效性.



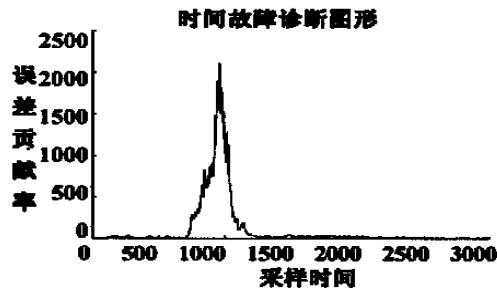
(A) PY154 批次 SPE 图



(B) PY154 批次 Hotelling  $T^2$  图



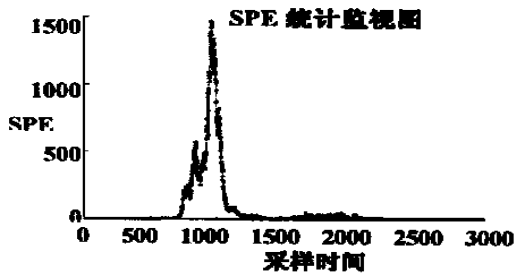
(C) 变量诊断图形



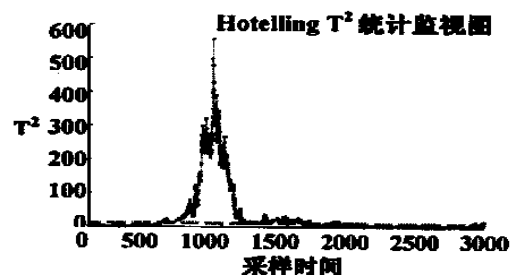
(D) 时间诊断图形

图 2 PY154 批次过程离线故障检测和诊断

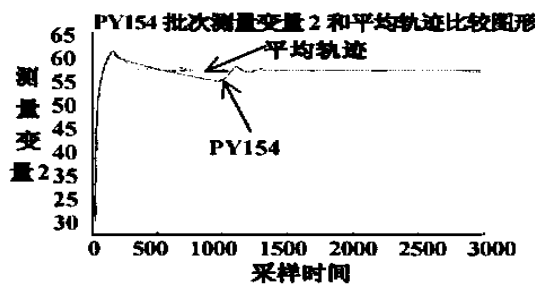
Fig. 2 Fault detection and diagnosis for PY154 off-line batch process



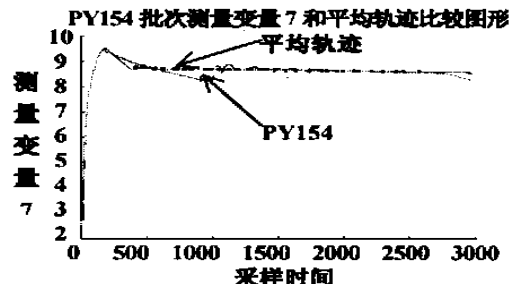
(A) PY154 批次 SPE 图



(B) PY154 批次 Hotelling  $T^2$  图



(C) 变量 2 轨迹图



(D) 变量 7 轨迹图

图 3 PY154 批次过程在线故障检测和诊断图

Fig. 3 Fault detection and diagnosis for PY154 on-line batch process

## 6 结论(Conclusion)

本文提出的最小窗口 PCA 方法是一种应用在复杂非线性间歇过程中具有实时在线监视和故障诊断功能的多元统计分析方法. 最小窗口 PCA 方法主要设计思想是将整个间歇过程分成若干子窗口数据空间, 在每个子窗口数据空间分别建立子数学统计模型, 多个子数学模型构成了整个间歇过程动态模型集合. 最小窗口 PCA 方法具有以下三个突出特点: (1) 非线性建模方式. 在每个子数据空间分别建立线性化子模型, 不同时刻切换不同的数学模型, 充分体现非线性动态特性, 因此最小窗口方法是一种非线性动态多模型的建模方式. (2) 在线实时监视和故障诊断功能. 最小窗口 PCA 方法在监视过程中将  $k$  时刻的数据集合和  $k$  时刻的子数学模型相比较, 只需要当前时刻数据即可, 无需知道整个批量过程全部数据, 易于实现实时在线策略. (3) 过程性能监视和故障诊断精度高. 最小窗口 PCA 方法在间歇过程监视和故障诊断过程中无须预报未来时刻测量变量时间轨迹, 既减少了计算误差又减少了舍入误差, 提高了监视和诊断精度. 最小窗口 PCA 方法的普遍意义在于, 其子模型的数据集合是最小数据块单元. 如果根据过程不同而采用不同窗口宽度下的数据集合作为数据单元, 或者根据具体情况自动改变窗口宽度的建模方法, 我们称之为滑动窗口 PCA 方法.

有关滑动窗口 PCA 方法的进一步讨论在另文给出.

## 参 考 文 献 (References)

- 1 Kresta J, MacGregor J F, Marlin T E. Multivariate Statistical Monitoring of Process Operating Performance. AICHE Meeting, 1989
- 2 MacGregor J F. Statistical Process Control of Multivariate Process. IFAC, ADCHEM (Advanced Control of Chemical Processes), 1994
- 3 Nomikos Paul and MacGregor John F. Monitoring Batch Processes Using Multiway Principal Component Analysis. AICHE Journal, August 1994, 40(8)
- 4 Dong Dong and McAvoy Thomas J. Batch Tracking via Nonlinear Principal Component Analysis. AICHE Journal August 1996, 42(8)

## 作者简介

赵立杰(1972-), 女, 分别于 1996 年和 1999 年在沈阳化工学院获得学士和硕士学位, 现为沈阳化工学院自动化系讲师. 研究领域为多元统计分析方法, 神经网络, 工业过程性能监视和故障诊断.

王 纲(1956-), 1991 年毕业于东北大学, 获硕士学位, 现为沈阳化工学院教授, 1996-1998 年曾赴美国德克萨斯大学作访问学者, 长期从事复杂工业过程控制和故障诊断理论和应用研究工作. 研究领域为工业过程建模, 性能监视和故障诊断, 现场总线等.

(上第 358 页)

- 2 Pican N, Alexandre F. Artificial Neural Networks for the Presetting of a Steel Temper Mill. IEEE Expert Intelligent System and Their Applications. 1996, 11(1): 22~27
- 3 万维汉, 万百五. 闪速炉的神经网络冰镍质量模型与稳态优化控制研究. 自动化学报. 1999, 28(6): 800~803
- 4 Darby M L, White D C. On-Line Optimization of Complex Process Units. Chem. Eng. Progress, Oct. 1998, 51~59
- 5 Joao Albuquerque, Vipin Gopal. Interior point SQP strategies for large-scale, structured process optimization

problems. Computers and Chemical Engineering 1999, 23: 543~554

- 6 周金辉, 邵惠鹤. 基于遗传算法和逐次二次规划的在线优化方法及应用. 信息与控制. 1999, 28(增): 289~292

## 作者简介

裴浩东(1970-), 男, 在职博士生. 研究领域为智能优化控制.

苏宏业(1969-), 男, 博士, 教授. 研究领域为大时域大时滞系统理论及应用、先进控制、智能优化控制等.

褚 键(1963-), 男, 博士, 教授博导, 长江学者. 研究领域为大时滞系统理论及应用、先进控制、智能优化控制等.