

递推 PLS 方法及其在过程监控中的应用*

苏金明¹, 李春富¹, 孙如田², 侯平智¹, 王建中¹

(1. 杭州电子科技大学 自动化学院, 浙江 杭州 310018; 2. 克拉玛依职业技术学院, 新疆 克拉玛依 833600)

摘要: 针对处理批量数据时传统的偏最小二乘(PLS)模型无法在线更新的问题, 提出了结合遗忘因子法的递推 PLS 算法, 进而更新 Q_a 控制限, 有效地克服了传统 PLS 中 Q_a 无法反映系统时变性的缺点。并对异丁烯酸甲酯(MMA)聚合反应过程中的数据进行了仿真分析。研究表明, 递推 PLS 方法大大降低了监控过程中的误报率和漏报率, 提高了监控系统的性能。

关键词: 递推偏最小二乘; 过程监控; 间歇过程

中图分类号: TP277

文献标识码: A

文章编号: 1001-4551(2010)03-0093-04

Recursive PLS method and its application in process monitoring

SU Jin-ming¹, LI Chun-fu¹, SUN Ru-tian², HOU Ping-zhi¹, WANG Jian-zhong¹

(1. School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China;

2. Karamay Vocational & Technical College, Karamay 833600, China)

Abstract: Aiming at the problem that recursive partial least square(PLS) model is not updated with process changes, which may lead to increased false alarm rate. Recursive PLS with forgetting factor method was proposed. When new data was available, the PLS model and the corresponding Q_a control limit were updated to accommodate the process change. A MMA polymerization batch process was studied. The simulation result demonstrates that recursive PLS method can greatly reduce the false alarm rate and omission rate in the monitoring process, and improve its performance.

Key words: recursive partial least square(PLS); process monitoring; batch processes

0 引 言

作为一种基于数据驱动回归模型的建模方法, 偏最小二乘(PLS)方法在化工过程中得到了广泛的应用。PLS 建模方法可以通过采用 Hotelling T^2 统计量和平方预测误差 Q 统计量(或称 SPE 统计量), 对工业过程进行性能监控。然而, 传统的 PLS 方法存在当过程特征或操作条件发生变化时不能及时进行模型更新的缺陷。为了能在线利用实时数据来更新模型, Heland^[1]等提出了一种递推 PLS(RPLS)方法, 将新数据矩阵 X 和 Y 用它们的负荷矩阵 P 和 Q 来表示, 然后将新数据与负荷矩阵相结合来更新模型。Qin^[2]对 RPLS 方法进行了改进, 将其扩展到块式递推部分最小二乘

法(block-wise recursive PLS), 并提出两种自适应算法: 移动窗口法和遗忘因子法。

本研究在对递推 PLS 方法进行深入分析的基础上, 结合遗忘因子法的基本思想, 对间歇过程进行建模, 并对其过程进行监控。当新的批次数据到来时用递推 PLS 方法对该模型进行更新, 从而适应过程参数的变化, 并将该方法用于异丁烯酸甲酯(MMA)聚合反应过程。

1 偏最小二乘(PLS)方法

PLS 方法通过将高维数据空间投影到低维特征空间, 以得到相互正交的特征向量, 再建立特征向量间的一元线性回归。假设 X 和 Y 分别为输入/输出数据矩

收稿日期: 2009-10-12

基金项目: 浙江省科技计划面上资助项目(2009C31161)

作者简介: 苏金明(1984-), 男, 浙江温州人, 主要从事生产过程的综合自动化方面的研究. E-mail: jinming126@163.com

通信联系人: 李春富, 男, 助理研究员. E-mail: lichufu@hdu.edu.cn

阵,其中 $X \in R^{n \times m}$, $Y \in R^{n \times p}$ 。偏最小二乘回归建模时对 X 和 Y 进行如下分解:

$$X = t_1 p_1^T + E_1 \quad (1)$$

$$Y = u_1 q_1^T + F_1 \quad (2)$$

其内部关系为:

$$U = TB \quad u_k = b_k t_k \quad (3)$$

$$b_k = u_k^T t_k / (t_k^T t_k) \quad (4)$$

式中 k —主元个数。

建立起来的模型为:

$$Y = UQ^T + F^* = TBQ^T + F^* \quad (5)$$

式中 E_1 和 F_1 —残差矩阵; F^* —PLS 模型的预测误差。

当 X 和 Y 之间的模型是线性模型时,所需要的特征向量个数等于模型的维数。有几种方法可用于选定所需的特征向量的个数。一种方法是给 $\|F_h\|$ 定义一个阈值,当 $\|F_h\|$ 小于这个阈值时,认为已经选取了足够多的特征向量。另一种方法是比较 $\|F_h\|$ 和 $\|F_{h-1}\|$, 当它们的差小于设定的阈值时,就可以停止增加新的特征向量。将两种方法结合起来往往会更为有效。

2 递推 PLS 算法

RPLS 算法使用 PLS 模型参数和新数据来修正模型,使得模型在并不完全遗忘旧信息的基础上适应过程的新变化^[3]。给定输入/输出矩阵 X 和 Y ,假设 X 的秩为 r ,采用如下形式表示相应的 PLS 模型:

$$\{X, Y\} \xrightarrow{\text{PLS}} \{T, W, P, B, Q\} \quad (6)$$

式中 T —输入的得分矩阵; W —输入的权值矩阵; P 和 Q —输入和输出的负荷矩阵; B —由输入输出得分向量间的回归系数构成的对角矩阵。

为推导方便,将特征向量 t_h ($h = 1, 2, \dots, r$) 归一化。于是 T 的各列标准正交。如果 $\text{rank}(X) = r \leq m$, 那么残差矩阵 $E_r = E_{r+1} = \dots = E_m = 0$ 。因而,当提取了 r 个特征向量后,则有:

$$X = TP^T + E_r = TP^T \quad (7)$$

$$Y = TBQ^T + F_r \quad (8)$$

由于输出矩阵 F_r 与前面所提取的得分向量 T 正交,即:

$$t_h^T F_r = 0, r \geq h$$

则由式(2)、式(3)可以得到:

$$X^T X = PT^T TP^T = PP^T \quad (9)$$

$$X^T Y = PT^T (TBQ^T + F_r) = PT^T TBQ^T + PT^T F_r = PBQ^T \quad (10)$$

当使用新数据 $\{X_1, Y_1\}$ 和旧数据 $\{X, Y\}$ 更新 PLS

模型时,整个数据矩阵可写为: $X_{\text{new}} = \begin{bmatrix} X \\ X_1 \end{bmatrix}$, $Y_{\text{new}} =$

$\begin{bmatrix} Y \\ Y_1 \end{bmatrix}$ 。则有:

$$X_{\text{new}}^T X_{\text{new}} = \begin{bmatrix} X \\ X_1 \end{bmatrix}^T \begin{bmatrix} X \\ X_1 \end{bmatrix} = \begin{bmatrix} P^T \\ P_1^T \end{bmatrix}^T \begin{bmatrix} P^T \\ P_1^T \end{bmatrix} \quad (11)$$

同理,可得:

$$X_{\text{new}}^T Y_{\text{new}} = \begin{bmatrix} X \\ X_1 \end{bmatrix}^T \begin{bmatrix} Y \\ Y_1 \end{bmatrix} = \begin{bmatrix} P^T \\ P_1^T \end{bmatrix}^T \begin{bmatrix} BQ^T \\ B_1Q_1^T \end{bmatrix} \quad (12)$$

对于输入/输出矩阵 X 和 Y ,其所有的 PLS 模型参数都可以由协方差矩阵 $X^T X$ 和 $X^T Y$ 计算得到^[4],所以

由式(6)和式(7)可知 $\left\{ \begin{bmatrix} P^T \\ P_1^T \end{bmatrix}, \begin{bmatrix} BQ^T \\ B_1Q_1^T \end{bmatrix} \right\}$ 和

$\left\{ \begin{bmatrix} X \\ X_1 \end{bmatrix}, \begin{bmatrix} Y \\ Y_1 \end{bmatrix} \right\}$ 有相同的 PLS 回归系数。也就是对现有

全部数据的回归等价于对新数据结合老数据而得到的 PLS 模型参数进行回归。

由此得到递推部分最小二乘(RPLS)算法,其步骤如下:

(1) 从现场采集数据,构造初始化数据矩阵 $\{X, Y\}$,并对数据进行标准化处理。

(2) 使用 PLS 回归算法得到 PLS 模型: $\{X, Y\} \xrightarrow{\text{PLS}} \{T, W, P, B, Q\}$ 。

(3) 当新数据 $\{X_1, Y_1\}$ 到来后,采用步骤(1)标准化处理中得到的均值和方差,按照步骤(1)对 $\{X_1, Y_1\}$

进行数据预处理。然后令 $X = \begin{bmatrix} P^T \\ X_1 \end{bmatrix}$, $Y = \begin{bmatrix} BQ^T \\ Y_1 \end{bmatrix}$,返回步骤(2)。

为了削弱老数据对建模的影响,可用遗忘因子与旧参数矩阵相乘,即 P 和 BQ^T 作如下处理:

$$\begin{cases} P^T = \lambda \cdot P^T \\ BQ^T = \lambda \cdot B \cdot Q^T \end{cases} \quad (13)$$

式中 λ —遗忘因子,用于调整过去时刻的数据对当前模型的影响。

将处理过的 P 和 BQ^T 同新数据组合,构成 PLS 回归的输入/输出矩阵,这样就得到了具有遗忘因子的适应性 RPLS 算法。

3 多变量统计过程监控

多变量统计监控方法^[5-6]首先对过程数据进行统计分析,提取出不相关或相对独立的特征向量和误差向量。过程变量的信息可以完全由特征向量和误差向量代替,从而将监控过程的变量问题转化为特征向量

和误差向量的监控问题。因为特征向量和误差向量之间互不相关,因而避免了错综复杂的相关性监控。多变量统计过程监控方法往往会提出一些特殊的统计指标,监控统计指标按统计规律融合了特征向量和误差向量的信息,监控指标的变化反应了特性向量和误差向量的变化,且监控指标本身的变化满足一定的概率分布特征。多变量监控的原理图如图 1 所示。

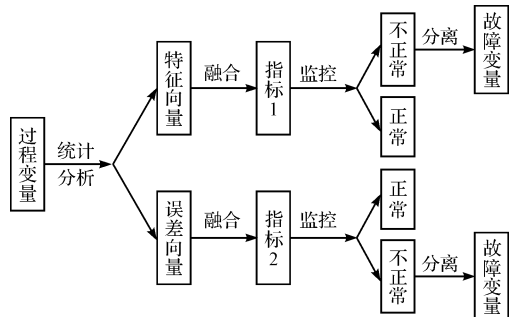


图 1 多变量监控的原理图

常用的多变量统计控制图有平方预测误差 (SPE) 图、 T^2 图、主元得分图、贡献图等。其中平方预测误差图中的统计量 SPE 在 i 时刻的值 e_i 是一个标量,它刻画了此时刻测量值 X_i 对主元模型的偏离程度。当用 PLS 算法得到模型后,第 i 时刻其值为:

$$SPE = e_i^2 = \sum_{j=1}^m (X_{ij} - \hat{X}_{ij})^2 \quad (14)$$

SPE 控制限可以按下式计算:

$$Q_a = \theta_1 \left[\frac{C_a \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \quad (15)$$

其中:

$$\theta_i = \sum_{j=k+1}^n \lambda_j^i \quad (i = 1, 2, 3) \quad (16)$$

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \quad (17)$$

式中 λ_i — X 的协方差矩阵的特征值; C_α —正态分布在检验水平为 α 下的临界值; C_α 与 h_0 同号; k —主元模型中所保留的主元个数; n —全部主元个数。

但是基于 PLS 建模的监控方法的缺陷是新的数据没有被有效地利用,模型和 Q 统计量的检测控制限不会在线更新,过程存在时变时不能适应过程的变换,因此不可避免的发生漏报和误报现象,大大降低了监控系统的有效性。而 RPLS 算法,在每次新数据到来时会重新更新模型和 Q 控制限,从而解决了上述问题。

4 应用研究

本研究以异丁稀酸甲酯 (MMA) 聚合过程作为研究对象,如图 2 所示。该反应为间歇聚合过程,反应器

具有一个夹套,通过调节反应器夹套的冷热水流量来控制反应器温度。用一个包括两个 PID 控制器 (TC) 的串级控制系统来控制反应器温度。通过采用不同的反应温度和不同的引发剂含量,可以得到不同规格的产品。产品数均分子量 M_n 和重均分子量 M_w 是重要的质量指标^[7-8],但不能在线测量,只能通过离线分析得到。通过在线测量的过程变量对它们进行监控,操作工能够及时得知产品质量,可以根据实际情况及时进行调整。

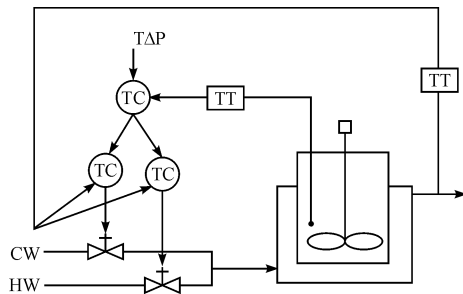


图 2 MMA 聚合反应过程

本研究共采集到 40 个批次的反应数据,前 20 批次数据作为初始数据来建立 PLS 模型。在检验水平 $\alpha = 0.95$ 下确定初始的 Q 和 Q_a ;当下一个批次的数据到来后,又会重新计算 Q 和 Q_a ;当 $Q < Q_a$ 时,调用 RPLS 算法重新建模,以此类推,这样就达到了实时更新模型和 Q_a 的目的。

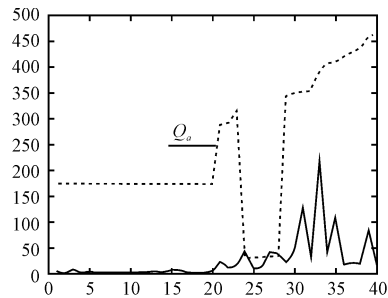


图 3 RPLS 模型的质量监控图

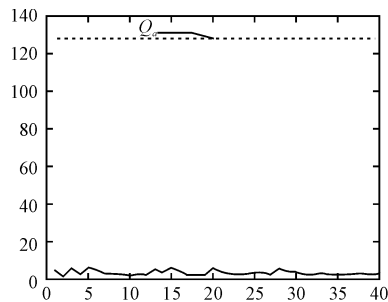


图 4 PLS 模型的质量监控图

RPLS 模型的监控过程中的 SPE 图如图 3 所示, PLS 模型的监控 SPE 图如图 4 所示。从图 3、图 4 中可
(下转第 99 页)

绝对误差 $< 0.1 \text{ dBm}$;

(2) 光功率在 $-53 \text{ dBm} \sim 0 \text{ dBm}$ 范围内,可清晰听到接收的语音信号;

(3) 光功率在 $-47 \text{ dBm} \sim 0 \text{ dBm}$ 范围内,能够准确接收到数字信号;

(4) 6 节充电电池供电时间(1 300 mAh) 达到 27 小时;

(5) 环境温度在 $-30 \text{ }^\circ\text{C} \sim 55 \text{ }^\circ\text{C}$ 范围内均能正常工作。

4 结束语

本研究研制的便携式智能光纤测试仪具备光功率测量和数字、音频信号传输多重功能,能够精确测量出光信号的强度并能够以数字的形式将测得的光功率值传给发送方,能够产生语音呼叫信号并传输呼叫方的语音和接收对方传来的数字和语音信号;采用最新的对数运算放大器来实现光功率测量,使得光功率测量值绝对误差小于 0.1 dBm 。

在降低电路噪声的基础上,通过单片机实现编/解

码、增强容错能力以保证数字传输的可靠性,解决了远距离数字通信过程中信号检测问题,并有效地提高了电路的信噪比。

参考文献(References):

- [1] 吴翼平. 现代光纤通信技术[M]. 北京:国防工业出版社, 2004.
- [2] 刘增基. 光纤通信[M]. 西安:西安电子科技大学出版社, 2001.
- [3] 范志刚. 光纤测试技术[M]. 北京:电子工业出版社, 2004.
- [4] 嵇晓强,戴明,张秋鄂. 基于 AD7705 的在线激光功率检测系统设计[J]. 现代电子技术, 2009, 32(6): 165 - 167.
- [5] 帅希士,帅词俊,杨成林,等. 基于 Delphi 的双通道光功率检测系统的设计[J]. 机械与电子, 2006(12): 44 - 46.
- [6] 林德健. 光纤元件的自动测试方法[J]. 国外电子测量技术, 1998(6): 30 - 31.
- [7] 王秀峰. OTDR 测试光纤故障点及应注意的几个问题[J]. 电线电缆, 2003(5): 30 - 32.
- [8] DAKIN J, CULSHAW B. Optical Fiber Sensors, Principles and Components, Vol. I[M]. London: Artech House, 1988.

[编辑:柴福莉]

(上接第 95 页)

以看出,基于 RPLS 模型的监控图中 SPE 控制限会利用 RPLS 进行递归更新,从而使得到的统计模型能够更真实地反映实时监控过程中的系统统计结构。对于第 24 个和第 27 个数据为故障数据的情况下,在图 4 中显示为正常,说明基于 PLS 模型的过程监控图不能准确的检测出故障;而基于 RPLS 方法的监控图显示这两个点的 Q 统计量已经超过控制限,为故障数据,因而能够准确的反映系统的性能。

5 结束语

本研究提出了一种与遗忘因子法相结合的 RPLS 方法来进行间歇过程质量监控,利用新获得的数据在线更新监控模型和 Q 统计量,使得模型能够适应过程的变换。该方法克服了传统 PLS 算法中不能实时更新 SPE 控制限的缺点,从而减少了监控系统中的误报率,提高了系统的可靠性。

参考文献(References):

- [1] HELLAND K, BERNTSEN H E, BORGES O S, et al. Re-

cursive algorithm for partial least squares regression [J]. **Chemometrics Intelligence Laboratory System**, 1992, 14 (1 - 3): 129 - 137.

- [2] QIN S J. Recursive PLS algorithms for adaptive data modeling [J]. **Computers & Chemical Engineering**, 1998, 22(4 - 5): 503 - 514.
- [3] 李凡,吴强,杨英华,等. 基于递推 PLS 的自适应钢温软测量模型 [J]. 控制工程, 2007(2): 147 - 150.
- [4] LINDGREN F, GELADI P, WORD S. The kernel algorithm for PLS [J]. **Chemometrics**, 1993, 7(1): 45 - 59.
- [5] 毛振华. 基于主元分析的自适应过程监控方法研究 [D]. 杭州:浙江大学信息科学与工程学院, 2008.
- [6] 赵忠盖. 基于 PCA 统计过程监控的若干问题研究 [D]. 无锡:江南大学通信与控制工程学院, 2007.
- [7] 张杰,阳宪惠. 多变量统计过程控制 [M]. 北京:化学工业出版社, 2000.
- [8] 李春富,叶昊,王桂增. 基于多向 PLS 方法的间歇过程质量预测 [J]. 系统仿真学报, 2004, 16(6): 1168 - 1170.

[编辑:柴福莉]