

动态 MPCA 在发酵过程监测与故障诊断中的应用 Fermentation Process Monitoring and Fault Detection Based on Dynamic MPCA

汪志锋^{1,2*}, 袁景淇²

WANG Zhi-Feng^{1,2*} and YUAN Jing-Qi²

1 上海第二工业大学自动化系, 上海 201209

2 上海交通大学自动化系, 上海 200240

1 Department of Automation, Shanghai Second Polytechnic University, Shanghai 201209, China

2 Department of Automation, Shanghai Jiaotong University, Shanghai 200240, China

摘 要 针对发酵过程非线性和时变特点, 提出了一种具有实时性的动态 MPCA 方法, 采用多模型非线性结构代替传统 MPCA 单模型线性化结构, 克服了后者不能处理非线性过程和实时性的问题, 并避免了 MPCA 在线应用时预报未来测量值带来的误差, 提高了发酵过程性能监测和故障诊断的准确性。对头孢菌素 C 发酵过程的拟在线仿真研究, 验证了基于动态 MPCA 的统计过程监测的有效性。

关键词 多方向主元分析(MPCA), 多模型, 发酵过程, 在线监测, 故障诊断

中图分类号 TP 273 **文献标识码** A **文章编号** 1000-3061(2006)03-0483-05

Abstract A dynamic multiway principle component analysis for on-line batch process monitoring and fault detection was proposed. It integrates the time-lagged windows of process dynamic behavior with the multiway principle component analysis (MPCA). Using multi-model instead of single model, the dynamic MPCA approach emphasizes particularly on-line process performance monitoring and fault detecting. On-line process monitoring of cephalosporin C fermentation was studied, the results demonstrate that the dynamic MPCA method is able to efficiently monitor performance of the fermentation process and exactly detect faults which results in extraordinary behavior of processes.

Key words multiway principle component analysis(MPCA), multi-model, fermentation, on-line monitoring, fault detection

在生物发酵过程中, 微生物的生长对环境变化极为敏感, 如果生产条件偏离了预先设定的最佳条件, 就会影响其内部的代谢过程, 降低产生菌的质量和数量, 甚至得不到所要求的产物。如果能在异常出现的初期及时给出预警信息, 从而采取相应的措施, 将能减少损失, 甚至避免异常的发生。但是生物发酵过程复杂, 很多关键变量如基质浓度、生物量需

要离线分析, 滞后时间长, 难以及时地对发酵趋势和故障做出预警^[1]。近年来多变量统计技术被广泛用于发酵过程故障诊断^[2,3], 如多方向主元分析(MPCA), 它具有不依赖于过程数学模型的特点, 明显区别于基于数学模型或知识的其他故障诊断方法, 但常规 MPCA 中的 T^2 图和 SPE 图只能监测过程是否发生了变化, 不能直接提供引起这些变化的

Received: December 15, 2005; Accepted: February 21, 2006.

This work was supported by a grant from the National Natural Sciences Foundation of China (No. 60574038).

* Corresponding author. Tel: 86-21-50851891; E-mail: zhifeng.wang@163.com

国家自然科学基金资助项目(No. 60574038).

原因和故障发生的时刻,对于批生产过程来说,它就难以保证在线状态监测和故障诊断的实时性^[4]。本文针对批过程生产的特点,提出一种具有实时性的动态 MPCA 方法,并结合某厂实际生产数据,将该技术用于生产工艺和过程控制远不如传统抗生素(如青霉素)成熟的头孢菌素 C 发酵生产中,验证了该方法的准确性和有效性。

1 动态 MPCA

1.1 多方向主元分析方法(MPCA)

批过程体现为重复性的生产过程,其数据集合比连续生产过程数据集合多一维“批量”元素,具有序贯特性,因此可用三维数据矩阵 $X(I \times J \times K)$ 代表批过程数据集合,其中: I 表示批量数, J 表示变量数, K 表示采样点数。MPCA 将三维数据矩阵沿时间轴方向切割为批量和变量数据块 $X(I \times J)$, 每个数据块依次向右水平排列,形成一个新的二维数据矩阵 $X(I \times KJ)$ ^[5], 见图 1。

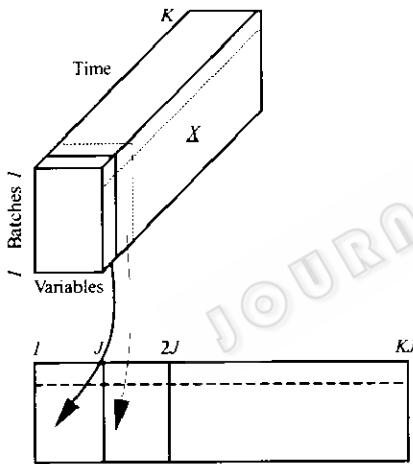


图 1 MPCA 数据矩阵分解

Fig. 1 Unfolding of a three-way array by MPCA

三维数据矩阵展开成二维数据矩阵后, MPCA 数据处理和分析过程等同于普通 PCA 方法。展开的数据矩阵 X 被分解为得分向量和载荷向量乘积之和,再加上残差矩阵 E , 见式(1)~(2)。

$$X = \sum_{r=1}^A t_r \otimes P_r + E \quad (1)$$

$$E = \sum_{r=A+1}^M t_r \otimes P_r \quad (2)$$

其中 t_r 是系统主元, $r = 1, 2, \dots, M (= KJ)$, 亦称得分向量, 相互正交, 表征整个批操作过程中不同批次过程之间相互关联的信息; P_r 是主元特征向量, 又称载荷向量, 亦相互正交, 代表过程变量时间序列相互关联的信息; A 是系统中保留的主元个数, 由交叉

确认准则确定^[5], 一般可要求保留的主元累积方差贡献率 $\geq 85\%$ 。式(1)右侧第一项代表批过程主要变化信息, 是批次、过程变量和时间的函数, 第二项 E 表征批过程数据的次要变化, 应在最小二乘意义上最小。

1.2 动态主元分析方法

MPCA 方法通过抽取测量变量平均轨迹可以使问题部分线性化, 但是并没有从本质上解决非线性对象的多元统计建模问题。动态主元分析方法建模的主要设计思路如下: 首先将数据用 MPCA 的方法进行处理, 将三维数据矩阵变换成二维数据矩阵, 然后将数据空间划分成一系列子数据空间集合, 最后在所有子数据空间分别建立子统计模型。具体过程如下:

(a) 子数据空间构造。将图 1 中的二维数据矩阵 $X(I \times KJ)$, 用移动数据窗口技术^[6] 将 X 划分成一系列子数据空间 $(I \times dJ)$, $k = d, d+1, \dots, k, \dots, K$, 表示 I 个批次、 J 个过程测量变量在采样点 $[k-d+1, \dots, k]$ 上的数据集合, 即从 k 点回溯 $d-1$ 个采样点, 其中 d 值的大小要适中, 太大于模型难以及时跟踪过程的变化; 太小又不能将过去状态对过程的影响带入模型, d 值的具体确定见文献[3]。这样, 原二维数据矩阵 X 就可表示为这一系列子数据空间的集合, 见式(3)。

$$X = \{ X_d^d(I \times dJ) X_{d+1}^d(I \times dJ) \dots X_k^d(I \times dJ) \dots X_{K-1}^d(I \times dJ) X_K^d(I \times dJ) \} \quad (3)$$

(b) 子数据空间分解。首先将子数据空间集合 $X_k^d(I \times dJ)$ 量化成均值为 0、方差为 1 的标准化数据矩阵 X_k , 然后用 MPCA 分别建立子统计模型, 见式(4)~(5)。

$$X_k = \sum_{r=1}^A t_{r,k} \otimes P_{r,k} + E_k \quad (4)$$

$$E_k = \sum_{r=A+1}^M t_{r,k} \otimes P_{r,k} \quad (5)$$

其中, $P_{r,k}$ 为第 k 个子数据空间的载荷矩阵, $t_{r,k}$ 为第 k 个子数据空间的得分向量, E_k 是第 k 个子数据空间的残差矩阵。

2 监测与故障诊断原理

基于主元分析法进行状态监测和故障诊断的基本思想是, 根据正常工况下的过程历史数据建立主元模型, 即找出能够表达正常工况下过程各变量之

间的因果关系的低维主元成份,一旦过程的实时测量数据与建立的主元模型不符,就可以判断过程中已有故障发生,再通过对测量数据中各变量变化对主元模型的误差贡献率的分析,进行故障诊断。

对新的测量样本 X_{new} 进行故障检测时,首先要对 X_{new} 进行分组,每 d 个采样点数据为一组,如采样点 $1 \sim d$ 的数据作为第 d 组,采样点 $2 \sim (d+1)$ 的数据作为第 $(d+1)$ 组,以此类推。然后根据不同分组来选择相应的子模型进行故障检测。这里不失一般性,假设要对第 i 罐批第 k 组样本数据进行检测,那么对应的参考模型是第 k 个子模型,按式(4)、式(5)计算出样本主元值 t_{new} 及 X_{new} 的估计值 \hat{X}_{new} 和残差 E_{new} ,最后利用得到的 t_{new} 、 \hat{X}_{new} 和 E_{new} 进行统计假设检验,根据检验结果,可以监视过程状态是否正常和判定有无故障发生。常用的统计变量包括:平方预测误差 SPE 和 Hotelling T^2 。 SPE 是残差矩阵 E 每行的平方和,表示每次采样在变化趋势上与模型的误差, SPE 是表征模型外部数据变化的一种测度,子模型 k 的 SPE_k 定义为:

$$SPE_k = E_k E_k^T = X_k (I - P_{k-1} P_{k-1}^T) X_k^T \quad (6)$$

Hotelling T^2 是得分向量的标准平方和,表示每个采样在变化幅值上偏离模型的程度, T^2 是表征模型内部变化的一种测度,子模型 k 的 T_k^2 定义为:

$$T_k^2 = t_k \lambda_k^{-1} t_k^T = X_k P_{k-1} \lambda_k^{-1} P_{k-1}^T X_k^T \quad (7)$$

式中: λ_k^{-1} 为对角矩阵,对角元素为模型中保留的 A 个主元特征向量所对应的特征值的逆。一般认为过程正常运行下的测量数据 X_k 满足正态分布,如果 SPE_k 、 T_k^2 满足式(8)~(9),则说明该时刻批过程处于正常操作状态,否则认为在第 k 个时刻过程异常。

$$SPE_k < SPE_{k,\alpha} \quad (8)$$

$$T_k^2 < T_{k,\alpha} \quad (9)$$

式(8)~(9)右侧的数值分别是第 k 个子数据空间的 SPE 、 T^2 对应于置信度 α 的统计控制限。这样,通过比较第 k 组数据和第 k 个子模型参数之间的接近程度,就可实现批过程在线实时监视和故障诊断。

3 拟在线研究

3.1 参考数据库建立

本文以某厂头孢菌素 C 生产车间 53 个历史罐批(包括正常罐批 50 个和异常罐批 3 个)为依据,从中选取 20 个正常罐批组成参考数据库,作为参考数

据库的历史罐批须在相同的工艺条件下,具有较好的代表性,能够反映头孢菌素 C 生产过程规律的正常罐批数据,即极端异常罐批和故障罐批的数据不进入数据库;另外 2 个罐批(罐批号编为 21 和 22,21 为异常罐批)作为模型检验样本。工业头孢菌素 C 发酵过程中关键变量(如产物浓度和基质浓度等)是过程监控和控制的重要窗口覆盖的输入变量的累计量,如产量(P)、糖消耗量(S)、前体消耗量(PAA)、氨水消耗量($\text{NH}_3 \cdot \text{H}_2\text{O}$)、硫酸消耗量[$(\text{NH}_4)_2\text{SO}_4$]、氧消耗量(O_2)和二氧化碳产生量(CO_2)等,为此,我们选择前 5 个主要变量来综合表征头孢菌素 C 发酵菌体生长和产物合成状况。参考数据库覆盖 53 个采样点,共计 212 h(离线采样周期 4 h)。这样,模型参考数据库可以写成三维数据矩阵 $X(20 \times 5 \times 53)$ 。然后将三维数据矩阵 $X(20 \times 5 \times 53)$ 沿时间轴方向分割成子数据矩阵,取 $d=5$,依次展开得到 49 个子数据块 $X_k^d(20 \times 25)(k=d, d+1, \dots, K)$,共建立 49 个子统计模型,根据子统计模型依次分析每个子数据空间主元方差贡献率,按照方差贡献率 $\geq 85\%$ 原则来确定子数据空间保留的主元个数。文中进行仿真研究时,考虑到头孢菌素 C 发酵过程前期波动较大,拟在线监测的起点是发酵开始后的第 80h。

3.2 拟在线分析

图 2 是基于传统 MPCA 方法的 SPE 和 Hotelling T^2 图形,图中实线和虚线是根据给定的参考数据库建确定的分别对应置信度为 99% 和 95% 控制限。图中检验样本罐批 21 的 SPE 和 T^2 都超出了 99% 控制限,故判断该罐批过程发生了异常,而罐批 22 在正常范围内。不过这些都是操作结束时做出的判断。那么对于 21 罐何时开始严重偏离正常范围的呢?可以采用动态 MPCA 来进行分析,图 3、4 分别给出了罐批 21 和 22 的 SPE 和 Hotelling T^2 拟在线监视图形。从图 3 中可以看出, SPE 和 T^2 在 160~188h 区间均出现了超限,因此,可以判断罐批 21 从 160h 开始发生故障;而罐批 22 的均在控制限内,视为该罐批运行正常。比较图 2 和图 3、4,基于传统 MPCA 方法和动态 MPCA 方法对检验样本罐批 21、22 的过程监视结果是一致的。但动态 MPCA 在不需要发酵过程完整数据集的情况下,实现了实时故障诊断,而传统 MPCA 原则上只适合于事后诊断。

结合 SPE 贡献率图可以进一步确定故障原因,图 5 是罐批 21 在 172 h 的 SPE 贡献率图,图中表明,前体消耗量(PAA)对 SPE 的累积贡献率远远超

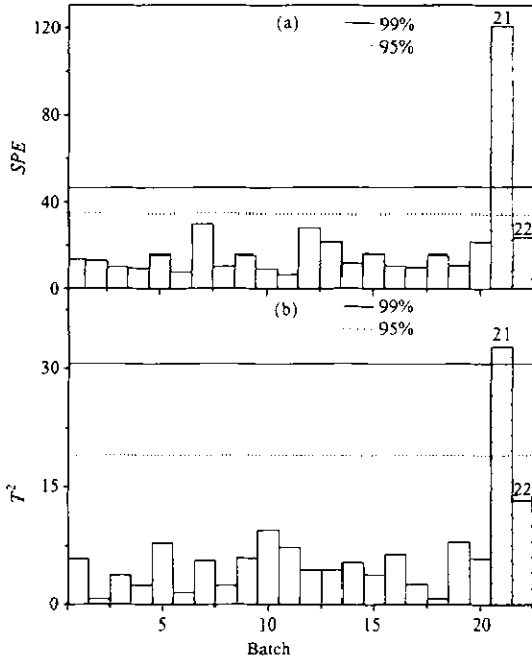


图2 基于传统 MPCA 方法对罐批 21,22 的统计监视图
Fig. 2 Statistical control chart of on-line monitoring by using traditional MPCA for batch 21 and 22

补料量,从而改善发酵过程,避免更大的经济损失。

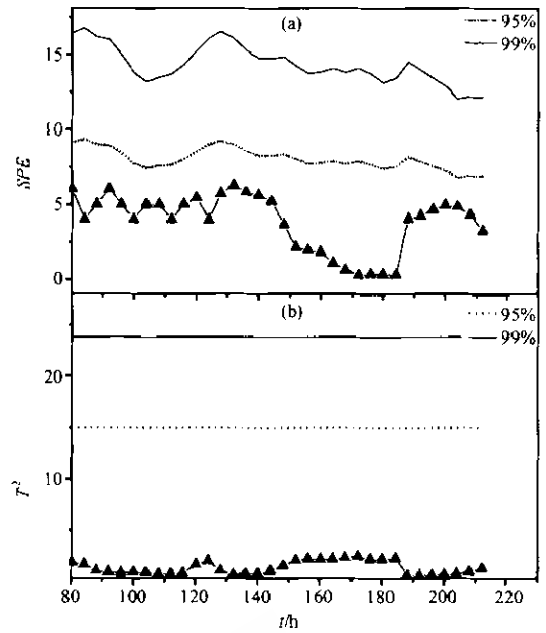


图4 基于动态 MPCA 方法对罐批 22 的统计监视图
Fig. 4 Statistical control chart of monitoring by using dynamic MPCA for batch 22

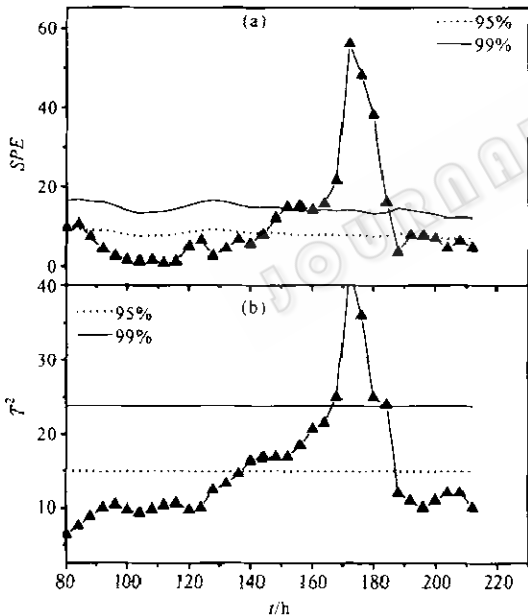


图3 基于动态 MPCA 方法对罐批 21 的统计监视图
Fig. 3 Statistical control chart of monitoring by using dynamic MPCA for batch 21

出其它变量,可以认为,前体消耗量异常是引起该时段头孢菌素 C 发酵过程异常的主要原因。图 6 给出了检验样本罐批和参考数据库前体消耗速率轨迹(标准化后的数据),通过对比发现罐批 21 前体消耗速率明显低于参考罐批,进一步确认了前体的补料量异常是造成罐批 21 发酵过程故障原因。在实际生产过程中,可以根据以上预警信息,调整相应的

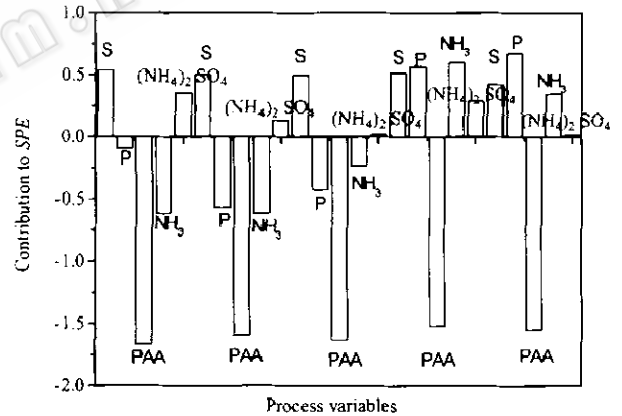


图5 罐批 21 过程变量对 SPE 贡献图
Fig. 5 Contribution of process variables to SPE for batch 21

4 结语

基于传统 MPCA 的过程监控有以下两个问题:(1)它仅考虑数据的静态关系而非动态关系,它不仅默认为每个时刻的变量值独立于其它时刻,而且默认为同一时刻不同变量之间相互独立,该方法仅能给出当前罐批与正常罐批之间的相似程度的判定;(2)用于在线监测时,需对所有相关过程变量进行预测,这时,预报误差会导致误检测。与此形成对比的是,动态 MPCA 方法采用多模型建模方式,具有分段线性逼近非线性动态特性。它在监视过程中将 k 时刻的数据集合和 k 时刻的子统计模型相比较,只需

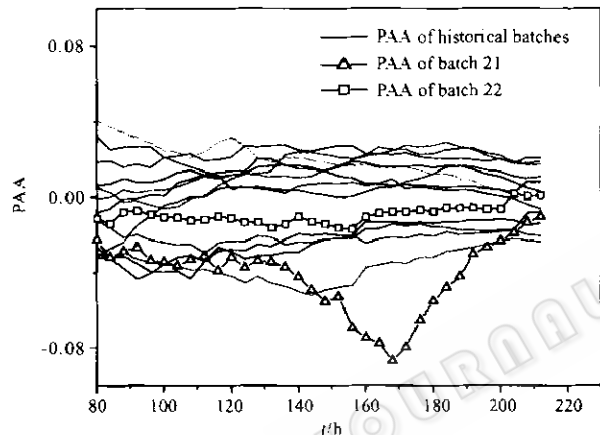


图 6 PAA 消耗速率(已作标准化处理)

Fig. 6 PAA consumption rate (normalized)

当前时刻数据,无需预报有关变量的走势,故和其它需要利用未来状态预报值的诊断技术相比,检测精度高、实时性强。

REFERENCES(参考文献)

- [1] Li YF(李运锋), Yuan JQ(袁景淇). Neural network detection of abnormalities in fed-batch fermentation. *Chinese Journal of Biotechnology*(生物工程学报), 2005, **21**(1):102 - 106
- [2] Louwerse DJ, Smilde AK. Multivariate statistical process control of batch processes based on three-way models. *Chemical Engineering Science*, 2000, **55**:1225 - 1235
- [3] Nomikos P, MacGregor JF. Monitoring batch process using multiway principal component analysis. *AIChE Journal*, 1994, **40**(8):1361 - 1375
- [4] Ku W, Storer RH, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 1995, **30**:179 - 196
- [5] Sarolta A, Robert D Kinley. Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision. *TRENDS in Biotechnology*, 2001, **19**(2):53 - 62
- [6] Yuan JQ, Vanrolleghem PA. Rolling learning-prediction of product formation in bioprocesses. *Journal of Biotechnology*, 1999, **69**:47 - 62