

基于时序模型迁移的间歇过程监测建模方法

刘昊知¹, 胥布工¹, 高福荣²

(1. 华南理工大学 自动化科学与工程学院, 广州 510641; 2. 香港科技大学 化学工程系, 香港)

摘要: 针对工业生产过程频繁变更操作条件的需求, 提出一种基于时序模型迁移的方法, 通过从旧过程数据中挖掘并提取相似性信息, 可以快速建立新过程的监测模型. 将该建模方法与时序模型迁移相结合应用于青霉素发酵仿真过程, 实验结果验证了该方法的有效性.

关键词: 间歇过程; 确定性因素分解; 时序模式组合; 时序模型迁移; 多阶段划分

中图分类号: TP277

文献标志码: A

Modeling method of monitoring for batch process based on time series model migration

LIU Hao-zhi¹, XU Bu-gong¹, GAO Fu-rong²

(1. College of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China; 2. Department of Chemical Engineering, Hong Kong University of Science and Technology, Hong Kong, China. Correspondent: XU Bu-gong, E-mail: aubgxu@scut.edu.cn)

Abstract: Aiming at the problem that the operating conditions frequently change to meet the production process requirements, a batch process monitoring strategy of modeling based on time series model migration is proposed. The combination of the modeling method and time series model migration is applied to the penicillin fermentation process of simulation. The experimental results show the effectiveness of the proposed method.

Keywords: batch process; certainty factor decomposition; time series pattern; time series model migration; multi-stage separation

0 引言

作为工业生产中一种重要的生产方式, 间歇过程已被广泛应用于精细化工、生物制药、聚合物反应等领域. 为了满足市场和顾客的不同需求, 根据工艺要求的不同, 过程的操作条件或属性值会频繁地发生改变, 以得到不同的产品结构或特性. 若能从旧过程大量蕴含生产状况的观测数据中挖掘并提取出有价值的信息, 则能有效加快新过程建模的速度.

近年来, 陆续有学者围绕过程操作条件变更和产品质量特性之间的变化展开研究. Jaeckle 等^[1-3]提出了在不同对象间传递信息的方法. Garcia-Munoz 等^[4]对该方法加以改进, 提出了一种新的主元回归方法来解决产品转移的问题. Lu 等^[5-8]提出了基于过程相似性的模型迁移的概念和对应不同相似性类型的迁移策略. 针对过程监控的模型迁移, Facco 和 Tomba

等^[9-10]率先采用主元分析建模的方法, 对连续系统的过程监控模型迁移进行了研究.

本文提出一种基于时序模型迁移的间歇过程监控建模方法, 当设定值发生改变时, 只需要新过程一个批次数据, 通过对旧过程多个批次数据迁移, 即可实现快速建模. 本文首先对间歇过程的相似性及差异性进行分析; 然后, 建立观测变量的时序模型, 针对不同时序模式蕴含的特点给出相应的迁移策略; 最后, 给出间歇过程基于负载矩阵变化的阶段划分方法以实现过程阶段的精确划分. 将所提出的建模方法应用于青霉素发酵仿真过程, 结果验证了该方法的合理性和有效性.

1 相似性和差异性

实际间歇过程存在不同程度的非线性, 当系统的设定值变化后, 系统的动态特性和静态特性都会发生

收稿日期: 2014-09-10; 修回日期: 2015-02-28.

作者简介: 刘昊知(1980—), 男, 博士生, 从事复杂工业过程建模与优化、过程监测的研究; 胥布工(1956—), 男, 教授, 博士生导师, 从事无线传感器/执行器网络、网络化控制系统等研究.

改变,但是过程自身的物理特性没变,新旧过程存在一定的相似性. 本文的迁移方法主要适用于存在以下几种相似性的间歇过程.

1) 阶段相似性. 受设定值变更的影响,观测变量的运行轨迹会产生一定程度的变化,由于间歇过程的多阶段特征,这种变化往往是阶段性的,称其为“阶段相似性”.

2) 模式相似性. 设定值改变的只是观测变量的运行轨迹,变量自身的变化规律并未受其影响,即变量的物理特性保持不变,因此将这种特性称为“模式相似性”.

3) 慢时变波动相似性. 除了设定值变更的影响因素外,还存在其他因素的影响会导致过程逐渐变化,由于实际生产过程总是遵循某种规律在批次间不断地缓慢变化,称其为“慢时变波动”.

2 时序模型迁移策略

统计过程监控技术是建立在过程变量的观测数据基础上的,数据样本按照时间顺序采集,构成了传统的时间序列. 本文通过对时间序列的分析,建立不同类型的时序模型,再结合相应的迁移策略实现模型迁移.

2.1 建立时序模型

根据 Cramer 分解定理^[11],任何一个时间序列 x_t 都可以分解为两部分的叠加,即

$$x_t = v_t + \varepsilon_t. \quad (1)$$

其中: v_t 是由多项式决定的确定性趋势成分, ε_t 是平稳的零均值误差成分.

确定性因素分解法^[12-14]是常用的时间序列分析方法. 不同的序列波动特征可以归纳成 4 类因素的综合影响: 1) 长期趋势; 2) 突然跳跃; 3) 周期波动; 4) 随机波动.

时间序列通常很少同时存在 4 类因素的影响,这就形成了时间序列的不同组合模式. 本文围绕下面 4 种常见的组合模式进行研究:

1) 长期趋势模式

$$x_t = T_t; \quad (2)$$

2) 突然跳跃模式

$$x_t = J_t; \quad (3)$$

3) 含随机波动的周期模式

$$x_t = C_t + I_t; \quad (4)$$

4) 含随机波动的趋势模式

$$x_t = T_t + I_t. \quad (5)$$

2.2 时序模型迁移策略

2.2.1 长期趋势模式的迁移

长期趋势模式是工业过程中最常见的一种时间序列. 假设 $X^{\text{base}} \in R^{I \times K}$ 为旧过程任意一个过程变量的多个批次观测数据, $x^{\text{new}} \in R^K$ 为新过程同一变量的一个批次数据. I 表示过程批次, K 表示采样点个数. 基于模式相似性,新过程变量的序列依然为长期趋势模式,即

$$\begin{cases} x_{i,k}^{\text{base}} = T_k, \\ x_k^{\text{new}} = T_k. \end{cases} \quad (6)$$

其中: $x_{i,k}^{\text{base}}$ 为旧过程变量第 i 个批次在 k 采样点的观测值, x_k^{new} 为新过程变量在 k 采样点的观测值.

由于机理特性的变化,新过程的变量轨迹发生了偏移,需要根据偏移值将旧过程变量的所有批次进行迁移. 根据慢时变波动相似性,新过程和旧过程第 1 个批次的数据存在波动相似的对应关系,因此首先计算两者的纵向偏移值

$$\gamma_k = x_k^{\text{new}} - x_{1,k}^{\text{base}}, \quad (7)$$

其中 γ_k 为纵向迁移因子.

由于新过程变量批次间的慢时变波动规律和旧过程具有相似性的特点,利用求出的纵向迁移因子 γ_k ,将旧过程变量的所有批次纵向迁移,即

$$x_{i,k}^{\text{new}} = x_{i,k}^{\text{base}} + \gamma_k, \quad (8)$$

其中 $x_{i,k}^{\text{new}}$ 为旧过程变量纵向迁移后的值.

2.2.2 突然跳跃模式的迁移

若时间序列存在突然跳跃因素,当受到设定值变更的影响时,虽然新过程变量整体满足模式相似性,但峰值有可能发生横向偏移,不符合模式相似性的条件,则需要先将旧过程序列横向迁移,使旧过程的峰值和新过程的峰值重合在同一时刻上,有

$$x_{i,k}^{\text{base}'} : \begin{cases} x_{i,[1+(k-1)\delta_L]}^{\text{base}'} = x_{i,k}^{\text{base}}, \\ k = 1, 2, \dots, k_p^{\text{base}}; \\ x_{i,[K-(K-k)\delta_R]}^{\text{base}'} = x_{i,k}^{\text{base}}, \\ k = k_p^{\text{base}} + 1, \dots, K; \end{cases} \quad (9)$$

$$\delta_L = \frac{k_p^{\text{new}} - 1}{k_p^{\text{base}} - 1}, \quad \delta_R = \frac{K - k_p^{\text{new}}}{K - k_p^{\text{base}}}. \quad (10)$$

其中: $x_{i,k}^{\text{base}'}$ 为旧过程变量横向迁移后的值, δ_L 为左横向迁移因子, δ_R 为右横向迁移因子, k_p^{base} 为旧过程变量的峰值采样点, k_p^{new} 为新过程变量的峰值采样点.

横向迁移后的序列整体和局部均满足模式相似性,可以对其进行纵向迁移,即

$$\begin{cases} \gamma_k = x_k^{\text{new}} - x_{1,k}^{\text{base}'}, \\ x_{i,k}^{\text{new}} = x_{i,k}^{\text{base}'} + \gamma_k. \end{cases} \quad (11)$$

2.2.3 含随机波动的周期模式的迁移

假设新过程变量的波动周期为 p , 一个批次内含 m 个周期; 旧过程变量的波动周期为 q , 一个批次内含 n 个周期. 当 $m \neq n$ 时, 表明序列的周期因设定值变更而发生了变化. 需要先对序列进行横向迁移, 使新旧过程变量的周期相同, 满足模式相似性, 有

$$x_{i,k}^{\text{base}'} : x_{i,[1+(k-1)\delta]}^{\text{base}'}$$

$$x_{i,k}^{\text{base}'} \begin{cases} m < n, k = 1, 2, \dots, mq + 1; \\ m > n, k = 1, 2, \dots, nq + 1; \end{cases} \quad (12)$$

$$\delta = p/q. \quad (13)$$

其中 δ 为横向迁移因子.

当周期模式除了周期波动外还存在随机波动时, 由于随机波动不受慢时变波动的影响, 不存在相似性, 仅对周期波动进行纵向迁移, 随机波动保持不变, 即

$$\begin{cases} \gamma_k = C_k^{\text{new}} - C_{1,k}^{\text{base}'}, \\ x_{i,k}^{\text{new}} = x_{i,k}^{\text{base}'} + \gamma_k, \end{cases} \quad (14)$$

其中 C_k^{new} 和 $C_{1,k}^{\text{base}'}$ 分别是新过程变量的周期波动函数和旧过程变量横向迁移后的周期波动函数, 可通过三角函数拟合得到.

2.2.4 含随机波动的趋势模式的迁移

当趋势模式除了长期趋势外还存在随机波动时, 仅对长期趋势进行纵向迁移, 随机波动保持不变, 即

$$\begin{cases} \gamma_k = T_k^{\text{new}} - T_{1,k}^{\text{base}'}, \\ x_{i,k}^{\text{new}} = x_{i,k}^{\text{base}'} + \gamma_k, \end{cases} \quad (15)$$

其中 T_k^{new} 和 $T_{1,k}^{\text{base}'}$ 分别是新旧过程变量的长期趋势函数, 可通过二次多项式拟合得到.

2.3 基于时序模型迁移的新过程建模步骤

本文提出的基于时序模型迁移的建模步骤描述如下.

1) 获得旧过程的正常数据 $X^{\text{base}}(I \times J \times K)$, 3 个维数分别表示间歇批次、过程变量数以及每次间歇操作采样点个数. 基于统计方法建立旧过程的监控模型.

2) 当设定值变更后, 采集新过程一个批次的正常数据 $X^{\text{new}}(1 \times J \times K)$. 用旧过程的模型对新过程进行监测, 通过监测结果, 找出不满足阶段相似性的变量 $x_{j,k}^{\text{base}}(1 \times I)$.

3) 对该变量建立相应的时序模型, 通过迁移策略将旧过程变量进行迁移, 迁移后得到新过程的建模变

量 $x_{j,k}^{\text{new}}(1 \times I)$.

4) 对迁移后的数据重新建立监控模型, 再次对新过程进行监控, 如果出现报警, 则转步骤 3).

5) 将监测结果为正常工况的新模型用于新过程的在线监测. 对于再来的新过程数据, 可通过自适应更新^[15]的方法对已建立的监测模型进行校正和完善.

需特别注意的是, 过程数据的检测精度是保证模型迁移后监测精度的前提条件, 在检测时, 为使过程数据满足检测精度, 应尽量减少误差, 提高检测精度. 采用误差分析, 通过数理统计的方法分析过程数据, 从而降低误差对检测精度的影响.

3 基于时段划分的过程建模及在线监测

间歇过程蕴含多操作阶段的特性, 每个阶段都有不同的物理特性和主导变量. 本文采用一种基于负载矩阵变化的分段方法建立过程的子时段 PCA 模型^[16].

3.1 基于时段划分的建模

采集间歇过程的正常过程数据 $X(I \times J \times K)$, 若将三维矩阵 X 沿 K 方向切割, 则数据标准化后可以得到时间片矩阵 $\tilde{X}_k(I \times J)$. 分析每一个 $\tilde{X}_k(I \times J)$ 的负载矩阵, 进而建立子操作时段的 PCA 模型.

选用 k -means 聚类算法将 K 个时间片矩阵划分为 C 个子时段, 第 c 个子操作时段的 PCA 模型定义如下:

$$\begin{cases} \tilde{T}_k = \tilde{X}_k \tilde{P}_c, \\ \hat{\tilde{X}}_k = \tilde{T}_k (\tilde{P}_c)^T, \\ \tilde{E}_k = \tilde{X}_k - \hat{\tilde{X}}_k. \end{cases} \quad (16)$$

3.2 在线监测

在线监测时, 获得当前时刻的采样数据 x , 计算主成分向量 t 及预测误差向量 e , 有

$$\begin{cases} t = x \tilde{P}_c, \\ e = x - x \tilde{P}_c (\tilde{P}_c)^T. \end{cases} \quad (17)$$

计算 T^2 统计量和 SPE 统计量, 判断是否超出控制限, 即

$$\begin{cases} T^2 = t^T (S_c)^{-1} t, \\ \text{SPE} = e^T e. \end{cases} \quad (18)$$

4 青霉素发酵仿真过程的应用研究

4.1 实验数据获取

通过青霉素生产仿真软件 PenSim2.0^[17], 可以有效验证本文提出的时序模型迁移的方法. 实验选用如下 7 个过程变量: 底物浓度、生物量浓度、青霉素浓度、培养基体积、二氧化碳浓度、反应产生的热量、冷

却水流量. 各批次反应周期设定为400h, 采样间隔为1h. 在其他设定值默认的情况下, 人为地设定初始培养基容量在100~101.9L间波动, 对应产生20个旧过程批次, 它们包含了批次间慢时变波动规律. 之后再再将底物浓度设定为13g/L, 培养基容量在100~102L间波动, 产生21个新过程批次, 第1个批次用于模型迁移, 后20个批次用于迁移后的模型检验.

4.2 旧过程建模

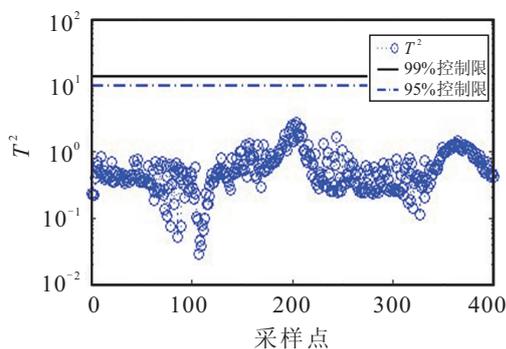
将三维建模数据阵 $X(20 \times 7 \times 400)$ 按时间片进行二维展开, 数据标准化后得到400个时间片矩阵. 分别对每个时间片进行主元分析, 采用 k -means聚类算法按照主成分投影方向的相似性进行聚类, 划分为5个阶段.

4.3 时序模型迁移

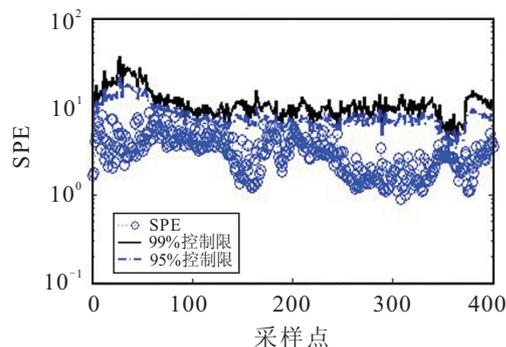
用旧过程建立的模型监控新过程, 通过报警显示, 找出不满足阶段相似性的过程变量, 分别建立相应模式的时序模型, 如表1所示. 通过时序模型迁移,

表1 旧过程迁移的阶段及变量

迁移阶段	迁移变量	时序模型
1	1	长期趋势模式
	2、6	长期趋势模式
2~3	4	突然跳跃模式
	3、5、7	含随机波动的趋势模式
4	4	突然跳跃模式



(a) T^2 统计量



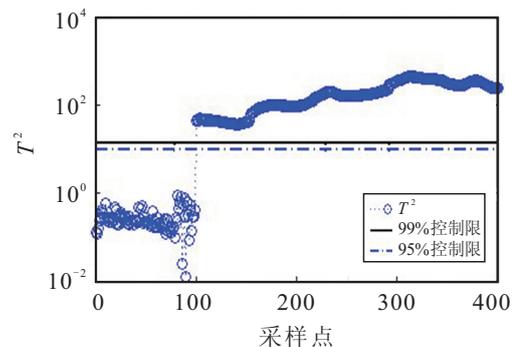
(b) SPE 统计量

图1 变量迁移后对新过程的监测曲线

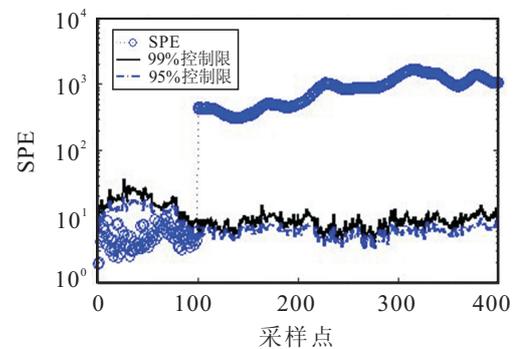
对迁移后的过程批次建立监控模型, 再次对新过程第1批次进行监控, 如图1所示. 可以看出统计量均未超出控制限.

4.4 在线监测

从新过程前20个批次的正常数据中任选一组, 从第100个采样点开始, 人为地引入故障异常, 反应产生的热量阶跃性增长30%, 以验证新模型的监测能力. 从图2的监测结果可以看出, 当反应产生的热量异常时, 两个监测指标均明显地超出控制限, 表明了新模型能及时监测出过程的异常情况.



(a) T^2 统计量



(b) SPE 统计量

图2 故障在线监测曲线

5 结 论

为满足市场需求, 生产工业过程需频繁变更操作条件, 以得到不同的产品结构或特性. 为了减少时间和成本的浪费, 本文提出了一种基于时序模型迁移的间歇过程监控建模策略, 只需一个新过程正常批次的的数据, 通过对旧过程变量建立时序模型, 利用模型迁移即可实现新过程的快速建模. 实验结果表明了该策略的有效性.

参考文献(References)

- [1] Jaeckle C M, MacGregor J F. Product design through multivariate statistical analysis of process data[J]. AIChE J, 1998, 44(5): 1105-1118.
- [2] Jaeckle C M, MacGregor J F. Industrial applications of product design through the inversion of latent variable

- models[J]. *Chemometrics and Intelligent Laboratory Systems*, 2000, 50(2): 199-210.
- [3] Jaeckle C M, MacGregor J F. Product transfer between plants using historical process data[J]. *AIChE J*, 2000, 46(10): 1989-1997.
- [4] García-Munoz S, MacGregor J F, Kourti T. Product transfer between sites using joint-Y PLS[J]. *Chemometrics and Intelligent Laboratory Systems*, 2005, 79(1/2): 101-114.
- [5] Lu J D, Gao F R. Process modeling based on process similarity[J]. *Industrial and Engineering Chemistry Research*, 2008, 47(1): 1967-1974.
- [6] Lu J D, Gao F R. Model migration with inclusive similarity for development of a new process model[J]. *Industrial and Engineering Chemistry Research*, 2008, 47(23): 9508-9516.
- [7] Lu J D, Yao K, Gao F R. Process similarity and developing new process models through migration[J]. *AIChE J*, 2009, 55(9): 2318-2328.
- [8] Lu J D, Yao Y, Gao F R. Model migration for development of a new process model[J]. *Industrial and Engineering Chemistry Research*, 2009, 48(21): 9603-9610.
- [9] Facco P, Tomba E, Bezzo F, et al. Transfer of process monitoring models between different plants using latent variable techniques[J]. *Industrial and Engineering Chemistry Research*, 2012, 51(21): 7327-7339.
- [10] Tomba E, Facco P, Bezzo F, et al. Combining fundamental knowledge and latent variable techniques to transfer process monitoring models between plants[J]. *Chemometrics and Intelligent Laboratory Systems*, 2012, 116(1): 67-77.
- [11] Cramer H, Leadbetter M R. Stationary and related random processes[M]. New York: John Wiley & Sons, 1967.
- [12] 韩冬梅, 高铁梅. 基于结构时间序列模型的季节调整方法研究[J]. *数量经济技术经济研究*, 2000, 17(3): 41-44. (Han D M, Gao T M. The seasonal adjustment method study on the structure time series model[J]. *J of Quantitative & Technical Economics*, 2000, 17(3): 41-44.)
- [13] 陈飞, 高铁梅. 结构时间序列模型在季节调整方面的应用[J]. *系统工程理论与实践*, 2007, 27(11): 7-14. (Chen F, Gao T M. The application of the structure time series model on seasonal adjustment[J]. *Systems Engineering-Theory & Practice*, 2007, 27(11): 7-14.)
- [14] 张金艳, 郭鹏江. 确定性时间序列模型及ARIMA模型的应用[J]. *西安邮电学院学报*, 2009, 14(3): 128-132. (Zhang J Y, Guo P J. Application of certainty time series analysis model and ARIMA model[J]. *J of Xi'an University of Post and Telecommunications*, 2009, 14(3): 128-132.)
- [15] Zhao C H, Wang F L, Gao F R, et al. Adaptive monitoring method for batch processes based on phase dissimilarity updating with limited modeling data[J]. *Industrial and Engineering Chemistry Research*, 2007, 46(14): 4943-4953.
- [16] Lu N Y, Gao F R, Wang F L. A sub-PCA modeling and on-line monitoring strategy for batch processes[J]. *AIChE J*, 2004, 50(1): 255-259.
- [17] Birol G, Undey C, Cinar A. A modular simulation package for fed-batch fermentation: Penicillin production[J]. *Computers & Chemical Engineering*, 2002, 26(11): 1553-1565.

(责任编辑: 李君玲)