

7-2-2020

Online Product Quality Prediction for Multi-phase Based on Local Model

Li Yuan

1. *College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China;*

Yayun Yan

1. *College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China;*

Xiaochu Tang

2. *Shenyang Aerospace University, Automation College, Shenyang 110136, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Online Product Quality Prediction for Multi-phase Based on Local Model

Abstract

Abstract: For offline quality prediction accuracy for batch process, an online prediction method for multi-phase product quality was proposed based on the local model. According to the repetitive cycle of batch process, *batch process could be divided into stable phase and transitional phase using the repeatability factor. The least squares support vector machine (LSSVM) model was established in stable phase using the time slice of same phase position, and the LSSVM model was established in transitional phase using optimal subset based on diffusion distance*, which made the natural properties of current stable phase and transitional phase be similar to the natural properties of historical stable phase and transitional phase respectively, and the product quality which is difficult to be measured could be obtained. The application to penicillin fermentation process generated in Pensim simulation platform shows that the method based on multi-phase online prediction has better predictive performance than overall offline prediction.

Keywords

repeatability factor, multi-phase, diffusion distance, least squares support vector machine, local prediction model

Recommended Citation

Li Yuan, Yan Yayun, Tang Xiaochu. Online Product Quality Prediction for Multi-phase Based on Local Model[J]. Journal of System Simulation, 2016, 28(4): 966-971.

基于局部模型的多阶段在线产品质量预测

李元¹, 燕亚运¹, 唐晓初²

(1. 沈阳化工大学信息工程学院, 辽宁 沈阳 110142; 2. 沈阳航空航天大学自动化学院, 辽宁 沈阳 110136)

摘要: 针对间歇过程产品质量离线预测精度不高的问题, 提出了一种基于局部模型的多阶段在线预测产品质量方法。利用间歇过程的周期重复性, 采用重复因子将间歇过程划分为稳定阶段和过渡阶段。对稳定阶段采用相同相位时间片建立最小二乘支持向量机(LSSVM)模型, 过渡阶段采用扩散距离(Diffusion Distance)选取最优子集建立 LSSVM 模型, 使得当前样本的稳定阶段和过渡阶段的属性能分别和历史数据集的稳定阶段和过渡阶段的属性相似, 预测出不易测量的产品质量。通过在 Pensim 仿真平台青霉素发酵过程中的应用表明, 与整体离线预测相比, 基于局部模型的多阶段在线预测方法有更好的预测性能。

关键词: 重复因子; 多阶段; 扩散距离; 最小二乘支持向量机; 局部预测模型

中图分类号: TP29

文献标识码: A

文章编号: 1004-731X (2016) 04-0966-07

Online Product Quality Prediction for Multi-phase Based on Local Model

Li Yuan¹, Yan Yayun¹, Tang Xiaochu²

(1. College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China;

2. Shenyang Aerospace University, Automation College, Shenyang 110136, China)

Abstract: For offline quality prediction accuracy for batch process, an online prediction method for multi-phase product quality was proposed based on the local model. According to the repetitive cycle of batch process, *batch process could be divided into stable phase and transitional phase using the repeatability factor. The least squares support vector machine (LSSVM) model was established in stable phase using the time slice of same phase position, and the LSSVM model was established in transitional phase using optimal subset based on diffusion distance*, which made the natural properties of current stable phase and transitional phase be similar to the natural properties of historical stable phase and transitional phase respectively, and the product quality which is difficult to be measured could be obtained. The application to penicillin fermentation process generated in Pensim simulation platform shows that the method based on multi-phase online prediction has better predictive performance than overall offline prediction.

Keywords: repeatability factor; multi-phase; diffusion distance; least squares support vector machine; local prediction model

引言

间歇过程在现代制造业和化工领域发挥了重

要作用, 近些年应用数据驱动的多元统计分析方法研究间歇过程的过程监视与质量预测也引起了很大关注^[1]。传统的多向主元分析(MPCA)^[2]和多向偏最小二乘(MPLS)^[3]广泛应用于间歇生产过程中, 然而这些传统的方法却忽视了间歇过程多阶段特征的特殊性, 如果间歇过程存在多阶段特性时, 每个阶段都会有自身的数据表征机制, 因此, 过程在不



收稿日期: 2014-12-13 修回日期: 2015-02-15;
基金项目: 国家自然科学基金重大项目(61490701),
国家自然科学基金(60774070, 61174119);
作者简介: 李元(1964-), 女, 辽宁沈阳, 教授, 博士,
研究方向为基于数据驱动复杂过程故障诊断; 燕亚运
(1990-), 男, 江苏泰州, 硕士生, 研究方向为质量预
测与控制。

<http://www.china-simulation.com>

• 966 •

同的阶段也会表现出不同的行为, 即使在每个阶段, 变量的关系在不同的时间段也会改变, 这是由于过程的动态性和时变性或不稳定的数据特征。针对间歇过程多阶段的特点, 基于多阶段的质量分析也被关注, 通过多阶段的质量预测可以找出产品质量和特定子阶段之间的关联以及子阶段中影响产品质量的关键过程变量, 且基于多阶段的在线质量预测不仅模型简单实用, 而且精度可靠。

对于间歇过程主要有 3 种方法可进行阶段划分^[4]。第 1 种是基于先验知识的, 这也是研究中用到最多的。然而这种方法的缺点是如果已知的知识不足以划分阶段, 将不能继续进行^[5]。第 2 种方法是基于过程分析技术, 但它是在某些特定的所需的功能事先已知的假设下进行的^[6]。最近另一种基于间歇过程数据分析的方法已经被提出, 其中有 2 个典型的方法: 一个是基于变量相关性的变化和基于 PCA 模型预测能力检测阶段划分点的 sub-PCA 方法^[7]。然而这种方法的主要缺点是任意 2 个相邻阶段间的过渡阶段没有很好的考虑。而过渡是指在 2 个特定阶段, 在一个时间段的数据特性是频繁改变的。在实际过程中, 大多数间歇过程从一个阶段到另一个阶段都会有明显的过渡行为, 这个过渡时间段的信息对建模和质量控制都会有很大的影响。另一个是基于 PCA 相似因子法^[8], 它能识别间歇过程的过渡阶段。然而这 2 种方法的共同缺点是阶段的数目先要指定, 这跟实际过程行为不匹配。因此本文提出基于间歇过程的重复性进行阶段辨识, 不同批次间有相似的行为, 尤其是在稳定阶段。相比较稳定阶段, 过渡阶段的重复性就很低, 意味着过渡阶段的不确定性比稳定阶段高。只要计算出每个时间间隔的重复性, 就可以将间歇过程划分为不同的稳定阶段和过渡阶段。

基于阶段划分的结果, 可在线实时预测产品的质量。目前, 比较常用的预测方法有人工神经网络法^[9-10]、线性回归分析法^[11-12]、支持向量机(SVM)^[13-14]和最小二乘支持向量机(LSSVM)^[15-16]等。神经网络是常用的时序预测方法, 有较强

的非线性预测能力, 但由于需要的训练样本较大, 而容易陷入局部极值, 导致预测的效果不稳定。线性回归分析建立了两组变量间的线性因果关系便于分析, 但可能忽略了交互效应和非线性的因果关系。LSSVM 是 SVM 的一种改进, 是支持向量机在二次损失函数下的一种形式, 它采用求解线性方程算法得到支持向量, 取代了二次规划求解优化问题, 求解的速度也变快很多, 很适合用来建立局部预测模型。

综上所述, 首先通过重复因子将间歇过程划分为不同的稳定阶段和过渡阶段。然后提出稳定阶段采用相同相位时间片建立 LSSVM 模型, 预测稳定阶段待测样本的质量输出, 而用扩散距离提取每个过渡阶段的最优子集并建立 LSSVM 的局部预测模型, 用来预测当前过渡阶段待测样本质量输出, 最后将本文方法与整体离线预测的误差率进行比较。

1 阶段辨识和在线质量预测

本节首先介绍扩散距离的基本概念, 然后基于扩散距离定义重复因子将间歇过程划分为稳定阶段和过渡阶段, 再根据稳定阶段和过渡阶段不同的自然属性, 分别建立局部预测模型在线预测产品质量。

1.1 阶段辨识

1.1.1 扩散距离

扩散距离^[17-18]是应用随机过程中的马尔可夫过程, 通过扩散映射(Diffusion Map)将数据集映射到高维特征空间。

假设 X 是一个数据 $X = \{x_1, x_2, \dots, x_n\} \in R^d$, 设 $I = \{1, 2, \dots, n\}$, 其中指数 d 代表变量数, n 表示 X 的样本数。如果假设每个 X 的样本点是一个加权图 G 的一个节点, 权重函数表示为 $k(x_i, x_j), \forall i, j \in I$, 它满足对称性和保正性, 即 $k(x_i, x_j) = k(x_j, x_i)$ 和 $k(x_i, x_j) \geq 0$ 。权重函数 k 选择高斯核函数:

$$k_{\beta} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\beta}\right) \quad (1)$$

这里的 β 是高斯核的宽度参数。在图 G 中节点 x 的度表示为:

$$\text{deg}(x) = \sum_{s \in X} k_{\beta}(x, s) \quad (2)$$

从节点 x_i 到节点 x_j 的一步转移概率:

$$p(x_i, x_j) = \frac{k_{\beta}(x_i, x_j)}{\text{deg}(x_i)} \quad (3)$$

它的概率代表马尔可夫随机游走过程的概率, 例如如果在图 G 中只有一个节点, 那么这个节点的转移概率就是 1, 它只能回到自己原来的位置, 因为没有其他路线可走, 对于一个转移了 t 步的转移概率定义它为一歩转移概率的 t 次方即 P^t , 表示为 $P_t(x_i, x_j)$ 。那么 x_i 与 x_j 之间的扩散距离表示如下:

$$D_t^2(x_i, x_j) = \sum_{s \in X} \frac{(P_t(x_i, s) - P_t(x_j, s))^2}{q(s)} \quad (4)$$

而这里的 q 计算如下:

$$q(x) = \frac{\text{deg}(x)}{\sum_{s \in X} \text{deg}(s)} \quad (5)$$

扩散距离反映了数据之间连接的几何内涵, 一对节点间的路径越多, 那么它们之间的扩散距离就越小, 如图 1 所示有两类数据集, A, C 都属于第一类, B 属于第二类, 虚线表示扩散距离, 实线表示欧氏距离, 由图可以看出 A 到 B 和 C 的欧式距离是一样的, 但 A 到 C 之间的路径比 A 到 B 的路径要多, 那么 A 到 C 的扩散距离要小于 A 到 B 的扩散距离。

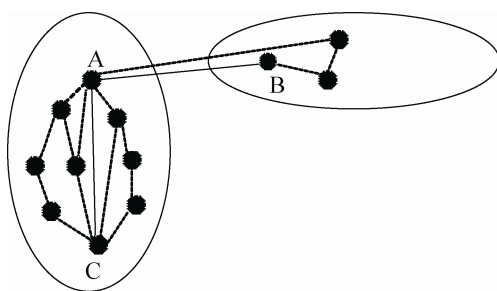


图 1 扩散距离和欧式距离

1.1.2 重复因子

假设一个间歇过程包含 I 个批次, J 个变量, 每个批次采样 K 次。这样历史数据集就可以用一

个三维数据矩阵表示 $X(I \times J \times K)$ 。将三维数据按批次展开成二维的数据矩阵 $X(I \times JK)$ 。

为了估计每个时间片的重复性, 定义重复因子如下:

$$RF_k(i, j) = D_t^2(x_i^k, x_j^k) \quad (6)$$

这里的 $i, j = 1, 2, \dots, I$, $k = 1, 2, \dots, K$, x_i^k 是第 k 个时间片矩阵 $X_k(I \times J)$ 的第 i 个向量。在特定时间片两批次间的重复性通过计算两批次数据向量的扩散距离得到。因此扩散距离越小, 重复性越高, RF 值越小。如果 2 个批次有不同的数据行为, 则两个批次间的扩散距离较大, RF 值就会较大, 相反则会较小。每个时间片的重复因子是该时间片内所有批次间的重复因子值的平均值^[19]:

$$\overline{RF}_k = \frac{1}{I^2} \sum_i \sum_j RF_k(i, j) \quad (7)$$

在间歇过程中, 由于从一个阶段到另一个阶段间的过渡阶段有更高的不确定性存在, 因此在过渡阶段的 RF 值一般会大于稳定阶段的 RF 值。

1.2 在线预测间歇过程的质量输出

当阶段被辨识后, 如图 2 所示, 细线代表历史批次, 粗线代表当前批次, 红叉代表这一批次的一个采样时刻, 总共分稳态—过渡—稳态 3 个阶段, 在示意图中稳态的红叉都在一条竖直线上, 而过渡的红叉不一定, 即在稳定阶段相同采样时刻是相似的, 但对于过渡阶段并不一定是相同采样时刻最相似, 很有可能发生错位。这是由于对于每一个稳定阶段, 数据的相关结构在同一采样间隔是相似的, 因此可以选择相同相位时间片的数据建立预测模型。然而, 在过渡阶段, 相同采样时刻不同批次数据的相关性不同, 而从一个时间片到另一个时间片数据相关性也将发生改变, 这种情况下, 单个预测模型可能不足以代表整个周期模型, 因为它几乎不能捕获所有相关的信息。本文提出一种局部建模方法可以更好的在线预测过渡阶段的产品质量, 局部预测模型通过扩散距离和 LSSVM 方法相结合建立。该方法主要分 3 步完成: (1) 当获得一个新的

样本时, 比较新样本过程输入与历史数据过程输入的扩散距离。(2) 选取与新样本过程输入扩散距离最小(即最相关的)的 L 个历史样本作为 LSSVM 局部建模的训练样本。(3) 根据当前建立的局部模型预测当前新样本的质量输出。具体流程如图 3 所示。

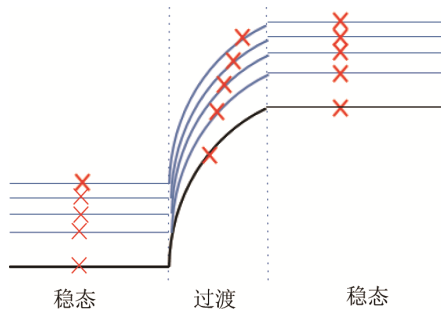


图 2 相似性示意图

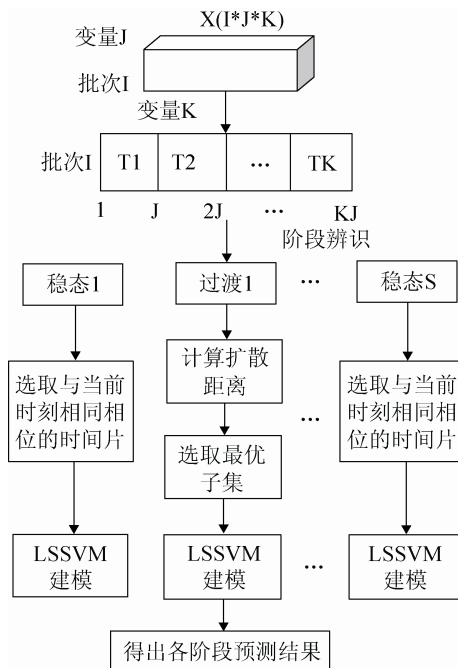


图 3 在线局部模型质量预测流程图

设所选取的 (X_L, Y_L) 作为 LSSVM 的训练集的过程输入和质量输出, 那么如果有 m 个时间片就需要建立 m 个 LSSVM 模型, $Lopt$ 值表示 L 的最优值, 可由下面的验证误差最小化得到^[20]:

$$E_k = \|y_k - f(x_k)\| = \left\| y_k - \sum_{i=1}^L \alpha_i K(x_i, x_k) + b \right\| \quad (8)$$

这里的高斯核函数, $K(u_i, u_j) = \exp(-\|u_i - u_j\|^2 / \sigma^2)$ 是选择的高斯核函数, σ 是核宽参数, k 是验证集大小,

α_i 是拉格朗日系数。然而实际上 $Lopt$ 的选取需要考虑计算负担和在线建模需求, 因此 $Lopt$ 的选取是个权衡问题。但总体上 LSSVM 建模所需的训练样本数要远远小于全部样本建模的训练样本数, 所以计算量是可以保证的。因此当前样本数据 u_c 的最优预测输出计算如下:

$$f(u_c) = \langle w, \varphi(u_c) \rangle + b = \sum_{i=1}^{Lopt} \alpha_i K(u_i, u_c) + b \quad (9)$$

2 青霉素发酵过程的应用

青霉素发酵过程是一种典型的非线性、多阶段间歇生产过程, 它是青霉素产生菌在合适的培养基、温度、pH 值、搅拌等发酵条件下生长和合成抗生素的代谢活动。由于青霉素发酵过程的严重非线性和不确定性, 研究青霉素发酵过程有着重要的现实意义, 本文将在 Pensim 仿真平台上进行研究^[21]。青霉素发酵工艺流程如图 4 所示, 青霉素发酵过程从操作角度主要分为 2 个阶段, 初始阶段是间歇过程, 这一过程是底物的快速消耗和菌体的迅速成长过程, 并无青霉素产物产生。当菌体消耗完初始加入的葡萄糖底物后反应进入下一个阶段。第 2 阶段是半间歇补料操作阶段, 该阶段为青霉素的合成和菌体死亡阶段。青霉素发酵过程较为复杂, 每个微小操作都会影响产品质量, 且产品质量又难以测量, 所以根据已知的过程变量在线预测产品质量具有实际的意义。

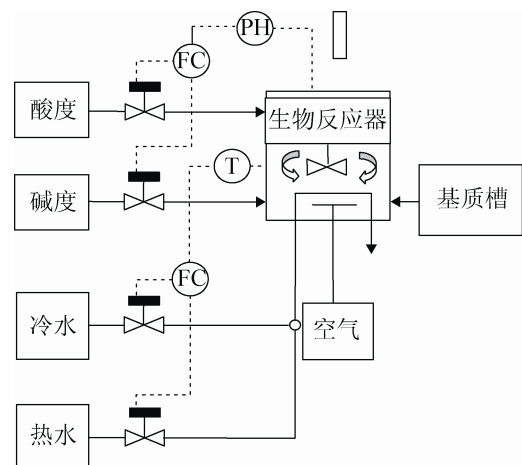


图 4 青霉素发酵流程图

本文选取 20 个正常批次数据, 10 个过程变量和 2 个质量变量(如表 1 所示), 每个批次采样数 400(每 1 h 采样 1 次)作为历史数据集。

表 1 青霉素发酵过程的过程变量和质量变量

No.	过程变量
1	通风率
2	搅拌功率
3	底物流加速率
4	底物流加温度
5	溶解氧浓度
6	生物量浓度
7	二氧化碳浓度
8	PH
9	发酵反应器温度
10	反应产生的热量
No.	质量变量
1	菌体浓度
2	青霉素浓度

首先要对这 20 个批次进行阶段辨识, 得到重复因子如图 5 所示。由于过渡过程的重复因子要大于稳定过程的重复因子, 粗线代表分界线, 分界线上面的就是辨识出的过渡阶段, 分界线下面的左边对应了青霉素过程的第 1 阶段, 右边是第 2 阶段。由图可以辨识出 1~29 为第 1 阶段, 30~44 为过渡阶段, 45~400 为第 2 阶段。定义分界线:

$$L = \frac{1}{N} \sum_i^N \overline{BF} \quad (10)$$

本文是预测出当前批次的产品质量, 根据历史数据阶段划分结果同样可以将当前批次划分为 3 个阶段, 对于第 1 阶段和第 3 阶段都采用与历史数据集相同相位的时间片建立 LSSVM 模型预测, 而过渡过程采用扩散距离选取最优子集建立局部模 LSSVM 型的方法, 用该方法在线预测菌体浓度和青霉素浓度, 并与整体离线预测结果进行比较(如图 6, 7 所示)。

图 6、7 中蓝色代表实际值, 红线代表分阶段在线预测值, 绿线代表整体离线预测值, 由图中的局部放大图可以看出分阶段在线预测 2 个阶段转折处的值更接近实际值。这是由于整体离线建模时

并没有考虑到过渡阶段变量的不稳定性。在稳定阶段操作时段的不同采样时刻的过程变量和质量变量之间的相关关系并没有显著变化, 但在过渡阶段却体现出不同的统计关系。因此必须找到跟这个阶段相关的过程数据建立质量预测模型才能较为精确的预测出产品质量。

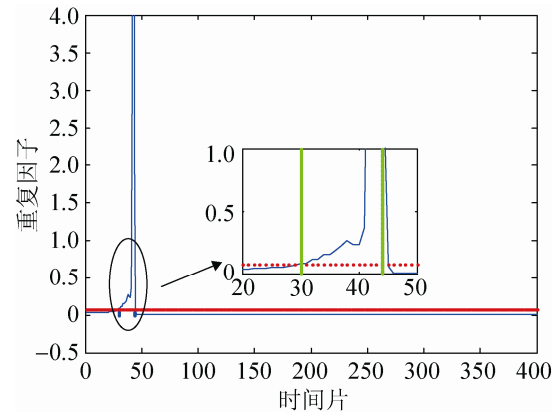


图 5 阶段辨识

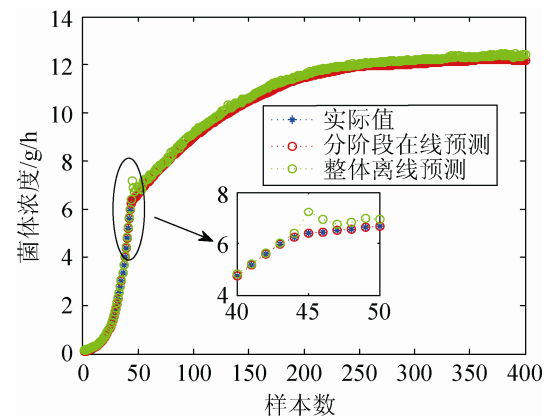


图 6 菌体浓度预测

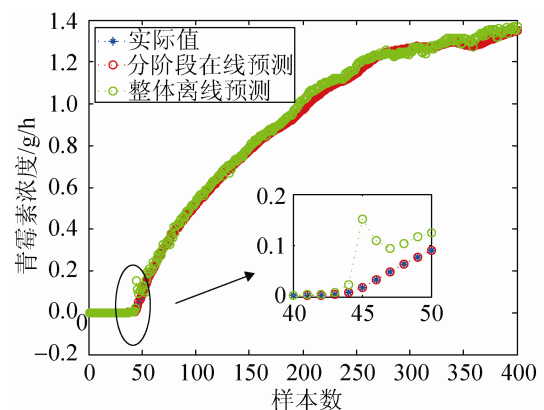


图 7 青霉素浓度预测

图 8, 9 分别给出了分阶段在线预测和整体离线预测的菌体浓度和青霉素浓度误差图, 并进行了比较。其中红色代表分阶段在线预测误差, 蓝线代表整体离线预测误差。预测误差由实际值减去预测值后取绝对值得到, 从图 8、9 可以看出蓝线基本都在红线上方, 即分阶段在线预测要比整体离线预测的预测精度要高。由此可以得出并不是用于预测的数据越多, 预测精度越好, 有时数据越多, 预测精度反而越差。这是因为如果过程变量和质量变量之间的相关关系没有发生明显变化, 那么数据的统计关系也不会发生显著变化, 过程的内部机制没有发生本质改变, 因此采用对应相位时间片建模能更好的反应过程变量与质量变量间的相关关系, 且过程内部机制没有发生本质改变, 但整体离线建模, 随着时刻的变化, 过程内部机制也会发生变化, 此时并不能很好的反应过程变量与质量变量间的相关关系, 预测精度会较差。

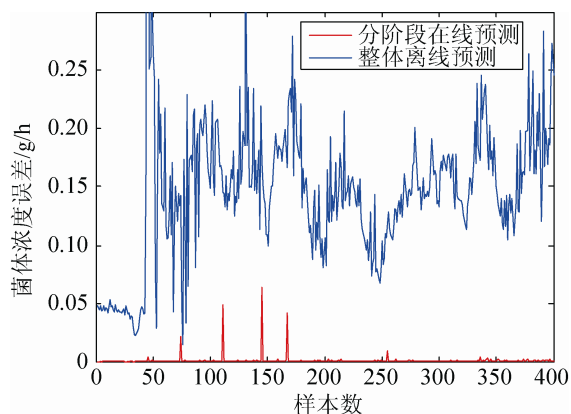


图 8 菌体浓度误差

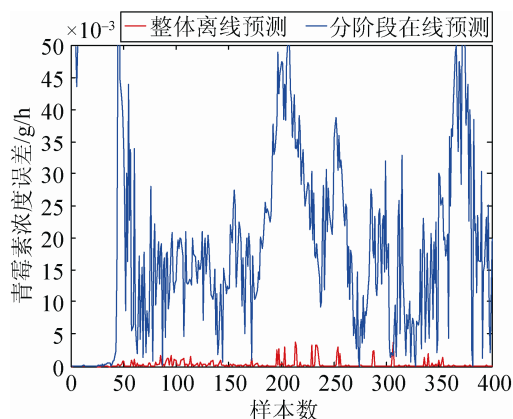


图 9 青霉素浓度误差

3 结论

本文提出了分阶段在线预测产品质量方法, 主要通过扩散距离选取最优子集, 建立局部模型预测过渡阶段产品质量, 选取相同相位时间片建立局部模型预测稳定阶段产品质量的方法。由于局部建模需要的样本量很小, 该方法降低了模型的复杂性, 而且对变量不稳定的过渡阶段也有很好的预测效果。这种多阶段在线预测方法预测青霉素的质量指标表现出较好的预测性, 是一种较为有效的在线预测产品质量的方法。

参考文献:

- [1] 周东华, 李钢, 李元. 数据驱动的工业过程故障诊断技术 [M]. 北京: 科学出版社, 2011.
- [2] P Nomikos, J F MacGregor. Monitoring batch processes using multiway principal component analysis [J]. *AIChE Journal* (S0001-1541), 1994, 40(8): 1361-1375.
- [3] P Nomikos, J F MacGregor. Multi-way partial least square in monitoring batch processes [J]. *Chemometrics and Intelligent Laboratory Systems* (S0169-7439), 1995, 30(1): 97-108.
- [4] Y Yao, F R Gao. A survey on multistage/multiphase statistical modeling methods for batch processes [J]. *Annual Reviews in Control* (S1367-5788), 2009, 33(2): 172-183.
- [5] S P Reinikainen, A Hoskuldsson. Multivariate statistical analysis of a multi-step industrial process [J]. *Analytica Chimica Acta* (S0003-2670), 2007, 595(1-2): 248-256.
- [6] X T Doan, R Srinivasan. Online monitoring of multi-phase batch processing phase-based multivariate statistical process control [J]. *Computers and Chemical Engineering* (S0098-1354), 2008, 32(1): 230-243.
- [7] N Y Lu, F R Gao. Stage-based process analysis and quality prediction for batch processes [J]. *Industrial and Engineering Chemistry Research* (S0888-5885), 2005, 44(10): 3547-3555.
- [8] Y Yao, F R Gao. Phase and transition based batch process modeling and online monitoring [J]. *Journal of Process Control* (S0959-1524), 2009, 19(5): 816-826.
- [9] Philippe Lauret, Eric Fock, Rija N. Randrianarivony. Bayesian neural network approach to short time load forecasting [J]. *Energy Conversion and Management* (S0196-8904), 2008, 49(5): 1156-1166.
- [10] 张捷, 薄煜明. 基于神经网络预测的网络控制系统故障检测 [J]. *南京理工大学学报*, 2010, 34(1): 19-23.

(下转第 978 页)