

基于统计特征的不等长间歇过程故障诊断研究*

郭金玉, 赵璐璐, 李元

(沈阳化工大学 信息工程学院, 沈阳 110142)

摘要: 为了提高不等长间歇过程故障诊断的性能,同时降低算法的复杂度,提出了一种基于统计特征的不等长间歇过程故障诊断算法。首先计算每个不等长批次的均值、方差、偏度、峭度和任意两个变量间的欧氏距离,并将这些统计特征组合成一个等长的特征向量;然后运用主元分析(PCA)进行过程监视。半导体工业实例的仿真结果表明,与传统的多向主元分析(MPCA)方法相比,基于统计特征的不等长间歇过程故障诊断算法的故障诊断率提高15%,故障检测时间减少了0.002 s,因此该算法具有很好的故障诊断性能。

关键词: 故障诊断; 不等长间歇过程; 统计特征; 多向主元分析

中图分类号: TP277

文献标志码: A

文章编号: 1001-3695(2014)01-0128-03

doi:10.3969/j.issn.1001-3695.2014.01.030

Fault diagnosis for uneven-length batch processes based on statistic features

GUO Jin-yu, ZHAO Lu-lu, LI Yuan

(College of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China)

Abstract: In order to improve the fault diagnosis performance of the uneven-length batch processes, and decrease the complexity of the algorithm, this paper presented fault diagnosis method based on statistic features for uneven-length batch processes. Firstly, it calculated the means, variance, skewness, kurtosis and the Euclidean distance between two variables for each uneven-length batch. Secondly, it combined these statistic features into an even-length feature vector. Lastly, it used principal component analysis (PCA) to the feature vectors for monitoring the batch processes. The monitoring results of an industrial example show that compared with traditional MPCA, the fault diagnosis method based on statistic features for uneven-length batch processes increases 15% of the fault diagnosis rate and reduces 0.002 second of the fault diagnosis time, so it has good fault detection performance.

Key words: fault diagnosis; uneven-length batch processes; statistic features; multiway principal component analysis (MPCA)

0 引言

间歇过程一直是生产高附加值产品的重要生产过程。化工间歇过程主要是生产精细的化工产品,对间歇过程反应机理的监视、物料平衡的控制以及生产节奏的把握都是直接影响和制约生产的能力与水平、产量与效益的关键要素。多元统计分析是进行间歇过程监控和故障诊断的一种有效方法。Nomikos等人^[1,2]于1995年提出了用多向主元分析(MPCA)和多向偏最小二乘(multiway partial least squares, MPLS)^[3]解决间歇过程的建模和监视,这种方法随后得到了很大的发展^[4~8]。为了运用传统的统计分析方法,假设批次数据是等长的,因此操作事件在批次上是同步化的。然而,这种假设通常是不成立的^[9,10]。

解决间歇过程批次不等长问题通常采用最短长度法。最短长度法直接按最短批次数据轨迹的长度切割其余批次,这种方法简单,但使数据轨迹的过程信息大量丢失,并且使点对点数据的相关性降低,导致此数据的可靠性降低。Kassidas等人^[10]运用动态时间规整(dynamic time wrapping, DTW)解决间

歇过程批次不等长问题。DTW运用动态规划的原理进行多元批次轨迹的同步化,以得到其间的最短距离。在批次的轨迹中,它会适时地转换、扩张或压缩一些局部片段,最终实现轨迹的同步匹配。这种方法虽然按照轨迹中点与点的模式进行动态匹配,但其处理过程的复杂性与其离线性导致其实际应用的困难。Rothwell等人^[11]表明DTW方法在故障检测方面的效果劣于最短长度法和指示变量法。Lu等人^[12]提出一种基于阶段子PCA的建模方法来解决批次不等长问题,其中每个阶段建立两个PCA模型,一个用于阶段划分,另一个用于过程监控。文献[13]运用高斯混合模型提取不等长间歇过程的阶段和过渡过程信息。文献[14]提出改进的统计分析和在线监控方案,用于处理不等长的多阶段间歇过程。在划分阶段的基础上,根据数据特性的变化,把不规则的批次分成不同的组,然后对同一阶段所有不同的组建立共同模型和特定模型。然而在大规模的生产过程中这些方法需要建立成千上万个模型,其计算量和存储量都非常大,阻止了这些方法在大规模生产过程中的在线应用。为了提高不等长间歇过程故障诊断的性能,同时降低算法的复杂度,本文提出了一种基于统计特征的不等长间歇过程的故障诊断算法。

收稿日期: 2013-03-24; **修回日期:** 2013-04-28 **基金项目:** 国家自然科学基金重点资助项目(61034006);国家自然科学基金资助项目

(61174119);辽宁省博士启动基金项目(20131089);辽宁省教育厅资助项目(L2012139)

作者简介: 郭金玉(1975-),女,山东高唐人,副教授,博士,主要研究方向为故障诊断、生物特征识别(shandong401@sina.com);赵璐璐(1986-),女,硕士研究生,主要研究方向为故障诊断;李元(1964-),女,教授,博士,主要研究方向为故障诊断。

1 不等长问题

在每个批次($i=1,2,\dots,I$)的运行中,假设在持续时间不固定的操作循环中,在 $k(k=1,2,\dots,K_i)$ 个时刻在线测量 J 个过程变量,形成每个不规则的批次数据集合,表示为 $X_i(J \times K_i)$ 。图1中参考批次 $X_i(J \times K_i)(i=1,\dots,I)$ 有不同的轨迹长度 K_i 。

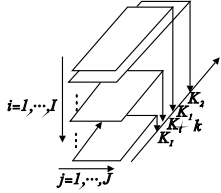


图1 不等长批次过程的三维数据描述

2 基于统计特征的不等长间歇过程故障诊断

基于统计的方法用统计特征组成的随机模型来重新定义和衡量原始间歇过程数据。常见的统计特征包含均值、方差、偏度、峭度和欧氏距离等。假定用 X_k 表示每个批次的样本数据:

$$X_k = [x_1 \quad x_2 \quad \dots \quad x_m] = \begin{bmatrix} x_1(k-w+1) & x_2(k-w+1) & \dots & x_m(k-w+1) \\ x_1(k-w+2) & x_2(k-w+2) & \dots & x_m(k-w+2) \\ \vdots & \vdots & & \vdots \\ x_1(k) & x_2(k) & \dots & x_m(k) \end{bmatrix} \quad (1)$$

其中: m 为变量个数; w 为批次中样本数; k 为批次中最大时刻。

每个批次的均值 μ 和方差 ν 分别为^[15]

$$\mu_i = \frac{1}{w} \sum_{l=0}^{w-1} x_i(k-l) \quad (2)$$

$$\nu_i = \frac{1}{w} \sum_{l=0}^{w-1} (x_i(k-l) - \mu_i)^2 \quad (3)$$

每个批次不同变量之间的欧氏距离 $d_{i,j}$ 为

$$d_{i,j} = \frac{1}{w} \sum_{l=0}^{w-1} \|x_i(k-l) - x_j(k-l)\|^2 \quad (4)$$

每个批次的偏度 r_i 和峭度 k_i 分别为^[15]

$$\gamma_i = \frac{\frac{1}{w} \sum_{l=0}^{w-1} [x_i(k-l) - \mu_i]^3}{\left(\frac{1}{w} \sum_{l=0}^{w-1} [x_i(k-l) - \mu_i]^2 \right)^{\frac{3}{2}}} \quad (5)$$

$$\kappa_i = \frac{\frac{1}{w} \sum_{l=0}^{w-1} [x_i(k-l) - \mu_i]^4}{\left(\frac{1}{w} \sum_{l=0}^{w-1} [x_i(k-l) - \mu_i]^2 \right)^2} - 3 \quad (6)$$

每个批次的统计特征用 $F = [\mu_i \quad \nu_i \quad d_{i,j} \quad r_i \quad k_i]$ 表示,因此所有的统计特征组合成 $4m + \frac{m \times (m-1)}{2}$ 维的特征向量:

$$F = [\mu_1, \nu_1, d_{1,2}, d_{1,3}, \dots, d_{1,m}, d_{2,3}, d_{2,4}, \dots, d_{2,m}, \dots, d_{m-1,1}, d_{m-1,2}, \dots, d_{m-1,m}, \kappa_1, \dots] \quad (7)$$

由此可见,不等长间歇过程的统计特征向量 F 是等长的。

对标准化后的统计特征向量 F 进行主元分析,如果只取前 k 个主元,得到下面的主元模型:

$$F = t_1 p_1^T + t_2 p_2^T + \dots + t_k p_k^T + E \quad (8)$$

式中: t 被称为得分(score)向量; p_i 被称为载荷(loading)向量; F 的得分向量也叫做 F 的主元。

一旦建立了过程监视和故障检测的模型,就需要计算操作限和警告限。典型的监视控制图包括预测误差平方(squared prediction error, SPE)图和主元得分图。当主元模型的SPE或

得分超出控制限时,就认为过程中出现了不正常情况。

主元模型的预测误差平方和SPE也被称做 Q 统计量。

Q 统计量定义为

$$Q = ee^T = X(I - P_k P_k^T) X^T \quad (9)$$

其中: e 是残差矩阵; $P_k = [p_1 p_2 \dots p_k]$; I 为 $n \times n$ 的单位阵。SPE或 Q 统计量代表了数据中没有被主元模型所解释的变化。当SPE过大时,说明过程中出现了故障现象,从过程正常运行数据所建立的模型已不再适用。控制限的计算是建立在一定的假设基础上的,当检验水平为 α 时,SPE控制限按式(10)计算^[16]:

$$Q_{UCL} = a(b + cz_\alpha)^d \quad (10)$$

其中: z_α 为标准正态分布的 100α 百分点,其他参数如下:

$$a = \sum_{i=k+1}^n \lambda_i, b = 1 + [\theta_2 h_0 (h_0 - 1)] / a^2$$

$$c = (\sqrt{2\theta_2 h_0}) / a, d = 1/h_0$$

$$\theta_2 = \sum_{i=k+1}^n \lambda_i^2, \theta_3 = \sum_{i=k+1}^n \lambda_i^3, h_0 = (1 - 2a\theta_3) / (3\theta_2^2) \quad (11)$$

基于统计特征的不等长间歇过程故障诊断算法的具体步骤为:

a)建立正常状态的模型

(a)收集正常操作时各批次历史数据。

(b)计算每个不等长批次的均值、方差、偏度、峭度和任意两个变量间的欧氏距离,并将这些统计特征组合成一个等长的特征向量。

(c)建立PCA模型。

(d)确定统计量SPE的控制限。

b)故障诊断

(a)计算新批次等长的统计特征向量,并对其进行标准化。

(b)计算新批次数据的统计量SPE。

(c)根据SPE值是否超过建模控制限来判断新的批次是否异常。如果SPE值超过控制限,则新的批次是故障的;否则是正常的。

3 仿真实验

本文应用半导体工业实例——A1堆腐蚀过程比较不同故障诊断方法的性能。半导体工业数据由108个正常硅片和21个故障硅片构成。由于两个批次过程数据的大量丢失,本文应用107个正常硅片和20个故障硅片进行仿真研究。关于故障的更多描述详见文献[17]。由于文献[17]中的步骤4和5是主要的腐蚀步骤,仿真仅应用这两个步骤中的数据点,进行故障诊断的17个非定点过程变量如表1所示。这个物理问题表明这些变量应该与过程和最终产品的状态有关。正如前面所指出的,半导体过程有些独特的特性,这些特性包括不等长的批次时间、不等长的阶段及过程漂移和转变。两个例子(变量EndPt A和TCP Load)表明了这些特性,如图2所示。

表1 A1堆腐蚀过程的故障诊断变量

序号	过程变量	序号	过程变量
1	BC13	10	RF Pwr
2	C12 Flow	11	RF Impedance
3	RF Btm Pwr	12	TCP Tuner
4	RF Btm Rfl Pwr	13	TCP Phase Err
5	EndPt A	14	TCP Impedance
6	Pressure	15	TCP Top Pwr
7	RF Tuner	16	TCP Load
8	RF Load	17	Vat Valve
9	RF Phase Err		

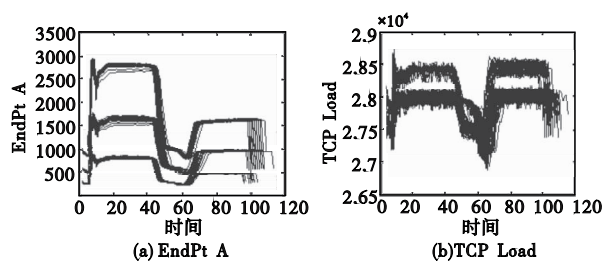


图2 数据集中的两个变量

仿真实验中,每个批次有 17 个过程变量。每个批次是不等长的,持续时间在 95 ~ 112 s 之间变化。随机抽取 96 个正常批次为建模数据,其余 11 个正常批次为校验数据。为了消除传感器中初始的波动影响,去除开始的 5 个样本,保留 85 个样本以适应最短的批次。对 96 个等长的正常批次数据运用 MPCA 方法进行建模,并对 11 个校验批次数据和 20 个故障批次进行故障检测。MPCA 方法的 SPE 检测结果如图 3 所示。MPCA 方法检测出了故障 4、7、12 和 13,然而,由于它们的数值较大,没有显示在图 3 中。

运用本文方法进行故障诊断,不需要数据预处理。直接计算 96 个不等长批次的均值、方差、偏度、峭度和任意两个变量间的欧氏距离,并将这些统计特征组合成一个等长的特征向量,然后运用 PCA 进行过程监视。本文方法的 SPE 检测结果如图 4 所示。本文方法检测出了故障 4 和 12,同样没有显示在图 4 中。两种方法的 SPE 检测结果对比如表 2 所示。从表 2 可以看出,本文方法和传统 MPCA 对正常批次的 SPE 检测率均为 100%;本文对故障批次的 SPE 故障检测率为 80%,而传统的 MPCA 的检测率为 65%,证明了运用本文方法进行故障检测的准确率远远高于传统的 MPCA 方法,大大提高了故障检测率,减少了故障的误报率和漏报率。

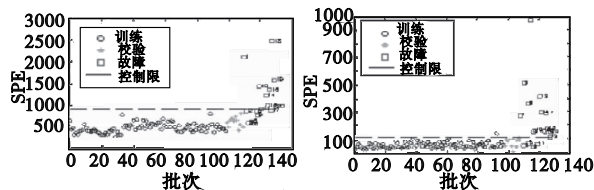


图3 MPCA方法的SPE检测

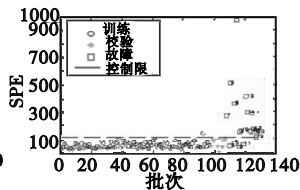


图4 本文方法的SPE检测

表2 SPE 监测结果对比

方法	监测项目				故障批次 检测比例/%
	校验批次 总数	检测的校验 批次总数	校验批次 检测比例/%	故障批次 总数	
MPCA	11	11	100	20	65
本文	11	11	100	20	80

表 3 是在 MATLAB 7.1 平台下传统 MPCA 和本文方法故障检测所需的 CPU 时间 (CPU: Pentium Dual-core 2.93 GHz, RAM: 1.96 GB)。由表 3 可知, MPCA 方法需要计算 1445×1445 维的协方差矩阵,而本文方法协方差矩阵的维数仅为 204×204 ,减少了存储空间。与 MPCA 方法相比,本文方法降低了特征向量的维数,但是故障检测时间仅比 MPCA 方法减少了 0.002 s。这是由于本文方法在运用 MPCA 进行故障诊断之前,需要先计算统计特征向量,增加了时间。

表3 MPCA 与本文方法故障检测对比

方法	特征向量大小	协方差矩阵大小	故障检测时间/s
MPCA	1445×3	1445×1445	0.031 9
本文	204×3	204×204	0.031 7

4 结束语

传统的 MPCA 方法可以有效地对等长批次生产过程中存在的故障进行检测和诊断,但是由于间歇过程具有间歇性、操作的不确定性、反应过程的多模态及不等长等特性,致使 MPCA 不能直接应用于间歇过程故障诊断。本文提出的基于统计特征的不等长间歇过程故障诊断方法可以有效地解决这些问题,使故障检测结果更加准确。同时,本文算法减少了计算量和存储空间,降低了故障检测时间。

参考文献:

- [1] NOMIKOS P, MACGREGOR J F. Monitoring batch processes using multiway principal component analysis[J]. *AIChE Journal*, 1994, 40(8): 1361-1375.
- [2] NOMIKOS P, MACGREGOR J F. Multivariate SPC charts for monitoring batch processes[J]. *Technometrics*, 1995, 37(1): 41-59.
- [3] NOMIKOS P, MACGREGOR J F. Multi-way partial least squares in monitoring batch processes[J]. *Chemometrics and Intelligent Laboratory Systems*, 1995, 30(1): 97-108.
- [4] CAMACHO J, PICÓ J. Multi-phase principal component analysis for batch processes modelling[J]. *Chemometrics and Intelligent Laboratory Systems*, 2006, 81(2): 127-136.
- [5] CAMACHO J, PICÓ J. Online monitoring of batch processes using multi-phase principal component analysis[J]. *Journal of Process Control*, 2006, 16(10): 1021-1035.
- [6] HUA Kun-lun, YUAN Jing-qi. Multivariate statistical process control based on multiway locality preserving projections[J]. *Journal of Process Control*, 2008, 18(7-8): 797-807.
- [7] YU Jie, QIN S J. Multiway Gaussian mixture model based multiphase batch process monitoring[J]. *Industrial & Engineering Chemistry Research*, 2009, 48(18): 8585-8594.
- [8] GUO Jin-yu, LI Yuan, WANG Guo-zhu, et al. Batch process monitoring based on multilinear principal component analysis[C]//Proc of International Conference on Intelligent Systems and Design and Engineering Applications. 2010: 413-416.
- [9] CHANG Yu-qing, LU Yun-song, WANG Fu-li, et al. Sub-stage PCA modelling and monitoring method for uneven-length batch processes[J]. *The Canadian Journal of Chemical Engineering*, 2012, 90(1): 144-152.
- [10] KASSIDAS A, MACGREGOR J F, TAYLOR P A. Synchronization of batch trajectories using dynamic time warping[J]. *AIChE Journal*, 1998, 44(4): 864-875.
- [11] ROTHWELL S G, MARTIN E B, MORRIS A J. Comparison of methods for dealing with uneven length batches[C]//Proc of the 7th International Conference on Computer Application in Biotechnology. 1998: 387-392.
- [12] LU Ning-yun, GAO Fu-rong, YANG Yi, et al. PCA-based modeling and on-line monitoring strategy for uneven-length batch processes[J]. *Industrial & Engineering Chemistry Research*. 2004, 43(13): 3343-3352.
- [13] YAO Yuan, DONG Wei-wei, ZHAO Lu-ping, et al. Multivariate statistical monitoring of multiphase batch processes with uneven operation durations[J]. *The Canadian Journal of Chemical Engineering*, 2012, 90(6): 1383-1392.
- [14] ZHAO Chun-hui, MO Sheng-yong, GAO Fu-rong, et al. Statistical analysis and online monitoring for handling multiphase batch processes with varying durations[J]. *Journal of Process Control*, 2011, 21(6): 817-829.
- [15] WANG Jin, HE Q P. Multivariate statistical process monitoring based on statistics pattern analysis[J]. *Industrial & Engineering Chemistry Research*, 2010, 49(17): 7858-7869.
- [16] GARCIA-ALVAREZ D, FUENTE M J, SAINZ G. I. Fault detection and isolation in transient states using principal component analysis[J]. *Journal of Process Control*, 2012, 22(3): 551-563.
- [17] WISE B M, GALLAGHER N B, BUTLER S W, et al. A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process[J]. *Chemometrics*, 1999, 13(3-4): 379-396.