

基于 GMM 的间歇过程故障检测

王静¹ 胡益¹ 侍洪波¹

摘要 对间歇过程的多操作阶段进行划分时, 往往会被离群点和噪声干扰, 影响建模的精确性, 针对此问题提出一种新的方法: 主元分析-多方向高斯混合模型 (Principal component analysis-multiple Gaussian mixture model, PCA-MGMM) 建模方法. 首先用最短路法对数据进行等长处理, 融合不同展开方法相结合的处理方式消除数据预估问题; 利用主元分析方法将数据转换到对故障较为敏感的低维子空间中, 得到主元的同时消除了离群点和噪声的干扰; 通过改进的高斯混合模型 (Gaussian mixture model, GMM) 算法对各阶段主元进行聚类, 减少了运算量的同时自动得到最佳高斯成分和对应的统计分布参数; 最后将局部指标融合为全局概率监控指标, 实现了连续的在线监控. 通过一个实际的半导体制造过程的仿真研究验证了所提方法的有效性.

关键词 间歇过程, 多阶段操作, 故障检测, 高斯混合模型, 全局概率指标

引用格式 王静, 胡益, 侍洪波. 基于 GMM 的间歇过程故障检测. 自动化学报, 2015, 41(5): 899-905

DOI 10.16383/j.aas.2015.c130680

Fault Detection for Batch Processes Based on Gaussian Mixture Model

WANG Jing¹ HU Yi¹ SHI Hong-Bo¹

Abstract A new novel of principal component analysis-multiple Gaussian mixture model (PCA-MGMM) method is proposed in this article to handle the problem about outliers and noise interference, which affects the accuracy of modeling when dividing multiple operation phases in batch processes. At first, a shortest length approach is used to solve the problem of unequal data, while a method of hybrid unfolding of a multi-way data matrix is used to eliminate data estimate problem. Secondly, using PCA sequentially to achieve a low-dimensional representation of the original data space, the operation not only gets the principal but also eliminates the outliers and noise interference. After that the modified algorithm of Gaussian mixture model (GMM) is adopted to automatically optimize the number of Gaussian components and estimate their statistical distribution parameters so as to reduce the computational complexity. Finally, the online monitoring is guaranteed to be continuous by using a global probability index integrated by local probability indices of each operation. The effectiveness and flexibility of the proposed method is validated through an empirical study on a real semiconductor process.

Key words Batch processes, multiple operation phases, fault detection, Gaussian mixture model (GMM), global probability index

Citation Wang Jing, Hu Yi, Shi Hong-Bo. Fault detection for batch processes based on Gaussian mixture model. *Acta Automatica Sinica*, 2015, 41(5): 899-905

间歇过程已经被广泛应用到半导体、化工、发酵、制药等许多领域, 是现代工业一种重要的生产方式, 与连续生产相比, 过程更加复杂、多变, 操作条件要求更加苛刻, 即使微小的异常状况都会影响最终产品的产量和质量. 间歇过程研究缺少一定的基础知识, 过程模型的细节很难获得, 这些

又加大了间歇过程监控的挑战性^[1-2]. 所以找到有效的间歇过程监控方法对间歇过程进行故障检测与监控具有重要意义. 数据不等长和多阶段操作是间歇过程的固有特征, 是实际生产过程中不容忽视的问题, 所以针对具有多阶段操作特性的间歇生产过程进行正确的操作阶段的划分, 对实现间歇过程监控具有很重要的作用, 是正确反映间歇过程多阶段操作特性的关键. 多元统计过程监控技术中的多向主元分析 (Multiple direction principal component analysis, MPCA) 和多向偏最小二乘 (Multiple direction partial least squares, MPLS) 方法已经被广泛运用于监控间歇过程^[3-4], 这两种方法均假设过程数据来自同一个操作阶段, 没有考虑间歇过程的多阶段操作特性, 张佳等^[5]提出的段 MPCA 法监测间歇过程的故障主要针对过程本身

收稿日期 2013-07-16 录用日期 2014-03-25
Manuscript received July 16, 2013; accepted March 25, 2014
国家自然科学基金 (61374140) 资助
Supported by National Natural Science Foundation of China (61374140)

本文责任编辑 刘德荣
Recommended by Associate Editor LIU De-Rong
1. 华东理工大学化工过程先进控制和优化技术教育部重点实验室 上海 200237

1. Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai 200237

的动态特性将间歇过程分成多阶段, 没有针对间歇过程的多阶段特性进行划分, 常玉清等^[6]提出的基于多时段 MPCA 模型的间歇过程监测方法研究, 通过三步法对间歇过程多阶段进行划分, 但对阶段的划分需要一定的先验知识. 近年来高斯混合模型 (Gaussian mixture model, GMM) 在运动图像处理领域得到了广泛的应用, 这种方法不需要先验知识, 能够通过对过程数据的学习和训练自动获得多个高斯模型的信息, 而且它已经被成功地引入到过程监控领域^[7-8], 为解决间歇过程的多阶段操作问题提供了很好的参考价值. Yu 等^[8]提出了一种基于 GMM 算法的划分, 而直接对大量历史数据进行 GMM 训练, 常常会被离群点、噪声等因素干扰, 很难得到精确的模型. 基于以上原因, 本文提出一种新的主元分析-多方向高斯混合模型 (Principal component analysis-multiple Gaussian mixture model, PCA-MGMM) 建模方法, 在消除数据预估的基础上, 用主元分析方法对数据进行降维, 得到得分向量, 在低维子空间中用 GMM 对得分向量进行训练, 通过期望最大化 (Expectation maximization, EM) 算法和 Figueiredo-Jain (F-J) 算法自动对多阶段进行划分, 得到较为精确的高斯模型, 最后将局部概率指标融合为全局概率监控指标, 实现连续的在线监控.

1 间歇过程数据处理

1.1 数据不等长问题的处理

间歇过程数据是一个三维数据矩阵: $X (I \times J \times K)$, 其中, I 为批次数, J 为过程变量个数, K 为采样时刻数, 在实际生产过程中, 由于各种各样的原因, 不可能达到完全的重复生产, 或是由于采样时刻持续时间不等, 使得间歇过程各个批次的长度不完全等长, 如图 1 所示.

数据不等长问题是间歇过程不可忽视的特点, 所以在建模之前对不等长数据进行等长处理是非常有必要的, 本文采用“最短长度法”对不等长数据进行等长处理, 具体做法^[9-10]: 在所有的间歇操作数据批次中寻找最短的一次操作数据, 然后以这批数据作为标准, 将其他批次的数据进行截取, 使所有批次的数据具有最短的数据长度, 即得到等长数据.

1.2 消除数据预估问题

根据建模的需要, 须将已处理的等长三维数据处理成二维数据. 首先按批次方向^[5, 9]将历史建模数据 $X (I \times J \times K)$, 展开成二维矩阵 $X (I \times JK)$,

然后对展开后的数据进行标准化处理:

$$\bar{x}_{j,k} = \frac{1}{I} \sum_{i=1}^I x_{i,j,k} \tag{1}$$

$$\delta_{j,k} = \sqrt{\frac{1}{I-1} \sum_{i=1}^I (x_{i,j,k} - \bar{x}_{j,k})^2} \tag{2}$$

$$\bar{x}_{i,j,k} = \frac{x_{i,j,k} - \bar{x}_{j,k}}{\delta_{j,k}} \tag{3}$$

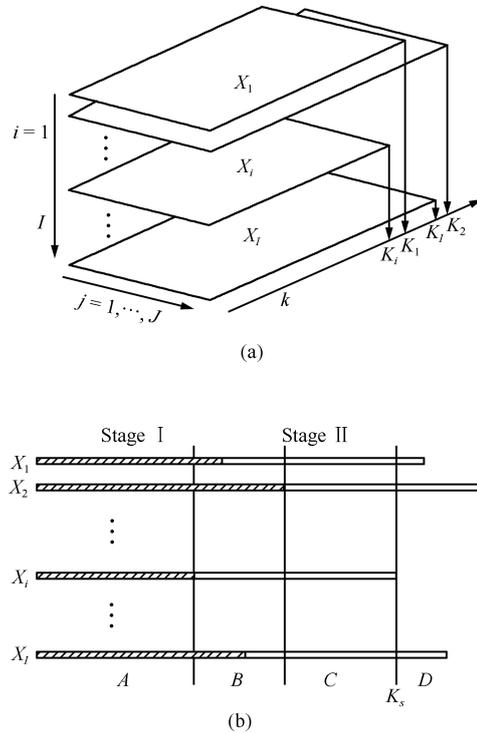


图 1 不等长间歇过程数据描述

Fig. 1 Description of the uneven data

处理完的数据实际上是抽取了间歇过程正常操作下过程变量的平均运行轨迹, 在一定程度上降低了变量轨迹中的非线性和动态特性对建模的影响, 并突出了间歇过程不同操作批次之间的变化信息. 接着将标准化后的数据重新排列成三维数据 $X (I \times J \times K)$, 然后按变量方向^[9]展开成二维矩阵 $X (J \times IK)$, 然后对展开后的数据进行标准化处理:

$$\bar{x}_{i,k} = \frac{1}{J} \sum_{j=1}^J x_{i,j,k} \tag{4}$$

$$\delta_{i,k} = \sqrt{\frac{1}{J-1} \sum_{j=1}^J (x_{i,j,k} - \bar{x}_{i,k})^2} \tag{5}$$

$$\bar{x}_{i,j,k} = \frac{x_{i,j,k} - \bar{x}_{i,k}}{\delta_{i,k}} \tag{6}$$

这种展开方法是针对每个变量进行标准化的, 利用的是每个变量在所有间歇操作的所有时间上的平均值和方差, 避免了在线应用时对未来测量值进行预估的问题, 而且考虑了变量在整个操作时间上的相关性和变化信息.

2 离线建模策略

2.1 主元分析算法

由于作为训练的历史数据量比较大, 而且直接对大量的历史数据进行建模, 往往会受到离群点和噪声的干扰, 使得到的模型不够精确, 本文利用主元分析 (Principal component analysis, PCA) 的方法^[11] 对数据进行处理, 将数据从高维空间投影到了对故障敏感的低维空间, 并得到了能够代表原历史数据信息的得分向量 T . 此方法将数据从原始空间映射到主元空间不会改变数据的分布特性, 而且能够捕捉变量的相关性, 利用 PCA 方法对标准化后的所有历史数据进行建模处理:

$$X^{J \times K} = t_1 p_1^T + t_2 p_2^T + \cdots + t_m p_m^T + E = TP^T + E \quad (7)$$

$$T = XP \quad (8)$$

其中, $P (J \times R)$ 为负载矩阵, $T (KI \times R)$ 为主元得分矩阵, $E (J \times IK)$ 为残差矩阵, R 为根据方差贡献率保留的主元数. 用此算法得到的得分向量进行建模大大减少了计算量, 同时也提高了模型的精确性.

2.2 基于 F-J 算法的高斯混合模型

对于间歇过程数据多阶段操作的特点, 可以假设来自每一个阶段的原始数据满足局部高斯分布^[8, 12], 用高斯混合模型建模方法对多阶段的间歇过程数据进行建模是很合适的.

传统的高斯混合模型算法在聚类时利用 EM 算法的不断迭代进行, 但混合模型的似然函数不是单模态的, 所以 EM 算法对初始值比较敏感, 而且不能自动确定模型的个数. 而本文提出的基于 Figueiredo 等提出的 F-J 算法^[13] 的 GMM 建模方法是一种无监督的聚类方法, 对初始值的敏感度较低, 它的优点是: 在没有先验知识的情况下, 根据每一阶段的原始数据满足局部高斯分布的条件, 采用最小信息长度 (Minimum message length, MML) 判据, 从任意数目的高斯成分开始训练, 通过 EM 算法的不断迭代逐步去掉权重为零的项, 自适应地调节高斯成分的数目, 并得到每个阶段所对应的高斯模型的均值和方差, 来确定最佳模型的个数.

在 PCA 算法得到的子空间中, 假设 $t_{\text{train}} \in \mathbf{R}^m$ 是来自间歇过程的 m 维历史数据主元, 它的概率密

度可用高斯混合模型^[8, 13-14] 表示为

$$p(t_{\text{train}}) = \sum_{k=1}^K p(t_{\text{train}}) p(t_{\text{train}}|k) = \sum_{k=1}^K \alpha_k g(t_{\text{train}}|\mu_k, \sum_k) \quad (9)$$

式中, K 是高斯成分的数目, α 是第 k 个高斯成分的权重, $\mu_1, \sum_1, \cdots, \mu_k, \sum_k$ 分别表示局部高斯模型的均值和方差.

通过期望最大化算法的不断迭代, 来自动确定后验概率和相应的分布参数. 采用任意初值 $\alpha^{(0)}, \mu^{(0)}, \sum^{(0)}$, 期望步 (E 步):

$$p^{(s)}(C_k|t_{\text{train}}) = \frac{\alpha_k^{(s)} g(t_{\text{train}}|\mu_k^{(s)}, \sum_k^{(s)})}{\sum_{k=1}^K \alpha_k^{(s)} g(t_{\text{train}}|\mu_k^{(s)}, \sum_k^{(s)})} \quad (10)$$

其中, $p^{(s)}(C_k|t_{\text{train}})$ 表示在第 s 步迭代中, 主元 t_{train} 属于第 k 个高斯成分的后验概率.

最大化步 (M 步):

$$\mu_k^{(s+1)} = \frac{\sum_{j=1}^n p^{(s)}(C_k|t_{\text{train}}) t_{\text{train}}}{\sum_{j=1}^n p^{(s)}(C_k|t_{\text{train}})} \quad (11)$$

$$\alpha_k^{(s+1)} = \frac{\sum_{j=1}^n p^{(s)}(C_k|t_{\text{train}})}{n} \quad (12)$$

$$\sum_k^{(s+1)} = \frac{\sum_{j=1}^n p^{(s)}(C_k|t_{\text{train}}) (t_{\text{train}} - \mu_k^{(s+1)}) (t_{\text{train}} - \mu_k^{(s+1)})^T}{\sum_{j=1}^n p^{(s)}(C_k|t_{\text{train}})} \quad (13)$$

式中, $\mu_k^{(s+1)}, \sum_k^{(s+1)}, \alpha_k^{(s+1)}$ 分别表示第 $s+1$ 步迭代中, 第 k 个高斯成分的均值、方差和先验概率.

3 全局监控策略

3.1 集成监控指标

集成监控指标不同于传统的单一模型的监控指标 T^2 和 SPE 控制图^[11, 15], 而是融合了全部模型的信息. 式 (9) 中的混合概率密度函数是多个局部高斯成分的加权和. 其中权重表示历史数据中各高斯模型在全局模型中所占的权重, 表示先验概率. 对于新样本来说, 只要根据推断出的模型参数计算出它

属于各个模型的后验概率,从而判断新样本所属模型.作为监控的全局指标^[7,16]:

$$P_{\text{global}} = \sum_{k=1}^K \beta(C_k | t_{\text{test}}) P_{\text{local}}^{(k)}(t_{\text{test}}) \quad (14)$$

式中, $\beta(C_k | t_{\text{test}})$ 是测试数据的得分向量属于第 k 个模型的后验概率, $P_{\text{local}}^{(k)}(t_{\text{test}})$ 表示对于第 k 个模型的局部概率指标.每个采样点的后验概率根据贝叶斯公式来计算:

$$\beta(C_k | t_{\text{test}}) = \frac{\alpha_k p(t_{\text{test}} | C_k)}{p(t_{\text{test}})} = \frac{\alpha_k p(t_{\text{test}} | C_k)}{\sum_{i=1}^K \alpha_i p(t_{\text{test}} | C_i)} \quad (15)$$

局部概率指标可以根据 t_{test} 到每个模型的马氏距离的概率来推出:

$$P_{\text{local}}^{(k)}(t_{\text{test}}) = \chi_m^2(D_{t_{\text{test}} \in C_k}) = \chi_m^2(t_{\text{test}}^T \sum t_{\text{test}}) \quad (16)$$

其中, m 表示低维子空间的维数,即得分向量 t_{test} 变量的个数.由式(16)可得,局部概率指标实际上将局部马氏距离通过 χ^2 分布转化为概率值,所以原本由马氏距离所反映的故障特征依然可以体现在监控图中.

3.2 具体操作流程

离线建模操作过程:

1) 采集间歇过程中的正常历史数据集的 87 批正常数据作为训练数据,并对其进行等长和展开处理;

2) 用 PCA 算法进行降维处理,通过式(7)得到训练数据的得分向量;

3) 在 PCA 子空间中用基于 F-J 算法的 GMM 对得分向量 t_{train} 进行训练,通过 EM 算法即式(10)~(13)的不断迭代,得到最优高斯成分的个数 K 以及模型参数 $\mu_1, \sum_1, \dots, \mu_k, \sum_k$;

4) 计算各高斯模型的先验概率 $\alpha_1, \dots, \alpha_k$.

在线监控操作过程:

5) 对新的采样点 x_{test} ,采用历史数据的均值和方差进行标准化,然后用 PCA 算法进行处理,得到新的得分向量 t_{test} ;

6) 利用式(15)和式(16)计算出对于每个高斯成分的后验概率 $\beta(C_k | t_{\text{test}})$ 和局部概率指标 $P_{\text{local}}(t_{\text{test}})$,判断所检测的样本属于哪个高斯成分;

7) 根据式(14)计算出新的采样点的全局监控指标 P_{global} ,如果 $P_{\text{global}} > 0.95$,则表明新的采样点为故障点.

本文监控策略的计算量主要集中在离线建模阶段,保证了模型的精确性,在线监控时均为简单的线性计算,对于一般的工业计算机来说完全可以保证在线监控的实时性.

3.3 维数的选择

在离线建模阶段,通过算法可以自动获得各阶段的先验概率以及各高斯成分的均值和方差,而算法中的参数设置集中在 PCA 降维时维数的保留,选择维数时要兼顾“维数之灾”和数据的信息损失.

本文采用 GMM 建模,降维是数据预处理的环节,与建立监控模型的降维算法不同,只要在降维的同时保留历史数据的聚类特征即可.本文利用主元贡献率累积和百分比的方法确定维数,前 k 个主元的贡献率为

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^m \lambda_i} \quad (17)$$

其中, k 为维数, m 为变量个数.最终确定维数为 7.

4 仿真实验

半导体生产制造越来越受重视,对于它的监控也逐渐成为产业界的一个热点问题.仿真所采用的数据是从一个实际的半导体制造过程采集而来,一共进行了 3 组实验,选取了 17 个过程变量对过程进行监测,共得到 129 批数据,其中包括 108 批正常数据和 21 批故障数据^[17].通过对数据的初步分析,有 2 批数据明显缺失,所以去掉其中 2 批数据,即共有 107 批正常数据,20 批故障数据.为了验证方法的有效性,随机选取 87 批正常数据作为训练样本,剩余的 20 批正常数据和 20 批故障数据作为测试样本,每一批数据含有 85 个时刻采样点,得到一个 3 维数据矩阵 X ($87 \times 17 \times 85$).

对所选择的训练样本经过预处理后由 PCA 算法在保留数据分布特性的基础上将数据从高维空间投影到对故障较敏感的低维空间,在所得的 PCA 子空间中对得分向量用 GMM 进行训练,通过 F-J 算法和 EM 算法的不断迭代自动得到最佳高斯成分个数 6,用时为 1.052283 s.聚类结果如图 2 所示.

用 20 批正常测试样本对所建立的模型进行检验,检验结果如图 3 所示.

图 3 中横线表示 99% 控制限,曲折线表示正常样本点的仿真曲线.由图 3 可看出,正常测试样本点均在控制限以下,没有出现误报现象,说明所建的 GMM 模型没有损失对正常批次数据监测的能力.

为了更好地说明本文所提方法的有效性,与传

统的 MPCA 方法进行比较, 表 1 给出了两种方法分别对 20 个故障的检测结果, 从表中数据可以看出, 本文所提的 PCA-MGMM 方法对多数故障的检测效果明显高于 MPCA 方法, 其中 MPCA 方法只对故障 8, 故障 16, 故障 19 的检测率好于本文的方法, 同时在表中用粗体标出 MPCA 方法检测率高于 PCA-MGMM 的故障. 图 4 和图 5 分别给出了两种方法对故障 11 和故障 20 的监控图, 其中用横线表示控制限.

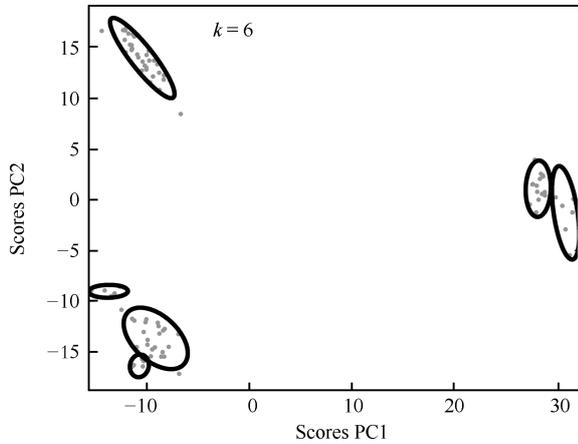


图 2 最佳模型个数

Fig. 2 Clustering results of the training data

表 1 两种方法对故障批次的监测结果 (%)

Table 1 Monitoring results of the two methods (%)

故障批次编号	MPCA (T^2)	MPCA (SPE)	MPCA-MGMM (BIP)
1	74.7	84.2	96.5
2	51.6	2.1	77.8
3	11.6	8.5	85.1
4	100	100	100
5	5.3	10.9	17.8
6	1.7	14.7	22.3
7	98.7	90.4	100
8	51.3	27.6	35.2
9	13.6	41.1	75.6
10	66.3	78.2	89.3
11	50.5	38.2	95.1
12	100	100	100
13	100	98.5	100
14	65.1	75.2	90.2
15	57.8	38.3	63.3
16	46.2	75.8	75.6
17	78.9	75.3	86.2
18	28.2	63.1	65.7
19	92.3	98.9	95.6
20	82.1	86.3	96.7

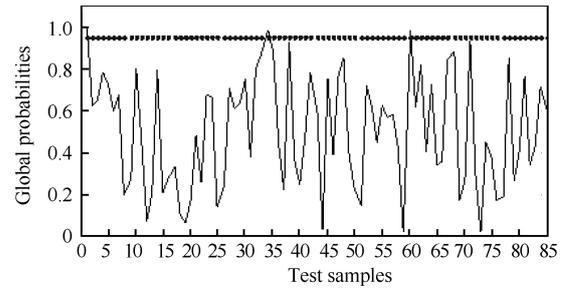
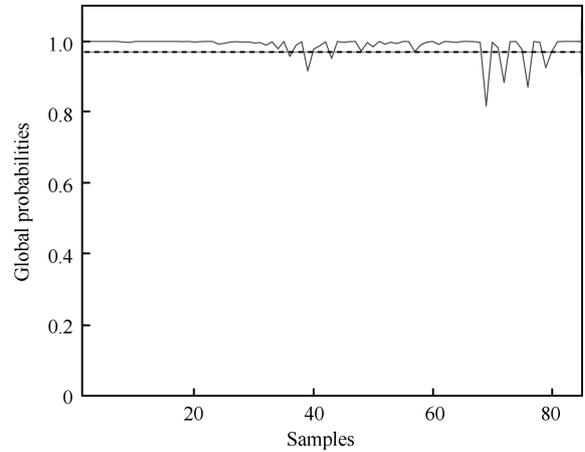
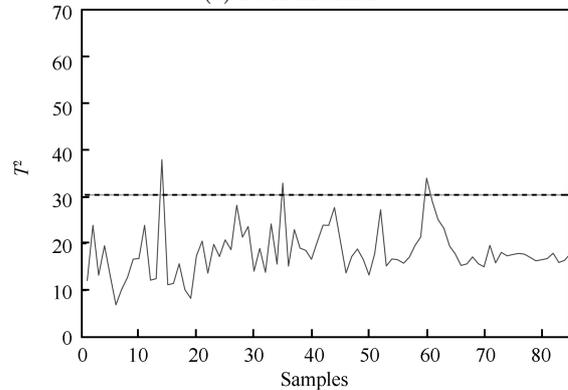


图 3 正常批次数据监测结果

Fig. 3 Monitoring results of the normal batch



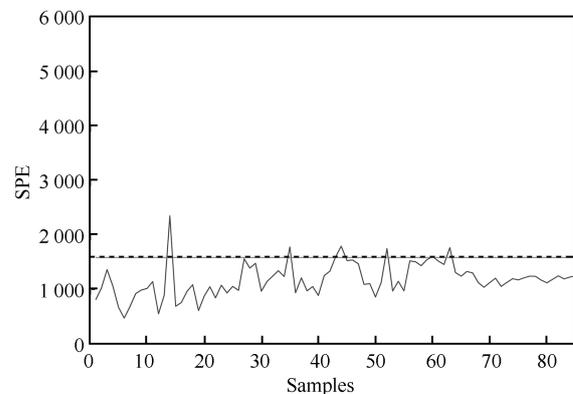
(a) PCA-MGMM



(b) MPCA

图 4 2 种方法对故障 11 的在线监控图

Fig. 4 Monitoring results of Fault 11



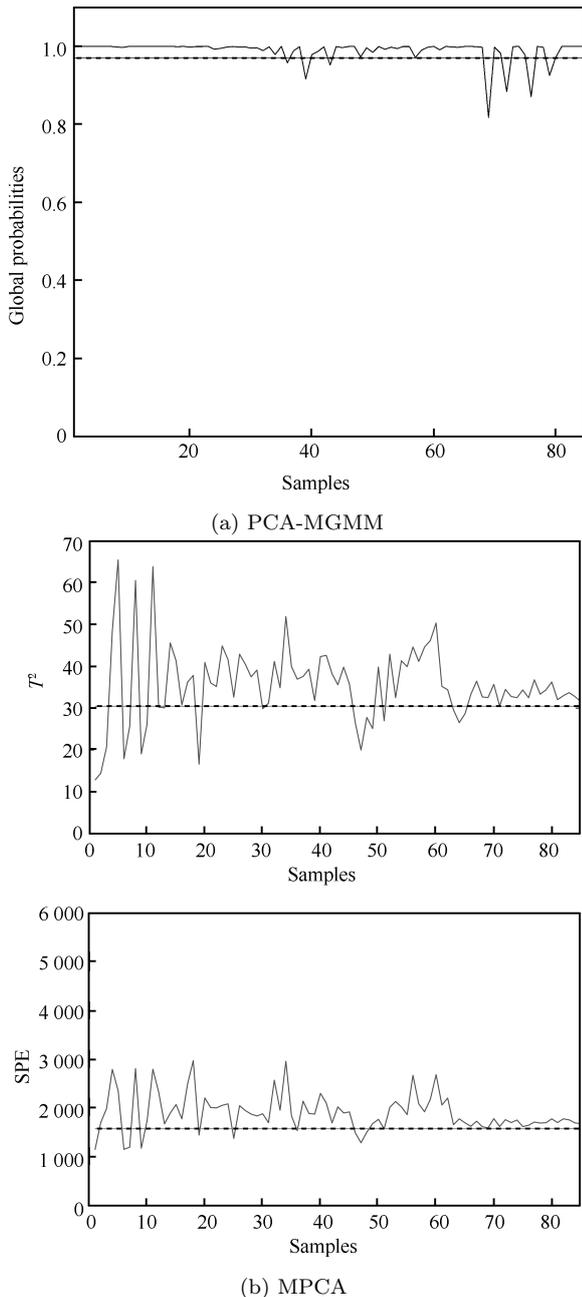


图 5 2 种方法对故障 20 的在线监控图

Fig. 5 Monitoring results of Fault 20

半导体的故障批次样本中, 所有采样时刻的采样点都是故障点, 由本文所提方法的仿真曲线可看出, 故障采样点在控制限以上的概率明显大于 MPCA 方法, 即本文的故障检测率大于 MPCA 方法, 具体体现在: 图 4 中 MPCA 方法的 SPE 和 T^2 两种统计量的结果均显示故障点在控制限以下的概率比较大, 漏报率较高, 而 PCA-MGMM 方法的漏报率相对较低. 图 5 中的两种方法均有一定的漏报率, 本文方法的优越性体现在及时性, 对故障更为敏感, PCA-GMM 方法在最初采样时刻就检测出了故

障点并及时地反映到图中, 而 MPCA 方法的统计量均有一定的延时, 不能及时地将故障点检测出. 同时新方法的监控指标实现了连续地在线监控, 能更好地给出故障报警信号, 所以本文所提的监控策略对于间歇过程产品质量的监控和控制具有更高的可靠性.

5 结论

本文主要针对间歇过程在建模过程中受离群点和噪声的干扰的问题, 提出一种新的建模方法和监控策略, 在消除了数据预估的同时通过 PCA 算法降维并得到主元向量, 接着用基于 F-J 算法和 EM 算法的 GMM 建模方法对主元向量进行训练. 此方法最大的特点是将多元统计方法与模式识别方法相结合应用到故障检测中, 这种相融合的方法不仅避免了对大量数据的处理, 降低了运算量和离群点的影响, 提高了所建模型的精确性, 而且所建模型数目是通过 F-J 算法和 EM 算法在不需要先验知识的情况下自动获得的, 即使缺少先验信息时也很适用. 所用的全局监控指标融合了局部信息的集成监控指标, 实现了连续的在线监控, 提高了故障的检测率, 而且本文所提的监控方法计算量主要集中在离线建模阶段, 保证了在线监控的实时性. 实际的半导体制造实验过程的监控结果更好地说明了新方法的有效性和可行性.

References

- 1 Cinar A, Parulekar S, Undey C, Birol G. *Batch Fermentation: Modeling, Monitoring, and Control*. New York: CRC Press, 2003.
- 2 Lee J M, Yoo C K, Lee I B. Enhanced process monitoring of fed-batch penicillin cultivation using time-varying and multivariate statistical analysis. *Journal of Biotechnology*, 2004, **110**(2): 119–136
- 3 Nomikos P, MacGregor J F. Monitoring batch processes using multi-way principal component analysis. *AIChE Journal*, 1994, **40**(8): 1361–1375
- 4 Kourti T, Nomikos P, MacGregor J F. Analysis, monitoring and fault diagnosis of batch processes using multi-block and multiway PLS. *Journal of Process Control*, 1995, **5**(4): 277–284
- 5 Zhang Jia, Sun Wei, Zhao Jin-Song, Sun Mei-Hong. Fault detection of batch process based on multi-phase multi-way principal component analysis. *Computers and Applied Chemistry*, 2010, **27**(3): 298–302
(张佳, 孙巍, 赵劲松, 孙美红. 多段 MPCA 法监测间歇过程的故障. *计算机与应用化学*, 2010, **27**(3): 298–302)
- 6 Chang Yu-Qing, Wang Zhu, Tan Shuai, Wang Fu-Li, Yang Jie. Research on multistage-based MPCA modeling and monitoring method for batch processes. *Acta Automatica*

- Sinica*, 2010, **36**(9): 1312–1320
(常玉清, 王姝, 谭帅, 王福利, 杨洁. 基于多时段 MPCA 模型的间歇过程监测方法研究. *自动化学报*, 2010, **36**(9): 1312–1320)
- 7 Yu J, Qin S J. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE Journal*, 2008, **54**(7): 1811–1829
- 8 Yu J, Qin S J. Multiway Gaussian mixture model based multiphase batch process monitoring. *Industrial and Engineering Chemistry Research*, 2009, **48**(18): 8585–8594
- 9 Zhao Chun-Hui, Wang Fu-Li, Yao Yuan, Gao Fu-Rong. Phase-based statistical modeling, online monitoring and quality prediction for batch processes. *Acta Automatica Sinica*, 2010, **36**(3): 366–374
(赵春晖, 王福利, 姚远, 高福荣. 基于时段的间歇过程统计建模、在线监测及质量预报. *自动化学报*, 2010, **36**(3): 366–374)
- 10 Rothwell S G, Martin E B, Morris A J. Comparison of methods for dealing with uneven length batches. In: *Proceedings of the 1998 International Conference on Computer Application in Biotechnology*. Osaka, Japan: Elsevier, 1998. 387–392
- 11 Zhou Dong-Hua, Li Gang, Li Yuan. *Fault Diagnosis Technology of Data-driven in Industrial Process*. Beijing: Science Press, 2011. 39–41
(周东华, 李钢, 李元. 数据驱动的工业过程故障诊断技术. 北京: 科学出版社, 2011. 39–41)
- 12 He Ming, Feng Bo-Qin, Ma Zhao-Feng, Fu Xiang-Hua. A unsupervised rough clustering method based on Gaussian mixture model. *Journal of Harbin Institute of Technology*, 2006, **38**(2): 256–259, 322
(何明, 冯博琴, 马兆丰, 傅向华. 一种基于高斯混合模型的无监督粗糙聚类方法. *哈尔滨工业大学学报*, 2006, **38**(2): 256–259, 322)
- 13 Figueiredo M A T, Jain A K. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(3): 381–396
- 14 Choi S W, Park J H, Lee I B. Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. *Computers and Chemical Engineering*, 2004, **28**(8): 1377–1387
- 15 Chiang L H, Russell E, Braatz R D. *Fault Detection and Diagnosis in Industrial Systems*. New York: Springer Verlag, 2001. 36–37
- 16 Xie X, Shi H B. Dynamic multimode process modeling and monitoring using adaptive Gaussian mixture models. *Industrial and Engineering Chemistry Research*, 2012, **51**(15): 5497–5505
- 17 Barry M W, Neal B G, Stephanie W B, Daniel D W J, Gabriel G B. A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. *Journal of Chemometrics*, 1999, **13**(3–4): 379–396



王 静 华东理工大学化工过程先进控制和优化技术教育部重点实验室硕士研究生. 主要研究方向为间歇过程故障检测与诊断.

E-mail: rain12027@163.com

(**WANG Jing** Master student at the Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology. Her research interest covers fault detection and diagnosis of batch processes.)



胡 益 华东理工大学化工过程先进控制和优化技术教育部重点实验室博士研究生. 2008 年获得华东理工大学信息学院学士学位. 主要研究方向为基于核偏最小二乘法的过程监控技术.

E-mail: hu_y8612@126.com

(**HU Yi** Ph. D. candidate at the Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology. He received his bachelor degree from East China University of Science and Technology in 2008. His research interest covers process monitoring based on kernel partial least squares.)



侍洪波 华东理工大学化工过程先进控制和优化技术教育部重点实验室教授. 2000 年获得华东理工大学控制理论与控制工程专业博士学位. 主要研究方向为工业过程建模, 控制与优化, 化工过程故障检测与诊断. 本文通信作者.

E-mail: hbshi@ecust.edu.cn

(**SHI Hong-Bo** Professor at the Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology. He received his Ph. D. degree from East China University of Science and Technology in 2000. His research interest covers modeling, control, optimization, fault detection and diagnosis of industrial processes. Corresponding author of this paper.)