

# 基于核主元分析-主元分析的多阶段间歇过程故障监测与诊断

齐咏生<sup>1</sup>, 王 普<sup>2</sup>, 高学金<sup>2</sup>

(1. 内蒙古工业大学 电力学院, 内蒙古 呼和浩特 010051; 2. 北京工业大学 电子信息与控制工程学院, 北京 100124)

**摘要:** 具有过渡特性的多阶段间歇过程故障监测是一个复杂的问题, 既需要考虑稳定阶段下的故障监测, 也需要考虑不同阶段间的过渡故障监测. 为克服传统硬划分方法导致误警和漏报率高的缺陷, 同时也为实现更加精确、有效的故障监测与诊断, 提出一套完整的基于核主元分析-主元分析(KPCA-PCA)的多阶段间歇过程故障监测与诊断策略. 该方法依据数据相似度实现阶段划分, 定义模糊隶属度辨识相邻阶段间的过渡, 最后对稳定阶段和过渡过程分别建立具有时变协方差的PCA和KPCA故障监测与诊断模型. 通过对青霉素发酵过程的仿真平台及工业应用研究表明, 该方法具有更可靠的监控性能, 能及时、准确的检测出过程中存在的异常情况.

**关键词:** 主元分析; 核主元分析; 故障诊断; 间歇过程

**中图分类号:** TP277      **文献标识码:** A

## Fault detection and diagnosis of multiphase batch process based on kernel principal component analysis-principal component analysis

QI Yong-sheng<sup>1</sup>, WANG Pu<sup>2</sup>, GAO Xue-jin<sup>2</sup>

(1. College of Electric Power, Inner Mongolia University of Technology, Huhhot Inner Mongolia 010051, China;

2. School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China)

**Abstract:** Fault detection in multiple phase processes is a complicated problem, because it is needed in both the steady phase and the transition from phase to phase. To overcome the hard-partition and misclassification problems, and also to monitor batch processes more accurately and efficiently, we propose a novel strategy for fault monitoring and diagnosing in batch processes based on the kernel principal component analysis-principal component analysis (KPCA-PCA). In this work, a phase division algorithm is designed based on the similarity index between different time-slice data matrices of batch processes, following by a fuzzy membership grade transition identification step. The steady phase ranges and the transition ranges are then modeled by PCA with time-varying covariance structures and KPCA separately. Results of simulation study and industrial application to penicillin fermentation process clearly demonstrate the effectiveness and feasibility of the proposed method, which detects various faults more promptly with desirable reliability.

**Key words:** principal component analysis; kernel principal component analysis; fault diagnosis; batch process

### 1 引言(Introduction)

多阶段是大多数间歇过程的一个固有特征, 过程的每个阶段都有不同的过程特征及过程主导变量, 而且过程变量相关关系并非跟随时间时刻变化, 而是跟随过程操作进程或过程机理特性变化呈现分阶段性<sup>[1]</sup>. 如发酵过程按细菌生长周期可划分为停滞期、指数生长期、静止期等阶段; 注塑过程可划分为注射、保压和冷却等操作阶段. 目前, 传统多主元分析(MPCA)方法虽已在间歇过程中取得了一些成功应用<sup>[2-4]</sup>, 但将一个完整批次数据作为一个统计样本来建立MPCA监控模型, 忽视了间歇过程中的局部行为特征, 很难揭示过程变量相关关系的变化, 例如采用一个线性模型对多阶段间歇过程进行

描述, 或者模型只能准确描述出某一或某几个阶段的特性, 导致在其余阶段下出现大量误报警; 或者模型涵盖所有阶段的操作范围, 控制限宽松, 导致在某些阶段下出现故障也不能及时报警, 出现大量漏报警情况. 此外由于MPCA本质上是一种线性化建模方法, 因此对于非线性较强的多阶段间歇过程, 该方法表现出较差的有效性. 针对过程的多阶段特性, Jie等提出基于混合高斯模型划分阶段的方法<sup>[5]</sup>; Camacho等提出采用线性局部模型逼近过程行为特征的方法来实现阶段划分的目的<sup>[6]</sup>; Lu等提出间歇过程子时段划分算法<sup>[7-8]</sup>, 以及基于子时段PCA(sub-PCA)模型的故障监测方法.

然而, 上述各类方法均属于硬分类方法, 在每个

子阶段将数据看成一个完整的对象处理,不能很好的反映过渡阶段特性的变化,从而造成相邻阶段的过渡过程特性变化对监控结果产生很大影响。这是因为相比各稳定阶段的主要运行模态,模态间的过渡虽不代表主流的过程操作机理特性,但却是一种普遍现象且是一种重要的过程行为。这种过渡模态表现为一种渐变趋势,不仅体现在过程变量的变化上,更体现在过程变量相关关系变化上。从总的趋势来看,在过渡开始初期,各时间点的过程特性与前一阶段的运行特性比较接近;随着过渡的不断进行,在过渡后期,过程特性慢慢过渡到后一阶段的运行模态。但应指出,过渡中的过程潜在特性并不总是处于严格递变中,且批次间相应的过渡过程常表现为时间轴上的不同步。鉴于过渡区域与稳定阶段具有显著不同的运行模态和潜在特性,有必要将过渡区域分离出来,单独建模分析其过程特性的发展变化。

此外,由于过程变量相关关系呈现分阶段性使得过程的非线性特性往往与过程阶段密切相关,采用分阶段PCA实质是非线性过程的局部线性化,因此在一定程度上能较好的解决过程的非线性问题。然而,对于阶段之间的过渡过程,由于各阶段轨迹不同步以及过渡过程的存在使得过渡阶段往往具有更强的非线性,因此,为建立更精确、有效的过程监控模型,过渡过程的非线性必需加以考虑。

针对以上问题,提出一套完整的间歇过程故障监控与诊断策略。对每个稳定阶段建立不同的统计模型,使每个稳定模型可以涵盖该阶段运行特性。对于稳定阶段间的过渡,考虑到随时间推移的过渡过程特性,通过提取变量间的非线性信息,建立过渡子模型反映过渡阶段变量相关关系的变化。

## 2 多元统计方法(Multivariate statistical method)

### 2.1 数据集的相似性度量(Similarity analysis of the process data)

假设具有相同列数的任意两个数据矩阵 $X_1 \in \mathbb{R}^{K_1 \times J}$ 和 $X_2 \in \mathbb{R}^{K_2 \times J}$ ,其中: $X_i$ 是经过标准化处理的数据, $J$ 为变量数, $K_i$ 为采样点数。则将上述两个数据阵混合后获得的协方差矩阵 $M$ 可表示如下:

$$M = \frac{1}{K_1 + K_2 - 1} [X_1^T X_1 + X_2^T X_2] = \frac{K_1 - 1}{K_1 + K_2 - 1} M_1 + \frac{K_2 - 1}{K_1 + K_2 - 1} M_2, \quad (1)$$

其中 $M_i$ 代表 $X_i$ 的协方差矩阵,对 $M$ 进行特征值分解,得到正交矩阵 $P_0$ 满足

$$MP_0 = P_0 \Lambda, \quad (2)$$

定义转换矩阵 $P = P_0 \Lambda^{-\frac{1}{2}}$ ,可知,矩阵 $P$ 满足 $P^T M P = I$ ,利用 $P$ 对数据集 $X_i$ 进行如下变换可

得到新矩阵 $Y_i$ :

$$Y_i = \sqrt{\frac{K_i - 1}{K_1 + K_2 - 1}} X_i P = \sqrt{\frac{K_i - 1}{K_1 + K_2 - 1}} X_i P_0 \Lambda^{-\frac{1}{2}}, \quad (3)$$

矩阵 $Y_i$ 的协方差阵进一步可表示为

$$S_i = \frac{1}{K_i - 1} Y_i^T Y_i = \frac{K_i - 1}{K_1 + K_2 - 1} P^T M_i P. \quad (4)$$

由式(1)–(4)可知

$$S_1 + S_2 = P^T M P = I. \quad (5)$$

对 $S_i$ 进行特征值分解,并利用式(5)可得到

$$S_2 v_1^j = (I - S_1) v_1^j = (1 - \lambda_1^j) v_1^j, \quad (6)$$

$v_1^j$ 是 $S_1$ 进行特征值分解后得到的第 $j$ 个特征向量,即 $S_1$ 与 $S_2$ 具有相同的特征向量,且其对应的特征值满足如下关系:

$$\lambda_1^j = 1 - \lambda_2^j. \quad (7)$$

由式(6)和式(7)可知, $Y_1$ 和 $Y_2$ 的主元方向是一致的,只是主元所解释变化信息的大小顺序刚好相反。依据对称性,如果 $X_1$ 和 $X_2$ 足够相似,应有 $\lambda_1^j$ 接近0.5;相反如果 $X_1$ 和 $X_2$ 差异很大,则有 $\lambda_1^j$ 接近1或0。据此,可定义两个数据集相似性指标 $D$ 为<sup>[9]</sup>

$$D = \text{diss}(X_1, X_2) = \frac{4}{J} \sum_{j=1}^J (\lambda_1^j - 0.5)^2 = \frac{4}{J} \sum_{j=1}^J (\lambda_2^j - 0.5)^2. \quad (8)$$

可以看到 $D$ 取值范围为 $0 \leq D \leq 1$ 。

### 2.2 具有时变协方差的改进MPCA方法

(Improved MPCA with time-varying covariance structures)

传统MPCA方法是将取自成功历史批次的三维矩阵 $X(I \times J \times K)$ ( $I, J, K$ 分别表示批次数、变量数和采样点数)展开成二维矩阵,之后应用PCA方法建立过程监控模型。目前,可行的展开方式主要有两种:一是基于批次方向展开的NM-MPCA方法,该方法优势是能抽取过程变量平均运行轨迹,一定程度上可消除变量沿时间方向上的非线性,突出不同批次间的变化信息;其主要缺陷为在线应用时,需对当前采样时刻到操作周期结束时刻的变量测量值进行估计,由此引入的估计误差会导致较高的误警率,降低算法的可靠性。另一种是基于变量方向展开的W-MPCA方法,该方法不需要估计未来测量值,不存在估计误差,但由于忽略了不同时刻变量的相关性,且不能突出批次间的方差信息,因此对故障不敏感,故障检测的快速性和灵敏性差<sup>[4]</sup>。为克服以上缺陷,本文采用一种将上述两种展开方式相结合的改进方

法<sup>[2]</sup>. 即首先对三维矩阵沿批次方向展开进行数据标准化, 一定程度上消除过程变量沿时间方向的非线性; 之后将数据矩阵重新沿变量方向排列为二维矩阵 $X(KI \times J)$ , 最后对新的二维矩阵 $X$ 建立PCA模型. 在建立监控模型后, 需计算监控统计量来判定过程有无故障发生. 在本文方法中, 为充分考虑主元得分向量的动态特性, 提出采用时变协方差计算 $T^2$ 统计量. 即将获得的主元矩阵 $T(KI \times R)$ 依据采样时刻划分为 $K$ 个数据子块 $T_k(I \times R)$ ,  $k = 1, \dots, K$ . 利用 $T_k$ 计算 $T^2$ 统计量:

$$T^2 = t_{\text{new},k} A_k^{-1} t_{\text{new},k}^T \sim \frac{R(I^2 - 1)}{I(I - R)} F_{R, I-R, \alpha}, \quad (9)$$

其中:  $A_k$ 为对应 $k$ 时刻数据子块 $T_k$ 的协方差矩阵,  $t_{\text{new},k}$ 表示新批次第 $k$ 时刻的主元得分向量. 这里, 采用时变的 $A_k$ 代替固定的 $A$ ( $A$ 为主元矩阵 $T$ 的协方差阵), 由此获得 $T^2$ 统计量可充分考虑主元得分向量的动态特性, 一定程度上可提升故障检测灵敏度.

### 2.3 KPCA监控方法(KPCA monitoring)

传统的主元分析法是基于原始特征的一种线性变换, 当原始数据存在非线性属性时, 使用PCA分析后留下的显著成分就不能反映这种非线性属性. 而核主元分析(KPCA)通过非线性变换将输入空间样本映射到高维特征空间, 是一种能从数据样本中提取非线性特征的有效方法. 该算法的基础是去中心化核矩阵. 设输入样本集 $X_s = \{x_1, x_2, \dots, x_N\}$ , 其中 $x_k \in \mathbb{R}^m$ ,  $N$ 为样本数. 通过非线性映射 $\phi$ 将输入数据从原空间映射到高维特征空间 $F$ , 记为 $\phi(x_k)$ . 定义核矩阵 $K = [K_{ij}]$ ,

$$K_{ij} = k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle, \quad (10)$$

可实现特征空间中两个向量的内积运算, 避免了求解非线性映射 $\phi$ 的问题. 对于任意测试向量 $x$ , 它在特征空间中的主元得分 $t$ 可按下式计算:

$$t_k(x) = \sum_{i=1}^N \alpha_i^k \langle \phi(x_i), \phi(x) \rangle = \sum_{i=1}^N \alpha_i^k k(x_i, x), \quad (11)$$

其中:  $k = 1, \dots, p$ ,  $p$ 为保留主元数,  $\alpha$ 为归一化后的核矩阵的特征向量. 基于KPCA的 $T^2$ 和SPE统计量及控制限计算方法, 可参见文献[10].

### 3 多阶段 KPCA-PCA 监控策略(KPCA-PCA multiphase monitoring strategy)

如图1所示, 为本文提出监控策略的原理示意图. 具体描述如下:

#### 3.1 阶段划分(Phase division)

1) 三维建模数据矩阵沿批次方向展开进行数据标准化, 并分割为 $K$ 个时间片数据子块 $X_i(I \times J)$ ,  $i = 1, 2, \dots, K$ .

2) 对每个 $X_i$ 依据下式计算相似性指标:

$$D_i(k) = \text{diss}(X_i, X_j), \quad (12)$$

其中  $\begin{cases} k = j, & j = 1, 2, \dots, i-1, \\ k = j-1, & j = i+1, \dots, K. \end{cases}$  将 $D_i$ 作为聚

类的输入样本. 需要强调一点, 如果将采样数据直接进行聚类, 由于过渡过程存在、阶段轨迹不同步及一些批次存在不连续跳变点(扰动点), 使得阶段划分往往无法获得令人满意的结果, 甚至不能划分阶段. 而本文方法依据时间片数据块的相似性指标进行聚类, 能获得更为合理、有效的分类结果, 且对一些异常点扰动具有更好的鲁棒性.

3) 采用模糊 $C$ 均值聚类(FCM)进行阶段初步划分. 依据最大隶属度原则, 将过程首先分割为 $c$ 个阶段, 之后在每个阶段采用单变量控制图检测最大隶属度值的离群点, 由于连续离群点主要发生在阶段开始与结束时刻, 因此可据此辨识过渡阶段的开始与结束时刻. 最后将移除过渡过程的阶段确定为相应的稳定阶段. 这里, 之所以能够采用单变量控制图实现对过渡阶段的辨识, 是因为过程变量相关关系并非随时间时刻变化, 而是跟随过程操作进程或机理特性而变化呈现分时段性. 在每个稳定阶段, 时间片数据矩阵之间表现出很高的相似度; 而在稳定阶段与相邻过渡阶段间的数据相似度则表现出显著的不同. 因此, 每个阶段开头与结尾的采样点能够很容易被作为连续离群点而检测到.

4) 采用迭代法确定单变量控制图的控制限. 每次迭代中, 对当前数据采用传统单变量控制图计算方法确定控制限, 之后移除超过控制限的离群点并更新数据, 重复上述步骤直到控制限收敛. 由于控制限的确定是基于统计方法而不是由主观经验决定, 因此过渡过程的辨识更加客观、合理.

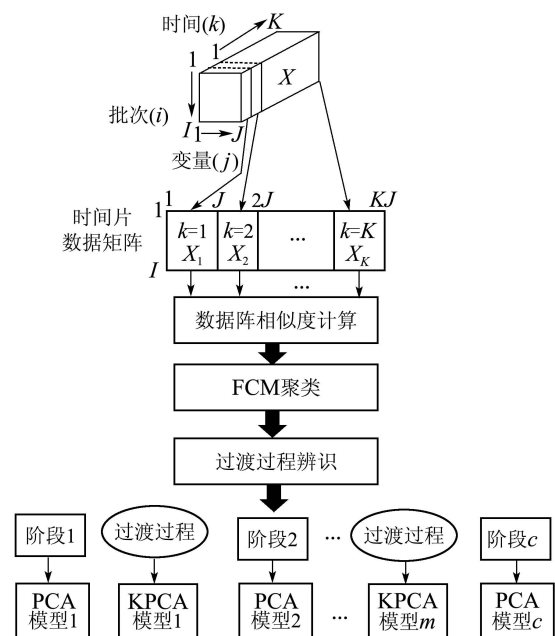


图1 KPCA-PCA算法原理示意图

Fig. 1 Illustration of KPCA-PCA algorithm

### 3.2 阶段建模(Phase modeling)

#### 3.2.1 稳定阶段建模(Steady phase modeling)

在划分好稳定和过渡过程后, 对各阶段建立相应的监控模型, 具体步骤如下:

1) 将每个稳定阶段三维矩阵 $X_c(I \times J \times k_c)$ 沿变量方向展开为二维矩阵 $X_c(Ik_c \times J)$ , 其中 $k_c$ 为第 $c$ 个稳定阶段所包含的采样点个数;

2) 对 $X_c(Ik_c \times J)$ 建立PCA模型, 采用交叉检验法确定保留主元个数. 将主元矩阵 $T_c$ 划分为 $k_c$ 个子块 $T_i^c$ , 计算相应的时变协方差矩阵 $\Lambda_i^c(i = 1, \dots, k_c)$ ;

3) 计算模型每个时刻 $T^2$ 、SPE统计量及控制限.

#### 3.2.2 过渡过程建模(Transition modeling)

1) 对各过渡阶段的三维数据 $X_m(I \times J \times k_m)$ , 按变量方向排列成二维数据矩阵 $X_m(Ik_m \times J)$ , 其中 $k_m$ 代表过渡阶段 $m$ 包含的采样点个数;

2) 对输入样本 $X_m$ 建立KPCA模型; 然而, 实际计算中发现, 由于建立KPCA模型时需要计算和存储核矩阵并对核矩阵进行特征值分解, 因此当建模样本数据过大时, 计算将十分耗时且需要大容量内存. 而本文算法中批次数据沿变量方向展开, 当过渡过程比较长即样本容量 $Ik_m$ 很大时, 也存在上述问题. 为此本文算法中引入特征采样的方法来解决该问题<sup>[11]</sup>. 考虑如下定理: 分割 $(n \times n)$ 核矩阵 $K_n$ 为

$$K_n = \begin{bmatrix} K_{n-1} & K_{n-1,n} \\ K_{n-1,n}^T & k_{n,n} \end{bmatrix}, \quad (13)$$

其中:  $K_{n-1} = \{k(x_i, x_j)\}$ ,  $i, j = 1, \dots, n-1$ ,  $k_{n,n} = k(x_n, x_n)$ ,  $K_{n-1,n} = \{k(x_i, x_n)\}$ ,  $i = 1, \dots, n-1$ . 假定 $K_{n-1}$ 满秩, 如果满足 $\delta = k_{n,n} - K_{n-1,n}^T K_{n-1}^{-1} K_{n-1,n} = 0$ , 则 $K_n$ 降秩为 $n-1$ . 实际过程中, 考虑到噪声的影响, 设定一个很小的阈值 $\varepsilon > 0$ , 当 $\delta \leq \varepsilon$ 时, 可认为 $K_n$ 和 $K_{n-1}$ 具有近似相同的秩. 由此可得到一种实现特征采样的方法, 总结如下:

a) 假设初始样本基包含任意一个样本, 即 $d = 1$ , 计算相应的核矩阵 $K_d$ ;

b) 逐个检验样本, 计算 $\delta_{d+1}$ , 如果 $\delta_{d+1} \leq \varepsilon$ , 则该样本不加入样本基 $FS$ ; 否则令 $d = d+1$ , 将该样本加入样本基 $FS$ , 并修改相应的核矩阵和核矩阵的逆;

c) 对所有样本检验后得到样本基 $FS = \{fs_1, fs_2, \dots, fs_d\}$ 和 $d$ . 其中 $\varepsilon$ 的选取很重要, 太大,  $K(FS, FS)$ 不能全面反映样本在特征空间中的信息; 太小, 则导致过大的 $d$ 值或使 $K_d$ 接近奇异. 由算法可知, 显然 $d$ 与 $\varepsilon$ 成负相关关系. 因此可根据 $\varepsilon$ - $d$ 曲线中的转折点确定 $\varepsilon$ 取值.

3) 采用累计方差贡献率法确定KPCA的主元个数, 计算模型每个时刻的 $T^2$ 、SPE统计量及控制限.

### 3.3 故障的在线监测与诊断(Online monitoring and fault diagnosis)

#### 3.3.1 在线监测(Online monitoring)

1) 新批次在线监测时, 对当前时刻采样数据 $x_{new,k}$ 采用相应时刻建模数据进行标准化;

2) 根据当前采样时刻进行模型切换, 依据所选模型计算当前时刻的 $T^2$ 和SPE统计量;

3) 检查 $T^2$ 、SPE统计量是否超出各自控制限. 若统计量出现超限现象, 则表明过程中可能出现故障, 此时可结合 $T^2$ 和SPE贡献图分析故障产生原因, 实现故障的进一步隔离和诊断.

#### 3.3.2 故障诊断(Fault diagnosis)

1) 稳定阶段的故障诊断. 由于稳定阶段建立的是沿变量方向展开的MPCA模型, 因此可直接应用随时间变化的贡献图<sup>[2]</sup>实现故障的进一步诊断.

2) 过渡过程的故障诊断. 对于过渡阶段建立的KPCA模型, 由于无法找到一个从高维的特征空间到低维输入空间的逆映射, 因此传统的贡献图方法无法应用. 为此, Cho等曾提出一种基于核函数梯度的贡献图方法用于连续过程的故障诊断<sup>[12]</sup>. 本文对该方法进行了进一步扩展研究, 将该方法推广应用到间歇过程中, 实现过渡过程的故障隔离与诊断. 另外, 在实际应用中发现, 由于核函数梯度贡献图法需要对核矩阵中每一项进行求导运算, 需要 $O(N^3)$ 次运算( $N$ 为样本数量), 导致运算量过大, 非常耗时. 为此本文对该方法提出了改进措施, 即采用经过特征采样后的核矩阵代替原始数据形成的核矩阵作为求解贡献图梯度的核矩阵, 由于 $d \ll N$ , 因此极大地减少了运算量, 可实现实时的贡献图诊断. 此外, 通过仿真研究也表明采用特征采样的核矩阵计算贡献图对故障的隔离与诊断不会产生影响.

## 4 仿真实验(Simulation verification)

美国Illinois州立理工学院过程监控与技术小组开发了青霉素仿真平台PenSim2.0, 它为青霉素间歇过程监控和故障诊断研究提供了一个标准平台<sup>[13]</sup>. 青霉素发酵过程每个批次的反应时间为400h, 采样间隔为1h, 选择10个过程变量(通风率、搅拌功率、底物补料温度、溶氧浓度、发酵罐体积、二氧化碳浓度、pH值、发酵罐温度、反应产生热量、冷水流速)来监测过程运行. 为更符合实际情况, 所有测量变量均加入了测量噪声, 共产生了35个正常批次作为初始的模型参考数据库, 得到三维数据矩阵 $X(35 \times 10 \times 400)$ , 将三维数据矩阵沿批次方向进行预

处理,并切割为400个时间片数据阵 $X_i(35 \times 10)$ ,按式(12)计算相似度指标 $D_i$ 作为聚类输入样本.

为表明基于数据相似度阶段划分算法的优势,比较了基于原始数据直接聚类的阶段划分结果.图2和图3分别给出采用 $K$ -mean和FCM算法对每个时刻的采样数据直接进行聚类的结果.由图可知,分类效果很差,2,3阶段几乎无法实现有效划分.而图4为基于相似度指标 $D_i$ 进行聚类的结果.显然,分类效果很理想,且与发酵过程的机理特性基本吻合,表明在青霉素发酵过程中,过程变量相关关系并非随时间时刻变化,而是跟随过程机理特性的变化呈现分阶段性.图5中, $\text{Sim}(k, c)$ 代表第 $k$ 个时间片矩阵属于第 $c$ 个阶段的隶属度值.由图可知,随着过程机理特性的变化 $\text{Sim}(k, c)$ 将发生变化.在每个阶段的中段, $\text{Sim}(k, c)$  ( $k \in c$ )接近于1表明当前时刻的时间片数据阵与该阶段具有很高的相似度;而在阶段的开始和结尾时刻, $\text{Sim}(k, c)$  ( $k \in c$ )的值是逐步减小的渐变过程,表明相邻阶段之间存在一定的过渡.如图6所示,在阶段初步划分之后,依据单变量图实现对过渡过程的辨识.即在每个阶段计算该阶段时间片数据阵的不相似度 $1 - \text{Sim}(k, c)$ 作为单变量控制图输入,通过检测阶段开始与结束时刻的连续离群点,确定过渡过程区间.图7形象地展示了相邻阶段的过渡过程特征,图中:实线表示过渡阶段时间片矩阵属于第1阶段,点划表示过渡阶段时间片矩阵属于第2阶段,点线表示过渡阶段时间片矩阵属于第3阶段.采用上述方法最终确定各阶段的采样区间为:稳定阶段(1-46), (70-183), (218-400);过渡阶段(47-69), (184-217).最后,对各阶段建立相应监控模型.

为验证本文算法的有效性,针对不同过程变量及故障类型生成多个故障批次进行算法测试.各故障类型见表1.其中,每类故障均产生3个不同初始值、不同幅度的故障批次,并以均值作为该类故障的最终评价指标.在监测性能方面与NM-MPCA, sub-PCA<sup>[7]</sup>方法进行了比较.另外,在实验中使用主频2.83 G,内存4 G的DELL OptiPlex960;在本文算法中,离线建模阶段所需时间(包括阶段划分、过渡辨识、MPCA及KPCA建模时间之和)为450 s左右,在线应用时单个样本检测所需时间为1 s左右,因而完全能满足一般工业过程故障监测实时性要求.

表2给出了3种方法的监控性能比较,可以看出,本文方法对于各类故障的检测均是有效的,且在3种方法中误警率(I型误差率)最低,表明本文方法在一定程度上可提升监控系统的可靠性.对于故障批次,本文方法能够在较小的漏报警率(II型误差率)下,实现故障的快速、准确检测.而NM-MPCA和sub-PCA方法则给出较高的漏报率.

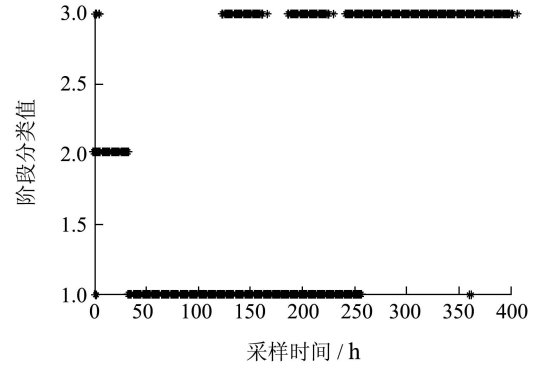


图2  $K$ 均值直接聚类结果

Fig. 2  $K$ -mean clustering result

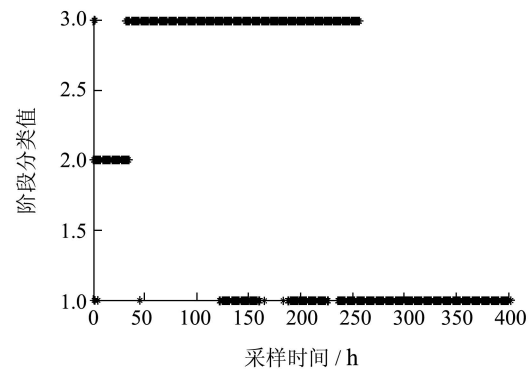


图3 FCM直接聚类结果

Fig. 3 FCM clustering result

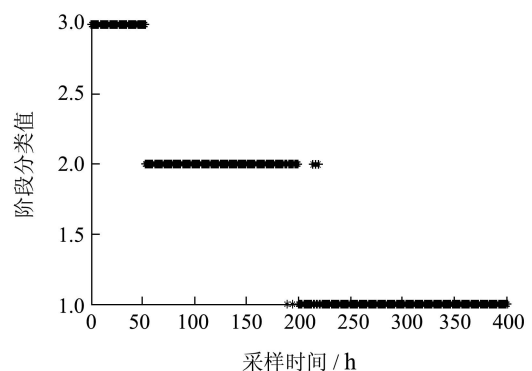


图4 本文算法聚类结果

Fig. 4 The proposed method clustering result

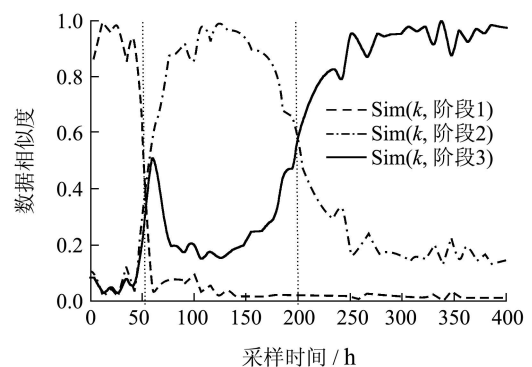


图5 过程相关性变化曲线

Fig. 5 Process nature changing with the similarity

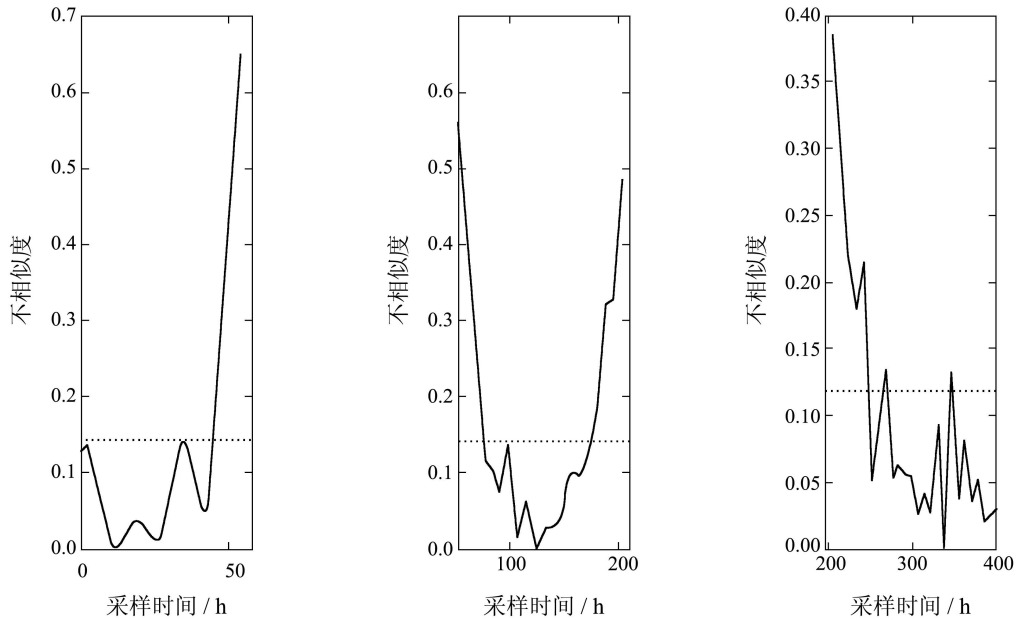


图 6 过渡过程辨识

Fig. 6 Transition ranges identification

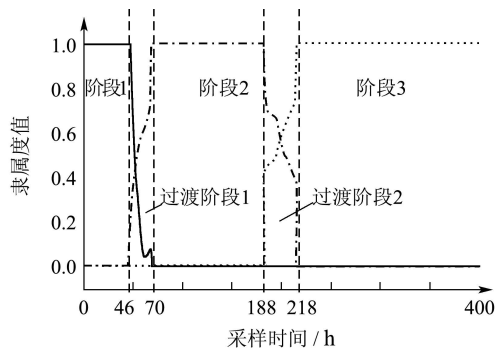


图 7 阶段划分示意图

Fig. 7 Sketch map of membership grades

表 1 仿真中用到的故障类型

Table 1 Fault types introduced in process

故障编号	过程变量	故障类型
故障1	底物补料速率	阶跃扰动
故障2	搅拌功率	阶跃扰动
故障3	通风率	阶跃扰动
故障4	底物补料速率	斜坡扰动
故障5	搅拌功率	斜坡扰动
故障6	通风率	斜坡扰动

表 2 比较3种方法的监控结果

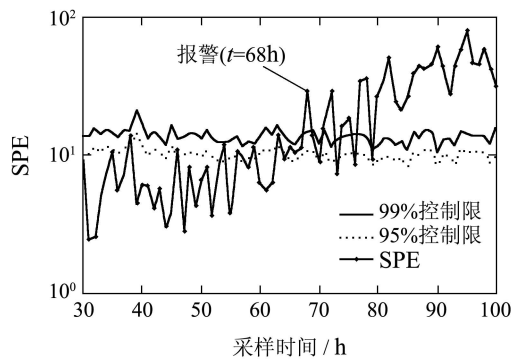
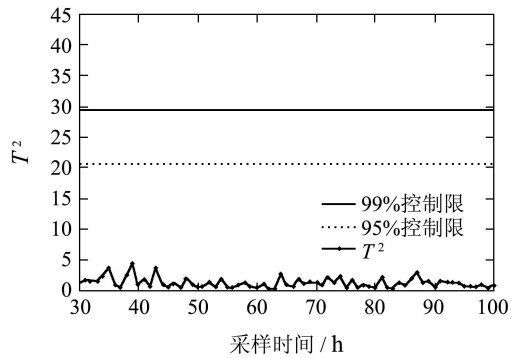
Table 2 Summary of monitoring results for three methods

工况	I 型误差率/%			II 型误差率/%		
	NM-MPCA	sub-PCA	KPCA-PCA	NM-MPCA	sub-PCA	KPCA-PCA
正常批次	5.81	2.52	1.87	—	—	—
故障1	1.54	0.79	0.67	10.46	45.08	6.6
故障2	5.71	1.43	1.67	6.97	0	0
故障3	7.71	2.8	4.1	12.33	0.9	0.78
故障4	3.84	1	1	22.4	40.7	12
故障5	3.13	1.36	1	7.9	39.9	2
故障6	3.25	0.75	0.75	51.25	44.2	4.5

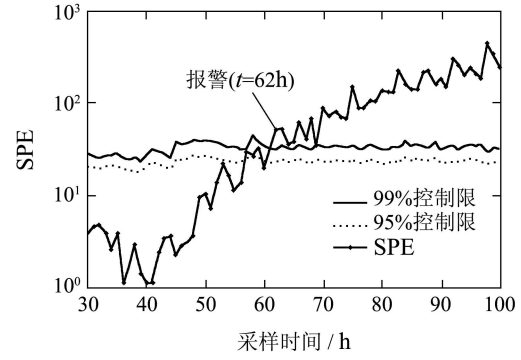
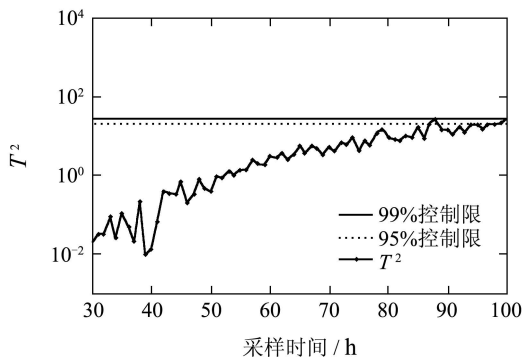
图8是对故障类型5的批次进行监测的结果. 该故障批次为搅拌功率在47h加入斜率为1.2%的斜坡扰动直到反应结束. 由图可知, 本文方法在48h几乎在故障发生的同时就指示了异常情况的发生, 比NM-MPCA方法和sub-PCA方法分别提前了大约20h和14h. 在 $T^2$ 监控图中, sub-PCA方法对

故障的检测比本文方法滞后了大约55h, 而NM-MPCA方法的 $T^2$ 图中未检测到任何异常发生. 分析可知, 本次故障恰好发生在过渡阶段1内, 由于sub-PCA方法严格将过程划分为不同的子阶段, 割裂了相邻阶段的联系, 不能很好的反映出过渡阶段的过程特性, 因此, 对故障的检测具有较大的滞

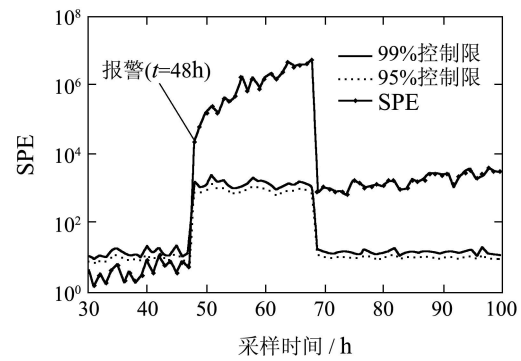
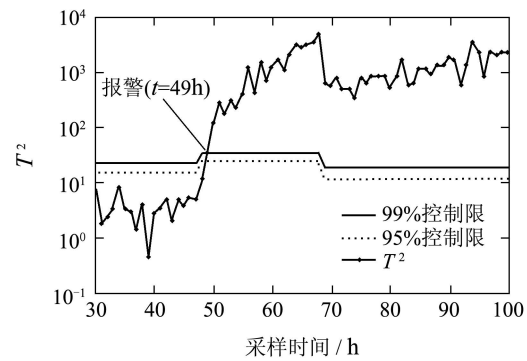
后;而NM-MPCA方法将整个批次数据作为一个整体来处理,或者不能准确描述出所有阶段的特性;或者涵盖所有阶段的操作范围,控制限宽松;最终导致在某些阶段下出现故障也不能及时报警,出现大量漏报情况.可见,过程过渡相关特性的变动对监测结果有较大影响,必须加以考虑.在检测到故障后需要对故障进行隔离与诊断,如图9为采用核函数梯度方法得到的过渡阶段1的统计量贡献图,  $T^2$ 和SPE贡献图给出了相同的诊断结果,即过程的故障是由于变量2——搅拌功率的异常所导致.本例中在计算核函数梯度时采用经过特征采样的核矩阵,如图10所示给出了阈值 $\epsilon$ 的确定过程,当 $10^{-1} \geq \epsilon \geq 10^{-5}$ 时,得到 $d = 55$ 及相应的样本基FS.通过测试发现,基于核函数的贡献图计算时间仅需15s左右,而不经特征提取的计算时间则需要大约18 min.



(a) NM-MPCA



(b) sub-PCA



(c) 本文方法

图8 采用传统MPCA、sub-PCA和本文方法监控测试批次的结果

Fig. 8 Monitoring results using MPCA, sub-PCA and the proposed method for test batch

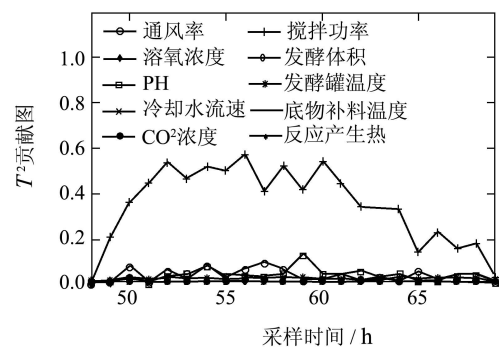


图9(a)  $T^2$ 统计量贡献图

Fig. 9(a) Contribution plots to  $T^2$

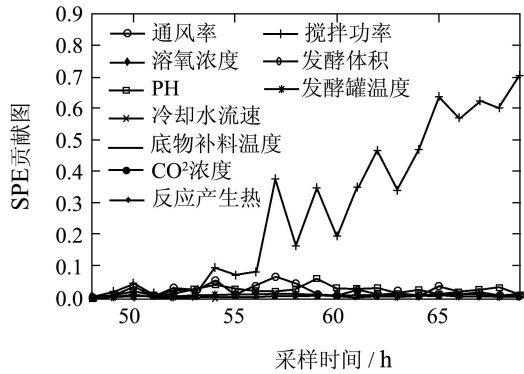


图 9(b) SPE统计量贡献图

Fig. 9(b) Contribution plots to SPE

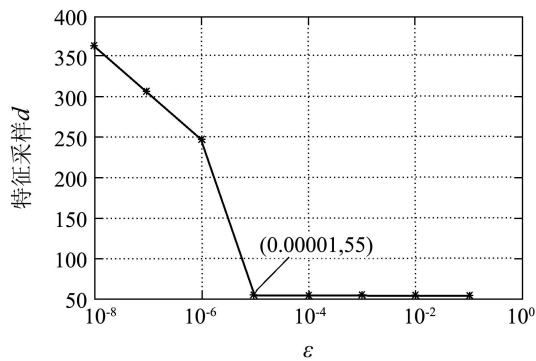


图 10  $\epsilon$ - $d$  曲线

Fig. 10  $\epsilon$ - $d$  curve

以上仿真结果表明基于KPCA-PCA的多阶段监控模型在准确度和鲁棒性方面均优于NM-MPCA和基于硬划分的sub-PCA方法。

### 5 工业应用研究(Industrial application study)

本节给出了本文方法在河北某制药公司青霉素发酵过程监测中的应用。该过程发酵周期为 212h, 采样间隔为 4h, 选择 9 个主要过程变量(体积、搅拌电流、温度、空气流量、pH 值、补糖量、补硫酸铵量、补氨量、补苯乙酸量)来综合表征青霉素发酵菌体生长和产物合成状况。选取 24 个正常批次作为模型参考数据库, 得到三维数据矩阵  $X(24 \times 9 \times 53)$ , 按本文方法划分阶段, 最终确定各阶段及过渡过程采样区间为:

稳定阶段(1-5), (10-20), (27-53);

过渡阶段(6-9), (21-26)。

图 11 给出了对正常生产批次 185 进行监测的结果。由图可知, sub-PCA 及 NM-MPCA 方法的 SPE 监控图误警率分别为 3.7% 和 9.4%, 显著偏离了控制限设定值 1%, 此外, 由于存在连续的误报警很可能误导操作员做出错误判断, 给正常生产带来不必要影响; 而本文方法误警率低于 1%, 实现对

185批次的正确监测。

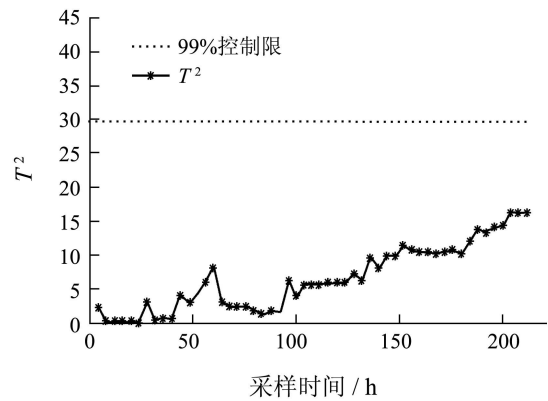
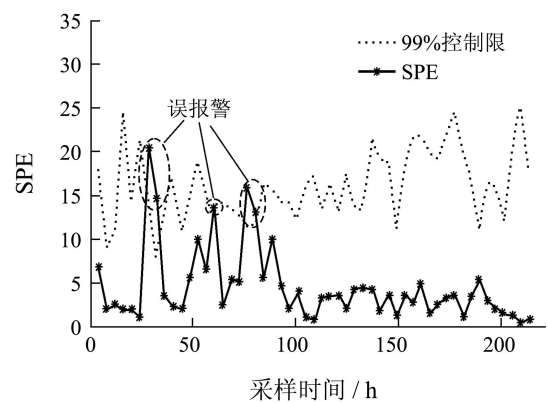
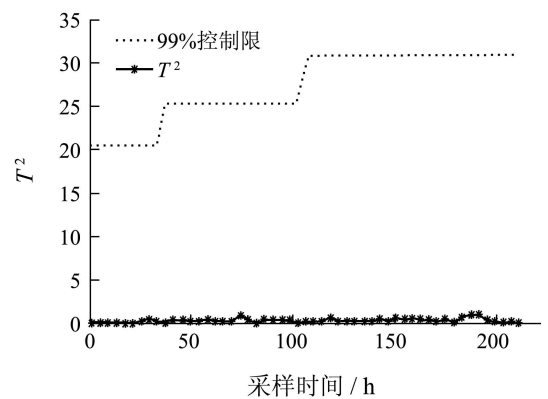


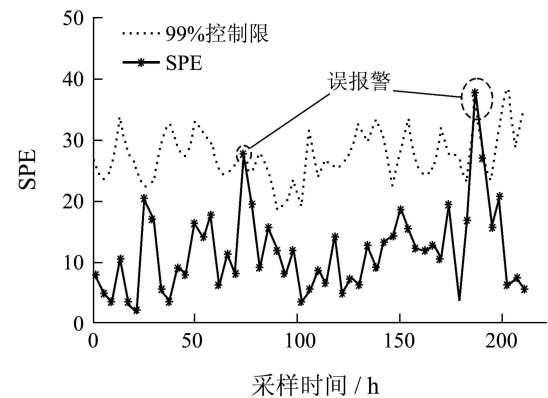
图 11(a) NM-MPCA



(a) NM-MPCA



(b) sub-PCA



(b) sub-PCA



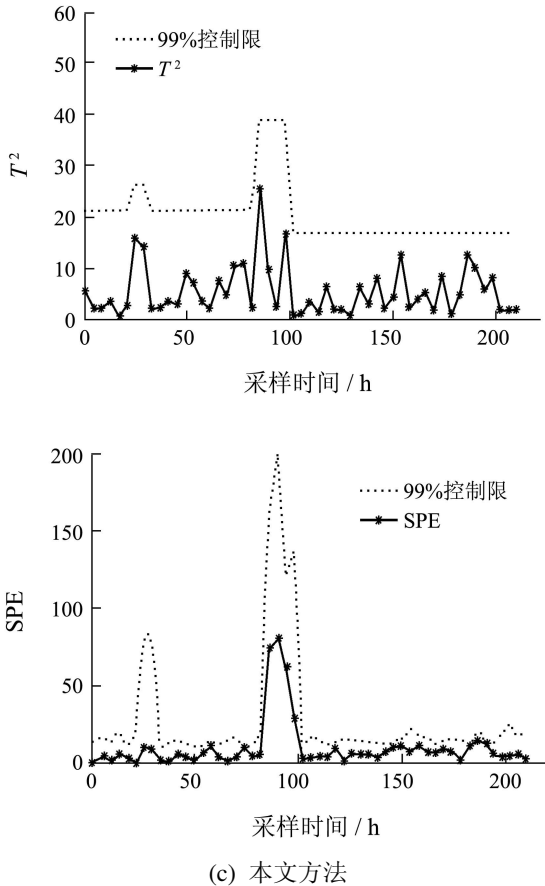


图 11 采用传统MPCA、sub-PCA和本文方法监测正常批次185的结果

Fig. 11 Monitoring results for MPCA, sub-PCA and the proposed method in normal batch 185

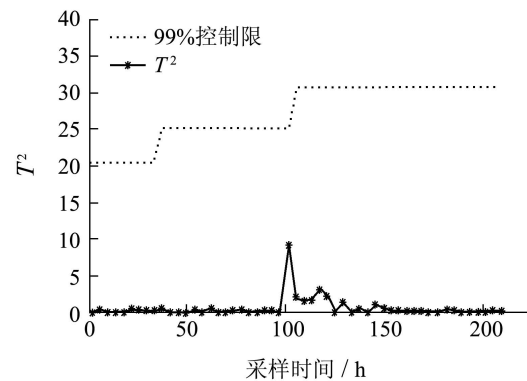
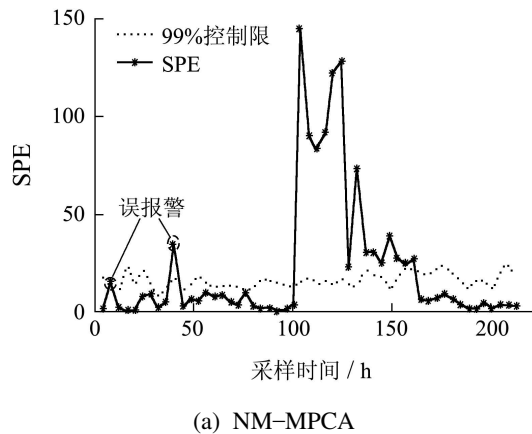
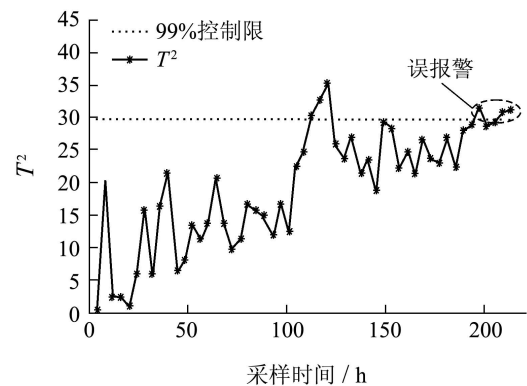
图12给出了对故障批次198进行监测的结果. 该批次在发酵中期, 由于机械原因导致通气量减小, 随后经操作员及时调整, 过程回到正常工况. 由于最终产品效价符合生产要求, 因此若从最终产品质量的角度考虑, 该批次是合格的. 由图可知, 在SPE监控图中, 本文方法和sub-PCA均在100 h检测到故障, 相比NM-MPCA提前4 h, 且NM-MPCA和sub-PCA方法均有不同程度的误报警现象. 在 $T^2$ 监控图中, sub-PCA方法未检测到故障发生, 而MPCA方法对故障的检测存在明显的滞后和误报警.

此外, 可以看到, 在随后的过程监测中MPCA方法不能很快摆脱该故障的影响, 并给出过程持续故障的监测结果. 造成这种现象的原因是MPCA将批次数据看成一个完整的对象进行处理, 不具备遗忘性, 从而导致过程前期的异常工况对后续的监测结果产生较大影响. 这就类似PID控制器中的积分作用太强而造成过程超调一样, 如果操作员

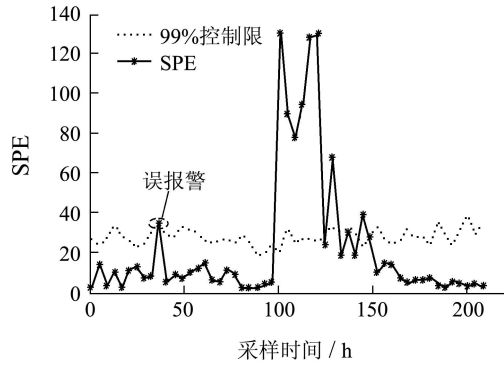
此时盲目按照MPCA给出的监测结果调整生产, 势必会造成过程不稳定. 而本文方法则能快速识别过程异常, 客观反映在故障发生后操作员为消除故障所做出努力, 准确反映最终产品质量情况, 避免发生误警.

当检测到故障后, 需要对故障进行进一步定位与诊断, 如图13中给出了本文方法在稳定阶段3随时间变化的贡献图, 从 $T^2$ 和SPE贡献图可得到基本相同的诊断结果, 即发现过程故障是由通气量异常所导致.

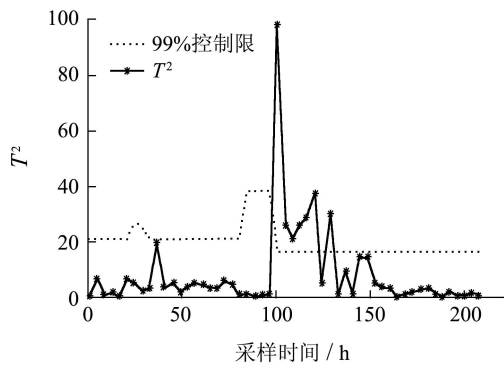
以上对工业发酵过程实际生产数据的应用可知, 相比MPCA和sub-PCA, 本文方法能够对过程做出更加合理的评价, 减少过程的误警率.



(a) NM-MPCA



(b) sub-PCA



(c) 本文方法

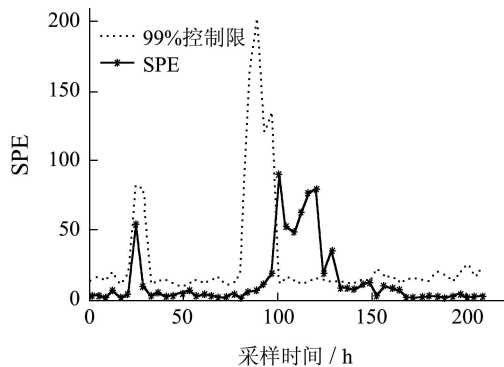


图 12 采用传统MPCA、sub-PCA和本文方法监测故障批次198的结果

Fig. 12 Monitoring results for MPCA, sub-PCA and the proposed method in fault batch 198

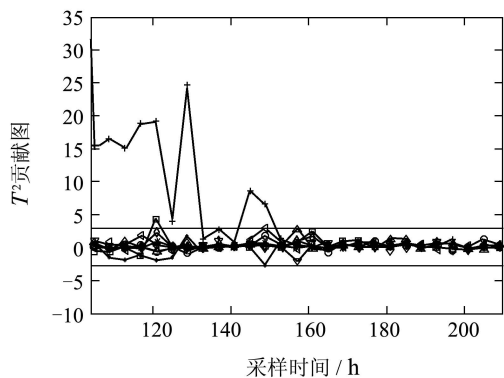


图 13 稳定阶段3对应的统计量贡献图

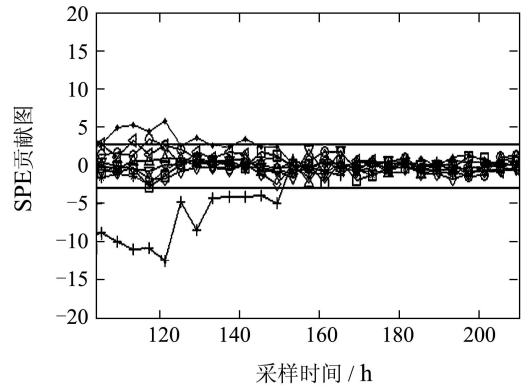


图 13 稳定阶段3对应的统计量贡献图

Fig. 13 Contribution plots in steady phase 3

### 6 结论(Conclusions)

本文提出一种新的软过渡KPCA-PCA的监控策略. 旨在克服相邻子类边界数据划分的不合理性及过渡过程的非线性等问题, 改善过程监测的可靠性和灵敏度. 当过程从一个操作阶段过渡到另一个操作阶段时, 可降低在线监控过渡数据的误报率和漏报率. 通过青霉素发酵仿真及实际工业生产过程的应用表明该策略能较好的反映各阶段的特征多样性, 对于解决多阶段间歇过程的故障监测问题, 具有一定的实用价值.

### 参考文献(References):

- [1] 赵春晖, 王福利, 姚远, 等. 基于时段的间歇过程统计建模、在线监测及质量预报 [J]. 自动化学报, 2010, 36(3): 366 – 374. (ZHAO Chunhui, WANG Fuli, YAO Yuan, et al. Phase-based statistical modeling, online monitoring and quality prediction for batch processes [J]. *Acta Automatica Sinica*, 2010, 36(3): 366 – 374.)
- [2] 齐咏生, 王普, 高学金. 一种基于改进MPCA的间歇过程监控与故障诊断方法 [J]. 化工学报, 2009, 60(11): 2838 – 2846. (QI Yongsheng, WANG Pu, GAO Xuejin. Batch process monitoring and fault diagnosis based on improved multi-way principal component analysis [J]. *Journal of Chemical Industry and Engineering*, 2009, 60(11): 2838 – 2846.)
- [3] 李荣雨, 荣冈. 基于故障映射向量和结构化残差的主元分析(PCA)故障隔离 [J]. 控制理论与应用, 2008, 25(6): 1099 – 1104. (LI Rongyu, RONG Gang. Principal component analysis(PCA) of fault isolation based on fault mapping vector and structured residual [J]. *Control Theory & Applications*, 2008, 25(6): 1099 – 1104.)
- [4] AGUADO D, FERRER A. Multivariate SPC of a sequencing batch reactor for wastewater treatment [J]. *Chemometrics and Intelligent Laboratory Systems*, 2007, 85(1): 82 – 93.
- [5] YU J, QIN S J. Multiway Gaussian mixture model based multiphase batch process monitoring [J]. *Industrial & Engineering Chemistry Research*, 2009, 48(18): 8585 – 8594.

- [6] CAMACHO J, PICO J. Multi-phase principal component analysis for batch processes modeling [J]. *Chemometrics and Intelligent Laboratory Systems*, 2007, 81(2): 127 – 136.
- [7] LU N, GAO F. Sub-PCA modeling and on-line monitoring strategy for batch processes [J]. *American Institute of Chemical Engineers Journal*, 2004, 50(1): 255 – 259.
- [8] ZHAO C, WANG F. Improved batch process monitoring and quality prediction based on multiphase statistical analysis [J]. *Industrial & Engineering Chemistry Research*, 2008, 47(3): 835 – 849.
- [9] ZHAO C, WANG F. Adaptive monitoring method for batch processes based on phase dissimilarity updating with limited modeling data [J]. *Industrial & Engineering Chemistry Research*, 2007, 46(14): 4943 – 4953.
- [10] CUI P L, LI J H. Improved kernel principal component analysis for fault detection [J]. *Expert Systems with Applications*, 2008, 34(2): 1210 – 1219.
- [11] TIAN X M, ZHANG X L, DENG X G. Multiway kernel independent component analysis based on feature samples for batch process monitoring [J]. *Neurocomputing*, 2009, 72(7/9): 1584 – 1596.
- [12] CHO J H, LEE J M. Fault identification for process monitoring using kernel principal component analysis [J]. *Chemical Engineering Science*, 2005, 60(1): 279 – 288.
- [13] BIROL G, UNDEY C, CINAR A. A modular simulation package for fed-batch fermentation: penicillin production [J]. *Computers and Chemical Engineering*, 2002, 26(11): 1553 – 1565.

#### 作者简介:

齐咏生 (1975–), 男, 博士, 目前研究方向为复杂工业过程建模与优化控制、故障监测与诊断, E-mail: qyslyt@163.com;

王普 (1962–), 男, 教授, 博士生导师, 目前研究方向为工业过程控制与优化、复杂系统控制和计算机控制系统, E-mail: wangpu@163.com;

高学金 (1973–), 男, 副教授, 硕士生导师, 目前研究方向为复杂工业过程建模与优化控制、生物传感器, E-mail: gaouxuejin@bjut.edu.cn.