

基于多时段 MPCA 模型的间歇过程监测方法研究

常玉清^{1,2} 王姝^{1,2} 谭帅² 王福利^{1,2} 杨洁²

摘要 针对间歇过程的多时段特性, 提出一种新的生产操作时段划分方法. 该方法利用反映过程特性变化的主成分个数、负载矩阵以及主成分矩阵的变化实现间歇过程子时段的三步划分. 根据各时间片主成分个数不同, 对生产操作时段进行粗划分. 为了更客观地反映负载矩阵以及主成分矩阵的相似性, 提出了基于加权负载向量夹角余弦的负载矩阵相似度度量以及基于加权主成分欧氏距离的主成分矩阵相似度度量方法. 以相似度最小原则, 对时间片矩阵进行奖惩竞争聚类, 进而实现了生产操作子时段的细划分. 将基于改进时段划分方法的 MPCA 建模应用于注塑成型过程在线监测, 实验结果验证了该方法的有效性.

关键词 间歇过程, 主成分分析, 操作时段划分, 过程监测, 注塑成型过程

DOI 10.3724/SP.J.1004.2010.01312

Research on Multistage-based MPCA Modeling and Monitoring Method for Batch Processes

CHANG Yu-Qing^{1,2} WANG Shu^{1,2} TAN Shuai² WANG Fu-Li^{1,2} YANG Jie²

Abstract A new operation stage separation method is proposed for the substage separation of multistage batch processes. Based on the changes in principle component number, loading matrixes, and principle component matrixes, which reveal evolvement of the underlying process behavior, a three-step substage separation method is realized. First, rough separation of operation substage is executed by the difference of principle component number. To reflect objectively the similarity of the loading matrixes and the similarity of the principle component matrixes, two improved similarity distances, based on weighted cosine of the angle between loading vectors and weighted Euclidean distances, are introduced respectively. According to the criterion of minimum similarity distances, time-slice matrixes are sorted using the rival penalized competitive learning algorithm to realize separation of operation substage more particularly for batch processes. The effectiveness of the proposed method is illustrated by applying it to the MPCA modeling and on-line monitoring of the injection mold process.

Key words Batch processes, principal component analysis (PCA), operation stage separation, process monitoring, injection mold process

主成分分析 (Principal component analysis, PCA)^[1-3] 和偏最小二乘 (Partial least square, PLS)^[4-5] 方法作为多元统计过程建模的核心技术, 只需要利用正常的生产过程数据建立模型, 在处理高维数据时具有很大的优势, 因此在工业生产过程中的应用越来越广泛. 20 世纪 90 年代中期 Nomikos 和 MacGregor 提出的多向主成分分析 (Multiway PCA, MPCA) 和多向偏最小二乘 (Multiway PLS, MPLS)^[5-9] 方法首次将多元统计分析方法成功应用

于间歇生产过程, 使得基于多元统计技术的间歇过程建模、过程监测及故障诊断等内容成为当今控制领域的研究热点之一.

多操作阶段是许多间歇过程的一个固有特征, 例如批生物发酵过程可以按细菌生长周期分为延滞期、生长期、稳定期和死亡期; 注塑成型生产过程按照操作过程可以分为注射、保压和冷却三个阶段. Kosanovich 等^[10] 在 1994 年将 MPCA 方法应用于聚合物反应工业过程, 提出针对具有不同特性的反应时段分别建立 MPCA 模型, 可以更加准确、有效地监测生产过程的运行状态. Lu 等^[11] 和 Zhao 等^[12] 分别在 2004 年和 2007 年提出了基于 K-means 的间歇过程子时段划分方法, 将间歇过程又进一步划分为多操作时段和过渡时段, 使得对间歇过程特性的分析和研究更加深入.

综上, 针对具有多操作阶段特性的间歇生产过程, 如何进行正确的操作时段及过渡时段的划分, 以及如何将不同时段模型进行合理融合, 是正确反映多时段间歇过程特性的关键. 本文提出一种改进

收稿日期 2009-10-09 录用日期 2010-04-19
Manuscript received October 9, 2009; accepted April 19, 2010
国家重点基础研究发展计划 (973 计划) (2009CB320601), 国家自然科学基金 (60774068) 资助
Supported by National Basic Research Program of China (973 Program) (2009CB320601) and National Natural Science Foundation of China (60774068)
1. 东北大学流程工业综合自动化教育部重点实验室 沈阳 110819
2. 东北大学信息科学与工程学院 沈阳 110819
1. Key Laboratory of Integrated Automation of Process Industry, Ministry of Education, Northeastern University, Shenyang 110819
2. School of Information Science and Engineering, Northeastern University, Shenyang 110819

的间歇过程时段划分方法用于实现间歇过程的操作时段及过渡时段的精确划分. 该方法的基本思想是: 不同的操作时段, 过程特性不同, 反映过程特性变化的变量空间投影方向和主成分得分就会不同. 本文首先根据主成分分析后主成分个数的变化对过程操作时段进行粗划分; 然后, 再根据变量空间投影方向的变化, 对过程操作时段进行第二步划分; 最后, 根据主成分得分的变化, 对过程操作时段进行第三步划分. 在后两步时段细划分过程中, 为了反映变量投影方向和主成分的特性不同, 并强调各主成分方向及主成分的重要性有差异, 本文提出了改进的相似度度量方法, 并基于该方法实现样本的奖惩竞争聚类. 在对间歇过程进行合理时段划分之后, 本文提出一种基于多时段 MPCPA 模型的间歇过程监测方法, 实现对间歇生产过程的监测. 最后, 本文将提出的方法应用于注塑成型过程, 实现注塑成型过程的在线监测.

1 MPCPA 建模

1.1 间歇过程数据的二维时间片展开

假设一个间歇生产过程, 其建模数据表示为最常用的三维矩阵形式 $\hat{X}(I \times J \times K)$, 其中 I 、 J 和 K 分别表示间歇操作次数 (也称批次)、过程变量个数以及每一次间歇操作过程的采样时刻个数. 为了利用 MPCPA 方法实现间歇过程建模, 首先需要将建模数据进行二维展开. 不同特性的间歇过程, 建模需求不同, 数据展开方式也不同, 大致具有六种方式^[13].

本文假设各个批次的操作时间相同, 并对间歇过程数据进行二维时间片展开 (也称 D 展开), 进而得到 K 个时间片矩阵 $X_k(I \times J)$, $k = 1, 2, \dots, K$.

1.2 MPCPA 建模

对间歇过程数据进行二维展开之后, MPCPA 建模和 PCA 建模的步骤没有任何区别. MPCPA 建模的工作对象是 K 个时间片矩阵 $X_k(I \times J)$, $k = 1, 2, \dots, K$, 按式 (1) 所示的主成分分析方法分别对其进行主成分分析后, 得到 K 个负载矩阵 $P_k(J \times a_k)$, $k = 1, 2, \dots, K$, 其中 a_k ($a_k < J$) 为对第 k 个时间片矩阵进行主成分分析后, 根据主成分累计贡献率方法确定所保留的主成分个数^[1]. 基于二维时间片展开的 MPCPA 建模原理如图 1 所示.

$$\begin{aligned} X_k &= T_k P_k^T + E_k \\ \hat{X}_k &= T_k P_k^T \\ E_k &= X_k - \hat{X}_k \end{aligned} \quad (1)$$

1.3 基于 MPCPA 的过程监测模型

基于 PCA 方法的过程监测是通过监视两个多

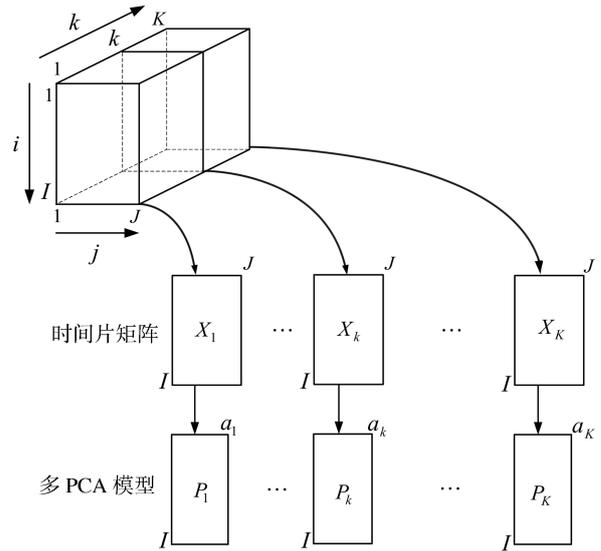


图 1 MPCPA 建模

Fig. 1 MPCPA modeling

元统计量, Hotelling - T^2 和残差子空间的 Q 统计量, 来实现对生产过程运行状态是否异常的实时监测. 其中, 第 k 个时间片第 i 个批次的过程变量 $\mathbf{x}_{k,i}(1 \times J)$ 所对应的 T^2 统计量定义如下:

$$T_{k,i}^2 = \mathbf{t}_{k,i} S_k^{-1} \mathbf{t}_{k,i}^T \quad (2)$$

其中, $(1 \times a_k)$ 维向量 $\mathbf{t}_{k,i} = X_{k,i} P_k$ ($i = 1, 2, \dots, a_k$) 为第 k 个时间片第 i 个批次的过程数据 $X_{k,i}(1 \times J)$ 的主成分变量. 对角矩阵 $S_k = \text{diag}\{\lambda_{k,1}, \lambda_{k,2}, \dots, \lambda_{k,a_k}\}$ 是由建模数据集 X_k 的协方差矩阵 $X_k^T X_k$ 的前 a_k 个特征值所构成.

Q 统计量, 也称为预测误差平方和指标 (Squared prediction error, SPE), 是测量偏离主成分模型的距离值, 定义如下:

$$SPE_{k,i} = (\mathbf{x}_{k,i} - \hat{\mathbf{x}}_{k,i})(\mathbf{x}_{k,i} - \hat{\mathbf{x}}_{k,i})^T \quad (3)$$

T^2 统计量的控制限可以利用 F 分布采用式 (4) 计算^[1]. 其中, A 为主成分个数, I 为建模批次, α 为显著性水平.

$$T_\alpha^2 \sim \frac{A(I-1)}{I-A} F_{A,I-A,\alpha} \quad (4)$$

SPE 统计量的控制限可由下式计算^[8]:

$$\begin{cases} SPE_{k,\alpha} = g_k \chi_{h_k,\alpha}^2 \\ g_k = \frac{v_k}{2m_k} \\ h_k = \frac{2m_k^2}{v_k} \end{cases} \quad (5)$$

其中, m_k 是建模数据集中所有批次的测量数据在第 k 个时刻 SPE 值的均值, v_k 是对应的方差.

2 基于改进多时段划分的 MPCA 建模及在线过程监测方法

2.1 改进的间歇过程的时段划分

当生产过程处于不同的操作时段时,其运行状态和操作目标往往有所差异,那么用来表征生产过程特性的过程变量信息就会表现出不同的特征. Lu 和 Zhao 等提出利用主成分分析后未舍弃任何信息的负载矩阵 $P_k(J \times J)$, $k = 1, 2, \dots, K$ 的变化来进行时段识别^[11-12],并针对不同的子时段分别建立子模型的 MPCA 建模方法.但是,在原始变量空间中如果包含噪声数据或是过程变量之间存在严重的相关性时,噪声和冗余信息的存在会对操作时段识别结果正确性产生影响.特别是变量冗余信息的存在,会扩大相关变量对阶段识别所起的作用,这将导致错误的阶段识别结果.另外,从一个具有 2 个变量的数据分布(见图 2)可以明显看出,三个时间片 k_1 , k_2 和 k_3 的数据具有不同的分布特性,应该属于生产操作的 3 个不同时段.然而,利用文献 [11-12] 提出的基于负载矩阵变化的分段方法虽然可以把 k_3 与另外两个时间片区分开来,但却会把具有相同变异方向、不同变异幅值的两个时间片 k_1 , k_2 两组数据分在同一个操作子时段内,这在反映生产过程操作阶段特性上明显是不合理的.另外,由于 k_1 , k_2 两组数据的变化幅值有较大差异,将这两组数据分在一个时间段内,会对 T^2 统计量的计算产生错误的影响.可见,仅利用负载矩阵进行时段划分是不充分的.

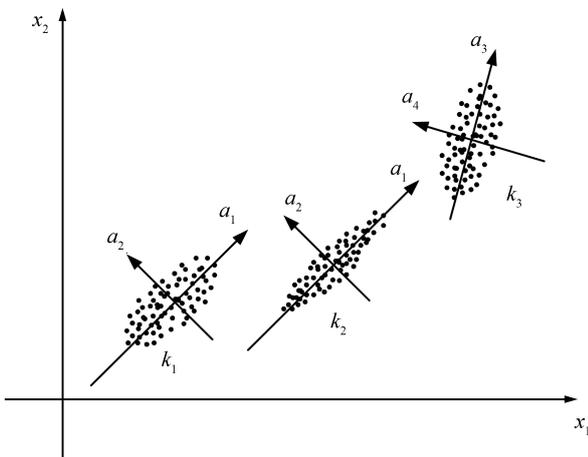


图 2 具有不同分布的三个时间片数据示意图

Fig. 2 Illustration of three time-slice data with different distributions

综上,本文提出一种利用主成分分析后的负载矩阵 $P_k(J \times a_k)$, $k = 1, 2, \dots, K$ 和主成分矩阵 $T_k(I \times a_k)$, $k = 1, 2, \dots, K$ 的双重变化来进行阶段

识别的新方法;同时为了实现更准确的时段划分,本文还提出了改进的负载矩阵及主成分矩阵相似度度量方法.本文提出的阶段识别算法的主要步骤及原理如下(假设建模数据均为标准化后的正常数据):

步骤 1. 根据连续时间片主成分分析后所具有主成分个数的不同,对操作时段进行粗划分.

当连续时间片具有相同的主成分个数时,就把这些连续的时间片划分为一个时段.

个别时刻由于测量噪声等原因的影响,会出现“跳类”现象.凡是持续时间小于 r_1 (结合实际生产工艺确定的一个较小的整数值,由于第一步时段划分为粗划分,所以 r_1 的值不必取得太小) 个时刻与其前后相邻时段的采样时刻主成分个数不相同的采样点均定义为第一步时段划分过程中的“跳类”点.对于“跳类”点,将其归于与其相邻时刻所属的时段内.

划分后的时段记为时段 a , 时段 b , \dots , 相应时段内的主成分个数分别记为 a_a, a_b, \dots .

该时段划分的主要目的在于: 1) 时段粗划分,根据主成分个数差异,粗略地划分出生产过程特性差异较大的几个主要稳定生产时段,区分不同的过程特性; 2) 减少第二步时段划分的计算量; 3) 为 T^2 统计量控制限的计算奠定基础.

很容易理解,当生产过程中两个不同时刻具有相同的主成分个数,但变量变异方向和幅值有较大的差异时,这两个时刻一定处在生产操作的不同时段.因此,对生产过程按照主成分个数不同进行操作时段的粗划分之后,还需要对具有相同主成分个数的各个时段进行细划分,以更细致地反映生产过程的多时段特性.

步骤 2. 根据主成分分析后负载矩阵的变化情况,即根据过程变量变化方向不同,对上一步划分后的各个时段进行第二步细划分.

1) 采用奖惩竞争算法^[14]对时段 a 内的时间片矩阵按照负载矩阵的相似性进行聚类分析.

奖惩竞争算法的基本思想是:不仅对获胜的数据中心加以修正来适应输入样本,而且竞争的次胜者要受到惩罚,使其远离输入样本.这样,多余的初始数据中心就会被最终移出样本区域,即奖惩竞争算法可以自动选择适当的类数和类中心位置,保证分类的合理性.这就是我们选择奖惩竞争学习方法对学习样本集进行分类的原因.

文献 [14] 是基于欧氏距离最小原则的一类样本聚类方法.由于负载矩阵描述的是过程变量的变异方向,向量夹角应该比欧氏距离更能反映方向间的相近性,因此本文提出改进的负载矩阵相似性度量方法,并基于该相似性度量值最小原则代替欧氏距离最小原则,对同一粗时段内的时间片矩阵进行奖

惩罚竞争聚类划分. 负载矩阵 $P_i(J \times a_a) = [\mathbf{p}_{i,1} \ \mathbf{p}_{i,2} \ \cdots \ \mathbf{p}_{i,a_a}]$ 和 $P_j(J \times a_a) = [\mathbf{p}_{j,1} \ \mathbf{p}_{j,2} \ \cdots \ \mathbf{p}_{j,a_a}]$ 之间的相似性度量表示为

$$d_{ij}^p = 1 - \sum_{l=1}^{a_a} \gamma_l \frac{|\mathbf{p}_{il}^T \mathbf{p}_{jl}|}{\|\mathbf{p}_{il}\| \cdot \|\mathbf{p}_{jl}\|} \quad (6)$$

式中的 γ_l 为加权系数, 用以强调不同投影方向的不同重要性. γ_l 按下式取值

$$\gamma_l = \frac{1}{\sum_{h=1}^{a_a} \frac{1}{h}}, \quad l = 1, 2, \dots, a_a \quad (7)$$

从式 (7) 可以看出, $\sum_{l=1}^{a_a} \gamma_l = 1$, 并且有 $1 > \gamma_1 > \gamma_2 > \dots > \gamma_{a_a} > 0$. a_a 为主成分分析后保留的主成分个数, 也是投影方向的个数.

式 (6) 中的第 2 部分表示负载矩阵 P_i 和 P_j 中 a_a 个投影方向的夹角余弦值的加权和, 该值是一个小于等于 a_a 的数值. 两个负载矩阵越相近, 该值越接近 a_a . 可见, $d_{ij}^p \geq 0$, 式 (6) 的值越接近 0, 表示负载矩阵 P_i 和 P_j 间的相似度越高.

由于不同的操作时段也可能具有相同的变量变异方向, 因此对时段 a 进行聚类以后, 再按如图 3 所示的原理, 根据生产操作时刻的连续性进行时段的二次划分. 从图 3 可以看出, 时段 A 和时段 F (第 1 类), 时段 B 和时段 D (第 2 类) 虽然被聚在同一类内, 但是由于操作时刻并不连续, 所以彼此应属于不同的操作时段.

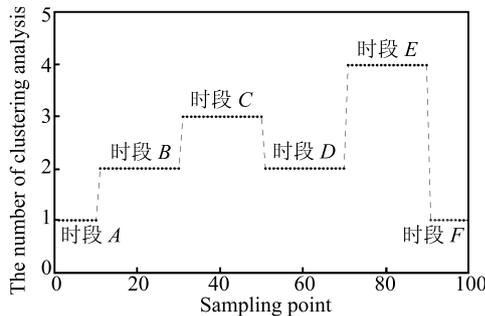


图 3 时段 a 聚类分析后的时段划分示意图

Fig. 3 Illustration of stage separation after classification analysis for stage a

凡是持续时间小于 r_2 个时刻, 或者持续时间大于 r_2 但与其前后相邻时段的负载矩阵相似度相差较大的采样点, 均定义为第二步时段划分过程中的“跳类”点. 对于个别时刻的“跳类”点, 仍将其归于与其相邻时刻所属的时段内.

2) 与上一步同理, 分别对其他的时段 b , 时段 c 等所有粗划分后的时段进行聚类分析, 再进行时段

的二次划分.

假设对所有时段细分之后得到操作子时段 A, B, \dots , 相应时段内的主成分个数分别记为 a_A, a_B, \dots . 可以看出, 经过第二步时段划分之后, 每个子时段内的时间片数据均具有相同主成分个数和相同的变量变异方向.

步骤 3. 与步骤 2 类似, 再根据主成分矩阵的变化情况, 对步骤 2 细分后的子时段进行第三步细分, 用于划分由变量变化幅值不同描述的操作时段变化.

1) 第二步时段划分后, 采用改进奖惩竞争算法对时段 A 内的时间片矩阵, 按照主成分矩阵的相似性进行聚类分析.

由于不同的投影方向提取的过程变量变异信息多少不同 (第一主成分提取的变异信息最多), 本文提出采用式 (8) 所示的加权欧氏距离来描述主成分矩阵 $T_i(I \times a_A) = [\mathbf{t}_{i,1} \ \mathbf{t}_{i,2} \ \cdots \ \mathbf{t}_{i,a_A}]$ 和 $T_j(I \times a_A) = [\mathbf{t}_{j,1} \ \mathbf{t}_{j,2} \ \cdots \ \mathbf{t}_{j,a_A}]$ 之间的相似性, 进行奖惩竞争聚类.

$$d_{ij}^t = \sum_{l=1}^{a_A} w_l \|\mathbf{t}_{i,l} - \mathbf{t}_{j,l}\| \quad (8)$$

式中的 w_l 为加权系数, 用以强调不同主成分的重要性. w_l 按下式取值

$$w_l = \frac{1}{\sum_{h=1}^{a_A} \frac{1}{h}}, \quad l = 1, 2, \dots, a_A \quad (9)$$

从式 (9) 可以看出, $\sum_{l=1}^{a_A} w_l = 1$, 且 $1 > w_1 > w_2 > \dots > w_{a_A} > 0$. a_A 为第二步时段划分后, 操作时段 A 的主成分个数, 也是投影方向的个数; $d_{ij}^t \geq 0$, 式 (8) 的值越小表示主成分矩阵 T_i 和 T_j 间的相似度越高.

与步骤 2 类似, 进行聚类分析后, 再按照生产操作时刻的连续性进行时段的第三次细分.

凡是持续时间小于 r_3 个时刻, 或者持续时间大于 r_3 但与其前后相邻时段的主成分矩阵相似度相差较大的采样点, 均定义为第三步时段划分过程中的“跳类”点. 对于个别时刻的“跳类”点, 仍将其归于与其相邻时刻所属的时段内.

2) 与上一步同理, 分别对时段 B 、时段 C 等所有的时段进行细分.

对所有时段细分之后, 得到最终的 N 个操作子时段 $1, 2, \dots, N$.

在时段划分过程中, 由于有些过渡时段持续时间较短, 所以“跳类”和“过渡”往往难于辨别. 本文结合如下共识实现对“跳类”点的判断: 1) 某一

稳定时段中间不会存在过渡点,可能会出现“跳类”点; 2) 过渡时段的特性不会发生“跳”变,通常是其前后相邻时段的过渡变化,但是“跳类”点的特性通常是“跳”变; 3) “跳类”通常是短时刻出现,通常是单一时刻出现,而“过渡”往往相对时间较长。

通过上述三步时段划分之后,当某一子时段或连续几个时段相对于其他时段所含时间片较少时,表示该子时段为生产操作时段间的过渡时段. 因此,本文提出的生产操作时段划分算法简单地实现了生产操作时段以及过渡时段的划分,避免了硬划分算法^[10]过渡段模型精度不高以及软划分算法^[11]复杂、计算量大且需要设置较多参数的缺陷. 更为重要的是,本文提出的时段划分算法更加细致地反映了生产过程的时段特性.

2.2 基于改进多时段划分方法的 MPCMA 建模的步骤

对生产过程进行操作时段划分之后,每个时段都具有不同的运行特征,因此应针对不同的子时段建立不同的 PCA 模型. 在同一时段内,由于运行特征相近,因此可以采用统一的代表模型. 本文提出的基于多时段 MPCMA 建模步骤如图 4 所示. 具体描述如下:

- 1) 获得 I 个批次的正常运行状态数据,按第 2.1 节所述方式,对其进行二维时间片展开,并对矩阵中的每一列数据进行标准化处理(0 均值和 1 标准差).
- 2) 对标准化后的时间片数据分别进行 PCA,得到负载矩阵 $P_i(J \times a_i)$, $i = 1, 2, \dots, K$ 及主成分矩阵 $T_i(I \times a_i)$, $i = 1, 2, \dots, K$.
- 3) 按照第 2.1 节所描述的方法对生产过程进行三步操作时段的细划分,得到操作子时段 $1, 2, \dots, i, \dots, N$, 假设第 i 个操作子时段对应的生产运行时刻为 $(k_{i-1} + 1) \sim k_i$, 其中 $k_0 = 0, k_N = K$.
- 4) 按式 (10) 计算每个子时段内的奇异值对角阵,用于在线监测时 T^2 统计量的计算.

$$\bar{S}_i = \frac{1}{k_i - k_{i-1} - 1} \sum_{j=(k_{i-1}+1)}^{k_i} S_j, \quad i = 1, 2, \dots, N \quad (10)$$

从式 (10) 可以看出,若把类似于图 2 中 k_1, k_2 的两种或多种具有不同变化幅值的时间片数据分在同一时段内,就会影响式 (10) 的计算结果,最终影响 T^2 的正确计算. 按照本文提出的方法进行时段划分之后,再对各子时段分别建模可以有效地解决这个问题.

- 5) 按式 (11) 计算每个子时段内的负载矩阵重心 $\bar{P}_i(J \times a_i)$, $i = 1, 2, \dots, N$.

$$\bar{P}_i = \frac{1}{k_i - k_{i-1} - 1} \sum_{j=(k_{i-1}+1)}^{k_i} P_j, \quad i = 1, 2, \dots, N \quad (11)$$

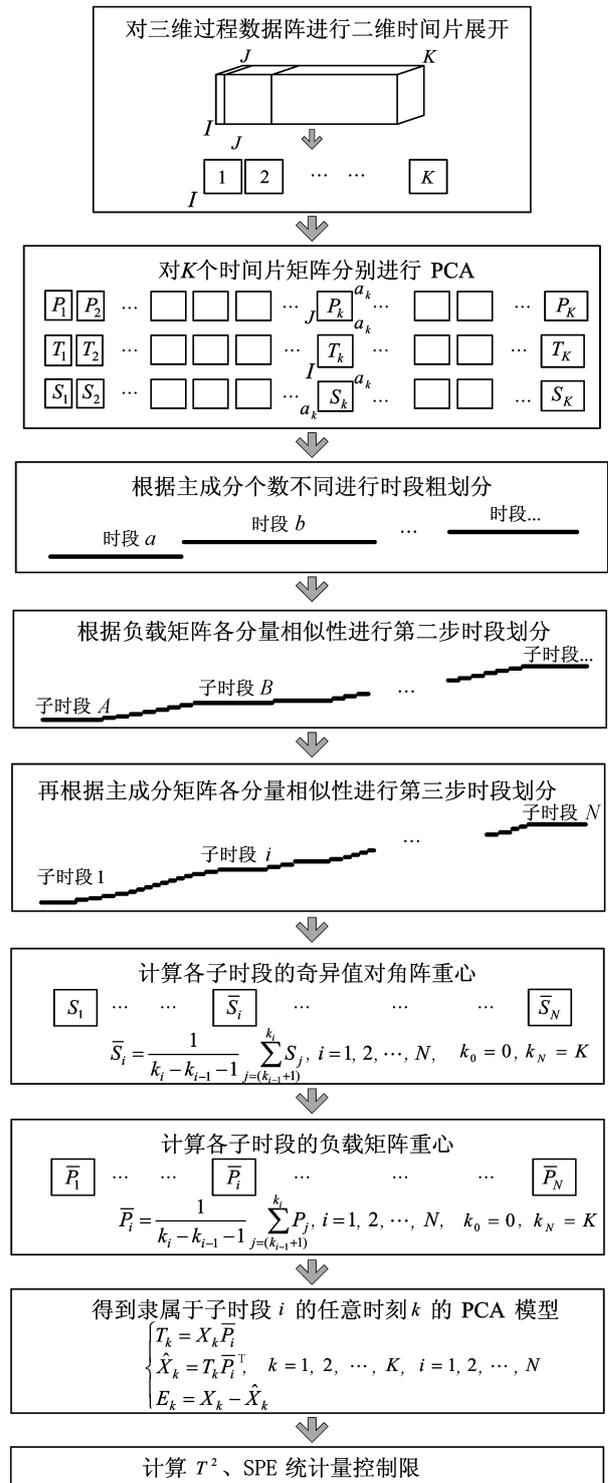


图 4 基于多时段 MPCMA 建模方法示意图
Fig. 4 Illustration of multistage-based MPCMA modeling method

综上所述, 生产操作过程中隶属于某一时段 i 的任意时刻 k , 都可以采用时段 i 所对应的负载矩阵重心 \bar{P}_i 来近似描述. 因此, 隶属于时段 i 的任意时刻 k 的 PCA 模型可以表示为

$$\begin{cases} T_k = X_k \bar{P}_i \\ \hat{X}_k = T_k \bar{P}_i^T \\ E_k = X_k - \hat{X}_k \end{cases} \quad (12)$$

6) 按第一步时段粗划分后的各子时段主成分个数不同, 根据式 (4) 分别计算各子时段的 T^2 统计量控制限.

7) 按式 (5) 计算各时间片的 SPE 统计量 $SPE_k, k = 1, 2, \dots, K$.

2.3 基于改进多时段划分方法的在线监测

在线过程监测时, 当实时获得当前时刻的采样数据 $\mathbf{x}(1 \times J)$ 后, 按下述步骤实现生产过程的在线监测:

- 1) 获取当前时刻的过程变量 $\mathbf{x}(1 \times J)$.
- 2) 判断当前时刻所处的操作子时段 i .
- 3) 利用操作子时段 i 的负载矩阵 $\bar{P}_i(J \times a_i)$, 按式 (11) 计算 X 所对应的主成分向量 $\mathbf{t}(1 \times a_i)$ 以及预测误差向量 $\mathbf{e}(1 \times J)$.

$$\begin{cases} \mathbf{t} = \mathbf{x} \bar{P}_i \\ \mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} = \mathbf{x}(I - \bar{P}_i \bar{P}_i^T) \end{cases} \quad (13)$$

式中, I 为 $J \times J$ 的单位阵.

4) 按式 (14) 计算数据 X 对应的 T^2 和 SPE 统计量.

$$\begin{cases} T^2 = \mathbf{t} \bar{S}_i^{-1} \mathbf{t}^T \\ SPE = \mathbf{e} \mathbf{e}^T \end{cases} \quad (14)$$

5) 判断 T^2 和 SPE 统计量是否超出相应的控制限. 若两者均未超出, 判定当前过程测量数据正常; 否则判定当前过程发生异常.

3 注塑过程中的应用研究

3.1 注塑过程简介

注塑过程是一个典型的具有多时段特性的间歇生产过程^[15]. 一个完整的注塑过程由闭模、注射座前进、注射、保压、塑化、冷却、开模、制品顶出等操作组成. 注射成型是热塑性或热固性塑料制品的主要成型方法之一. 在整个注塑操作过程中, 注射段、保压段和冷却段是决定塑料制品质量的三个操作阶段. 表征上述三个操作阶段生产状态的温度、压力、行程等 13 个过程变量及其工程单位如表 1 所示.

表 1 注塑过程的过程变量

Table 1 Process variables of injection process

变量描述	单位	变量描述	单位
喷嘴压力	bar	螺杆旋转速度	RPM
螺杆行程	mm	液压阀开度 1	%
注射速度	mm/s	液压阀开度 2	%
油缸压力	bar	模具温度	°C
塑化压力	bar	喷嘴温度	°C
背压	bar	料桶温度	°C
模腔压力	bar		

3.2 实验数据的获取

在同样的生产条件下, 以注射速度 24 mm/s、保压压力 200 bar、保压时间 8s、冷却时间 13s 为操作条件, 获取 35 组正常工况下的 13 个过程变量相关数据. 在整个操作过程中, 每个操作批次以 20 ms 为采样周期, 共获得 1100 个采样值. 在获取的数据中, 前 30 组数据用于建模, 后 5 组数据用于模型检验. 过程监测建模的三维数据阵表示为 $\hat{X}(30 \times 13 \times 1100)$.

3.3 子时段划分

将三维建模数据阵 $\hat{X}(30 \times 13 \times 1100)$ 按时间片进行二维展开, 得到 1100 个时间片矩阵 $X_1(30 \times 13), X_2(30 \times 13), \dots, X_{1100}(30 \times 13)$. 分别对每个时间片进行 PCA, 每个时间片矩阵按累计贡献率 95% 取主成分个数. 按主成分个数不同, 对该过程进行第一步时段粗划分 ($r_1 = 10$) 的结果如图 5 所示. 对照实际过程分析, 时段 a 、 b 属于注射段, 时段 c 为保压段, 时段 d 为塑化段, 时段 e 为冷却段. 可见, 按照主成分个数的区别, 可以大致划分出生产过程的不同操作阶段.

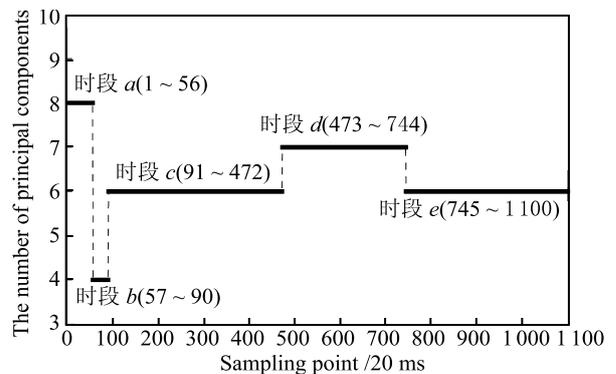


图 5 第一步时段划分后的结果

Fig. 5 The first substage separation results

在时段 $a \sim e$ 内, 分别采用奖惩竞争算法, 按照主成分投影方向的相似性进行第二步时段细划分 ($r_2 = 5$), 划分结果如图 6 所示. 从图 6 可以看出, 主成分个数相同, 各个主成分投影方向并不一定相同, 并且这种现象大多出现在粗时段的边缘. 例如: 时段 E 和 F 就是从时段 c 的后边缘划分出来的. 这些被再一次划分出来的小时段虽然主成分个数与原始时段相同, 但其投影方法却发生了变化.

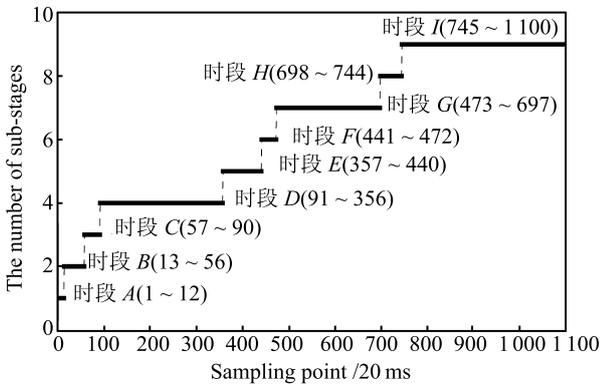


图 6 第二步时段划分后的结果

Fig. 6 The second substage separation results

在时段 $A \sim I$ 内, 再分别采用奖惩竞争算法, 按照主成分的相似性进行第三步时段细划分 ($r_3 = 2$), 共计得到 16 个操作子时段, 划分结果如图 7 所示. 从图 7 可以看出, 在主成分个数相同、投影方向相同的情况下, 由于主成分得分不同, 在一些时段的边缘又再一次划分出了新的时段.

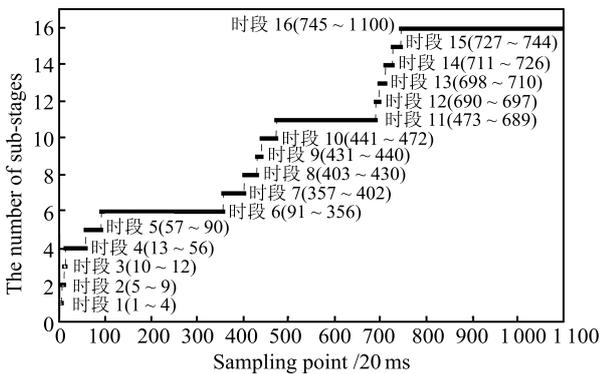


图 7 第三步时段划分后的结果

Fig. 7 The third substage separation results

从上述时段划分结果可以看出, 在稳定的生成操作时段内, 主成分投影方向和主成分得分是极其相似的. 当生产过程从一个生成操作时段向新时段过渡时, 主成分投影方向和主成分得分两者至少有一个会发生变化. 从图 7 可以看出, 一个间歇生产过程, 可以认为是由操作时段以及操作时段间的若干

过渡时段组成的.

从图 7 可以看出, 某些过渡时段持续时间很短 (例如: 子时段 1 只有 4 个时刻 80 ms), 但是这些小的子时段是有其存在的必要性的. 对于一个生产过程, 其过渡时段的长短并不是统一规定的. 如果一个间歇过程生产周期为 72 小时 (例如: 批生物发酵过程), 那么就没有必要将过渡时段细致到秒级或毫秒级, 细致到分钟级 (甚至小时级) 就可以满足过程特性描述的要求. 但是对于一个只有 20 s 左右生产周期的注塑生产过程, 为了能采用 PCA 线性建模方法准确地建立过程监测模型, 细致到文中的时段划分程度就非常必要了. 如果不对这些较短的过渡时段建立独立的监测模型, 即将 2 个或多个相邻的过渡时段建立 1 个 PCA 过程监测模型, 必然会忽略掉这些过渡时段的非线性特性, 因此势必会影响对该过渡时段的过程监测结果.

3.4 在线监测

从 5 组正常工况的测试数据中任选一组, 模拟实际过程的在线监测, 监测结果如图 8 所示. 从监测结果可以看出, 两个监测指标 T^2 和 SPE 都处在正常的统计控制限之内.

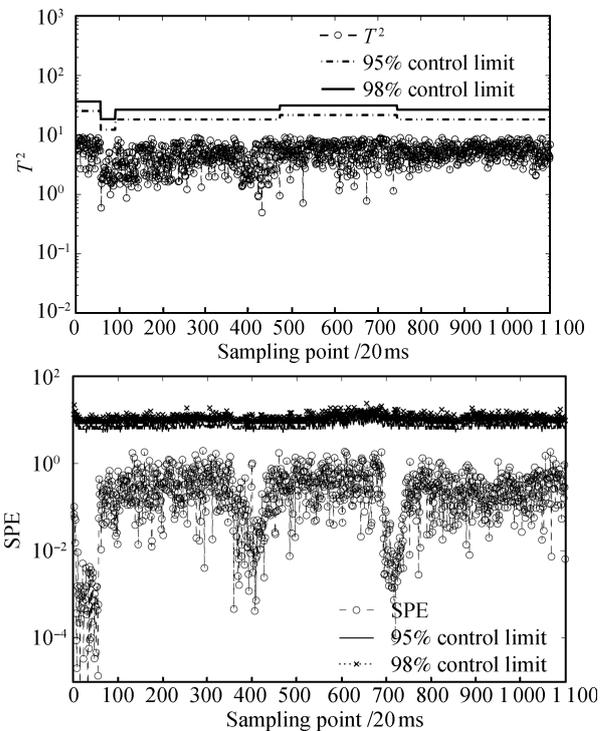


图 8 正常批次的过程在线监测图

Fig. 8 Process monitoring charts for a normal batch

另选一组测试数据, 从第 400 个采样点到第 800 个采样点之间通过将模具温度数据置 0 的方法, 模拟模具温度传感器断路, 人为引入传感器故障, 并对

过程进行在线监测, 监测结果如图 9 所示. 从图 9 可以看出, 当模具温度传感器断路故障时, 两个监测指标均比较明显地超出控制限, 及时地指示出过程出现异常.

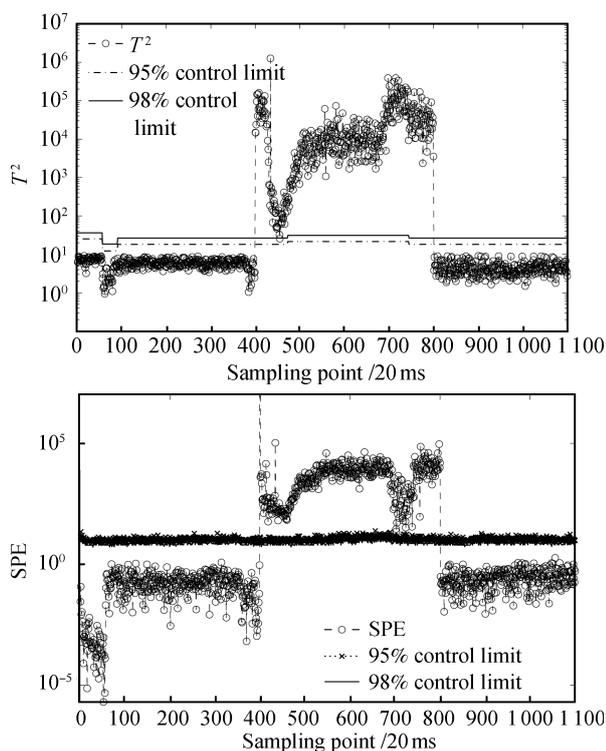


图 9 模具温度故障在线监测图

Fig. 9 Process monitoring charts for fault of mold temperature

从图 9 可以看出, 多个 PCA 模型联合进行在线监测时, 由于不同时间段的变量投影空间不同, 因此对同样大小的故障会产生不同的反应, 这对故障的进一步解释和定位是非常不利的. 所以在进行故障定位的时候应该选择合适的生产时段进行. 由于复杂工业生产过程出现故障的时候, 往往会发生故障变量的传递, 也就是说一个变量的故障可能会导致其他过程变量随之变化. 所以, 在实现故障定位时, 最可靠的故障定位信息是故障刚刚出现的时候. 因此, 故障定位应该根据系统异常刚出现的那个时段特性所对应的过程模型和数据进行.

4 结论

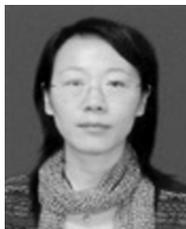
本文提出了一种新的间歇过程时段划分方法. 该方法根据各时间片主成分个数的不同、变量信息变异方向不同以及变量变异的范围不同, 对生产操作时段进行了三步划分. 针对负载矩阵和主成分矩阵描述变量变异的不同特性, 同时为了强调不同主成分具有不同的地位, 提出了改进的负载矩阵和主

成分矩阵相似性度量方法, 基于该方法实现了时间片样本的奖惩竞争聚类. 最后, 将基于改进时段划分方法的 MPCA 建模应用于注塑成型过程的在线监测. 从实验结果可以看出, 该方法能够有效地实现过程监测, 快速检测异常状况.

References

- 1 Jackson J E. *A User's Guide to Principal Components*. New York: Wiley, 1991
- 2 Kourti T, MacGregor J F. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics and Intelligent Laboratory Systems*, 1995, **28**(1): 3–21
- 3 Wang X Z. *Data Mining and Knowledge Discovery for Process Monitoring and Control*. London: Springer, 1999
- 4 Geladi P, Kowalshi B R. Partial least squares regression: a tutorial. *Analytica Chimica Acta*, 1986, **185**(1): 1–17
- 5 Hoskuldsson A. PLS regression methods. *Journal of Chemometrics*, 1988, **2**(3): 211–228
- 6 Lu Ning-Yun, Wang Fu-Li, Gao Fu-Rong, Wang Shu. Statistical modeling and online monitoring for batch process. *Acta Automatica Sinica*, 2006, **32**(3): 400–410 (陆宁云, 王福利, 高福荣, 王姝. 间歇过程的统计建模与在线监测, *自动化学报*, 2006, **32**(3): 400–410)
- 7 Nomikos P, MacGregor J F. Monitoring batch processes using multiway principal component analysis. *American Institute of Chemical Engineers Journal*, 1994, **40**(8): 1361–1375
- 8 Nomikos P, MacGregor J F. Multi-way partial least squares in monitoring batch processes. *Chemometrics and Intelligent Laboratory Systems*, 1995, **30**(1): 97–108
- 9 Nomikos P, MacGregor J F. Multivariate SPC charts for monitoring batch processes. *Technometrics*, 1995, **37**(1): 41–59
- 10 Kosanovich K A, Piovoso M J, Dahl K S. Multi-way PCA applied to an industrial batch process. In: *Proceedings of the American Control Conference*. Baltimore, Maryland: IEEE, 1994. 1294–1298
- 11 Lu N Y, Gao F R, Wang F L. A sub-PCA modeling and online monitoring strategy for batch processes. *American Institute of Chemical Engineers Journal*, 2004, **50**(1): 255–259
- 12 Zhao C H, Wang F L, Lu N Y, Jia M X. Stage-based soft-transition multiple PCA modeling and online monitoring strategy for batch processes. *Journal of Process Control*, 2007, **17**(9): 728–741
- 13 Westerhuis J A, Kourti T, MacGregor J F. Comparing alternative approaches for multivariate statistical analysis of batch process data. *Journal of Chemometrics*, 1999, **13**(3-4): 397–413

- 14 Xu L, Krzyzak A, Oja E. Rival penalized competitive learning for clustering analysis, RBF net and curve detection. *IEEE Transactions on Neural Networks*, 1993, 4(4): 636–649
- 15 Zhao C H, Wang F L, Mao Z Z, Lu N Y, Jia M X. Quality prediction based on phase-specific average trajectory for batch processes. *American Institute of Chemical Engineers Journal*, 2008, 54(3): 693–705



常玉清 东北大学副教授. 主要研究方向为复杂工业过程建模、过程监测及质量预测方法. 本文通信作者.

E-mail: changyuqing@mail.neu.edu.cn
(**CHANG Yu-Qing** Associate professor at Northeastern University. Her research interest covers process modeling, process monitoring, and quality

prediction and their applications in industry process. Corresponding author of this paper.)



王 姝 东北大学博士研究生. 主要研究方向为统计过程建模、过程监测和故障诊断方法.

E-mail: wangshu@ise.neu.edu.cn
(**WANG Shu** Ph.D. candidate at Northeastern University. Her research interest covers multivariate statistical modeling, process monitoring, and fault

diagnosis and their applications in industry process.)



谭 帅 东北大学博士研究生. 主要研究方向为统计过程建模、过程监测和故障诊断方法. E-mail: 94100542@qq.com

(**TAN Shuai** Ph.D. candidate at Northeastern University. Her research interest covers multivariate statistical modeling, process monitoring, and fault diagnosis and their applications in industry process.)



王福利 东北大学教授. 主要研究方向为复杂工业过程建模与优化、故障诊断.

E-mail: wangfuli@mail.neu.edu.cn
(**WANG Fu-Li** Professor at Northeastern University. His research interest covers modeling and optimization of complex system, and failure diagnosis.)



杨 洁 东北大学硕士研究生. 主要研究方向为统计过程建模、过程监测和故障诊断方法. E-mail: wnn-yj@163.com

(**YANG Jie** Master student at Northeastern University. His research interest covers multivariate statistical modeling, process monitoring, and fault diagnosis and their applications

in industry process.)