

一种新的间歇过程多模式辨识方法

关守平, 付冲, 侯金芬

(东北大学 信息科学与工程学院, 辽宁 沈阳 110819)

摘 要: 针对间歇过程的多模式划分问题, 提出了一种基于主角度相似性比较的多模式划分新方法, 有效克服了噪声或冗余数据对模式划分的影响. 该方法的基本思想是利用 PCA 对间歇数据按时间轴进行主成分建模, 然后利用主角度这一用于比较子空间相似度的方法进行主元模型相似性比较, 从而对各个模型和过渡过程进行有效辨识和划分; 在此基础上, 对上述方法进行了深入分析, 改进并完善了主角度相似性划分标准, 使这一方法更趋完善. 仿真结果检验了所提方法的有效性.

关 键 词: 多模式划分; 间歇过程; 主元分析; 主角度; 子空间相似性

中图分类号: TP 273

文献标志码: A

文章编号: 1005-3026(2015)03-0327-05

A New Multi-mode Identification Method of Batch Process

GUAN Shou-ping, FU Chong, HOU Jin-fen

(School of Information Science & Engineering, Northeastern University, Shenyang 110819, China. Corresponding author: GUAN Shou-ping, E-mail: guanshouping@ise.neu.edu.cn)

Abstract: A new method of multi-mode partition in the batch processes was presented on the basis of comparing the similarity of principal angles, which effectively overcomes the defects of noise or redundant data impaction on the phase partition. The main idea is to set up the principal component models of the batch data along time axis by employing the principal component analysis (PCA) approach, and then uses the principal angle method to compare the similarity of the principal component models, finally identify the steady phases and the transition phases effectively. Furthermore, an improved strategy is proposed to complete the partition criterion of the principal angle similarity. The simulation results demonstrate the effectiveness of the proposed method.

Key words: multi-mode partition; batch process; principal component analysis (PCA); principal angle; subspace similarity

多模式是许多间歇过程的一个固有特征. 过程的每个模式都有不同的过程主导变量和过程特征, 且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过程机理特性的变化呈阶段性^[1]. 因此, 对多模式间歇过程的统计建模和在线监控, 不仅要分析过程的整体运行状况是否正常, 更应深入分析过程的每一个操作子阶段是否正常.

传统的多向主元分析 (MPCA: multiway principal component analysis) 方法^[2]是将一个批次的所有数据当作一个统计样本来建立 MPCA 模型, 这样就忽视了间歇生产中的局部过程行为

特征, 很难揭示过程变量相关关系的变化, 容易造成较高的漏报率和误报率.

为克服传统 MPCA 存在的问题, 并充分考虑到间歇过程的多模式特性, 文献[3-6]从不同角度提出了间歇过程的多模式划分方法, 并取得了一定的应用效果. 但是上述各类方法均属于硬分类方法, 即在每个子时段将数据看成一个完整的对象处理, 不能很好地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检测结果产生很大影响. 在过程相关性发生变化的过渡状态时, 可能导致“误分类”现象, 增加漏报和误报概率.

针对硬分类方法存在的问题,研究者提出了软过渡的多模式划分方法,即建立过渡阶段的动态软过渡模型,如文献[7]提出了一种软过渡多 PCA 监控方法,采用模糊 C 均值聚类算法实现阶段划分;文献[8-9]提出了基于 PCA 方法建立过程的时间片模型,通过负载矩阵的变化来进行时段划分和软过渡过程辨识。

对于基于主元分析(PCA)的模式划分方法,若利用未舍弃任何信息的负载矩阵的变化来进行模式识别,则当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模式识别结果正确性产生影响;特别是变量冗余信息的存在,将导致错误的模式识别结果。

本文提出了一种基于主角度比较的间歇过程软过渡的多模式划分方法,可有效避免上述问题。该方法在分析两个模型之间的相似度时运用的是载荷矩阵,有效消除了噪音数据对模式分类的干扰。其基本思想是利用 PCA 对间歇数据按时间轴进行主成分建模,然后借用主角度这一比较子空间相似度的概念,建立主元模型之间相似度指标,从而对各个模型和过渡过程进行有效辨识和划分;在此基础上,继续完善了主角度相似度划分标准,使这一方法更趋完善。仿真结果检验了所提方法的有效性。

1 基于主角度的模式划分

在正常模式下,采集 I 个批次的正常间歇过程数据 $X(I \times J \times K)$ 对数据进行预处理,按照 K 个时刻分成 K 个矩阵 $X_k(I \times J) (k=1, 2, \dots, K)$, 对每个时间片的数据进行 PCA 建模,得到 K 个 PCA 模型,并且得到 K 个主元负荷矩阵 $P_i (J \times a_i) (i=1, 2, \dots, K)$, 其中 $a_i (a_i < J)$ 为第 i 个时间片矩阵的主元个数。对于任意两个相邻的负荷矩阵 $P_i (J \times a_i)$ 和 $P_{i+1} (J \times a_{i+1}) (i=1, 2, \dots, K-1)$, 假设 $a_i = \dim(P_i) \geq \dim(P_{i+1}) = a_{i+1} \geq 1$, 则它们之间的基于主角度的相似度指标 MSI (model similarity index) 可以按下式计算:

$$\text{MSI} = (1/a_{i+1}) \sum_{k=1}^{a_{i+1}} \cos^2 \theta_k = (1/a_{i+1}) \sum_{k=1}^{a_{i+1}} \lambda_k (P_{i+1}^T P_i P_i^T P_{i+1}) = (1/a_{i+1}) \text{trace}(P_{i+1}^T P_i P_i^T P_{i+1}) \quad (1)$$

这里, $a_i = \dim(P_i) \geq \dim(P_{i+1}) = a_{i+1}$, 如果是相反情况, $a_i = \dim(P_i) \leq \dim(P_{i+1}) = a_{i+1}$, 则式(1)中的 a_{i+1} 用 a_i 来代替, 算子 $\lambda_k(A)$ 表示矩阵 A 的第 k 个最大的特征值。因此,按照式(1)所

述, a_{i+1} 个主角度的余弦值之平方和与矩阵 $P_{i+1}^T P_i P_i^T P_{i+1}$ 的所有特征值之和相等, 即与矩阵 $P_{i+1}^T P_i P_i^T P_{i+1}$ 的迹(trace)相同。这表明, 可以通过式(1)简明地刻画 2 个 PCA 模型之间的相似度。此外, 显而易见, MSI 介于 0 与 1 之间, 当且仅当 $P_i \perp P_{i+1}$ 时, $\text{MSI} = 0$; 当且仅当 $P_i \supseteq P_{i+1}$ 时, $\text{MSI} = 1$ 。很明显, 当 MSI 越接近 1, 说明模型越相近, 可以融合为同一模式; 相反 MSI 越接近于 0, 说明模型之间差异性越大, 应分为不同模式。根据具体情况可以采用不同的阈值, 这里假设阈值为 θ 。当 $\text{MSI} \geq \theta$ 时, 划为同一模式; 当 $\text{MSI} < \theta$ 时, 则划为不同模式。对所有的 K 个数据片两两相邻的数据进行比较后就可得到 C 个大类, 整个过程如图 1 所示。

C_i 表示第 i 个类中采集时刻的模式数量。在具体建模时, 由于现场采集的数据难免受到各种原因的影响, 采集到的个别数据会发生错误, 也就是数据中出现离群点。在用本文所述方法进行建模时, 这样的离群点就表现在两个相同模式之间出现的一两个时刻的短暂的不同模式, 出现这类情况, 就可以去除这些离群点, 舍弃这些模式。但是对于 2 个稳定模式之间出现的较大量短暂的不同模式, 这是过渡模式, 就不能舍弃, 要对其建立过渡过程模型。通过这种模式划分得到的 C 个模式称为父类模式, 其划分过程如图 1 所示。

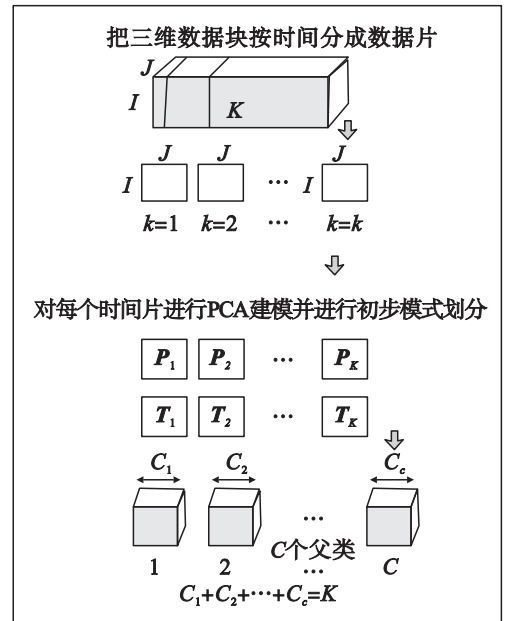


图 1 PCA 建模与父类划分

Fig. 1 PCA modeling and parent classes division

2 模式划分的改进

需要指出, 仅仅通过主角度来衡量两个子空

间的相似度是不够的,其缺陷表现在主角度会把一些主元负荷向量顺序不一致的模式划分为相同模式,而实际上它们是不同的模式. 比如,图 2 表示的是具有 2 个变量的数据分布图,可以明显看出,4 个时间片 1,2,3,4 的数据具有不同的分布特性,应该属于生产操作的 4 个不同操作模式. 但是通过基于主角度的求解矩阵相似度的方法,只能区分成两大类模式,其中 1,2,3 是一类模式,4 单独作为一种模式;但是 1,2,3 属于不同的模式,因此需要对 1,2,3 设计新的相似度指标,继续进行模式划分.

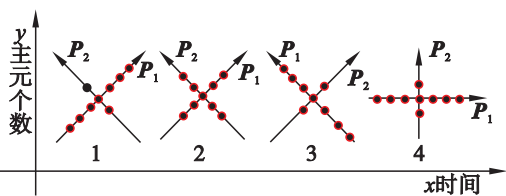


图 2 具有不同分布的 4 个 PCA 模型

Fig. 2 Four PCA models with different distributions

进一步研究发现,1,2 具有相同的主元负荷方向,这样比较两个模型的第一主元方向,看是否一致,一致则继续比较第二主元方向,直到有不一样主元方向出现为止. 若所有主元的负荷向量方向均保持一致,则属于同类模式,若有一个负荷方向不一致则为不同模式,这样 1,2 就会被划分为一种模式,而 3 又被划分为另一种模式;接下来在 1,2 中继续比较,这次按照主元得分的比重进行比较,从图中可以观察到 1,2 之间主元得分占的比重并不一样,所以它们也就被划分为不同的操作模式. 下面具体介绍这种改进的划分方法,并把这 4 种模式全部划分出来.

基于主角度的划分方法得到的 C 个父类如果已经把模式划分的比较合理,则可以停止继续划分,否则要在各父类内部继续进行细分. 每个大类内部仍然按照时刻顺序进行两两比较,在之前获得的 PCA 主元模型的基础上重新定义 MSI 指标,如式(2)所示:

$$MSI' = \sum_{k=1}^{a_{i+1}} \frac{\lambda_{a_{i+1},k}}{\sum_{m=1}^{a_{i+1}} \lambda_{a_{i+1},m}} (\mathbf{P}_{a_i,k}^T \mathbf{P}_{a_{i+1},k})^2. \quad (2)$$

其中: a_{i+1} 为第 $i+1$ 个主元模型的主元个数,同样假设 $a_i = \dim(\mathbf{P}_i) \geq \dim(\mathbf{P}_{i+1}) = a_{i+1}$; $\mathbf{P}_{a_{i+1},k}$ 为第 $i+1$ 个主元负荷矩阵第 k 个负荷向量方向. 这里的 MSI' 称之为狭义 MSI,以区别上一节中的广义 MSI. 由表达式可以看出,当所有 $\mathbf{P}_{a_i,k}$ 与 $\mathbf{P}_{a_{i+1},k}$, $k=1,2,\dots,a_{i+1}$ (这里假设 $a_{i+1} < a_i$, 假如

$a_{i+1} > a_i$ 则用 a_i 替换 a_{i+1}) 均相等时,即 $(\mathbf{P}_{a_i,k}^T \mathbf{P}_{a_{i+1},k})^2 = 1, k=1,2,\dots,a_{i+1}$ 时,有

$$MSI' = \sum_{k=1}^{a_{i+1}} \frac{\lambda_{a_{i+1},k}}{\sum_{m=1}^{a_{i+1}} \lambda_{a_{i+1},m}} (\mathbf{P}_{a_i,k}^T \mathbf{P}_{a_{i+1},k})^2 = \sum_{k=1}^{a_{i+1}} \frac{\lambda_{a_{i+1},k}}{\sum_{m=1}^{a_{i+1}} \lambda_{a_{i+1},m}} = 1. \quad (3)$$

否则 $MSI' < 1$. 当然 MSI' 越接近于 1, 表明 2 个模型越是相近. $\lambda_{a_{i+1},k} / \sum_{m=1}^{a_{i+1}} \lambda_{a_{i+1},m}$ 代表权值因子,以突显不同主元负荷分量,第一主元负荷方向的对应的权值最大,第二主元负荷方向次之,以此类推. 这里采用新的阈值 θ' , 当 $MSI' \geq \theta'$, 2 个模型划为一类模式; 当 $MSI' < \theta'$ 时, 2 个模型划为不同模式. 对所有的 C 个大类重新进行细分后, 就得到第二代分类——子类, 其个数为 C' , 以区别于 C 个父类.

在 C' 个子类的基础上继续细分, 此时每个子类中的所有模型都拥有相同的主元负荷方向, 且顺序都是一样的, 但其相同方向的得分向量并不一定完全一致, 也就是说在同一类的模式下还可以继续划分出不同模式. 这里采用的划分基准是得分向量在相应负荷方向的比重是否一致.

在每个子类内部, 对于每个时刻而言, 对应第 i 个时间片数据的负荷矩阵 $\mathbf{P}_i (J \times a_i)$ 的得分矩阵为 $\mathbf{T}_i (I \times a_i)$, 其中 J 表示变量的个数, a_i 为对应主元模型所保留的主元个数. 则新的评价基准 TMSI 定义为

$$TMSI_{i,i+1} = \sum_{k=1}^{a_{i+1}} \frac{\lambda_{a_{i+1},k}}{\sum_{m=1}^{a_{i+1}} \lambda_{a_{i+1},m}} |t_{a_i,k} - t_{a_{i+1},k}|. \quad (4)$$

式中: $\lambda_{a_{i+1},k}$ 表示某个子类中第 $i+1$ 个主元模型的第 k 个特征值; $t_{a_{i+1},k}$ 表示该特征值所对应的得分向量; $TMSI_{i,i+1}$ 表示相邻两个模型的得分矩阵相似度. 由定义可知, 当 2 个模型完全一致时, 有 $TMSI_{i,i+1} = 0$, 否则 $TMSI_{i,i+1} > 0$, 且 $TMSI_{i,i+1} < 1$. 当 $TMSI_{i,i+1}$ 大于一定的阈值时, 就可以判定其存在较大差异性, 从而继续分成不同模式, 否则仍划为同一模式. 最终共获得 C'' 个模式, 区别于父类和子类的划分, 这里 C'' 称之为孙类.

3 仿真实验与结果分析

考虑以下线性系统, 该系统有 7 个测量变量:

$$\mathbf{x}(t) = \mathbf{s}(t) \cdot \mathbf{A} + \mathbf{v}(t);$$
$$\mathbf{A} = \begin{bmatrix} 0.95 & -0.48 & 0 & 0 & 0.92 & 0 & 0 \\ 0 & -0.89 & 0 & 0 & 0 & 0.93 & -0.89 \\ 0 & 0 & 0.82 & -0.79 & 0 & 0.91 & 0 \end{bmatrix},$$
$$\mathbf{s} = [(s_1 + r_1), (s_2 + r_2), (s_3 + r_3)].$$

其中: s_i 是不相关的随机信号,服从 $[0,1]$ 之间的均匀分布;测量变量 \mathbf{x} 被噪声 \mathbf{v} 污染; \mathbf{v} 是零均值方差为 0.05 的高斯噪声. 过程工况的改变由 r_i 产生. r_i 的变化直接影响到 \mathbf{x} 的改变,因此,通过调整 r_i 可以改变过程在不同模式下运行.

通过对 r_i 的设置引入模式改变. 在第 401 个采样点处引入 r_1 斜坡上升信号,在第 550 个采样点处引入 r_2 阶跃信号,在第 941 个采样点处引入 r_3 加速下降信号,在第 1 481 个采样点处引入 r_4 先上升后下降信号. 令整个过程运行 1 000 个批次,每个批次采样 2 000 次,把采集的数据用来建立历史统计模型. 运用上节的总结分类方法,对这 2 000 个主元模型依次划分,划分结果如图 3 所示.

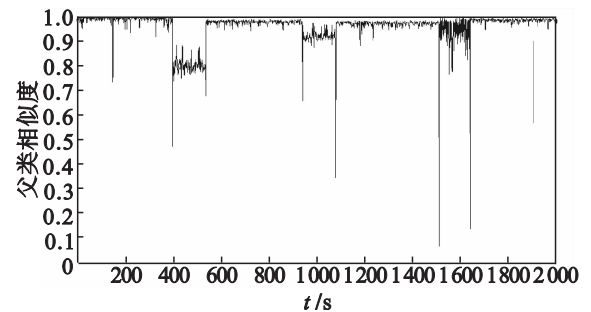


图 3 父类相似度比较结果
Fig. 3 Comparison results of parent-classes similarity

通过仿真结果可以看出,2 000 个主元模型可以明显地划分为 4 个稳定的模式和 3 个过渡模式,划分结果完全反应模型的不同模式,这表明这种基于主角度相似度的划分方法是有效的. 每个稳定模式内部的主元模型表现出很高的相似度,但模式一出现了奇异点,出现时刻在第 160 s 时刻,所以可以去除该模型;过渡阶段模型的相似度要低一些,表现出不太稳定的特点,这也符合过渡阶段的特性.

在第一步的基础上,继续在各个模式的内部进行细分,第二步子类的划分结果如图 4 所示. 在第二个模式中又划分出一个子模式,实现了子类的划分. 继续在第一步,第二步的基础上进行孙类的划分,模式分类的结果与第二步是相似的,如图 5 所示.

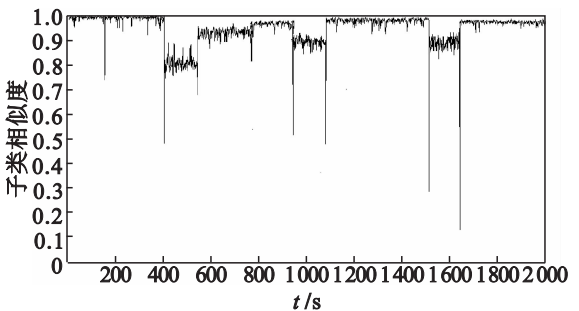


图 4 子类相似度比较结果
Fig. 4 Comparison results of subclasses similarity

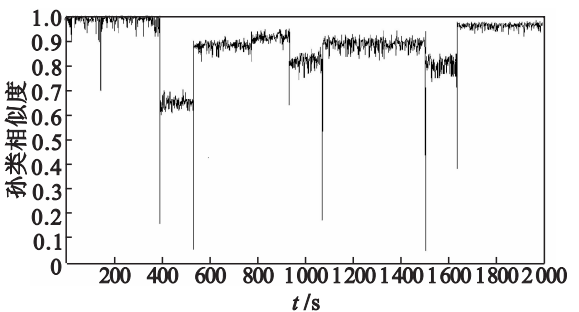


图 5 孙类相似度比较结果
Fig. 5 Comparison results of grandson-classes similarity

从本例来看,基于主角度的划分方法的结果,以及进行的第二步的划分结果是理想的. 第三步的划分与第二步是相似的,说明本例中例子只需要两步划分即可达到划分要求,证明了上述理论的正确性. 模式划分结果如表 1 所示.

表 1 各模式主元个数及统计阈值					
Table 1	Principal component number and statistical threshold				
模型号	模型一	模型二	模型三	模型四	模型五
保留主元	3	3	3	4	4
SPE 统计阈值	0.267 3	0.194 4	0.173 5	0.095 4	0.131 3
T ² 统计阈值	4.403 1	4.083 1	4.126 3	5.445 1	6.446 1

4 结 语

本文提出一种基于主角度相似度比较的新模式辨识方法,不但能实现软划分(辨识出过渡阶段),而且算法简单,便于实现. 该方法先按时间轴方向对每个时刻建立一个 PCA 模型,然后建立基于主角度的相似度指标,比较相邻时刻两两之间的主元模型相似度,最后辨识出各个模式. 针对主角度会把一些主元负荷向量顺序不一致的模式划分为相同模式的问题,文中对基于主角度相似度比较的分类方法进行了改进,使这一方法更趋完善. 经过仿真实验检验了所提方法的有效性.

(下转第 358 页)