

基于偏最小二乘的质量相关多模态故障检测技术

孔祥玉, 曹泽豪[†], 杜柏阳, 罗家宇

(火箭军工程大学 导弹工程学院, 西安 710025)

摘要: 偏最小二乘 (PLS) 算法通常适用于稳定工况下的工业过程故障检测. 在日趋复杂的工业过程中, 过程数据通常不满足正态分布, 存在非线性、动态、多模态等问题. 针对多模态问题, 已有大量模态区分方法可用, 但这些方法都未考虑质量相关因素, 因此并不适用于质量相关类算法. 为此, 针对质量相关类算法提出新的质量相关模态区分规则, 该规则通过核模糊聚类对添加线性递增时间变量的数据在时间方向上进行初步的聚类, 再通过质量相关指标进一步准确划分模态; 同时, 过程复杂化导致静态控制限不能满足故障检测的需求, 现存的动态控制限适用范围具有一定的局限性, 可通过改进动态控制限将其推广为广义动态综合控制限. 实验中, 先是基于两种非线性偏最小二乘模型将新方法应用于青霉素发酵过程故障检测中, 极大减少了漏报率和误报率. 最后, 通过数值仿真实验验证了添加线性递增时间变量的合理性.

关键词: 偏最小二乘; 故障检测; 多模态; 质量相关; 控制限; 非线性

中图分类号: TP273

文献标志码: A

Quality-related multimodal fault detection technique based on partial least squares

KONG Xiang-yu, CAO Ze-hao[†], DU Bo-yang, LUO Jia-yu

(Department of Missile Engineering, Rocket Force University of Engineering, Xi'an 710025, China)

Abstract: The partial least squares (PLS) algorithm is generally suitable for the fault detection of industrial processes under stable conditions. In increasingly complex industrial processes, process data does not satisfy the normal distribution and there are nonlinear, dynamic, and multimodal problems. For the multimodal problem, a large number of modality discrimination methods are available. But these methods do not consider quality-related factors, and therefore these methods do not apply to quality-related algorithms. For this reason, this paper proposes a new quality-related modal differentiation rule for quality-related algorithms. This rule uses kernel fuzzy clustering to perform preliminary clustering on the increase data for linearly increasing time variables, and then further quantifies the modality accurately by the quality related indicators. At the same time, the complexity of the process makes the static control limit unable to meet the requirements of fault detection. The applicable range of the existing dynamic control limit has certain limitations. By improving the dynamic control limit, this paper extends it to the generalized dynamic integrated control limit. In the experiment, firstly based on two nonlinear partial least squares models, the proposed methods are applied to the fault detection of penicillin fermentation process, which greatly reduce the false negative rate and false positive rate. Finally, a numerical simulation experiment verifies the rationality of adding a linear incremental time variable.

Keywords: partial least squares; fault detection; multimodal; quality-related; control limit; nonlinear

0 引言

在工业过程监控领域, 以主元分析 (PCA)^[1]、独立成分分析 (ICA)^[2] 以及偏最小二乘 (PLS)^[3] 等为核心的多元统计过程监控 (Multivariate statistical process monitoring, MSPM) 方法, 因其在处理高维、高度耦合数据时具有独特的优势, 越来越受到研究人员和现场工程师的青睐^[4]. 由于 MSPM 只需要正常工况下的

历史数据来建立模型而不需要过程机理知识, 目前基于这类多变量统计模型的过程监测、故障诊断、控制器设计、质量控制等算法层出不穷, 并且已成功地应用到各种连续生产过程中^[5].

近年来, 工业过程向着非线性、动态、多模态等复杂化方向发展. 大多数工业在生产过程中都有多种操作以及多个阶段, 并采用多种制造策略以及多种不

收稿日期: 2018-03-11; 修回日期: 2018-06-26.

基金项目: 国家自然科学基金项目 (61673387, 61374120).

责任编委: 高会军.

[†]通讯作者. E-mail: 578021174@qq.com.

同的原材料. 这些复杂化已产生两种较为严重的问题: 1) 多模态问题. 由于工业过程的复杂化, 过程本身通常是非稳定的, 存在大量模态转变, 这导致工业过程中有多个稳定模态, 采用传统MSPM整体建模的检测效果较差. 2) 故障检测控制限问题. 传统MSPM监控统计量的控制限是在稳定工况中假定过程数据满足正态分布情况下确定的^[6], 工业过程复杂化导致了从过程获取的数据不符合正态分布^[7], 采用传统故障检测控制限进行过程监控效果极差.

针对多模态问题, 学者们提出了两个不同的方向: 对不同模态的数据或模型处理后整合为一个整体, 然后对总体进行监控^[8-9]; 区分模态后分别建立监控模型^[10-11]. 学者们沿用分模态监控思想对高斯混合模型(Gaussian mixture model, GMM)的多模态问题进行了深入研究^[12-15]. 近年来, 基于隐马尔可夫模型(Hidden markov models, HMM)多模态过程监控方法成为研究的热点^[16-20], HMM模型不仅可以对数据的多模态进行建模, 还可以捕捉模态转移概率. 从一个模态到另一个模态之间往往存在过渡过程, 多模态特性的工业过程最容易在模态切换这一动态渐变过渡过程中发生故障^[21], 因此有必要区分出模态间的过渡模态. 针对这一问题, Singhal等^[22]、Zhao等^[23]、Yao等^[24]以及谭帅等^[5]相继提出了过渡模型的区分方法及监控方法. 最近, 谭帅等^[25]提出可以从历史数据中寻找与当前数据特征相似的数据进行补充, 解决了新出现的模态过程短期内无法获得充足建模数据这一难题. Zhang等^[26]提出了过渡模态建立综合监测模型的新方法, 而张淑美等^[27]通过改进的K-means聚类算法给出了建模数据模态自适应识别技术.

总结以上文献提出的方法, 可得出解决多模态问题的3个步骤: 1) 通过合适的聚类算法对多模态数据进行聚类, 初步划分模态; 2) 给出进一步划分模态的指标, 对模态进行进一步的细化分; 3) 对划分后各模态数据分别建立监控模型.

在第1个步骤中, 文献[8]采用分层聚类技术, 文献[11]采用启发式平滑聚类(HSC)技术, 文献[19]采用新的聚类技术正交非负矩阵因子(Orthogonal nonnegative matrix factorization, ONMF), 文献[23]采用模糊聚类(FCM)对加载矩阵进行分类而对数据进行分类, 而文献[27]采用改进的K-means聚类算法. 目前复杂的工业过程往往都是非线性动态系统, 以上聚类算法并没有考虑到数据的非线性特性. 本文通过核模糊聚类算法(KFCM)针对数据的非线性进行有效的模态聚类, 初步完成模态的划分.

在第2个步骤中, 基于GMM方法的文献[12-15]基于整体数据提取GMM模型指标, 将数据分为概率分布集群. 基于HMM方法同样基于整体数据提取HMM模型指标对数据区分模态. 不同于GMM和HMM的方法, 文献[22, 24]采用了PCA模型主元得分之间的角度指标进行模态相似度辨识. 而上述这些指标都基于数据整体进行提取, 没有考虑到质量相关因素. 本文从数据中提取出质量相关指标, 针对质量相关类算法提出质量相关的模态区分规则.

在第3个步骤中, 文献[5, 8-11, 21-25]采用线性PCA算法建立监控模型, 文献[12-20]采用GMM和HMM并建立监控模型. 这些模型的建立都未考虑数据的非线性特性, 且在故障检测时控制限也采用传统的“一线制”, 没有考虑到系统动态特性对控制限的影响. 本文采用非线性PLS算法建立监控模型, 并针对工业复杂化造成的故障检测控制限问题, 通过改进动态控制限提出广义动态综合控制限.

综合以上3个步骤, 本文形成一套非线性、动态质量相关的多模态故障检测技术.

1 非线性偏最小二乘

复杂工业过程内部必然存在严重的非线性特性, 传统线性算法不能够很好地解决非线性问题. 为此, 本文采用非线性偏最小二乘中的核偏最小二乘(KPLS)算法和核全潜结构投影(TKPLS)算法.

1.1 核偏最小二乘算法及其故障检测技术

Rosipal等^[28]基于核函数理论以及Cover定理提出了核偏最小二乘(KPLS)算法. KPLS算法通过核函数将原始变量由低维空间映射到再生核希尔伯特空间(RKHS), 即高维特征空间, 使非线性数据在高维特征空间中呈现线性关系, 然后在RKHS中利用核函数建立线性PLS模型.

假设过程数据为 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T \in \mathbf{R}^{n \times m}$, 质量数据为 $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]^T \in \mathbf{R}^{n \times p}$, \mathbf{X} 和 \mathbf{Y} 是标准化后的数据矩阵. 定义从 n 维原始数据空间向RKHS的非线性映射为 ϕ , 则 \mathbf{X} 可由 ϕ 映射为 $\Phi_{\text{raw}} = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_n)]^T \in \mathbf{R}^{n \times M}$, KPLS就是从 Φ_{raw} 和 \mathbf{Y} 中依次提取 A 组潜向量 \mathbf{t} 和 \mathbf{u} 的算法, 潜变量个数 A 一般由交叉验证^[29]得到. 由于非线性映射 ϕ 很难求解, 可通过核函数来代替. 定义 $\mathbf{K} \in \mathbf{R}^{n \times n}$ 为核矩阵, \mathbf{K}_{raw} 代表 $\Phi_{\text{raw}} \Phi_{\text{raw}}^T$, 即

$$K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle, \quad (1)$$

其中核函数 $K(\cdot)$ 代表非线性映射函数在RKHS中的

点积运算, $i, j = 1, 2, \dots, n$. 应用最广泛的核函数是高斯核函数, 可以通过调整高斯核半径 σ 将有限样本在 RKHS 中线性分开. 本文所选高斯核函数如下:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\sigma)^2). \quad (2)$$

为消除平均效应, 需要对 \mathbf{K}_{raw} 进行零均值处理, 定义 $\mathbf{1}_n$ 为元素为 1 的 n 维列向量, \mathbf{I}_n 为 $n \times n$ 的单位矩阵, 则 \mathbf{K}_{raw} 的零均值处理为

$$\mathbf{K} = (\mathbf{I}_n - (1/n)\mathbf{1}_n\mathbf{1}_n^T)\mathbf{K}_{\text{raw}}(\mathbf{I}_n - (1/n)\mathbf{1}_n\mathbf{1}_n^T). \quad (3)$$

可以构建如下模型:

$$\begin{cases} \Phi = \hat{\Phi} + \Phi_r = \mathbf{T}\mathbf{P}^T + \Phi_r, \\ \mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{Y}_r = \mathbf{T}\mathbf{Q}^T + \mathbf{Y}_r. \end{cases} \quad (4)$$

其中: \mathbf{T} 为得分矩阵, \mathbf{P} 和 \mathbf{Q} 为负载矩阵, Φ_r 和 \mathbf{Y}_r 为残差部分, 权重矩阵 $\mathbf{R} = \Phi^T \mathbf{U}(\mathbf{T}^T \mathbf{K} \mathbf{U})^{-1} \in \mathbf{R}^{M \times A}$.

假设测试数据为 $\mathbf{x}_{\text{new}} \in \mathbf{R}^m$, 则测试数据的核化形式为 $\mathbf{k}_{\text{new}}^{\text{raw}} = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_{\text{new}})$, $\mathbf{k}_{\text{new}}^{\text{raw}}$ 的零均值处理为

$$\mathbf{k}_{\text{new}} = (\mathbf{I}_n - (\frac{1}{n})\mathbf{1}_n\mathbf{1}_n^T) (\mathbf{k}_{\text{new}}^{\text{raw}} \mathbf{I}_n - (\frac{1}{n})\mathbf{k}_{\text{raw}}\mathbf{1}_n\mathbf{1}_n^T). \quad (5)$$

新的得分向量为

$$\mathbf{t}_{\text{new}} = \mathbf{R}^T \phi(\mathbf{x}_{\text{new}}) = (\mathbf{U}^T \mathbf{K} \mathbf{T})^{-1} \mathbf{U}^T \mathbf{k}_{\text{new}}. \quad (6)$$

故障检测统计量 T^2 和 SPE 的详细参数如表 1 所示, 详细参数及 SPE 统计量公式参见文献 [30].

表 1 KPLS 统计量公式和传统控制限

| 统计量 | 公式 | 传统控制限 |
|-------|---|--|
| T^2 | $\mathbf{t}_{\text{new}}^T \mathbf{\Lambda}^{-1} \mathbf{t}_{\text{new}}$ | $\frac{A(n^2 - 1)}{n(n - A)} F_{A, n-A, \alpha}$ |
| SPE | $\ \phi_r(\mathbf{x}_{\text{new}})\ ^2$ | $g\chi_{h, \alpha}^2$ |

1.2 核全潜结构投影算法及其故障检测技术

KPLS 与传统 PLS 有着同样的缺点^[31], Peng 等^[30] 通过改进 KPLS 算法提出了核全潜结构投影 (TKPLS) 算法. TKPLS 将 $\hat{\Phi}$ 分解为 Φ_y 和 Φ_o , 将 Φ_r 分解为 Φ_{rp} 和 Φ_{rr} , Φ_y 代表 $\hat{\Phi}$ 中与 \mathbf{Y} 相关的部分, 主要用于预测 \mathbf{Y} ; Φ_o 代表 $\hat{\Phi}$ 与 \mathbf{Y} 正交的部分; Φ_{rp} 代表 Φ_r 中变异最大的部分; Φ_{rr} 代表 Φ 中的残差部分, 最终模型如下:

$$\begin{cases} \Phi = \mathbf{T}_y \mathbf{P}_y^T + \mathbf{T}_0 \mathbf{P}_0^T + \mathbf{T}_r \mathbf{P}_r^T + \Phi_{rr} = \\ \Phi_y + \Phi_o + \Phi_{rp} + \Phi_{rr}, \\ \mathbf{Y} = \mathbf{T}_y \mathbf{Q}_y^T + \mathbf{Y}_r. \end{cases} \quad (7)$$

其中: \mathbf{T}_y 是 \mathbf{T} 中与 \mathbf{Y} 直接相关的部分, \mathbf{T}_0 是原始 \mathbf{T} 矩阵中与 \mathbf{Y} 正交的成分, \mathbf{T}_r 是原始残差 Φ_r 中含有较大变化方差部分的得分, Φ_{rr} 是 Φ 中最终的残差部分代表噪声, \mathbf{P}_y 、 \mathbf{P}_o 、 \mathbf{P}_r 和 \mathbf{Q}_y 为对应部分的负载矩阵. 若

测试数据为 $\mathbf{x}_{\text{new}} \in \mathbf{R}^m$, 可得 $\mathbf{k}_{\text{new}}^{\text{raw}}$ 及其零均值形式 \mathbf{k}_{new} , 则新的得分向量如下:

$$\mathbf{t}_{y\text{new}} = \mathbf{Q}_y^T \mathbf{Y}^T \mathbf{T} (\mathbf{U}^T \mathbf{K} \mathbf{T})^{-1} \mathbf{U}^T \mathbf{k}_{\text{new}}, \quad (8)$$

$$\mathbf{t}_{o\text{new}} =$$

$$\mathbf{V}_o^{-1/2} \mathbf{D}_o^T [\mathbf{I} - \mathbf{T}_y (\mathbf{T}_y^T \mathbf{T}_y)^{-1} \mathbf{T}_y^T] \mathbf{T} \mathbf{T}^T \mathbf{K}.$$

$$[\mathbf{I} - \mathbf{T} \mathbf{T}^T \mathbf{T}_y (\mathbf{T}_y^T \mathbf{T}_y)^{-1} \mathbf{Q}_y^T \mathbf{Y}^T] \mathbf{T} (\mathbf{U}^T \mathbf{K} \mathbf{T})^{-1} \mathbf{U}^T \mathbf{k}_{\text{new}}, \quad (9)$$

$$\mathbf{t}_{r\text{new}} = \mathbf{V}_r^{-1/2} \mathbf{D}_r^T (\mathbf{I} - \mathbf{T} \mathbf{T}^T).$$

$$[\mathbf{I} - \mathbf{K} \mathbf{T} (\mathbf{U}^T \mathbf{K} \mathbf{T})^{-1} \mathbf{U}^T] \mathbf{k}_{\text{new}}. \quad (10)$$

其中 \mathbf{V} 和 \mathbf{D} 分别是对应 $\mathbf{K}_o = \Phi_o \Phi_o^T$ 和 $\mathbf{K}_r = \Phi_r \Phi_r^T$ 两个核矩阵的特征值对角阵和特征向量矩阵. 故障检测统计量 T_y^2 、 T_o^2 、 T_r^2 和 SPE 的公式以及控制限如表 2 所示, 详细参数及 SPE 统计量公式参见文献 [30].

表 2 TKPLS 统计量公式和传统控制限

| 统计量 | 公式 | 传统控制限 |
|---------|---|--|
| T_y^2 | $\mathbf{t}_{y\text{new}}^T \mathbf{\Lambda}_y^{-1} \mathbf{t}_{y\text{new}}$ | $\frac{A_y(n^2 - 1)}{n(n - A_y)} F_{A_y, n-A_y, \alpha}$ |
| T_o^2 | $\mathbf{t}_{o\text{new}}^T \mathbf{\Lambda}_o^{-1} \mathbf{t}_{o\text{new}}$ | $\frac{A_o(n^2 - 1)}{n(n - A_o)} F_{A_o, n-A_o, \alpha}$ |
| T_r^2 | $\mathbf{t}_{r\text{new}}^T \mathbf{\Lambda}_r^{-1} \mathbf{t}_{r\text{new}}$ | $\frac{A_r(n^2 - 1)}{n(n - A_r)} F_{A_r, n-A_r, \alpha}$ |
| SPE | $\ \phi(\mathbf{x}_{\text{new}})\ ^2$ | $g\chi_{h, \alpha}^2$ |

2 质量相关多模态区分方法及故障检测技术

从复杂工业过程中获取的数据往往具有严重的非线性问题, 本文采用核模糊聚类 (KFCM) 对非线性数据进行初步的模态区分. 文献 [5, 8-27] 提出的模态区分方法在划分模态时都没有考虑质量变量, 这些模态区分方法都是为了质量无关算法而提出的, 不适用于质量相关算法. 为了划分质量相关的模态, 本文专门为质量相关算法提出了以质量相关 T^2 统计量为标准的质量相关模态区分规则. Hotelling- T^2 统计量用来表征过程变量关系的变化情况, SPE 统计量用来衡量过程测量信息偏离模型的程度, T^2 统计量的变化代表着过程发生了重大变化或系统状态发生了变化^[13]. 如果 T^2 和 SPE 统计量已超过控制限, 则表明此时的生产过程出现异常; 如果生产过程一直处于正常状态, 而 T^2 和 SPE 统计量也已超过各自的控制限, 则表明系统的稳态工作点发生了变化, 即系统从一个正常工作状态转换到另一个正常工作状态^[32], 也就是过程发生了模态的变更. 其中 T^2 统计量为样本的马氏距离, 马氏距离用来确定每个样本的互信息, 不仅测量距离, 而且反映变量的相关性^[33], 因此, 完全可以用 T^2 统计量作为模态转换的衡量标准. 在 PLS 算

法中通常比较关注质量相关部分,因此,可以通过正常过程数据获取质量相关部分的 T^2 或 T_y^2 统计量,通过这些质量相关统计量及控制限的变化进一步划分模态.

在这里有必要阐明质量相关模态区分的必要性.首先模态转变是一个渐变过程^[5],模态转变的实质是系统内部变量间的关系由当前状态转换到另一个状态的过程.尽管变量关系最终可以达到一个平衡的稳态,即完成模态转换,但在转换过程中各个变量间的变化并不是完全同步进行的.也就是质量相关部分模态的转变并不完全和整体模态转变同步,将整体指标划分的模态用于质量相关算法将无法达到较好的效果.因此,有必要提取质量相关指标进行质量相关的模态区分,然后对质量相关模态分别建立监控模型.

2.1 核模糊聚类算法

核模糊聚类(KFCM)算法与KPLS和TKPLS算法相同,都是先将原始非线性数据映射到RKHS中线性化,然后在RKHS利用传统的线性算法对数据进行处理.相比一些传统的非核非模糊聚类算法,核模糊聚类(KFCM)算法有以下优点:1)模糊聚类算法扩展了隶属度的取值范围,有着更好的聚类效果与数据表达能力^[34];2)不需要事先给定聚类中心,只需要给定中心数目,便可自适应地得到聚类中心;3)可以克服非线性问题,对于存在严重非线性的数据依然能够起到很好的聚类效果^[35].

假设 ϕ 为映射函数,规定 c 个聚类中心,过程数据 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T \in \mathbf{R}^{n \times N}$, \mathbf{X} 为标准化后的数据,聚类中心为 $\mathbf{Y}_c = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_c]^T \in \mathbf{R}^{c \times N}$,则RKHS中Euclidean距离可表示为

$$d_F(\mathbf{x}_i, \mathbf{y}_j) = \sqrt{|\phi(\mathbf{x}_i) - \phi(\mathbf{y}_j)|^2} = \sqrt{\phi^T(\mathbf{x}_i)\phi(\mathbf{y}_i) - 2\phi^T(\mathbf{x}_i)\phi(\mathbf{y}_j) + \phi^T(\mathbf{y}_j)\phi(\mathbf{y}_j)} = \sqrt{K(\mathbf{x}_i, \mathbf{x}_i) - 2K(\mathbf{x}_i, \mathbf{y}_j) + K(\mathbf{y}_j, \mathbf{y}_j)}. \quad (11)$$

其中: $K(\cdot)$ 为高斯核函数, $K(\mathbf{x}_i, \mathbf{x}_i) = 1$, $K(\mathbf{y}_j, \mathbf{y}_j) = 1$, $d_F(\mathbf{x}_i, \mathbf{y}_j) = \sqrt{2[1 - K(\mathbf{x}_i, \mathbf{y}_j)]}$.该距离就是KFCM算法的聚类度量标准.定义 $u_j(\mathbf{x}_i)$ 是第 i 个样本属于第 j 类的程度,即隶属度,KFCM的目标函数为

$$J_m(u_j(\mathbf{x}_i), \mathbf{y}_j) = \sum_{j=1}^c \sum_{i=1}^n [u_j(\mathbf{x}_i)]^m d^2(\mathbf{x}_i, \mathbf{y}_j), \quad c \leq n, \quad (12)$$

其中 $m \in (1, \infty)$ 是模糊加权幂指数^[36],用于控制聚类结果模糊程度,通常取 $m = 2$.定义 $\mathbf{U} \in \mathbf{R}^{c \times n}$ 为模糊 c 划分矩阵,其元素为 $u_j(\mathbf{x}_i)$, $i = 1, 2, \dots, n, j = 1, 2, \dots, c$,KFCM算法即为求得使目标函数(12)达

到最小的 \mathbf{U} 和 \mathbf{Y}_c ,其约束条件为各聚类隶属度之和为1,即 $\sum_{j=1}^c u_j(\mathbf{x}_i) = 1$.通过拉格朗日乘数法求得该最优化问题的解,得到聚类中心 \mathbf{y}_j 和隶属度 $u_j(\mathbf{x}_i)$ 的迭代更新公式如下:

$$\mathbf{y}_j = \frac{\sum_{i=1}^n [u_j(\mathbf{x}_i)]^m K(\mathbf{x}_i, \mathbf{y}_j) \mathbf{x}_i}{\sum_{i=1}^n [u_j(\mathbf{x}_i)]^m K(\mathbf{x}_i, \mathbf{y}_j)}, \quad (13)$$

$$u_j(\mathbf{x}_i) = \frac{([1 - K(\mathbf{x}_i, \mathbf{y}_j)])^{-1/(m-1)}}{\sum_{k=1}^c ([1 - K(\mathbf{x}_i, \mathbf{y}_k)])^{-1/(m-1)}}. \quad (14)$$

KFCM算法可自适应迭代求得最终的 \mathbf{U} 和 \mathbf{Y}_c .

2.2 质量相关模态区分规则

质量相关模态区分规则基于KFCM算法.由于工业过程模态随着时间的递进发生变更或过渡,应该通过KFCM算法对过程数据在时间方向上进行聚类.确定聚类数后,KFCM将过程数据的模态进行初步的划分;同时通过非线性PLS算法从过程数据中提取出质量相关指标 T^2 或 T_y^2 ;然后制定规则,通过相邻模态的 T^2 或 T_y^2 统计量及控制限之间的关系,进一步确定质量相关部分的模态转变时刻,合并相邻且相似的模态.

1) 时间约束变量添加.先对过程数据标准化,然后在KFCM算法对建模数据 $\{\mathbf{X}, \mathbf{Y}\}$ 聚类前,需要在数据中加入一列时变变量 t 得到聚类数据 $\{t, \mathbf{X}, \mathbf{Y}\}$, t 的元素为随时间线性递增的数值,数值范围由标准化后的数据波动大小决定. t 的添加使数据在聚类时有一个时间方向上的约束,防止不连续时段数据的误聚类.

2) 聚类数的确定.由系统操作转变次数或过程经历的生物、物理或化学阶段来确定过程的模态个数 p ,相邻模态之间存在过渡过程,因此聚类数确定为 $2p - 1$.

3) 相邻模态精确划分以及合并规则.

Step 1: 将加入时间约束变量的数据进行KFCM聚类,得到 $2p - 1$ 个聚类结果,即 $2p - 1$ 个模态,其中包括过渡模态;然后通过 $2p - 1$ 个聚类结果,确定这 $2p - 1$ 个模态的时间分段,完成模态初步识别.

Step 2: 将质量相关的 T^2 或 T_y^2 统计量 \mathbf{s}_T 按照时间分段分为 $2p - 1$ 个统计量分段,构建出静态统计量窗口 $\mathbf{s}_S(k)$, $k = 1, 2, \dots, 2p - 1$.

Step 3: 求解 $2p - 1$ 个 $\mathbf{s}_S(k)$ 的控制限 $T_{C\mathbf{s}_S(k)}$,即 $T_{C\mathbf{s}_S(k)} = g_{\mathbf{s}_S(k)} \chi_{h_{\mathbf{s}_S(k)}, \alpha}^2$,并令 $k = 0$.为了防止噪声以及小扰动造成区分结果不精确问题的发生,可在

该步骤后采用中值滤波算法通过调整滤波参数,在不改变统计量大体趋势的情况下,对统计量进行中值滤波,去除统计量曲线上对模态区分有影响的毛刺。

Step 4: 令 $k = k + 1$, 如果两个相邻模态统计量 $s_S(k)$ 和 $s_S(k + 1)$ 的控制限满足如下条件:

$$\max(T_{Cs_S(k)}, T_{Cs_S(k+1)}) \leq (\delta + 1) \min(T_{Cs_S(k)}, T_{Cs_S(k+1)}), \quad (15)$$

其中 $\delta \in (0, 0.3)$ 为合并系数, 则合并该相邻模态并返回 Step 4; 否则, 延长相邻两个模态 $s_S(k)$ 和 $s_S(k + 1)$ 中较小的 T_{Cs_S} 控制限, 即延长 $\min(T_{Cs_S(k)}, T_{Cs_S(k+1)})$. 如果延长线与另一个模态中的统计量曲线不相交, 则合并该相邻模态, 返回 Step 4.

Step 5: 如果延长线在另一个模态中与统计量曲线相交, 则交点的横坐标即为质量相关模态转变时刻; 返回 Step 4, 直到所有相邻模态区分完毕, 得到质量相关的模态区分时间标尺。

Step 6: 由于时间约束变量 t 是人为增加的变量, 只是为了聚类时在时间方向上有个约束. t 并不是模型数据, 因此要将聚类数据 $\{t, X, Y\}$ 中的 t 去除, 得到建模数据 $\{X, Y\}$, 然后按照模态区分时间标尺将 $\{X, Y\}$ 划分为分段数据 $\{X, Y\}_k, k = 1, 2, \dots, 2p - 1$. 考虑到模态转换的渐变特性, 可适当在当前模态数据前增加少量滞后的历史数据, 用以克服时变干扰, 然后将分段数据 $\{X, Y\}_k$ 用于建立各模态的监控模型。

2.3 质量相关的多模态故障检测技术及要点分析

完成质量相关模态区分后, 可分别建立各模态的监控模型. 当获得在线数据时, 可通过新数据所属时段匹配所属的监控模型, 也可通过聚类的方式匹配所属的监控模型. 图 1 为基于质量相关模态区分规则的多模态故障检测技术图, 其中质量相关模态区分规则中有两个要点需要进一步详细说明。

1) 添加时间约束变量的必要性. 工业过程复杂化导致同一模态的过程数据并不满足正太分布, 数据的分布特性具有非平稳特性. 模态变更通常随着时间变化, 如果不加入时间约束变量, 则在聚类时部分样本数据会错误聚类到其他模态中. 加入的时间约束变量可作为过程数据的一个新的维度, 通过这一时间维度拉远时间间隔较长样本间的距离, 在时间方向上对聚类形成一定的约束, 可减少误聚类. 然而, 这一约束变量的数值不宜过大, 否则会造成模态在时间方向上均分的误聚类. 为此, 给出时间约束变量 t 的选取方法, 参照图 2. 图 2 中的大量波形为标准化后的数据波形, 取所有数据的最小值及第 1 个样本时刻为时

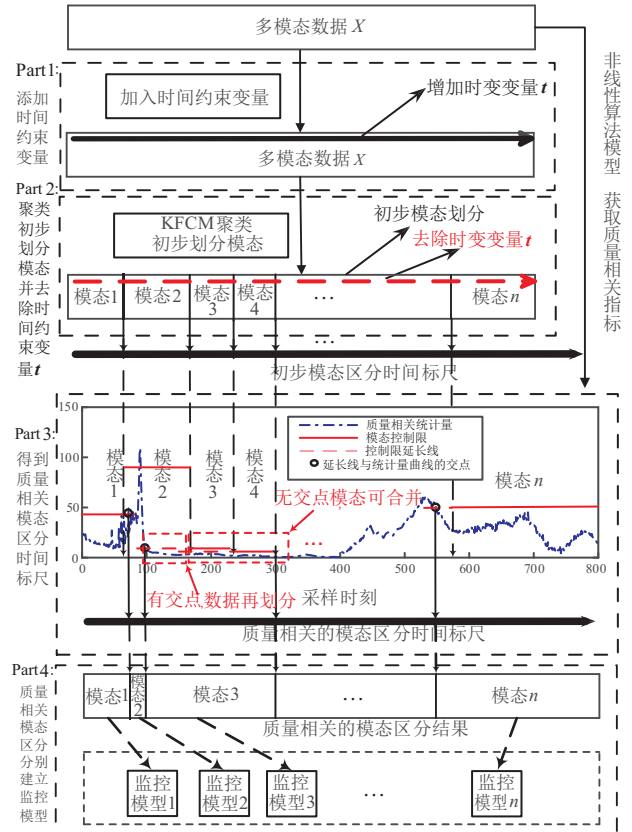


图 1 质量相关的多模态故障检测技术

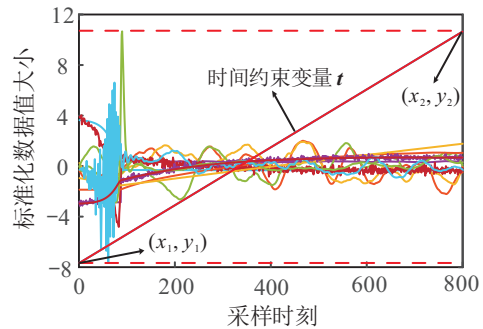


图 2 时间约束变量 t 的选取图

间约束变量 t 的起始坐标 (x_1, y_1) , 取所有数据的最大值和最后一个样本时刻为时间约束变量 t 的终止坐标 (x_2, y_2) , 最终得到数值大小合理的时间约束变量 t .

2) 进一步划分模态及合并模态的必要性. 聚类算法通常具有一定的偶然性, 因此聚类只能提供一个初步的模态划分参考. 为排除偶然性, 可行的办法有两个: 一是进行大量的重复性实验, 取一个平均值; 二是在聚类的基础上, 采用更具有代表性的质量相关 T^2 统计量指标进一步划分模态及合并相邻且相似的模态. 显然第 2 种方法更为精准. 本文给出了两种合并模态的情况, 即模态细化分方法: ① 根据合并系数 (式 (15)) 初步判断模态合并的必要性. 式 (15) 的分子为相邻模态较大的控制限与较小的控制限之差, 分母为较小控制限的数值, 如果这一比值小于一定限度表

明两个模态控制限差别很小,代表着相邻模态的分布大致相似,因此需要合并. ②通过延长相邻模态中较小的控制限来决定相邻模态合并及相邻模态数据的再划分. 如果控制限较大模态的统计量也能够全部位于较小控制限下方,则较小控制限就不会与控制限较大模态中的统计量曲线相交,这说明两种模态分布非常相似. 如果两个相邻模态存在明显的差异,其质量相关统计量曲线必然存在较大的高低差别. 这使得较小控制限的延长线势必会与较大控制限模态统计量曲线相交,交点的横坐标即为质量相关模态更精准的模态转变时刻,可参见图1中的Part3. 较高的质量相关统计量曲线的一部分仍然在较小控制限下方,说明这部分数据理应划分到较小控制限模态中. 因此将这部分满足小控制限的数据合并到小控制限模态区间中,实现相邻模态数据的再划分.

3 广义动态综合控制限

传统的 T^2 控制限仅与保留的主元个数和置信度有关,因而不能满足复杂工业过程监控要求^[37]. 文献^[37]虽然给出了动态数据窗口控制限算法,定义了统计量动态数据窗口的概念,但该方法适用范围有限,在以下两个方面存在不足: 1) 动态数据窗口内仅包括历史滞后数据,仅引入滞后数据会导致当前控制限的更新跟不上统计量的变化,严重时会引起漏报误报; 2) 通过权值 ω 直接将总体的控制限与动态窗口控制限结合在一些工业过程监控中不太适用,例如当总体控制限大于一部分动态控制限,并且同时小于另一部分动态控制限时,将无法通过权值 ω 获得满意的效果. 为了解决问题1), 本文通过历史量来预测超前量,并将超前量一并纳入滑动数据窗口,推广了动态数据窗口的概念. 为了解决问题2), 按照模态区分规则得到的模态区分时间标尺将统计量 s_T 分段,求得分段的控制限 $T_{C s_{S_{new}(b)}}$,然后将不同分段的传统控制限通过权值 ω 与对应分段的动态控制限加权组合,提高了动态综合控制限的应用范围; 最终提出了适用范围更广泛的广义动态综合控制限,简称综合控制限.

3.1 兼顾历史和未来的动态滑动窗口控制限

动态统计量滑动窗口定义如下:

$$s_k = \{\xi_{k-d}, \dots, \xi_{k-1}, \xi_k, \xi_{k+1}, \dots, \xi_{k+l}\}.$$

其中: s_k 为一个长度为 $d+l+1$ 的连续样本数据点的统计量窗口; ξ_k 为对应于离线建模数据中 k 时刻 T^2_{mod} 或 SPE_{mod} 统计量的值;参数 d 为滞后量, l 为超前量,通过两个参数可以调整当前控制限与统计量的匹配程度,在线检测时超前量可通过历史部分进行预

测. k 时刻 T^2 动态滑动窗口控制限定义为

$$T_{DC}(k) = g_{T^2k} \chi^2_{h_{T^2k}, \alpha}. \quad (16)$$

其中 g_{T^2k} 和 h_{T^2k} 可近似取 $g_{T^2} = b_{T^2k}/(2a_{T^2k})$ 和 $h_{T^2k} = 2a^2_{T^2k}/b_{T^2k}$, a_{T^2k} 和 b_{T^2k} 代表 s_{T^2k} 的均值和方差, s_{T^2k} 为 T^2_{mod} 统计量的动态统计量窗口. 同理,定义SPE动态滑动窗口控制限定义为

$$SPE_{DC}(k) = g_{SPEk} \chi^2_{h_{SPEk}, \alpha}. \quad (17)$$

3.2 综合控制限

由质量相关模态区分规则得到的质量相关模态区分时间标尺将整体的统计量 s_T 分段为各个模态的统计量 $s_{S_{new}(b)}$, $b = 1, 2, \dots, B$,其中 B 为划分及合并后的模态个数. 求得每个 $s_{S_{new}(b)}$ 的 T^2 模态控制限为

$$T_{C s_{S_{new}(b)}} = g_{C s_{S_{new}(b)}} \chi^2_{h_{C s_{S_{new}(b)}}, \alpha}.$$

将 $T_{C s_{S_{new}(b)}}$ 与该模态下 T^2 的动态控制限 $T_{DC}(k)$ 通过权值 ω 结合起来,得到动态综合控制限 T_{TC} ,简称为综合控制限,即

$$T_{TC}(k, b) = \omega T_{C s_{S_{new}(b)}} + (1 - \omega) T_{DC}(k). \quad (18)$$

同理,综合动态控制限SPE为

$$SPE_{TC}(k, b) = \omega SPE_{C s_{S_{new}(b)}} + (1 - \omega) SPE_{DC}(k). \quad (19)$$

可通过调整 ω 使综合控制限处于最佳位置,防止噪声以及扰动的干扰,减少误报率,如图3所示.

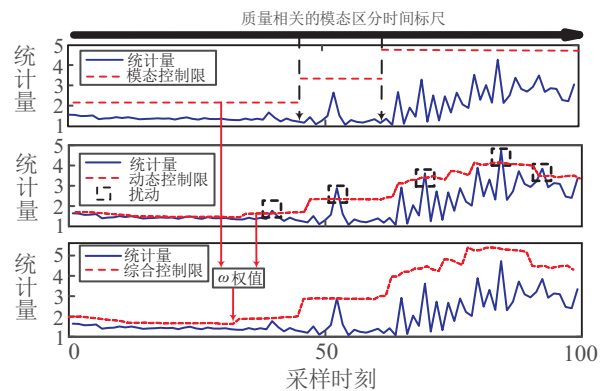


图3 综合控制限

4 青霉素发酵仿真研究

青霉素是通过菌体的生命活动产生的,比一般的化学工业过程复杂.青霉素的生产过程属于半间歇半连续的方式,是一种复杂的且含有较多相互耦合关系变量的化学反应过程,数据具有时变性、非线性、多模态和不确定性^[38],因此本文采用青霉素发酵过程数据对所提出方法的有效性进行验证. 数据来源于Pensim 2.0仿真软件,青霉素发酵过程详细介绍可参见文献^[38].

4.1 仿真实验初始化

青霉素广泛用于医药品中,是一种有效的杀菌抗生素,其质量关系着用药患者的生命安全,因此需要对生产过程中青霉素的质量指标进行严格的监控.青霉素发酵 Pensim 2.0 仿真软件产生的数据中有 18 个变量如表 3 所示.

表 3 Pensim 2.0 仿真数据变量

| 标号 | 变量 | 标号 | 变量 |
|----|--------|----|----------|
| 1 | 采样时间 | 10 | 反应器体积 |
| 2 | 通风速率 | 11 | 排气二氧化碳浓度 |
| 3 | 搅拌速率 | 12 | PH 值 |
| 4 | 底物流加速率 | 13 | 温度 |
| 5 | 补料温度 | 14 | 产生热 |
| 6 | 底物浓度 | 15 | 酸流加速率 |
| 7 | 容解氧浓度 | 16 | 碱流加速率 |
| 8 | 菌体浓度 | 17 | 冷水流加速率 |
| 9 | 产物浓度 | 18 | 热水流加速率 |

本文设定 Pensim 2.0 仿真时间 400 h, 采样时间 0.5 h, 获取 800×18 维的多批次正常数据作为离线建模数据. Pensim 2.0 仿真中共有 3 种故障, 前两种故障检测难度较小, 仅通过调节核函数参数即可实现完美的效果, 而故障 3 检测难度较大, 因此设定故障类型为阶跃型的底物流加速率故障 3, 故障幅值为 -50 , 故障引入时间 200 h, 故障结束时间 400 h, 获得 800×18 维的多批次故障数据作为在线检测数据. 本文通过大量实验排除偶然性, 验证了新方法的有效性. 下面给出实验过程及结果分析.

在 KPLS 和 TKPLS 算法中, 本文选取变量 2 到 7 以及 10 到 14 共 11 个过程变量作为 800×11 维的过程数据 \mathbf{X} , 选取 8 和 9 共 2 个质量变量作为 800×2 维的质量数据 \mathbf{Y} . 由交叉验证与主元个数相关理论算得主元个数 A 为 5, A_y 为 2, A_o 为 3, A_r 为 6. 青霉素发酵过程从生物进化过程来看分为 4 个生理阶段: 反应滞后期、菌体的指数增长期、青霉素合成期和菌体衰亡(自溶)期, 即发酵过程存在 4 个不同阶段, 不同阶段具有不同的模态, 即 p 为 4. 因此在采用 KFCM 算法对 800×14 维变量数据矩阵 $\{t, \mathbf{X}, \mathbf{Y}\}$ 聚类时, 聚类个数确定为 $2p - 1$, 即聚类数为 7, 其中 \mathbf{X} 与 \mathbf{Y} 为标准化数据. 其次采用 800×13 维变量数据矩阵 $\{\mathbf{X}, \mathbf{Y}\}$ 构建整体的 KPLS 和 TKPLS 监控模型. 当 $\{\mathbf{X}, \mathbf{Y}\}$ 分模态后, 采用分段数据 $\{\mathbf{X}, \mathbf{Y}\}_k$ 构建分模态的 KPLS 和 TKPLS 监控模型, $k = 1, 2, \dots, 2p - 1$. 以上 3 种采用核函数的 KFCM、KPLS 和 TKPLS 算法都采用高斯核函数, 其高斯核半径 σ 取值都为 150; 模态区分规则中合并系数 $\delta = 0.3$, 并规定 KFCM 的 m 参数为 2.

4.2 仿真实验

首先, 实验通过 KFCM 对离线建模数据 $\{\mathbf{X}, \mathbf{Y}\}$ 在时间方向上进行聚类, 将数据初步划分为 7 个模态数据 $\{\mathbf{X}, \mathbf{Y}\}_k, k = 1, 2, \dots, 2p - 1$; 其次, 通过离线建模数据 $\{\mathbf{X}, \mathbf{Y}\}$ 建立整体的 KPLS 和 TKPLS 模型, 并按照 7 个划分阶段将建模质量相关的 T^2 和 T_y^2 统计量分段建立模态控制线, 如图 4 和图 5 所示; 然后, 按照模态区分规则得到质量相关的模态区分时标尺, 通过质量相关时间标尺将 $\{\mathbf{X}, \mathbf{Y}\}$ 分为当前标尺规定模态个数的分段数据 $\{\mathbf{X}, \mathbf{Y}\}_k$; 最后, 采用重新划分的分段数据 $\{\mathbf{X}, \mathbf{Y}\}_k$ 建立各个区分模态的监控模型对过程进行监控. 以下实验部分所有控制限的置信度设为 0.99, 其中模态转变时刻非整数取整即可, 每个模态开始可适当增加少量滞后数据, 克服时变扰动.

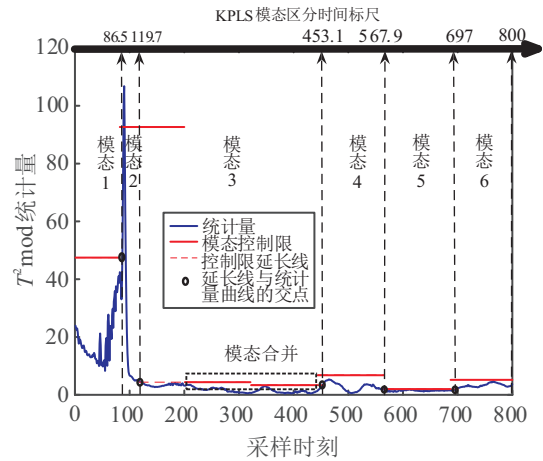


图 4 KPLS 模态区分

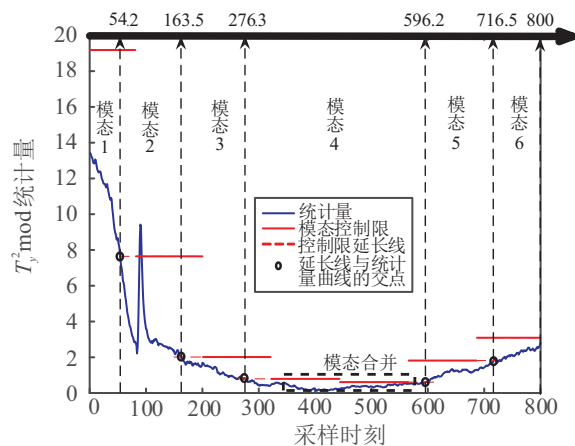


图 5 TKPLS 模态区分

为了验证广义动态综合控制限具有更广的适用性, 将其与文献 [37] 的控制限应用于 KPLS 进行仿真对比. 结果表明, 广义的综合控制在极端条件下依然适用, 如图 6 所示. 由图 6 可看出: 文献 [37] 仅引入滞后信息, 使得控制限跟不上统计量的变化, 引起误报, 而且直接通过权值 ω 将整体控制限引入动态控制限,

抬高了故障区域的检测控制限,引起大量漏报.改进后的广义动态综合控制限有效解决了这些问题,仿真参数如图6标注,之后实验的综合控制限采用相同参数.

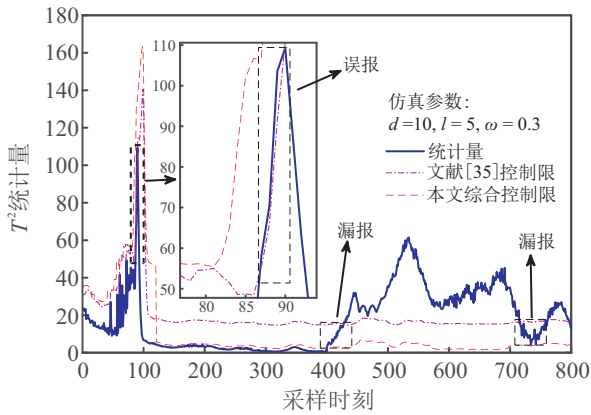


图6 两种控制限故障检测对比结果

本文分别将传统控制限和综合控制限应用到KPLS和TKPLS模型中进行仿真对比,如图7和图8所示.

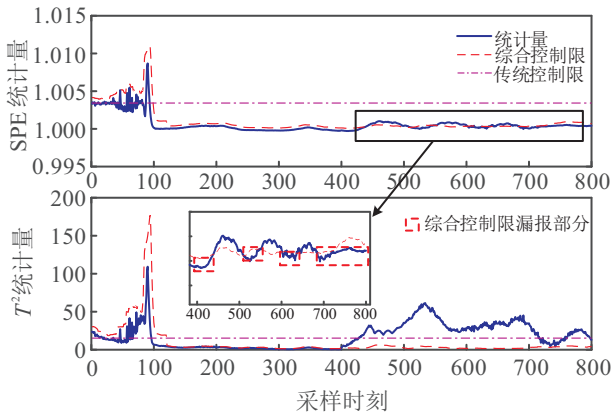


图7 传统、综合控制限KPLS故障检测对比结果

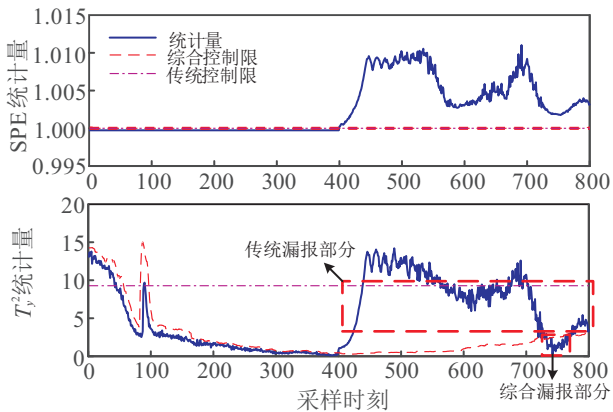


图8 传统、综合控制限TKPLS故障检测对比结果

图7中,传统的“一线”控制限在SPE部分中存在大量误报,甚至全部漏报.在 T^2 部分中存在大量误报和漏报.图8中 T_y^2 部分存在大量误报和漏报.相比于传统的控制限,本文的综合控制限大大减少了误报和漏报.尽管综合控制限有效降低了漏报率和误报率,

但在图7中SPE部分存在较高的故障漏报以及图8中 T_y^2 部分存在少量漏报.分析原因如下:上述实验采用的建模方法都是整体建模,并未考虑青霉素生产过程的多操作以及青霉菌生物进化多阶段造成的多模态问题.因此,基于本文提出的质量相关多模态划分规则分别建立了每个模态的KPLS和TKPLS监控模型,并采用提出的综合控制限进行故障检测,其检测结果如图9和图10所示.

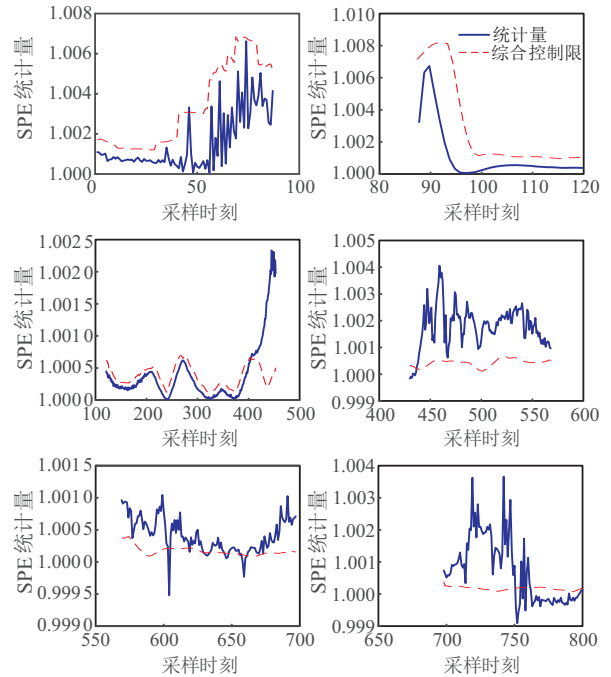


图9 区分模态KPLS的SPE故障检测指标结果

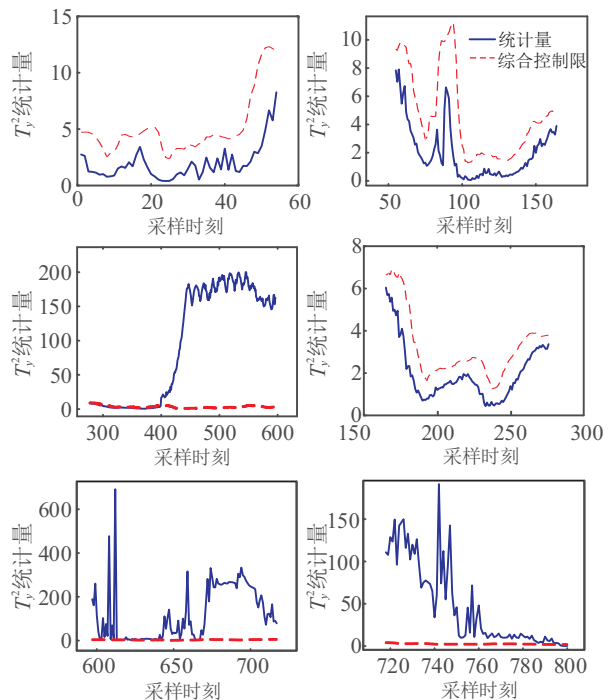


图10 区分模态TKPLS的 T_y^2 故障检测指标结果

对比图7中的SPE统计量部分和图9,可看出图7中SPE统计量部分样本数400~450、525~550以及

600~645大量漏报,甚至在700~800完全漏报;而图9区分模态后,样本数由400~800的故障检测漏报率明显降低.对比图8中 T_y^2 统计量部分和图10,图8中 T_y^2 部分样本数700~800中存在一定的漏报;而图10从样本700~800的漏报率大大降低.可见,区分模态建模方法的故障检测效果明显优于传统的整体建模方法.为了更好地对比这些差异,给出表4和表5的故障检测结果.

表4 KPLS统计量公式和传统控制限 %

| KPLS | T^2 误报 | T^2 漏报 | SPE 误报 | SPE 漏报 |
|-------|----------|----------|--------|--------|
| 传统控制限 | 13.5 | 18.5 | 9.25 | 100 |
| 综合控制限 | 0 | 0 | 0.5 | 60 |
| 分模态监控 | 0 | 0 | 1.25 | 16.25 |

表5 TKPLS统计量公式和传统控制限 %

| TKPLS | T^2 误报 | T^2 漏报 | SPE 误报 | SPE 漏报 |
|-------|----------|----------|--------|--------|
| 传统控制限 | 11.25 | 59.75 | 0 | 0 |
| 综合控制限 | 0 | 9.25 | 0 | 0 |
| 分模态监控 | 0 | 3.5 | 0 | 0 |

对比表4和表5数据可看出,相对于传统控制限,本文提出的广义综合控制限大大降低了漏报率和误报率.相对于整体建立监控模型的方法,基于本文提出的质量相关模态区分规则建立模态监控子模型的方法,更加有效地降低了漏报率和误报率,验证了所提出方法的有效性.

4.3 数值仿真实验

为了验证增加时间约束变量 t 的合理性,本节设计了一个数值仿真实验,从数值计算角度解释了增加约束变量 t 对聚类的影响.仿真模型如下:

$$\begin{cases} \text{mode 1} = \text{normrnd}(0, 5, 2, 100), \\ \text{mode 2} = \text{normrnd}(5.5, 5, 2, 100), \\ \text{data} = [\text{mode 1}^T; \text{mode 2}^T]. \end{cases} \quad (20)$$

其中: $\text{normrnd}(\cdot)$ 为 Matlab 自带的正态分布随机函数,生成了均值为0和5.5,方差都为5的两组2维100个样本的仿真数据,即 mode 1 和 mode 2 . mode 1 为模态1的仿真数据, mode 2 为模态2的仿真数据, data 为未加时间约束变量 t 的原始仿真数据.由于模型数值都是随机生成的,进行了大量数值仿真实验已排除结论的偶然性.时间约束变量 t 生成方式如下:

$$\begin{cases} n = \text{size}(\text{data}, 1), \\ \text{mi} = \min(\min(\text{data})), \\ \text{ma} = \max(\max(\text{data})), \\ \text{step} = (\text{mi} - \text{ma}) / n, \\ \mathbf{t} = [\text{mi} : \text{step} : (\text{ma} - \text{step})]^T. \end{cases} \quad (21)$$

其中: n 为数据 data 的样本数, mi 为 data 中的最小值, ma 为 data 中的最大值, step 为 t 的选取步长, t 为样本数200的列向量.基于 t ,可以构造增加时间约束变量 t 的增广数据矩阵 $\mathbf{t}_{\text{data}} = [t, \text{data}]$,加 t 的数据波形如图11所示.

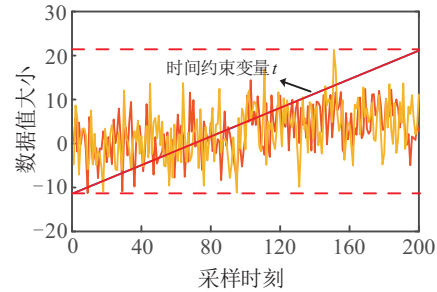


图11 加 t 的数据波形图

原始数据 data 为2维的数据,因此便于生成二维平面分布图,见图12.增广的数据 \mathbf{t}_{data} 为3维数据,变量 t 的增加使得原始数据 data 在第3维坐标方向拉开距离,使得原始的2维数据在3维空间中变得更加容易聚类.为了证明这一结论,给出了不加 t 和加 t 的聚类效果图,如图13~图15所示.其中聚类数为2,高斯核参数为150.图13为原始数据的聚类效果图,与原始数据分布图12对比,可见图13中存在大量误聚类.图14为加 t 的数据聚类3D图的侧面图.对比图14与图12,可见加 t 后的误聚类样本只有2个(在图14中由虚线框圈出).为了更好地展现增加 t 的合理性,给出聚类3D效果图,如图15所示.可见增加了第3维时间变量 t 明显地拉大了两个模态的数据距离,因此聚类效果更佳.

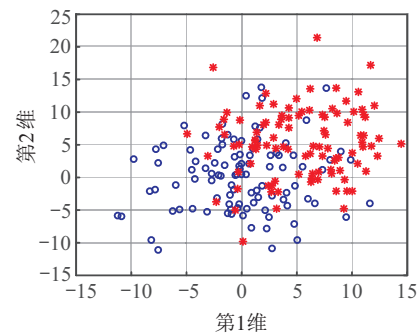


图12 不加 t 的原始数据聚类效果图

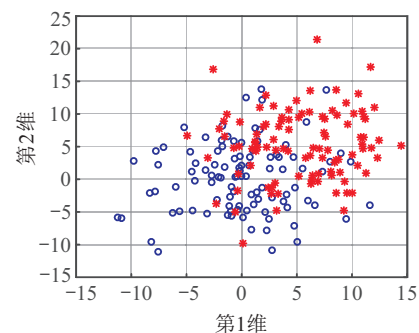
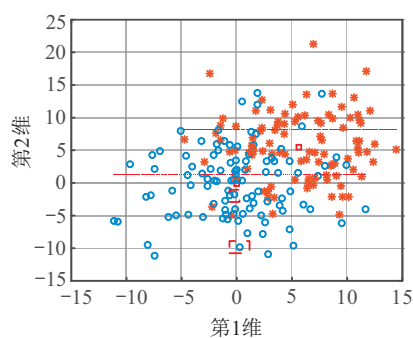
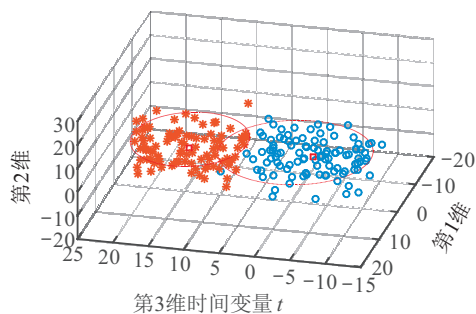


图13 原始数据分布图

图14 加 t 的数据聚类3D效果侧面图图15 加 t 的原始数据聚类3D效果图

5 结论

多模态过程监控是目前工业过程监控中具有挑战性的难题,不但要考虑稳定模态的监控,还要考虑稳定模态间过渡过程的监控.模态变更主要体现为变量间相关关系的变化.而在质量相关领域,质量相关变量相关关系的变化并不完全与整体变量相关关系的变化同步,要区分质量相关模态更是困难.质量相关 T^2 统计量恰恰表征着质量相关变量间的相关关系,这正是本文解决质量相关模态区分难题的突破点.在进行分模态监控时,稳定模态监控采用传统的“一线制”控制限,然而传统控制限并不适用具有较大波动特性的过渡模态,因此考虑采用滑动窗口构造动态控制限.而仅采用历史数据滑动窗口构造的动态控制限在波动较大的过渡模态过程监控中的效果较差,本文考虑引入超前量,并与对应模态的传统控制限进行加权组合,得到广义动态控制限,最终实现较好的监控效果.综合本文提出的质量相关模态区分规则以及广义动态控制限,本文提供了一套完整的质量相关多模态过程监控技术,具有一定实用价值.

参考文献(References)

- [1] Jolliffe I T. Principal component analysis and factor analysis[M]. New York: Springer, 1986.
- [2] Hyvärinen A, Oja E. Independent component analysis: Algorithms and applications[J]. Neural Networks, 2000, 13(4): 411-430.
- [3] Wold S, Sjöström M, Eriksson L. PLS-regression: A basic tool of chemometrics[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 58(2): 109-130.

- [4] 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 C H, Wang F L, Yao Y, et al. Phase-based statistical modeling, online monitoring and quality prediction for batch processes[J]. Acta Automatica Sinica, 2010, 36(3): 366-374.)
- [5] 谭帅, 王福利, 常玉清, 等. 基于差分分段 PCA 的多模态过程故障监测[J]. 自动化学报, 2010, 36(11): 1626-1636.
(Tan S, Wang F L, Chang Y Q, et al. Fault detection of multi-mode process using segmented PCA based on differential transform[J]. Acta Automatica Sinica, 2010, 36(11): 1626-1636.)
- [6] Peng X, Tang Y, Du W, et al. Multimode process monitoring and fault detection: A sparse modeling and dictionary learning method[J]. IEEE Trans on Industrial Electronics, 2017, 64(6): 4866-4875.
- [7] Liu J P, Beyca O F, Rao P K, et al. Dirichlet process gaussian mixture models for real-time monitoring and their application to chemical mechanical planarization[J]. IEEE Trans on Automation Science and Engineering, 2017, 14(1): 208-221.
- [8] Hwang D H, Han C. Real-time monitoring for a process with multiple operating modes[J]. Control Engineering Practice, 1999, 7(7): 891-902.
- [9] Lane S, Martin E B, Kooijmans R, et al. Performance monitoring of a multi-product semi-batch process[J]. J of Process Control, 2001, 11(1): 1-11.
- [10] Kosanovich K A, Dahl K S, Piovoso M J. Improved process understanding using multiway principal component analysis[J]. Industrial & Engineering Chemistry Research, 1996, 35(1): 138-146.
- [11] Chen J, Liu J. Mixture principal component analysis models for process monitoring[J]. Industrial & Engineering Chemistry Research, 1999, 38(4): 1478-1488.
- [12] Choi S W, Park J H, Lee I B. Process monitoring using a gaussian mixture model via principal component analysis and discriminant analysis[J]. Computers & Chemical Engineering, 2004, 28(8): 1377-1387.
- [13] Thissen U, Swierenga H, De Weijer A P, et al. Multivariate statistical process control using mixture modelling[J]. J of Chemometrics, 2005, 19(1): 23-31.
- [14] Yu J, Qin S J. Multimode process monitoring with bayesian inference-based finite gaussian mixture models[J]. AIChE J, 2008, 54(7): 1811-1829.
- [15] 谭帅, 常玉清, 王福利. 基于GMM的多模态过程模态识别与过程监测[J]. 控制与决策, 2015, 30(1): 53-58.
(Tan S, Chang Y Q, Wang F L. Mode identification and process monitoring for multiple mode processes based on GMM[J]. Control and Decision, 2015, 30(1): 53-58.)
- [16] Yu J. Hidden Markov models combining local and global information for nonlinear and multimodal process

- monitoring[J]. *J of Process Control*, 2010, 20(3): 344-359.
- [17] Rashid M M, Yu J. Hidden Markov model based adaptive independent component analysis approach for complex chemical process monitoring and fault detection[J]. *Industrial & Engineering Chemistry Research*, 2012, 51(15): 5506-5514.
- [18] Ning C, Chen M, Zhou D. Hidden markov model-based statistics pattern analysis for multimode process monitoring: An index-switching scheme[J]. *Industrial & Engineering Chemistry Research*, 2014, 53(27): 11084-11095.
- [19] Wang F, Zhu H, Tan S, et al. Orthogonal nonnegative matrix factorization based local hidden markov model for multimode process monitoring[J]. *Chinese J of Chemical Engineering*, 2016, 24(7): 856-860.
- [20] Afzal M S, Tan W, Chen T. Process monitoring for multimodal processes with mode-reachability constraints[J]. *IEEE Trans on Industrial Electronics*, 2017, 64(5): 4325-4335.
- [21] Lu N, Gao F, Wang F. Sub-PCA modeling and on-line monitoring strategy for batch processes[J]. *AIChE J*, 2004, 50(1): 255-259.
- [22] Singhal A, Seborg D E. Pattern matching in historical batch data using PCA[J]. *IEEE Control Systems*, 2002, 22(5): 53-63.
- [23] Zhao C, Wang F, Lu N, et al. Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes[J]. *J of Process Control*, 2007, 17(9): 728-741.
- [24] Yao Y, Gao F. Phase and transition based batch process modeling and online monitoring[J]. *J of Process Control*, 2009, 19(5): 816-826.
- [25] 谭帅, 王福利, 彭俊, 等. 多模态过程中新模态过程建模方法[J]. *控制与决策*, 2012, 27(8): 1241-1245. (Tan S, Wang F L, Peng J, et al. Modeling algorithm for new mode in multi-mode process[J]. *Control and Decision*, 2012, 27(8): 1241-1245.)
- [26] Zhang Y, Li S. Modeling and monitoring between-mode transition of multimodes processes[J]. *IEEE Trans on Industrial Informatics*, 2013, 9(4): 2248-2255.
- [27] 张淑美, 王福利, 谭帅, 等. 多模态过程的全自动离线模态识别方法[J]. *自动化学报*, 2016, 42(1): 60-80. (Zhang S M, Wang F L, Tan S, et al. A fully automatic offline mode identification method for multi-mode processes[J]. *Acta Automatica Sinica*, 2016, 42(1): 60-80.)
- [28] Rosipal R, Trejo L J. Kernel partial least squares regression in reproducing kernel hilbert space[J]. *J of Machine Learning Research*, 2001, 2(12): 97-123.
- [29] Wold S. Cross-validatory estimation of the number of components in factor and principal components models[J]. *Technometrics*, 1978, 20(4): 397-405.
- [30] Peng K, Zhang K, Li G. Quality-related process monitoring based on total kernel PLS model and its industrial application[J]. *Mathematical Problems in Engineering*, 2013, 2013(4): 1-14.
- [31] Zhou D, Li G, Qin S J. Total projection to latent structures for process monitoring[J]. *AIChE J*, 2010, 56(1): 168-178.
- [32] 常玉清, 袁勇, 王福利. 改进块式递推偏最小二乘建模方法及应用[J]. *控制与决策*, 2009, 24(8): 1243-1246. (Chang Y Q, Yuan Y, Wang F L. Improved block-wise recursive partial least square modeling method and its application[J]. *Control and Decision*, 2009, 24(8): 1243-1246.)
- [33] Qin Y, Zhao C, Zhang S, et al. Multimode and multiphase batch processes understanding and monitoring based on between-mode similarity evaluation and multimode discriminative information analysis[J]. *Industrial & Engineering Chemistry Research*, 2017, 56(34): 9679-9690.
- [34] 李侃, 刘玉树. 模糊核聚类的自适应算法[J]. *控制与决策*, 2004, 19(5): 595-597. (Li K, Liu Y. Fuzzy kernel clustering self-adaptive algorithm[J]. *Control and Decision*, 2004, 19(5): 595-597.)
- [35] Khanlari M, Ehsanian M. An improved KFCM clustering method used for multiple fault diagnosis of analog circuits[J]. *Circuits Systems & Signal Processing*, 2017, 36(9): 1-23.
- [36] Huang H C, Chuang Y Y, Chen C S. Multiple kernel fuzzy clustering[J]. *IEEE Trans on Fuzzy Systems*, 2012, 20(1): 120-134.
- [37] 王天真, 刘远, 汤天浩, 等. 基于相对主元分析的动态数据窗口故障检测方法[J]. *电工技术学报*, 2013, 28(1): 142-148. (Wang T Z, Liu Y, Tang T H, et al. Dynamic data window fault detection method based on relative principal component analysis[J]. *Trans on China Electrotechnical Society*, 2013, 28(1): 142-148.)
- [38] Birol G, Ündey C, Cinar A. A modular simulation package for fed-batch fermentation: Penicillin production[J]. *Computers & Chemical Engineering*, 2002, 26(11): 1553-1565.

作者简介

孔祥玉(1967—), 男, 教授, 博士生导师, 从事系统特征提取、自适应信号处理等研究, E-mail: xiangyukong@126.com;

曹泽豪(1993—), 男, 硕士生, 从事系统特征提取、多元统计过程监控等研究, E-mail: 578021174@qq.com;

杜柏阳(1990—), 男, 博士生, 从事多元特征提取、自适应信号处理的研究, E-mail: duboyange@gc@163.com;

罗家宇(1994—), 男, 博士生, 从事多元特征提取、多元统计过程监控的研究, E-mail: 540629964@qq.com.

(责任编辑: 孙艺红)