局部近邻标准化偏最小二乘的多模态间歇过程故障检测

李 元^{1†}, 马雨含¹, 张 成², 冯立伟²

(1. 沈阳化工大学 信息工程学院, 辽宁 沈阳 110142; 2. 沈阳化工大学 数理系, 辽宁 沈阳 110142)

摘要:本文针对多模态间歇过程数据多中心和模态方差差异明显的问题,提出了一种基于局部近邻标准化偏最 小二乘方法.首先,采用统计模量方法处理间歇过程数据,再利用局部近邻标准化方法将统计模量后的训练数据进 行高斯化处理,建立偏最小二乘监控模型,确定控制限;然后,同样对统计模量后的测试数据进行局部近邻标准化处 理,再计算测试数据的高斯偏最小二乘监控指标,进行过程监视及故障检测.最后,通过数值实例和青霉素发酵过 程验证方法有效性.实验结果表明所提方法解决了故障样本近邻集跨模态问题,对多模态数据具有更好的故障检 测能力.

关键词:局部近邻标准化;偏最小二乘;多模态间歇过程;故障检测

引用格式: 李元, 马雨含, 张成, 等. 局部近邻标准化偏最小二乘的多模态间歇过程故障检测. 控制理论与应用, 2020, 37(5): 1109 – 1117

DOI: 10.7641/CTA.2019.80725

Fault detection for multi-modal batch process based on the local neighborhood standardization partial least squares

LI Yuan^{1†}, MA Yu-han¹, ZHANG Cheng², FENG Li-wei²

College of Information Engineering, Shenyang University of Chemical Technology, Shenyang Liaoning 110142, China;
 Department of Science, Shenyang University of Chemical Technology, Shenyang Liaoning 110142, China)

Abstract: In this paper, a local neighborhood standardization partial least squares (LNS-PLS) method is proposed to solve the problem of multi center and the distinctly different modal variance in multi-modal batch process data. Firstly, statistical pattern method is used for batch process data, and the local nearest neighbor standardization (LNS) method is used to transform the training data after statistical pattern into Gaussian distribution. The partial least squares (PLS) model is established and the control limits of T^2 and squared prediction error (SPE) are determined. Next, the LNS standardized is performed on the test data of statistical pattern, and the new Gaussian PLS monitoring indexes are calculated for process monitoring and fault detection. Finally, the effectiveness of the algorithm is verified by the simulation experiment of numerical example and penicillin fermentation process. The results show that the proposed method solves the problem of the neighbor set of fault samples spanning two modes and has better fault detection ability for multi-modal data.

Key words: local neighborhood standardization; partial least squares; multi-modal batch process; fault detection

Citation: LI Yuan, MA Yuhan, ZHANG Cheng, et al. Fault detection for multi-modal batch process based on the local neighborhood standardization partial least squares. *Control Theory & Applications*, 2020, 37(5): 1109 – 1117

1 引言

近年来,工业生产过程的复杂性不断提高,产品与 技术不断更新换代,为了保证生产过程的安全性、可 靠性以及产品质量,及时而有效的故障检测与诊断的 理论的研究具有重要意义.基于数据驱动的多元统计 分析的故障检测与诊断方法已成功应用于化工过程 和生物过程的检测与监视,并逐渐成为工业界和学术 界的研究热点^[1-7]. 偏最小二乘法(partial least squares, PLS)^[8-10]是多 元统计分析方法的一种, PLS算法利用输入对输出的 解释预测作用选取特征向量, 更适合应用于质量相关 的故障检测. 该方法已广泛应用于过程建模、监控和 故障诊断等领域^[11]. Zhou等^[12-14]提出了全潜结构投 影法 (total projection to latent structures, T–PLS), 将 PLS的输入空间分解为4个不同的子空间来提高质量 相关故障和质量无关的故障的检测精度. 但偏最小二

收稿日期: 2018-09-24; 录用日期: 2019-11-19.

[†]通信作者. E-mail: li-yuan@mail.tsinghua.edu.cn; Tel.: +86 13082424115.

本文责任编委: 王郸维.

国家自然科学基金项目(61490701, 61673279)资助.

Supported by the National Natural Science Foundation of China (61490701, 61673279).

乘算法和全潜结构投影法为单模态方法,应用于多模态过程时检测效果不佳.

多模态过程[15]是重要的化工和生物生产过程,具 有多中心、变量非高斯性和非线性等特点.数据的多 模态性使得其故障诊断面临可靠性与准确性低等问 题. Zhao 等^[16-17]提出了基于主元分析(principal component analysis, PCA)算法和偏最小二乘法算法的多 模型方案应用于多模态过程监控. 高斯混合模型[18] (Gaussian mixture model, GMM)在多模态过程监控中 得到了成功的应用.但高斯混合模型的子模型个数事 先未知,难以应用于复杂工业过程中.因此对多模态 过程建立单一监控模型逐渐成为研究热点. 马贺贺 等^[19]提出一种基于马氏距离局部离群因子(local outlier factor, LOF)的方法实现了单个模型对多模态 过程的监控.针对多模态过程的非线性和多中心等特 征, He 等^[20]提出了基于k近邻(k-nearest neighbor, kNN)的故障检测方法. kNN算法^[21] 通过统计局部距 离信息,能够有效降低非线性和多中心的影响,提高 故障检测效率.但两个模态数据方差差异明显时, kNN算法的监控性能明显降低.传统的局部近邻 标准化(local neighborhood standardization, LNS)方 法[22]使用样本近邻集的均值和方差标准化原始数据, 不同模态的数据经过处理之后能够近似满足高斯分 布,解决了各个模态方差差异明显问题,但当故障位 于两个模态之间时,故障样本的近邻会跨越两个模态, 导致无法分离正常样本和故障样本,针对故障样本的 近邻集跨模态问题, Ma等^[22]提出基于局部近邻标准 化策略的主元分析(local neighborhood standardization principal component analysis, LNS-PCA)方法应用 于连续过程故障检测.本文针对间歇过程数据 多中心、模态方差差异明显和故障样本近邻集跨模 态的问题,提出一种基于局部近邻标准化偏最小 二乘 (local neighborhood standardization partial least squares, LNS-PLS)的多模态间歇过程故障检测方法.

注 后文将传统的近邻标准化方法用LNS表示, Ma提出的方法用MLNS表示.

2 基本算法

2.1 PLS算法

PLS算法利用潜变量提取来克服噪声和消除变量 间的相关性,并通过正常生产过程数据准确捕捉质量 变量与过程变量之间的关系^[11-12]. PLS算法具体如 下:

对m个过程变量进行n次独立采样,得到输入矩 阵 $X \in \mathbb{R}^{n \times m}$,同时对p个质量变量进行n次独立采样 得到输出矩阵 $Y \in \mathbb{R}^{n \times p}$. PLS对X和Y空间的分解如 下^[23]:

$$\begin{cases} X = TP^{\mathrm{T}} + E, \\ Y = UQ^{\mathrm{T}} + F, \end{cases}$$
(1)

其 中: $T = [t_1 \cdots t_d] \in \mathbb{R}^{n \times d}$ 和 $P = [p_1 \cdots p_d] \in \mathbb{R}^{m \times d}$ 分别是X的得分矩阵和负载矩阵, *E*是X的残 差矩阵, $U = [u_1 \cdots u_d] \in \mathbb{R}^{n \times d}$ 和 $Q = [q_1 \cdots q_d] \in \mathbb{R}^{p \times d}$ 分别是Y的得分矩阵和负载矩阵, *F*是Y的残差 矩阵.

PLS迭代算法^[23]的目标函数如下:

 $\begin{cases} J_{\text{PLS}}(w_i, c_i) = \max \operatorname{cov}(t_i, u_i) = \max w_i^{\text{T}} X^{\text{T}} Y c_i, \\ \text{s.t. } \|w_i\| = 1, \ \|c_i\| = 1, \end{cases}$

其中: $w_i n c_i$ 是投影向量, $t_i = X w_i$, $u_i = Y c_i$. 由于 得分矩阵T无法由 $W = [w_1 \cdots w_d] n X$ 直接得出, 定义 $R = [r_1 \cdots r_d]$, $r_1 = w_1$, d为主元个数. 对于 i > 1的情况有

$$r_i = \prod_{j=1}^{i-1} (I_m - w_j p_j^{\mathrm{T}}) w_i.$$
(3)

得分矩阵T获得如下:

$$T = XR.$$
 (4)

R, P和W有如下关系:

$$R = W(P^{\mathrm{T}}W)^{-1},$$

$$P^{\mathrm{T}}R = R^{\mathrm{T}}P = I_{d}.$$
(5)

PLS算法对输入X分解如下:

$$\begin{cases} x = \hat{x} + \tilde{x}, \\ \hat{x} = PR^{\mathrm{T}}x, \\ \tilde{x} = (I - PR^{\mathrm{T}})x. \end{cases}$$
(6)

PLS模型通常用平方预测误差(squared prediction error, SPE)和Hotelling的统计量 T^2 来监控过程是否发生异常. T^2 统计量和SPE统计量的计算方法如下:

$$T^{2} = t^{\mathrm{T}} \Lambda^{-1} t,$$

SPE = $\|\tilde{x}\|^{2},$ (7)

其中 $\Lambda = \frac{1}{n-1}T^{T}T$.可采用核密度估计^[24]等方法确定控制限.

2.2 PLS方法的不足

采用PLS算法进行故障检测时,过程数据需要在 检测之前进行标准化处理.目前常用的是z-score标 准化方法^[12,23]. PLS算法的基本假设为数据具有多元 高斯分布,而多模态数据z-score标准化后仍具有多 模态特征,不满足该假设.设计一个数值例子来解释 这种情况:模态1: x_1 服从 $N(5,0.5), x_2$ 服从 $N(10, 0.7), y=10x_1-5x_2;模态2: x_1$ 服从 $N(20,2.5), x_2$ 服 从 $N(10,2.2), y=10x_1-5x_2$.在两个模态间设置一 个故障点(11.5, 10). 图1中z-score标准化后数据仍具 有多模态特征,故障样本位于两个模态之间. 图2中 PLS算法没有检测出故障,说明原始数据具有多模态 结构时PLS算法存在不足.







Fig. 2 Fault detection results of PLS algorithm

2.3 MLNS方法

多模态数据集具有两个特征,一是各模态数据中 心不重合,二是数据离散程度不同,即各模态方差不 同.采用局部近邻标准化(LNS)^[22]方法可以解决这两 个问题.LNS标准化和z-score标准化的主要区别是 LNS方法使用用每个样本局部近邻集的均值和方差, 而z-score方法使用全部样本集的均值和方差.但 LNS方法无法解决故障样本近邻集跨模态问题,文献 [22]提出的MLNS方法解决了这一问题.两种方法具 体过程如下:

找到每个样本的k近邻,并计算每个样本到其各个 近邻的欧式距离. 假设 x_i 是X的一个样本, x_i^k 表示 x_i 的第k个近邻, $d(x_i, x_i^k)$ 是 x_i 和 x_i^k 之间的欧氏距离, $NN(x_i) = \{x_i^1, x_i^2, \dots, x_i^k\}$ 为样本 x_i 的近邻集. 样本 x_i 可以按以下方法两种方法标准化:

LNS方法:

$$Z_i = \frac{x_i - m(NN(x_i))}{s(NN(x_i))},\tag{8}$$

其中 $m(NN(x_i))$ 和 $s(NN(x_i))$ 分别是 x_i 的近邻集 $NN(x_i)$ 的均值和方差.

MLNS方法:

$$Z_{i} = \frac{x_{i} - m(NN(x_{i}^{1}))}{s(NN(x_{i}^{1}))},$$
(9)

其中 $m(NN(x_i^1))$ 和 $s(NN(x_i^1))$ 分别是 x_i 的第一近邻 x_i^1 的近邻集 $NN(x_i^1)$ 的均值和方差.

LNS方法无法解决故障样本近邻集跨模态问题. 采用第1.2节例子验证LNS方法和MLNS方法的效果. 图3(a)中LNS方法故障样本的近邻集跨越两个模态, 此时故障样本的近邻集方差较大,由式(8)可知,故障 样本标准化后的数值接近正常样本的模态中心,导致 故障样本和正常样本无法分离.由图3(b)可以清晰 地看出故障样本湮没在正常样本之中.图3(a)中 MLNS方法故障样本的第一近邻的近邻集位于一个模 态.图3(c)中MLNS方法标准化后正常样本离散程度 基本一致,且与故障样本完全分离.由图3(d)可知,标 准化后的变量近似服从零均值的高斯分布,这是因为 近邻均值总是非常接近变量值,变量x1减去其近邻均 值后在零附近.





图 3 LNS和MLNS方法效果 Fig. 3 Effect of LNS and MLNS methods

3 基于MLNS-PLS的多模态间歇过程故障 检测

本文采用统计模量^[25-26]方法处理间歇过程数据. 统计模量方法是用统计特征矩阵重新定义原始间歇 过程数据,一般所用的统计特征为均值和方差.PLS算 法应用于多模态间歇过程故障检测时效果不佳.LNS 方法解决了这一问题,但仍存在缺陷,当故障样本的 近邻集跨模态时,该方法无法分离正常样本和故障样 本.本文使用MLNS方法处理数据,提出了一种基于 局部近邻标准化偏最小二乘(MLNS-PLS)方法改进 PLS算法处理多模态间歇过程数据的不足.MLNS标 准化后变量近似服从零均值的高斯分布,符合PLS算 法基本假设.基于MLNS-PLS方法的故障检测包括 离线建模和在线检测两部分,具体步骤如下:

离线建模:

 在训练集X中寻找每个样本x_i的k个局部近邻 集NN(x_i),在训练集Y中寻找每个样本y_i的k个局部 近邻集NN(y_i);

 $m(NN(x_i^1))$ 和方差 $s(NN(x_i^1))$ 来标准化 x_i ,同理得 到标准化后的 y_i ;

3) 在MLNS处理后的数据集上,利用式(3)-(4)确 定PLS的得分矩阵T,利用式(6)计算残差矩阵;

4) 计算每个样本的SPE统计量和T²统计量及控 制限.

在线检测:

1) 对于待检测样本 x_j 和 y_j ,分别在训练集X和Y中找到 x_j 和 y_j 的第一近邻 x_i^1 和 y_i^1 ;

2) 用 x_j^1 的近邻集的均值 $m(NN(x_j^1))$ 和方差 $s(NN(x_i^1))$ 来标准化 x_i ,同理得到标准化后的 y_i ;

3)利用负载矩阵将MLNS方法标准化后的数据投影到PLS的主元空间和残差空间;

4) 计算待检测样本 x_j 和 y_j 的SPE统计量和 T^2 统 计量,并与控制限比较确定检测结果.

4 实验应用研究

4.1 数值实例的分析与实验

使用多模态数值实例分析MLNS-PLS方法的检测性能,同时与PLS算法、*k*NN算法和LNS-PLS方法的结果进行对比分析.所用模型如下:

模态1:

$$\begin{cases} x_1 \sim N(1, 0.2), \\ x_2 \sim N(1, 0.1), \\ x_3 \sim N(1, 0.05), \\ y_1 = 10x_1 - 5x_2 - x_3, \\ y_2 = 2x_1 - 5x_2 - 3x_3. \end{cases}$$
(10)

模态 2:

$$\begin{cases} x_1 \sim N(-0.7, 1), \\ x_2 \sim N(-0.7, 0.5), \\ x_3 \sim N(-0.7, 0.5), \\ y_1 = 10x_1 - 5x_2 - x_3, \\ y_2 = 2x_1 - 5x_2 - 3x_3, \end{cases}$$
(11)

其中: x_1, x_2, x_3 是输入X的3个变量, y_1, y_2 是输出Y 的两个变量.取500组训练数据,两个模态各250组; 取10组校验数据,两个模态各5组;在3个变量上设置 偏移,产生5个故障点 $f_1(-3, -2, -3), f_2(0, 0.5, 0.5),$ $f_3(0.7, 0.1, 0.3), f_4(0.5, 1.2, 1.3), f_5(1, 0.5, 1).$

检测过程中主元数为2, kNN算法的近邻数为3, LNS-PLS方法和MLNS-PLS方法的近邻数为10. 仿 真结果如图4-9所示. 图中黑色实线是置信度为99% 的控制限. 原始数据所有变量分布见图4, 图中两个模 态数据方差差异明显, 模态1数据分布密集, 模态2数 据分布较为稀疏, 故障1靠近模态2, 故障2和故障3在 两个模态之间, 故障4和故障5靠近模态1.



Fig. 5 Data scatter mapping



图 6 z-score-PLS算法检测结果

Fig. 6 Fault detection result of z-score-PLS algorithm



图 7 kNN算法检测结果

Fig. 7 Fault detection result of kNN algorithm







图 9 MLNS-PLS算法检测结果 Fig. 9 Fault detection result of MLNS-PLS algorithm

图6中PLS算法未检测出故障,且正常样本的统计 量差异明显.图7中kNN算法检测出故障1和故障2,当 两个模态疏密程度不一致时,两个模态样本的D²统计 量差异明显,此时控制限由稀疏模态决定,故障4和故 障5为靠近密集模态的微弱故障,无法被检测.

图8中LNS-PLS方法未检测出故障2和故障3.因为故障2和故障3的近邻集跨越两个模态,以故障2为例,它的前两个近邻属于第二模态,而第三近邻属于第1模态,使得故障2的近邻集方差较大,由式(8)可知,标准化后的故障2靠近正常样本中心,导致故障2和正常样本无法分离.从图5(a)中可以看出LNS方法标准化后故障2和故障3位于正常样本中心周围.

图5(b)清楚的表明MLNS标准化消除了数据的模态结构信息,故障样本和正常样本完全分离.因此图9中,综合T²和SPE两个指标,MLNS-PLS方法检测出所有故障.

4.2 青霉素过程的故障检测

青霉素生产过程是一个典型的非线性、多模态间歇生产过程.其发酵过程可分3个阶段:菌体快速生长阶段、菌体合成青霉素阶段和菌体自溶阶段.本文基于Pensim仿真平台^[27-28],验证基于MLNS-PLS故障检测方法的有效性.青霉素发酵过程流程如图10所示.

Pensim仿真平台有5个输入变量可以控制发酵过 程参数变化,9个过程变量是菌体合成及生长中产生 的,5个质量变量影响青霉素的产量.本文选取7个过 程变量作为输入和5个质量变量作为输出,变量选 取^[8]见表1.本文使用Pensim仿真平台生成数据,仿真 时间设定为400h,采样时间设定为1h.在正常工况下 生成两个模态数据:模态1初始CO₂浓度设定为0.5, 通过调整初始变量值生产50个批次,为稀疏模态;模 态2初始CO₂浓度设定为1,通过微调初始变量值生 产75个批次,为密集模态.从两个模态中各随机选 取5个批次作为校验样本,剩余115个批次作为训练样本.图11中可看出正常批次的CO₂浓度分为两个模态.



图 10 青霉素发酵过程流程图 Fig. 10 Flow chart of penicillin fermentation process

表1 青霉素发酵过程变量选取

| Table 1 | Variable selecti | on of penicillin | fermentation |
|---------|------------------|------------------|--------------|
| | process | | |

| | • | | |
|----|-----------------------------|--------------|------|
| 序号 | 变量(单位) | 变量范围 | 变量类型 |
| 1 | 通气率(L/h) | 3~10 | 过程变量 |
| 2 | 搅拌功率(W) | $20 \sim 50$ | 过程变量 |
| 3 | 底物流速率(L/h) | 0.035~0.045 | 过程变量 |
| 4 | 底物流加温度(K) | 296~298 | 过程变量 |
| 5 | 培养基容量(L) | 100~190 | 过程变量 |
| 6 | CO ₂ 浓度(mmole/L) | 0.5~1.0 | 过程变量 |
| 7 | 发酵反应器温度(K) | 298~300 | 过程变量 |
| 8 | 底物浓度(g/L) | $5 \sim 50$ | 质量变量 |
| 9 | 容氧浓度(mmole/L) | 1.16 | 质量变量 |
| 10 | 菌体浓度(g/h) | $0 \sim 0.2$ | 质量变量 |
| 11 | 青霉素浓度(g/h) | 默认 | 质量变量 |
| 12 | 反应产生的热量(kcal/h) | 默认 | 质量变量 |
| | | | |



Fig. 11 CO₂ concentration of normal samples

Pensim仿真平台能对前3个变量(通气率、搅拌功 率和底物流速率)引入故障,故障类型有阶跃和斜坡两 种,并可进一步设定两种故障的幅度、引入时间和终 止时间.为了测试方法的有效性,本文在实验中生产12个故障批次,12个故障批次的故障类型及程度见表2.

表 2 青霉素发酵过程中设置的12个批次故障

 Table 2
 Twelve faults of penicillin fermentation process

| 故障批次 | 故障变量 | 故障类型 | 故障尺度 | 初始CO ₂ 浓度 |
|------|------|------|------|----------------------|
| f1 | 1 | 阶跃 | 15% | 0.5 |
| f2 | 1 | 阶跃 | -30% | 0.5 |
| f3 | 1 | 斜坡 | 3 | 0.5 |
| f4 | 1 | 斜坡 | -3 | 1 |
| f5 | 2 | 阶跃 | -30% | 0.5 |
| f6 | 2 | 阶跃 | 30% | 1 |
| f7 | 2 | 阶跃 | 45% | 1 |
| f8 | 2 | 阶跃 | 50% | 1 |
| f9 | 2 | 斜坡 | -30 | 1 |
| f10 | 3 | 阶跃 | 13% | 0.5 |
| f11 | 3 | 阶跃 | 6% | 1 |
| f12 | 3 | 阶跃 | 20% | 1 |
| | | | | |

青霉素发酵过程数据是间歇过程三维数据矩阵,本文采用统计模量方法将数据处理成二维矩阵.使用PLS算法、*k*NN算法、LNS-PLS方法和MLNS-PLS方法对青霉素发酵过程进行仿真实验和对比分析.实验中PLS算法、LNS-PLS方法和MLNS-PLS方法的主元数取3.*k*NN算法的近邻数取10,标准化方法为z-score方法.LNS-PLS方法和MLNS-PLS方法的近邻数取15.实验结果见图12-16和表3.



Fig. 12 Data scatter mapping





图 13 z-score-PLS算法检测结果

Fig. 13 Fault detection result of z-score-PLS algorithm



图 14 kNN算法检测结果

Fig. 14 Fault detection result of kNN algorithm



图 15 LNS-PLS算法检测结果







Fig. 16 Fault detection result of MLNS-PLS algorithm

表 3 12个故障批次的检测结果 Table 3 Fault detection results of 12 batches

| お暗 | PLS | | ŀNN | LNS-PLS | | MLNS-PLS | |
|-------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| HXIT | T^2 | SPE | WININ | T^2 | SPE | T^2 | SPE |
| f1 | × | \checkmark | \checkmark | × | \checkmark | × | \checkmark |
| f2 | \checkmark | \checkmark | \checkmark | × | \checkmark | × | \checkmark |
| f3 | \checkmark | \checkmark | \checkmark | × | \checkmark | × | \checkmark |
| f4 | × | × | \checkmark | × | \checkmark | × | \checkmark |
| f5 | \checkmark | \checkmark | \checkmark | × | \checkmark | \checkmark | \checkmark |
| f6 | × | × | \checkmark | × | × | \checkmark | \checkmark |
| f7 | × | × | × | × | × | \checkmark | \checkmark |
| f8 | × | × | \checkmark | × | × | \checkmark | \checkmark |
| f9 | \checkmark |
| f10 | × | \checkmark | \checkmark | × | \checkmark | \checkmark | \checkmark |
| f11 | × | × | × | \checkmark | \checkmark | \checkmark | \checkmark |
| f12 | × | × | × | \checkmark | \checkmark | \checkmark | \checkmark |
| 故障检测率 | 0.50 | | 0.75 | 0.75 | | 1.00 | |

图12中正常样本分为两个模态,并且两个模态的 数据离散程度不同.图13中PLS算法的 T^2 和SPE共检 测出6个故障,多模态数据不服从多元高斯分布,故 PLS算法检测效果不佳. 图14中kNN 算法未检测出故 障7、故障11和故障12.因为当两个模态数据离散程 度不同时, kNN 算法的控制限 D^2_{α} 由方差大的第1模态 决定,故障11为靠近密集模态的微弱故障(见图12),它 的统计量 D^2 在两个模态正常样本的统计量之间,无法 被检测. 故障7和故障12同理. 图15中LNS-PLS方法 未检出故障6、故障7和故障8,因为这3个故障样本的 近邻集跨越两个模态. 以故障7为例, 其第1近邻和第2 近邻分别是第9号和27号样本都属于第1模态,而其第 3近邻为第104号样本属于第2模态.此时故障7的近邻 集方差较大,由式(8)可知,标准化后的故障7接近正常 样本的模态中心,导致故障7和正常样本无法分离,故 障6和故障8同理.图16中MLNS-PLS方法检测出全 部故障,且两个模态正常样本的统计量差异不大.综 上所述, MLNS-PLS方法解决了两个模态数据方差 差异明显和故障样本近邻集跨模态的问题.

5 结论

针对PLS算法无法有效应用于多模态间歇过程故

障检测的问题,本文提出一种局部近邻标准化偏最小 二乘(MLNS-PLS)方法.先利用MLNS方法处理统计 模量后的数据,避免了故障样本的近邻集跨模态的情况;再利用PLS算法进行故障检测,改善了PLS对多模 态间歇过程数据的检测效果.将MLNS-PLS对多模 态间歇过程数据的检测效果.将MLNS-PLS方法应 用于青霉素发酵过程的故障检测,实验结果表明, MLNS-PLS方法能保证较好的检测效果,该方法可 以应用到其他工业和生物领域的多模态间歇过程的 故障检测中.

参考文献:

 ZHOU Donghua, LI Gang, LI Yuan. Data-driven Industrial Process Fault Detection and Diagnosis Technology. Beijing: Science Press, 2011: 1 – 76.

(周东华,李钢,李元.数据驱动的工业过程故障检测与诊断技术.北 京:科学出版社,2011:1-76.)

- [2] LI Y, ZHANG X M. Diffusion maps based k-nearest-neighbor rule technique for semiconductor manufacturing process fault detection. *Chemometrics and Intelligent Laboratory Systems*, 2014, 136(15): 47 – 57.
- [3] WANG Guozhu, LIU Jianchang, LI Yuan, et al. Fault diagnosis of industrial processes based on weighted k-nearest neighbor reconstruction analysis. *Control Theory & Applications*, 2015, 32(7): 873 – 880. (王国柱,刘建昌, 李元,等. 加权k最近邻重构分析的工业过程故障 诊断. 控制理论与应用, 2015, 32(7): 873 – 880.)
- [4] DING S X. Data-driven design of monitoring and diagnosis systems for dynamic processes: a review of subspace technique based schemes and some recent results. *Journal of Process Control*, 2014, 24(2): 431 – 449.
- [5] GE W, WANG J, ZHOU J, et al. Incipient fault detection based on fault extraction and residual evaluation. *Industrial & Engineering Chemistry Research*, 2015, 54(14): 3664 – 3677.
- [6] LI Yuan, LIU Yadong, ZHANG Cheng, et al. Fault detection of batch process based on diffusion k-nearest neighbors distance. Control Theory & Applications, 2015, 32(12): 1653 1659.
 (李元,刘亚东,张成,等.基于扩散k近邻距离的间歇过程故障诊断. 控制理论与应用, 2015, 32(12): 1653 1659.)
- [7] LI Yuan, MA Yuhan, GUO Jinyu. On-line fault detection based on dynamic multiway local outlier factor. *Application Research of Computers*, 2017, 34(11): 3259 – 3261, 3266.
 (李元,马雨含,郭金玉.基于动态多向局部离群因子的在线故障检 测. 计算机应用研究, 2017, 34(11): 3259 – 3261, 3266.)
- [8] ZHONG B, WANG J, ZHOU J, et al. Quality-related statistical process monitoring method based on global and local partial leastsquares projection. *Industrial & Engineering Chemistry Research*, 2016, 55(6): 1609 – 1622.
- [9] HIRAI T, KANO M. Adaptive virtual metrology design for semiconductor dry etching process through locally weighted partial least squares. *IEEE Transactions on Semiconductor Manufacturing*, 2015, 28(2): 137 – 144.
- [10] ZHAO Xiaoqiang, XUE Yongfei. Fault detect algorithm of chemical process based on kernel T-PLS. *CIESC Journal*, 2013, 64(12): 4608 4614.
 (赵小强,薛永飞. 基于核T-PLS的化工过程故障检测算法. 化工学 报, 2013, 64(12): 4608 4614.)
- [11] SHI Huaitao, LIU Jianchang, ZHANG Yu, et al. Fault detection method based on relative-transformation partial least squares. *Chinese Journal of Scientific Instrument*, 2012, 33(4): 816-822. (石怀涛, 刘建昌, 张羽, 等. 基于相对变换PLS的故障检测方法. 仪 器仪表学报, 2012, 33(4): 816-822.)

- [12] ZHOU D, LI G, QIN S J. Total projection to latent structures for process monitoring. *AiChE Journal*, 2010, 56(1): 168 – 178.
- [13] LI G, ALCALA C F, QIN S J, et al. Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the tennessee eastman process. *IEEE Transactions on Control Systems Technology*, 2011, 19(5): 1114 – 1127.
- [14] LI G, QIN S J, ZHOU D. Output relevant fault reconstruction and fault subspace extraction in total projection to latent structures models. *Industrial & Engineering Chemistry Research*, 2010, 49(19): 9175 – 9183.
- [15] KOURTI T, MACGREGOR J F. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics and Inteligentl Laboratory Systems*, 1995, 28(1): 3 – 21.
- [16] ZHAO S J, ZHANG J, XU Y M. Monitoring of processes with multiple operating modes through multiple principle component analysis models. *Industrial & Engineering Chemistry Research*, 2004, 43(22): 7025 – 7035.
- [17] ZHAO S J, ZHANG J, XU Y M. Performance monitoring of processes with multiple operating modes through multiple PLS models. *Journal of Process Control*, 2006, 16(7): 763 – 772.
- [18] YU J, QIN S J. Multiway Gaussian mixture model based multiphase batch process monitoring. *Industrial & Engineering Chemistry Research*, 2009, 48(18): 8585 – 8594.
- [19] MA Hehe, HU Yi, SHI Hongbo. Fault detection of complex chemical processes using Mahalanobis distance-based local outlier factor. *CI-ESC Journal*, 2013, 64(5): 1674 – 1682. (马贺贺, 胡益, 侍洪波. 基于马氏距离局部离群因子方法的复杂化 工过程故障检测. 化工学报, 2013, 64(5): 1674 – 1682.)
- [20] HE Q P, WANG J. Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, 2007, 20(4): 345 – 354.
- [21] FENG Liwei, ZHANG Cheng, LI Yuan, et al. Fault detection strategy of standard-distance-based k nearest neighbor rule in multimode processes. *Control Theory & Applications*, 2019, 36(4): 553 560.
 (冯立伟, 张成, 李元, 等. 基于标准距离k近邻的多模态过程故障检测策略. 控制理论与应用,2019, 36(4): 553 560.)

- [22] MA H, HU Y, SHI H. A novel local neighborhood standardization strategy and its application in fault detection of multimode processes. *Chemometrics & Intelligent Laboratory Systems*, 2012, 118(7): 287 – 300.
- [23] LI G, QIN S J, ZHOU D. Geometric properties of partial least squares for process monitoring. *Automatica*, 2010, 46(1): 204 – 210.
- [24] CHEN Q, WYNNE R J, GOULDING P, et al. The application of principal component analysis and kernel density estimation to enhance process monitoring. *Control Engineering Practice*, 2000, 8(5): 531 – 543.
- [25] WANG J, HE Q P. Multivariate statistical process monitoring based on statistics pattern analysis. *Industrial & Engineering Chemistry Research*, 2010, 49(17): 7858 – 7869.
- [26] HE Q P, WANG J. Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes. *Aiche Journal*, 2015, 57(1): 107 – 121.
- [27] MORI J, YU J. Quality relevant nonlinear batch process performance monitoring using a kernel based multiway non-Gaussian latent subspace projection approach. *Journal of Process Control*, 2014, 24(1): 57 – 71.
- [28] GE Z Q, SONG Z H. Online monitoring and quality prediction of multiphase batch processes with uneven length problem. *Industrial* & *Engineering Chemistry Research*, 2014, 53(2): 800 – 811.
- 作者简介:

李元 教授,博士生导师,目前研究方向为基于数据驱动的复杂 工业过程故障诊断, E-mail: li-yuan@mail.tsinghua.edu.cn;

马雨含 硕士研究生,目前研究方向为基于数据驱动的复杂工业

过程故障诊断, E-mail: 2483230405@qq.com;

张 成 副教授,博士研究生,目前研究方向为基于数据驱动的复 杂工业过程故障诊断,E-mail: zcgg_2005@126.com;

冯立伟 讲师,硕士,目前研究方向为基于数据驱动的复杂工业过程故障诊断, E-mail: feng-li-wei@163.com.