

基于时空近邻标准化和局部离群因子的复杂过程故障检测

冯立伟^{1,2}, 李元^{2†}, 张成^{1,2}, 谢彦红^{1,2}

(1. 沈阳化工大学 数理系, 辽宁 沈阳 110142;

2. 沈阳化工大学 技术过程故障诊断与安全性研究中心, 辽宁 沈阳 110142)

摘要: 针对复杂过程数据的非线性、动态性和中心漂移等特征, 提出了基于时空近邻标准化和局部离群因子的故障检测方法(TSNS-LOF). 首先使用训练样本在时空两个方向上的近邻集来标准化训练样本; 然后在标准样本集上计算样本的局部离群因子, 并确定其上分位点作为检测控制限, 进行在线故障检测. 时空近邻标准化解决了复杂过程数据的非线性、动态性和中心漂移的问题; 局部离群因子通过度量样本的相似度实现了故障样本和正常样本的分离. 将TSNS-LOF应用于田纳西-伊斯曼过程(TE)过程进行故障检测实验, 结果表明相对于主元分析、动态主元分析、 k 近邻、局部离群因子等方法, TSNS-LOF对故障预警更加及时且具有更高的故障检测率. 理论分析和仿真实验说明TSNS-LOF方法适用于具有动态性或多模态特性或两者兼具的过程故障检测, 能够更好地保障生产过程的安全性和产品的高质量.

关键词: 时空近邻标准化; 局部离群因子; 模型; 主元分析; 过程控制

引用格式: 冯立伟, 李元, 张成, 等. 基于时空近邻标准化和局部离群因子的复杂过程故障检测. 控制理论与应用, 2020, 37(3): 651 – 657

DOI: 10.7641/CTA.2019.90007

Time-space neighborhood standardization-local outlier factor based fault detection for complex process

FENG Li-wei^{1,2}, LI Yuan^{2†}, ZHANG Cheng^{1,2}, XIE Yan-hong^{1,2}

(1. College of Science, Shenyang University of Chemical Technology, Shenyang Liaoning 110142, China;

2. Research Center for Technical Process Fault Diagnosis and Safety,

Shenyang University of Chemical Technology, Shenyang Liaoning 110142, China)

Abstract: A fault detection method based on time-space nearest neighborhood standardization and local outlier factor (TSNS-LOF) was proposed to deal with the problem of nonlinear, dynamic and mean drift of complex process data. Firstly, training samples are normalized by using the time-space nearest neighborhood set; then the local outlier factor of samples are calculated on standard sample set, and the upper quantile is determined as the detection control limit, and the online fault detection is performed. The time-space nearest neighborhood standardization overcomes the difficulties of the nonlinearity, dynamics and mean drift. The local outlier factor measures the similarity of samples, to separate the fault samples and the normal samples. The fault detection experiment of Tennessee Eastman Process was carried out. The results showed that TSNS-LOF is timelier for the early fault warning, and has higher detection rate than principal component analysis, dynamic principal component analysis, k nearest neighbor rule and local outlier factor methods. The theoretical analysis and simulation experiments showed that the TSNS-LOF method is suitable for fault detection of dynamics or multiple or both operating faults and ensures the safety of the production process and high quality of products.

Key words: time-space neighborhood standardization; local outlier factor; model; principal component analysis; process control

Citation: FENG Liwei, LI Yuan, ZHANG Cheng, et al. Time-space neighborhood standardization-local outlier factor based fault detection for complex process. *Control Theory & Applications*, 2020, 37(3): 651 – 657

1 引言

伴随着现代化学工业产品在生产和生活各领域中广泛应用, 如何保证生产过程的安全性和产品的高质

量成为一个重要问题. 基于数据驱动的故障检测和过程监控技术在化学工业过程中已经得到了广泛的应用. 主元分析(principal component analysis, PCA)、偏

收稿日期: 2019-01-03; 录用日期: 2019-07-08.

†通信作者. E-mail: li-yuan@mail.tsinghua.edu.cn; Tel.: +86 24-89385010.

本文责任编辑: 陈增强.

国家自然科学基金重点项目(61490701), 国家自然科学基金项目(61673279)资助.

Supported by the Key Project of the National Natural Science Foundation of China (61490701) and the National Natural Science Foundation of China (61673279).

最小二乘方法和独立元分析是经典的故障检测方法^[1-12]. 它们都采用了统计量 T^2 和SPE对过程进行监控, 但这两个统计量都假定数据来源于独立同分布的单模态生产过程.

当过程具有动态特性时, 前后时刻样本存在的时序相关性导致数据无法满足独立性的要求, 针对此问题, Ku等提出了动态主元分析(dynamic principal component analysis, DPCA)^[13]. Rato等^[14-15]研究了DPCA中时滞参数选取问题. DPCA通过构造时滞矩阵将不同时刻样本间的时序相关性转化为变量间的相关性, 成功消除了时序相关性的影响. 但是, DPCA类方法并不适用于多模态过程.

针对过程的非线性和多模态特征, He等提出了基于 k 近邻的故障检测方法(k nearest neighbor rule, k NN), 并成功检测出半导体蚀刻工艺过程的大部分故障^[16]. 针对各模态方差不同的多模态过程, 冯等提出基于标准距离 K 近邻方法^[17]. 但是, 当过程具有强动态性时, k NN类方法的检测性能显著降低. 非线性、多模态过程的另一类监控策略是标准化方法. Ma等提出局部近邻标准化(local neighborhood standardization, NS)的故障检测方法^[18-19]. 冯等提出统计模量和局部近邻标准化的局部离群因子故障检测方法^[20]. NS类方法将多模态数据融合为单模态数据, 但亦没有考虑时序相关性的问题.

Breunig等提出一种基于局部离群因子(local outlier factor, LOF)的离群点检测方法^[21]. LOF使用样本的相对密度来度量其离群程度, 可以用来区分故障样本和正常样本, 能够实现非线性过程、单模态过程或多模态过程(包括方差显著不同)的故障检测^[22-24]. Ma等提出马氏距离局部离群因子的故障检测方法^[22]. 刘等提出基于局部密度估计的多模态故障检测方法^[23]. Zhu等提出多模态批次过程LOF监控, 并应用于青霉素发酵过程中^[24]. 但是, 上述LOF类方法均未考虑过程动态性的影响.

针对复杂生产过程的非线性、多模态和动态特征, 本文提出一种基于时空近邻标准化和局部离群因子(time-space nearest neighborhood standardization and LOF, TSNS-LOF)的故障检测方法, 并在数值模拟过程和化工生产过程中进行了检验.

2 局部离群因子方法

LOF方法使用局部离群因子表征样本的离群程度, 是一种建立在可达距离基础上的基于相对密度的离群点检测算法, 能够对无动态性的非线性、多模态过

程进行故障检测. LOF方法计算过程如下:

首先, 寻找样本 \mathbf{x} 的前 k 近邻集, 并计算 \mathbf{x} 到每个近邻样本的第 k 可达距离

$$rd_k(\mathbf{x}, \mathbf{x}^{(f)}) = \max\{d_k(\mathbf{x}^{(k)}), d(\mathbf{x}, \mathbf{x}^{(f)})\}, \quad (1)$$

其中: $d_k(\mathbf{x}^{(k)})$ 是 \mathbf{x} 和其第 k 近邻 $\mathbf{x}^{(k)}$ 间的距离, $d(\mathbf{x}, \mathbf{x}^{(f)})$ 是 \mathbf{x} 和其第 f 近邻 $\mathbf{x}^{(f)}$ 间的距离.

其次, 计算样本 \mathbf{x} 的局部可达密度

$$lrd_k(\mathbf{x}) = \frac{1}{\frac{1}{k} \sum_{f=1}^k rd_k(\mathbf{x}, \mathbf{x}^{(f)})}. \quad (2)$$

最后, 计算样本 \mathbf{x} 的局部离群因子

$$\text{LOF}(\mathbf{x}) = \frac{1}{k} \sum_{f=1}^k \frac{lrd_k(\mathbf{x}^{(f)})}{lrd_k(\mathbf{x})}. \quad (3)$$

在所有训练样本的LOF值的基础上使用核密度估计等方法确定其所服从的分布, 并将其分布的上分位点作为检测模型的控制限. 当一个在线样本的LOF值大于控制限时, 判定过程发生故障, 否则为正常.

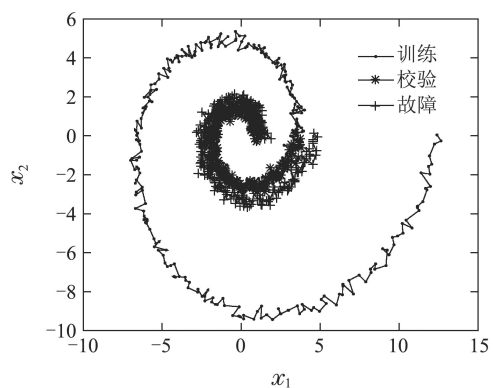
在LOF方法中, 近邻距离的计算能够消除数据的非线性、多模态特性, 使数据集融合为单模态数据; 相对密度的计算能够调整数据的疏密程度使各模态数据的方差近似相等.

当过程变量随时间具有显著变化时, 其数据中心逐渐漂移且方差逐渐变化, 这导致LOF方法对部分故障会出现漏报. 使用一个动态非线性的数值模拟过程进行阐述. 过程有两个输入变量 θ , t 和两个输出变量 x_1 , x_2 , 模型如下:

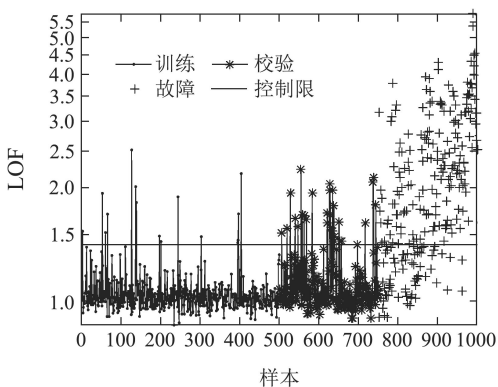
$$\begin{cases} x_1 = e^{at} \cos \theta + e_1, \\ x_2 = e^{at} \sin \theta + e_2, \\ \theta : 0 \sim 4\pi, t : 0 \sim 4\pi, \end{cases} \quad (4)$$

其中: a 为过程参数, 本文取为0.2, e_1, e_2 为随机噪声. 过程正常运行产生500个样本作为建模数据. 再次, 让过程正常运行, 但从 2π 时刻起在变量 t 上添加幅值为 -4 的阶跃型扰动信号, 共产生500个样本作为测试数据. 所有数据的空间分布见图1(a), 正常数据的中心和方差随时间逐渐变化, 位于后半阶段的故障样本靠近前半阶段的正常数据.

使用LOF方法对该过程进行故障检测实验, 选取近邻个数为4, 取97%的置信度, 检测结果见图1(b). 从图1(a)可看出有大量故障非常接近正常样本导致其可达距离较小, 所以其局部离群因子小于正常样本, 因此被LOF误报为正常样本.



(a) 原始数据



(b) LOF

图1 动态非线性过程监控

Fig. 1 Monitoring for dynamic and nonlinear process

3 基于时空近邻标准化局部离群因子的故障检测方法

当故障靠近正常样本时, 为了避免LOF方法的误报, 可以先采用标准化策略对数据进行处理, 使故障样本远离正常样本. Ma等^[19]提出了NS-LOF方法, 使用样本的最近邻样本的近邻集标准化当前样本, 再使用LOF方法进行检测. NS-LOF的步骤如下: 首先寻找样本 \mathbf{x} 的最近邻样本 $\mathbf{x}^{(1)}$, 其次寻找 $\mathbf{x}^{(1)}$ 的前 N 近邻样本集, 计算此近邻集的均值 $m(N(\mathbf{x}^{(1)}))$ 和标准差 $s(N(\mathbf{x}^{(1)}))$, 最后使用式(5)对 \mathbf{x} 进行标准化:

$$\bar{\mathbf{x}} = \frac{\mathbf{x} - m(N(\mathbf{x}^{(1)}))}{s(N(\mathbf{x}^{(1)}))}. \quad (5)$$

使用NS方法对前一节图1(a)中的过程数据进行处理, 选取近邻个数为4, 结果见图2(a). 由于后半阶段发生的故障样本的最近邻样本处于前半阶段, 导致其近邻集也都来自前半阶段样本集, 所以该部分故障样本在NS标准化后混入了正常样本, NS标准化出现了错误. 因此, 所用的近邻样本应来自于样本的相近时刻的正常样本才是合理的, 所以本节提出数据处理方法——时空近邻标准化(TSNS).

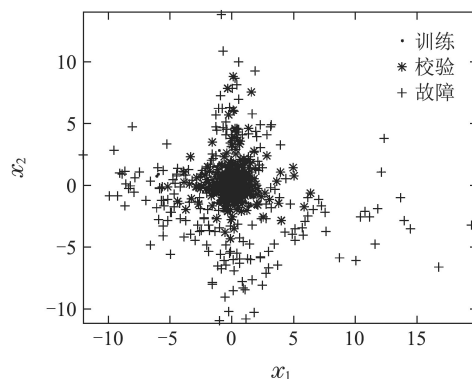
首先, 在训练集中寻找样本 \mathbf{x} 在时间方向上的前 N 近邻样本集 $N(\mathbf{x}) = \{\mathbf{x}_1^{(x)}, \dots, \mathbf{x}_t^{(x)}, \dots, \mathbf{x}_N^{(x)}\}$, 再寻找 $\mathbf{x}_t^{(x)}$ 在空间方向上的前 K 近邻样本集 $N(\mathbf{x}_t^{(x)}) =$

$\{\mathbf{x}_1^{(xt)}, \dots, \mathbf{x}_f^{(xt)}, \dots, \mathbf{x}_K^{(xt)}\}$, 使用式(6)–(7)计算此样本集的均值 $m(N(\mathbf{x}_t^{(x)}))$ 和标准差 $s(N(\mathbf{x}_t^{(x)}))$, 最后使用式(8)将 \mathbf{x} 标准化为标准样本:

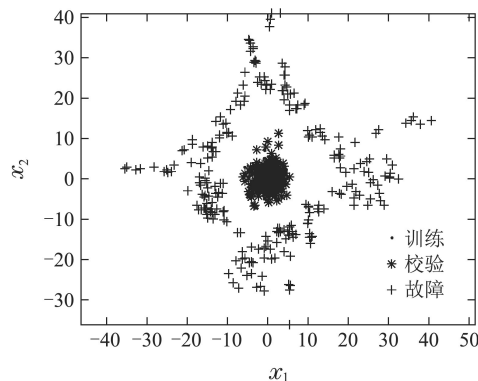
$$m(N(\mathbf{x}_t^{(x)})) = \frac{1}{K} \sum_{f=1}^K \mathbf{x}_f^{(xt)}, \quad (6)$$

$$s(N(\mathbf{x}_t^{(x)})) = \sqrt{\frac{1}{K-1} \sum_{f=1}^K (\mathbf{x}_f^{(xt)} - m(N(\mathbf{x}_t^{(x)})))^2}, \quad (7)$$

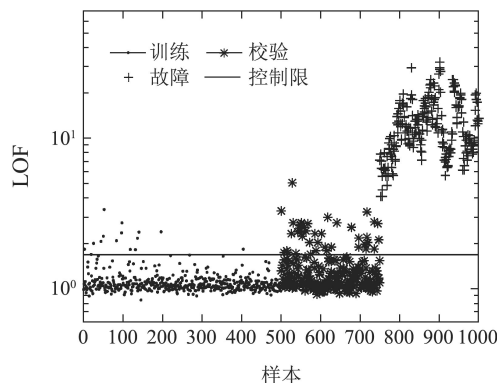
$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{t=1}^N \frac{\mathbf{x} - m(N(\mathbf{x}_t^{(x)}))}{s(N(\mathbf{x}_t^{(x)}))}. \quad (8)$$



(a) NS标准化后数据



(b) TSNS标准化后数据



(c) TSNS-LOF

图2 TSNS-LOF方法对动态非线性过程的监控

Fig. 2 Monitoring of TSNS-LOF method in dynamic and nonlinear process

对于正常样本而言,其时间和空间近邻样本都是相似的,经TSNS处理后样本位于坐标原点周围.对于故障样本而言,它离其同时刻的正常样本较远,所以其近邻距离较大,经TSNS处理后,其标准样本离坐标原点较远.总之,TSNS方法能够实现故障与正常样本在空间上的分离.由于TSNS方法使用了相近时刻样本的近邻集,能够避免只使用空间近邻导致的错误.

时空近邻标准化中的两个近邻参数 N 和 K 的选取原则:要求训练数据经时空近邻标准化后数据具有最佳的正态性.此时每个样本之间是独立的,也就是没有时序相关性,即消除了过程的动态性.

使用TSNS方法对前一节的数据进行标准化处理,其中 $N = 4$, $K = 5$,结果见图2(b).TSNS方法实现了故障样本和正常样本的分离.

在经TSNS处理后得到的标准样本集 \bar{X} 上使用LOF方法进行故障检测.首先,计算样本 \bar{x} 到其前 k 近邻集中所有近邻样本 \bar{x}^f 的第 k 可达距离

$$rd_k(\bar{x}, \bar{x}^f) = \max\{d_k(\bar{x}^k), d(\bar{x}, \bar{x}^f)\}. \quad (9)$$

其次,计算样本 \bar{x} 的局部可达密度

$$lrd_k(\bar{x}) = \frac{1}{\frac{1}{k} \sum_{f=1}^k rd_k(\bar{x}, \bar{x}^f)}. \quad (10)$$

接下来,计算样本 \bar{x} 的局部离群因子

$$\text{LOF}(\bar{x}) = \frac{1}{k} \sum_{f=1}^k \frac{lrd_k(\bar{x}^f)}{lrd_k(\bar{x})}. \quad (11)$$

最后,根据训练样本的LOF值使用核密度估计等方法确定其分布,将其分布的相应于置信度 α 的上分位数作为检测模型的控制限.TSNS-LOF具体计算过程如下:

1) 离线建模.

a) 采集正常样本组成模型的训练集 \mathbf{X} ,对每个样本 \mathbf{x}_i 寻找其时间方向上的前 N 近邻集 $N(\mathbf{x}_i) = \{\mathbf{x}_1^{(i)}, \dots, \mathbf{x}_t^{(i)}, \dots, \mathbf{x}_N^{(i)}\}$,再对 $N(\mathbf{x}_i)$ 中的每个样本 $\mathbf{x}_t^{(x)}$,寻找其空间方向上前 K 近邻集 $N(\mathbf{x}_t^{(i)}) = \{\mathbf{x}_1^{(it)}, \dots, \mathbf{x}_f^{(it)}, \dots, \mathbf{x}_K^{(it)}\}$.

b) 使用式(6)–(8)进行对 \mathbf{x}_i 进行时空近邻标准化,获得标准样本 $\bar{\mathbf{x}}_i$,组成标准训练集 $\bar{\mathbf{X}}$.

c) 对 $\bar{\mathbf{X}}_i$ 中每个样本 $\bar{\mathbf{x}}_i$ 按式(9)–(11)计算其局部离群因子 LOF_i .

d) 对全体训练样本的 LOF_i 使用核密度估计方法确定相应于置信度 α 的检测模型控制限 LOF_α .

2) 在线检测.

a) 在 \mathbf{X} 中,寻找在线新样本 \mathbf{x} 在时间方向上的前 N 近邻集 $N(\mathbf{x}) = \{\mathbf{x}_1^{(x)}, \dots, \mathbf{x}_t^{(x)}, \dots, \mathbf{x}_N^{(x)}\}$,再对 $N(\mathbf{x})$ 中的每个样本 $\mathbf{x}_t^{(x)}$,查询其空间方向上的前 K 近

邻集 $N(\mathbf{x}_t^{(x)}) = \{\mathbf{x}_1^{(xt)}, \dots, \mathbf{x}_f^{(xt)}, \dots, \mathbf{x}_K^{(xt)}\}$.

b) 使用式(6)–(8)对 \mathbf{x} 进行时空近邻标准化操作,获得标准样本 $\bar{\mathbf{x}}$.

c) 按式(9)–(11)计算 $\bar{\mathbf{x}}$ 的局部离群因子 LOF_x .

d) 当 $\text{LOF}_x > \text{LOF}_\alpha$ 时,则判定样本 \mathbf{x} 为故障,否则为正常.

图2(c)为参数 $k = 4$ 时TSNS-LOF方法对前一节所述过程的故障检测结果,检测出全部故障.

4 TE过程故障检测

田纳西-伊斯曼过程(Tennessee-Eastman process)是一个典型的化工联合生产过程.已有大量的过程控制和故障诊断方法^[2,6,8,12,15,18,25]将其作为算法的评价测试平台.TE过程主要由反应器、冷凝器、压缩机、分离器和汽提塔等5个操作单元组成^[15,18].在生产过程中,4种气态物料A, C, D和E反应生产出两种产品G和H,伴有一种副产品F,另外在进料中还混有少量的惰性气体B^[15,18].

TE过程包含12个控制变量,19个连续测量变量和22个成分测量变量^[15,18].本文选取全部41个测量变量用于故障检测.过程在模态1(G:H = 50:50)下正常运行48 h后,切换到模态3(G:H = 90:10)下再运行48 h,前后分别以3 min和6 min为间隔采集数据组成训练集;为了检验TSNS-LOF方法的性能,产生一组测试集:过程在模态1下运行48 h,从8 h至结束引入7类故障,以3 min为采样间隔产生测试数据,故障描述见表1.

表1 TE过程故障描述

Table 1 Fault descriptions for TE process

批次	故障描述	类型
f1	A/C进料比	阶跃
f2	成分B	阶跃
f3	A,B,C进料比	随机
f4	反应动态	慢偏移
f5	进料C温度	随机
f6	反应器冷却水入口温度	随机
f7	反应器冷却水阀	粘住

使用PCA, DPCA, k NN, LOF与TSNS-LOF对TE过程进行监控,并检测故障,所用方法的置信水平均取为97%.图3给出了测试集中故障批次f6的检测图,表2分别给出了所用方法对测试集中全部故障批次的故障检测率和误报率(最后一行).

PCA和DPCA主元个数分别取为9和20^[14],DPCA中的时滞参数取为2^[14]. k NN的近邻个数 $k = 4$.TSNS-LOF方法的近邻个数 $N = 3$, $K = 4$, $k = 4$.从表2可看出所有方法对测试集的故障f1, f2, f3, f4的检

测率均较高; PCA, DPCA, k NN和LOF方法对故障f5, f6, f7的检测率较低, 而TSNS-LOF方法检测率最高.

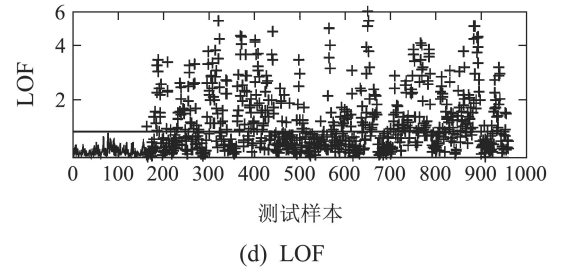
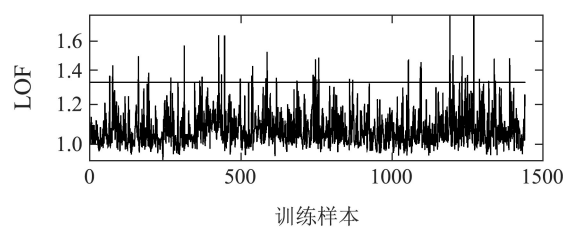
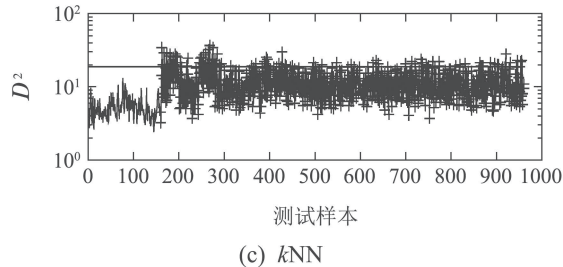
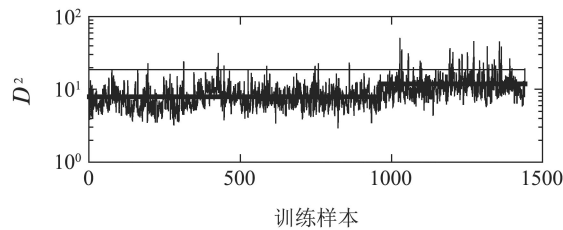
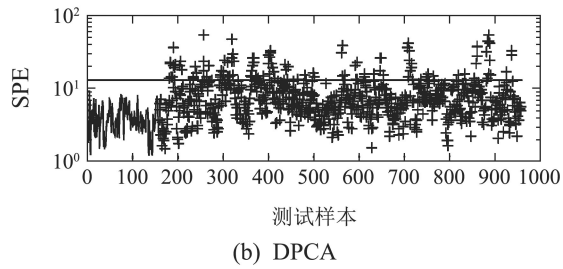
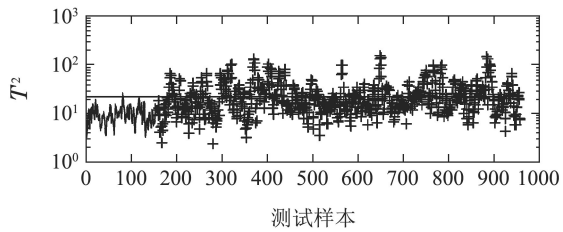
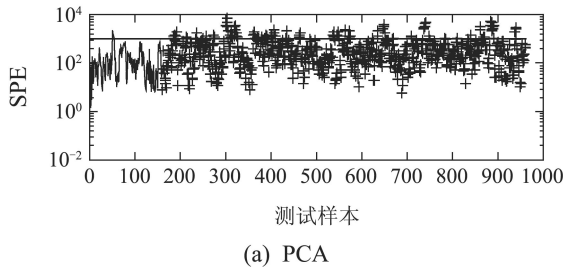
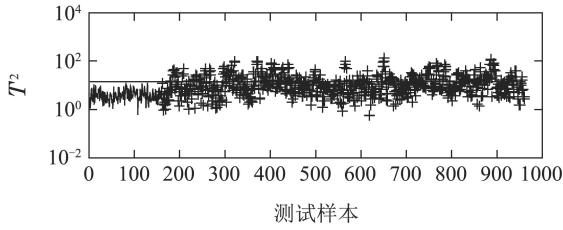


Fig. 3 测试集中批次f6的故障检测结果

Fig. 3 Fault detection results for batch f6 in test set

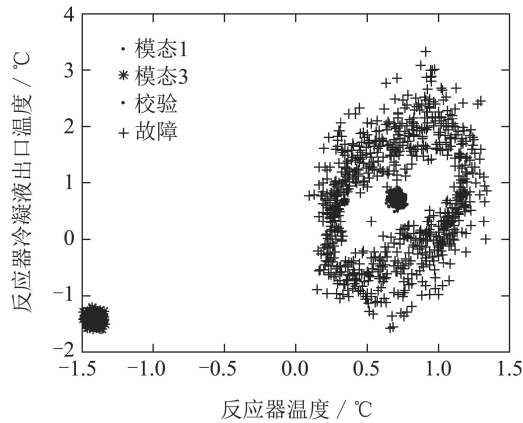
表2 测试集的故障检测率和误报率(%)

Table 2 Fault detection rate and false alarm rate of test set (%)

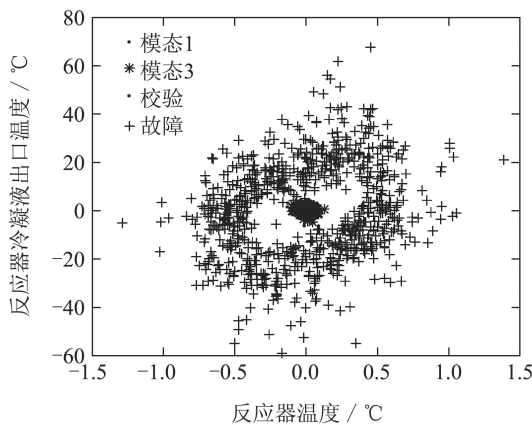
故障	PCA		DPCA		k NN	LOF	TSNS-LOF
	T^2	SPE	T^2	SPE	D^2	LOF	LOF
f1	51.9	98.9	59.9	98.8	99.6	99.8	99.6
f2	90.9	98.0	96.7	97.4	99.0	99.1	99.1
f3	95.6	97.6	97.0	97.0	98.4	98.5	98.6
f4	98.6	97.6	98.4	97.4	98.6	98.4	98.6
f5	4.6	4.9	10.9	0.25	2.0	5.7	83.4
f6	20.1	33.5	42.8	22.5	41.5	49.9	83.9
f7	2.3	22.1	1.8	12.5	11.6	31.1	99.1
f1~f7	0	2.5	0.6	0.0	0.6	0.6	0.6

图4(a)给出了正常样本和故障f7的反应器温度和反应器冷凝液出口温度, TE过程正常数据分为两个模态, 故障f7位于正常样本周围. PCA, DPCA方法的统计量都要求过程数据服从单模态的多元高斯分布, 它们的 T^2 控制限为在各自元空间内中心在原点的超椭球面. 而很多故障位于两个模态数据之间, 即比较接近数据中心导致其 T^2 值较小, 它们被错误地包含在超椭球面内部, 导致误报为正常, 所以 T^2 对f5, f6, f7的检测率较低. SPE控制限为残差空间内的超球面, 由于大多数故障位于两个模态数据之间, 导致其到数据中心的欧式距离较小, 也被包含在超球面内部, 仍被

误报为正常,所以SPE对f5, f6, f7的检测率也较低. DPCA方法虽然能够提取过程的动态特性,但是其统计量亦有单模态的多元高斯分布的假设,所以其对多模态过程的部分故障的检测率也比较低.



(a) 原始数据的变量图



(b) TSNS处理后数据的变量图

Fig. 4 TE过程数据图

Fig. 4 TE process data diagrams

k NN方法和LOF方法能够对一般的多模态过程进行故障检测. LOF方法能够提取疏密程度信息,所以其对f5, f6, f7的检测率比 k NN方法略高. 但它们对f5, f6, f7的检测率仍然较低. 下面以故障f7的第201个样本分析 k NN方法和LOF方法漏报的原因. 此样本的故障主要发生在第9个变量(反应器温度)上见图4(a)中方框点. 所有训练样本的统计量中变量贡献值的平均值见图5,在其变量贡献值中,虽然第9个变量贡献值高于正常样本的,但是其余变量的贡献较小,导致其变量贡献值之和(即统计量)小于训练样本的,也就低于控制限,所以不能被LOF方法检测出, k NN方法的原因类似.

TSNS-LOF方法使用正确时刻的近邻集信息将数据标准化为单模态数据见图4(b),同时其避免了故障样本标准化出错的问题,所以TSNS-LOF方法的检测率最高且检测比较及时,见图3(e)和表2.

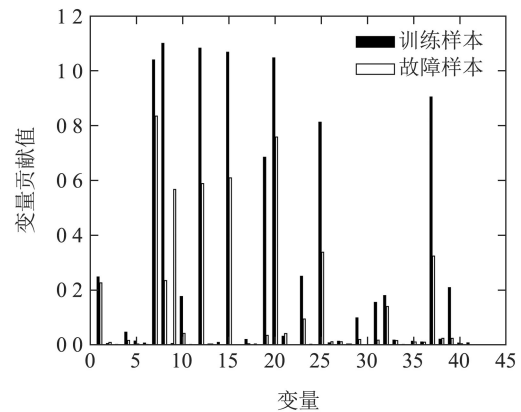


Fig. 5 LOF中变量贡献图

Fig. 5 Contribution diagram of variables in LOF

5 结论

通过数值模拟过程分析了LOF在动态过程故障检测中存在的不足. 针对复杂过程的非线性、动态性和多模态的特点,提出了TSNS-LOF的故障检测方法. TSNS克服了非线性、动态性和多中心的不利影响. 将TSNS-LOF应用于TE过程故障检测过程中,并与PCA, DPCA, k NN, LOF方法进行比较. 实验结果表明TSNS-LOF对动态性或多模态过程故障具有较高的检测能力.

参考文献:

- [1] 周东华, 李钢, 李元. 数据驱动的工业过程故障诊断技术: 基于主元分析与偏最小二乘的方法. 北京: 科学出版社, 2011.
- [2] CHERRY G A, QIN S J. Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis. *IEEE Transactions on Semiconductor Manufacturing*, 2006, 19(2): 159 – 172.
- [3] YU J, QIN S J. Multimode process monitoring with bayesian inference-based finite gaussian mixture models. *Aiche Journal*, 2008, 54(7): 1811 – 1829.
- [4] ZHAO C, YAO Y, GAO F, et al. Statistical analysis and online monitoring for multimode processes with between-mode transitions. *Chemical Engineering Science*, 2010, 65(22): 5961 – 5975.
- [5] LEE J M, YOO C K, SANG W C, et al. Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 2004, 59(1): 223 – 234.
- [6] CHENG C Y, HSU C C, CHEN M C. Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes. *Industrial and Engineering Chemistry Research*, 2011, 49(5): 2254 – 2262.
- [7] ZHANG Y, LI S, HU Z. Improved multi-scale kernel principal component analysis and its application for fault detection. *Chemical Engineering Research & Design*, 2012, 90(9): 1271 – 1280.
- [8] ZHOU D, LI G, QIN S J. Total projection to latent structures for process monitoring. *Aiche Journal*, 2010, 56(1): 168 – 178.
- [9] ROSIPAL R, TREJO L J. Kernel partial least squares regression in reproducing kernel hilbert space. *Journal of Machine Learning Research*, 2001, 2(3): 97 – 123.
- [10] ZHANG Y, HU Z. Multivariate process monitoring and analysis based on multi-scale KPLS. *Chemical Engineering Research & Design*, 2011, 89(12): 2667 – 2678.

- [11] BACH F R, JORDAN M I. Kernel independent component analysis. *Journal of Machine Learning Research*, 2002, 3(1): 1 – 48.
- [12] GE Z, SONG Z. Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors. *Industrial & Engineering Chemistry Research*, 2007, 46(7): 2054 – 2063.
- [13] KU W, STORER R H, GEORGAKIS C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 1995, 30(1): 179 – 196.
- [14] RATO T J, REIS M S. Advantage of using decorrelated residuals in dynamic principal component analysis for monitoring large-scale systems. *Industrial & Engineering Chemistry Research*, 2013, 52(38): 13685 – 13698.
- [15] RATO T J, REIS M S. Fault detection in the tennessee eastman benchmark process using dynamic principal components analysis based on decorrelated residuals (DPCA-DR). *Chemometrics & Intelligent Laboratory Systems*, 2013, 125(7): 101 – 108.
- [16] 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.
- [17] 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.)
- [18] MA H, HU Y, SHI H B. A novel local neighborhood standardization strategy and its application in fault detection of multimode processes. *Chemometrics & Intelligent Laboratory Systems*, 2012, 118(7): 287 – 300.
- [19] MA H, HU Y, SHI H B. Fault detection and identification based on the neighborhood standardized local outlier factor method. *Industrial & Engineering Chemistry Research*, 2013, 52(6): 2389 – 2402.
- [20] FENG Liwei, ZHANG Cheng, LI Yuan, et al. Local outlier factor fault detection method based on statistics pattern and local nearest neighborhood standardization. *Journal of Computer Applications*, 2018, 38(4): 946 – 970. (冯立伟, 张成, 李元, 等. 基于统计模量和局部近邻标准化的局部离群因子故障检测方法. *计算机应用*, 2018, 38(4): 965 – 970.)
- [21] BREUNIG M M, KRIEGEL H P, NG R T, et al. LOF: identifying density-based local outliers. *ACM SIGMOD International Conference on Management of Data*. Dalles USA, ACM, 2000: 93 – 104.
- [22] 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.)
- [23] LIU Bangli, MA Yuxin, SHI Hongbo. Multimode process monitoring based on local density estimation. *CIESC Journal*, 2014, 65(8): 3071 – 3081. (刘帮莉, 马玉鑫, 侍洪波. 基于局部密度估计的多模态过程故障检测. *化工学报*, 2014, 65(8): 3071 – 3081.)
- [24] ZHU J, WANG Y, ZHOU D. Batch process modeling and monitoring With local outlier factor. *IEEE Transactions on Control Systems Technology*, 2018, 27(1): 1 – 13.
- [25] TAN S, WANG F, PENG J, et al. Multimode process monitoring based on mode identification. *Industrial & Engineering Chemistry Research*, 2011, 51(1): 374 – 388.

作者简介:

冯立伟 硕士, 讲师, 目前研究方向为过程故障检测, E-mail: feng-li-wei@163.com;

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

张成 硕士, 副教授, 目前研究方向为过程故障诊断分析, E-mail: zhangcheng@syuct.edu.cn;

谢彦红 教授, 目前研究方向为过程控制与故障诊断, E-mail: xyh666256@163.com.