

# 控制与决策

Control and Decision

## 基于双权重多邻域保持嵌入的间歇过程故障检测

姚红娟, 赵小强, 李炜, 惠永永

引用本文:

姚红娟, 赵小强, 李炜, 等. 基于双权重多邻域保持嵌入的间歇过程故障检测[J]. *控制与决策*, 2021, 36(12): 3023–3030.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.0659>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### [基于局部信息增量与MPLS的质量相关故障检测方法](#)

Quality-related fault detection method based on local information increment and MPLS

*控制与决策*. 2021, 36(7): 1647–1654 <https://doi.org/10.13195/j.kzyjc.2019.1402>

### [基于批次图像化的卷积自编码故障监测方法](#)

Fault detection of batch image-based convolutional autoencoder

*控制与决策*. 2021, 36(6): 1361–1367 <https://doi.org/10.13195/j.kzyjc.2019.1342>

### [基于自适应混合核典型变量分析的工业过程质量相关故障检测](#)

Quality-related fault detection for industrial processes based on adaptive mixed kernel canonical variable analysis

*控制与决策*. 2021, 36(4): 801–807 <https://doi.org/10.13195/j.kzyjc.2020.0690>

### [基于相互邻近度的密度峰值聚类算法](#)

Density peaks clustering based on mutual neighbor degree

*控制与决策*. 2021, 36(3): 543–552 <https://doi.org/10.13195/j.kzyjc.2019.0795>

### [基于交叉熵的改进NPE间歇过程故障检测算法](#)

Improved NPE batch process fault detection algorithm based on cross entropy

*控制与决策*. 2021, 36(2): 411–417 <https://doi.org/10.13195/j.kzyjc.2019.0725>

# 基于双权重多邻域保持嵌入的间歇过程故障检测

姚红娟, 赵小强<sup>†</sup>, 李 炜, 惠永永

1. 兰州理工大学 电气工程与信息工程学院, 兰州 730050;
2. 兰州理工大学 甘肃省工业过程先进控制重点实验室, 兰州 730050;
3. 兰州理工大学 国家级电气与控制工程实验教学中心, 兰州 730050)

**摘要:** 针对间歇过程数据的动态特性带来的故障检测问题, 提出一种双权重多邻域保持嵌入 (double weight multiple neighborhoods preserving embedding, DWMNPE) 算法. 首先, 为每个样本点寻找时间近邻来描述样本点之间的时序相关关系; 其次, 定义角度近邻, 并为样本点寻找角度近邻和距离近邻, 以表征样本点在空间上的相似性, 通过提取这 3 种不同的流形特征, 充分表征数据的局部结构特征; 再次, 构造一种新的目标函数, 在考虑误差最小的同时兼顾 3 种近邻的顺序信息, 可防止 NPE 算法在计算重构权重时丢失近邻顺序信息, 在解决数据动态性的同时充分提取原始数据的本质局部结构; 最后, 对降维数据构造局部离群因子 (local outlier factor, LOF) 统计量进行监控, 消除数据非高斯特性对监控效果的不利影响. 数值例子和青霉素发酵过程仿真结果验证了 DWMNPE 方法对动态性间歇过程故障检测的有效性.

**关键词:** 间歇过程; 故障检测; 动态性; 邻域保持嵌入; 角度近邻; 局部离群因子

中图分类号: TP277

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.0659

开放科学(资源服务)标识码(OSID):



**引用格式:** 姚红娟, 赵小强, 李炜, 等. 基于双权重多邻域保持嵌入的间歇过程故障检测[J]. 控制与决策, 2021, 36(12): 3023-3030.

## Fault detection of batch process based on double weight and multiple neighborhoods preserving embedding

YAO Hong-juan, ZHAO Xiao-qiang<sup>†</sup>, LI Wei, HUI Yong-yong

1. College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. Gansu Key Laboratory of Advanced Control for Industrial Processes, Lanzhou University of Technology, Lanzhou 730050, China;
3. National Experimental Teaching Center of Electrical and Control Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

**Abstract:** Aiming at the problem of fault detection caused by the dynamic characteristic of batch process data, a double weight multiple neighborhoods preserving embedding (DWMNPE) algorithm is proposed. Firstly, by finding time neighbors for each sample point, the time correlations between samples are reflected. By defining angle neighbors, sample points are reconstructed to represent the similarity in the space by finding time neighbors, angle neighbors and distance neighbors for sample points. Three different manifold features can fully extract the essential structure of original data. Then, considering the minimum error and three kinds of neighbor order information, a new objective function is constructed to further prevent the loss of neighbor order information when the reconstructing weights are calculated using the NPE algorithm. The data dynamic is solved, meanwhile, the essential local structure is achieved. Finally, the LOF statistic of the dimensionality reduction data is constructed to monitor the process and eliminate the bad influence of data non-Gaussian for monitoring effect. The results of a numerical example and the penicillin fermentation process simulation demonstrate that the DWMNPE algorithm is effective for fault detection in dynamic batch process.

**Keywords:** batch process; fault detection; dynamic; NPE; angle neighbor; local outlier factor

收稿日期: 2020-05-29; 修回日期: 2020-09-03.

基金项目: 国家自然科学基金项目(61763029, 61763027); 国防基础科研项目(JCKY2018427C002); 甘肃省高等学校产业支撑引导项目(2019C-05); 甘肃省工业过程先进控制重点实验室开放基金项目(2019KFJJ01).

责任编辑: 赵春晖.

<sup>†</sup>通讯作者. E-mail: xqzhao@lut.cn.

## 0 引言

近年来,间歇过程已被广泛应用于注塑、生物制药、半导体制造等高附加值产品生产领域,是一种重要的工业生产方式<sup>[1-2]</sup>.一旦生产过程发生故障而没有及时报警,将造成巨大的经济损失和人员伤亡.因此,对间歇过程进行及时、准确的故障检测以保障其安全、稳定生产具有重要的经济价值和社会意义<sup>[3]</sup>.

在间歇过程故障检测研究中,基于大量历史数据的多元统计方法是一类重要的故障检测方法<sup>[4]</sup>,多向主成分分析(multiway principal component analysis, MPCA)<sup>[5]</sup>、多向偏最小二乘(multiway partial least squares, MPLS)<sup>[6]</sup>等都是典型的多元统计方法,此类方法通过降维提取原始数据的特征信息,并建立统计量监测过程运行状态,因此,维度约简算法的特征提取能力直接影响到过程的监控性能.

数据降维的目的是提取高维原始数据的本质特征,其中局部几何结构表征了相邻数据点之间的空间分布关系,是一种重要的特征信息.传统的降维算法如主成分分析算法(principal component analysis, PCA)通过保留数据的方差信息进行降维,保持的是数据的全局结构,忽略了数据的局部结构,因此,基于流形学习的降维算法得到广泛关注和研究. Tenenbaum等<sup>[7]</sup>提出了等距映射(isometric feature mapping, ISOMAP); Belkin等<sup>[8]</sup>提出了基于谱图理论的拉普拉斯特征映射(Laplacian eigenmaps, LE); Roweis<sup>[9]</sup>提出了局部线性嵌入(locally linear embedding, LLE); He等<sup>[10]</sup>提出了邻域保持嵌入(neighborhood preserving embedding, NPE),该算法是LLE的线性逼近,相较于LLE算法,NPE的优势在于可以计算出精确的线性投影矩阵,使得实时性较好,因此在人脸识别<sup>[11]</sup>、生物医学<sup>[12]</sup>等领域得到了成功的应用,而且近年来,NPE也被引入到故障检测邻域<sup>[13-16]</sup>.

间歇过程数据具有动态特性,不满足传统方法进行监测时过程采样独立的假设,会导致高漏报率的存在.针对这一问题,文献[17]最早提出了动态主元分析(dynamic principal component analysis, DPCA)算法,通过包含当前时刻和历史时刻采样值的增广矩阵建立故障检测模型,考虑了过程数据之间的时序相关性,提高了故障检测率.此后,一系列基于增广矩阵的多元统计方法被用于动态过程监控中,取得了一定的故障检测效果<sup>[18-19]</sup>.文献[20]提出了一种改进的动态核主元分析(dynamic kernel principal component analysis, DKPCA)故障检测方法,首先通过不可区分度筛选出相关程度较大的变量以减少数据量,接着

构建增广矩阵,最后通过核主元分析(kernel principal component analysis, KPCA)建模,该方法对微小故障表现出更好的故障检测能力.但是,以上方法揭示的是数据的全局结构,忽略了数据的局部结构,导致特征提取效果欠佳,直接影响了故障检测效果. NPE算法由于能够在降维的同时挖掘数据的局部结构,被广泛用于过程监测中,但是它没有考虑过程数据的时序相关性.对此,文献[21]借鉴DPCA的思想,将增广矩阵应用于NPE算法,提出了动态多向NPE算法,并成功应用于间歇过程故障检测,然而间歇过程数据丰富,采用增广矩阵使得算法的计算效率降低.文献[22]提出了时序扩展的NPE(TNPE)算法,为每一个样本点挑选 $k$ 个时间近邻重构样本点,以挖掘出采样数据之间的相关信息,通过统计量监控过程性能,但由于降维后的数据不服从高斯分布,用传统的统计量进行监控是不准确的.局部离群因子(local outlier factor, LOF)表征测试样本对训练样本的偏离程度,由于故障数据可以视为训练样本的异常点,可以利用LOF构建统计量进行故障检测,LOF统计量无需数据满足高斯分布的假设,具有更好的监控效果<sup>[23]</sup>.

综上,本文提出一种双权重多邻域保持嵌入(double weight multiple neighborhood preserving embedding, DWMNPE)的故障检测方法.该方法通过定义角度近邻,并挑选距离近邻、时间近邻、角度近邻重构样本点,在时间和空间上完成对样本点的重构.为了进一步考虑近邻顺序信息对重构权重的影响,构建新的目标函数,同时考虑近邻顺序信息和重构误差,有效提取数据的本质流形特征,最后通过构建LOF统计量,实现对动态间歇过程的监控.

## 1 邻域保持嵌入(NPE)

NPE作为一种局部流形学习算法,其基本思想是:对于 $R^m$ 空间数据集 $X(x_1, \dots, x_n)$ 中的每一个数据点,可以由其相邻的数据点进行线性重构,并能够找到一个低维映射 $A(a_1, \dots, a_d)(d \leq m)$ ,使得数据点的线性重构得到最好的保持,从而得到低维空间数据集 $Y(y_1, \dots, y_n)$ ,且满足 $Y = A^T X$ . NPE算法具体步骤如下:

1) 构造邻接图. 计算样本点之间的欧氏距离,并按距离从小到大排序,挑选前 $k$ 个样本点作为近邻点,并将样本点与其所有近邻点一一进行连线,构建邻接图.

2) 计算权重矩阵. 设 $W$ 为权重矩阵,若从样本点 $x_i$ 到样本点 $x_j$ 有一条边连接,则这条边上的权重系数为 $w_{ij}$ ,若从 $x_i$ 到 $x_j$ 无连接,则 $w_{ij} = 0$ . 在NPE算法中,所有样本点均能够通过其近邻点的线性组合进

行重构,因此,权重矩阵  $W$  通过计算重构误差最小化得到,有

$$\Phi(W) = \min \sum_{i=1}^n \left\| x_i - \sum_{j=1}^n w_{ij} x_j \right\|^2, \quad (1)$$

其中约束条件为  $\sum_{j=1}^n w_{ij} = 1$ , 即权重矩阵的行(列)元素的和为1.

3) 计算投影矩阵. NPE算法的目的是在数据约简的同时,保持数据的局部流行结构不变,即在高维空间  $R^m$  中重构数据点  $x_i$  的权重系数,可以在低维空间中重构对应的数据点  $x_j$ . 因此,投影矩阵  $A$  通过求解下式的最优解得到:

$$\begin{aligned} A &= \arg \min \sum_{i=1}^n \left\| y_i - \sum_{j=1}^n w_{ij} y_j \right\|^2 = \\ &\arg \min y^T (I - W)^T (I - W) y = \\ &\arg \min a^T X (I - W)^T (I - W) X^T a, \end{aligned} \quad (2)$$

其中约束条件为  $y^T y = a^T X X^T a = 1$ .

引入拉格朗日函数,并通过一定的代数运算,将式(2)的最优化问题转换为下式的广义特征值求解问题:

$$X M X^T a = \lambda X X^T a, \quad (3)$$

其中  $M = (I - W)^T (I - W)$ . 求解最小的  $d$  个特征值 ( $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d$ ) 所对应的特征向量构成  $A(a_1, \dots, a_d)$ .

## 2 局部离群因子(LOF)

局部离群因子(LOF)是由 Breunig 等<sup>[24]</sup> 首次提出用于寻找数据集中离群点的方法,其通过计算样本点的局部离群因子来衡量其离群程度. 样本集  $X \in R^{n \times m}$  中任一样本  $x_i$  的局部离群因子计算步骤如下,其中  $n$  为样本点数,  $m$  为变量个数.

首先,在欧氏距离度量下找到样本点  $x_i$  的  $k$  个近邻点,构成局部近邻集  $N(x_i) = \{x_i^1 \dots x_i^f \dots x_i^k\}$ ,  $f = 1, 2, \dots, k$ . 样本  $x_i$  与其近邻集中距其最远的第  $k$  个近邻之间的距离  $d(x_i, x_i^f)$  称为  $x_i$  的邻域半径,记为  $r(x_i)$ . 样本  $x_i$  与其近邻点  $x_i^f$  之间的可达距离定义为

$$\text{reach}_d(x_i, x_i^f) = \max\{r(x_i), d(x_i, x_i^f)\}, \quad (4)$$

则样本  $x_i$  的局部可达密度为

$$\text{Ird}(x_i) = \frac{k}{\sum_{f=1}^k \text{reach}_d(x_i, x_i^f)}. \quad (5)$$

样本  $x_i$  的局部离群因子定义为  $x_i$  的所有近邻的

平均局部可达密度与样本点  $x_i$  局部可达密度的比值,即

$$\text{LOF}(x_i) = \frac{1}{k} \sum_{f=1}^k \frac{\text{Ird}(x_i^f)}{\text{Ird}(x_i)}, \quad (6)$$

其中  $\text{Ird}(x_i^f)$  为样本  $x_i$  近邻点  $x_i^f$  的局部可达密度. 从式(6)可以看出:当  $x_i$  不是离群点时,其与近邻点具有相近的局部可达密度,则其局部离群因子  $\text{LOF} \approx 1$ ; 当  $x_i$  是离群点时,  $x_i$  具有比其近邻点较小的局部可达密度,则其局部离群因子  $\text{LOF} > 1$ .

## 3 基于DWMNPE算法的故障检测

### 3.1 双权重多邻域保持嵌入(DWMNPE)算法

NPE算法通过构建样本点的多个局部邻域来描述其特征结构,因此,邻域划分的好坏直接影响了NPE算法的特征提取能力. NPE算法通过欧氏距离选择邻域点,没有考虑采样点之间的时序相关性. 通常,在较短的采样时间内,样本点变化较小,可以认为每个邻域内的样本点之间构成线性相关的关系<sup>[22]</sup>. 因此,本文以采样时间邻近的样本点作为一种近邻点来描述样本点之间的动态相关信息,时间近邻表达了在时间序列上的相近,却忽略了在空间上的表达. NPE算法通过欧氏距离度量样本在空间上的相似性,但是这种方式并不充分. 如图1所示,样本点  $x_i$  与样本点  $x_j, x_k$  之间的距离  $d_1, d_2$  相等,但  $x_i$  与  $x_j, x_k$  之间的夹角不同,可见样本之间的空间相似性并不相同,说明仅仅通过欧氏距离大小来衡量样本之间的空间相似性是不准确的. 为此,通过样本间角度的大小来定义另一种空间相似性的度量指标,即角度近邻,定义如下:

$$\cos \theta_{ij} = \frac{x_i x_j^T}{\|x_i\| \|x_j\|}. \quad (7)$$

其中  $\theta_{ij}$  是样本点  $x_i$  与  $x_j$  ( $j = 1, 2, \dots, n, j \neq i$ ) 之间的角度,样本间夹角余弦值越大,夹角越小,样本间相似性越大.

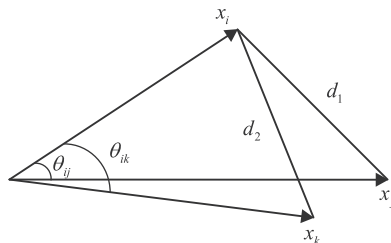


图1 样本间角度示意图

为了能够在时间和空间上充分表达样本的结构特征,本文根据样本点间欧氏距离的大小、采样时间的顺序、角度余弦的大小分别为样本点寻找距离、时间以及角度3种近邻点,从3种不同的角度来衡量

样本点之间的相似性,更加准确地表征样本的本质特征.

NPE算法在用近邻点重构样本点时,仅仅通过重构误差求解重构系数矩阵 $W$ ,未考虑近邻距离大小的顺序,一般而言,相比于距离较远的近邻点,距离较近的近邻点应该在重构样本点时具有更大的权重.本文通过下式描述的动态多变量数值过程<sup>[25]</sup>来分析这一问题:

$$\begin{cases} x(t) = Ax(t-1) + Bu(t-1), \\ y(t) = x(t) + v(t), \\ u(t) = Cu(t-1) + Dw(t-1). \end{cases} \quad (8)$$

其中: $x \in R^{3 \times 1}$ 为状态变量, $u \in R^{2 \times 1}$ 为输入变量, $y \in R^{3 \times 1}$ 为输出变量; $v$ 为0均值、0.02方差的高斯测量噪声, $w$ 为0均值、0.05方差的高斯动态噪声;系数矩阵 $A$ 、 $B$ 、 $C$ 、 $D$ 分别取值为

$$A = \begin{bmatrix} 0.118 & -0.191 & 0.287 \\ 0.847 & -0.264 & 0.943 \\ -0.333 & 0.514 & -0.217 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 3 & -4 \\ -2 & 1 \end{bmatrix}, \\ C = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix}, D = \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix}.$$

取 $[u_1(t), u_2(t), y_1(t), y_2(t), y_3(t)]$ 为监控变量,通过式(8)产生500个数据用于仿真动态过程.在NPE算法中,设置近邻点数 $k$ 和降维空间维度 $d$ 分别为12和2.NPE算法可求得权重矩阵,其中第1个样本点的12个近邻点重构系数的绝对值如图2所示.从图2可以看到,重构权值最大的为第2个近邻点,重构权值最小的为第9个近邻点,即并非越相似的近邻点的重构系数越大,可知NPE算法权重的计算忽略了近邻的顺序信息.

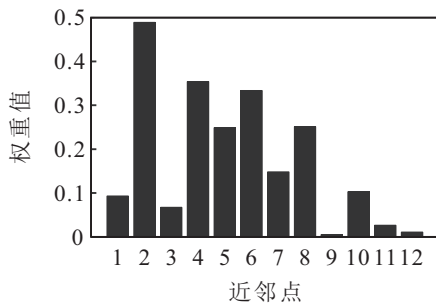


图2 第1个样本点的重构权重值

DWMNPE算法在求解重构权重时,不仅考虑了重构误差,还考虑到3种近邻的顺序信息,目标函数构造如下:

$$\min J = \min(\alpha J_E + (1 - \alpha) J_O). \quad (9)$$

$J_E$ 是NPE的目标函数,表征了重构误差信息,具体如下式所示:

$$J_E = \sum_{i=1}^n \left\| y_i - \sum_{j=1}^s w_{E_{ij}} y_j \right\|^2 = A^T X (I - W_E)^T (I - W_E) X^T A. \quad (10)$$

其中: $W_E$ 是通过样本点与其近邻点重构误差最小计算得到的; $s$ 是近邻点总数,包括距离近邻、时间近邻以及角度近邻.

$J_O$ 是表征重构顺序信息的目标函数,具体如下式所示:

$$J_O = \sum_{i=1}^n \left\| y_i - \sum_{j=1}^s w_{O_{ij}} y_j \right\|^2 = A^T X (I - W_O)^T (I - W_O) X^T A. \quad (11)$$

其中 $W_O$ 为表征数据距离顺序、时间顺序以及角度顺序信息的权重矩阵,借鉴高斯核, $W_O$ 定义为

$$W_{O_{ij}} = \begin{cases} \exp(-\|x_i - x_j\|^2 / \beta), & x_j \text{为} x_i \text{的距离近邻;} \\ \exp(-\|t_i - t_j\|^2 / \beta), & x_j \text{为} x_i \text{的时间近邻;} \\ \exp\left(-\left\|\frac{1}{\cos \theta_{ij}}\right\|^2 / \beta\right), & x_j \text{为} x_i \text{的角度近邻;} \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

此外,在计算权重 $W_O$ 前,需要对3种近邻顺序信息进行z-score标准化处理,以消除变量的量纲对结果的影响.

综上,DWMNPE的目标函数为

$$J = \alpha J_E + (1 - \alpha) J_O = \alpha A^T X (I - W_E)^T (I - W_E) X^T A + (1 - \alpha) A^T X (I - W_O)^T (I - W_O) X^T A = A^T X [\alpha (I - W_E)^T (I - W_E) + (1 - \alpha) (I - W_O)^T (I - W_O)] X^T A = A^T X \tilde{M} X^T A. \quad (13)$$

其中:约束条件为

$$y^T y = a^T X X^T a = 1, \\ \tilde{M} = \alpha (I - W_E)^T (I - W_E) + (1 - \alpha) (I - W_O)^T (I - W_O).$$

投影矩阵 $A$ 通过求解如下最优化问题得到:

$$A = \arg \min a^T X \tilde{M} X^T a. \quad (14)$$

引入拉格朗日函数,并通过一定的代数运算,将式(14)的最优化问题转换为下式的广义特征值求解问题:

$$X \tilde{M} X^T a = \lambda X X^T a. \quad (15)$$

求解最小的 $d$ 个特征值( $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d$ )所对应的特征向量构成 $A(a_1, \dots, a_d)$ .

### 3.2 基于降维数据的LOF故障检测

传统方法如PCA、NPE等通过 $T^2$ 和SPE统计量检测过程运行状况时,需要过程数据满足高斯分布的假设,然而实际过程数据并不满足这一假设,若继续采用 $T^2$ 和SPE统计量监控过程运行状况,必将减低故障检测率.因此,本文采用基于密度的方法对降维后的数据建立故障检测模型,构造LOF统计量并通过核密度估计(KDE)确定其控制限.相比于传统的基于距离的统计量,LOF统计量从样本的密度出发,通过离群程度判断检测样本,对数据的分布没有依赖性,避免了 $T^2$ 和SPE统计量对数据分布的局限,因此,具有更好的故障检测能力.

### 3.3 故障检测步骤

基于DWMNPE算法的故障检测包括离线建模和在线检测两部分.

#### 1) 离线建模.

step 1: 在正常工况下,获取若干批次正常工况数据作为训练样本 $X(I \times J \times K)$ , $I$ 为批次数, $J$ 为变量数, $K$ 为采样点数;

step 2: 将 $X(I \times J \times K)$ 沿着批次方向展开并进行标准化处理得到二维矩阵 $X(I \times JK)$ ,然后将其按变量方向排列为矩阵 $X(IK \times J)$ ;

step 3: 通过求解式(15)所示广义特征值问题,得到投影矩阵 $A$ ,并根据 $Y = A^T X$ ,得到二维训练样本的低维表示 $Y$ ;

step 4: 对于降维矩阵 $Y$ ,计算其局部离群因子(LOF),并采用核密度估计(KDE)确定其控制限.

#### 2) 在线检测.

step 1: 在故障工况下,得到新的待检测数据

$X_{new}$ ,并根据离线建模得到的训练样本均值和方差进行展开和标准化处理;

step 2: 根据离线建模得到的投影矩阵 $A$ ,得到 $x_{new}$ 的低维表示 $y_{new}$ ,在降维数据集 $Y$ 中挑选其邻域并计算其LOF值;

step 3: 判断统计量值是否超过控制限,若超过,则表明过程有故障发生,反之,表明过程正常.

## 4 案例研究及分析

### 4.1 数值例子

通过式(8),产生20批次正常条件下的数据作为训练样本,每一批数据由500个采样点构成.在实际的间歇生产过程中,批次过程不可能完全重复.为了对实际间歇过程更好地模拟,将每一批次数据的变量相关系数稍微改动.此外,对变量 $u_2$ 从第201个采样点开始到整个过程结束添加幅值为 $0.01 \times (t - 200)$ 的斜坡故障,产生一批故障数据用于验证故障检测效果.

将PCA、DPCA、NPE、TNPE以及DWMNPE算法用于上述数值模拟的间歇过程故障检测中,结果如图3所示,图中蓝色“.”代表200个正常工况测试数据,红色“o”代表300个故障工况测试数据,图3(a)~(e)中椭圆为 $T^2$ 统计量控制限,图(f)中椭圆为LOF统计量控制限.图3(a)、(b)分别为PCA和DPCA算法的检测图,从图3(a)可以看到,大部分故障数据在椭圆线内部,可知PCA算法无法准确检测到故障的发生,出现了大量的漏报;从图3(b)可以看到,部分故障数据在椭圆线之外,可知DPCA算法能够检测到大部分故障,具有一定的故障检测效果.图3(c)、(d)分别为NPE

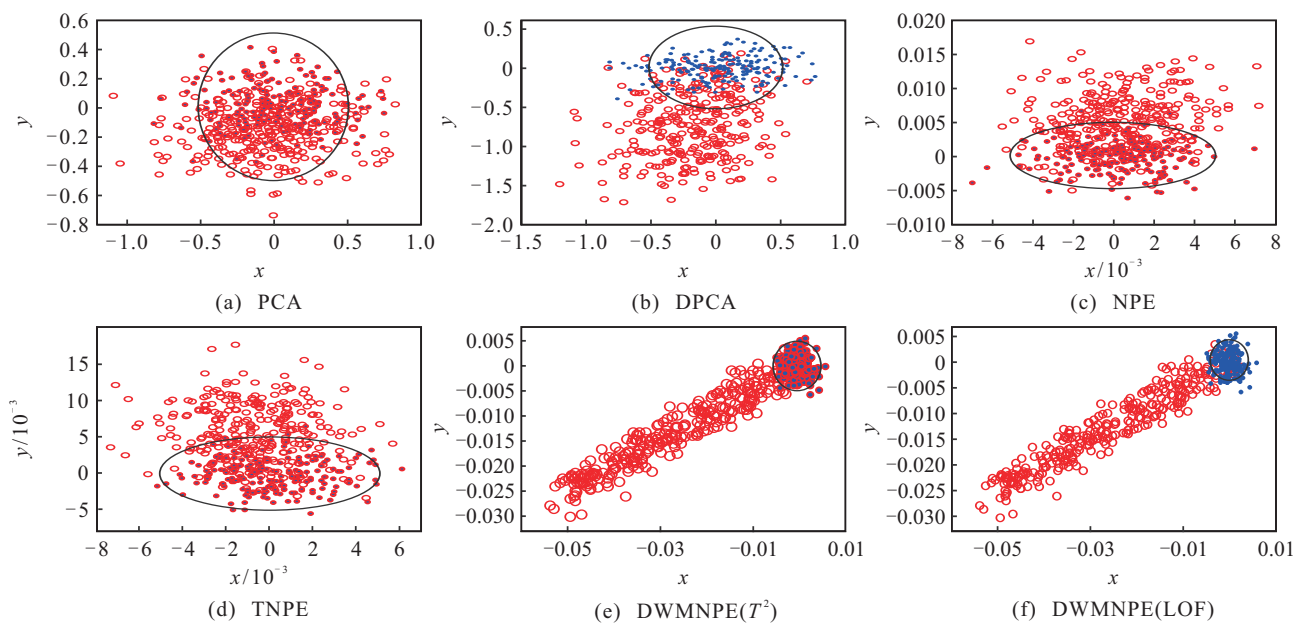


图3 数值例子统计量检测

和TNPE算法的检测图,从图3(c)可以看到,部分故障数据在椭圆线之外,可知NPE算法相比于PCA算法,具有更好的故障检测效果;从图3(d)可以看到,TNPE算法同样具有较好的故障检测效果.图3(e)、(f)分别为DWMNPE算法的 $T^2$ 统计量检测图和LOF统计量检测图,从图中可以看出,相比于PCA、DPCA、NPE、TNPE算法,DWMNPE算法具有较小的漏报率,表明DWMNPE算法有效地解决了NPE算法未考虑样本近邻顺序信息的问题,使得特征提取更加准确,且其LOF统计量的故障漏报率低于统计量,表现出更好的故障检测效果.

#### 4.2 青霉素发酵过程

为了进一步验证本文所提算法的有效性,通过青霉素生产制备过程仿真平台Pensim2.0<sup>[26]</sup>产生间歇过程数据.设置反应时间为400h,采样时间为1h,通过设置不同但都在正常范围内的初始条件产生30批正常工况数据,选取18个变量中的10个变量(通风速率(L/h)、搅拌速率(r/min)、底物流加速率(L/h)、补料温度(K)、底物浓度(g/L)、溶解氧浓度(mmol/L)、反应器体积(L)、排气二氧化碳浓度(mmol/L)、pH值和发酵罐温度(K))作为监控变量,构成训练样本集,

为了更加符合实际间歇过程,所有变量均加入了高斯白噪声.此外,本文通过设置不同故障类型和幅度产生6个故障批次数据,如表1所示.

表1 青霉素发酵过程的6个故障批次

故障	变量名	类型	幅值/%	引入时间/h
1	通风速率/(L/h)	阶跃	8	150~250
2	通风速率/(L/h)	斜坡	2	150~400
3	搅拌速率/(r/min)	阶跃	8	200~400
4	搅拌速率/(r/min)	斜坡	8	200~400
5	底物流加速率/(L/h)	阶跃	8	150~350
6	底物流加速率/(L/h)	斜坡	0.5	250~350

PCA、NPE、DPCA、TNPE以及DWMNPE算法对表1中6种故障的检测率如表2所示.通过对检测结果的对比分析可知,对于故障1,上述5种算法均能完全检测到故障的发生,但对于其他5种故障,本文所提方法具有最高的故障检测率,故障检测延时最小,可知,本文所提算法优于PCA、NPE、DPCA以及TNPE算法.故障5存在较大检测延时的原因主要在于反应底物通过相关变量传播的速度较慢,使得其故障特征难以及时显现.与阶跃故障相比,斜坡故障(如故障2、4、6)因其过程变量变化缓慢而更难被检测到.

表2 青霉素发酵过程的6个故障批次的故障检测率

%

故障	PCA		NPE		DPCA		TNPE		DWMNPE		
	$T^2$	SPE	$T^2$	SPE	$T^2$	SPE	$T^2$	SPE	$T^2$	SPE	LOF
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.680	0.880	0.832	0.920	0.861	0.960	0.930	0.940	0.988	0.984	0.992
3	0.986	0.962	0.991	0.988	0.994	0.992	0.996	0.993	0.998	0.996	0.999
4	0.593	0.582	0.596	0.584	0.781	0.762	0.801	0.796	0.913	0.921	0.929
5	0.359	0.378	0.342	0.338	0.534	0.523	0.632	0.628	0.796	0.788	0.812
6	0.941	0.936	0.950	0.954	0.962	0.958	0.968	0.961	0.972	0.967	0.980

图4为PCA、NPE、DPCA、TNPE及DWMNPE算法在故障2(通风速率在150~400采样点添加的斜坡扰动)下的检测结果.其中:图4(a)为PCA算法的统计量检测图,从图中可知,其 $T^2$ 和SPE分别在第200采样点和176采样点开始检测到故障,故障检测存在较大的延迟;图4(b)为NPE算法的统计量检测图,从图中可知,其 $T^2$ 和SPE分别在第192采样点和169采样点开始检测到故障,虽与PCA算法相比,其故障检测时间均有所提前,但从故障被引入的时间来看,仍存在较大的延迟;图4(c)为DPCA算法的统计量检测,从图中可知,其 $T^2$ 和SPE分别在第185采样点和158采样点开始检测到故障,DPCA由于考虑到了过程数据的动态性,相比于算法PCA以及NPE,故障检测时

间有所提前,但仍存在一定的延迟;图4(d)为TNPE算法的统计量检测图,从图中可知,其 $T^2$ 和SPE分别第167采样点和165采样点开始检测到故障,故障检测同样存在一定的延迟;图4(e)为DWMNPE算法的 $T^2$ 和SPE统计量检测图,从图中可知,其 $T^2$ 和SPE分别第153采样点和154采样点开始检测到故障,与PCA、NPE、DPCA以及TNPE相比,DWMNPE算法故障检测响应最快,能够在故障发生时及时检测到故障;图4(f)为DWMNPE算法的LOF统计量检测图,从图中可知,其LOF统计量在第152个采样点开始检测到了故障的发生,比其他统计量具有更快的响应速率,表现出更为满意的故障检测效果.

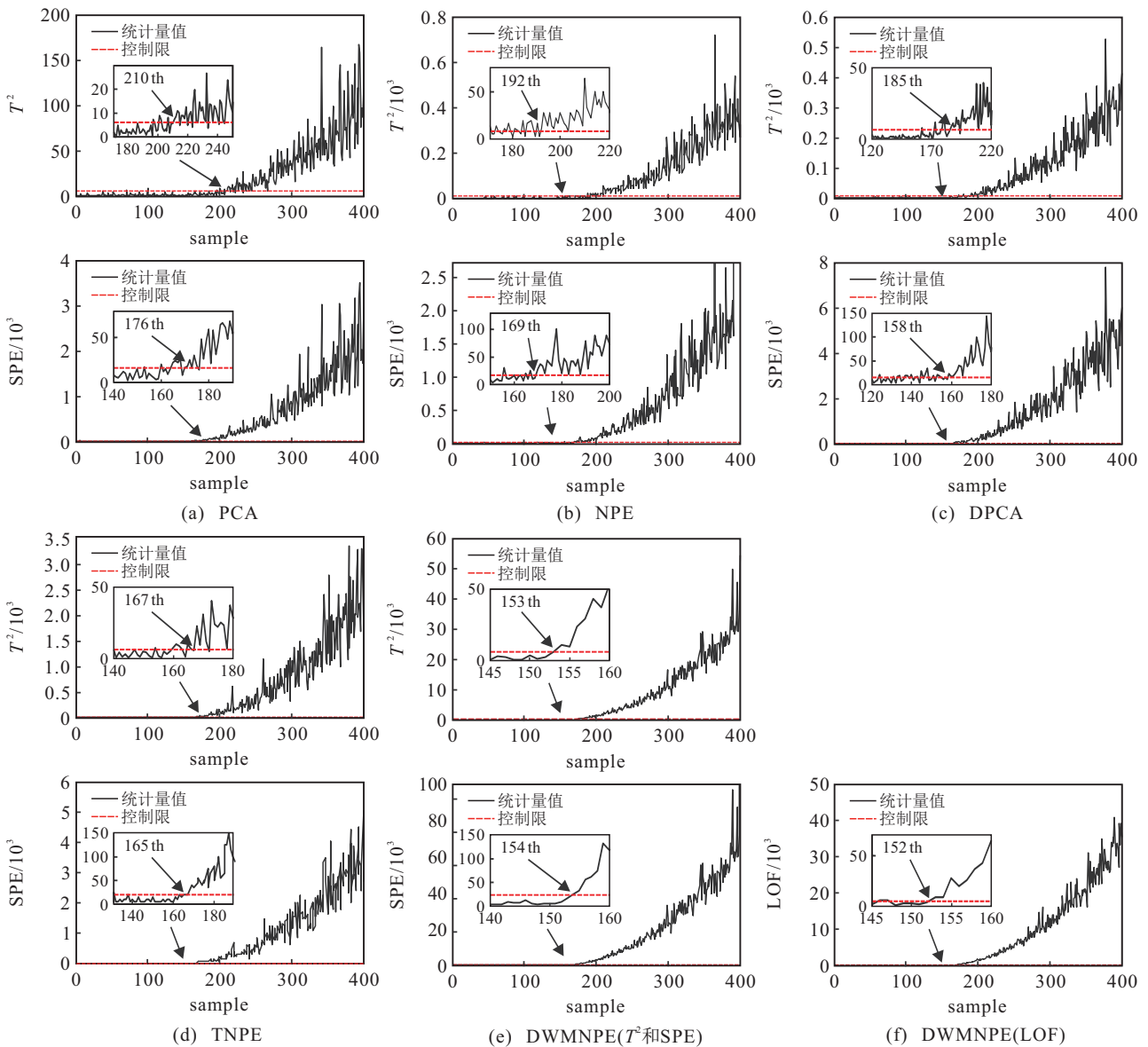


图 4 青霉素发酵过程在故障2下的检测结果

### 5 结论

本文提出了一种双权重多邻域保持嵌入算法用于间歇过程故障检测. 首先定义角度近邻, 并为样本点挑选距离近邻、时间近邻、角度近邻在时间和空间上充分描述样本点的局部特征; 接着, 在NPE算法考虑误差最小化的基础上, 考虑近邻的顺序信息, 定义一种新的目标函数, 在提取样本本质结构特征的同时解决其动态性问题; 最后, 构建LOF统计量完成对间歇过程的监控, 消除传统 $T^2$ 和SPE统计量对数据满足高斯分布的假设. 通过数值例子和青霉素发酵过程的仿真实验表明, 与PCA、NPE、DPCA、TNPE相比, 本文所提算法具有更令人满意的故障检测效果.

#### 参考文献(References)

[1] 赵春晖, 余万科, 高福荣. 非平稳间歇过程数据解析与状态监控——回顾与展望[J]. 自动化学报, 2020,

46(10): 2072-2091.

(Zhao C H, Yu W K, Gao F R. Data analytics and condition monitoring methods for nonstationary batch processes — Current status and future[J]. Acta Automatica Sinica, 2020, 46(10): 2072-2091.)

[2] 陆宁云, 王福利, 高福荣, 等. 间歇过程的统计建模与在线监测[J]. 自动化学报, 2006, 32(3): 400-410.

(Lu N Y, Wang F L, Gao F R, et al. Statistical modeling and online monitoring for batch processes[J]. Acta Automatica Sinica, 2006, 32(3): 400-410.)

[3] Ge Z Q, Song Z H, Gao F R. Review of recent research on data-based process monitoring[J]. Industrial & Engineering Chemistry Research, 2013, 52(10): 3543-3562.

[4] 常鹏, 王普, 高学金. 基于统计量模式分析的T-KPLS间歇过程故障监控[J]. 化工学报, 2015, 66(1): 265-271.

(Chang P, Wang P, Gao X J. Fault monitoring batch process based on statistics pattern analysis of T-KPLS[J]. CIESC Journal, 2015, 66(1): 265-271.)

- [5] Tong C D, Lan T, Shi X H. Fault detection and diagnosis of dynamic processes using weighted dynamic decentralized PCA approach[J]. *Chemometrics and Intelligent Laboratory Systems*, 2017, 161: 34-42.
- [6] 孔祥玉, 曹泽豪, 安秋生, 等. 偏最小二乘线性模型及其非线性动态扩展模型综述[J]. *控制与决策*, 2018, 33(9): 1537-1548.  
(Kong X Y, Cao Z H, An Q S, et al. Review of partial least squares linear models and their nonlinear dynamic expansion models[J]. *Control and Decision*, 2018, 33(9): 1537-1548.)
- [7] Tenenbaum J B, de Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. *Science*, 2000, 290(5500): 2319-2323.
- [8] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering[J]. *Advances in Neural Information Processing Systems*, 2002, 14(6): 585-591.
- [9] Roweis S T. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000, 290(5500): 2323-2326.
- [10] He X F, Cai D, Yan S C, et al. Neighborhood preserving embedding[C]. *Proceeding of the 10th IEEE International Conference on Computer Vision*. Beijing: IEEE, 2005: 1208-1213.
- [11] Chen X, Zhang J S. A novel maximum margin neighborhood preserving embedding for face recognition[J]. *Future Generation Computer Systems*, 2012, 28(1): 212-217.
- [12] Li J X, Zhang Z, Li X, et al. Kernel-based learning for biomedical relation extraction[J]. *Journal of the American Society for Information Science and Technology*, 2008, 59(5): 756-769.
- [13] 李锋, 汤宝平, 董绍江. 基于正交邻域保持嵌入特征约简的故障诊断模型[J]. *仪器仪表学报*, 2011, 32(3): 621-627.  
(Li F, Tang B P, Dong S J. Fault diagnosis model based on feature compression with orthogonal neighborhood preserving embedding[J]. *Chinese Journal of Scientific Instrument*, 2011, 32(3): 621-627.)
- [14] 杨健, 宋冰, 谭帅, 等. 时序约束NPE算法在化工过程故障检测中的应用[J]. *化工学报*, 2016, 67(12): 5131-5139.  
(Yang J, Song B, Tan S, et al. Time constrained NPE for fault detection in chemical processes[J]. *CIESC Journal*, 2016, 67(12): 5131-5139.)
- [15] 赵小强, 惠永永. 基于TTGNPE算法的间歇过程监控[J]. *控制与决策*, 2017, 32(3): 557-562.  
(Zhao X Q, Hui Y Y. Batch process monitoring based on TTGNPE algorithm[J]. *Control and Decision*, 2017, 32(3): 557-562.)
- [16] 惠永永, 赵小强. 基于WICA-WGNPE的高斯-非高斯联合指标间歇过程监控[J]. *仪器仪表学报*, 2018, 39(1): 190-199.  
(Hui Y Y, Zhao X Q. Batch process monitoring with Gaussian and non-Gaussian joint indicator based on WICA-WGNPE[J]. *Chinese Journal of Scientific Instrument*, 2018, 39(1): 190-199.)
- [17] Ku W F, Storer R H, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 1995, 30(1): 179-196.
- [18] Zhang N, Tian X M, Cai L F, et al. Process fault detection based on dynamic kernel slow feature analysis[J]. *Computers & Electrical Engineering*, 2015, 41: 9-17.
- [19] 王光, 孙程远, 尹坤. 基于动态全主成分回归质量相关的故障检测[J]. *信息与控制*, 2017, 46(6): 671-676.  
(Wang G, Sun C Y, Yin S. Quality-related fault detection approach based on dynamic total principal component regression[J]. *Information and Control*, 2017, 46(6): 671-676.)
- [20] 翟坤, 杜文霞, 吕锋, 等. 一种改进的动态核主元分析故障检测方法[J]. *化工学报*, 2019, 70(2): 296-302.  
(Zhai K, Du W X, Lyu F, et al. Fault detect method based on improved dynamic kernel principal component analysis[J]. *CIESC Journal*, 2019, 70(2): 296-302.)
- [21] Hu K L, Yuan J Q. Statistical monitoring of fed-batch process using dynamic multiway neighborhood preserving embedding[J]. *Chemometrics and Intelligent Laboratory Systems*, 2008, 90(2): 195-203.
- [22] Miao A M, Ge Z Q, Song Z H, et al. Time neighborhood preserving embedding model and its application for fault detection[J]. *Industrial & Engineering Chemistry Research*, 2013, 52(38): 13717-13729.
- [23] 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[J]. *控制理论与应用*, 2020, 37(3): 651-657.  
(Feng L W, Li Y, Zhang C, et al. Time-space neighborhood standardization-local outlier factor based fault detection for complex process[J]. *Control Theory & Applications*, 2020, 37(3): 651-657.)
- [24] Breunig M M, Kriegei H P, Ng R T, et al. LOF: Identifying density-based local outliers[C]. *ACM SIGMOD International Conference on Management of Data*. Dalles: ACM, 2000: 93-104.
- [25] Zhang H Y, Tian X M, Deng G. Batch process monitoring based on multiway global preserving kernel slow feature analysis[J]. *IEEE Access*, 2017, 5: 2696-2710.
- [26] Birol G, Ündey C, Çinar A. A modular simulation package for fed-batch fermentation: Penicillin production[J]. *Computers & Chemical Engineering*, 2002, 26(11): 1553-1565.

## 作者简介

姚红娟(1995—), 女, 博士生, 从事间歇过程故障检测与诊断的研究, E-mail: hjyao01@163.com;

赵小强(1969—), 男, 教授, 博士生导师, 从事故障检测与诊断、图像处理等研究, E-mail: xqzhao@lut.cn;

李炜(1963—), 女, 教授, 博士生导师, 从事复杂系统建模、动态系统故障诊断与容错控制等研究, E-mail: liwei@lut.cn;

惠永永(1992—), 男, 讲师, 博士, 从事过程监控与故障诊断等研究, E-mail: hyylut@163.com.

(责任编辑: 齐 霖)