

DOI:10.19651/j.cnki.emt.2211085

基于自适应时序窗口加权 k 近邻的故障检测方法^{*}冯立伟^{1,3} 顾欢^{2,3} 孙立文^{2,3} 李元³(1. 沈阳化工大学理学院 沈阳 110142; 2. 沈阳化工大学计算机科学与技术学院 沈阳 110142;
3. 辽宁省化工过程智能化技术重点实验室 沈阳 110142)

摘要: 为了解决在工业生产过程中时序和多阶段的问题,提出了一种基于正交局部保持投影(OLPP)和自适应时序窗口加权 k 近邻(ATSWKNN)的故障检测方法。首先,采用 OLPP 方法,在保持样本近邻关系的基础上,将原始数据投影到低维特征空间;其次,在特征空间中选取某类时序窗口,并计算时序平方距离;然后,将窗口内样本到其空间上近邻集的平均累积平方距离的倒数作为权重;最后,构造统计量对过程进行监控。OLPP-ATSWKNN 通过时序信息的提取和窗口内权重的引入降低了过程的自相关性和解决了多阶段的统计差异问题。此外,自适应的窗口切换策略解决了阶段切换时统计指标异常的问题。通过对数值模拟过程和青霉素发酵过程的监控实验,检验了 OLPP-ATSWKNN 的监控性能,监控结果显著优于经典方法。

关键词: 多阶段过程;加权 k 近邻;时序窗口;故障检测

中图分类号: TP277 **文献标识码:** A **国家标准学科分类代码:** 510.8010

Fault detection method based on adaptive timing sequence
window weighted k nearest neighborsFeng Liwei^{1,3} Gu Huan^{2,3} Sun Liwen^{2,3} Li Yuan³(1. College of Science, Shenyang University of Chemical Technology, Shenyang 110142, China;
2. College of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110142, China;
3. Liaoning Key Laboratory of Intelligent Technology for Chemical Process Industry, Shenyang 110142, China)

Abstract: In order to solve the timing sequence and multi-stage problems in industrial production process, a fault detection method based on orthogonal local preserving projection (OLPP) and adaptive timing sequence window weighted k nearest neighbor (ATSWKNN) was proposed. Firstly, basing on maintaining the sample nearest neighbor relationship, the original data are projected into the low-dimensional feature space by using OLPP. Secondly, a certain kind of timing window is selected in the feature space, and the timing sequence square distance is calculated. Then, the reciprocal of the average cumulative square distance between the sample in the window and its spatial nearest neighbor set is taken as the weight. Finally, statistics are constructed to monitor the process. OLPP-ATSWKNN reduces the autocorrelation of process and solve the problem of multi-stage statistical difference by extracting time series information and introducing weight within the window. In addition, the problem of abnormal statistical indicators during phase switching is solved by adaptive window switching strategy. The monitoring performance of OLPP-ATSWKNN is verified by monitoring the numerical simulation process and penicillin fermentation process, and the monitoring results are significantly better than the classical methods.

Keywords: multistage process; weighted k nearest neighbor; timing sequence window; fault detection

0 引言

随着现代技术的快速发展,工业生产过程逐渐趋于复杂。一方面表现为阶段内部及阶段间具有时序相关性特

征;另一方面体现为具有多个生产阶段。为了提高产品的品质和生产效能,对复杂工业过程进行监控有着重要意义^[1-3]。

当工业过程中数据呈现线性结构时,基于多元统计技

收稿日期:2022-08-16

* 基金项目:国家自然科学基金(61673279)、辽宁省自然科学基金(2019-MS-262)、辽宁省教育厅基金(LJ2019013)项目资助

术的主成分分析(principal component analysis, PCA)方法得到了广泛应用^[4-6]。为了有效监控非线性过程,基于核函数思想的主成分分析(kernel principal component analysis, KPCA)及其一类方法被提出^[7-8]。两类方法在提取数据的主要特征时,只考虑了数据的全局信息,其局部近邻结构发生了改变。为了使局部近邻结构得到保持,出现了正交局部保持投影方法(orthogonal locality preserving projections, OLPP)^[9]。然而三类方法对时序相关性均缺乏考虑。

针对过程中数据存在时序相关性的问题,很多学者使用动态主成分分析(dynamic principal component analysis, DPCA)方法^[10]。而李明虎等^[11]在 DPCA 的基础上使用核技巧运用了动态核主成分分析(dynamic kernel principal component analysis, DKPCA)。该方法虽然考虑了数据的时序相关性和非线性问题,但是当数据的数量规模较大时,DKPCA 的检测性能降低。为了解决这一问题,Zhang 等^[12]将特征向量选择(feature vector selection, FVS)和 DKPCA 结合。此方法使用 FVS 减少数据的监控变量,然后捕获数据的非线性和时序相关性。

上述方法仅适用于单个阶段过程。在过程具有多阶段特征时,数据的多中心特性不满足上述方法的多元高斯分布的假设。因此衍生出应用于多阶段过程的监控方法,从而避免了多元高斯分布的前提假设,例如高斯混合模型(Gaussian mixture model, GMM)和 k 近邻(k nearest neighbor, KNN)等方法^[13-17]。Chen 等^[18]基于阶段内部距离最小和阶段间距离最大的准则以及指数加权算法对 GMM 模型进行修正,提高了 GMM 的故障检测能力。冯立伟等^[19]在 KNN 的基础上引入权重使方差不同的多阶段数据处于同一标准,提出了加权 k 近邻技术(weighted k nearest neighbors, WKNN)。如果故障数据在空间上距离不同时刻的正常数据较近,WKNN 方法寻找到错误的近邻,使得监控能力降低。究其原因在于该方法丢失了时序信息。

针对监控过程中数据具有时序相关性、多阶段等特征的问题,Song 等^[20]提出了改进的动态邻域保持嵌入(improved dynamic neighborhood preserving embedding, IDNPE)方法。此方法通过时滞矩阵和一种包含局部标准差信息的距离,考虑了时序相关性以及各阶段内和阶段间的不同尺度的问题。但 IDNPE 频繁更新数据的距离矩阵从而增加了计算负担,以致系统不能及时实现故障预警。本文提出一种基于正交局部保持投影(orthogonal locality preserving projections, OLPP)和自适应时序窗口加权 k 近邻(adaptive timing sequence window weighted k nearest neighbors, ATSWKNN)的故障检测方法。首先,通过 OLPP 的投影矩阵,在保持样本的近邻关系不变的基础上,将数据投影到低维特征空间;其次,根据时序信息,构造时序窗口;然后,计算窗口内样本到其空间近邻的平方距离作

为权重,进而得到时序加权平方距离和 WD 。在监控过程中使用差商矩阵作为窗口切换判据;最后,使用统计量 WD 对过程进行故障检测。

1 加权 k 近邻方法

WKNN 方法的基本思想是寻找全部样本在空间上的双层近邻结构。将样本 \mathbf{x} 的第二层近邻的平均累积欧氏距离的倒数作为第一层近邻欧氏距离 d_i^* 的权重 W_i , 构造统计量:

$$D^2 = \frac{1}{K} \sum_{i=1}^K W_i d_i^* \quad (1)$$

其中, K 为第一层近邻个数。虽然在故障检测时 WKNN 能使统计量处于同一尺度,但是在过程中存在时序特征时,WKNN 未考虑时序信息,导致漏报部分故障。值得注意的是,该方法还存在多次计算高维样本间欧氏距离的高计算量问题,而通过特征提取能够有效地消除此缺点。

2 正交局部保持投影-自适应时序窗口加权 k 近邻方法

2.1 正交局部保持投影

为了提高监控方法的运行效率,又能保持局部样本的时间序列,故本文采用 OLPP 方法对样本进行预处理。对过程进行监控,采集到 n 个等长时刻的正常样本 $\mathbf{X}_{n \times m}$, 其中 m 表示变量数。OLPP 在保持样本的局部近邻关系不变的基础上,将 \mathbf{X} 从高维(m)空间映射到低维($l, l < m$)空间。其目标函数为:

$$\underset{\mathbf{a}}{\operatorname{argmin}} \{ f(\mathbf{a}) = \frac{\mathbf{a}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{a}}{\mathbf{a}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{a}} \} \quad (2)$$

其中, $\mathbf{D} = \operatorname{diag}(\sum_{i=1}^n S_{ij})$, $\mathbf{L} = \mathbf{D} - \mathbf{S}$, \mathbf{S} 为加权矩阵,

$S_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}}$, $t \in R$ 。为了保持近邻关系不变,添加约束条件-投影向量间正交:

$$\mathbf{a}_r^T \mathbf{a}_1 = \mathbf{a}_r^T \mathbf{a}_2 = \dots = \mathbf{a}_r^T \mathbf{a}_{r-1} = 0 \quad (3)$$

OLPP 经过迭代计算获得矩阵 $\mathbf{P}_{OLPP} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r]$, 选取前 l 列作为投影矩阵 $\mathbf{P} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_l]$, 最后将 \mathbf{X} 通过 \mathbf{P} 投影为 \mathbf{X}' 。

$$\mathbf{X}' = \mathbf{X} \mathbf{P} \quad (4)$$

OLPP 主要步骤:

1) 构造邻接图: ϵ -近邻法或 k 近邻法构造邻接图。

2) 构建加权矩阵 \mathbf{S} : 根据邻接图构建 \mathbf{S} 。

3) 迭代计算投影向量: 计算 $(\mathbf{X} \mathbf{D} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{L} \mathbf{X}^T$ 的最小非零特征值所对应的特征向量 \mathbf{a}_1 和迭代计算第 r ($r = 2, 3, \dots, m$) 个 $\mathbf{M}^{(r)}$ 的最小非零特征值所对应的特征向量 \mathbf{a}_r , 然后组成 $\mathbf{P}_{OLPP} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r]$ 。其中 $\mathbf{M}^{(r)} = \{\mathbf{I} - (\mathbf{X} \mathbf{D} \mathbf{X}^T)^{-1} \mathbf{A}^{(r-1)} [\mathbf{B}^{(r-1)}]^{-1} [\mathbf{A}^{(r-1)}]^T\} (\mathbf{X} \mathbf{D} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{L} \mathbf{X}^T$, $\mathbf{A}^{(r-1)} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{r-1}]$, $\mathbf{B}^{(r-1)} = [\mathbf{A}^{(r-1)}]^T (\mathbf{X} \mathbf{D} \mathbf{X}^T)^{-1} \mathbf{A}^{(r-1)}$ 。

4) OLPP 投影:由包含数据主要特征的前 l 列的 \mathbf{P}_{OLPP} 组成投影矩阵 $\mathbf{P} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_l]$, 将 \mathbf{X} 经过 \mathbf{P} 映射为低维特征空间的 \mathbf{X}' 。

2.2 自适应时序窗口加权 k 近邻

WKNN 能够处理多阶段问题,但忽略了过程的时序信息。当在时刻上相距较远的样本在空间上较近时, WKNN 寻找到错误的近邻,导致漏报部分故障。为避免此错误,通过对 WKNN 引入时序信息进行改造,构建 ATSWKNN 方法。

由于样本间存在时序相关性,本文定义三类宽度为 K_1 的时序窗口:向前、向后和中间时序窗口。以 \mathbf{X}' 中样本 \mathbf{x}'_i 为例,其向前时序窗口:

$$FTW = [\mathbf{x}'_{i+1}, \mathbf{x}'_{i+2}, \dots, \mathbf{x}'_{i+K_1}]$$

向后时序窗口:

$$BTW = [\mathbf{x}'_{i-K_1}, \dots, \mathbf{x}'_{i-2}, \mathbf{x}'_{i-1}]$$

中间时序窗口:

$$ITW = [\mathbf{x}'_{i-o}, \dots, \mathbf{x}'_i, \dots, \mathbf{x}'_{i+h}]$$

其中, $o+h = K_1, o$ 或 $h \in \{1, 2, \dots, K_1-1\}$ 。计算 \mathbf{x}'_i 与其某类窗口内第 j 样本 \mathbf{x}'_{i+j} 的差分 $\mathbf{x}'_i - \mathbf{x}'_{i+j}$ 。由差分得到时序平方距离:

$$d_{ij} = \|\mathbf{x}'_i - \mathbf{x}'_{i+j}\|^2 \tag{5}$$

由于各阶段间以及阶段内部样本在空间上的疏密程度存在差异,导致不同阶段的 d_{ij} 处于不同水平,所以引入权重对 d_{ij} 进行调节使之处于同一尺度。对窗口内第 j 样本 \mathbf{x}'_{i+j} 寻找其在空间上的前 K_2 个近邻样本 $[\mathbf{x}'_{(i+j,1)}, \mathbf{x}'_{(i+j,2)}, \dots, \mathbf{x}'_{(i+j,K_2)}]$ 。计算 \mathbf{x}'_{i+j} 到其第 k 近邻的平方距离 $d_{i+j}^k = \|\mathbf{x}'_{i+j} - \mathbf{x}'_{(i+j,k)}\|^2$ 。将 K_2 个 d_{i+j}^k 的均值倒数作为权重:

$$W_{ij} = 1 / \frac{1}{K_2} \sum_{k=1}^{K_2} (d_{i+j}^k)^2 \tag{6}$$

将权重 W_{ij} 赋予 d_{ij} 。再将窗口内所有样本的加权距离作平均,构成平均时序加权平方距离:

$$WD_i = \frac{1}{K_1} \sum_{j=1}^{K_1} (\sqrt{W_{ij}} d_{ij})^2 \tag{7}$$

图 1 为 3.1 节数值模拟过程中两个阶段各 500 个训练样本的统计指标 WD 。式(7)中的时序平方距离 d_{ij} 通过对样本的差分操作降低样本的相关性,使各阶段内样本处于稳态分布,无明显变化趋势;空间权重 W_{ij} 调节阶段间及阶段内样本的方差差异,使图 1 中两个阶段的 WD 处于同一尺度。

样本的每个变量的每个差分近似服从均值为 0 的正态分布,不同阶段样本的权重与标准差均不同,通过两者相互调节,使得式(7)中 $\sqrt{W_{ij}} d_{ij}$ 近似服从正态分布。于是可以得到 WD 近似服从自由度为 K_1 的非标准卡方分布。使用最小二乘法迭代获得非标准卡方分布的概率密度函数:

$$f(x) = \frac{a}{2^{K_1/2} \Gamma(K_1/2)} e^{bx/2} c x^{-1+K_1/2} + d \tag{8}$$

式中: Γ 为 Gamma 函数。设定监控的置信度为 α , 将非标准卡方分布的上 α 分位点作为控制限。

若监控过程中始终使用同一类窗口,比如向前窗口,过程运行到每个阶段末端时,则会出现当前样本的向前时序窗口内的部分样本属于下一阶段的情况。由于不同阶段中变量的数值范围显著不同,导致式(7)中时序平方距离 d_{ij} 的值偏大,从而出现对正常样本的误报,如图 1 中黑点。若此时使用中间时序窗口可避免此错误。因此在转换阶段时,引入自适应技术,实现时序窗口的自动切换。若开始时使用向前窗口,当发生阶段转换时,自适应地切换为相应的中间时序窗口;在完成阶段转换后,再次自适应地切换回向前时序窗口。若监控全程使用向后或中间时序窗口,也存在同样的误报问题,亦需进行相似的窗口切换。

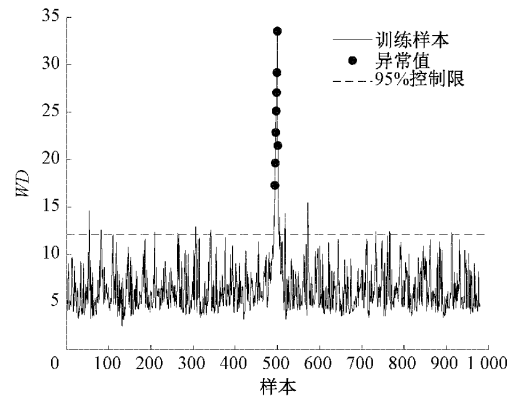


图 1 训练样本的异常统计值

对于如何自适应地进行窗口切换的判定,本文构造二维差商矩阵 \mathbf{DQ} 作为阶段切换的判别指标。每个变量的每步向前差商 \mathbf{DQ} 都近似服从正态分布,将其标准差的倍数设定为判别阶段切换的阈值,如 5 倍标准差 5σ 。若样本的某一变量从第 j 步至第 K_1 步差商都超过各自阈值,那么判定该时刻后发生阶段转换,则需调整时序窗口类别。

2.3 OLPP-ATSWKNN 的监控过程

OLPP-ATSWKNN 方法的过程监控步骤如下,流程如图 2。

1) 离线建模

(1) 计算 OLPP 的投影矩阵 $\mathbf{P} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_l]$ 。按式(4)将训练样本 \mathbf{X} 投影到 l 维空间,得到 \mathbf{X}' ;

(2) 选取某类时序窗口,计算 \mathbf{X}' 的差商矩阵 $\mathbf{DQ}_{X'}$ 。再计算 $\mathbf{DQ}_{X'}$ 中每个变量的标准差 σ , 并设定 5σ 作为窗口切换的阈值。若差商值大于阈值,则切换时序窗口类别,否则窗口类别维持不变;

(3) 按式(5)~(7)计算样本 \mathbf{x}'_i 的时序加权距离 $WD_{X'}$;

(4) 计算非标准卡方分布的上 α 分位点作为 WD 的控制限 WD_α 。

2) 在线监控

(1) 当采集到新样本 y 时,按式(4),将 y 通过投影矩阵 P 映射为 y' ;

(2) 按照训练时选取的时序窗口,计算 y' 的差商矩阵 $DQ_{y'}$;若其差商值大于 5σ ,则切换时序窗口类别,否则窗口类别维持不变;

(3) 对于样本 y' ,使用第 i 时刻训练样本 x'_i 的时序窗口,按式(5)~(7)计算其时序加权距离 $WD_{y'}$;

(4) 若 $WD_{y'} > WD_c$,则判定该样本为故障并发出警报,否则为正常样本。

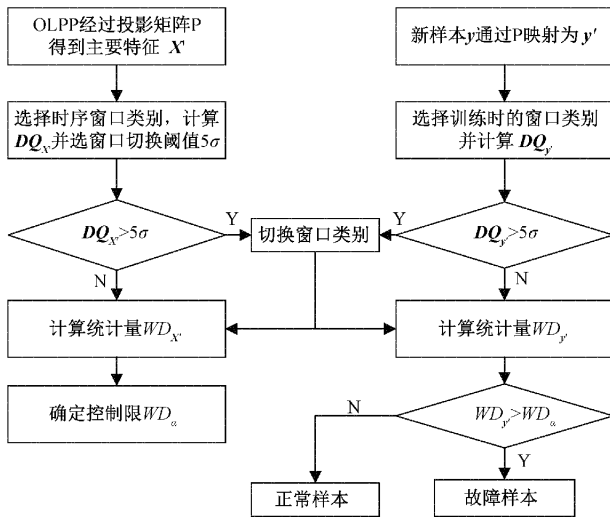


图2 OLPP-ATSWKNN 的流程

3 仿真实验研究

本节通过对具有多阶段和时序特征的数值模拟过程和青霉素发酵过程进行分析 OLPP-ATSWKNN 的检测性能,并与非线性监控方法 KPCA、OLPP,动态监控方法 DPCA,多阶段监控方法 WKNN 和多阶段时序监控技术 IDNPE 作对比分析。

3.1 数值模拟过程

本文使用的多阶段时序数值过程含有 2 个监控变量 $u = [u_1, u_2]^T$,其模型如式(9)所示,此模型是基于文献[20]改进而来。

$$u(i) = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & -0.115 \end{bmatrix} u(i-1) + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (9)$$

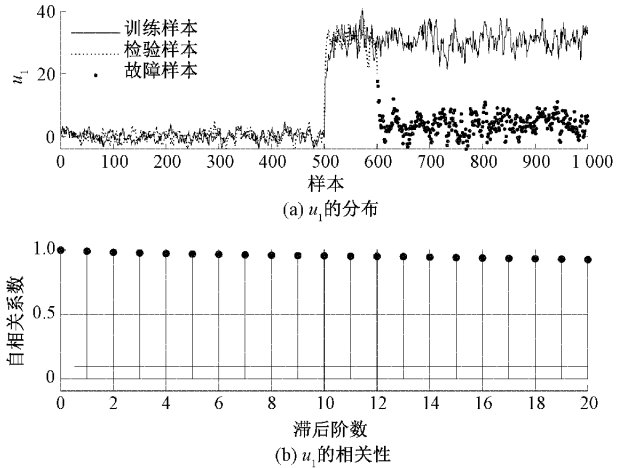
阶段 1: $e_1 \sim \mathcal{N}(0, 1), e_2 \sim \mathcal{N}(0, 1)$

阶段 2: $e_1 \sim \mathcal{N}(11, 2), e_2 \sim \mathcal{N}(11, 2)$

给定过程变量初始值 $u(1) = [0, 0.1]^T$ 。过程每一时刻迭代一次,阶段 1 持续 500 时刻;从 501~1 000 时刻运行在阶段 2。为了检验本文方法,使过程再次正常运行,从 601 时刻开始在监控变量 u_1 上引入幅值为 9.5 的阶跃故障。

此过程主要具有两个特性:阶段性和时序相关性。过

程中全部变量具有显著的阶段性:中心漂移和方差差异等,如图 3(a)。过程在第 i 时刻的样本是由第 $i-1$ 时刻所确定的,换句话说第 i 时刻的样本对后续时刻的样本有强烈影响,即过程变量具有显著自相关性。图 3(b)表明变量 u_1 的相关性超过 90% 的步数至少为 20 步。

图3 变量 u_1 的分布和相关性

使用经典方法 KPCA、DPCA 和 OLPP 对过程进行监控。结果显示 T^2 的检测率分别为 26.5%、5.8% 和 0%, SPE 的检测率为 12.0%、8.8% 和 0%。3 种方法检测率低的原因在于未考虑多阶段对监控结果的影响。

WKNN 的检测率为 50.5%。WKNN 通过引入权重将两个阶段的样本统计量统一到同一尺度,如图 4 中训练样本的统计量。图 3(a)表明故障样本与第 1 阶段的正常样本在空间上比较接近,因此在计算其加权近邻距离时,错误地将第 1 阶段的样本作为其近邻,丢失了时序信息,故出现漏报现象,如图 4 中故障样本。

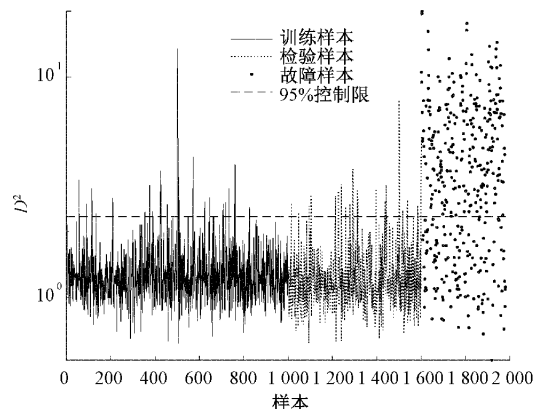


图4 WKNN 故障检测结果

IDNPE 方法在主元和残差空间中的检测率分别为 9.5% 和 9.8%,检测结果如图 5。虽然变量 u_1 在 601 时刻后显著偏离正常的轨迹,但仍处于阶段 1 的正常数值区间。所以大部分故障样本的统计指标低于控制限。

OLPP-ATSWKNN 方法的检测率为 100%。自适应

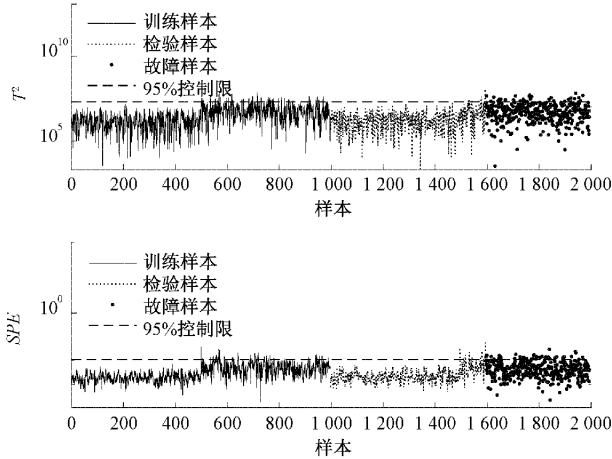


图 5 IDNPE 故障检测结果

时序窗口技术使两个阶段中样本的近邻处于各自阶段内。在计算统计量时,此方法利用权重将其统一标准,即两个阶段统计量的均值和方差近似相等,如图 6(a)中两个处于不同阶段中训练样本的统计量无显著差异。故障样本在时间尺度上处于第 2 阶段,计算其时序加权平方距离时,使用了正确的近邻,其 WD 远大于正常样本的 WD , 所以故障样本被成功检出,如图 6(a)所示。OLPP-ATSWKNN 通过时序窗口技术剔除了过程的自相关性,如图 6(b)所示。此外,自适应地进行窗口调整,避免了在切换阶段时的统计错误。

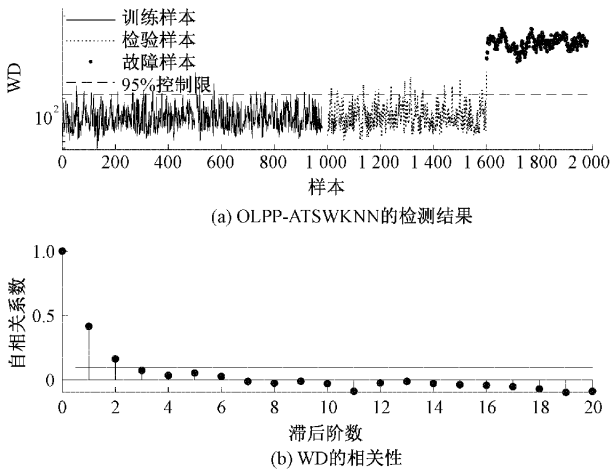


图 6 OLPP-ATSWKNN 的检测结果和 WD 的相关性

以下对统计量 WD 的卡方分布假设进行拟合优度检验。根据 WD 的观察值将其分为 q ($q = \lfloor \sqrt{n} \rfloor$) 组,获得每组中观测值的频率 f_i ; 根据假设的卡方分布计算理论上 WD 落于每组的概率作为理论值 p_i 。构造检验统计量 $\Psi = \sum_{i=1}^q \frac{(f_i - np_i)^2}{np_i}$ 。根据皮尔森^[21]定理知 Ψ 服从自由度为 $q-1$ 的卡方分布。计算自由度为 $q-1$ 的非标准卡方分布的拒绝域阈值为 $\Psi_f = 18.6$, 而 WD 的 $\Psi_{np} = 12.3 < 18.6$ 。因此 WD 近似服从非标准的卡方分布,如图 7 所示。

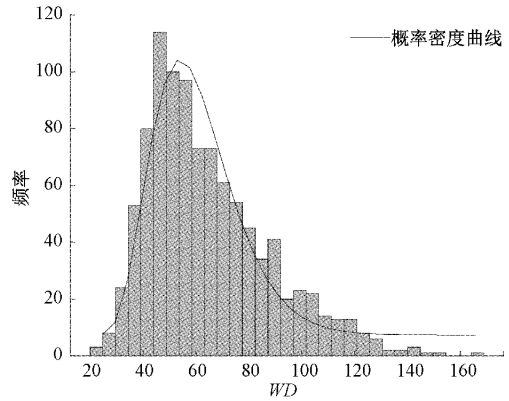


图 7 统计量的卡方分布拟合

3.2 青霉素发酵过程

本文使用的青霉素发酵仿真平台已被大量应用于故障检测与诊断方法的验证^[22-24]。此过程具有以下特征:1)多阶段性:0~43 h 为菌体生长阶段,其后为青霉素生产阶段,如图 8(a)所示,过程变量显著不同;2)时序相关性:当前过程状态对后续状态有重要影响,如图 8(b)所示 3 个变量具有显著自相关性。在整个过程中影响青霉素发酵的主要因素有 PH 值、温度、底物溶度和二氧化碳浓度等。本文选取对发酵过程有重要影响的 9 个因素作为监控变量,如表 1 所示。

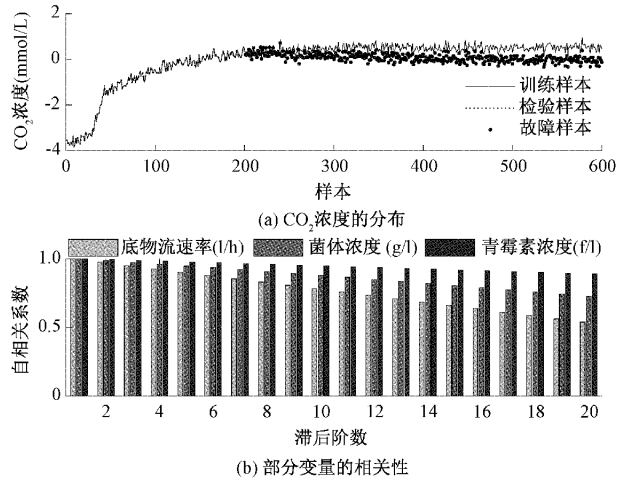


图 8 CO_2 浓度的分布和部分变量的相关性

表 1 青霉素发酵过程监控变量

序号	变量名	序号	变量名
1	空气流量/(L/h)	6	菌体浓度/(g/L)
2	搅拌功率/W	7	青霉素浓度/(f/L)
3	底物流速率/(L/h)	8	CO_2 浓度/(mmol/L)
4	底物流温度/K	9	反应热/kcal
5	溶解氧浓度/(mmol/L)		

在正常条件下,设置生产时间为 600 h,采样间隔为 1 h。过程正常运行到 200 h 时,在酸流速上分别设置两类

故障:1)引入幅度为6%的阶跃故障,如图8(a)中黑点部分;2)引入斜坡变化为0.02 L/h的故障。使用OLPP-

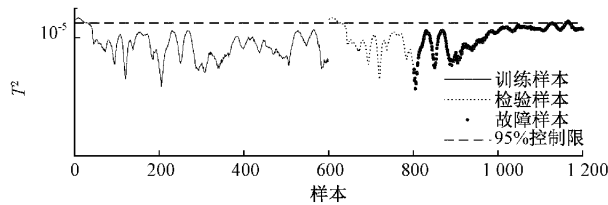
ATSWKNN、KPCA、OLPP、DPCA、WKNN和IDNPE方法对过程进行监控,结果如表2。

表2 青霉素发酵过程故障检测率

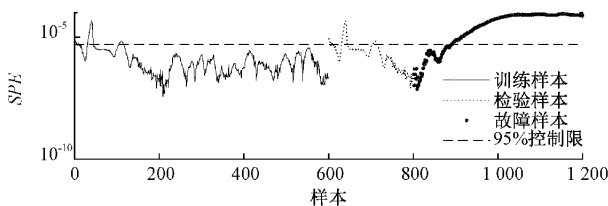
%

方法	KPCA		OLPP		DPCA		WKNN	IDNPE		OLPP-ATSWKNN
	T^2	SPE	T^2	SPE	T^2	SPE	D	T^2	SPE	WD
阶跃故障	1.3	0	0	0	0	0	14.0	0	7.0	98.2
斜坡故障	2.5	77.5	57.3	37.8	3.0	76.9	88.0	71.0	85.3	97.9

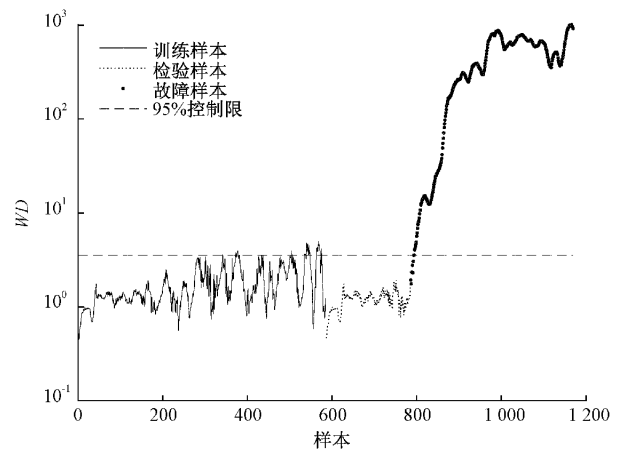
表2表明KPCA、OLPP和DPCA对阶跃故障的检测结果较差。此故障属于第2阶段的微弱故障,其统计值较小,导致故障被第1阶段的正常样本所淹没,因此故障难以被分离出来。对于斜坡故障3种方法的 T^2 检测率均较低,而 SPE 检测率较高。第1阶段的统计值高于第2阶段的统计值,导致第1阶段的统计值错误地决定了控制限。由于故障样本处于第2阶段且初始幅度微小,因此斜坡初始时期的部分故障难以检测出来。图9(a)为KPCA方法检测结果,DPCA和OLPP检测结果类似。此外,统计量不是稳态分布,仍具有自相关性。



(a) KPCA的检测结果



(b) WKNN的检测结果



(c) OLPP-ATSWKNN的检测结果

图9 斜坡故障的检测结果

由表2和图9(b)可知,WKNN方法对斜坡故障比较成功,但对阶跃故障不理想。该方法忽略了时序信息,错误地选取近邻样本,导致阶跃故障的检测率较低。IDNPE方法与WKNN的故障检测效果相似。该方法不仅考虑了样本的时序相关性,而且采用了包含局部结构信息的新距离,对于斜坡故障取得了较佳的监测性能。但在监控过程中IDNPE需要更新样本间距离,增加了系统的运行负担。

OLPP-ATSWKNN对两类故障的检测率均较高,检测结果如表2和图9(c)。此方法采用自适应时序窗口不仅考虑了时序信息,而且在转换阶段时自适应地切换了窗口类别。时序窗口将处于第2阶段的故障样本在时间上与第1阶段样本区分开,显著降低了统计量的自相关性。自适应切换窗口避免了阶段切换处对正常样本的误报,降低了误报率。此外,虽然两个阶段间和阶段内的样本离散程度不同,但是引入的权重将统计量统一到同一标准。实验结果表明时序窗口和统计量发挥了预期作用。将IDNPE和OLPP-STAWKNN对9个校验样本以及9个斜坡故障样本的检测时间作对比,如图10。由图可知本文方法对于故障识别具有较强的及时性。

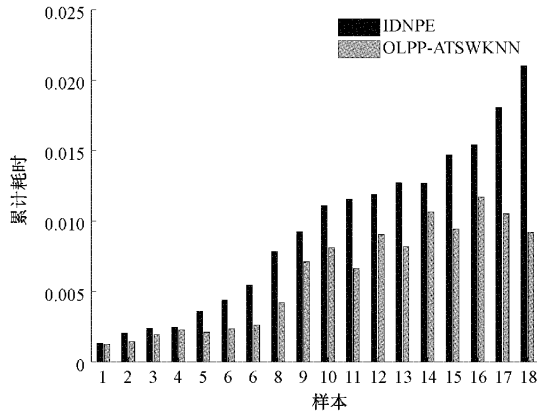


图 10 IDNPE 和 OLPP-ATSWKNN 检测累计耗时

4 结 论

针对工业生产过程监控中时序和多阶段的问题,本文提出一种基于正交局部保持投影-自适应时序窗口加权 k 近邻的故障检测方法。时序窗口中差分消除了过程的自相关特征;时序加权距离统一了多阶段的控制限;自适应技术保证实时采用合适的窗口类型。使用此方法对工业生产过程进行监控,能够更好地保障生产安全,有效提高生产效能。

参 考 文 献

- [1] 祁斌,赵灏. 基于神经网络的直流微电网故障诊断方法研究[J]. 国外电子测量技术, 2021, 40(4): 121-124.
- [2] 吴君凤. 基于孪生支持向量机的民用航空发动机故障诊断[J]. 电子测量技术, 2020, 43(22): 132-136.
- [3] 高学金, 马东阳, 韩华云, 等. 基于 DAE 和 TCN 的复杂工业过程故障预测[J]. 仪器仪表学报, 2021, 42(6): 140-151.
- [4] WANG J, SWARTZ C L E, CORBETT B, et al. Supply chain monitoring using principal component analysis [J]. Industrial & Engineering Chemistry Research, 2020, 59(27): 12487-12503.
- [5] 赵帅, 宋冰, 侍洪波. 基于加权互信息主元分析算法的质量相关故障检测 [J]. 化工学报, 2018, 69(3): 962-973.
- [6] 李元, 张昊展, 唐晓初. 基于多模态数据全信息的概率主成分分析故障检测研究[J]. 仪器仪表学报, 2021, 42(2): 75-85.
- [7] 李锦冰, 韩冰, 冯守渤, 等. 基于分块核主成分分析和支持向量机的故障检测 [J]. 控制理论与应用, 2020, 37(4): 847-854.
- [8] 刘光达, 董梦坤, 张守伟, 等. 基于 KPCA-SVM 的表面肌电信号疲劳分类研究 [J]. 电子测量与仪器学报, 2021, 35(10): 1-8.
- [9] 魏永合, 马步芳, 刘炜, 等. 基于 VMD 与正交局部保持

投影的齿轮故障诊断[J]. 组合机床与自动化加工技术, 2020(3): 3-6, 12.

- [10] KWON J, OH H S, LIM Y. Dynamic principal component analysis with missing values[J]. Journal of Applied Statistics, 2020, 47(9/12): 1957-1969.
- [11] 李明虎, 李钢, 钟麦英. 动态核主元分析在无人机故障诊断中的应用[J]. 山东大学学报(工学版), 2017, 47(5): 215-222.
- [12] ZHANG Q, LI P, LANG X, et al. Improved dynamic kernel principal component analysis for fault detection[J]. Measurement, 2020, 158: 1-11.
- [13] SHAO W M, GE Z Q, YAO L, et al. Bayesian nonlinear Gaussian mixture regression and its application to virtual sensing for multimode industrial processes [J]. IEEE Transactions on Automation Science and Engineering, 2020, 17(2): 871-885.
- [14] LIU J X, LIU T, CHEN J H. Sequential local-based Gaussian mixture model for monitoring multiphase batch processes [J]. Chemical Engineering Science, 2018, 181: 101-113.
- [15] 朱雄卓, 张瀚文, 杨春节. 基于高斯混合模型的 MWPCA 高炉异常监测算法 [J]. 化工学报, 2021, 72(3): 1539-1548.
- [16] SONG B, TAN S, SHI H B, et al. Fault detection and diagnosis via standardized k nearest neighbor for multimode process[J]. Journal of the Taiwan Institute of Chemical Engineers, 2020, 106: 1-8.
- [17] ZHU W B, SUN W, ROMAGNOLI J. Adaptive k nearest neighbor method for process monitoring [J]. Industrial & Engineering Chemistry Research, 2018, 57(7): 2574-2586.
- [18] CHEN J S, ZHANG X Y, ZHANG N, et al. Fault detection for turbine engine disk using adaptive Gaussian mixture model [J]. Proceedings of the Institution of Mechanical Engineers, 2017, 231(10): 827-835.
- [19] 冯立伟, 张成, 李元. 基于 PC-WKNN 的多工况间歇过程故障检测方法研究 [J]. 计算机应用研究, 2018, 35(4): 1130-1134.
- [20] SONG B, MA Y X, SHI H B. Multimode process monitoring using improved dynamic neighborhood preserving embedding [J]. Chemometrics & Intelligent Laboratory Systems, 2014, 135: 17-30.
- [21] ROLKE W, GONGORA C G. A chi square goodness of fit test for continuous distributions against a known alternative [J]. Computational Statistics, 2020, 36(3): 1885-1900.
- [22] ZHU J L, WANG Y Q, ZHOU D H, et al. Batch

process modeling and monitoring with local outlier factor[J]. IEEE Transactions on Control Systems Technology, 2019, 27(4): 1552-1565.

- [23] LIBOTTE G B, LOBATO F S, PLATT G M, et al. Robust multi-objective singular optimal control of penicillin fermentation process[J]. Global Journal of Researches in Engineering, 2020, 20(3): 23-30.
- [24] ABBASI M A, KHAN A Q, MUSTAFA G, et al. Data-driven fault diagnostics for industrial processes: An application to penicillin fermentation process[J]. IEEE Access, 2021, 9: 65977-65987.

作者简介

冯立伟, 硕士, 讲师, 主要研究方向为基于数据驱动的复杂过程故障检测。

E-mail: feng-li-wei@163.com

顾欢, 硕士, 主要研究方向为故障检测与诊断。

E-mail: 2293186473@qq.com

孙立文, 硕士, 主要研究方向为故障检测与诊断。

E-mail: 1165652254@qq.com

李元(通信作者), 博士, 教授, 主要研究方向为基于数据驱动的复杂过程故障检测。

E-mail: li-yuan@mail.tsinghua.edu.cn