

基于独立元的 k 近邻故障检测策略

张成^{1,2}, 高宪文^{1†}, 徐涛², 李元², 逢玉俊²

(1. 东北大学 信息科学与工程学院, 辽宁 沈阳 110819;

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: k 近邻故障检测(fault detection based on k nearest neighbors, FD- k NN)方法能够提高具有非线性和多模态特征过程的故障检测率. 由于系统故障通常由潜隐变量异常变化引起, 而该类型故障并不能被观测数据直观表现, 因此直接在观测变量上执行FD- k NN方法, 其故障检测率降低. 本文旨在提高FD- k NN方法针对潜隐变量故障的检测能力, 提出基于独立元的 k 近邻故障检测方法. 首先, 通过对观测数据应用独立元分析(independent component analysis, ICA)方法, 获得独立元矩阵; 接下来在独立元矩阵中应用FD- k NN方法进行故障检测. 这等同于直接监控过程潜隐变量的变化, 可以提高过程故障检测率. 通过非线性实例仿真实验, 证明本文方法检测潜隐变量故障是有效的; 同时, 在半导体蚀刻工艺过程的仿真实验中, 与主元分析(principal component analysis, PCA)方法、核主元分析(kernel principal component analysis, KPCA)方法、基于主元分析的 k 近邻故障检测(principal component-based k nearest neighbor rule for fault detection, PC- k NN)方法和FD- k NN方法进行对比, 实验结果进一步验证了本文方法的有效性.

关键词: k 近邻; 独立元分析; 主元分析; 故障检测; 间歇过程

引用格式: 张成, 高宪文, 徐涛, 等. 基于独立元的 k 近邻故障检测策略. 控制理论与应用, 2018, 35(6): 805 – 812

中图分类号: TP277 **文献标识码:** A

Fault detection strategy of independent component-based k nearest neighbor rule

ZHANG Cheng^{1,2}, GAO Xian-wen^{1†}, XU Tao², LI Yuan², PANG Yu-jun²

(1. College of Information Science and Engineering, Northeastern University, Shenyang Liaoning 110819, China;

2. Research Center for Technical Process Fault Diagnosis and Safety, Shenyang University of Chemical Technology, Shenyang Liaoning 110142, China)

Abstract: Fault detection based on k nearest neighbors (FD- k NN) method is able to improve the fault detection rate (FDR) in a process with nonlinear and multimode characteristic. Since some faults are caused by some abnormal change of latent variables and they are difficultly recognized through the observed variables, when FD- k NN is implemented directly in observed data set, its detection result is disappointed. Aiming to improve the fault detection ability of FD- k NN on abnormal change of latent variables, a k nearest neighbors fault detection strategy based on independent component analysis (ICA) is proposed in this paper. First, implement ICA in observed data set to obtain an IC matrix, in which all variables are independent. Then, the conventional FD- k NN is implemented to detect faults in the proposed IC matrix. When FD- k NN is implemented in IC matrix, it means that some latent variables are monitored by FD- k NN. Hence, the faults occurring on latent variables are able to be detected by FD- k NN. The efficiency of the proposed strategy is implemented in a simulated case and in the semiconductor manufacturing processes. The experimental results indicate that the proposed method outperforms PCA (principal component analysis), KPCA (kernel principal component analysis), k -nearest neighbor rule based on PCA (PC- k NN) and FD- k NN.

Key words: k nearest neighbors; independent component analysis; principal component analysis; fault detection; batch process

Citation: ZHANG Cheng, GAO Xianwen, XU Tao, et al. Fault detection strategy of independent component-based k nearest neighbor rule. *Control Theory & Applications*, 2018, 35(6): 805 – 812

收稿日期: 2017-06-09; 录用日期: 2017-12-12.

†通信作者. E-mail: gaonianwen@ise.neu.edu.cn; Tel: +86 18698865637.

本文责任编辑: 王鞞维.

国家自然科学基金项目(61573088, 61490701, 61673279), 辽宁省教育厅项目(LZ2015059, L2015432)资助.

Supported by the National Natural Science Foundation of China (61573088, 61490701, 61673279) and the Education Department Research Project of Liaoning Province (LZ2015059, L2015432).

1 引言(Introduction)

在现代工业生产过程中,产品质量与过程安全等因素促使过程故障检测与诊断技术迅速发展.由于分布式控制系统可以自动测量和存储大量的多变量采样数据,因此多变量统计过程监控方法已经成为当今监控领域研究的重点.

主元分析(principal component analysis, PCA)是一种典型的、基于数据驱动的多变量统计过程监控方法;已经被广泛地应用在化工生产过程并且取得了可喜的成果^[1-4].PCA通过正交变换将高维观测空间分解成主元子空间和残差子空间,并且分别应用Hotelling's T^2 和平方预测误差(squared prediction error, SPE)实现对上述两个子空间的过程监控.由于PCA本质上是一种静态的线性变换方法,当其被应用到非线性系统监控时,该方法通常会导致大量误报情况.为了解决非线性过程的监控问题,一些基于PCA的非线性方法已经被提出,同时获得了快速发展^[5-7].近些年,基于核理论的主元分析方法(kernel principal component analysis, KPCA)已经有效地被用于解决非线性系统的过程监控问题^[8-11].

上述方法,如PCA和KPCA,适用于监控单模态过程;将其应用于多模态过程通常存在较高的误报情况^[12].为了监视多模态过程,He等提出一种基于 k 最近邻故障检测(fault detection based on k nearest neighbors, FD- k NN)方法^[13].FD- k NN方法以当前样本与其前 k 最近邻样本距离之和作为统计指标来识别当前样本状态.该方法只依靠样本及其近邻的距离相似程度来监控过程状态,降低了多模态对过程监控的影响.然而,当样本具有较高维数时,FD- k NN需要频繁计算高维数据间的距离,这将产生大量的时间消耗,同时对存储空间也提出更高的要求.针对上述缺陷,He等提出一种基于PCA和FD- k NN相结合的检测(principal component-based k nearest neighbor rule for fault detection, PC- k NN)方法^[14].首先,应用PCA将数据集分成主元子空间和残差子空间;接下来,在主元子空间中执行FD- k NN方法.PC- k NN在低维主元子空间中计算样本间距离,这将降低距离计算的复杂度,提高计算效率.然而在监控过程中,检测结果的重要性高于计算效率.由于PCA并不是一种距离保持降维方法,也就是说,在降维过程中,故障点可能由于空间位置关系转变为正常点,这将导致检测结果的不准确性.综上,相对检测结果而言,PC- k NN不具有优于FD- k NN的检测性能.

FD- k NN方法应用当前样本与其近邻的欧式距离相似度量当前样本状态,因此变量的尺度变换是首要考虑的问题.通常FD- k NN与 Z -score标准化方法结合使用,通过 Z -score方法调节监控变量尺度来提高FD- k NN的检测性能.然而,观测变量的标准化处

理并不能实现对过程潜隐变量的同步统一标准化处理.这将导致分布集中的潜隐变量方向上的微弱故障,较难被FD- k NN方法检测出来.

近年来,独立元分析(independent component analysis, ICA)被引入过程监控^[15].ICA作为PCA的延伸,是从多维统计数据中发现隐含因素或成分的一种方法^[16-17].ICA区别于其他方法的特点是利用ICA方法得到的独立成分在统计上独立.ICA方法一方面充分利用了高阶统计量,另一方面可从观测数据中进一步提取出相互独立的潜隐变量,这些潜隐变量更能本质地反映过程特征.为了提高FD- k NN检测潜隐变量故障的能力,本文提出一种基于独立元的 k 最近邻故障检测(independent component-based k nearest neighbors rule for fault detection, FD- k NN)方法.首先,对观测数据进行独立元分析,获得独立元矩阵.接下来,在独立元矩阵中执行FD- k NN.本文的余下部分安排如下:第2节简要叙述ICA和FD- k NN方法;第3节给出基于独立元的 k 最近邻故障检测基本框架;在第4节中,给出关于非线性数值例子,说明本文方法的有效性;第5节给出本文方法在半导体蚀刻工艺过程中仿真应用,同时与PCA、KPCA、FD- k NN和PC- k NN等方法进行对比分析,进一步验证本文方法的有效性;第6节给出结论和相关展望.

2 独立元分析与 k 近邻故障检测方法 (Independent component analysis and fault detection based on k nearest neighbors)

2.1 独立元分析(Independent component analysis)

设 \boldsymbol{x} 为观测向量,包含 n 个观测变量.ICA将 \boldsymbol{x} 分解为

$$\boldsymbol{x} = \boldsymbol{s}A + \boldsymbol{e}, \quad (1)$$

其中: $\boldsymbol{s} \in \mathbb{R}^{1 \times r}$ 为独立元向量; $A \in \mathbb{R}^{r \times n}$ 为未知混合矩阵; $\boldsymbol{e} \in \mathbb{R}^{1 \times n}$ 为误差向量,且 $r \leq n$.ICA方法是由观测向量 \boldsymbol{x} 估计 A 和 \boldsymbol{s} ,进而得到分离矩阵 W 和重构独立元向量 $\hat{\boldsymbol{s}}$,使其满足

$$\hat{\boldsymbol{s}} = \boldsymbol{x}W. \quad (2)$$

一般进行ICA算法之前,先要对数据进行中心化和白化预处理,以去除观测变量的相关性,使预处理后的观测变量的协方差阵为单位阵.

1) 中心化.对观测变量按照式(3)进行中心化处理,

$$x_i = x_i - \mu_i, \quad (3)$$

其中 μ_i 为变量 x_i 均值.

2) 白化.首先,对中心化向量 \boldsymbol{x} 的协方差矩阵进行奇异值分解:

$$E\{\boldsymbol{x}^T \boldsymbol{x}\} = UAU^T, \quad (4)$$

其中: A 为 \boldsymbol{x} 的协方差矩阵的特征值对角阵, U 为与特征值对应的特征向量矩阵.

接下来, 选取

$$Q = UA^{-\frac{1}{2}}, \quad (5)$$

$$z = xQ, \quad (6)$$

则 z 的协方差矩阵满足 $E\{z^T z\} = I$, 称 z 为白化向量.

由式(1)和式(6)可得, $z = xQ = sAQ = sB$, 其中 $B = AQ$ 为分离矩阵且为正交矩阵. 于是, 得到独立元 s 估计值

$$\hat{s} = xQB^T. \quad (7)$$

由式(2)和式(7)可知 B 与 W 满足以下关系:

$$W = QB^T. \quad (8)$$

文献[18]给出获取独立元的算法FastICA. FastICA是一种快速寻优迭代算法; 基于负熵最大的FastICA算法以负熵最大作为一个搜寻方向, 可以实现顺序地提取独立元. 文献[19]提出依据分离矩阵 W 中行的Euclidian范数(L_2)贡献确定独立元个数的方法.

2.2 k 近邻故障检测方法 (Fault detection based on k nearest neighbors)

FD- k NN方法的基本原理是使用样本与其近邻的距离度量样本之间的差异. 首先, 在训练样本空间中查找样本 x 的前 k 个近邻 $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$. 其次, 计算 x 与其前 k 个近邻样本的距离平方和:

$$D^2 = \sum_{i=1}^k d_i^2, \quad (9)$$

其中 d_i 为样本 x 与其第 i 近邻 $x^{(i)}$ 的欧式距离. 最后, 根据非中心的卡方分布, 确定置信度为 α 的检测控制限 D_α^2 [13]. 当待检测样本的 D^2 大于控制限 D_α^2 时判定其为故障样本, 否则判定其为正常样本. FD- k NN具体计算过程如下:

离线建模过程包含以下3步:

第1步 在训练样本集 X 中查找每个样本 x 的前 k 个近邻样本;

第2步 应用式(9)计算 x 的 k 近邻距离 D^2 ;

第3步 确定训练模型的控制限 D_α^2 .

同样, 在线检测过程也包含3步:

第1步 在训练数据集 X 中寻找待检测样本 x_* 的前 k 个近邻样本;

第2步 利用式(9)计算 x_* 的 k 近邻距离 D_*^2 ;

第3步 将 D_*^2 与控制限 D_α^2 比较: 若 $D_*^2 > D_\alpha^2$, 判定 x_* 为故障, 否则 x_* 为正常.

3 基于独立元的 k 近邻故障检测方法 (Independent component-based k nearest neighbors rule for fault detection)

FD- k NN是一种有效的多变量过程监控方法, 已经成功地被应用于非线性和多模态系统的过程监视

中, 并取得了令人满意的结果. 由于变量尺度与量纲差异影响, 通常在过程监控之前对采样数据需要进行有效的标准化处理. Z -score是一种常用的数据预处理方法, 通过调整变量的均值与方差, 能够使不同量纲变量在同一尺度下进行比较. 因此, 传统方法中通常将 Z -score与监控方法结合使用, 提高经典算法对过程的监控能力, 例如FD- k NN, PCA和KPCA等.

Z -score与FD- k NN相结合的方法对变量相互独立的过程进行监控是有效的. 例如, 图1(a)中变量 x 和 y 相互独立且服从二元高斯分布, 均值与标准差分别为 $[0, 0]$ 和 $[5, 1]$. 点 f_1 和 f_2 为过程离群点, 分别沿变量 x 和 y 方向偏离数据中心4倍标准差的距离. 经过 Z -score处理后的数据散点图如图1(b)所示, 变量 x 和 y 被调整为同一尺度, 同时离群点 f_1 和 f_2 的故障程度近似相同. 在经过标准化处理后的数据集上应用FD- k NN方法, 可以有效检测出以上两个离群点, 同时可以统计出两个故障点的故障程度近似相同. 如果没有进行 Z -score的数据预处理过程, 直接在图1(a)所示数据集中应用FD- k NN, 在计算距离过程中, 变量 x 对距离的贡献往往会淹没变量 y 相应的贡献, 故离群点 f_2 将会被误报为正常点, 降低了FD- k NN方法的检测性能.

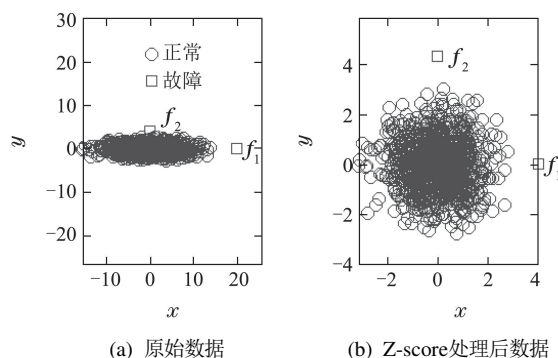


图1 说明例子

Fig. 1 Illustration case

图1中 x 和 y 同时为观测变量和潜隐变量, 对观测变量进行标准化处理等同于对潜隐变量执行相应的处理方法. 对具有相关性的观测变量进行 Z -score处理通常并不能实现对潜隐变量尺度的同步调整. 例如观测变量 x 和 y 与潜隐变量 s 和 t 满足以下关系:

$$\begin{cases} x = s - t, \\ y = s + t, \end{cases}$$

如果对观测变量 x 和 y 进行 Z -score处理, 显然并不等同于对潜隐变量 s 和 t 进行了 Z -score处理.

为了提高FD- k NN故障检测率, 本节提出一种基于独立元的 k 近邻故障检测方法. 首先, 应用ICA将观测矩阵进行独立化处理, 获取过程潜隐变量; 其次在独立元矩阵中执行FD- k NN故障检测方法, 进行过程监控.

FD-IkNN方法分为模型建立和在线检测两个阶段, 具体流程如下:

离线建模过程包含以下4步:

第1步 应用式(3)和式(6), 中心化和白化训练数据集 X ;

第2步 计算分离矩阵 W ;

第3步 由式(7)计算训练数据独立元矩阵 \hat{S} ;

第4步 在数据集 \hat{S} 上执行FD-kNN方法, 确定显著性水平为 α 的检测控制限 D_α^2 ;

同样, 在线检测过程包含以下4步:

对待检测样本 \mathbf{x}_* ,

第1步 应用式(3)和(6), 中心化和白化样本 \mathbf{x}_* ;

第2步 应用式(7), 计算样本 \mathbf{x}_* 的独立元向量 $\hat{\mathbf{s}}_*$;

第3步 依据式(9), 计算样本 $\hat{\mathbf{s}}_*$ 在 \hat{S} 中 k 近邻距离平方和 D_*^2 ;

第4步 比较 D_*^2 和 D_α^2 : 若 $D_*^2 > D_\alpha^2$, \mathbf{x}_* 为故障样本; 否则, \mathbf{x}_* 为正常样本.

FD-IkNN方法通过对训练数据进行独立性分析, 能够消除过程变量的相关性, 降低变量不同尺度变化对样本与其近邻距离计算的影响, 提高过程故障检测效率.

4 数值实例(Numerical case)

在本例将通过一个包含4个观测变量的非线性实例验证本文方法的有效性, 具体模型如式(10)所示.

$$\begin{cases} x_1 = 5s - t, \\ x_2 = 3s + t, \\ x_3 = e^{0.2s} + t^2, \\ x_4 = s^5, \end{cases} \quad (10)$$

其中潜隐变量 s, t 分别在区间 $[-50, 50]$ 和 $[0, 1]$ 上服从均匀分布.

依据式(10), 随机生成500个样本构成训练集 $X_{500 \times 4}$; 随机生成100个样本用于校验本文方法性能. 通过人工调整潜隐变量 t 初始值, 设置两个阶段故障, 共包含200个故障点. 对于训练、校验和故障数据的潜隐变量 s, t 分布如图2所示.

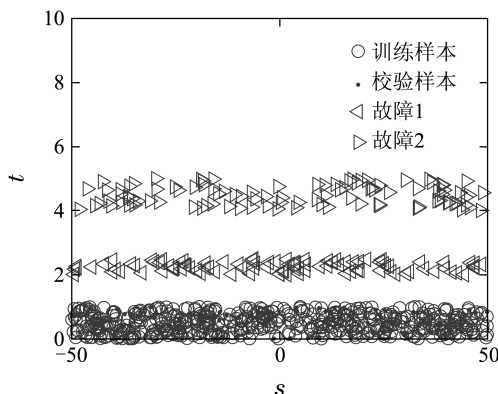


图2 潜隐变量 s, t 散点图

Fig. 2 Scatter diagram of s, t

首先, 传统的PCA方法在本例中被执行. 按照85%方差累计贡献率, 2个主元被选取用于过程监视, T^2 和SPE检测结果如图3所示. 由于系统具有非线性结构, 这导致 T^2 和SPE两种监控统计均不能有效检测出系统故障.

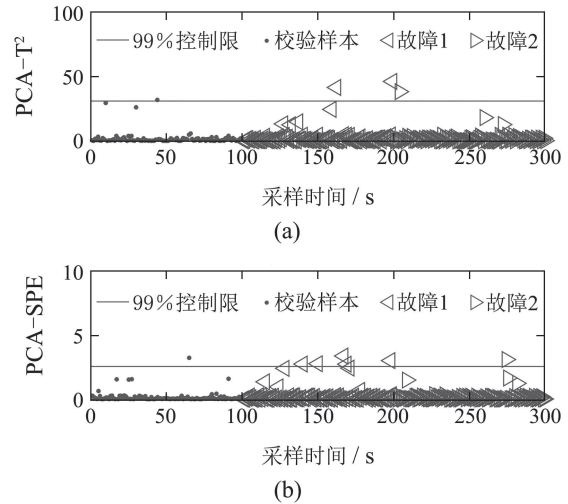


图3 T^2 和SPE控制图(PCA)

Fig. 3 Control charts of T^2 and APE in PCA

接下来, 应用FD-kNN方法在观测数据集上进行故障检测. 在执行FD-kNN之前, 采用Z-score方法对数据集进行标准化预处理. 对训练数据建模分析及试验参数寻优, 设置参数 $k = 3$, FD-kNN的检测结果如图4所示. 容易看出, 由潜隐变量 t 变化引起的故障不能被FD-kNN有效检测. 由于Z-score方法只是对观测数据进行标准化处理, 调整了观测变量的尺度变化, 系统的相关性仍然保持在标准化的数据集中, 上述原因导致故障不能被FD-kNN检测出来.

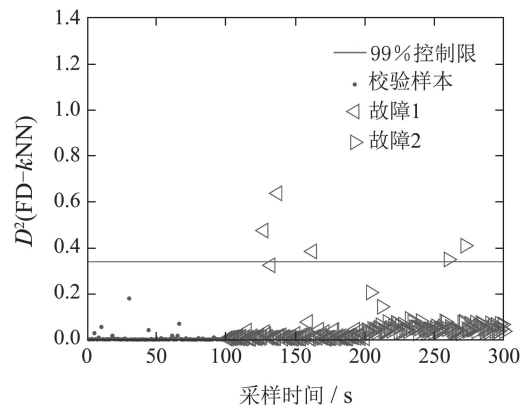


图4 D^2 控制图(FD-kNN)

Fig. 4 Control chart of D^2 in FD-kNN

最后, FD-IkNN方法在本例中被执行检测故障. 首先对观测数据集进行独立性分析, 计算独立元矩阵, 其中依据Euclidian范数(L_2)贡献方法, 选取独立元个数为1. 然后, 执行FD-kNN方法, 其中 $k = 3$, 检测结果如图5所示. 可以看出本文方法能够准确识别出两

类故障; 同时, 可由控制图分析, 两类故障具有不同的故障尺度, 这与由两个潜隐变量设置故障的过程相一致, 进一步证明了本文方法的有效性. 图6给出FD- k NN和FD- Ik NN两种方法针对本例中300个测试样本点在线监控的时间统计, 可以看出本文方法较FD- k NN方法能够更加及时准确检测出过程故障.

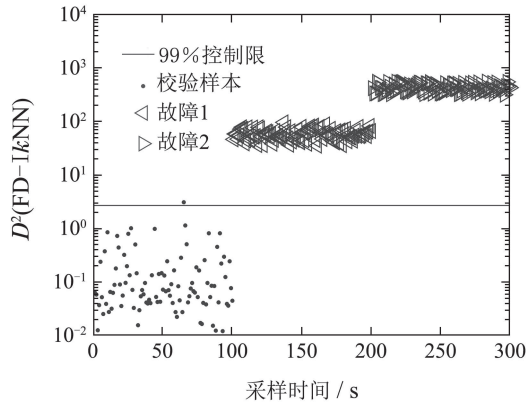


图 5 D^2 控制图(FD- Ik NN)

Fig. 5 Control chart of D^2 in FD- Ik NN

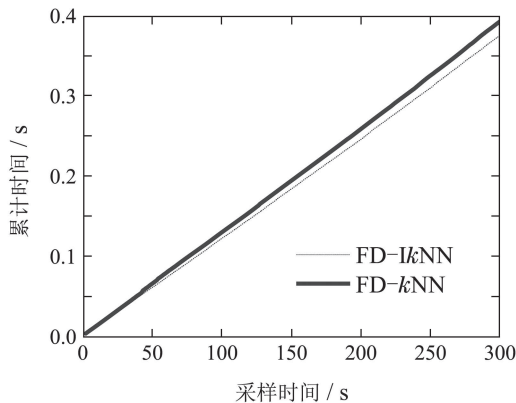


图 6 故障检测累积时间

Fig. 6 Accumulated time of fault detection

5 工业实例(Industry case)

数据采自半导体铝堆蚀刻工艺过程, 包括来自3个不同试验的晶片蚀刻数据: 107个正常晶片, 21个故障晶片^[20]. 故障变量名称及其参数设置可以参考文献^[21]. 原始数据包括40个变量, 本文只使用其中的19个变量监控生产状态, 这些变量与产品的生产过程和最终状态密切相关^[21].

5.1 数据分析(Data analysis)

半导体蚀刻工艺数据采自间歇生产过程, 蚀刻工艺具有如下特点^[21]: 批次宽度不等长: 在107个正常批次中, 批次持续生产时间由95 s变化到112 s; 工序宽度不等长: 各批次数据采自蚀刻过程第四步和第五步, 不同批次中的第4步铝蚀刻持续的时间不同, 由44 s变化至52 s; 进程漂移: 对于蚀刻过程, 由于材料不同、蚀刻范围不同等原因导致进程轨迹漂移. 数据集除具有不等长特点外还具有三向结构(批次 $I \times$ 变量

$J \times$ 时间 K), 如图7(a)所示. 本文方法FD- Ik NN不能直接应用于原始数据集(图7(a))进行故障检测, 因此在进行故障检测之前, 进行数据展开处理是必要的. 文献[22-23]提出一种基于统计模量的间歇过程数据预处理方法. 在本例中, 统计模量方法首先被执行对批次数据集进行展开处理, 具体过程如下. 对于一个包含 m 个采样时刻和 n 个监视变量的批次数据集, 首先分别计算每个变量的均值与方差; 然后, 由所有变量的均值与方差将当前批次数据映射成一个统计模量向量 $sp = [c, v]$, 其中 c 和 v 分别是均值向量和方差向量, 具体过程如图7(a)和图7(b)所示.

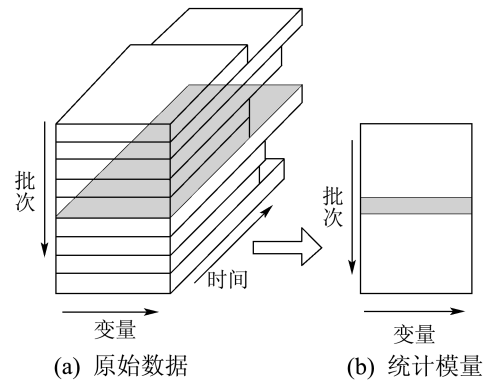


图 7 统计模量框架

Fig. 7 SPA framework

5.2 过程检测结果与分析(Result and analysis)

为了验证本文方法在间歇生产过程中故障检测的有效性, 首先在正常数据集中随机选择101个批次数据构成训练集, 余下6个批次数据构成校验数据集; 其次, 传统的PCA, KPCA, FD- k NN和PC- k NN方法依次被执行, 与本文方法进行对比分析. PCA和KPCA的主元数均通过85%的方差累计贡献率进行确定; 在KPCA中, 核函数选用高斯核函数

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{c}\right),$$

其中窗宽参数 c 通过文献[24]中的交叉验证方法确定; FD- k NN、PC- k NN和FD- Ik NN中近邻数 k 通过多次试验寻优确定; 各种方法控制限均采用核密度估计方法确定, 阈值设置为99%^[25-26]. 各种参数详细设置见表1.

表 1 参数设置

Table 1 Parameter setting in different methods

	主元数PCs	近邻数 k	窗宽参数 c
PCA	14	—	—
KPCA	15	—	500
FD- k NN	—	3	—
PC- k NN	14	3	—
FD- Ik NN	5	3	—

PCA和KPCA方法在主元子空间和残差子空间分别采用 T^2 和SPE统计量进行过程监控. 在99%控制限标准下, PCA的 T^2 和SPE的检测率分别为14.3%和80.9%; KPCA的 T^2 和SPE的检测率分别为4.76%和85.71%, 检测结果见图8和9. 本例中, 半导体蚀刻工艺数据集具有多模态结构, 这一特点将制约单模型方法的检测性能, 所以PCA和KPCA在本例中具有较低检测率.

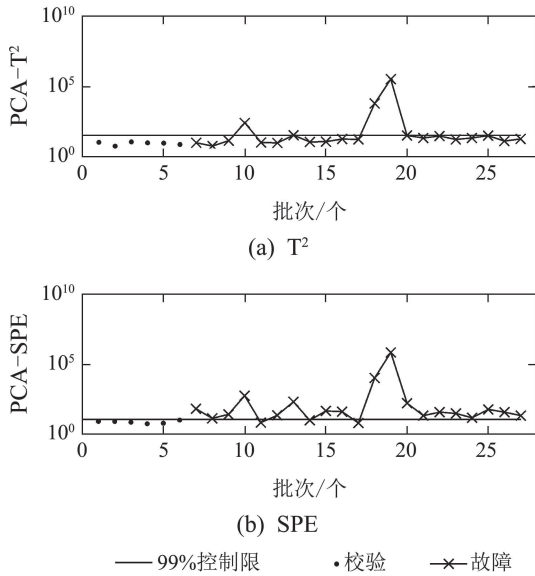


图8 PCA故障检测结果

Fig. 8 Fault detection results of PCA

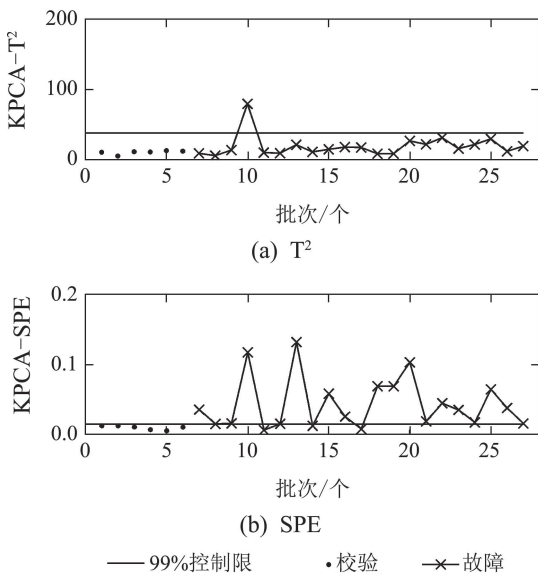


图9 KPCA故障检测结果

Fig. 9 Fault detection results of KPCA

FD- k NN和FD- lk NN方法检测结果由图10给出. 可以看出FD- lk NN具有优于FD- k NN的检测性能. FD- k NN虽然能够有效处理具有非线性和多模态特征的故障检测问题, 但是通常需要过程变量服从独立高斯分布. 在FD- k NN方法前执行 Z -score方法对观测变量进行标准化处理, 所得被标准化数据仍然具有相

关性; 同时 Z -score对观测数据标准化并不等同于标准化潜在变量的尺度, 因此, FD- k NN对分布密集的潜在变量的微弱故障识别能力仍然较低. 由图10(b)可以看出, 相比其他方法, FD- lk NN具有最高的检测性能, 故障检测率达到100%. 通过本例实验可知, 故障的发生通常较难反映在观测变量上; 换句话说, 观测变量通常是潜在变量的综合体现; 独立潜在变量的变化通常是故障产生的根源. 本文方法FD- lk NN正是对独立的潜在变量进行监控, 提升了对过程故障的检测能力. PC- k NN方法是指在主元子空间运行FD- k NN方法. 本例针对PC- k NN方法也进行了测试: 总共4个故障点被检测出来, 结果如图11所示. PCA方法是依靠原始数据的线性变换寻找主元子空间, 这种线性变换并不能使得样本与其近邻的距离得到保持; 换句话说, 这种线性变换可能使得本来为故障的样本变换成正常样本.

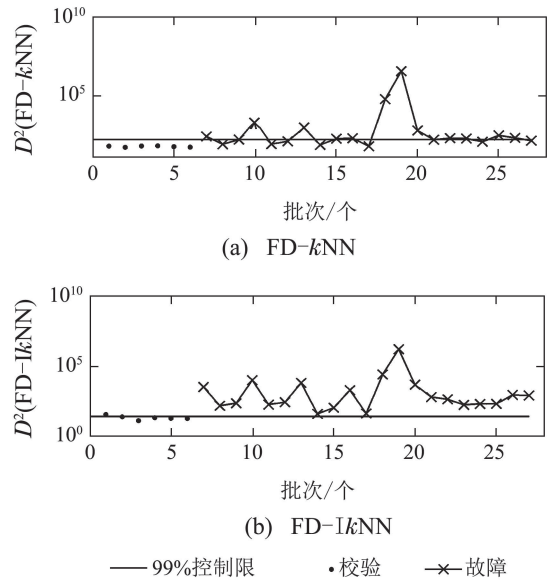


图10 FD- k NN and FD- lk NN检测结果

Fig. 10 Fault detection results of FD- k NN and FD- lk NN

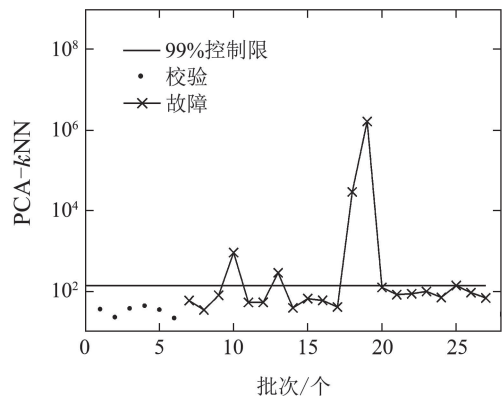


图11 PC- k NN故障检测结果

Fig. 11 Fault detection result of PC- k NN

综上, PC- k NN方法不具有优于FD- k NN的检测

性能. 从计算效率来看, 在FD- k NN中5个独立元被选取, 相比包含14个主元的PC- k NN而言, FD- k NN具有更高的计算效率. 图12给出FD- k NN, FD- k NN和PC- k NN3种方法在本试验中的累积时间统计图. 由图12可以看出, 本文方法具有最高的计算效率, 这为过程故障检测的实时性提供了保障.

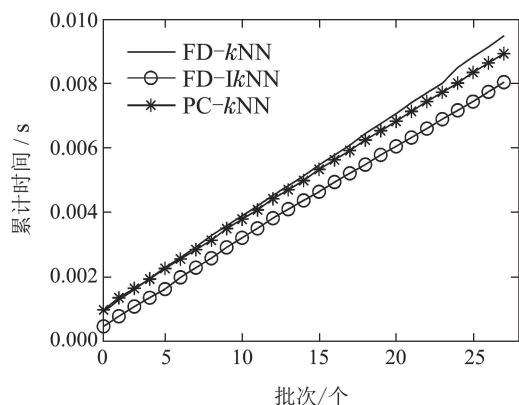


图12 故障检测累积时间

Fig. 12 Accumulated time of fault detection

6 结论(Conclusions)

本文为提高FD- k NN方法对潜隐变量故障的检测率, 提出基于独立元分析的 k 近邻故障检测方法. 通过一个模拟实例和一个工业实例测试并与传统监控方法PCA, KPCA, FD- k NN和PC- k NN进行对比, 实验结果验证了本文方法的有效性. 本文方法从检测性能上分析优于FD- k NN, 但是其具有较高的计算复杂性. 因此, 降低复杂的计算过程是本文方法的下一步研究问题. 同时, 关于近邻数 k 的确定方法也是一个开放研究的问题, 本文是通过交叉验证的方法确定近邻数, 确定过程较为复杂. 接下来在FD- k NN方法中将重点研究近邻数的确定问题.

参考文献(References):

- [1] ZHOU Donghua, LI Gang, LI Yuan. *Data driven Based Process Fault Detection and Diagnosis Technology* [M]. Beijing: Science Press, 2011.
(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 [M]. 北京: 科学出版社, 2011.)
- [2] SHAMS M A B, BUDMAN H M, DUEVER T A. Fault detection, identification and diagnosis using CUSUM based PCA [J]. *Chemical Engineering Science*, 2011, 66(20): 4488 – 4498.
- [3] QI Yongsheng, WANG Pu, GAO Xuejin. Fault detection and diagnosis of multiphase batch process based on kernel principal component analysis-principal component analysis [J]. *Control Theory & Applications*, 2012, 29(6): 69 – 79.
(齐咏生, 王普, 高学金. 基于核主元分析-主元分析的多阶段间歇过程故障监测与诊断 [J]. 控制理论与应用, 2012, 29(6): 69 – 79.)
- [4] CONG Ya, GE Zhiqiang, SONG Zhihuan. Multi-rate principle component analysis for process monitoring [J]. *Journal of Shanghai Jiaotong University*, 2015, 49(6): 762 – 767.
(丛亚, 葛志强, 宋执环. 多采样率主元分析的过程故障检测 [J]. 上海交通大学学报, 2015, 49(6): 762 – 767.)
- [5] JD L. Nonlinear principal component analysis [J]. *Artificial Intelligence Methods in the Environmental Sciences*, 2011, 20(4): 173 – 190.
- [6] LI N, YANG Y. Ensemble kernel principal component analysis for improved nonlinear process monitoring [J]. *Industrial & Engineering Chemistry Research*, 2015, 54(1): 318 – 329.
- [7] JIA F, MARTIN E B, MORRIS A J. Non-linear principal components analysis for process fault detection [J]. *Computers & Chemical Engineering*, 1998, 22(12): S851 – S854.
- [8] LEE J M, YOO C K, SANG W C, et al. Nonlinear process monitoring using kernel principal component analysis [J]. *Chemical Engineering Science*, 2004, 59(1): 223 – 234.
- [9] WANG H, YAO M. Fault detection of batch processes based on multivariate functional kernel principal component analysis [J]. *Chemometrics & Intelligent Laboratory Systems*, 2015, 149: 78 – 89.
- [10] ZHANG Y. Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM [J]. *Chemical Engineering Science*, 2009, 64(5): 801 – 811.
- [11] ZhANG Jialiang, CAO Jianfu, GAO Feng, et al. Fault diagnosis of complex system based on nonlinear spectrum and kernel principal component analysis [J]. *Control Theory & Applications*, 2012, 29(12): 1558 – 1564.
(张家良, 曹建福, 高峰, 等. 结合非线性频谱与核主元分析的复杂系统故障诊断方法 [J]. 控制理论与应用, 2012, 29(12): 1558 – 1564.)
- [12] WANG G, LIU J, ZHANG Y, et al. A novel multimode data processing method and its application in industrial process monitoring [J]. *Journal of Chemometrics*, 2015, 29(2): 126 – 138.
- [13] HE Q P, WANG J. Fault detection using the k -nearest neighbor rule for semiconductor manufacturing processes [J]. *IEEE Transactions on Semiconductor Manufacturing*, 2007, 20(4): 345 – 354.
- [14] HE Q P, WANG J. Large scale semiconductor process fault detection using a fast pattern recognition based method [J]. *IEEE Transactions on Semiconductor Manufacturing*, 2010, 23(2): 194 – 200.
- [15] HYVÄRINEN A, HURRI J, HOYER P O. Independent component analysis [J]. *IEEE Transactions on Neural Networks*, 2004, 15(2): 529 – 529.
- [16] GAWEHNS D, WILDERJANS T F, PUTTEN C M V. The statistical analysis of neuronal data: comparing algorithms for independent component analysis [J]. *Isme Journal*, 2016, 2(7): 317 – 336.
- [17] ZHANG Y, AN J, ZHANG H. Monitoring of time-varying processes using kernel independent component analysis [J]. *Chemical Engineering Science*, 2013, 88(2): 23 – 32.
- [18] HYVÄRINEN A, OJA E. A fast fixed-point algorithm for independent component analysis [J]. *International Journal of Neural Systems*, 2000, 10(1): 1 – 8.
- [19] ZHAO Zhonggai, LIU Fei. Nonlinear process monitoring method based on kernel independent component analysis [J]. *Journal of System Simulation*, 2008, 20(20): 5585 – 5588.
(赵忠盖, 刘飞. 一种基于核独立元分析的非线性过程监控方法 [J]. 系统仿真学报, 2008, 20(20): 5585 – 5588.)

- [20] ZhOU Z, WEN C, YANG C. Fault detection using random projections and k -nearest neighbor rule for semiconductor manufacturing processes [J]. *IEEE Transactions on Semiconductor Manufacturing*, 2015, 28(1): 70 – 79.
- [21] WISE B M, GALLAGHER N B, BUTLER S W, et al. A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process [J]. *Journal of Chemometrics*, 1999, 13(3/4): 379 – 396.
- [22] HE Q P, WANG J. Statistics pattern analysis: a new process monitoring framework and its application to semiconductor batch processes [J]. *AICHE Journal*, 2015, 57(1): 107 – 121.
- [23] ZHANG Cheng, LI Yuan. Study on the fault-detection method in batch process based on statistical pattern analysis [J]. *Chinese Journal of Scientific Instrument*, 2013, 34(9): 2103 – 2110.
(张成, 李元. 基于统计模量分析间歇过程故障检测方法研究 [J]. 仪器仪表学报, 2013, 34(9): 2103 – 2110.)
- [24] THOMAS M, DE B K, DE M B. New bandwidth selection criterion for kernel PCA: approach to dimensionality reduction and classification problems [J]. *BMC Bioinformatics*, 2014, 15(1): 137.
- [25] MA H, HU Y, SHI H. Fault detection and identification based on the neighborhood standardized local outlier factor method [J]. *Industrial & Engineering Chemistry Research*, 2013, 52(6): 2389 – 2402.
- [26] DEHNAD K. Density estimation for statistics and data analysis [J]. *Technometrics*, 1986, 29(4): 296 – 297.

作者简介:

张成 (1979–), 男, 副教授, 博士研究生, 研究方向为复杂工业过程故障诊断, E-mail: zcgg_2005@126.com;

高宪文 (1955–), 男, 教授, 博士生导师, 研究方向为复杂工业生产过程智能建模等, E-mail: gaioxianwen@ise.neu.edu.cn;

徐涛 (1980–), 女, 讲师, 硕士, 研究方向为复杂工业过程故障诊断, E-mail: xvtao_2005@126.com;

李元 (1964–), 女, 教授, 博士生导师, 研究方向为复杂工业过程故障诊断, E-mail: li-yuan@mail.tsinghua.edu.cn;

逢玉俊 (1954–), 男, 教授, 研究方向为复杂工业过程故障诊断, E-mail: Pangyujun@syuct.edu.cn.