

# 基于局部保持嵌入- $K$ 近邻比率密度的半导体蚀刻过程故障诊断策略

张 成, 郑百顺, 郭青秀, 冯立伟, 李 元<sup>†</sup>

(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:** 针对多模态间歇过程存在数据维度高且方差差异较大的特征, 提出一种基于局部保持嵌入- $K$ 近邻比率密度(NPE- $K$ RD)规则的故障检测方法. 首先, 利用局部保持嵌入(NPE)方法将原始的高维数据投影到低维空间; 其次, 在低维空间通过计算样本的密度及其前 $K$ 近邻密度的均值来建立 $K$ 近邻比率密度( $K$ RD); 最后, 根据核密度估计法确定统计量控制限并进行故障诊断. NPE方法既能够在低维空间保持数据局部近邻结构, 又能够降低故障检测过程的计算复杂度. 通过引入比率密度, NPE- $K$ RD可以降低多模态方差结构差异对故障检测的影响, 提高过程故障检测率. 通过数值例子和半导体工业过程的仿真实验, 并与主元分析、 $K$ 近邻、局部保持嵌入等方法进行比较, 验证了本文方法的有效性.

**关键词:** 局部保持嵌入;  $K$ 近邻比率密度; 半导体蚀刻过程; 故障诊断; 多工况

**引用格式:** 张成, 郑百顺, 郭青秀, 等. 基于局部保持嵌入- $K$ 近邻比率密度的半导体蚀刻过程故障诊断策略. 控制理论与应用, 2020, 37(6): 1342 – 1348

DOI: 10.7641/CTA.2019.90271

## A novel fault diagnosis strategy based on neighborhood preserving embedding- $K$ nearest neighbor ratio density in semiconductor etching processes

ZHANG Cheng, ZHENG Bai-shun, GUO Qing-xiu, FENG Li-wei, LI Yuan<sup>†</sup>

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

**Abstract:** Aiming at characteristics with high dimensions and large differences in variance of multimodal batch process, a fault detection method based on neighborhood preserving embedding- $K$  nearest neighbor ratio density (NPE- $K$ RD) rule is proposed. Firstly, the raw high dimensional data are projected into a low dimensional space using neighborhood preserving embedding (NPE). Secondly,  $K$  nearest neighbor ratio density ( $K$ RD) is established by calculating the density of the sample and the mean of its  $K$  nearest neighbor in the low dimensional space. Finally, the control limit of statistics is determined by kernel density estimation method and fault diagnosis is carried out. NPE can not only maintain the local neighbor structure of data in the low dimensional space, but also reduce the computational complexity of the fault detection process. By introducing ratio density, NPE- $K$ RD can reduce the influence of multimode variance structure difference on fault detection and improve the fault detection rate of processes. Compared with principal component analysis,  $K$  nearest neighbor and neighborhood preserving embedding, the effectiveness of the proposed method is verified in a numerical cases and semiconductor industrial processes.

**Key words:** neighborhood preserving embedding;  $K$  nearest neighbor ratio density; semiconductor etching processes; fault diagnosis; multimode

**Citation:** ZHANG Cheng, ZHENG Baishun, GUO Qingxiu, et al. A novel fault diagnosis strategy based on neighborhood preserving embedding- $K$  nearest neighbor ratio density in semiconductor etching processes. *Control Theory & Applications*, 2020, 37(6): 1342 – 1348

收稿日期: 2019-04-19; 录用日期: 2019-12-26.

<sup>†</sup>通信作者. E-mail: li-yuan@mail.tsinghua.edu.cn; Tel.: +86 13082424115.

本文责任编辑: 陈增强.

国家自然科学基金项目(61490701, 61673279), 辽宁省自然科学基金项目(2019-MS-262), 辽宁省教育厅基金项目(LJ2019013)资助.

Supported by the National Natural Science Foundation of China (61490701, 61673279), the Liaoning Natural Science Fund Project (2019-MS-262) and the Liaoning Provincial Department of Education Fund Project (LJ2019013).

## 1 引言

随着现代工业的发展趋势逐渐转向大型化、复杂化,工业系统的安全性和可靠性成为亟待解决的问题。为了有效监控过程健康状态,近年来,深度学习算法及基于数据驱动的多元统计方法已经被广泛应用于生产过程故障诊断领域<sup>[1-4]</sup>。

基于数据驱动的主元分析(principle component analysis, PCA)是一种经典的线性降维算法<sup>[5]</sup>。近年来,以PCA为基础的各类衍生方法相继被提出,如动态PCA(dynamic PCA, DPCA)、贝叶斯融合动态PCA(integrated dynamic PCA, IDPCA)等<sup>[6-7]</sup>。通常基于PCA的一系列方法所获得的低维数据包含原始数据更多有效信息。为了在低维空间中保持原始数据的近邻结构,多种经典的非线性流形算法被提出,如拉普拉斯特征映射(laplacian eigenmaps, LE)、局部线性嵌入(locally linear embedding, LLE)和等距特征映射(isometric feature mapping, ISOMAP)等<sup>[8-10]</sup>。上述方法能够有效提取数据的流形结构,但是它们的非线性性质使其计算量较大<sup>[11]</sup>。因此,He等根据此类缺陷提出了局部保持嵌入(neighborhood preserving embedding, NPE)的算法<sup>[11]</sup>。与PCA相比较,NPE作为一种线性降维算法,能够在低维空间描述原始数据的局部流形结构,对异常值的敏感度低于PCA且适合实际应用<sup>[11]</sup>。然而,NPE与PCA同样应用 $T^2$ 和SPE统计量监控样本状态,文献[12]指出对于多模态和非线性过程,上述统计量不能有效完成状态监控。因此,NPE在监控多模态过程故障时不够完善<sup>[13]</sup>。

针对上述方法的局限性,He等提出的 $K$ 近邻( $K$  nearest neighbor, KNN)方法被应用到多模态过程故障检测中<sup>[12]</sup>。KNN通过寻找数据近邻特征,有效降低了非线性和多模态特征对故障检测的影响,提高过程故障检测率。然而,KNN方法受方差差异明显的多模态影响会导致故障漏报<sup>[14-15]</sup>。为了提高KNN方法的故障检测率,多种改进KNN方法被提出,但是此类方法都面临耗时且计算复杂度高的问题<sup>[16-17]</sup>。

针对方差差异显著的多模态间歇过程故障检测问题,本文提出局部保持嵌入-K近邻比率密度(neighborhood preserving embedding-K nearest neighbor ratio density, NPE-KRD)规则的故障检测方法。首先,利用NPE计算负载矩阵将高维数据投影到低维空间;其次,在低维空间计算每个数据点的 $K$ 近邻比率密度;最后,根据样本的 $K$ 近邻比率密度统计值进行故障检测。NPE-KRD不但继承NPE保持原始数据局部几何结构的优势,而且通过KRD方法能够有效处理以上多模态过程的故障检测问题。

## 2 局部保持嵌入(NPE)

假设原始数据集 $\mathbf{X}$ 含有 $m$ 个样本, $n$ 个变量。NPE

通过计算负载矩阵 $\mathbf{A}$ ,将高维空间中的 $m$ 个样本点投影到低维空间,得到新的数据集 $\mathbf{Y}^{m \times d}$ ,其中 $d$ 为数据降维后变量数。

**第1步** 通过优化式(1),计算权重矩阵 $\mathbf{W}$ :

$$\begin{cases} \min \sum_{i=1}^m \|\mathbf{x}_i - \sum_{j=1}^K w_{ij} \mathbf{x}_i^j\|^2, \\ \text{s.t.} \sum_{j=1}^K w_{ij} = 1, \end{cases} \quad (1)$$

其中: $\mathbf{x}_i$ 为原始样本点, $\mathbf{x}_i^j$ 为样本 $\mathbf{x}_i$ 的第 $j$ 个近邻, $w_{ij}$ 为样本点 $\mathbf{x}_i$ 与其近邻 $\mathbf{x}_i^j$ 的关系权重。

**第2步** 通过式(2)使样本的局部近邻结构得到保持,得到投影矩阵 $\mathbf{A}$ :

$$\begin{cases} \min \sum_{i=1}^m \|\mathbf{y}_i - \sum_{j=1}^m w_{ij} \mathbf{y}_i^j\|^2, \\ \text{s.t.} \mathbf{a}^T \mathbf{X}^T \mathbf{X} \mathbf{a} = 1, \end{cases} \quad (2)$$

其中 $\mathbf{y}_i = \mathbf{x}_i \mathbf{a}$ 为样本投影后的点。通过拉格朗日函数法可将最优化求解问题转化为广义特征矩阵中特征值问题,如式(3)所示:

$$\mathbf{X}^T \mathbf{M} \mathbf{X} \mathbf{a} = \lambda \mathbf{X}^T \mathbf{X} \mathbf{a}. \quad (3)$$

式(3)中 $l$ 个最小特征值所对应的特征向量构成投影矩阵,即 $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_l)$ ,样本 $\mathbf{x}_*$ 投影后可表示为

$$\mathbf{y}_* = \mathbf{x}_* \mathbf{A}. \quad (4)$$

NPE在过程监控中使用 $T^2$ 和SPE统计量,如式(5)和式(6)所示:

$$T_*^2 = \mathbf{y}_* \mathbf{\Lambda}^{-1} \mathbf{y}_*^T, \quad (5)$$

$$\text{SPE}_* = (\mathbf{x}_* - \mathbf{x}_* \mathbf{A} \mathbf{A}^T) (\mathbf{x}_* - \mathbf{x}_* \mathbf{A} \mathbf{A}^T)^T, \quad (6)$$

其中 $\mathbf{\Lambda}$ 是 $\mathbf{Y}$ 的协方差矩阵,并根据核密度估计法确定 $T^2$ 和SPE统计量控制限<sup>[18]</sup>。

## 3 基于局部保持嵌入-K近邻比率密度(NPE-KRD)的多模态故障检测方法

### 3.1 $K$ 近邻比率密度(KRD)

KNN方法通过计算样本与其前 $K$ 个近邻的距离平方和对过程进行监控。文献[12]指出KNN在处理方差差异较大的多模态过程存在一定缺陷。为了有效监控方差差异显著的多模态过程,本节提出了一种 $K$ 近邻比率密度( $K$  nearest neighbor ratio density, KRD)的方案。

目前,Rodriguez提出一种计算局部密度的方法<sup>[19]</sup>,具体过程如下:

首先,计算样本点 $\mathbf{x}_i$ 与样本点 $\mathbf{x}_j$ 之间的距离如式(7)所示:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|. \quad (7)$$

然后, 计算样本点的密度 $\rho_i$ , 如式(8)所示:

$$\rho_i = \sum_{j=1}^m \exp\left(-\frac{d(\mathbf{x}_i, \mathbf{x}_j)^2}{dc^2}\right), \quad (8)$$

其中:  $1 \leq j \leq m, j \neq i$ ,  $dc$ 为可按百分比参数求得的截止距离<sup>[20]</sup>.

为了有效监控多模态过程, 建立样本的统计量 $R$ , 如式(9)所示:

$$R = \frac{\omega}{\rho_i}, \quad (9)$$

其中 $\omega = \frac{1}{K} \sum_{j=1}^K \rho_i^j$ ,  $\rho_i^j$ 为样本 $\mathbf{x}_i$ 第 $j$ 近邻的密度.

由于各模态数据点的密度不同, 当正常样本的 $\rho_i$ 较小时,  $\omega$ 也相应较小. 当正常样本的 $\rho_i$ 较大时,  $\omega$ 也相应较大. 通过计算每个样本的统计指标 $R$ , 使得各模态中样本点对统计量的贡献度为同一水平且统计值在1附近波动, 实现故障样本与正常样本分离的目的. 与 $KNN$ 方法相比较,  $KRD$ 不仅可以降低多模态方差差异对故障检测的影响, 同时还能够有效监控过程密集模态的健康状态.

### 3.2 基于局域保持嵌入-K近邻比率密度(NPE-KRD)的多模态故障检测方法

在多模态故障检测过程中, 当样本维数较高时,  $KRD$ 方法会出现与 $KNN$ 方法同样的计算量大的问题. 因此, 先利用NPE方法进行维数约简, 接下来在低维空间使用 $KRD$ 进行故障检测. 其故障检测过程分为离线建模和在线检测, 如下所示:

离线建模具体步骤:

- 1) 利用Z-SCORE标准化方法处理得到的数据集 $\mathbf{X}$ ;
- 2) 利用式(3)计算 $\mathbf{A}$ , 经过式(4)得到降维后的数据集 $\mathbf{Y}$ ;
- 3) 通过式(9)计算样本 $\mathbf{y}_i$ 的 $KRD$ 统计值 $R$ ;
- 4) 根据核密度估计法得到控制限 $R_\alpha$ .

在线检测具体步骤:

- 1) 对测试样本 $\mathbf{x}_*$ 进行标准化;
- 2) 利用式(4)将 $\mathbf{x}_*$ 投影到低维子空间中, 得到样本点 $\mathbf{y}_*$ ;
- 3) 由式(9)得出 $\mathbf{y}_*$ 的 $KRD$ 统计值 $R_*$ ;
- 4) 若 $R_*$ 小于 $R_\alpha$ , 则 $\mathbf{x}_*$ 为正常样本; 否则,  $\mathbf{x}_*$ 为故障样本.

一旦过程发生故障, 需要诊断故障发生原因, 进而对故障进行准确定位. 相对于正常数据, 故障数据的分布轨迹与其存在本质上的差异, 即当样本密度小于正常样本密度时, 判定此样本为故障样本. 原因在于样本的某些变量发生变化使之与正常样本偏离. 实质上就是故障样本到其第 $K$ 近邻的距离大于正常样本到其第 $K$ 近邻的距离. 在故障诊断过程中, 变量的贡

献大是故障生成的主要原因. NPE-KRD诊断过程如下:

首先, 在低维空间内计算样本与其第 $K$ 近邻样本的距离, 如式(10)所示:

$$D = (\mathbf{y}_i - \mathbf{y}_i^K)(\mathbf{y}_i - \mathbf{y}_i^K)^T, \quad (10)$$

其中 $\mathbf{y}_i^K$ 为 $\mathbf{y}_i$ 的第 $K$ 近邻.

然后, 计算变量 $x_j$ 对 $D$ 的贡献, 如式(11)所示:

$$CON_{x_j} = (\mathbf{x}_i - \mathbf{x}_i^K) \mathbf{E}_j \boldsymbol{\Xi}, \quad (11)$$

其中:  $\boldsymbol{\Xi} = \mathbf{A} \mathbf{A}^T (\mathbf{x}_i - \mathbf{x}_i^K)^T$ ,  $\mathbf{E}_j = \boldsymbol{\Sigma}_j \boldsymbol{\Sigma}_j^T$ ,  $\boldsymbol{\Sigma}_j$ 为 $n$ 阶单位矩阵的第 $j$ 列.

### 4 数值例子

本节应用数值例子对NPE-KRD方法进行说明, 模型如下:

$$\begin{aligned} \text{模型 1: } & \begin{cases} x = 3t + e, \\ y = -2t + e, \end{cases} & t \sim U(-1, 1), \\ \text{模型 2: } & \begin{cases} x = 3t + e, \\ y = 40 - 2t + e, \end{cases} & t \sim U(-3, 12), \end{aligned}$$

其中每个模态均包括4个变量, 主要变量为 $x, y$ , 各模态其余2个变量和 $e$ 都是微小白噪声. 对模型中的两个模态各生成250个样本组成训练集合, 同时各生成50个样本组成校验集合, 并在密集模态1中通过对变量 $y$ 增加扰动生成100个故障样本. 数据分布特征如图1所示, 由图1可知两个模态数据具有不同的分散程度, 密集数据源于模态1, 稀疏数据源于模态2.

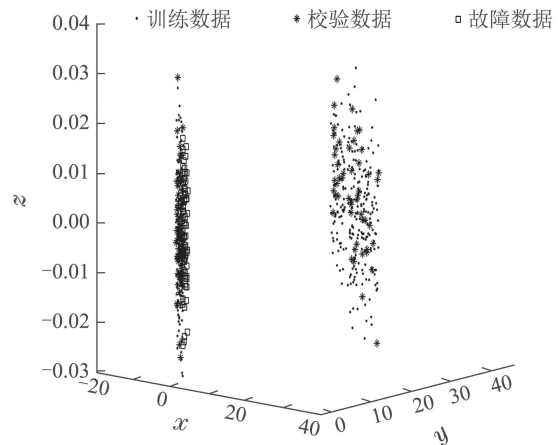


图1 数据散点图

Fig. 1 Scatter plots of samples

PCA, DPCA, NPE,  $KNN$ ,  $KRD$ 和NPE-KRD方法被应用到本节对该例进行故障检测. PCA, NPE及NPE-KRD的主元个数(principal components, PCs)选取为2, DPCA的PCs选取为3, 时滞参数为2. 本节例子近邻数 $K$ 选取为3.  $KRD$ 和NPE-KRD中的 $dc$ 设定为1.42和0.91.

由图2-4可知, PCA和NPE故障检测率为零,

DPCA故障检测率接近于零, 主要原因是NPE, PCA和DPCA在故障检测中都假设数据得分服从多元高斯分布, 如图5可知PCA得分分布曲线偏离直线轨迹, 因此, 本例多模态数据得分不满足假设条件. 由图6可知, KNN未能有效地在数据中提取出全部故障, 主要原因是KNN方法在处理方差显著的多模态过程时存在缺陷. KRD和NPE-KRD故障检测结果如图7和图8所示, 可看出两种方法的故障检测率相近. 相对而言, NPE-KRD方法检测效率更高, 如图9所示.

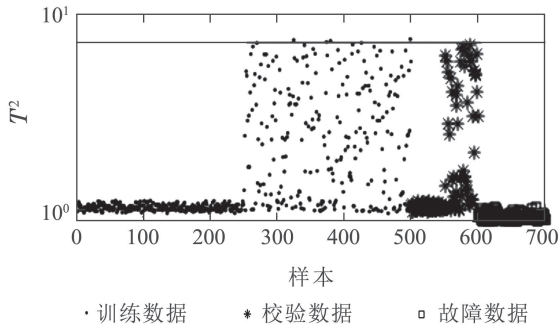


图 2 PCA检测结果

Fig. 2 Fault detection results using PCA

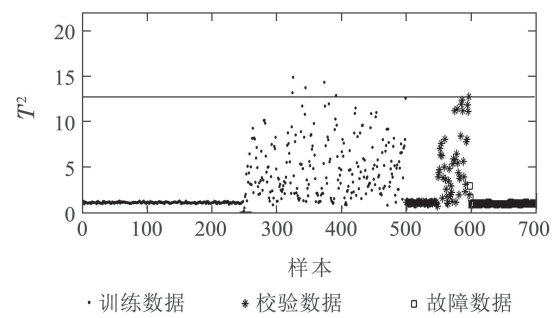


图 3 DPCA检测结果

Fig. 3 Fault detection results using DPCA

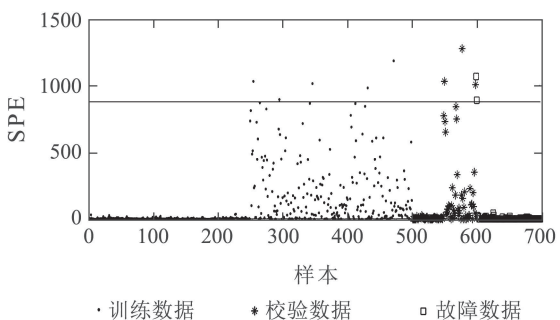


图 4 NPE检测结果

Fig. 4 Fault detection results using NPE

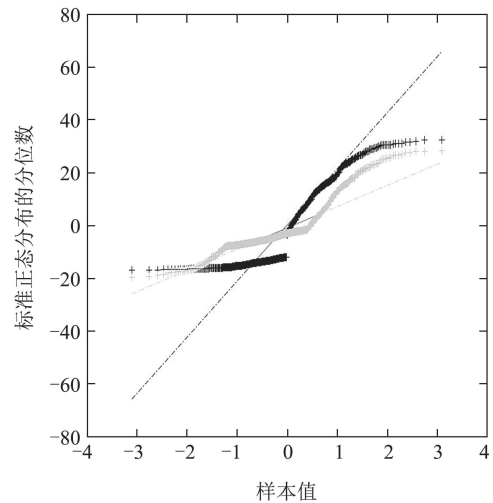
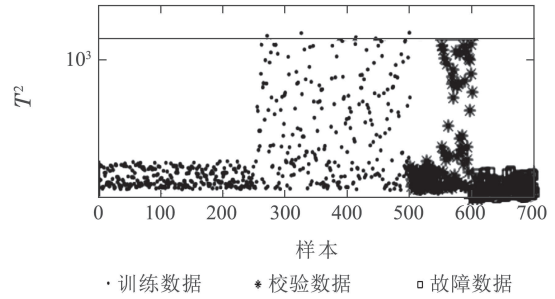


图 5 得分变量关系图

Fig. 5 Relation of score variables

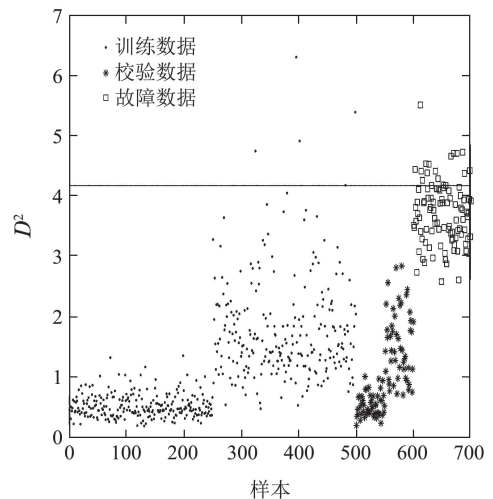


图 6 KNN检测结果

Fig. 6 Fault detection results using KNN

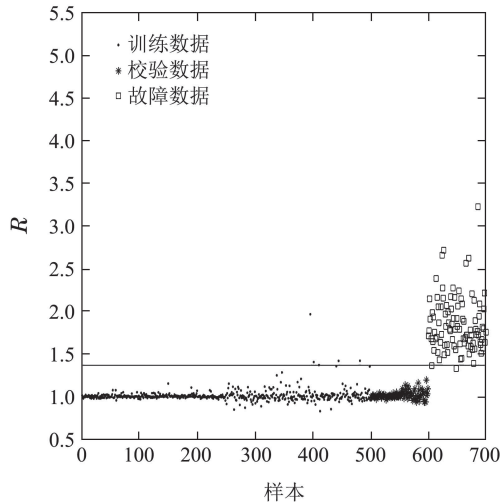


图7 KRD检测结果

Fig. 7 Fault detection results using KRD

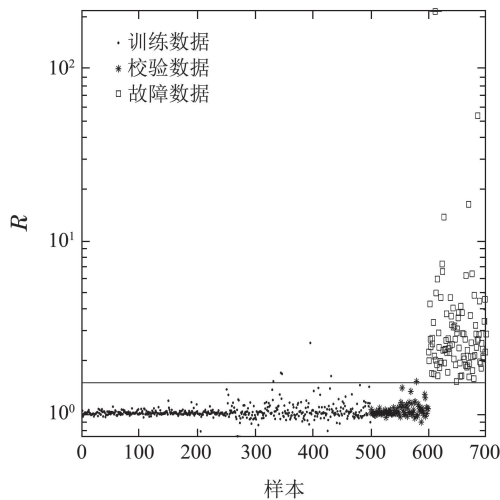


图8 NPE-KRD检测结果

Fig. 8 Fault detection results using NPE-KRD

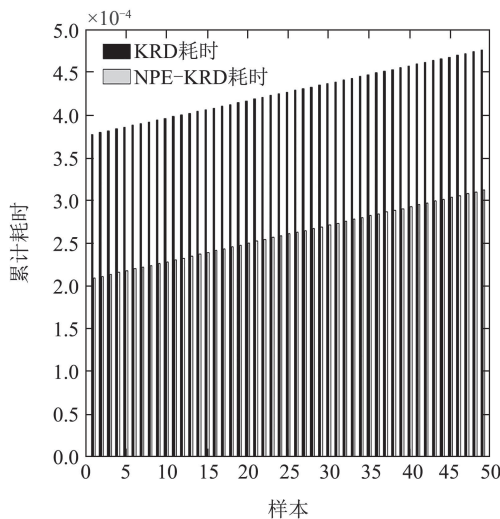


图9 KRD与NPE-KRD检测累计耗时

Fig. 9 Cumulative time of KRD and NPE-KRD

最后,对本例进行故障诊断,结果如图10所示.由图可知在所有变量中变量2的贡献占比较高,因此可

判定故障是由变量2发生异常变化所引起.本节中数值例子在变量 $y$ (即第2变量)上增加扰动生成故障,结果验证了诊断策略的有效性.

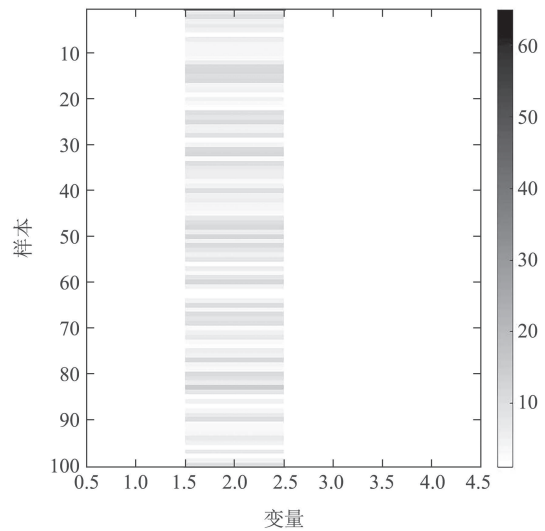


图10 变量贡献图

Fig. 10 Contribution charts of monitored variables

### 5 半导体工业过程故障检测实例

本节所应用的半导体数据源于美国德州仪器公司的半导体生产过程<sup>[21]</sup>,数据集包含3个实验过程的108个正常批次、21个故障批次.前34个批次源于第一工况,35到71批次源于第二工况,余下批次于第三工况.依据文献<sup>[22]</sup>,将正常样本分为训练集和校验集,其中训练集包含101个样本,校验集包含6个样本.应用PCA, DPCA, NPE, KNN, KRD以及NPE-KRD对该过程进行监控,参数设置如表1所示,检测率见表2.

表1 参数确定

Table 1 Parameter determinations

方法	PCs	$K$	$dc$
PCA	3	—	—
DPCA	3	—	—
NPE	14	3	—
KNN	—	3	—
KRD	—	18	9.12
NPE-KRD	14	15	1.15

表2 故障检测率

Table 2 Fault detection rates

方法	检测率/%
PCA- $T^2$	24
PCA-SPE	76
DPCA- $T^2$	19
DPCA-SPE	90
NPE- $T^2$	85
NPE-SPE	71
KNN	76
KRD	76
NPE-KRD	100

由于半导体数据具有多模态特征, 而PCA和NPE适用于单模态故障检测, 因此检测率较低; 此外, 3种工况分散程度不同引起KNN具有较低故障检测率. DPCA方法可以有效提取过程动态特征, 因而, 在半导体工业过程具有较好的故障检测效果. 但是本例中的多模态特征仍然制约该方法的监控效果.

相比于KRD方法, NPE-KRD具有更好的故障检测性能, 检测率达到100%, 如图11-12. 主要原因是本文方法利用NPE削弱了离群点的对故障检测的作用, 从而提高了故障检测率.

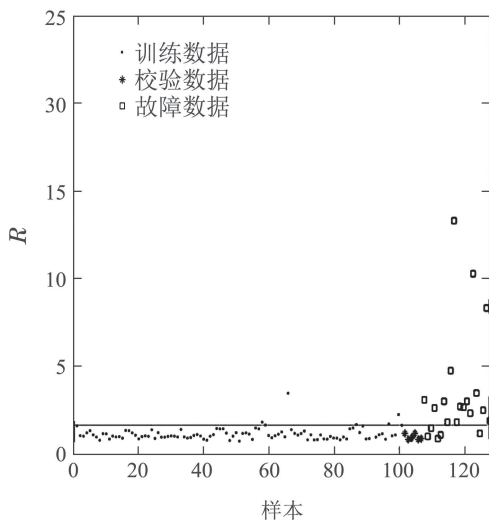


图 11 KRD检测结果

Fig. 11 Fault detection results using KRD

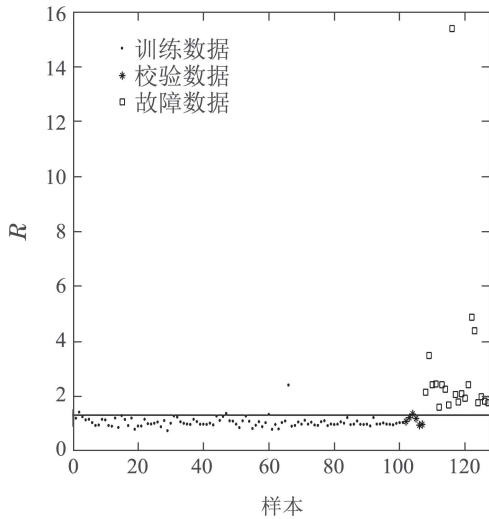


图 12 NPE-KRD检测结果

Fig. 12 Fault detection results using NPE-KRD

本节分别对故障1、故障4和故障10进行诊断, 其结果如图13-15所示. 针对故障1, 可看出变量13具有较大的贡献, 因此变量13是故障发生的主要原因. 由文献[23]知故障1的故障类型为TCP+50, 变量13对应变量名为TCP impedance, 符合半导体数据故障设置. 同理, 故障4和10的异常变量为26和13, 分别为变量

Chamber Pressure的方差和TCP tuner的均值. 诊断结果与文献[23]所述一致.

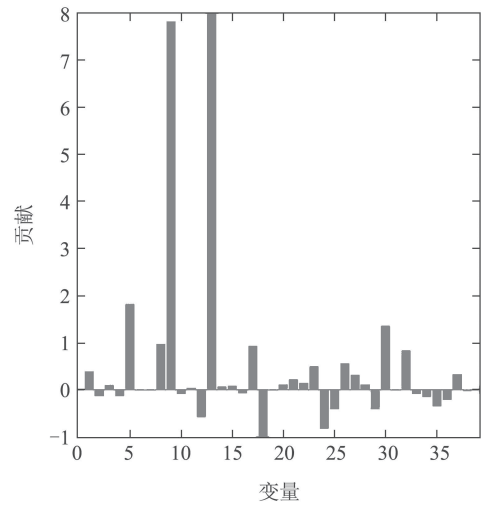


图 13 故障1的变量贡献图

Fig. 13 Contribution charts of variables for fault 1

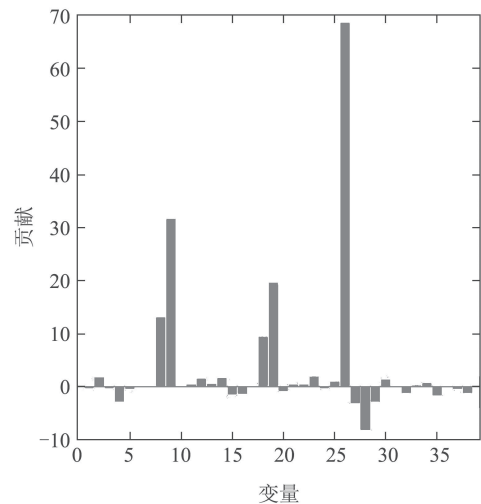


图 14 故障4的变量贡献图

Fig. 14 Contribution charts of variables for fault 4

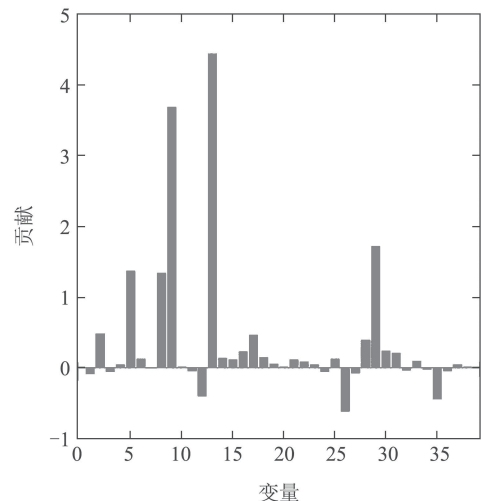


图 15 故障10的变量贡献图

Fig. 15 Contribution charts of variables for fault 10

## 6 结论

针对多模态质量监控问题, 本文提出了一种基于NPE-KRD的故障检测方法. NPE-KRD不仅降低了故障检测过程的计算复杂度, 还提高过程的故障检测率. 相比传统的NPE, KNN等方法, 本文方法更适合完成模态方差差异显著的过程监控. 数值例子及半导体工业过程的仿真实验验证了本文方法的有效性.

## 参考文献:

- [1] DING S X, ZHANG P, NAIK A, et al. Subspace method aided data-driven design of fault detection and isolation systems. *Journal of Process Control*, 2009, 19(9): 1496 – 1510.
- [2] LIU Y, HSEUH B F, GAO Z, et al. Flooding prognosis in packed columns by assessing the degree of steadiness (DOS) of process variable trajectory. *Industrial & Engineering Chemistry Research*, 2016, 55(40): 10744 – 10750.
- [3] DAI X, GAO Z. From model, signal to knowledge: a data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics*, 2013, 9(4): 2226 – 2238.
- [4] XUAN Q, CHEN Z, LIU Y, et al. Multi-view generative adversarial network and its application in pearl classification. *IEEE Transactions on Industrial Informatics*, 2019, 66(10): 8244 – 8252.
- [5] MOORE B C. Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 2003, 26(1): 17 – 32.
- [6] KU W, STORER R H, GEORGAKIS C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics & Intelligent Laboratory Systems*, 1995, 30(1): 179 – 196.
- [7] LIU Y, YU L, GAO Z, et al. Online flooding supervision in packed towers: an integrated data-driven statistical monitoring method. *Chemical Engineering & Technology*, 2018, 41(3): 436 – 446.
- [8] BELKIN M, NIYOGI P. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems*, 2001, 14(6): 585 – 591.
- [9] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 290(5500): 2323 – 2326.
- [10] TENENBAUM J B, DE V S, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000, 290(5500): 2319 – 2323.
- [11] HE X, DENG C, YAN S, et al. Neighborhood preserving embedding. *The 10th IEEE International Conference on Computer Vision*. Piscataway, NJ, USA: IEEE, 2005: 1208 – 1213.
- [12] 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.
- [13] WANG J, WEI H, CAO L, et al. Soft-transition sub-pca fault monitoring of batch processes. *Industrial & Engineering Chemistry Research*, 2013, 52(29): 9879 – 9888.
- [14] 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): 86 – 93.  
(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): 86 – 93.)
- [15] GUO J, WANG X, LI Y. kNN based on probability density for fault detection in multimodal processes. *Journal of Chemometrics*, 2018, 32(7). DOI: 10.1002/cem.3021.
- [16] ZHANG C, GAO X W, LI Y, et al. Fault detection strategy based on weighted distance of  $k$  nearest neighbors for semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, 2019, 32(1): 75 – 81.
- [17] HA D, AHMED U, PYUN H, et al. Multi-mode operation of principal component analysis with  $k$ -nearest neighbor algorithm to monitor compressors for liquefied natural gas mixed refrigerant processes. *Computers & Chemical Engineering*, 2017, 106: 96 – 105.
- [18] CREMERS D, OSHER S J, SOATTO S. Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *International Journal of Computer Vision*, 2006, 69(3): 335 – 351.
- [19] RODRIGUEZ A, LAIO A. Machine learning clustering by fast search and find of density peaks. *Science*, 2014, 344(6191): 1492 – 1496.
- [20] LI H, WANG H, FAN W. Multimode process fault detection based on local density ratio-weighted support vector data description. *Industrial & Engineering Chemistry Process Design and Development*, 2017, 56(9): 2475 – 2491.
- [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. *Journal of Chemometrics*, 1999, 13(3/4): 379 – 396.
- [22] ZHANG Cheng, GUO Qingxiu, LI Yuan. Batch process monitoring based on  $k$  nearest neighbors in discriminated kernel principle component space. *Journal of Computer Applications*, 2018, 38(336): 51 – 57.  
(张成, 郭青秀, 李元. 基于判别核主元空间 $k$ 近邻的批次过程监视. 计算机应用, 2018, 38(336): 51 – 57.)
- [23] 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.

## 作者简介:

**张成** 博士研究生, 副教授, 目前研究方向为复杂工业过程故障诊断, E-mail: zhangcheng@syuct.edu.cn;

**郑百顺** 硕士研究生, 目前研究方向为复杂工业过程故障诊断, E-mail: 2445181145@qq.com;

**郭青秀** 硕士研究生, 目前研究方向为复杂工业生产过程智能建模等, E-mail: 505087396@qq.com;

**冯立伟** 硕士, 讲师, 目前研究方向为复杂工业过程故障诊断, E-mail: 470348733@qq.com;

**李元** 博士生导师, 教授, 目前研究方向为复杂工业过程故障诊断, E-mail: li-yuan@mail.tsinghua.edu.cn.