

文章编号:1006-2467(2021)08-1001-08

DOI: 10.16183/j.cnki.jsjtu.2020.295

# 基于邻域保持嵌入的主多项式非线性过程故障检测

李 元, 姚宗禹

(沈阳化工大学 信息工程学院, 沈阳 110142)

**摘 要:** 针对化工过程的变量数据维数高、非线性的问题,提出基于邻域保持嵌入(NPE)-主多项式分析(PPA)的过程故障检测算法.应用 NPE 算法提取高维数据的低维子流形,能够解决传统的线性降维算法不能提取局部结构信息的问题,对维数进行约减.利用 PPA 法时,使用一组灵活的主多项式分量来描述数据,能够有效地捕捉过程数据中固有的非线性结构.在降维后的流形空间进行主多项式分析并建立 Hotelling's  $T^2$  和平方预测误差统计量模型,同时确定控制限以进行故障检测.最后,通过一组非线性数值实例和 Tennessee Eastman 化工过程数据,将 NPE-PPA 算法与传统的核主元分析法、PPA 法进行对比分析,验证所提算法的有效性及优越性.

**关键词:** 邻域保持嵌入;主多项式分析;非线性过程;故障检测

**中图分类号:** TP 277

**文献标志码:** A

## Principal Polynomial Nonlinear Process Fault Detection Based on Neighborhood Preserving Embedding

LI Yuan, YAO Zongyu

(College of Information Engineering, Shenyang University of Chemical Technology,  
Shenyang 110142, China)

**Abstract:** Aimed at the problem of high dimension and nonlinearity of variable data in chemical process, a process fault detection algorithm based on neighborhood preserving embedding(NPE)-principal polynomial analysis (PPA) is proposed in this paper. The NPE algorithm is used to extract low dimensional submanifolds of high dimensional data, which overcomes the problem that the traditional linear dimensionality reduction algorithm cannot extract local structure information, so as to reduce the dimensions. The PPA method is used to describe data by a set of flexible principal polynomial components, which can effectively capture the inherent nonlinear structure of process data. The principal polynomial analysis is conducted in the reduced manifold space, and Hotelling's  $T^2$  and square prediction error statistical models are established to determine the control limit for fault detection. Finally, compared with the traditional kernel principal component analysis and the PPA method, a group of nonlinear numerical examples and Tennessee Eastman chemical process data experiments are performed to verify the effectiveness and superiority of the NPE-PPA algorithm.

收稿日期:2020-09-14

基金项目:国家自然科学基金重大项目(61490701),国家自然科学基金项目(61673279)

作者简介:李 元(1964-),女,辽宁省沈阳市人,教授,博士生导师,现主要从事统计过程控制和基于数据驱动的过程故障监控与诊断研究,电话(Tel.):13082424115;E-mail:li-yuan@mail.tsinghua.edu.cn.

**Key words:** neighborhood preserving embedding (NPE); principal polynomial analysis (PPA); nonlinear process; fault detection

在现代工业中,为了提高过程系统运行安全性和可靠性的同时,保证化学工程工艺系统中的产品质量,必须对过程进行实时监控.数据驱动的过程监测和控制技术已广泛应用于各种工业过程<sup>[1-3]</sup>,包括化学品、药品和半导体制造等<sup>[4-8]</sup>.

传统主成分分析(PCA)是目前常用的多元统计方法,基于 PCA 的过程监测方法已成功地应用于大量工业过程中.然而,基于 PCA 的方法均假设过程是线性的,这限制了其在非线性工业过程中的应用.现代工业非线性过程大量存在,为了监测非线性过程,多种非线性过程检测方法相继被提出.

文献[9]通过在 PCA 中引入核技术提出了核主元分析(KPCA).而文献[10-11]提出基于 KPCA 的连续过程监测的非线性过程监测技术,KPCA 将非线性低维数据映射到高维空间,使其在高维空间中存在线性关系.相对于 PCA 方法而言,KPCA 方法在故障检测过程中可以提取系统中的非线性特征信息而不至于丢失重要信息,具有更好的非线性故障检测能力.但是,KPCA 没有明确地考虑数据的非线性几何结构可能驻留在流形上的可能性<sup>[12]</sup>,若参数设置不当,核方法的故障检测性能就无法得到保证.此外,计算核函数耗时较长,影响对实时检测的需求.

文献[13-14]提出一种基于自联想的 5 层神经网络的非线性主成分分析(NLPCA),使用自关联神经网络将数据映射到特征空间.文献[15]提出一种基于主曲线和神经网络相结合的非线性主成分分析法,并应用于非线性过程监测.然而,大部分存在的非线性方法都是基于神经网络的,需要离线和在线训练,而且为了计算主元,需要解决非线性优化问题,主元的数量必须在神经网络训练之前确定,给实际应用带来不便.

文献[16]提出一种基于等距离映射(ISOMAP)的故障检测方法.该方法用等距离映射进行非线性降维,采用自适应准则选取邻域参数,能够有效地挖掘过程的非线性特征.但是当流形曲率较大时,流形上的测地距离估计会产生较大的误差,导致嵌入结果产生变形.文献[17-18]提出主多项式分析(PPA),通过最小化相应正交子空间中的回归误差(或方差)来变形直线主成分,并将其应用于遥感数据处理.文献[19-20]将 PPA 法引入到工业过程的

故障检测和诊断技术上,使用一组灵活的主多项式分量来描述数据,与基于 PCA 的方法相比,PPA 法更能有效地捕捉过程数据中固有的非线性几何结构.虽然,PPA 法的提取曲线主成分可以包含大部分的数据变化,但是其只捕获过程数据的全局结构,不能有效地保留数据详细的局部结构信息.

针对工业过程数据的非线性特征,提出一种基于邻域保持嵌入(NPE)-主多项式分析的过程故障检测算法.使用 NPE 算法提取高维数据的低维子流形,并进行维数约减,同时保持邻域结构不变.将特征空间的数据应用主多项式分析建立故障检测模型,计算其检测统计量及控制限来进行故障检测.本研究以 Tennessee Eastman(TE)化工过程为背景进行仿真实验,为基于数据驱动的故障检测领域提供了实践基础.

## 1 NPE 算法

邻域保持嵌入算法的目的是在保持数据集局部流形结构不变的同时,将给定原始数据  $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_n]$ ,  $\mathbf{x}_i \in \mathbf{R}^D$  从  $D$  维空间  $\mathbf{R}^D$  映射到一个相对低维的特征空间  $\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \cdots \ \mathbf{y}_n]$ ,  $\{\mathbf{y}_i \mid \mathbf{y}_i \in \mathbf{R}^d\} (d < D)$  中,  $d$  为低维空间维数,并求出映射变换矩阵  $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_d] \in \mathbf{R}^{m \times d}$ . 算法流程如下<sup>[21]</sup>.

(1) 通过样本间的欧式距离确定每个样本的前  $k$  个近邻并构造邻域连接图.

(2) 计算邻接图中每一条边上的权重值构成权重矩阵  $\mathbf{F}$ ,确定每个样本点与其邻域点之间的重构权向量,并在低维空间中保持每个邻域的权值不变,使误差函数达到最小,可由下式计算:

$$\min \sum_{i=1}^m \left\| \mathbf{x}_i - \sum_{j=1}^k f_{ij} \mathbf{x}_i^j \right\|^2 \quad (1)$$

$$\text{s. t. } \sum_{j=1}^k f_{ij} = 1 \quad (2)$$

式中:  $\mathbf{x}_i$  为原始样本点;  $\mathbf{x}_i^j$  为样本  $\mathbf{x}_i$  的第  $j$  个近邻;  $f_{ij}$  为样本点  $\mathbf{x}_i$  与其近邻  $\mathbf{x}_i^j$  的关系权重.

(3) 计算特征映射矩阵.根据高维空间中的样本点与其近邻点之间的权值矩阵  $\mathbf{F}$ ,嵌入低维空间的投影矩阵  $\mathbf{A}$  可以通过求解损失函数的最小化问题可以得到:

$$J_{\text{opt}} = \min \sum_{i=1}^m \left\| \mathbf{y}_i - \sum_{j=1}^k f_{ij} \mathbf{y}_i^j \right\|^2 \quad (3)$$

式中:  $\mathbf{y}_i$  为低维空间样本点;  $\mathbf{y}_i^j$  为样本  $\mathbf{y}_i$  的第  $j$  个近邻。

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的特征值问题,如下式所示:

$$\mathbf{X}^T \mathbf{M} \mathbf{X} \mathbf{a}_i = \lambda \mathbf{X}^T \mathbf{X} \mathbf{a}_i \quad (4)$$

式中:  $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$ ,  $\mathbf{I}$  为同维度单位矩阵。求解式(4)中最小的  $d$  个特征值所对应的特征向量组成映射矩阵  $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_d] \in \mathbf{R}^{n \times d}$ 。

## 2 主多项式分析

PPA 是一种计算主多项式成分的顺序算法。在每一步的计算中,计算出最佳投影数据的向量,通过使用一组主多项式成分从过程数据中学习数据的低维表示,将 PCA 中的直线主成分更换为曲线的主多项式成分,可以更好地捕捉过程变量的非线性特征<sup>[19]</sup>。给定数据矩阵  $\mathbf{G}$  为  $m$  个变量,  $n$  个训练样本,PPA 对原始数据进行如下分解:

$$\boldsymbol{\alpha}_p = \mathbf{e}_p^T \mathbf{Z}_{p-1} \quad (5)$$

$$\mathbf{Z}_p = \mathbf{E}_p^T \mathbf{Z}_{p-1} - \widehat{\mathbf{M}}_p \quad (6)$$

式中:  $\widehat{\mathbf{M}}_p$  为估计的条件均值;  $\boldsymbol{\alpha}_p$  为  $\mathbf{X}_{p-1}$  在  $\mathbf{e}_p^T$  上的投影;  $\mathbf{e}_p^T \in \mathbf{R}^{(m-p+1)}$  为主要投影方向;  $\mathbf{Z}_{p-1}$  为上一步处理后的残差;  $\mathbf{E}_p^T \in \mathbf{R}^{(m-p) \times (m-p+1)}$  正交于  $\mathbf{e}_p^T$ ;  $\mathbf{Z}_p$  为将用于下一步的残差;  $p = \rho - 1 : 1$ , 为顺序执行步数,  $\rho$  为 PPA 中保留的有效主多项式分量的个数。当  $p=1$  时,  $\mathbf{G}_0 = \mathbf{G}$  为原始输入数据。PPA 采用了幂为  $r_p$  的多项式函数来估计条件均值,估计方程可定义为

$$\widehat{\mathbf{M}}_p = \mathbf{W}_p \mathbf{V}_p \quad (7)$$

$$\mathbf{V}_p = [\mathbf{v}_{p,1} \ \mathbf{v}_{p,2} \ \cdots \ \mathbf{v}_{p,n}] \quad (8)$$

式中:  $\mathbf{W}_p \in \mathbf{R}^{(m-p) \times (r_p+1)}$  为多项式函数矩阵;  $\mathbf{v}_{p,i} = [1 \ \boldsymbol{\alpha}_p \ \boldsymbol{\alpha}_p^2 \ \cdots \ \boldsymbol{\alpha}_p^{r_p}]^T$  为一组范德蒙德向量。  $\mathbf{W}_p$  可以通过求解最小二乘问题来计算,可表示为

$$\mathbf{W}_p = (\mathbf{E}_p^T \mathbf{Z}_{p-1}) \mathbf{V}^\dagger \quad (9)$$

式中:  $\mathbf{V}^\dagger$  为  $\mathbf{V}$  的伪逆。

## 3 基于低维子流行空间主多项式分析的故障检测

获得一个经 NPE 算法维数约减的数据  $\mathbf{y}^{\text{new}} \in \mathbf{R}^d$ , 根据式(5)和(6)可以得到第  $p$  步的主多项式得分  $\boldsymbol{\alpha}_p^{\text{new}}$  为

$$\boldsymbol{\alpha}_p^{\text{new}} = \mathbf{e}_p^T \mathbf{y}_{p-1}^{\text{new}} \quad (10)$$

$$\mathbf{y}_p^{\text{new}} = \mathbf{E}_p^T \mathbf{y}_{p-1}^{\text{new}} - \widehat{\mathbf{M}}_p = \mathbf{E}_p^T \mathbf{y}_{p-1}^{\text{new}} - \mathbf{W}_p \mathbf{V}_p \quad (11)$$

令  $\boldsymbol{\alpha}^{\text{new}} = [\boldsymbol{\alpha}_1^{\text{new}} \ \boldsymbol{\alpha}_2^{\text{new}} \ \cdots \ \boldsymbol{\alpha}_\rho^{\text{new}}]^T \in \mathbf{R}^\rho$  表示前  $\rho$  个主多项式分量的得分, PPA 的  $T^2$  统计量定义如下:

$$T_{\text{PPA}}^2 = (\boldsymbol{\alpha}^{\text{new}})^T \boldsymbol{\Lambda}_{\text{PPA}}^{-1} \boldsymbol{\alpha}^{\text{new}} \quad (12)$$

式中:  $\boldsymbol{\Lambda}_{\text{PPA}} \in \mathbf{R}^{\rho \times \rho}$  为对角元素为主多项式分量的方差对角矩阵。PPA 的平方预测误差(SPE)统计量定义如下:

$$\text{SPE}_{\text{PPA}} = \|\mathbf{y}^{\text{new}} - \hat{\mathbf{y}}^{\text{new}}\|^2 \quad (13)$$

式中:  $\hat{\mathbf{y}}^{\text{new}}$  为给定前  $\rho$  个主多项式分量后对  $\mathbf{y}^{\text{new}}$  重建的观测向量。  $\hat{\mathbf{y}}^{\text{new}}$  可通过如下递归求解获得:

$$\hat{\mathbf{y}}_{\rho-1} = [\mathbf{e}_\rho \ \mathbf{E}_\rho] \left[ \begin{array}{c} \boldsymbol{\alpha}_\rho^{\text{new}} \\ \mathbf{y}_\rho^{\text{new}} + \mathbf{W}_\rho \mathbf{v}_{\rho,i} \end{array} \right] \quad (14)$$

式中:  $\hat{\mathbf{y}}_\rho^{\text{new}} = \mathbf{0}_{(m-\rho) \times 1}$ , 经过迭代求得  $\hat{\mathbf{y}}^{\text{new}} = \hat{\mathbf{y}}_0^{\text{new}}$ 。

$T^2$  统计量的控制限可以计算如下:

$$\xi_{T^2} = \frac{(n-1)\rho}{(n-\rho)} F_\alpha(\rho, n-\rho) \quad (15)$$

式中:  $F_\alpha(\rho, n-\rho)$  为在显著性水平  $\theta$  上具有自由度  $\rho$  和  $n-\rho$  的  $F$  分布。

SPE 统计量的控制限可以计算如下:

$$\left. \begin{array}{l} \xi_{\text{SPE}} = g \chi_{h,\theta}^2 \\ g = \frac{\bar{v}}{2\bar{\mu}} \\ h = \frac{2\bar{\mu}^2}{\bar{v}} \end{array} \right\} \quad (16)$$

式中:  $\chi_{h,\theta}^2$  为在显著性水平  $\theta$  上具有自由度  $h$  的卡方分布;  $\bar{\mu}$  和  $\bar{v}$  为从正常运行状态数据中估计的  $\text{SPE}_{\text{PPA}}$  的均值和方差。

故障检测的两个阶段可以分为离线建模和在线监测。

(1) 离线建模。离线建模的步骤如下:

**步骤 1** 采集正常工况下的过程数据, 将数据进行标准化。

**步骤 2** 利用下式计算权重矩阵  $\mathbf{F}$ , 利用得到的权重矩阵根据式(4)计算映射矩阵  $\mathbf{A}$ 。

$$f_{ij} = 1/\text{Dis}(\mathbf{x}_i, \mathbf{x}_i^j) \quad (17)$$

$$j = 1, 2, \dots, k$$

式中:  $\text{Dis}(\mathbf{x}_i, \mathbf{x}_i^j)$  为样本  $\mathbf{x}_i$  与其近邻  $\mathbf{x}_i^j$  的欧氏距离。

**步骤 3** 在数据矩阵上构造 PPA 模型, 得到对应的主多项式分量与残差分量, 并在主多项式分量空间和残差空间分别计算  $T^2$  统计量与 SPE 统计量。

**步骤 4** 计算其对应的监测统计量的控制限  $\xi_{T^2}$  和  $\xi_{\text{SPE}}$ ; 保存训练模型中所获得的 4 个参数  $\mathbf{e}_p$ ,  $\mathbf{E}_p$ ,  $\mathbf{W}_p$  和  $\mathbf{V}_p$ 。

(2) 在线检测。在线检测的步骤如下:

**步骤 1** 获得一个新的样本  $\mathbf{x}^{\text{new}}$ , 然后使用建模数据的均值和方差对其进行标准化.

**步骤 2** 利用投影矩阵  $\mathbf{A}$  将  $\mathbf{x}^{\text{new}}$  投影到低维子空间中, 得到样本点  $\mathbf{y}^{\text{new}}$ .

**步骤 3** 利用训练模型中所学习的模型参数将样本点  $\mathbf{y}^{\text{new}}$  映射到主多项式空间和残差空间, 并在对应空间求得  $T^2$  与 SPE 统计量.

**步骤 4** 将上一步求得的统计量与离线建模步骤中的控制限进行对比, 若  $T^2$  与 SPE 中任一统计量超过其对应的控制限, 则认为该样本为故障样本.

4 实验验证

通过将基于 NPE-PPA 的故障检测方法应用于一个非线性数值实例和 TE 过程, 说明该方法的效率和优点, 并将应用结果与标准的基于 KPCA 和 PPA 的监测方法进行比较.

4.1 非线性数值实例仿真应用

给出了一个非线性仿真实例, 说明了该方法在故障检测方面的有效性. 非线性系统由以下方程描述:

$$\left. \begin{aligned} x_1 &= 0.576\,8s_1 + 0.376\,6s_2 + \sigma_1 \\ x_2 &= 0.738\,2s_1^2 + 0.056\,6s_2 + \sigma_2 \\ x_3 &= 0.829\,1s_1 + 0.400\,9s_2^2 + \sigma_3 \\ x_4 &= 0.651\,9s_1s_2 + 0.207\,0s_2 + \sigma_4 \\ x_5 &= 0.397\,2s_1 + 0.804\,5s_2 + \sigma_5 \end{aligned} \right\} \quad (18)$$

$$\left. \begin{aligned} s_1 &\sim \text{uniform}(2,5) \\ s_2 &\sim N(7,1) \end{aligned} \right\} \quad (19)$$

式中:  $s_1$  和  $s_2$  为系统控制变量;  $\sigma_1 \sim \sigma_5$  服从均值为

0, 标准差为 0.01 的高斯分布的 5 个独立白噪声. 通过仿真生成 400 个正常样本组成训练集进行建模, 另外生成 200 个正常样本和 200 个故障样本作为测试数据集. 为该数值仿真实例设定如下 2 种故障:

**故障 1** 对变量  $x_1$  从第 201 个样本至第 400 个样本上添加  $0.2(i-200)$  来引入斜坡故障,  $i$  为样本数.

**故障 2** 对变量  $x_5$  从第 201 个样本至第 400 个样本上添加幅值为 25% 的阶跃故障.

接下来, 使用前文所提到的数值实例分别应用 KPCA、PPA 和 NPE-PPA 这 3 种方法进行建模分析. 在 KPCA 中, 通过 85% 累计贡献率来确定主元个数, 在 PPA 和 NPE-PPA 模型中的主多项式成分和多项式度均设置为 2. 采用 KPCA、PPA 和 NPE-PPA 算法对故障 1 的检测结果如图 1 所示. 从图 1(a) 中可以看出, KPCA 的 SPE 统计量能够检测出故障的发生,  $T^2$  统计量完全不能检测出故障的发生. PPA 的 SPE 统计量能够检测出故障的发生,  $T^2$  统计量在故障发生一段时间后能够做出一定的反应, 但是故障检测的精确度低. NPE-PPA 的  $T^2$  统计量和 SPE 统计量在故障发生后能立即跳变到控制限以上, 并且持续报警, 能够迅速准确地检测到故障的发生.

采用 KPCA、PPA 和 NPE-PPA 算法对故障 2 的检测结果如图 2 所示. 从图 2(a) 中可以看出, KPCA 的两种统计量均不能有效地检测出故障的发生, PPA 只有 SPE 统计量能够检测出故障的发生, NPE-PPA 的两种统计量能够迅速准确地检测到故

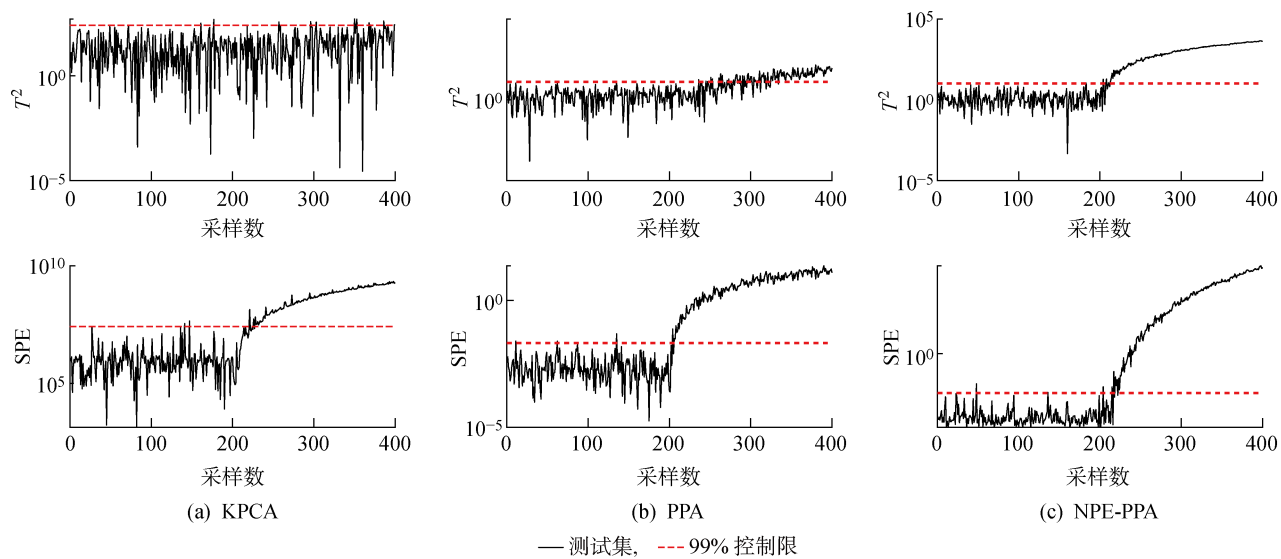


图 1 故障 1 的检测结果  
Fig. 1 Monitoring results of Fault 1

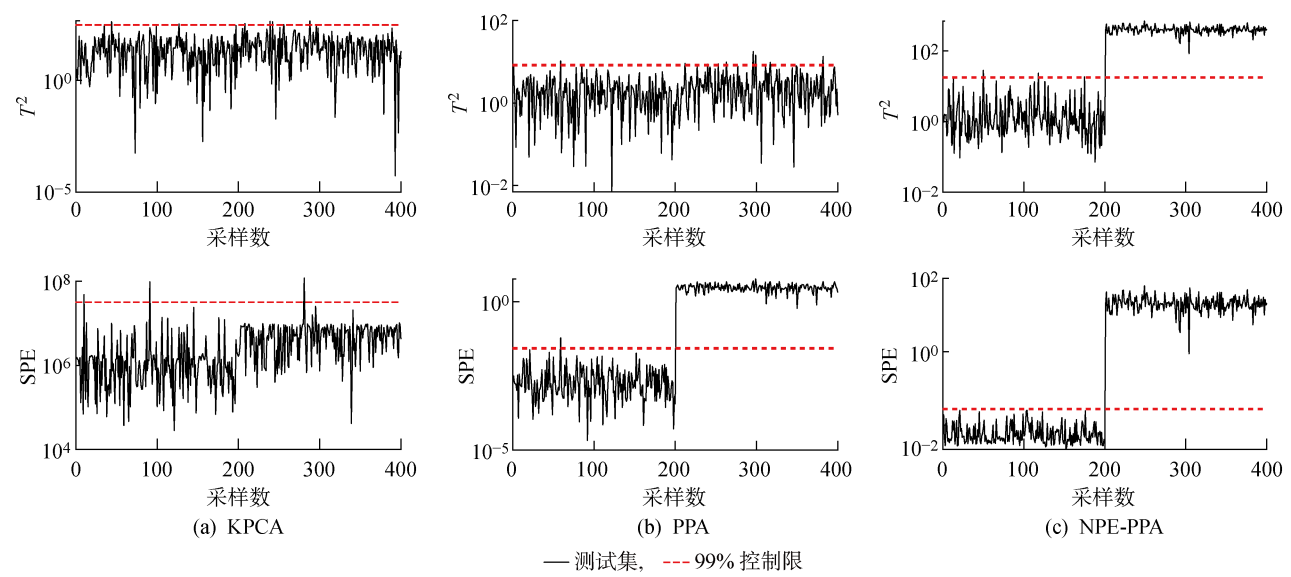


图 2 故障 2 的检测结果  
Fig.2 Monitoring results of Fault 2

障的发生. 仿真实验结果表明,邻域保持嵌入算法能够有效地保留数据的局部结构信息,提升主多项式分析处理数据的效果.

4.2 TE 过程仿真应用

TE 过程仿真应用平台是一个基于实际工业过程的仿真例子,用于测试过程系统工程中故障检测和诊断方法的效率<sup>[22-23]</sup>. Tennessee Eastman 过程最早由 Downs 和 Vogel 提出,其原型是美国伊斯曼化学公司的一个真实工业过程<sup>[24-25]</sup>. 过程主要包括 5 个操作单元:反应器、冷凝器、循环压缩机、气液分离器以及汽提塔. 整个过程中共涉及 8 种物料成分,分别为主要参加反应的气体进料 U、C、D、E;惰性不可溶进料 B;反应副产品 F 以及反应液态主产物 G 和 H. 在过程中共包含有 41 个测量变量和 12 个控制变量,其中,41 个测量变量又可以划分为 22 个过程测量变量以及 19 个成分测量变量.

该仿真平台总共预设了 21 种故障,具体的故障描述如表 1 所示,共包含阶跃型故障、随机变化故障、慢偏移故障、阀门粘滞故障、阀门卡死故障以及未知故障. 利用在正常操作条件下包含 500 个观测值的训练数据建立模型. 测试数据包含一组 21 个不同的过程故障,这些故障在样本 161 引入过程中,即过程在前 160 个样本中正常运行,然后从样本 161 到结束发生故障. 表 1 中列出了故障的详细描述,故障 3、9 和 15 是微小故障,由于反馈控制,对整个过程行为的影响不大.

将所提方法与基于 KPCA 和 PPA 算法的方法对 TE 过程的 21 种故障进行检测对比. 首先,建立

表 1 TE 过程的 21 种故障  
Tab.1 Induced 21 faults in TE process

故障编号	性质描述	变化类型
IDV1	物料 U/C 进料比改变,物料 B 含量不变	阶跃
IDV2	物料 U/C 进料比不变,物料 B 含量改变	阶跃
IDV3	物料 D 进料温度改变	阶跃
IDV4	反应器冷却入口温度改变	阶跃
IDV5	冷凝器冷却入口温度改变	阶跃
IDV6	物料 U 进料损失	阶跃
IDV7	物料 C 压力损失	阶跃
IDV8	物料 U、B、C 的组成比例改变	随机变量
IDV9	物料 D 进料温度改变	随机变量
IDV10	物料 C 进料温度改变	随机变量
IDV11	反应器冷却水入口温度改变	随机变量
IDV12	冷凝器冷却水入口温度改变	随机变量
IDV13	反应动力学参数改变	慢偏移
IDV14	反应器冷却阀门	粘住
IDV15	冷凝器冷却阀门	粘住
IDV16	未知	未知
IDV17	未知	未知
IDV18	未知	未知
IDV19	未知	未知
IDV20	未知	未知
IDV21	物流 4 阀门固定在恒定位置	恒定位置

统计监控模型,在建立 KPCA 模型时,使用主元贡献率 85%来确定其主元个数,在 PPA 方法和所提方法中使用的 PPA 模型通过交叉验证确定,主多项

式空间数量设置为 4,主多项式的幂设置为 3,检测统计量控制限置信度设置为 99%。

表 2 汇总了基于 KPCA、PPA 和 NPE-PPA 方法对 TE 过程的 21 种故障的故障检测结果,计算了所有 21 个故障的监测统计量的检测率. 针对每一个故障,检测率最高的值在表中用粗体突出表示. 对于故障 3、9 和 15 在众多文献中被认为很难检测出来,在本文的研究中也同样得到证实. 基于 KPCA 和 PPA 的方法在检测 6 个故障(故障 5、10、16、19、20 和 21)方面存在困难,大多数情况下的检测率小于 60%. 然而所提出的基于 NPE-PPA 的故障检测方法能够检测到所有 18 个故障(除故障 3、9 和 15 外),检出率高于 85%. 特别是故障 5、10 和 20,基于 NPE-PPA 方法的故障检出率比 KPCA 和 PPA 高出 2~3 倍.

为了说明基于 NPE-PPA 检测方法的优越性,分别比较了基于 KPCA、PPA 和 NPE-PPA 方法对故障 5 和 10 的检测结果,如图 3 和 4 所示. 故障 5 是 TE 过程中的冷凝器冷却入口温度发生变化形成的一种阶跃故障. 由图 3 可知,基于 KPCA 和基于 PPA 方法的  $T^2$  和 SPE 统计量成功地检测了样本 160~345 的故障,而在样本 346 之后未能检测到故障,基于 NPE-PPA 的  $T^2$  和 SPE 统计量都检测到样本 160 至过程结束的大部分故障,具有较高的检测率和较低的误报率. 故障 10 是 TE 过程中物料 C 进料温度改变形成的一种随机变化故障. 从图 4 可以清楚地观察到,所提出的基于 NPE-PPA 的方法比基于 KPCA 和 PPA 的方法更敏感,其  $T^2$  和 SPE 统计量的变化比基于 KPCA 和 PPA 统计量的变化

要显著得多,所提方法对于故障 10 的 SPE 统计量的检测率远高于其他监测统计量(见表 2)。

表 2 3 种方法对 TE 过程 21 个故障的检测率  
Tab. 2 Fault detection rates for 21 faults by using three methods in TE process

故障	KPCA		PPA		NPE-PPA	
	$T^2$	SPE	$T^2$	SPE	$T^2$	SPE
IDV1	0.99	<b>1</b>	0.99	<b>1</b>	0.98	<b>1</b>
IDV2	0.98	0.98	0.98	0.99	0.98	<b>0.99</b>
IDV3	0.01	0.08	0.12	<b>0.21</b>	0.06	0.20
IDV4	0.09	0.48	0.19	0.97	0.13	<b>0.99</b>
IDV5	0.22	0.30	0.31	0.39	<b>1</b>	<b>1</b>
IDV6	0.91	<b>1</b>	0.99	<b>1</b>	<b>1</b>	<b>1</b>
IDV7	0.99	<b>1</b>	0.48	<b>1</b>	0.71	<b>1</b>
IDV8	0.97	<b>0.98</b>	0.95	<b>0.98</b>	0.85	<b>0.98</b>
IDV9	0.01	0.09	0.15	<b>0.23</b>	0.04	0.16
IDV10	0.08	0.75	0.51	0.67	0.83	<b>0.93</b>
IDV11	0.30	0.54	0.34	0.72	0.13	<b>0.74</b>
IDV12	0.96	0.99	0.97	0.99	0.99	<b>1</b>
IDV13	0.94	0.95	0.94	0.97	0.95	<b>0.96</b>
IDV14	0.94	<b>1</b>	0.83	<b>1</b>	0.89	<b>1</b>
IDV15	0.01	0.10	0.16	0.25	0.07	<b>0.26</b>
IDV16	0.04	0.77	0.39	0.65	0.83	<b>0.95</b>
IDV17	0.69	0.88	0.78	0.94	0.87	<b>0.97</b>
IDV18	0.89	0.89	0.88	0.91	0.90	<b>0.92</b>
IDV19	0	0.45	0.05	0.32	0.60	<b>0.88</b>
IDV20	0.24	0.79	0.43	0.70	0.84	<b>0.92</b>
IDV21	0.29	0.37	0.27	0.50	0.43	<b>0.68</b>

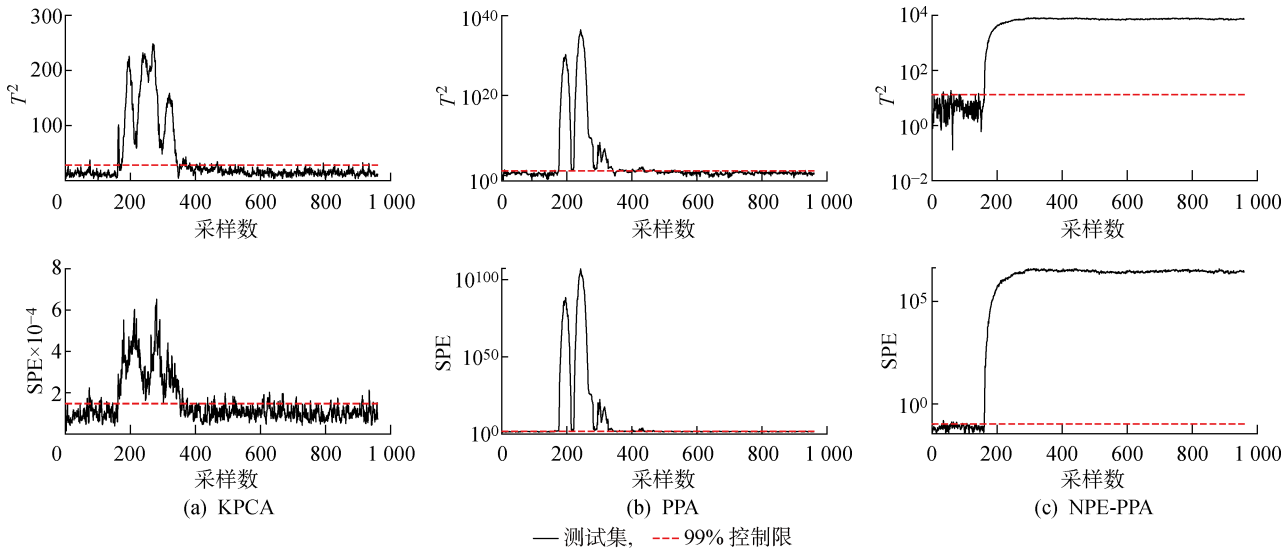


图 3 故障 5 的检测结果  
Fig. 3 Monitoring results of Fault 5



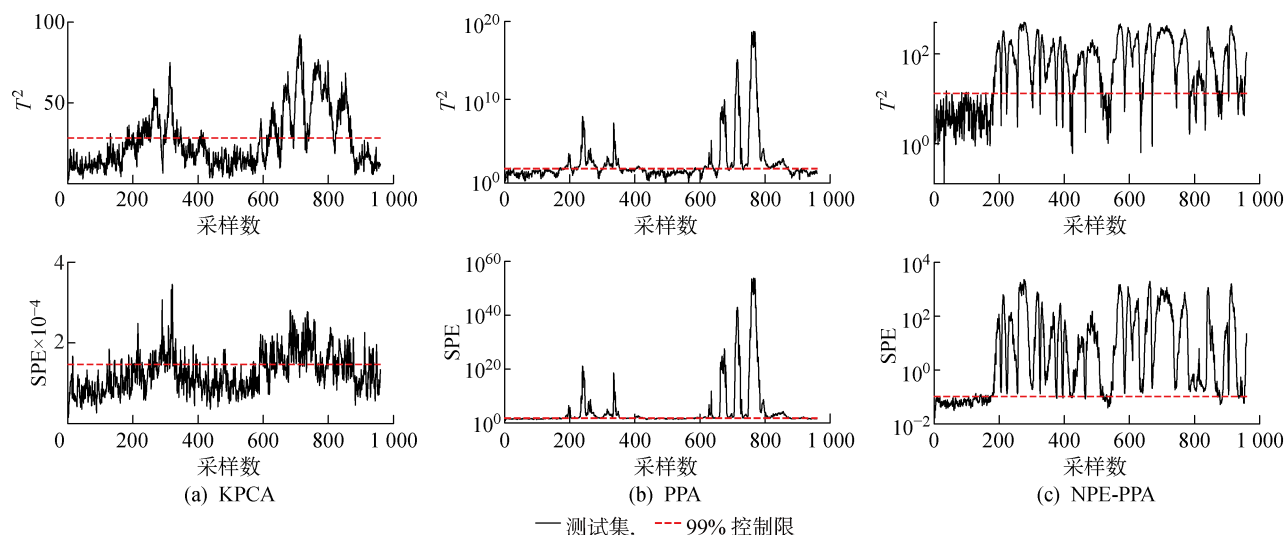


图4 故障10的检测结果

Fig.4 Monitoring results of Fault 10

由表2可知,与其他方法的监测统计方法相比,所提NPE-PPA方法的SPE统计量在TE过程的大多数故障中提供了最佳的监测结果,展示了所提出的基于NPE-PPA方法的故障检测性能.NPE-PPA算法的检测效果得到了非常明显的提升,说明了由NPE-PPA算法投影得到的特征空间包含了更多有效的信息,具有更好的特征提取能力,使得NPE-PPA方法具有更好的分类效果以及较好的故障检测性能。

## 5 结语

本文提出一种基于NPE-PPA的故障检测算法,通过邻域保持嵌入算法提取数据局部结构信息,再使用PPA方法通过主多项式分量来捕捉过程数据中潜在的非线性结构.本文通过将其应用于一组非线性数值实例和TE过程的工业基准过程,验证了所提出的基于NPE-PPA的故障检测算法的有效性与优越性.应用结果表明,所提出的基于NPE-PPA的故障检测算法优于基于KPCA和PPA的传统非线性故障检测方法。

## 参考文献:

- [1] SEVERSON K, CHAIWATANODOM P, BRAATZ R D. Perspectives on process monitoring of industrial systems[J]. *Annual Reviews in Control*, 2016, 42: 190-200.
- [2] MD NOR N, CHE HASSAN C R, HUSSAIN M A. A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems

- [J]. *Reviews in Chemical Engineering*, 2020, 36(4): 513-553.
- [3] JIANG Q C, YAN X F. Just-in-time reorganized PCA integrated with SVDD for chemical process monitoring[J]. *AIChE Journal*, 2014, 60(3): 949-965.
- [4] ZHANG X M, LI Y, KANO M. Quality prediction in complex batch processes with just-in-time learning model based on non-Gaussian dissimilarity measure [J]. *Industrial and Engineering Chemistry Research*, 2015, 54: 7694-7705.
- [5] LI Y, ZHANG X M. Variable moving windows based non-Gaussian dissimilarity analysis technique for batch processes fault detection and diagnosis[J]. *The Canadian Journal of Chemical Engineering*, 2015, 93(4): 689-707.
- [6] KANO M, NAKAGAWA Y. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry[J]. *Computers & Chemical Engineering*, 2008, 32(1/2): 12-24.
- [7] LUO L J, BAO S Y, MAO J F, *et al.* Monitoring batch processes using sparse parallel factor decomposition[J]. *Industrial & Engineering Chemistry Research*, 2017, 56(44): 12682-12692.
- [8] LI Y, ZHANG X M. Diffusion maps based k-nearest-neighbor rule technique for semiconductor manufacturing process fault detection[J]. *Chemometrics and Intelligent Laboratory Systems*, 2014, 136: 47-57.
- [9] SCHOLKOPF B, SMOLA A, MULLER K R. Non-linear component analysis as a kernel eigenvalue prob-

- lem[J]. **Neural Computation**, 1998, 10(5): 1299-1319.
- [10] LEE J M, YOO C, CHOI S W, *et al.* Nonlinear process monitoring using kernel principal component analysis[J]. **Chemical Engineering Science**, 2004, 59(1): 223-234.
- [11] CHOI S W, LEE C, LEE J M, *et al.* Fault detection and identification of nonlinear processes based on kernel PCA[J]. **Chemometrics and Intelligent Laboratory Systems**, 2005, 75(1): 55-67.
- [12] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. **Neural Computation**, 2003, 15(6): 1373-1396.
- [13] KRAMER M A. Nonlinear principal component analysis using autoassociative neural networks[J]. **AIChE Journal**, 1991, 37(2): 233-243.
- [14] MONAHAN A H. Nonlinear principal component analysis by neural networks: Theory and application to the Lorenz system[J]. **Journal of Climate**, 2000, 13(4): 821-835.
- [15] DONG D, MCAVOY T J. Nonlinear principal component analysis—Based on principal curves and neural networks[J]. **Computers & Chemical Engineering**, 1996, 20(1): 65-78.
- [16] 张妮, 田学民. 基于等距离映射的非线性动态故障检测方法[J]. **上海交通大学学报**, 2011, 45(8): 1202-1206.
- ZHANG Ni, TIAN Xuemin. Nonlinear dynamic fault detection method based on isometric mapping[J]. **Journal of Shanghai Jiao Tong University**, 2011, 45(8): 1202-1206.
- [17] LAPARRA V, TUIA D, JIMÉNEZ S, *et al.* Principal polynomial analysis for remote sensing data processing[C]//**2011 IEEE International Geoscience and Remote Sensing Symposium**. Vancouver, BC, Canada: IEEE, 2011: 12536130.
- [18] LAPARRA V, TULA D, JIMENEZ S, *et al.* Nonlinear data description with principal polynomial analysis[C]//**2012 IEEE International Workshop on Machine Learning for Signal Processing**. Santander, Spain: IEEE, 2012: 13117221.
- [19] ZHANG X M, KANO M, LI Y. Principal polynomial analysis for fault detection and diagnosis of industrial processes[J]. **IEEE Access**, 2018, 6: 52298-52307.
- [20] ZHANG X M, LI Y. Multiway principal polynomial analysis for semiconductor manufacturing process fault detection[J]. **Chemometrics and Intelligent Laboratory Systems**, 2018, 181: 29-35.
- [21] HE X F, CAI D, YAN S C, *et al.* Neighborhood preserving embedding[C]//**Tenth IEEE International Conference on Computer Vision**. Beijing: IEEE, 2005: 1208-1213.
- [22] DOWNS J J, VOGEL E F. A plant-wide industrial process control problem[J]. **Computers & Chemical Engineering**, 1993, 17(3): 245-255.
- [23] GE Z Q, SONG Z H. Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors[J]. **Industrial & Engineering Chemistry Research**, 2007, 46(7): 2054-2063.
- [24] YIN S, DING S X, HAGHANI A, *et al.* A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process[J]. **Journal of Process Control**, 2012, 22(9): 1567-1581.
- [25] JIANG L, SONG Z H, GE Z Q, *et al.* Robust self-supervised model and its application for fault detection[J]. **Industrial & Engineering Chemistry Research**, 2017, 56(26): 7503-7515.

(本文编辑:石易文)