

1-2-2019

## Chemical Process Monitoring Based on DNPE-SVDD

Xiaochun Han

*College of Electrical Engineering and Control Science, Nanjing TECH University, Nanjing 211816, China;*

Cuimei Bo

*College of Electrical Engineering and Control Science, Nanjing TECH University, Nanjing 211816, China;*

Yi Hui

*College of Electrical Engineering and Control Science, Nanjing TECH University, Nanjing 211816, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

---

## Chemical Process Monitoring Based on DNPE-SVDD

### Abstract

*Abstract: Chemical processes test dataset are high dimensional, and have the combined feature of nonlinear and dynamic characteristics. However, traditional linear dimension reduction algorithm cannot extract the local structure information and dynamic characteristics. The monitoring model of chemical process based on dynamic neighborhood preserving embedding-support vector data description (DNPE-SVDD) algorithm is proposed. With the superiority of NPE in nonlinear dimensionality reduction and SVDD in the detection of outliers, dimension is reduced by DNPE algorithm and the monitoring model of the manifold space with reduced dimension is established by SVDD algorithm. The Tennessee Eastman (TE) process is simulated using the proposed model. Compared with DPCA and DNPE algorithms, the simulation results show that the DNPE-SVDD has a higher accuracy of fault detection.*

### Keywords

neighborhood preserving embedding, support vector data description, dimensionality reduction, process monitoring

### Recommended Citation

Han Xiaochun, Bo Cuimei, Yi Hui. Chemical Process Monitoring Based on DNPE-SVDD[J]. Journal of System Simulation, 2018, 30(1): 184-190.

# 基于 DNPE-SVDD 的化工过程监控

韩晓春, 薄翠梅, 易辉

(南京工业大学电气工程与控制科学学院, 江苏 南京 211816)

**摘要:** 针对化工过程中检测数据变量维数高、非线性与动态特性相结合的特点, 而传统的线性降维算法不能提取局部结构信息和动态特性, 提出了基于动态邻域保持嵌入-支持向量数据描述 (DNPE-SVDD) 算法的化工过程监控模型。结合 DNPE 在非线性和 SVDD 在异常点检测的优势, 使用 DNPE 算法进行维数约减, 对降维后的流形空间采用 SVDD 算法建立监控模型, 通过 Tennessee Eastman (TE) 化工过程进行仿真研究, 同时与 DPCA、DNPE 算法对比验证所提算法的性能, 结果表明 DNPE-SVDD 能获得更高的故障检测准确率。

**关键词:** 邻域保持嵌入; 支持向量数据描述; 数据降维; 过程监控

中图分类号: TP277

文献标识码: A

文章编号: 1004-731X (2018) 01-0184-07

DOI: 10.16182/j.issn1004731x.joss.201801023

## Chemical Process Monitoring Based on DNPE-SVDD

Han Xiaochun, Bo Cuimei, Yi Hui

(College of Electrical Engineering and Control Science, Nanjing TECH University, Nanjing 211816, China)

**Abstract:** Chemical processes test dataset are high dimensional, and have the combined feature of nonlinear and dynamic characteristics. However, traditional linear dimension reduction algorithm cannot extract the local structure information and dynamic characteristics. The monitoring model of chemical process based on dynamic neighborhood preserving embedding-support vector data description (DNPE-SVDD) algorithm is proposed. With the superiority of NPE in nonlinear dimensionality reduction and SVDD in the detection of outliers, dimension is reduced by DNPE algorithm and the monitoring model of the manifold space with reduced dimension is established by SVDD algorithm. The Tennessee Eastman (TE) process is simulated using the proposed model. Compared with DPCA and DNPE algorithms, the simulation results show that the DNPE-SVDD has a higher accuracy of fault detection.

**Keywords:** neighborhood preserving embedding; support vector data description; dimensionality reduction; process monitoring

## 引言

随着化工过程控制系统的日趋复杂, 过程中数据变量存在强非线性、强耦合性和多模态的特性,

使得过程数据不能直接反映过程的运行状态, 故障检测的性能较低<sup>[1]</sup>。传统的线性降维算法, 如 PCA<sup>[2]</sup>、ICA<sup>[3]</sup>, 只限于线性变换, 描述数据的全局结构, 无法探测到隐藏在数据空间的内在非线性子流形, 而流形学习非线性降维算法能够发现高维数据集分布的内在规律, 保持了数据的局部近邻结构不变。其中, 邻域保持嵌入算法(NPE)<sup>[4]</sup>不仅能够找到高维训练数据的低维表示, 而且对于测试数



收稿日期: 2015-10-15 修回日期: 2015-11-09;  
基金项目: 国家自然科学基金(61203020, 61503181),  
江苏省自然科学基金(BK20140953), 江苏省高校自  
然科学基金(13KJB510013);  
作者简介: 韩晓春(1989-), 女, 山东莱芜, 硕士  
生, 研究方向为化工过程故障诊断。

<http://www.china-simulation.com>

据可以获得清晰映射,能很好的解决流形学习算法中存在的“out of sample”问题<sup>[5]</sup>,引起了人们的广泛关注。

NPE 是局部线性嵌入算法(LLE)<sup>[6]</sup>的线性表示,其思想是利用局部的线性结构来逼近全局的非线性特点,应用基础是数据样本独立分布,忽视了数据随时间变化的特征信息,而在实际的化工生产过程中,数据样本往往呈现出一定的时间相关性<sup>[7]</sup>。为了更好的体现过程数据的动态特性,对数据进行预处理,将扩展的动态特性矩阵代替原始数据集,然后使用邻域保持嵌入算法提取高维数据的低维子流形。支持向量数据描述(SVDD)<sup>[8]</sup>可以实现对目标数据集的超球形描述,可以用于样本分类和异类点检测,是一种单类分类器,具有较好的非线性处理能力。文献[9]提出了采用增量 LLE 算法来求取映射矩阵的方法,增加了计算量,降低了运算速度。文献[10]提出了一种基于 PCA-SVDD 的故障检测模型,采用 PCA 实现数据的线性降维,忽略了化工过程中非线性特性。

针对化工生产过程的非线性、强耦合性及其动态特性,本文采用 DNPE 提取过程数据的动态特性,将原始观测数据集降到低维子空间中,将特征空间的数据利用 SVDD 建立故障检测模型,并确定监控统计量及其控制限。本文应用 TE<sup>[11]</sup>化工过程,与 DPCA、DNPE 做对比,评估所提算法的监控性能。

## 1 动态邻域保持嵌入(DNPE)

### 1.1 DPNE 算法描述

给定  $D$  维数据样本集  $X=\{x_1, x_2, \dots, x_n\}$ ,  $x_i \in R^D$ , 通过动态邻域保持嵌入(DNPE)算法将  $X$  映射到数据集  $Y=\{y_1, y_2, \dots, y_n\}$ ,  $y_i \in R^d$ , ( $d < D$ )。DNPE 算法实现过程如下:

(1) 构造扩展矩阵。采用前  $h$  个观测值构造扩展矩阵  $X_A(h)$  如下:

$$X_A(h) = \begin{bmatrix} X_p & X_{p-1} & \dots & X_{p-h} \end{bmatrix} = \begin{bmatrix} x_1^T & x_0^T & \dots & x_{1-h}^T \\ x_2^T & x_1^T & \dots & x_{2-h}^T \\ \vdots & \vdots & \ddots & \vdots \\ x_n^T & x_{n-1}^T & \dots & x_{n-h}^T \end{bmatrix} \quad (1)$$

式中:  $h$  为动态步数,当  $h \in \{1, 2, 3\}$  时,模型能够满足系统的过程监控要求<sup>[12]</sup>。

(2) 构造邻域连接图。对每个样本点,用欧式距离确定它的  $k$  个近邻。

(3) 重构权值矩阵。在每个样本点及它的邻域点之间重构权向量,并在低维空间中保持每个邻域的权值不变,使误差函数  $\varepsilon(W_i)$  达到最小。其中

$$\varepsilon(W_i) = \arg \min \left\| x_i - \sum_{j=1}^k w_{ij} x_j \right\|^2 \quad (2)$$

$$\sum_{j=1}^k w_{ij} = 1 \quad (3)$$

如果  $x_j$  不属于  $x_i$  的近邻,则  $w_{ij}=0$ 。

(4) 计算特征映射矩阵。根据高维空间中的样本点与它的近邻点之间的权值矩阵  $W$ ,嵌入低维空间的投影矩阵  $A$  可以通过求解下式的最小化问题得到:

$$A_{opt} = \arg \min_a \left( \sum_{i=1}^n \left\| y_i - \sum_{j=1}^k W_{ij} y_j \right\|^2 \right) \quad (4)$$

利用拉格朗日乘子法来求解以上优化问题,即:

$$XMX^T a = \lambda XX^T a \quad (5)$$

式中  $M=(I-W)^T(I-W)$ 。求解式(5)中最小的  $d$  个特征值所对应的特征向量组成特征映射矩阵  $A(a_1, a_2, \dots, a_d) \in R^{D \times d}$ 。

### 1.2 统计量及控制限的确定

采用 DNPE 进行过程监控时,分别在特征空间和残差空间建立 Hotelling  $T^2$  和平方预测误差(SPE)统计量<sup>[13]</sup>。 $T^2$  统计量用来度量数据在模型空间的波动幅度,定义为:

$$T^2 = y_{new} \Lambda^{-1} y_{new}^T \leq \frac{d(n-1)}{n-d} F(d, n-d, \alpha) \quad (6)$$

其中,  $y_{new}=Ax_{new}$ ,  $\Lambda^{-1}=[Y^T Y/(n-1)]^{-1}$ , 为样本  $Y$

的协方差矩阵,  $F(d, n-d, \alpha)$ 为自由度为  $d$  和  $n-d$ , 置信水平为  $\alpha$  的  $F$  分布。SPE 用来度量过程的随机变化, 定义为:

$$SPE = \|x_{\text{new}} - \hat{x}_{\text{new}}\|^2 = e^T e \leq g \chi_h^2 \quad (7)$$

其中,  $g=b/2a$ ,  $h=2a^2/b$ ,  $a$  和  $b$  分别为训练样本 SPE 统计量的均值和方差。

## 2 SVDD

根据传统方法得到的  $T^2$  和 SPE 统计量指标采用向量和的形式来代替整个数据集的变量, 不可避免的增加了噪声的干扰, 因此采用 SVDD 算法构造监控统计量进行过程监控<sup>[14]</sup>。

SVDD 的基本思想是确定一个超球面, 使得所有训练样本尽可能多的包含在超球面内, 超球面由其中心  $a$  和半径  $R$  表示, 设训练数据集  $\{x_i\}$ ,  $x_i \in R^d, i=1, 2, \dots, n$ , 则满足条件的超平面可以通过下面的优化问题获得:

$$\min f(R, a, \xi) = R^2 + C \sum_{i=1}^n \xi_i \quad (8)$$

$$\text{s.t. } \|y_i - a\|^2 = R^2 + \xi_i \quad \xi_i \geq 0, i=1, 2, \dots, n$$

其中,  $C$  为错分样本的惩罚系数,  $\xi_i$  为松弛因子。

上式转化为 Lagrange 极值问题如下式所示:

$$L(R, a, \xi, \alpha, r) = \sum_{i=1}^n \alpha_i K(x_i \cdot x_i) - \sum_{i=1, j=1}^n \alpha_i \alpha_j K(x_i \cdot x_j) \quad (9)$$

其中,  $0 \leq \alpha_i \leq C$ ,  $\sum_{i=1}^n \alpha_i = 1$ 。

通过核函数将低维空间的非线性问题转化为高维空间的线性问题, 则超平面半径  $R$  可以表示为:

$$R^2 = K(x_k, x_k) - 2 \sum_{i=1}^n \alpha_i (x_i \cdot x_k) + \sum_{i=1, j=1}^n \alpha_i \alpha_j (x_i \cdot x_j) \quad (10)$$

当新的测试样本  $z$  满足下式条件时, 则测试点被接受, 否则被拒绝。

$$\|z - a\|^2 = (z \cdot z) - 2 \sum_{i=1}^n \alpha_i (z \cdot x_i) + \sum_{i=1, j=1}^n \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2 \quad (11)$$

在特征空间中构造统计量对过程进行监控, 定义统计量及其控制限如下:

$$L = \|z - a\|^2 \leq R^2 \quad (12)$$

## 3 基于 DNPE-SVDD 的故障检测步骤

### A. 离线建模

(1) 获取正常工况下的数据集并进行归一化处理;

(2) 确定动态步数  $h$ , 计算扩展矩阵  $X_A(h)$ ;

(3) 采用 DNPE 算法对扩展矩阵进行特征提取, 并获得特征映射矩阵  $A(a_1, a_2, \dots, a_d) \in R^{D \times d}$ ;

(4) 根据公式  $y = Ax$ , 计算低维嵌入空间  $Y = \{y_1, y_2, \dots, y_n\}$ ;

(5) 根据低维嵌入空间  $Y$  建立 SVDD 模型。

### B. 在线监控

(1) 获取新的观测数据集  $X_{\text{new}}$ , 并根据在正常工况下得到的均值和方差进行归一化处理;

(2) 根据确定的动态步数  $h$ , 计算新的扩展矩阵  $X_{A_{\text{new}}}(h)$ ;

(3) 根据公式  $y_{\text{new}} = Ax_{\text{new}}$ , 计算低维嵌入空间  $Y_{\text{new}}$ ;

(4) 计算  $Y_{\text{new}}$  中每个样本点到超平面中心的距离;

(5) 计算监控统计量  $L$ , 判断是否超过控制限。

## 4 仿真研究

### 4.1 实例仿真

本文采用一个典型的非线性模型验证算法的有效性, 具体模型结构如下<sup>[15]</sup>:

$$\begin{aligned} x_1 &= t + e_1 \\ x_2 &= t^2 - 3t + e_2 \\ x_3 &= -t^3 + 3t^2 + e_3 \end{aligned} \quad (13)$$

其中,  $t \in [0.01, 4]$ ,  $e_1, e_2, e_3$  是 3 个相互独立的噪

声信号, 服从均值为 0、方差为 0.5 的高斯分布。设定的故障类型如表 1 所示。

表 1 故障的具体信息  
Tab. 1 Specific information of the failure

故障	故障描述	故障类型
1	从 $t=200$ 时刻起给 $x_2$ 加入一个幅值为 -6 的信号	阶跃
2	从 $t=200$ 时刻起给 $x_1$ 加入一个 $0.02(t-200)$ 的信号	斜坡

对故障 1 的监控结果如图 1 所示, 图 1(a)、图 1(b) 分别采用 PCA、DNPE-SVDD 算法进行故障监控, 从图中可以看出, PCA 的  $SPE$  统计量能够检测出故障的发生,  $T^2$  统计量在故障发生时做出一定的反应, 但是故障检测的精确度低, 且故障发生一段时间后无法进行有效的检测。DNPE-SVDD 的  $L$  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, 能够迅速准确地监控故障的发生。

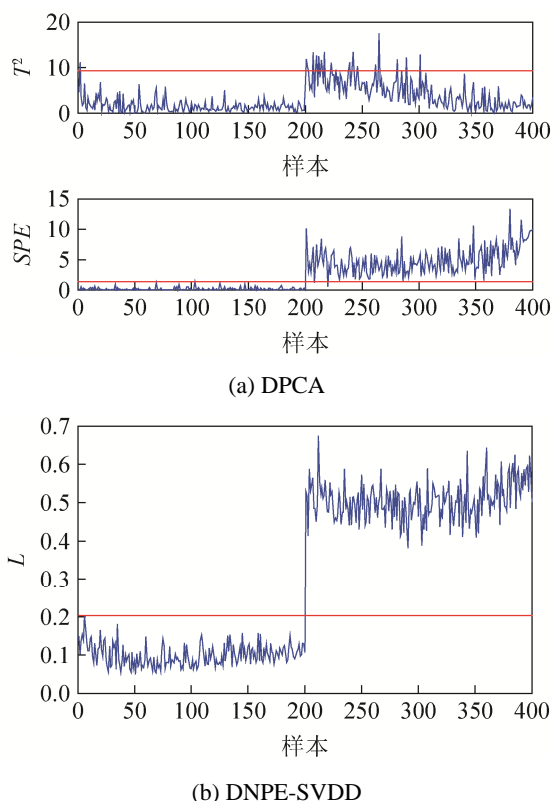


图 1 故障 1 的监控结果  
Fig.1 Monitoring results of fault 1

对故障 2 的监控结果如图 2 所示, 从图 2(a)

中可以看出, PCA 的  $T^2$  统计量能够检测出故障的发生,  $SPE$  统计量却检测不出故障, 不能表示数据集的残差空间。从图 2(b) 中可以看出, DNPE-SVDD 算法在故障发生后能迅速报警。

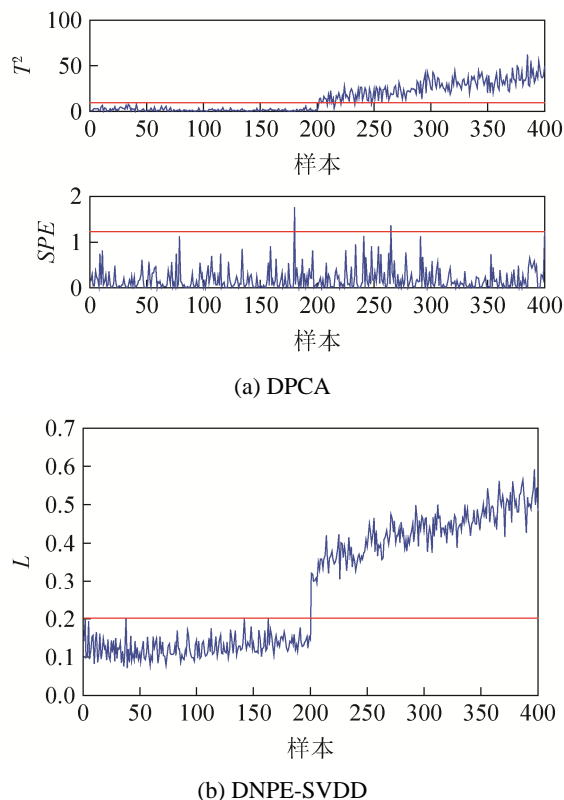


图 2 故障 2 的监控结果  
Fig.2 Monitoring results of fault 2

从数值案例仿真图中可以看出, DNPE-SVDD 能够补充 PCA 在  $T^2$  或  $SPE$  统计量监控的不足, 具有较好的非线性处理能力。

## 4.2 TE 过程仿真

田纳西-伊斯曼过程(TE)是一个实际工艺流程的标准测试过程, 有 4 种反应物、2 种产物、1 种恒量组分和 1 种副产物组成。其生产过程包括 5 个操作单元: 反应器、冷凝器、气液分离器、循环压缩机和汽提塔<sup>[16]</sup>。运行代码和故障数据来源于 <http://depts.washington.edu/control/LARRY/TE/download.html#Topics>。

TE 过程的故障仿真数据有 21 个故障状态, 在这些故障中, 有 16 个已知故障和 5 个未知故障状

态, 其中已知故障的故障类型有阶跃、随机变化、慢漂移和粘滞等。模拟正常工况时, 采集过程平稳运行下的 960 个数据样本作为训练集, 采集故障数据作为测试集时, 每种故障模式均采集 960 个样本点, 前 160 个点为正常样本, 从 161 个样本点开始引入故障。

本文将所提的方法与 DPCA 和 DNPE 算法相对比, 首先建立统计监控模型, 在建立 DPCA 模型时, 使用主元贡献度准则为 85% 确定主元个数, 在 DNPE 模型中, 设定近邻个数为  $k=33$ , 嵌入空间的维数为  $d=25$ , 作为对比, DNPE-SVDD 模型中设定相同的近邻个数和嵌入空间维数, 采用 99% 的控制限, 然后将这 3 种算法均对 TE 过程的 21 种故障进行了测试。

图 3 和图 4 分别绘出了这 3 种方法对故障 5、10 的过程监控图。故障 5 是受控制器补偿引起的故障, 表现为冷凝器冷却水入口温度阶跃变化, 故障 10 是流股 2 内物料 C 的温度发生变化引起的故障, 同 DPCA、DNPE 作对比, 所提算法能够在故障发生后迅速的检测, 并能维持报警状态, 持续地对故障进行检测, 从图中可以看出, DNPE-SVDD 算法的检测得到了非常明显的提升, 说明了 DNPE 算法投影得到的特征空间包含了更多的有效信息, 具有更好的特征提取能力, 使得 DNPE-SVDD 具有更好的分类效果, 较好的故障检测性能。

故障检测的漏报率是指没有检测出故障的样本数占有故障样本数的比值, 本文采用漏报率作为故障检测效果的评价标准, 计算了 TE 过程的 21 种故障的漏报率, 如表 2 所示。从表 2 中可以看出, 与其他两种算法相对比, DNPE-SVDD 具有更好的检测效果。针对每一个故障, 检测率最高的值在表中用粗体表示, 可以看出采用 DNPE 算法时, SPE 统计量的平均误报率(average)最低, 与 DNPE-SVDD 的  $L$  统计量的平均误报率相差不大。对于故障 3、9、15, 由于数据结构的特点, 所有的算法都不能准确的检测出故障, 而在比较除去故障 3、9、15 之后的平均误报率(average\*)时, DNPE-SVDD 在

这 3 种算法中最低, 具有较好的故障检测性能。

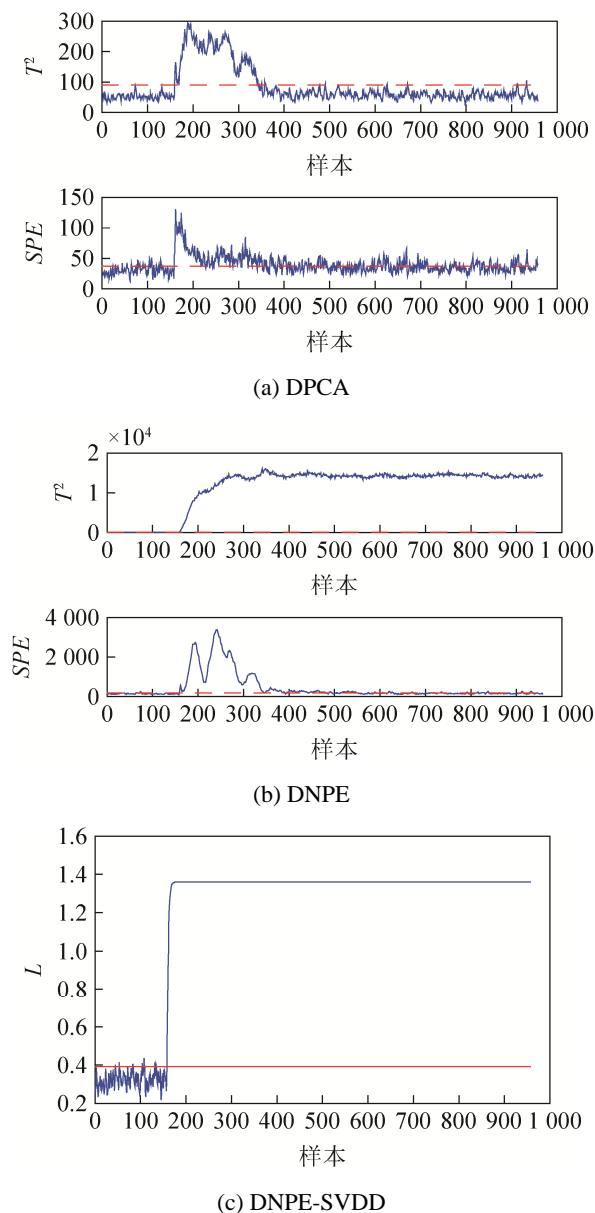
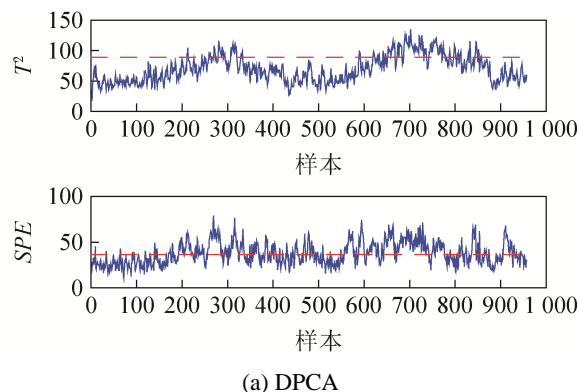


图 3 故障 5 的监控结果  
Fig.3 Monitoring results of fault 5



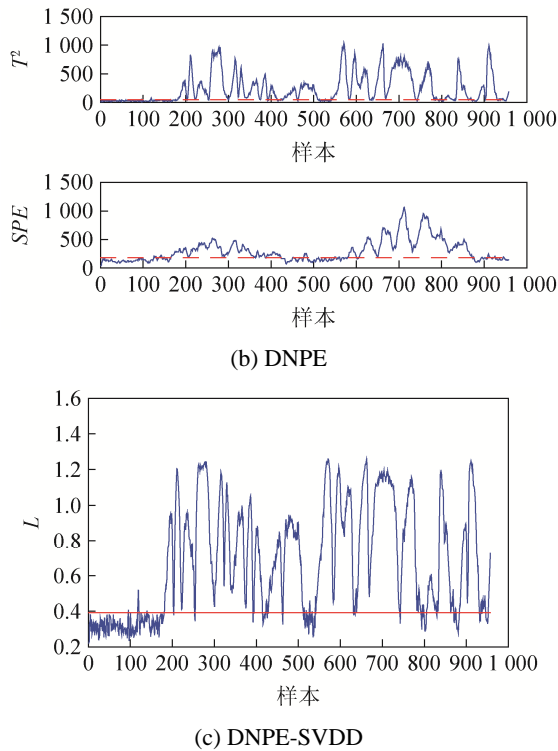


图4 故障10的监控结果

Fig.4 Monitoring results of fault 10

表2 TE过程故障检测的误报率

Tab. 2 False alarm rate of fault detection in TE process

故障类型	DPCA		DNPE		DNPE-SVDD
	$T^2$	$SPE$	$T^2$	$SPE$	
1	0.62	<b>0</b>	0.12	<b>0</b>	<b>0</b>
2	1.25	1.37	1	1.12	<b>0.75</b>
3	99.37	83.5	94.87	<b>57.75</b>	91.5
4	95.5	<b>0</b>	0.25	<b>0</b>	0.37
5	75	46.87	<b>0</b>	44	<b>0</b>
6	0.87	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
7	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
8	2.5	2.12	1.87	<b>0.5</b>	1.12
9	99.5	82.25	96.37	<b>58.12</b>	94
10	74.25	43.75	14.12	26.25	<b>12.25</b>
11	77.25	<b>6.5</b>	19.62	10.87	15.25
12	0.75	2.37	<b>0</b>	<b>0</b>	<b>0</b>
13	5.25	4.25	4.25	<b>3.5</b>	4.25
14	<b>0</b>	<b>0</b>	2.75	<b>0</b>	1.87
15	99.37	83.75	93.62	<b>62.62</b>	87.75
16	89.62	44.62	15	26.12	<b>12.37</b>
17	23.62	<b>2.37</b>	3.12	4.12	3.25
18	10.87	8.25	9.62	<b>4.87</b>	9.62
19	89.12	<b>28.87</b>	38.12	48	30.37
20	71.37	29.5	11.87	20.37	<b>11.25</b>
21	58.87	44	43.37	41	<b>37.5</b>
Average	46.42	24.49	21.42	<b>19.48</b>	19.69
Average*	37.59	14.71	9.174	12.81	<b>7.79</b>

## 5 结论

本文在 NPE 的基础上, 针对化工过程中测量值的动态特性, 采用动态邻域保持嵌入(DNPE)算法提取数据的特征值, 结合支持向量数据描述(SVDD)的分类有效性, 提出了 DNPE-SVDD 算法, 并将这一算法应用于故障监控中, 利用 SVDD 建立模型统计量进行故障监控。将 DPCA、DNPE 和 DNPE-SVDD 应用于 TE 模型进行对比分析, 结果表明 DNPE-SVDD 算法比其他算法具有更好的故障检测性能, 提高了分类准确率, 能有效的应用于过程监控中, 但是在 TE 仿真研究中, DNPE-SVDD 与其他算法一样, 对很难检测的故障 3、9、15 的检测率低。

## 参考文献:

- [1] 宋冰, 马玉鑫, 方永锋, 等. 基于 LSNPE 算法的化工过程故障检测[J]. 化工学报, 2014, 65(2): 620-627.  
Song Bing, Ma Yuxin, Fang Yongfeng, et al. Fault detection for chemical process based on LSNPE method [J]. Journal of Chemical Industry and Engineering (China), 2014, 65(2): 620-627.
- [2] 王健, 冯健, 韩志艳. 基于流形学习的局部保持 PCA 算法在故障检测中的应用[J]. 控制与决策, 2013, 28(5): 683-687.  
Wang Jian, Feng Jian, Han Zhiyan. Locally preserving PCA method based on manifold learning and its application in fault detection [J]. Control and Decision, 2013, 28(5): 683-687.
- [3] 张沐光, 宋执环. 一种基于独立元贡献度的子空间故障检测方法[J]. 控制理论与应用, 2010, 27(3): 296-302.  
Zhang Muguang, Song Zhihuan. Subspace fault detection method based on independent component contribution [J]. Control Theory & Applications, 2010, 27(3): 296-302.
- [4] Miao Aimin, Ge Zhiqiang, Song Zhihuan, et al. Time Neighborhood preserving embedding model and Its application for fault detection [J]. Industrial and Engineering Chemistry Research(S0888-5885), 2013, 52(38): 13717-13729.
- [5] 陈新忠, 胡汇涓, 王雪松. 基于加权近邻保持嵌入的高光谱数据降维方法[J]. 中国矿业大学学报, 2013, 42(6): 1066-1072.  
Chen Xinzong, Hu Huijuan, Wang Xuesong.



