

研究论文

DOI: 10.11949/j.issn.0438-1157.20170428

基于核相似度支持向量数据描述的间歇过程监测

王建林, 马琳钰, 刘伟旻, 邱科鹏, 于涛

(北京化工大学信息科学与技术学院, 北京 100029)

摘要: 基于支持向量数据描述的间歇过程监测方法选择历史过程数据中最大的核距离作为控制限, 忽略了高维空间中超球体的不规则性, 导致基于该方法的过程监测精度不高。针对上述问题, 提出了一种基于核相似度支持向量数据描述的间歇过程监测方法, 将间歇过程数据待监测样本与支持向量之间的核函数值作为相似度权重, 利用该相似度对不同时刻的支持向量球心距加权求和, 得到待监测间歇过程数据样本的动态控制限, 通过判断待监测样本的球心距是否超过其动态控制限, 实现间歇过程监测。所提方法综合考虑了超球体的不规则性和过程数据在高维空间分布的局部特性, 以及间歇过程数据待监测样本的时变性, 提高了间歇过程监测的准确性。利用数值仿真实验和半导体金属刻蚀实验验证了该方法的有效性。

关键词: 核相似度; 支持向量数据描述; 动态监测; 间歇过程

中图分类号: TQ 277

文献标志码: A

文章编号: 0438—1157 (2017) 09—3494—07

Batch process monitoring by kernel similarity-based support vector data description

WANG Jianlin, MA Linyu, LIU Weimin, QIU Kepeng, YU Tao

(College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: Kernel distance-based support vector data description (SVDD) for batch process monitoring exhibited poor monitoring precision by setting control limit from the largest kernel distance in historical process dataset but ignoring hyperspherical irregularity in high dimensional space. A kernel similarity based SVDD monitoring method was proposed for batch process monitoring. Kernel similarity was taken as kernel function value between support vectors and data samples for testing. The weighted summation of kernel similarity and distance of support vectors at various time points was utilized to set dynamic control limit for data samples of batch process to be monitored. Batch process monitoring was achieved by judging if kernel distance of test sample exceeded the dynamic control limit. This monitoring method considered irregularity of hypersphere, local distribution characteristics of process dataset in high dimensional space, and spontaneity of data samples, so that it could improve accuracy in batch process monitoring. Method effectiveness was demonstrated by numerical simulation and metal etching process in semiconductor manufacturing.

Key words: kernel similarity; support vector data description; dynamic monitoring; batch process

引言

间歇过程作为工业生产中的一种重要生产方

式, 具有生产灵活、产品附加值高等特点, 广泛应用于食品、精细化工、生物制药、金属加工等领域^[1-3]。间歇过程生产中出现异常, 会直接降低产

2017-04-19 收到初稿, 2017-05-27 收到修改稿。

联系人及第一作者: 王建林 (1965—), 男, 教授。

基金项目: 国家自然科学基金项目 (61240047); 北京市自然科学基金项目 (4152041)。

Received date: 2017-04-19.

Corresponding author: Prof. WANG Jianlin, wangjl@mail.buct.edu.cn

Foundation item: supported by the National Natural Science Foundation of China (61240047) and the Beijing Natural Science Foundation (4152041).

品质量,造成经济损失,甚至会引发安全事故,因此,对间歇过程实施过程监测,能够有效保障间歇过程安全生产、可靠运行,提高产品质量^[4-5]。

数据驱动的多元统计间歇过程监测方法通过获取单一或多个监测量在时间维度上的变化,并设定控制限,监测量超过控制限则认为过程发生了异常^[6-11]。MPCA^[12] (multiway principal component analysis)、MPLS^[13] (multiway partial least squares)、MICA^[14] (multiway independent component analysis)等方法均侧重间歇过程数据的多批次特性,基于变量投影分解原理实现过程监测;针对大多数间歇过程具有的多时段特性,相继出现了 sub-PCA^[15]、基于时段的 PLS^[16]等方法。然而,这些基于 PCA 与 PLS 的间歇过程监测方法需假设过程变量为正态分布且变量间线性相关,若过程数据结构较为复杂,不满足上述假设,这些方法的有效性及其过程监测精度会受到影响^[17-19]。对于间歇过程数据的非线性特性,核函数被引入到过程监测方法中,如 MKPCA^[20] (multiway kernel PCA)与 MKPLS^[21] (multiway kernel PLS)方法,利用核函数将非线性过程数据嵌入特征空间,使其在特征空间中呈现线性关系,然而这类方法的控制限设定仍需假设变量服从高斯分布,制约了间歇过程监测精度的提升。

支持向量数据描述 (support vector data description, SVDD) 是一种有效的数据描述和处理方法,通过非线性变换将正常数据样本空间映射到高维样本空间,实现对数据的分类,已在间歇过程监测中得到应用。Sun 等^[22]提出了一种基于核距离多变量控制图的过程监测方法 (kernel-distance-based multivariate control chart, K chart),该方法对过程数据分布不作假设,且不要求变量间线性相关,克服了传统过程监测方法的缺陷。但是该方法选择了历史过程数据中最大的核距离作为异常判定阈值,未考虑高维空间中超球体的不规则性以及过程数据在高维空间分布的局部特性,因此降低了间歇过程监测精度。Camci 等^[23]提出一种启发式参数选择方法,通过优化核函数的超参数来调节 K chart 方法的控制限;Ning 等^[24]在此基础上,同时考虑多个参数对控制限的影响,实现对控制限的优化设置。然而,上述基于参数优化方法的控制限沿用了支持向量的最大球心距,未考虑局部数据对超球结构的影响,忽略了超球体的不规整性。针对该问题,Sukchotrat 等^[25]提出基于支持向量球心距排序的控制限设计方法,通过权衡样本误分率与超球半径大

小,选择满足给定排序比例的超球半径均值作为异常控制限,获得了更为合理的监测控制限;Khediri 等^[26]在确定控制限时引入了可接受性参数,用以涵盖训练样本未能表达的部分数据,实现控制限的有效性修正。但这两种方法均引入了新的设定参数,对监测效果影响较大,也未给出参数选择方法,使其应用受到限制。

提出了一种基于核相似度 SVDD 的间歇过程监测方法,使用正常工况下获取的间歇过程数据作为训练样本,采用 SVDD 构造包含训练样本的最小超球体,将待监测样本与支持向量之间的核函数值作为相似度权重,利用该相似度对不同时刻的支持向量球心距加权求和,得到待监测间歇过程数据样本的动态控制限,通过判断待监测样本的球心距是否超过其动态控制限,实现间歇过程监测。

1 基于 SVDD 的过程监测

1.1 支持向量数据描述 (SVDD)

SVDD 是由 Tax 和 Duin^[27]提出的一种数据描述与处理方法,其核心是在高维空间中构造包含被描述对象的最小体积封闭超球。

给定训练数据集 $\mathbf{x}_i \in R^m, i=1,2,\dots,n$, 构造包含 $\mathbf{x}_i$ 的超球体转化为如式(1)的最优化问题。

$$\begin{aligned} \min_{R, \mathbf{a}, \xi} \quad & R^2 + C \sum_i \xi_i \\ \text{s.t.} \quad & \|\Phi(\mathbf{x}_i) - \mathbf{a}\|^2 \leq R^2 + \xi_i, \xi_i \geq 0, i=1,2,\dots,n \end{aligned} \quad (1)$$

其中, $\mathbf{a}$ 与 R 分别是超球体的球心和半径; C 和 ξ_i 分别为惩罚系数和松弛因子,是针对训练样本中可能含有的离群点而引入的参数。

式(2)为式(1)最优化问题的对偶形式;引入式(3)所示的高斯核函数能够简化内积运算,完成特征空间映射;惩罚系数 C 的参考值如式(4)所示,其中 f 为训练样本处于超球体外部的百分比, n 为训练样本总数,通过参数 C 可以协调球体体积与训练样本误分率。式(2)~式(4)中, α_i 为 Lagrange 乘子, K 为核函数, δ 为高斯核函数宽度参数。

$$\max_{\alpha_i} \quad \sum_i \alpha_i K(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (2)$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C, \sum_i \alpha_i = 1$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp \left[-(\mathbf{x}_i - \mathbf{x}_j)^2 / 2\delta^2 \right] \quad (3)$$

$$C = \frac{1}{nf} \quad (4)$$

对应的 Lagrange 乘子满足 $0 < \alpha_i < C$ 的 $\mathbf{x}_i$ 为处于

球体表面的支持向量, 对满足该条件的支持向量索引 i 进行筛选, 组成整数集合 s ; 由式(5)计算支持向量超球球心距, 即超球体的半径 R , 其中 $\mathbf{x}_o$ 为任一支持向量, $o \in s$ 。

$$R^2 = 1 - 2 \sum_i \alpha_i K(\mathbf{x}_i \cdot \mathbf{x}_o) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \quad (5)$$

利用式(6)计算测试样本 $\mathbf{z}$ 到球体中心 $\mathbf{a}$ 的距离, 当该距离小于或等于球体半径 R 时认为测试样本 $\mathbf{z}$ 与训练样本 $\mathbf{x}_i$ 为同一类数据。

$$\|\mathbf{z} - \mathbf{a}\|^2 = 1 - 2 \sum_i \alpha_i K(\mathbf{z} \cdot \mathbf{x}_i) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \quad (6)$$

SVDD 对数据没有高斯分布、相互独立或呈线性关系的要求, 适用范围较广^[28-29]; 其球形分类边界能够实现更高阶的数据边界描述。同时, SVDD 中测试样本 $\mathbf{z}$ 与中心 $\mathbf{a}$ 的距离是一种有效的监控量, 距离越大说明 $\mathbf{z}$ 点发生异常的可能性越大。

1.2 基于 SVDD 的过程监测

基于 SVDD 的过程监测方法利用正常过程数据构造训练样本, 读入待监测过程数据作为测试样本; 使用支持向量 $\mathbf{x}_{i \in s}$ 替代训练样本 $\mathbf{x}_i$ 简化原算式, 简化后的测试样本核距离 kd_z 计算式如(7), 若该距离小于最大的支持向量球心距 R_m , 则认为该过程点正常, 判别条件如式(8)。

$$kd_z = \sqrt{1 - 2 \sum_{i \in s} \alpha_i K(\mathbf{z} \cdot \mathbf{x}_i) + \sum_{i,j \in s} \alpha_i \alpha_j K(\mathbf{x}_i \cdot \mathbf{x}_j)} \quad (7)$$

$$kd_z \leq R_m \quad (8)$$

基于 SVDD 的过程监测方法能够处理高维、非线性过程数据, 不要求过程数据服从高斯分布, 相比传统方法能够提高过程监测水平。但是, 对于不同支持向量计算所得的不同球心距, 该方法选择了最大的支持向量球心距作为控制限, 没有考虑高维空间中超球体的不规则性, 以及过程数据在高维空间分布的局部特性和待监测过程数据样本的时变性, 限制了该方法的应用范围。

2 基于核相似度 SVDD 的间歇过程监测

2.1 核相似度 SVDD

由于高维空间中超球形状的不规则性, 各支持向量对应的超球半径各不相同, 现有的基于 SVDD 的控制限设置方法忽略了这种不规则性, 将所有支持向量对应的超球半径 R 取均值或最大值, 作为类别划分边界。对于不同子集之间数据特性差异较大的数据集, 这类固定的控制限无法兼顾到不同局部的特性, 因此导致较多的误判; 另外, 当数据集具

有时变特性, 固定的控制限无法适应其数值特征随时间的变化, 同样会出现较多误判。因此, 本研究提出核相似度 SVDD 方法, 根据测试样本在高维空间中与支持向量的位置相似性, 设定针对于该测试样本的判别控制限。

过程数据到特征空间的投影是通过核函数完成的, 向量之间的核函数值能够度量两个向量间的相似性。选择测试样本 $\mathbf{z}$ 与支持向量 $\mathbf{x}_{i \in s}$ 间的核函数值作为相似性的度量量, 核相似度权重 w_{kel} 的定义式如(9)所示, 由式(10)对核相似度权重进行标准化处理, 获得核相似度权值 $\hat{w}_{kel}$, 为控制限的选择提供依据, 其中 L 为支持向量个数。

$$w_{kel} = K(\mathbf{x}_{i \in s}, \mathbf{z}) \quad (9)$$

$$\hat{w}_{kel} = \frac{w_{kel}}{\sum_{l=1}^L w_{kel}(l)} \quad (10)$$

由式(11)计算得到测试样本 $\mathbf{z}$ 对应的判别限, 即每个支持向量的球心距与测试样本的核相似度加权求和, 得到的动态监测边界 R_z , 式中 $\mathbf{R}^T$ 表示 $\mathbf{R}$ 的转秩, $\mathbf{R}$ 为支持向量球心距集合, 如式(12)所示, R_l 为 $\mathbf{x}_{i \in s}$ 中第 l 个支持向量的球心距, 计算如式(13)所示。

$$R_z = \hat{w}_{kel} \mathbf{R}^T \quad (11)$$

$$\mathbf{R} = [R_l(\mathbf{x}_i), i \in s], l = 1, 2, \dots, L \quad (12)$$

$$R_l = \sqrt{1 - 2 \sum_{i \in s} \alpha_i K(\mathbf{x}_i, \mathbf{x}_{s(l)}) + \sum_{i,j \in s} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)} \quad (13)$$

对比式(7)获得的测试样本球心距 kd_z 和上述计算所得的动态监测边界 R_z , 由式(14)判断该样本点的核距离是否超出了该点对应的动态判别限。当满足式(14)时, 则判断样本点 $\mathbf{z}$ 的数据类型不同于训练数据集。

$$kd_z > R_z \quad (14)$$

核相似度 SVDD 改进了控制限的计算方式, 其训练样本计算量与 SVDD 相同, 获取动态监测边界 R_z 时不增加训练样本的计算量; 对过程数据在高维空间分布的局部特性和支持向量球心距造成的不规则性, 通过核函数量化测试样本与支持向量间的比邻关系, 动态调整各支持向量球心距的权重作为局部控制限的计算依据; 采用测试样本参与自身控制限计算, 使得控制限能够反映测试样本的时变性。

2.2 基于核相似度 SVDD 的间歇过程监测

核相似度 SVDD 同时考虑了超球模型的局部特征和测试样本的特性, 建立的动态异常判别阈值能够自动适应数据特性的变化, 实现间歇过程监测。图 1 为基于核相似度 SVDD 的间歇过程监测

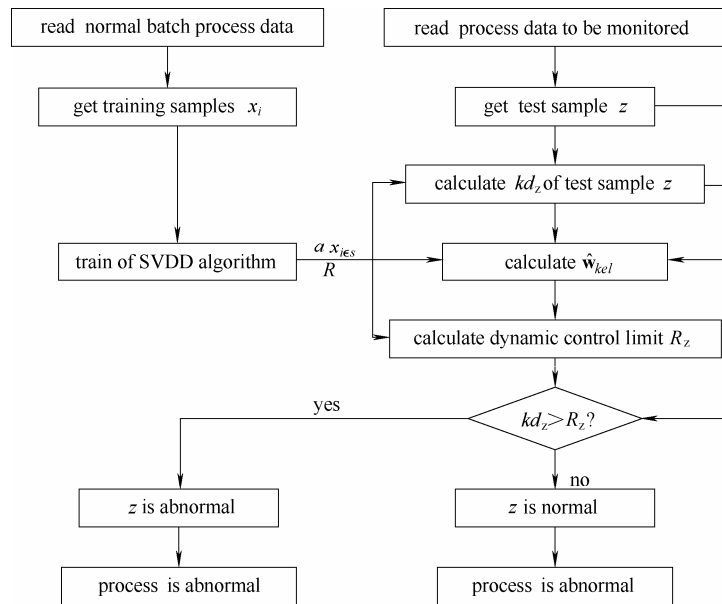


图1 基于核相似度 SVDD 的间歇过程监测流程

Fig.1 Flow chart of proposed method for batch process monitoring

流程图。

基于核相似度 SVDD 的间歇过程监测实现步骤如下：

(1) 在正常工况的间歇过程中，采集多批次过程数据 $\hat{\mathbf{X}}(I \times J \times K)$ ， I 为批次数， K 为每批次采样长度， J 为监测变量个数。将多个批次的过程数据沿变量方向展开，得到新的二维数据矩阵 $\mathbf{X}(IK \times J) = \{\mathbf{x}_i^{1 \times J}, i=1, 2, \dots, IK\}$ ，并将 $\mathbf{x}_i$ 作为训练样本。求解包含训练样本的最小体积超球，得到支持向量集合及其对应的 Lagrange 乘子 α_i ，进而计算每个支持向量所对应的超球球心距 R_i 。

(2) 读入待监测间歇过程数据，构造测试样本 $\mathbf{z}$ ， $\mathbf{z}$ 与单个训练数据样本具有相同的监测变量个数 J 。由式(6)计算 $\mathbf{z}$ 的球心距 kd_z 。

(3) 计算测试样本 $\mathbf{z}$ 与各支持向量 $\mathbf{x}_{is}^e$ 之间的核函数值 w_{kel} ，作为控制限的权重，并进行标准化处理，得到核相似度权重 $\hat{w}_{kel}$ 。将标准化的权重与支持向量球心距 R 加权求和，得到待监测样本所对应的监测动态边界阈值 R_z 。

(4) 对比待监测样本球心距 kd_z 与动态监测阈值 R_z ，若待监测样本球心距超过阈值，即判定该过程点存在异常。

3 实验验证

3.1 数值仿真

采用如下人工数据集^[30]对具有时段特性的批

次过程进行异常监测，该过程如式(15)所示，分 P1、P2 两个阶段。式中， i 为批次编号， k 为样本点数，3 个过程变量 x_1 、 x_2 、 x_3 由服从均值分布的随机变量 t 生成， t 在第 1、2 阶段的分布区间分别为 $[0.01, 2]$ 和 $[1.5, 4]$ ， v_1 、 v_2 、 v_3 是均值为零、方差为 0.01 的高斯噪声变量。训练样本均使用正常过程数据。

$$\begin{aligned}
 & x_1(i, k) = t(i, k) + v_1 \\
 \text{P1: } & x_2(i, k) = t^2(i, k) - 3t(i, k) + v_2 \\
 & x_3(i, k) = -t^3(i, k) + 3t^2(i, k) + v_3 \\
 & x_1(i, k) = t(i, k) + v_1 \\
 \text{P2: } & x_2(i, k) = -t^3(i, k) - 3t(i, k) + v_2 \\
 & x_3(i, k) = t^2(i, k) + 3t(i, k) + v_3
 \end{aligned} \tag{15}$$

在上述过程中引入异常点生成测试样本如式(16)，其中 K 为单批次样本点总数，设定 K 均为偶数。

$$\begin{aligned}
 \text{P1: } & x_2(f, k) = t^2(f, k) - 3t(f, k) - 0.5 + v_2 \\
 & k \in [K/4 + 1, K/2] \\
 \text{P2: } & x_1(f, k) = t(i, k) + 0.025(k - 75) + v_1 \\
 & k \in [3K/4 + 1, K]
 \end{aligned} \tag{16}$$

随机生成 5 组数据，每组的训练样本含 10 个正常批次、每批次 200 个样本点；每组的测试样本为单批次、长度 2000 点。表 1 给出监测结果，其中 Avg、Max 和 Dyn 分别代表支持向量球心距均值、K chart 和本文方法的误分率 (error rate)，其计算公式如式(17)，代表错误分类的测试实例个数占测试实例总数的比例，其中 FN (false negative) 为被错误分类的负例个数，即故障样本漏报个数；FP (false

表 1 不同样本在不同实验条件下的误分率与误分率标准差

Table 1 Misclassification rates with different parameters under deferent control limits

Parameter		$\delta=0.5$			$\delta=1$			$\delta=2$			$\bar{\text{err}}$	$\sigma(\text{err})$
		$C=0.01$	$C=0.1$	$C=1$	$C=0.01$	$C=0.1$	$C=1$	$C=0.01$	$C=0.1$	$C=1$		
data1	Avg	0.2415	0.5100	0.4115	0.3705	0.5090	0.5120	0.1990	0.4030	0.2870	0.3826	0.1189
	Max	0.5000	0.4535	0.2510	0.3585	0.3345	0.3025	0.0505	0.3800	0.2920	0.3247	0.1293
	Dyn	0.1035	0.3560	0.2430	0.1210	0.1795	0.1595	0.0240	0.2500	0.1435	0.1756	0.0969
data2	Avg	0.4375	0.4825	0.5055	0.2240	0.5105	0.5120	0.2410	0.2665	0.2735	0.3837	0.1283
	Max	0.3805	0.3435	0.3485	0.1680	0.3475	0.3470	0.0525	0.0610	0.1470	0.2439	0.2439
	Dyn	0.2770	0.3395	0.3540	0.0630	0.2065	0.2140	0.0625	0.0790	0.1230	0.1909	0.1156
data3	Avg	0.4495	0.4475	0.5135	0.2280	0.4230	0.5070	0.2000	0.4180	0.2830	0.3855	0.1179
	Max	0.3860	0.3210	0.3665	0.2205	0.3055	0.3630	0.0450	0.3865	0.1715	0.2851	0.1170
	Dyn	0.2865	0.2870	0.3350	0.0670	0.1180	0.2145	0.0295	0.2440	0.1570	0.1932	0.1063
data4	Avg	0.2975	0.4700	0.4615	0.4470	0.5055	0.4760	0.2810	0.2785	0.2800	0.3886	0.1003
	Max	0.3755	0.3360	0.3210	0.3405	0.3600	0.3095	0.1025	0.0870	0.1265	0.2621	0.1195
	Dyn	0.1125	0.3250	0.3135	0.1590	0.2445	0.1650	0.1055	0.0995	0.1055	0.1811	0.0907
data5	Avg	0.1940	0.1410	0.5110	0.1770	0.5090	0.5125	0.1870	0.2795	0.2890	0.3111	0.1570
	Max	0.4025	0.0655	0.4610	0.1090	0.3535	0.3395	0.0605	0.0920	0.3060	0.2433	0.1597
	Dyn	0.0780	0.0695	0.3645	0.0570	0.2345	0.2035	0.0360	0.1225	0.2080	0.1526	0.1080

positive) 为被错误分类的正例个数, 即正常样本的误报个数, N 为样本总数。表 1 最后两列为误分率均值 $\bar{\text{err}}$ 与标准差 $\sigma(\text{err})$ 。本文方法在不同的参数设定下有更小的误分率均值与标准差, 监测精度得到明显提升, 且监测效果受参数选择情况的波动较小, 性能更稳定。

$$\text{error rate} = (\text{FN} + \text{FP}) / N \quad (17)$$

图 2 给出了参数设定 $\delta=2$ 、 $C=0.4$ 下, 第 1 组数据 data1 的监测结果图, 其中 DIS 为测试样本球心距, R -dynamic、 R -average 和 R -max 分别为本文方法、支持向量球心距均值和 K chart 方法的控制限。可以明显看出球心距均值与 K chart 方法在一定程度上对异常敏感, 但无法兼顾每个阶段的数据特性, 产生较多漏报或者误报, 而本文方法定义的动态边界能够很好地描述不同阶段训练样本的局部特

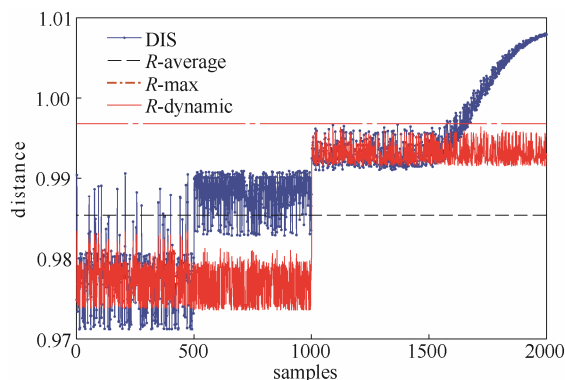


图 2 数值模型异常点监测效果

Fig.2 Monitoring result of numerical model

征, 并结合被测样本的时变性, 做到自动判别测试样本的阶段属性, 进而给出更加合理的动态控制限, 监测效果更佳。

3.2 半导体金属刻蚀工业过程

蚀刻过程是半导体制造工艺的重要环节之一, 是一个典型的非线性多时段和多工况的间歇过程。Lam9600TCP 金属蚀刻机^[31]的过程数据由 108 个正常晶片和 21 个故障晶片构成, 一个晶片制作过程为一个批次, 共 129 批次数据。该数据集来自不同的 3 个实验, 分为系列 L29、L31、L33, 数据有着不同均值和一定差异的协方差结构。数据集每个点记录了 21 个变量, 变量含义见表 2。在同样的实验条件下引入人为故障, 共生成 21 个故障批次, 故障描述见表 3。由于该数据集正常批次第 56 批次和故障批次的第 12 批次明显存在数据缺失, 故剔除不用。

表 2 半导体金属刻蚀过程变量编号及名称

Table 2 Variables of semiconductor etch process

No.	Variables	No.	Variables
1	* BCl3 Flow	11	* RF Pwr
2	* Cl2 Flow	12	* RF Impedance
3	* RF Btm Pwr	13	* TCP Tuner
4	* RF Btm Rfl Pwr	14	* TCP Phase Err
5	* Endpt A	15	* TCP Impedance
6	* He Press	16	* TCP Top Pwr
7	* Pressure	17	* TCP Rfl Pwr
8	* RF Tuner	18	* TCP Load
9	* RF Load	19	* Vat Valve
10	* RF Phase Err		

表3 金属刻蚀过程故障批次编号及故障设定

Table 3 Fault batches description of semiconductor

etch process			
Fault No.	Type	Fault No.	Type
1	TCP +50	12	RF +8
2	RF -12	13	BCl ₃ -5
3	RF +10	14	Pr +2
4	Pr +3	15	TCP -20
5	TCP +10	16	TCP -15
6	BCl ₃ +5	17	Cl ₂ -10
7	Pr -2	18	RF -12
8	Cl ₂ -5	19	BCl ₃ +10
9	He Chuck	20	Pr +1
10	TCP +30	21	TCP +20
11	Cl ₂ +5		

随机选择 30 个正常批次作为训练样本, 选择 19 个监控变量作为单个样本点输入, 对 20 个故障批次进行监测。设定误分率 $f=0.1$, 根据所选的训练样本长度 $n=3037$, 由式(4)得到 $C=0.0033$ 。选择核函数参数 $\delta=0.005$ 。在上述固定参数设定下进行监测, 表4给出了3种方法对20个故障批次的检测率, 图3、图4分别给出了故障批次5和故障批次19的监测结果图。

由于半导体金属刻蚀过程单批次数据在不同时段的特性差异较大, 支持向量球心距均值与核距

表4 金属刻蚀过程故障批次检测率

Table 4 Fault detection rates of different control limit for fault batches

Fault No.	Avg	Max	Dyn
1	0.9903	0.8835	0.9903
2	0.6415	0.1698	0.8679
3	0.7941	0.1961	0.8627
4	0.9892	0.6237	1
5	0.7653	0.1939	0.9796
6	0.6078	0.1569	0.7941
7	0.8019	0.4340	0.9906
8	0.5288	0.1058	0.8173
9	0.7576	0.1919	0.9495
10	1	0.7300	1
11	0.6947	0.2421	0.8526
13	0.7172	0.2121	0.8182
14	0.8696	0.4022	1
15	0.7800	0.4300	0.9900
16	0.8800	0.4100	0.9800
17	0.6667	0.2315	0.7870
18	0.8235	0.3431	0.9020
19	0.7451	0.3333	0.8627
20	0.8947	0.3158	0.9474
21	0.9278	0.3505	0.9691

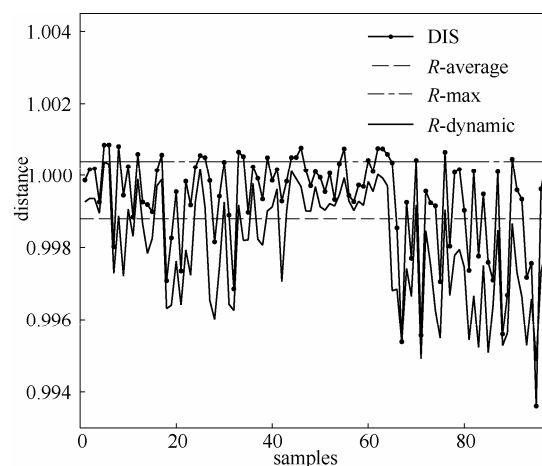


图3 No.5故障样本监测效果

Fig.3 Monitoring chart of fault batch No.5

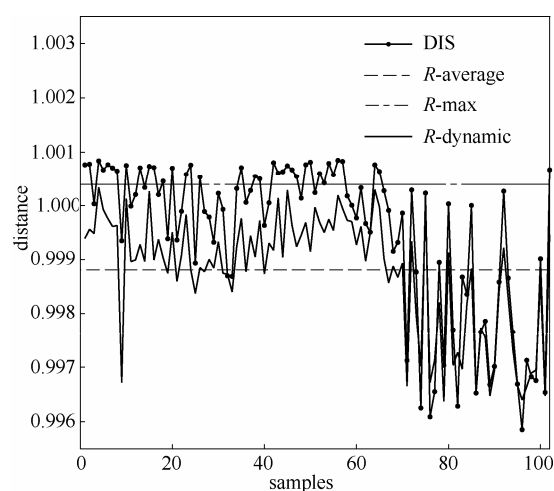


图4 No.19故障样本监测效果

Fig.4 Monitoring chart of fault batch No.19

离控制图难以在整个批次长度上都取得良好的监测效果,这两种方法均在过程后半程出现较多的漏报,同时,这两种控制限为不同批次提供数值相等的控制限,无法感知批次间的差异,因此监测精度较低。相比之下,本文方法获得的动态控制限能够很好地适应待监测样本的特性,无论是批次间还是批次内,都能对待监测样本点的球心距进行跟踪,在整个批次过程中都能保持较好的监测结果。

4 结 论

提出了一种基于核相似度 SVDD 的间歇过程监测方法。根据待监测过程数据的特性设定权重,灵活选择不同支持向量对应的球心距作为该监测点的控制限,无需新增人为设定参数,对不同时段特性差异较大的过程数据具有较好的适应性,并且能够很好地适应监测数据所具有的时变性,过程监测结果稳定。

与现有基于 SVDD 的间歇过程监测方法相比,本文所提方法在不增加训练样本计算量的同时对正常过程数据的描述更准确,为间歇过程监测提供了更合理的控制限,有效地提高了过程监测的准确性。

References

- [1] 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.
- [2] HU Y, MA H H, SHI H B. Enhanced batch process monitoring using just-in-time-learning based kernel partial least squares[J]. *Chemometrics and Intelligent Laboratory Systems*, 2013, **123**(3): 15-27.
- [3] RASHID M M, YU J. Nonlinear and non-Gaussian dynamic batch process monitoring using a new multiway kernel independent component analysis and multidimensional mutual information based dissimilarity approach[J]. *Industrial & Engineering Chemistry Research*, 2012, **51**(33): 10910-10920.
- [4] ZHAO C H, SUN Y X. Step-wise sequential phase partition (SSPP) algorithm based statistical modeling and online process monitoring[J]. *Chemometrics and Intelligent Laboratory Systems*, 2013, **125**(5): 109-120.
- [5] 赵春晖, 王福利, 姚远, 等. 基于时段的间歇过程统计建模、在线监测及质量预报[J]. *自动化学报*, 2010, **36**(3): 366-374.
ZHAO C H, WANG F L, YAO Y, *et al.* Phase-based statistical modeling, online monitoring and quality prediction for batch processes[J]. *Acta Automatica Sinica*, 2010, **36**(3): 366-374.
- [6] 张建明, 葛志强, 谢磊, 等. 基于支持向量数据描述的非高斯过程故障重构与诊断[J]. *化工学报*, 2009, **60**(1): 168-171.
ZHANG J M, GE Z Q, XIE L, *et al.* Non-Gaussian process monitoring and fault reconstruction and diagnosis based on SVDD[J]. *CIESC Journal*, 2009, **60**(1): 169-171.
- [7] 王培良, 葛志强, 宋执环. 基于迭代多模型 ICA-SVDD 的间歇过程故障在线监测[J]. *仪器仪表学报*, 2009, **30**(7): 1347-1352.
WANG P L, GE Z Q, SONG Z H. Online fault monitoring for batch processes based on adaptive multi-model ICA-SVDD[J]. *Chinese Journal of Scientific Instrument*, 2009, **30**(7): 1347-1352.
- [8] MACGREGOR J F, KOURTI T. Statistical process control of multivariate processes[J]. *Control Engineering Practice*, 1995, **3**(3): 403-414.
- [9] KITTIWACHANA S, FERREIRA D L S, LLOYD G R, *et al.* One class classifiers for process monitoring illustrated by the application to online HPLC of a continuous process[J]. *Journal of Chemometrics*, 2010, **24**(3/4): 96-110.
- [10] KITTIWACHANA S, FERREIRA D L S, FIDO L A, *et al.* Self-organizing map quality control index[J]. *Analytical Chemistry*, 2010, **82**(14): 5972-5982.
- [11] NING X H, TSUNG F G. A density-based statistical process control scheme for high-dimensional and mixed-type observations[J]. *IIE Transactions*, 2012, **44**(4): 301-311.
- [12] NOMIKOS P, MACGREGOR J F. Monitoring batch processes using multiway principal component analysis[J]. *AICHE Journal*, 1994, **40**(8): 1361-1375.
- [13] NOMIKOS P, MACGREGOR J F. Multi-way partial least squares in monitoring batch processes[J]. *Chemometrics & Intelligent Laboratory Systems*, 1995, **30**(1): 97-108.
- [14] CHANG K Y, LEE J M, VANROLLEGHEM P A, *et al.* On-line monitoring of batch processes using multiway independent component analysis[J]. *Chemometrics and Intelligent Laboratory Systems*, 2004, **71**(2): 151-163.
- [15] LU N Y, GAO F R, WANG F L. Sub-PCA modeling and on-line monitoring strategy for batch processes[J]. *AICHE Journal*, 2004, **50**(1): 255-259.
- [16] YAO Y, GAO F R. A survey on multistage/multiphase statistical modeling methods for batch processes[J]. *Annual Reviews in Control*, 2009, **33**(2): 172-183.
- [17] KANG J H, YU J, KIM S B. Adaptive nonparametric control chart for time-varying and multimodal processes[J]. *Journal of Process Control*, 2016, **37**: 34-45.
- [18] GANI W, TALEB H, LIMAM M. An assessment of the kernel-distance-based multivariate control chart through an industrial application[J]. *Quality and Reliability Engineering International*, 2011, **27**(4): 391-401.
- [19] GE Z Q, SONG Z H. Bagging support vector data description model for batch process monitoring[J]. *Journal of Process Control*, 2013, **23**: 1090-1096.
- [20] LEE J M, YOO C K, LEE I B. Fault detection of batch processes using multiway kernel principal component analysis[J]. *Computers & Chemical Engineering*, 2004, **28**(9): 1837-1847.
- [21] ZHANG Y W, HU Z Y. On-line batch process monitoring using hierarchical kernel partial least squares[J]. *Chemical Engineering Research and Design*, 2011, **89**(10): 2078-2084.
- [22] SUN R X, TSUNG F G. A kernel-distance-based multivariate control chart using support vector methods[J]. *International Journal of Production Research*, 2003, **41**(13): 2975-2989.
- [23] CAMCI F, CHINNAM R B, ELLIS R D. Robust kernel distance multivariate control chart using support vector principles[J]. *International Journal of Production Research*, 2008, **46**(18): 5075-5095.
- [24] NING X H, TSUNG F G. Improved design of kernel distance-based charts using support vector methods[J]. *IIE Transactions*, 2013, **45**(4): 464-476.
- [25] SUKCHOTRAT T, KIM S B, TSUNG F G. One-class classification-based control charts for multivariate process monitoring[J]. *IIE transactions*, 2009, **42**(2): 107-120.
- [26] KHEDIRI I B, WEIHS C, LIMAM M. Kernel k-means clustering based local support vector domain description fault detection of multimodal processes[J]. *Expert Systems with Applications*, 2012, **39**(2): 2166-2171.
- [27] TAX D M J, DUIN R P W. Support vector domain description[J]. *Pattern recognition letters*, 1999, **20**(11): 1191-1199.
- [28] TAX D M J, DUIN R P W. Support vector data description[J]. *Machine Learning*, 2004, **54**(1): 45-66.
- [29] SAKLA W, CHAN A, JI J, *et al.* An SVDD-based algorithm for target detection in hyperspectral imagery[J]. *Geoscience and Remote Sensing Letters, IEEE*, 2011, **8**(2): 384-388.
- [30] GE Z Q, GAO F R, SONG Z H. Batch process monitoring based on support vector data description method[J]. *Journal of Process Control*, 2011, **21**(6): 949-959.
- [31] YAO M, WANG H G, XU W L. Batch process monitoring based on functional data analysis and support vector data description[J]. *Journal of Process Control*, 2014, **24**(7): 1085-1097.