

基于 SVDD 的多时段间歇过程故障检测*

王建林, 马琳钰, 邱科鹏, 刘伟旻, 赵利强

(北京化工大学信息科学与技术学院 北京 100029)

摘要:支持向量数据描述(SVDD)不要求过程数据满足正态分布,已应用于间歇过程故障检测。现有的 SVDD 间歇过程故障检测方法采用聚类分析和模型识别划分间歇过程时段,多时段划分的准确性较低,制约了多时段间歇过程故障检测精度的提高。针对上述问题提出了一种基于 SVDD 的多时段间歇过程故障检测方法,利用 SVDD 的超球体半径值与支持向量个数的变化划分间歇过程的多时段,并建立了不同时段 SVDD 故障检测模型,使用待检测样本点的球心距与对应时段的超球体半径之差检测过程故障,实现了多时段间歇过程的时段划分与过程故障检测。发酵过程仿真实验和实际生产实验结果表明,该方法能够准确地划分间歇过程的多时段,并且能够针对不同时段进行故障检测,具有较高的检测率。

关键词: 间歇过程;支持向量数据描述;多时段;故障检测

中图分类号: TP277 TH165.3 文献标识码: A 国家标准学科分类代码: 510.80

Multi-phase batch processes fault detection based on support vector data description

Wang Jianlin, Ma Linyu, Qiu Kepeng, Liu Weimin, Zhao Liqiang

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

Abstract: Support vector data description (SVDD) has been applied to fault detection of batch processes without any restriction of process data distribution. However, the SVDD based traditional fault detection method of batch processes just simply divides the phase of batch processes using cluster analysis and model identification, and leads to rough phase division results which may compromise the accuracy of fault detection in multi-phase batch processes. To address this issue, a SVDD based fault detection method for multi-phase batch processes is presented. Firstly, the different phases are divided according to the change of hypersphere radius and support vectors of SVDD. Then, the SVDD-based fault detection models of different phases are established. Finally, fault detection is achieved by judging whether the difference between hypersphere radius of corresponding phase and the distance from the sample point to center of corresponding hypersphere exceeds the control limit. The experimental results of the fed-batch penicillin fermentation process show that the presented method can achieve phase division of multi-phase batch processes with a better accuracy, and further realize multi-phase fault detection of batch processes with a higher fault detection rate.

Keywords: batch processes; support vector data description (SVDD); multi-phase; fault detection

0 引言

间歇过程已广泛应用于生物制药、食品、半导体加工等工业生产领域。间歇过程生产中出现故障,直接影响生产过程的安全运行,导致产品质量降低,造成生产企业经济损失,因此,实现间歇过程故障检测,能够有效地保

障生产过程安全,提高产品质量^[1]。

数据驱动的多向主成分分析(principal component analysis, PCA)/偏最小二乘法(partial least squares, PLS)^[2-3]和多向核 PCA/PLS^[4-5]为代表的多元统计分析方法已用于间歇过程故障检测,取得了一定的故障检测效果,但这些方法均需假设过程变量服从正态分布,当过程数据不满足上述假设时,直接影响故障检测的精度;同

时这些方法将一次间歇操作的所有过程数据看作一个整体进行处理,建立整体间歇过程故障检测模型,忽略了间歇过程的多时段过程特性,当用于多时段间歇过程故障检测时,故障检测效果并不理想。

针对多时段间歇过程故障检测,出现了基于时段的多元统计分析的故障检测方法^[6],这类方法将多时段划分和多元统计分析故障检测建模相结合,采用聚类分析^[7-9]或模型识别^[10-12]划分间歇过程的多时段,分别建立多元统计分析的时段故障检测模型,实现多时段间歇过程故障检测;然而聚类分析的多时段划分结果严重依赖于聚类个数和聚类中心的选择和过程数据结构分析,容易陷入局部最优,且缺乏对时变过程特性的深入分析,难以确保多时段划分结果的时序性,直接影响时段划分结果的准确性,虽然模型识别的多时段划分考虑了过程特性的变化,克服了聚类分析划分时段的缺陷,但过程数据的分布特性仍然影响着多时段划分结果的准确性;同时多元统计分析的时段故障检测建模仍需假设过程变量服从正态分布且变量间线性相关,因此制约着故障检测精度的提升。

支持向量数据描述(support vector data description, SVDD)是一种高效的数据描述和处理方法,不需要假设过程数据服从正态分布及变量间线性相关,已在间歇过程故障检测中得到应用。王培良等人^[13]将独立成分分析用于过程数据降维和非高斯信息提取,建立了 SVDD 整体间歇过程故障检测模型,实现故障检测;Liu X. Q. 等人^[14]将 PCA 与神经网络方法相结合处理过程数据的非线性,构建了 SVDD 整体间歇过程故障检测模型,用于故障检测;Yao M 等人^[15]采用函数描述过程数据轨迹,对数据不等长和数据遗失进行处理,建立 SVDD 整体间歇过程故障检测模型,检测过程故障;然而这些方法均采用 SVDD 建立整体间歇过程故障检测模型,忽略了间歇过程不同时段的过程特性差异,制约了故障检测精度的提升。Khediri I. B. 等人^[16]采用核 k 均值聚类分析划分间歇过程的多时段,建立了 SVDD 时段故障检测模型,实现多时段间歇过程故障检测;Wang J 等人^[17]采用 PCA 分析和 k 均值聚类粗划分间歇过程的多时段,再利用 SVDD 识别过渡时段点进行时段二次细分并建立 SVDD 时段故障检测模型,提高了故障检测精度,但时段粗划分时仍需假设过程变量服从正态分布且变量间线性相关,影响了时段划分的准确性;Ge ZH. Q. 等人^[18]针对多时段间歇过程建立了 SVDD 时段故障检测模型,并利用 SVDD 在线识别过程数据所处时段,提高了故障检测的精度;Lv ZH. M. 等人^[19]利用 k 均值算法划分间歇过程的多时段,建立 SVDD 时段故障检测模型,检测多时段间歇过程故障;这些方法考虑了间歇过程的多时段过程特性,虽取得了较好的故障检测结果,但均采用聚类分析或模型识别

方法划分间歇过程的多时段,没有充分利用 SVDD 处理过程数据所具有的优势划分多时段,仅是利用 SVDD 建立时段故障检测模型,多时段划分结果的准确性仍然制约着故障检测精度的提升。

本文提出了一种基于 SVDD 的多时段间歇过程故障检测方法,结合过程数据的时序性和时变性,利用时间片数据样本集构建的 SVDD 超球体半径值与支持向量个数的变化划分间歇过程的多时段;建立了 SVDD 时段故障检测模型,使用待检测样本点的球心距与对应时段的超球体半径之差判定故障,同时实现了多时段间歇过程的时段划分和故障检测,并通过仿真实验和实际生产实验验证了所提方法的有效性。

1 支持向量数据描述

SVDD 是由 Tax D. M. J. 等人^[20]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体。

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。

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

式中: \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,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。式(2)中, α_i 为拉格朗日乘子, K 为核函数, δ 为高斯核函数宽度参数。

$$\begin{aligned} \max_{\alpha_i} \quad & \sum_i \alpha_i K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s. t.} \quad & 0 \leq \alpha_i \leq C, \sum_i \alpha_i = 1 \end{aligned} \quad (2)$$

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\delta^2}\right) \quad (3)$$

对应的拉格朗日乘子满足 $0 < \alpha_i < C$ 的 \mathbf{x}_i 为处于球体表面的支持向量,满足该条件的支持向量索引 i 组成整数集合 \mathbf{s} ;由式(4)计算支持向量超球体球心距,即超球体的半径 R ,其中 \mathbf{x}_o 为任一支持向量, $o \in \mathbf{s}$ 。

$$R^2 = 1 - 2 \sum_i \alpha_i K(\mathbf{x}_i \cdot \mathbf{x}_o) + \sum_{i,j} \alpha_i \alpha_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \quad (4)$$

利用式(5)计算测试样本 \mathbf{z} 到球体中心 \mathbf{a} 的距离 D ,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,如式(6)所示。

$$D = \sqrt{1 - 2 \sum_{i \in S} \alpha_i K(z \cdot x_i) + \sum_{i, j \in S} \alpha_i \alpha_j K(x_i \cdot x_j)} \quad (5)$$

$$D \leq D_{lim} = R \quad (6)$$

SVDD 仅依据目标数据集就能完成分类器的构建,对目标数据没有正态分布、相互独立或呈线性关系的要求,对复杂和稀疏数据集的描述性能优秀,常用来检测异常或离群值,特别是异常数据难以获得的情况,因其具有精确的数据描述边界,构造的故障控制限较为敏感。影响基于 SVDD 的间歇过程故障检测精度的主要因素为间歇过程的多时段特性,不同时段的数据特性差异使得采用单一故障控制限的故障检测精度较低。

2 基于 SVDD 的多时段间歇过程故障检测

2.1 基于 SVDD 的多时段间歇过程分段方法

使用 SVDD 描述不同的过程数据集将得到不同的超球体,具体体现在计算所得的支持向量与拉格朗日乘子的不同。如果训练样本足够表达过程数据源的特征,训练样本新增同一过程数据源的点作为新训练数据集时,由于之前计算所得的超球体已经足够包含新进点的特征,因此在相同参数条件下计算所得的超球体不会发生较大变化;相反地,读入特性差异较大的新数据点进行训练则会改变超球体形态。

基于 SVDD 的多时段间歇过程分段是将建立超球体模型所得的结构特征量如半径值、支持向量个数与分段相衔接,以这些特征量的变化说明不同时段间的变迁。根据超球体结构的变化来反映间歇过程的不同时段,训练计算时样本集依时序增加单个过程数据点分别进行 SVDD 建模,若读入当前数据点后,新的超球体模型相比于未读入该点之前的超球体模型发生了变化,则认为时段发生了改变,引起超球体结构变化的点属于下一时段。

基于 SVDD 的分段方法设定支持向量的个数和超球体半径,依次读入长度递增的数据样本集 Y^n 进行规划,如式(7)所示。

$$\begin{aligned} \max_{\alpha_i} \quad & \sum_{i=1}^n \alpha_i K(Y_i^n, Y_j^n) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(Y_i^n, Y_j^n) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \sum_{i=1}^n \alpha_i = 1 \end{aligned} \quad (7)$$

若读入某个样本点使得规划后超球体半径 R 小于最大球心距或使支持向量个数增加,则认为数据特性发生了变化,进入间歇过程的下一阶段。因此,定义控制变量 P 和 M , P 为支持向量个数, M 定义如式(8)所示。

$$M = \frac{R^2}{\max(D_k^2)} \quad (8)$$

式中: R 为当前 i 个样本点规划计算所得半径,由式(4)

计算可得; D_k 是样本集中每个样本点到球心的距离,定义如式(9)所示,式中 Y_k^n 表示样本子集 Y^n 中的第 k 个样本点。

$$D_k = \sqrt{1 - 2 \sum_{i=1}^n \alpha_i K(Y_k^n, Y_i^n) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(Y_i^n, Y_j^n)} \quad (9)$$

图1所示为基于 SVDD 的多时段间歇过程分段示意图。

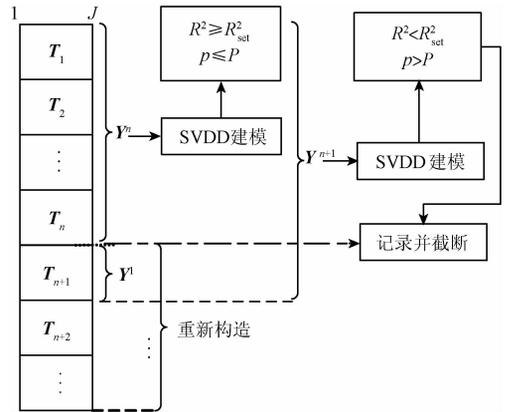


图1 基于 SVDD 的多时段间歇过程分段
Fig.1 Segmentation of multi-phase batch processes based on SVDD

每次规划超球体结束后计算该样本集对应支持向量个数 p , 如果 $p \leq P$ 且 $R^2 \geq M \cdot \max(D_k^2)$, 则读取下一个样本点扩充样本个数, 进入对扩充后新样本集的规划; 否则, 记录当前 R^2 和 D_k^2 , 截断样本集并将当前点作为上一时段的终点, 进入间歇过程下一时段, 然后对剩余样本点重复上述步骤, 直到规划遍历整个样本。一个单批次间歇过程的数据被划分为多个数据子段, 划分结果对应多个时段, 并且保证每个超球体能够包含绝大多数的样本, 这个包含比例由设定参数 M 决定。经上述处理最终将一个单批次过程数据样本划分为拥有相同支持向量个数的多个时段。

2.2 基于 SVDD 的多时段间歇过程故障检测方法

对某一批次间歇过程数据进行 SVDD 多时段间歇过程故障检测。取某一批次 $X_i \in R^{K \times J}, i = 1, 2, \dots, I$, 其中 I 为批次数, K 为样本个数, J 为变量个数; 首先按时间方向展开, $X_i = [T_1, T_2, \dots, T_K]^T \in R^{K \times J \times 1}, T_k \in R^{1 \times J}, k = 1, 2, \dots, K$ 。对单批次样本点 X_i 构造的样本集为 $Y, Y^1 = T_1, Y^2 = [T_1, T_2]^T, \dots, Y^n = [T_1, T_2, \dots, T_n]^T, \dots$, 其中 n 是不超过 K 的正整数; 对于 $Y^n, n = 1, 2, \dots, K$ 。图2所示为基于 SVDD 的多时段间歇过程分段数据样本集构建示意图。

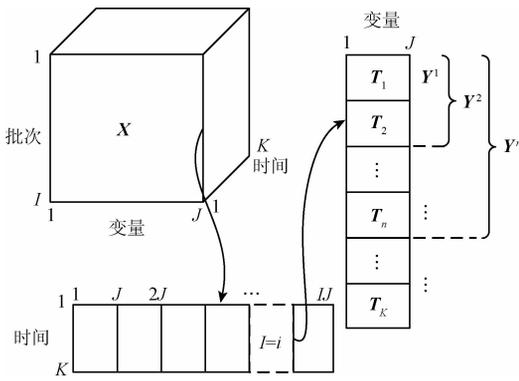


图 2 基于 SVDD 的多时段间歇过程分段数据样本集构建

Fig. 2 Construction of segmentation data set for multi-phase batch processes based on SVDD

图 3 所示为基于 SVDD 的多时段间歇过程分段流程图,首先读入前 n 个数据点 $\{T_1, \dots, T_n\}$ 构成 Y^n 进行 SVDD 超球体建模,对比当前计算所得 R^2 以及当前支持向量个数 p ,如果满足设定条件就认定超球体形态改变在可接受范围内,当前的 n 个点属于同一时段。再读入下一时刻点构成新的训练集 $Y^{n+1} = \{T_1, \dots, T_n, T_{n+1}\}$,进行新一轮的训练计算,得到超球体模型继续与分段条件对比,若不满足设定条件即认为该点为下一时段的起始点,在此处作记录并截断,进入新的循环中将 T_{n+1} 作为下一时段新的起点,重新开始构造长度递增的过程数据样本集,对样本集进行分段。

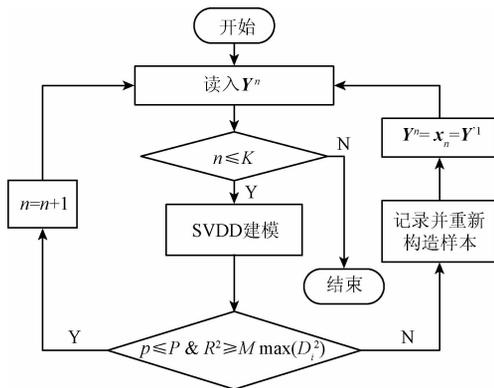


图 3 基于 SVDD 的多时段间歇过程分段流程
Fig. 3 Segmentation flow diagram of multi-phase batch processes based on SVDD

利用上述分段方法对训练批次样本进行分段,获得分段结果后,集成相应阶段的过程数据,进而分段建立模型。图 4 所示为训练样本分段结果集成与建模过程, n 个正常批次经基于 SVDD 的分段,被分别划分为 j 个子段,将每批次的第 1 时段集成到一起组成第 1 时段的训

练数据集,经 SVDD 训练建模,建立第 1 时段的模型 R_1 ,以此类推,直到对第 j 个时段建立模型 R_j ,得到每一时段各自的故障检测控制限。

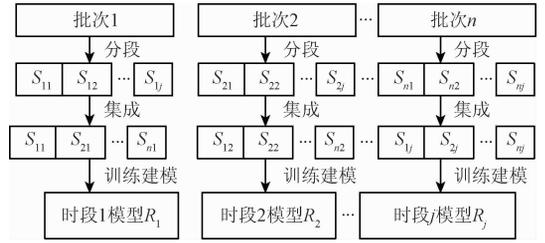


图 4 训练样本分段结果集成与建模
Fig. 4 Segmentation integration and modeling of sample sets

在同一 P, M 值设定下将待检测批次也进行分段,利用式(5)分别计算出各样本点到对应时段模型的球心距 D ,并与该点所属时段的超球体半径比较,超过半径便判为故障,故障判别如式(6)所示。

由于过程数据样本集的构造方式是依次读入后续的样本点,与是否已知一个批次的所有样本点无关,所以分段方法流程同样适用于在线过程数据样本分段,离线与在线分段只存在数据长度的差别。

实现间歇过程的在线故障检测,首先利用正常批次建立各时段的检测子模型,对待检测批次,程序每次读入最新得到的样本点,将该点与当前所有数据点作为训练数据集,带入本算法进行分段,确定该点所属子段之后,带入相应模型进行故障检测,从而判定过程故障是否存在。

基于 SVDD 的多时段间歇过程在线故障检测算法步骤如下:

- 1) 首先读入待检测批次的新进过程数据 X_i ,将该数据点与同批次的历史数据 Y^{n-1} 合并,得到长度为 n 的样本集 Y^n ;
- 2) 利用基于 SVDD 的间歇过程分段方法对该批次进行在线分段;
- 3) 分段结束后,查看该点的时段属性,找到该点所在时段所对应的检测子模型 R_j ;
- 4) 计算该新进点到检测子模型的超球体心距;
- 5) 判断该点球心距是否超过子模型球体半径,如果是,标记该点的警报值为 1,否则标记为 0,然后读入下一过程点,对下一时刻点进行在线故障检测。

3 仿真实验

利用 Pensim 仿真平台^[21]获取青霉素发酵过程的仿真数据,通过设定初始条件生成反映发酵过程批次间过

程数据差异的过程批次;选择通风率(L/h)、搅拌功率(W)、底物流加速率(L/h)、基质浓度(g/L)、溶氧量浓度(g/L)、生物质浓度(g/L)、青霉素浓度(g/L)、培养基容量(L)、二氧化碳浓度(g/L)、反应热(kcal)等10个主要过程变量,同时,为了模拟过程数据的噪声和扰动,对所有的过程变量加入高斯白噪声干扰;生成20个初始条件不同的正常批次过程数据分别作为训练样本1~20,同时产生6个故障批次,各批次的设定仿真时间均为400h,采样周期为0.5h,单个批次共800个采样点。表1所示为6个故障批次的编号及故障信息。

表1 青霉素发酵过程故障样本说明

Table 1 Fault batches description of penicillin process

故障批次编号	故障变量	故障类型	故障所在位置
1		斜坡, +5%	90~110
2		斜坡, +30%	160~180
3	通风率	斜坡, -20%	160~180
4		斜坡, +5%	320~340
5		阶跃, +30%	160~180
6		阶跃, +5%	320~340

3.1 分段参数与多时段划分

采用SVDD分段方法,设定不同的分段参数 P 和 M 对训练样本1进行分段,分段参数与分段结果如表2所示。

表2 不同分段参数的训练样本1分段结果

Table 2 Phase division results of training sample No. 1 with different parameters using SVDD

分段参数	$P=4, M=0.8$	$P=5, M=0.8$	$P=6, M=0.8$
	1~83	1~86	1~147
	84~233	87~272	148~348
分段结果	234~434	273~474	349~549
	435~635	475~676	550~751
	636~800	677~800	752~800

由表2实验结果可知,时段长度随着支持向量个数的增多而增加(最后一个时段因受批次总长度影响而减小),即支持向量个数 P 越小,分段间隔越小、越精细,支持向量个数越大,分段间隔增大。精细的分段会使时段

中数据特性更单一,能够建立更加精确的时段故障检测模型,对故障检测更有利;然而精细的分段会增加训练计算的成本,同时对故障检测结果的提升效果也存在瓶颈。本文中参数设定均为 $P=4, M=0.8$,高斯核参数为0.01,惩罚因子为0.1。训练样本1的分段结果如图5所示,可以看出本文方法能够正确识别青霉素发酵过程的时段,对时段间过渡点的定位可利用参数设置进行调整,能够对间歇过程进行合理的时段划分并进行故障检测,为各个时段提供不同阈值的故障检测控制限。

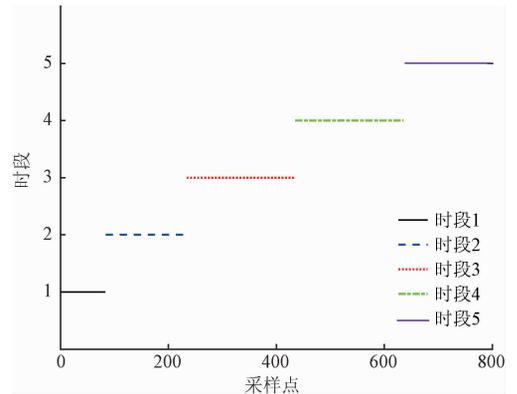


图5 SVDD分段的训练样本1分段结果

Fig. 5 Phase division results of training sample No. 1 using SVDD

3.2 方法比对

针对表1中的6个故障批次,将本文所提方法与基于 k 均值聚类分段的多向核主元分析(multiway kernel principal component analysis, MKPCA)多时段间歇过程故障检测方法、SVDD整体间歇过程故障检测方法、基于 k 均值聚类分段的SVDD时段间歇过程故障检测方法进行实验对比,验证本文所提方法的在线故障检测性能。

表3所示为上述4种方法对6个故障批次样本的在线故障检测结果。故障检测结果表明本文方法平均故障检测率为95.23%,优于其他3种方法,尤其对微小故障(故障批次1和故障批次4)的检测,明显提升了故障检测率。

表3 青霉素发酵过程故障批次样本在线检测结果

Table 3 Online monitoring results of penicillin fermentation process fault batches

(%)

故障类型	MKPCA 多时段故障检测		DIS 检测率		
	SPE	T^2	SVDD 整体故障检测	SVDD 时段故障检测	本文方法故障检测
1	4.76	9.52	47.62	66.67	85.71
2	90.48	85.71	90.47	95.24	100
3	90.48	57.14	61.90	71.43	95.24
4	76.19	0	47.62	66.67	90.48
5	90.48	85.71	100	100	100
6	71.43	0	100	100	100
平均	70.64	39.68	74.60	83.34	95.23

图 6~9 所示为上述 4 种方法针对编号 1 的青霉素发酵过程故障批次的在线故障检测结果。图 6 为基于 k 均值聚类分段的 MKPCA 多时段青霉素发酵过程故障检测方法的故障检测结果,基于 k 均值聚类分段方法将多时段青霉素发酵过程划分为 5 个时段, SPE 统计量和 T^2 统计量仅能分别检测出 2 个故障点和 1 个故障点;该方法虽然考虑了过程变量之间的非线性关系且实现了多时段故障检测,然而由于过程数据服从正态分布的假设制约了微小故障检测率的提升。图 7 所示为 SVDD 整体青霉素发酵过程故障检测方法的故障检测结果,在故障发生后的第 11 个采样点检测到故障,对微小故障检测有较大的延迟;该方法不需要假设过程数据服从正态分布及变量间线性相关,但整体间歇过程故障检测模型不能反映青霉素发酵过程的多时段特性,导致故障检测率降低。

图 8 所示为基于 k 均值聚类分段的 SVDD 时段青霉素发酵过程故障检测方法的故障检测结果,基于 k 均值聚类分段方法将多时段青霉素发酵过程划分为 5 个时段,对微小故障的检测仍然有 8 个采样点的延迟,该方法采用 k 均值聚类划分青霉素发酵过程的多时段,时段划分结果严重依赖于聚类个数和聚类中心的选择和过程数据结构分析,难以确保多时段划分结果的时序性和准确性,制约着故障检测率的提升。图 9 所示为本方法故障检测结果,SVDD 将多时段青霉素发酵过程划分为 5 个时段,分别对应 5 个故障检测模型的超球半径值,实线代表被检测批次各点的球心距,检测结果图中的报警量为 1 表示该点发生故障,为 0 表示无故障发生,故障检测结果表明,本方法在编号 1 的故障批次中仅有 2 个漏报点,有较高的故障检测率。

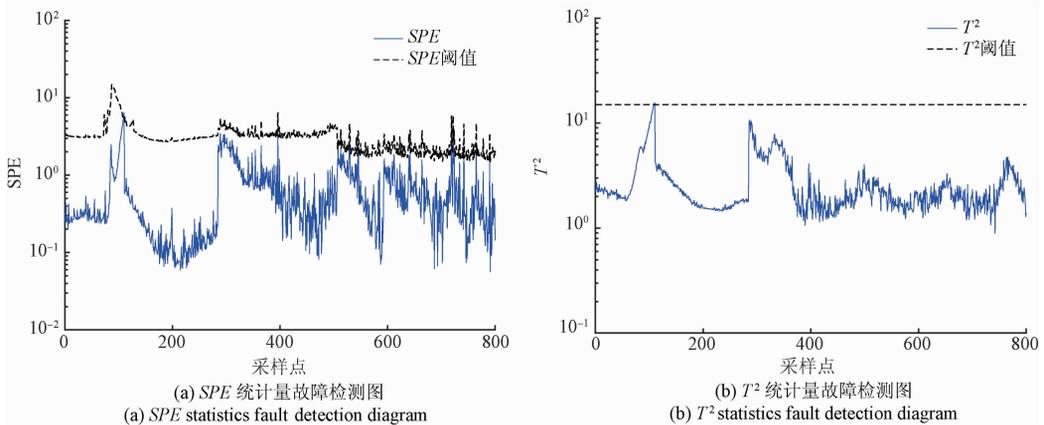


图 6 MKPCA 多时段青霉素发酵过程故障批次样本 1 检测结果

Fig. 6 Detection results of No. 1 penicillin fermentation process fault batch using phase-based MKPCA

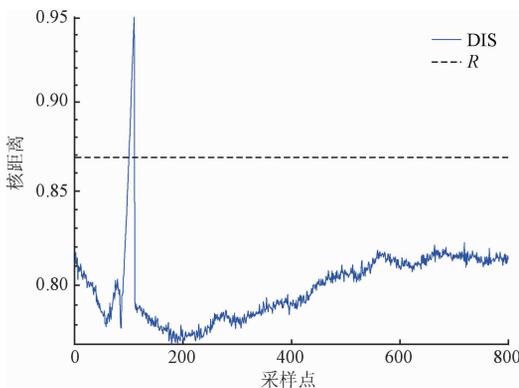


图 7 SVDD 整体青霉素发酵过程故障批次样本 1 检测结果

Fig. 7 Detection results of No. 1 penicillin fermentation process fault batch using single-phased SVDD

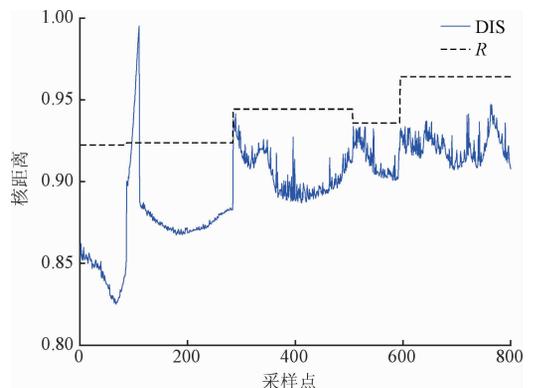


图 8 SVDD 时段青霉素发酵过程故障批次样本 1 检测结果

Fig. 8 Detection results of No. 1 penicillin fermentation process fault batch using phased-based SVDD

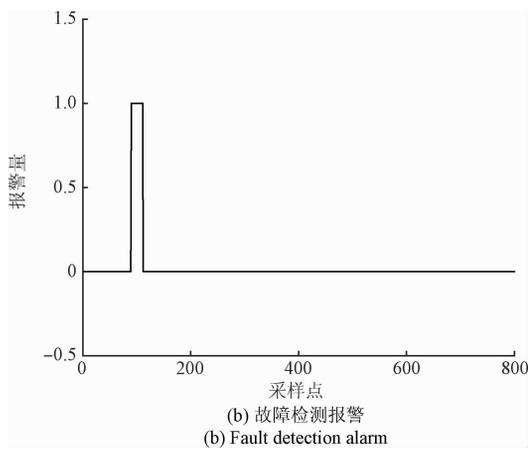
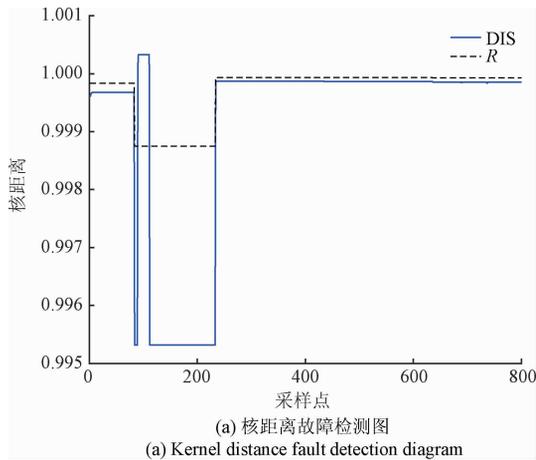


图9 本文方法青霉素发酵过程故障批次
样本1检测结果

Fig.9 Detection results of No.1 penicillin fermentation
process fault batch using the proposed method

4 发酵过程实验

金色链霉菌制备金霉素的发酵过程是一个典型非线性多时段的间歇过程。采集某生物制药厂金霉素发酵过程的20个批次现场数据作为训练样本1~训练样本20,每个批次包含温度(℃)、PH、溶氧量浓度(%)、空气流量(m^3/h)、补糖速率(L/h)、总糖浓度(%)等6个变量;每个批次的发酵时间为78 h,采样间隔为1 h。

设定分段参数 $P=4$, $M=0.999$,高斯核参数为0.01,惩罚因子为0.1,对训练样本进行分段。图10所示为金霉素发酵过程训练样本1分段结果。

故障批次为金霉素发酵过程的补糖速率和空气流量引入斜坡故障,故障批次描述如表4所示。

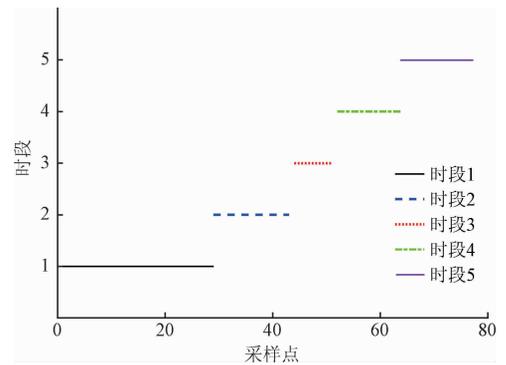


图10 SVDD分段的金霉素发酵过程训练批次
样本1分段结果

Fig.10 Phase division results of No.1 chlortetracycline
fermentation process training sample using SVDD

表4 金霉素发酵过程故障样本说明

Table 4 Fault batches description of chlortetracycline
batch process

故障批次编号	故障变量	故障类型	故障所在位置
1	补糖速率	斜坡, +20%	28~78
2	补糖速率	斜坡, +30%	28~78
3	空气流量	斜坡, +20%	5~78

表5所示为上述4种方法对3个故障批次样本的在线故障检测结果。故障检测结果表明本文方法平均故障检测率为94.31%,优于其他3种方法。

图11~14所示为上述4种方法针对编号1的金霉素发酵过程故障批次的在线故障检测结果。图11所示为基于 k 均值聚类分段的MKPCA多时段金霉素发酵过程故障检测方法的故障检测结果, SPE 统计量的漏报点多, T^2 统计量有11个采样点延迟;图12所示为SVDD整体金霉素发酵过程故障检测方法的故障检测结果,结果表明该方法几乎检测不到故障;图13所示为基于 k 均值聚类分段的SVDD时段间歇过程故障检测方法的故障检测结果,相比SVDD整体间歇过程故障检测方法和基于 k 均值聚类分段的MKPCA多时段间歇过程故障检测方法提前检测到故障,但仍有9个采样点的延迟,且出现了漏报点;图14所示为本方法的故障检测结果,金霉素发酵过程被划分为5个时段,分别对应5个故障检测模型的超球半径值,实线为被检测批次各点的球心距,报警量为1表示该点发生故障,为0表示无故障,故障检测结果表明,本方法没有漏报点,相比其他3种方法具有更高的检测率。

表 5 金霉素发酵过程故障批次样本的在线检测结果

Table 5 Online monitoring results of fault batches chlortetracycline fermentation process (%)

故障类型	MKPCA 多时段故障检测		DIS 检测率		
	SPE	T^2	SVDD 整体故障检测	SVDD 时段故障检测	本文方法故障检测
1	48.51	88.56	24.25	89.66	90.75
2	61.04	90.75	62.62	92.88	92.88
3	98.64	98.64	99.32	97.95	99.32
平均	69.40	92.65	62.06	93.50	94.31

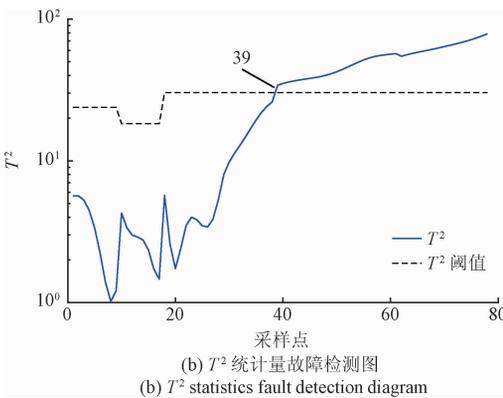
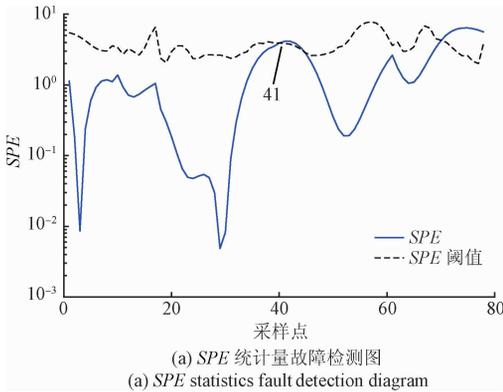


图 11 MKPCA 多时段金霉素发酵过程故障批次样本 1 检测结果

Fig. 11 Detection results of No. 1 chlortetracycline fermentation process fault batch using phase-based MKPCA

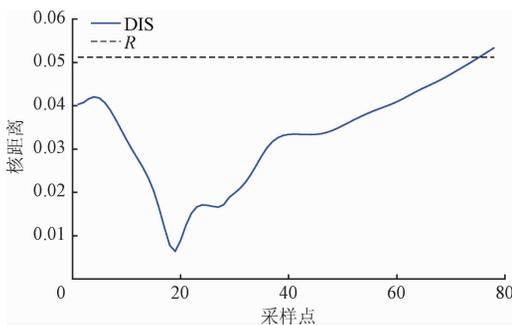


图 12 SVDD 整体金霉素发酵过程故障批次样本检测结果

Fig. 12 Detection results of chlortetracycline fermentation process fault batch using single-phased SVDD

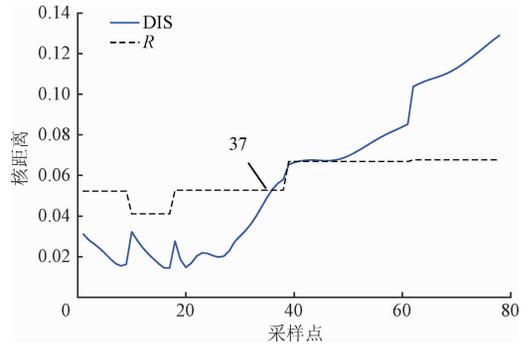


图 13 SVDD 时段金霉素发酵过程故障批次样本检测结果

Fig. 13 Detection results of chlortetracycline fermentation process fault batch using phased-based SVDD

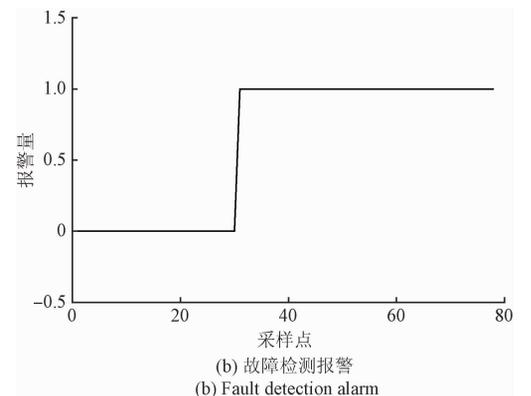
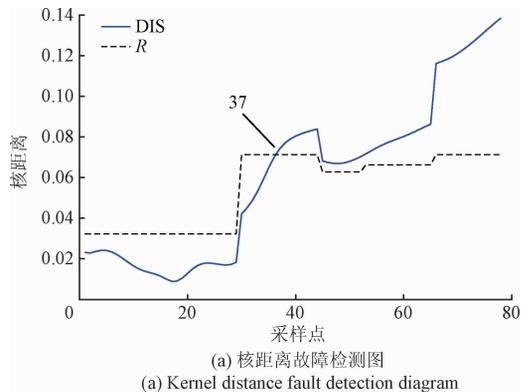


图 14 本文方法金霉素发酵过程故障批次样本 1 检测结果

Fig. 14 Detection results of No. 1 chlortetracycline fermentation process fault batch using the proposed method

5 结 论

本文提出了一种基于 SVDD 的多时段间歇过程故障检测方法,结合过程数据的时序性和时变性,利用时间片数据样本集构建的 SVDD 超球体半径值与支持向量个数的变化能够划分间歇过程的多时段;所建立的不同时段的 SVDD 故障检测模型,使用待检测样本点的球心距与对应时段的超球体半径之差判定故障,有效地实现了多时段间歇过程的故障检测;发酵过程仿真实验和实际生产实验验证表明:与现有的基于 k 均值聚类分段的 MKPCA 多时段间歇过程故障检测方法、SVDD 整体间歇过程故障检测方法、基于 k 均值聚类分段的 SVDD 时段间歇过程故障检测方法相比,本文所提方法有较高的故障检测灵敏度、延迟小,有效地提高了多时段间歇过程故障检测的准确性。

参考文献

- [1] GE Z Q, SONG Z H, GAO F R. Review of recent research on data-based process monitoring[J]. *Industrial & Engineering Chemistry Research*, 2013, 52(10): 3543-3562.
- [2] 朱荷蕾, 孟焯, 孙以泽, 等. 基于分段线性 MPCA 的充液氨过程建模及监控[J]. *仪器仪表学报*, 2015, 36(2): 271-279.
- ZHU H L, MENG ZH, SUN Y Z, et al. Modeling and monitoring of ammonia reactor filling process based on piecewise linear MPCA method[J]. *Chinese Journal of Scientific Instrument*, 36(2): 271-279.
- [3] CHEN J H, SONG C M, HSU T Y. Online monitoring of batch processes using IOHMM based MPLS [J]. *Industrial & Engineering Chemistry Research*, 2010, 49(6): 2800-2811.
- [4] 齐咏生, 王普, 高学金, 等. 改进 MKPCA 方法及其在发酵过程监控中的应用[J]. *仪器仪表学报*, 2009, 30(12): 2530-2538.
- QI Y SH, WANG P, GAO X J, et al. Application of an improved multi-way kernel principal component analysis method in fermentation process monitoring[J]. *Chinese Journal of Scientific Instrument*, 30(12): 2530-2538.
- [5] 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.
- [6] 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.
- [7] 张子羿, 胡益, 侍洪波. 一种基于聚类方法的多阶段间歇过程监控方法[J]. *化工学报*, 2013, 64(12): 4522-4528.
- ZHANG Z Y, HU Y, SHI H B. Multi-stage batch process monitoring based on a clustering method [J]. *CIESC Journal*, 2013, 64(12): 4522-4528.
- [8] LV Z M, YAN X F, JIANG Q C. Batch process monitoring based on just-in-time learning and multiple-subspace principal component analysis [J]. *Chemometrics and Intelligent Laboratory Systems*, 2014, 137(20): 128-139.
- [9] LUO L J, BAO S Y, MAO J F, et al. Phase partition and phase-based process monitoring methods for multiphase batch processes with uneven durations [J]. *Industrial & Engineering Chemistry Research*, 2016, 55(7): 2035-2048.
- [10] 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 CH 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.
- [11] ZHAO C H. Concurrent phase partition and between-mode statistical analysis for multimode and multiphase batch process monitoring [J]. *AIChE Journal*, 2014, 60(2): 559-573.
- [12] LIU J X, LIU T, ZHANG J. Window-based stepwise sequential phase partition for nonlinear batch process monitoring [J]. *Industrial & Engineering Chemistry Research*, 2016, 55(34): 9229-9243.
- [13] 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 ZH Q, SONG ZH 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.
- [14] LIU X Q, LI K, MCAFEE M, et al. Improved nonlinear PCA for process monitoring using support vector data description [J]. *Journal of Process Control*, 2011, 21(9): 1306-1317.
- [15] 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.

- [16] 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.
- [17] WANG J, WEI H T, CAO L, et al. Soft-transition sub-PCA fault monitoring of batch processes[J]. *Industrial & Engineering Chemistry Research*, 2013, 52 (29): 9879-9888.
- [18] GE ZH Q, SONG ZH H. Online monitoring and quality prediction of multiphase batch processes with uneven length problem[J]. *Industrial & Engineering Chemistry Research*, 2014, 53(2): 800-811.
- [19] LV ZH M, YAN X F. Hierarchical support vector data description for batch process monitoring[J]. *Industrial & Engineering Chemistry Research*, 2016, 55 (34): 9205-9214.
- [20] TAX D M J, DUIN R P W. Support vector data description [J]. *Machine Learning*, 2004, 54 (1): 45-66.
- [21] SUN W, MENG Y, PALAZOGLU A, et al. A method for multiphase batch process monitoring based on auto phase identification [J]. *Journal of Process Control*, 2011, 21(4): 627-638.

作者简介



王建林(通讯作者),1997年于天津大学获得博士学位,现为北京化工大学教授,主要研究方向为复杂工业过程智能检测与传感器技术、智能测控系统等。

E-mail:wangjl@ mail. buct. edu. cn

Wang Jianlin (Corresponding author)

received his Ph. D. degree from Tianjin University in 1997. Now he is a professor in Beijing University of Chemical Technology. His main research interests include intelligent measurement technology of complex industrial process, sensor technology, and intelligent measurement & control systems.



马琳钰,2014年于北京化工大学获得学士学位,现为北京化工大学硕士研究生,主要研究方向为复杂工业过程智能检测。

E-mail:2014200729@ grad. buct. edu. cn

Ma Linyu received her B. Sc. degree from Beijing University of Chemical Technology in

2014. Now she is a M. Sc. candidate in Beijing University of Chemical Technology. Her main research interests include intelligent measurement technology of complex industrial process, and sensor technology.