

# 支持向量数据描述及降幅重构方法在间歇过程故障分类中的应用研究

谢彦红<sup>1</sup>, 耿志成<sup>2</sup>, 李元<sup>2†</sup>

(1. 沈阳化工大学 数理系, 辽宁 沈阳 110142; 2. 沈阳化工大学 信息工程学院, 辽宁 沈阳 110142)

**摘要:** 针对间歇生产过程中的故障分类问题, 为进一步研究故障所属类型, 本文采用支持向量数据描述(support vector data description, SVDD)的方法. 在多种类型的故障数据库基础上, 应用SVDD建立对应故障种类的模型, 利用核函数求出各个模型超球面半径; 对于新的待分类故障样本, 先考察其与各个种类模型超球面球心的距离, 再比较此距离与半径的大小, 进而确定故障所属类型, 尤其是可能超出各个故障模型检测范围的待测故障样本, 对其进行降幅重构迭代, 确定其所属类型. 该方法不但能够准确识别独立发生的故障, 而且对于其他方法难以识别的多种并发的故障也能够有效地实现分类, 应用于数值仿真和青霉素发酵过程实验中, 验证了其有效性和准确性.

**关键词:** 间歇过程; 故障分类; 支持向量数据描述; 降幅重构; 核函数

中图分类号: TP277 文献标识码: A

## Research on the fault classification based on support vector data description and drop reconstruction in batch process

XIE Yan-hong<sup>1</sup>, GENG Zhi-cheng<sup>2</sup>, LI Yuan<sup>2†</sup>

(1. Department of Science, Shenyang University of Chemical Technology, Shenyang Liaoning 110142, China;

2. College of Information Engineering, Shenyang University of Chemical Technology, Shenyang Liaoning 110142, China)

**Abstract:** For batch process fault classification problem and further studies of failure of a type, Support Vector Data Description (SVDD) method is adopted in this paper. On the basis of the database of many types of faults, corresponding fault types of models are built applying SVDD, then super-spherical radius of each model is obtained using kernel function; for the new fault samples to be classified, examining its distance to the center of the sphere hyper-sphere various types of models, then comparing the size of this distance with radius, and to determine the type of fault, especially for the tested fault samples may be beyond the detection range in each fault model, performing iterative reconstruction decline to determine their types. The method not only can accurately identify faults occurred independently, but also can be effectively achieved for a variety of other methods are difficult to identify concurrent fault, and it is used in penicillin fermentation process and numerical simulation experiments to verify its validity and accuracy.

**Key words:** batch process; fault classification; support vector data description (SVDD); decline reconstruction; kernel function

### 1 引言(Introduction)

间歇过程是一种重要的工业生产方式, 具有生产过程本身反应复杂、产品生命周期有限等特点, 因此, 在间歇过程中, 先进的生产过程质量控制和监测技术已经变得越来越重要, 通过数据统计实现对整个生产过程的监测<sup>[1-2]</sup>, 并完成故障的识别、分类已经成为一种主要的研究课题.

目前, 故障分类方法主要有阈值法、推理法和分类器法. 阈值分类法原理比较简单<sup>[3]</sup>, 设置适当的阈值即可有效地实现故障分类, 但阈值的确定是此类方法

的难点; 推理法, 特别是模糊推理方法<sup>[4]</sup>, 由于能解决信息的不确定性与模糊性, 且能模拟人的推理过程, 近年来越来越多地应用于故障分类, 但确定模糊关系矩阵有着一定的限制; 分类器法是时下相当热门的分类方法, 其中人工神经网络(artificial neural network, ANN)具有较强的学习、泛化能力<sup>[5-6]</sup>, 神经网络应用于故障分类可提高分类速度与正确性, 但是高度的复杂性限制了其适用范围; 单值分类支持向量机(one-class support vector machine, one-class SVM)能够很好地实现单值分类问题<sup>[7-9]</sup>, 但是对于解决多分类问

收稿日期: 2014-06-11; 录用日期: 2014-11-11.

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

国家自然科学基金重点项目(61034006), 国家自然科学基金项目(61174119)资助.

Supported by State Key Program of National Natural Science Foundation of China (61034006) and National Natural Science Foundation of China (61174119).

题存在困难;支持向量聚类法(support vector clustering, SVC)是将对应的核特征空间数据映射回源数据空间<sup>[10-11]</sup>,通过核特征空间数据计算源数据的邻近矩阵,将数据划分为几部分,实现分类,具有较好的识别聚类中心和处理劣点的能力;基于聚类数据流的支持向量算法(support vector based algorithm for clustering data streams, SVStream)是将数据映射到核空间,把支持向量作为数据信息的总和<sup>[12]</sup>,构建集群边界,通过边界支持向量识别重叠聚类点以及检测劣点;基于K-means的支持向量数据描述(support vector data description based on K-means, SVDD K-means)通过K-means完成聚类分析<sup>[13]</sup>,对各个聚类簇应用SVDD算法建立对应模型;尽管SVC, SVStream, SVDD K-means 3种分类器算法有各自特点,能够以正常数据建模,实现分类或检测的目的,但是若以故障数据建模,尤其是多种并发故障,鉴于故障数据的不确定性,导致这3种方法对劣点检测的不准确性,并且由于其分类结果受权衡系数影响较大,不同的权衡系数会造成不同的分类结果,致使其对故障数据分类结果的不确定性和不准确性。

本文针对间歇生产过程已监测的故障<sup>[14]</sup>进行分类研究,运用在SVM基础之上发展的支持向量数据描述<sup>[15-17]</sup>(support vector data description, SVDD)方法与降幅重构相结合,解决故障分类问题,特别是多种故障并发的分类问题。由于SVDD方法建模具有相当高的灵敏度,对于超出建模范围的同类数据可能会产生误报的情况,因此对此类数据进行降幅重构<sup>[18-19]</sup>的处理,并且其结果受权衡系数以及劣点影响较小。将该方法应用到青霉素发酵过程<sup>[20]</sup>和数值仿真实验中,取得良好的分类效果。

## 2 基本理论(Basic theory)

### 2.1 SVDD的基本原理(The basic principle of SVDD)

支持向量数据描述(SVDD)是一种有效的数据处理方法,没有数据符合高斯分布的限制,并且能够应用到变量间是非线性关系的数据中,是目前流行的分类算法,被广泛应用于过程控制和信息检测等领域,它是由Tax和Duin提出<sup>[15-16]</sup>,通过非线性变换将正常数据样本映射到高维特征空间并建立一个模型,求取控制限,从而实现故障检测或者故障分类的目的。

若低维源故障数据空间为 $X = (x_1^T, x_2^T, \dots, x_n^T)$ ,通过非线性的变换映射到高维特征空间 $F$ ,即 $\varphi: X \rightarrow F$ ;然后在此特征空间中确定一个体积最小的超球面;为建立这样的最小体积的超球面,需要解决下面的最优化问题<sup>[7, 15-16]</sup>:

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

$$\text{s.t. } \|\varphi(x_i) - a\|^2 \leq R^2 + \xi_i, \xi_i \geq 0, i = 1, 2, \dots, n,$$

其中:  $R$ 是超球面的半径,  $a$ 是超球面的球心,  $C$ 是权

衡系数,  $\xi_i$ 是松弛因子。式(1)的最优化问题可以转化为解决相应的对偶问题,即下面的最优化问题:

$$\max_{\alpha_i} \sum_{i=1}^n \alpha_i K(x_i, x_i) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j), \quad (2)$$

$$\text{s.t. } 0 \leq \alpha_i \leq C, \sum_{i=1}^n \alpha_i = 1.$$

进一步求解该最优化问题,可以得出超球面的球心 $a$ 以及半径 $R_\alpha$ ,如下式:

$$a = \sum_{i=1}^n \alpha_i \varphi(x_i), \quad (3)$$

$$R_\alpha = \sqrt{1 - 2 \sum_{i=1}^n \alpha_i K(x_\alpha, x_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j)}, \quad (4)$$

其中:  $R_\alpha$ 是支持向量与球心的距离;  $\alpha_i$ 是每一数据样本的拉格朗日乘子;  $K(x_i, x_j)$ 是核函数且 $K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$ ,  $x_\alpha$ 是SVDD模型的支持向量,  $x_\alpha$ 的每个支持向量分别对应一个半径 $R_\alpha$ 。

本文采用核密度估计的方法,选取95%的统计控制限作为最终超球面半径 $R$ ,球心 $a$ 是不同支持向量的线性组合,只有非零的 $\alpha$ 值对应支持向量。

### 2.2 降幅重构(Decline and reconstruction)

由于SVDD模型建模采用的是故障数据库中提取的故障数据,这样的数据有一定的范围限制,不可能全部准确地涵盖所有同类的故障数据(例如:幅值大小不同);因此,在应用此模型实现分类的过程中,有可能对于超出建模范围的数据产生误报,影响分类效果。

本文采用降幅重构故障的方法避免上述的情况发生,将待分类故障数据(超出建模范围)构造在建模数据范围内。假设过程发生超出建模数据故障 $F_i (i = 1, 2, \dots, n)$ ,幅值为 $f$ ,此故障数据可以表示为 $x_{\text{fault}} = x^* + f\xi$ ,其中:  $x^*$ 为建模故障数据,  $\xi$ 为故障方向向量。当发生故障 $F_i$ 时,重构故障,  $x_{\text{fault}}$ 沿着故障方向接近建模故障数据 $x^*$ ,其到超球面球心的距离将逐渐减小至低于控制限,因此重构值可以优化式(5):

$$\arg \|\varphi(x_{\text{fault}} - f\xi) - a\|^2 \leq R_\alpha. \quad (5)$$

为满足式(5),作如下处理:

$$z = x_{\text{fault}} - \eta \Delta f \xi, \quad (6)$$

其中:  $\eta$ 为迭代次数,  $\Delta f$ 为超出建模范围故障数据与建模故障数据幅值之差。通过分析数据特征,将不同建模故障数据与正常数据幅值做差,当完成重构并能够判断 $z$ 所属类别时所进行的循环次数即为 $\eta$ 。

## 3 基于SVDD的间歇过程故障分类(Fault classification of batch process based on SVDD)

基于SVDD的间歇过程故障分类可分为两部分:模型建立和分类判别。

### 3.1 模型建立(Modeling)

假设间歇生产过程可能发生的故障状态是有限的, 分别称为  $s_1, s_2, \dots, s_N$ , 则称该间歇过程存在  $N$  个故障模式类, 将SVDD算法分别应用于  $N$  个故障类, 建立各自的SVDD模型, 从而构成  $N$  个不同的状态空间, 若间歇生产过程采集到的  $N$  种源故障数据是形如  $X_i(I \times J \times K)(i = 1, 2, \dots, N)$  的三维数据, 则

1) 数据预处理;

① 将三维数据  $X_i(I \times J \times K)$  按照  $I$  (批次) 展开, 得到二维数据  $\bar{X}_i(I \times JK)$ ;

② 将数据  $\bar{X}_i$  按照  $J$  (变量) 展开, 得到  $\bar{\bar{X}}_i(KI \times J)$ ;

2) 对变换后的  $\bar{\bar{X}}_i$  建立故障SVDD<sub>*i*</sub>模型。

### 3.2 分类判别(Classification and discrimination)

1) 对于新的待分类故障样本, 计算它与故障SVDD<sub>*i*</sub>模型球心  $a_i$  的距离, 其计算公式如下:

$$\text{Dist} = d(\varphi(\bar{x}_{\text{new}})) = \|\varphi(\bar{x}_{\text{new}} - a_i)\| = \sqrt{1 - 2 \sum_{i=1}^n \alpha_i K(\bar{x}_{\text{new}}, \bar{x}_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(\bar{x}_i, \bar{x}_j)}, \quad (7)$$

其中  $\bar{x}_{\text{new}}$  是待分类样本;

2) 将计算得到的Dist与  $R_i$  比较,  $R_i$  为各个超球面半径, 如果此数据样本到球心的距离不超过半径  $R_i$ , 则说明该数据属于此类故障; 反之, 如果此数据样本到球心的距离超过半径  $R_i$  则说明此数据属于其他类型故障, 或者说明此数据超出建模数据的检测范围, 应用式(6)对此数据进行重构, 经过多次迭代, 直至能够准确判断其属于哪一类故障为止。

基于SVDD的间歇过程建模及故障分类过程如图1所示。

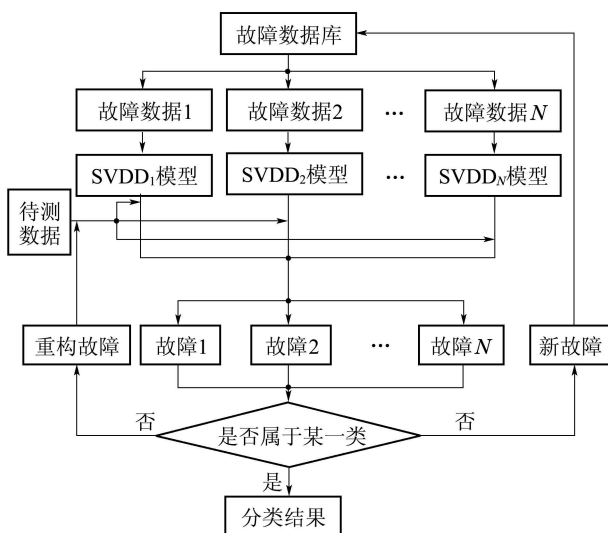


图1 基于SVDD的间歇过程故障分类流程图  
Fig. 1 Fault classification flowchart of batch process based on SVDD

## 4 故障分类的实例研究(Case study of fault classification)

### 4.1 数值仿真(Numerical simulation)

#### 4.1.1 建立故障数据库(Establishing a faulted database)

该仿真系统由5个变量组成, 其变量分别按照如式(8)定义:

$$\begin{cases} x_1 = t + e_1, \\ x_2 = t^2 + t + e_2, \\ x_3 = t^3 + 2t + 1 + e_3, \\ x_4 = t^2 + 0.2 \cos 2\pi t + e_4, \\ x_5 = t + 0.3 \cos 2\pi t + e_5, \end{cases} \quad (8)$$

式中:  $e_i(i = 1, 2, 3, 4, 5)$  是分布在  $[-0.1, 0.1]$  上的独立白噪声变量;  $t$  是定义在  $[-1, 1]$  上的变量, 采用400个数据样本点作为正常数据; 在此基础上, 在原有  $x_1$  分别增加0.1, 0.2, 0.3, 0.4, 0.5的阶跃故障, 与  $x_2, x_3, x_4, x_5$  组成变量1(故障1)的故障数据库1, 同理, 在原有  $x_2$  分别增加0.1, 0.2, 0.3, 0.4, 0.5的阶跃故障, 与  $x_1, x_3, x_4, x_5$  组成变量2(故障2)的故障数据库2, 在原有  $x_3$  分别增加0.1, 0.2, 0.3, 0.4, 0.5的阶跃故障, 与  $x_1, x_2, x_4, x_5$  组成变量3(故障3)的故障数据库3。

#### 4.1.2 数据预处理(Data preprocessing)

按照第3.1.1节所述方法建立3个故障数据库, 则每个数据库批次  $I = 5$ , 变量  $J = 3$ , 样本  $K = 400$ ; 把三维数据按照批次展开, 可以得到  $5 \times 1200$  的二维数据, 然后再按照变量进行展开, 可以得到  $2000 \times 3$  的二维数据。

#### 4.1.3 故障分类建模(Modeling fault classification)

应用SVDD方法对以上3个故障数据库分别建模。在建模以及分类过程中, 首先要确定各个参数的大小, 经过试验得出权衡系数  $C$  的大小为0.12, 本文中选取的高斯核函数如下:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\delta}\right). \quad (9)$$

本文对变量  $x_1, x_2, x_3$  各自独立加入0.25的阶跃故障作为待分类样本, 此时  $\delta = 3$ , 分类结果如图2-4所示。

图2-4中: (a) 为变量1故障模型分类结果图, (b) 为变量2故障模型分类结果图, (c) 为变量3故障模型分类结果图, 图中的短横线表示各自模型的控制限(分类线), 若样本点在其以下则说明属于此类别, 反之, 不属于此类别; 在图中可以明显看出, 3类故障模型能够准确地识别  $x_1 + 0.25, x_2 + 0.25, x_3 + 0.25$  的故障, 达到分类的目的, 由此可见, 运用SVDD方法建立的故障模型具有精确的识别、分类能力。

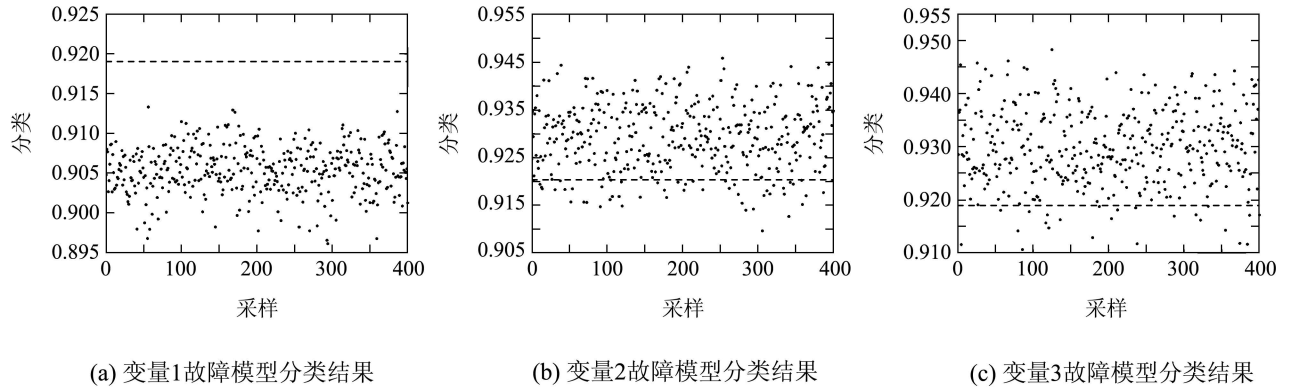


图 2 3类故障模型对 $x_1 + 0.25$ 的阶跃故障分类结果图

Fig. 2 Classification results of step fault about three types of fault model for  $x_1 + 0.25$

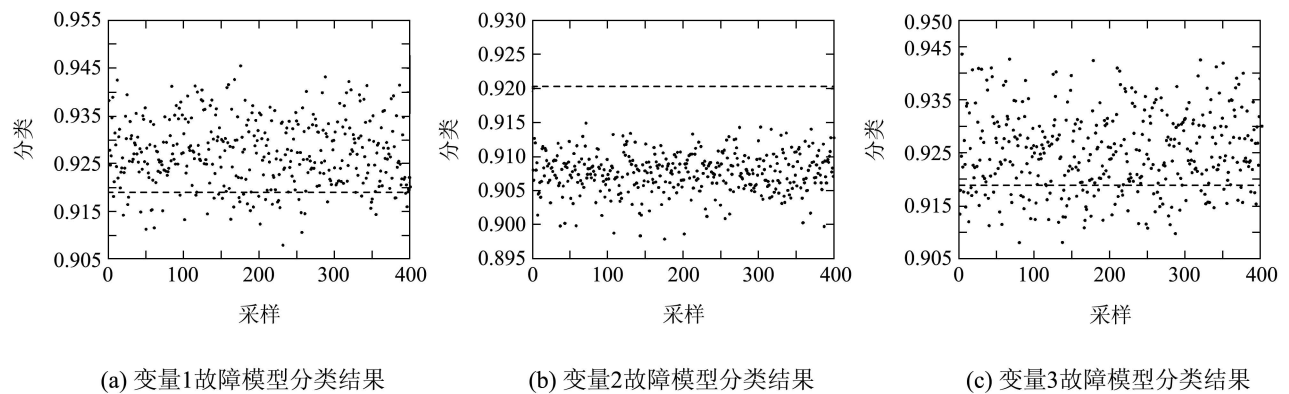


图 3 3类故障模型对 $x_2 + 0.25$ 的阶跃故障分类结果图

Fig. 3 Classification results of step fault about three types of fault model for  $x_2 + 0.25$

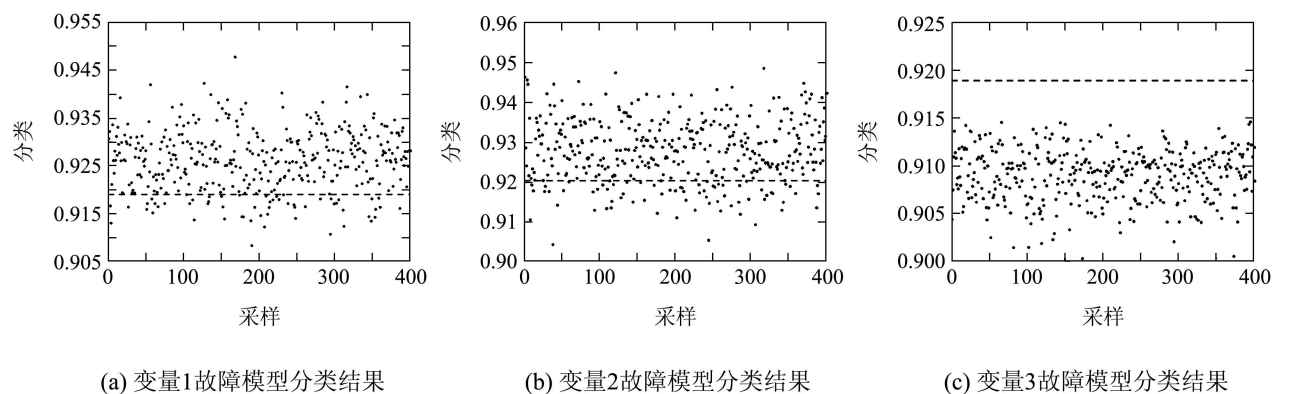


图 4 3类故障模型对 $x_3 + 0.25$ 的阶跃故障分类结果图

Fig. 4 Classification results of step fault about three types of fault model for  $x_3 + 0.25$

在实际生产过程中, 故障往往不是独立产生, 会出现多种故障并发的情况, 本文采用 $x_1 + 0.32$ ,  $x_2 + 0.27$ 两种故障并发以及 $x_1 + 0.32$ ,  $x_2 + 0.27$ ,  $x_3 + 0.25$ 三种故障并发为示例此时 $\delta = 1$ , 其分类结果如图5-6所示。

图5-6中: (a)为变量1故障模型分类结果图; (b)为变量2故障模型分类结果图; (c)为变量3故障模型结果图。在图5中, 能够明显看出对于 $x_1 + 0.32$ ,

$x_2 + 0.27$ 的两种并发故障, 变量1故障模型与变量2故障模型可以识别出其故障类型, 分别有1%和7.5%的误报率, 而变量3故障模型是无法识别这种并发故障, 说明变量3未发生故障; 在图6中, 3类故障模型都能够识别 $x_1 + 0.32$ ,  $x_2 + 0.27$ ,  $x_3 + 0.25$ 的3种并发故障, 误报率分别为1%, 7.5%, 6.5%, 具有较高的识别率, 由此说明, 本文所用方法建立的故障模型同样适用于多种故障并发的情况, 并且能够保证其较高的识别度。

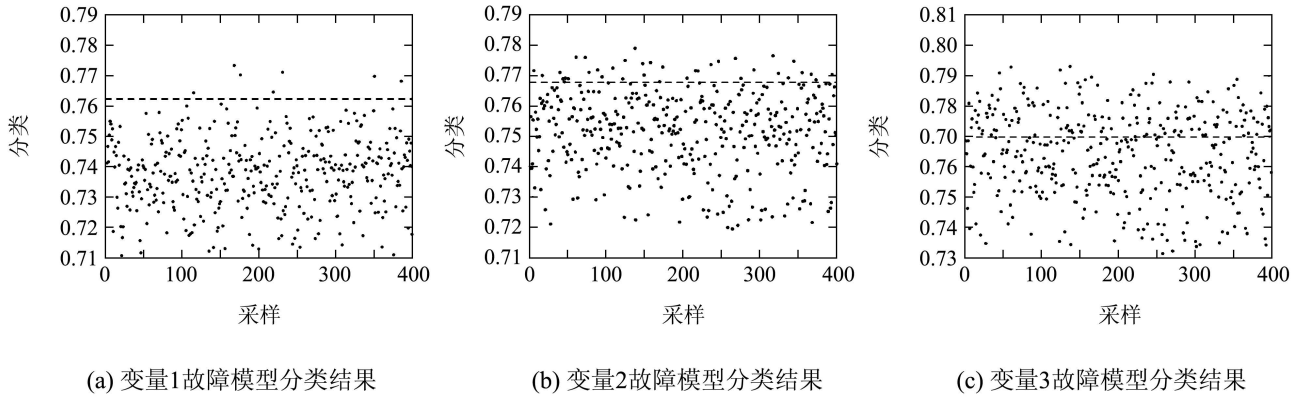


图 5 3类故障模型对  $x_1 + 0.32, x_2 + 0.27$  两种故障并发分类结果图

Fig. 5 Classification results of step fault about three types of fault model for  $x_1 + 0.32, x_2 + 0.27$

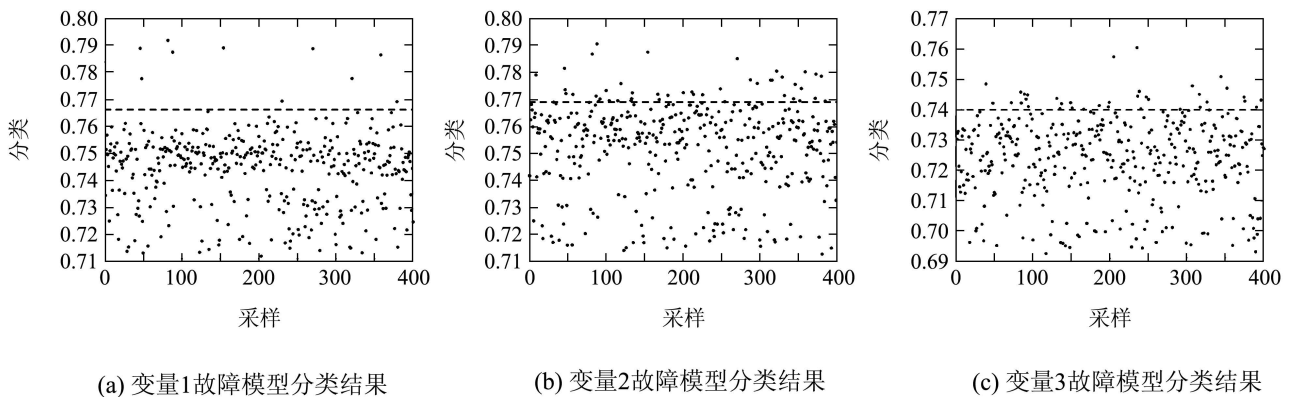


图 6 3类故障模型对  $x_1 + 0.32, x_2 + 0.27, x_3 + 0.25$  3种故障并发分类结果图

Fig. 6 Classification results of step fault about three types of fault model for  $x_1 + 0.32, x_2 + 0.27, x_3 + 0.25$

## 4.2 青霉素发酵过程故障分类的应用研究(Application of fault classification in penicillin fermentation)

### 4.2.1 数据收集(Data Collection)

本文利用青霉素仿真软件对变量1, 变量2分别产生5批次的故障训练样本作为故障数据库和2个批次的待分类故障数据, 都设置为阶跃故障, 400个采样时刻, 变量1的5个批次训练样本设置的故障分别为1%, 2%, 5%, 10%, 15%, 2个待分类批次的故障为7%, 18%, 变量2的5个批次训练样本设置的故障分别为1%, 3%, 5%, 8%, 10%, 2个待分类批次的故障为6%, 5%, 其中: 变量1的7%故障在40~120时刻发生, 变量2的6%故障在80~160时刻发生, 变量1的18%故障与变量2的15%故障是为验证前文提到的超出建模数据的情况, 在0~400时刻发生. 青霉素共有17个变量, 选取其中9个变量, 如表1所示.

### 4.2.2 数据预处理(Data preprocessing)

对产生的2个故障数据库进行预处理, 每个数据库批次  $I = 5$ , 变量  $J = 9$ , 样本  $K = 400$ ; 把三维数据按照批次展开, 可以得到  $5 \times 3600$  的二维数据, 然

后再按照变量进行展开, 可以得到  $2000 \times 9$  的二维数据.

表 1 变量名称

Table 1 Varibale names

变量号	变量名称
1	曝气量
2	搅拌功率
3	底物浓度
4	生物量浓度
5	培养基体积
6	CO <sub>2</sub> 浓度
7	PH值
8	温度
9	冷却水流量速率

### 4.2.3 分类结果与分析(Classification results and analysis)

应用SVDD方法对青霉素变量1、变量2两个故障数据库分别建模, 并进行故障分类, 其在建模范围内的数据的分类结果如图7-8所示.

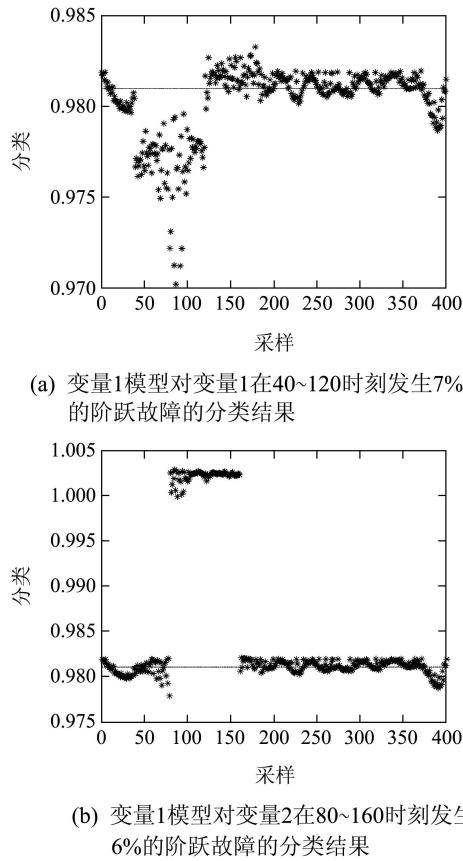


图 7 青霉素变量1模型分类结果图

Fig. 7 Classification results of model for penicillin variable 1

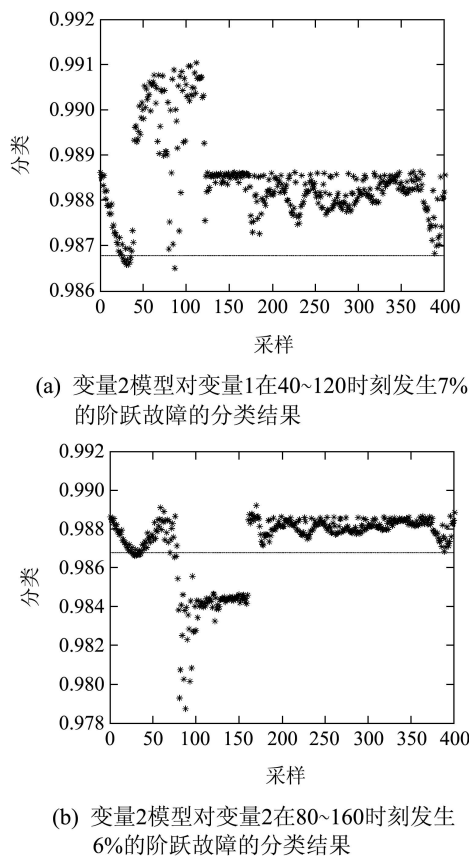


图 8 青霉素变量2模型分类结果图

Fig. 8 Classification results of model for penicillin variable 2

图7中: (a)为变量1模型对变量1在40~120时刻发生7%的阶跃故障的分类结果, (b)为变量1模型对变量2在80~160时刻发生6%的阶跃故障的分类结果, 横线为模型的控制限(分类线), 由于此模型是故障模型, 只为实现对故障的分类, 所以将正常数据划分为“故障数据”, 从图中可以看出, 青霉素变量1模型能够有效地判断出哪一变量发生故障, 并将其准确分类.

图8中: (a)为变量2模型对变量1在40~120时刻发生7%的阶跃故障的分类结果, (b)为变量2模型对变量2在80~160时刻发生6%的阶跃故障的分类结果, 从结果中能够看出, 青霉素变量2模型具有较好的分类效果, 能够准确地识别变量2所发生的故障, 由此可见, 模型分类结果比较理想, 能够完成分类的目的. 对于超出建模范围数据, 重构后的分类结果如图9-10所示.

图9-10中: (a)为模型对故障重构前的分类结果, (b)为模型对故障重构后的分类结果, 在图中可以看出, 对于超出建模范围的数据在重构之前, 对应模型无法对数据进行分类, 在重构之后, 能够有效地识别故障类型, 完成分类, 并且具有比较高的准确率.

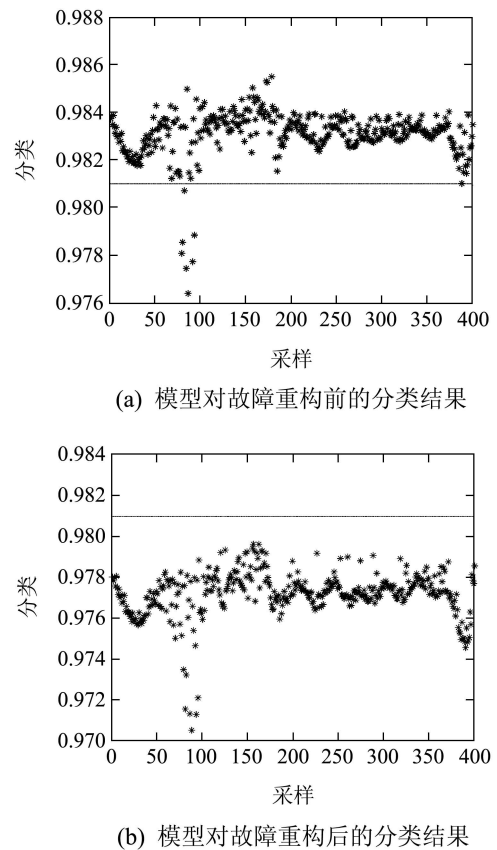


图 9 青霉素变量1模型对变量1的18%故障重构分类结果图  
Fig. 9 Classification results of fault reconstruction for 18% of penicillin variable 1 with its model

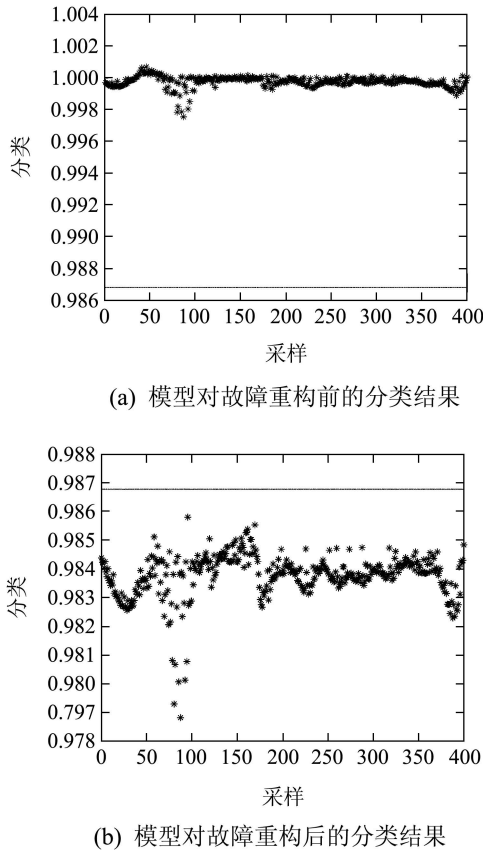


图 10 青霉素变量2模型对变量2的15%故障重构分类结果图  
Fig. 10 Classification results of fault reconstruction for 15% of penicillin variable 2 with its model

#### 4.2.4 错位重构(Malposed reconstruction)

在实际情况下,往往不会在重构的开始就能够准确地将超出建模范围的故障数据重构到对应的模型当中,即错位重构,从而会导致重构过程陷入反复循环,因此,需要设定合适的阈值作为迭代的终止条件,计算与分析过程如下:若超出建模范围的故障数据集 $X_{new} = (x_{new1}^T, x_{new2}^T, \dots, x_{newn}^T)$ ,不妨设 $x_{newn}^T$ 为错位重构故障数据,根据式(7)计算其到球心的距离为

$$Dist = \sqrt{1 - 2 \sum_{i=1}^n \alpha_i K(x_{newn}, x_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j)}$$

通过分析及计算可知式中 $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j)$ 已知,令其为 $C_1$ ,则

$$Dist = \sqrt{1 + C_1 - 2 \sum_{i=1}^n \alpha_i K(x_{newn}, x_i)}$$

当发生错位重构时,应用式(6)得出 $\sum_{i=1}^n \alpha_i K(x_{newn}, x_i)$ 逐渐趋于0,因此 $Dist = \sqrt{1 + C_1}$ 趋于常数,经过

计算与实验,本文设定的阈值 $R_\sigma = 1.0113$ .这里以青霉素变量1的错位重构为例,其结果如图11所示.

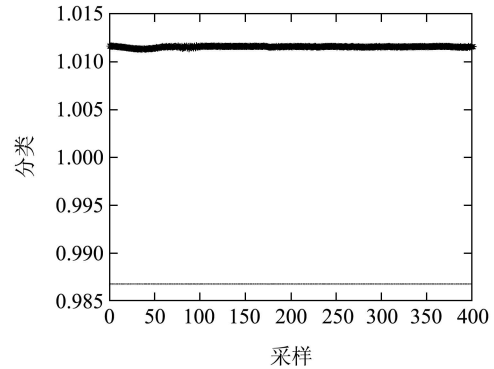


图 11 错位重构后分类结果图

Fig. 11 Classification results after malposed reconstruction

出现图11中的情况则判定为错位重构,将数据依次代入其他模型检验,直至找到对应的模型,完成分类;若故障样本按照各个模型重构之后都无法完成分类,则说明其不属于其中任何一类,需要对故障数据库进行更新.

#### 5 结论(Conclusions)

本文利用SVDD并结合故障重构的方法,建立了阶跃类型的故障模型,并应用到数值仿真和青霉素发酵过程中,实验结果表明该方法在间歇过程中具有良好的故障分类能力,不仅能够识别独立的故障类型,而且对于多种并发的故障,无需再次建立并发故障模型就能实现分类,大大减少了计算时间,同时也具有较高的准确率,这在实际的工业生产中有较强的应用价值.

#### 参考文献(References):

- [1] 周东华,李钢,李元. 数据驱动的工业过程故障检测与诊断技术 [M]. 北京: 科学出版社, 2011. (ZHOU Donghua, LI Gang, LI Yuan. *Fault Detection and Diagnosis in Industrial Process Based on Data-Driven* [M]. Beijing: Science Press, 2011.)
- [2] 赵春晖,王福利. 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 Chunhui, WANG Fuli. Phase-based statistical modeling, on-line monitoring and quality prediction for batch processes [J]. *Acta automatica sinica*, 2010, 36(3): 366 - 374.)
- [3] 段建东,张保会,周艺,等. 基于暂态量的超高压输电线路故障选相 [J]. 中国电机工程学报, 2006, 26(3): 1 - 6. (DUAN Jiandong, ZHANG Baohui, ZHOU Yi, et al. Transient-based faulty phase selection in EHV transmission lines [J]. *Proceedings of the CSEE*, 2006, 26(3): 1 - 6.)
- [4] PRADHAN A K, ROUTRAY A, PATI S, et al. Wavelet fuzzy combined approach for fault classification of a series-compensated transmission line [J]. *IEEE Transactions on Power Delivery*, 2004, 19(4): 1612 - 1618.

- [5] MAHANTY R N, DUTTA G P B. Application of RBF neural network to fault classification and location in transmission lines [J]. *IEEE Proceedings: Generation, Transmission and Distribution*, 2004, 151(2): 201 – 212.
- [6] SILVA K M, SOUZA B A, BRITO N S D. Fault detection and classification in transmission lines based on wavelet transform and ANN [J]. *IEEE Transactions on Power Delivery*, 2006, 21(4): 2058 – 2063.
- [7] 邓乃阳, 田英杰. 数据挖掘中的新方法——支持向量机 [M]. 北京: 科学出版社, 2004.  
(DENG Naiyang, TIAN Yingjie. *A new method of data mining—Support Vector Machines* [M]. Beijing: Science Press, 2004.)
- [8] CIOBANU D. Using SVM for classification [J]. *Acta Universitatis Danubius OEconomica*, 2012, 8(5): 209 – 224.
- [9] 李仁兵, 李艾华, 白向峰, 等. 支持向量机的进化多核设计 [J]. 控制理论与应用, 2011, 28(6): 793 – 798.  
(LI Renbing, LI Aihua, BAI Xiangfeng, et al. Evolutionary multiple kernels design for support vector machines [J]. *Control Theory & Applications*, 2011, 28(6): 793 – 798.)
- [10] BEN-HUR A, HORN D, SIEGELMANN H T, et al. Support vector clustering [J]. *Journal of Machine Learning Research*, 2001, 2(2): 125 – 137.
- [11] WANG C D, LAI J H, HUANG D. Incremental support vector clustering [C] // *Proceedings of the 2011 IEEE ICDM Workshop on Large Scale Visual Analytics*. Vancouver, Canada: Data Mining Workshops, 2011: 839 – 846.
- [12] WANG C D, LAI J H, HUANG D, et al. SVStream: a support vector based algorithm for clustering data streams [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2011, 25(6): 1410 – 1424. <http://dx.doi.org/10.1109/TKDE>.
- [13] 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.
- [14] 谢彦红, 耿志成, 李元. SVDD在间歇过程监测中的应用研究 [J]. 计算机与应用化学, 2013, 30(12): 1401 – 1405.  
(XIE Yanhong, GENG Zhicheng, LI Yuan. Research on the application of support vector data description method in batch process monitoring [J]. *Computers and applied chemistry*, 2013, 30(12): 1401 – 1405.)
- [15] TAX D M J, DUIN R P W. Support vector domain description [J]. *Pattern Recognition Letters*, 1999, 20(11): 1191 – 1199.
- [16] TAX D M J, DUIN R P W. Support vector data description [J]. *Machine Learning*, 2004, 54(1): 45 – 66.
- [17] 徐图, 何大可. 超球体多类支持向量机理论 [J]. 控制理论与应用, 2009, 26(11): 1293 – 1297.  
(XU Tu, HE Dake. Theory of hypersphere multiclass SVM [J]. *Control Theory & Applications*, 2009, 26(11): 1293 – 1297.)
- [18] DUNIA R, QIN S J. Subspace approach to multidimensional fault identification and reconstruction [J]. *AIChE Journal*, 1998, 44(8): 1813 – 1831.
- [19] 曹伟, 孙明. 基于角度修正迭代学习的离散时变系统故障诊断 [J]. 控制理论与应用, 2012, 29(11): 1495 – 1500.  
(CAO Wei, SUN Ming. Novel fault diagnosis for discrete time-varying system based on angle correction iterative learning [J]. *Control Theory & Applications*, 2012, 29(11): 1495 – 1500.)
- [20] 刘世成, 王海青, 李平. 青霉素生产过程的在线统计检测与产品的质量 [J]. 计算机与应用化学, 2006, 23(3): 227 – 232.  
(LIU Shicheng, WANG Haiqing, LI Ping. On-line statistical monitoring and production quality control of penicillin cultivation process [J]. *Computers and Applied Chemistry*, 2006, 23(3): 227 – 232.)

#### 作者简介:

**谢彦红** (1964–), 女, 教授, 硕士生导师, 主要研究方向为过程控制与故障诊断, E-mail: xyh666256@163.com;

**耿志成** (1988–), 男, 硕士研究生, 主要研究方向为过程控制与故障诊断, E-mail: gengzhicheng1988@163.com;

**李元** (1964–), 女, 教授, 博士生导师, 主要研究方向为过程控制与故障诊断, E-mail: li-yuan@mail.tsinghua.edu.cn.