

DOI: 10.13382/j.jemi.B2205723

改进价值函数的多阶段间歇过程故障检测*

赵小强^{1,2,3} 徐蓉蓉¹

(1. 兰州理工大学电气工程与信息工程学院 兰州 730050; 2. 甘肃省工业过程先进控制重点实验室 兰州 730050;
3. 兰州理工大学国家级电气与控制工程实验室教学中心 兰州 730050)

摘要:针对现有的阶段划分策略没有同时考虑间歇过程的动态性和多阶段特性,而导致过程检测效果不佳的问题,提出了一种改进的结合价值函数的多向核主成分分析(CVF-MKPCA)算法。首先,对间歇过程的三维数据进行相应的方向展开,并且通过构建扩展矩阵来提取间歇过程数据之间的动态特性;其次,构建一种改进的结合价值函数,评估互异时间序列信息之间的结构相似性;然后,根据动态结构相似性的评估要求,利用自下向上的搜索方法进行阶段划分,再采用MKPCA方法进行阶段建模,最后,通过构造出一种新的 *combine* 统计量对各阶段进行故障检测。所提算法在青霉素发酵仿真过程中故障误报率在控制限为95%时为3.40%,在控制限为99%时为7.98%,与对比方法相比误报率分别降低了2.12%和1.26%,证明了所提方法具有更优越的故障检测性能。

关键词:多阶段;间歇过程;结合价值函数;动态结构相似性

中图分类号: TP277; TN06 **文献标识码:** A **国家标准学科分类代码:** 510.80

Fault detection of multistage batch process based on improved value function

Zhao Xiaoqiang^{1,2,3} Xu Rongrong¹

(1. College of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou 730050, China; 3. National Experimental Teaching Center of Electrical and Control Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: Aiming at the problem that the existing stage division strategy does not consider the dynamic and multi-stage characteristics of batch processes at the same time, resulting in poor process detection effect, an improved multi-phase kernel principal component analysis based on combined value function (CVF-MKPCA) algorithm is proposed. Firstly, the three-dimensional data of the batch process are expanded in the corresponding direction, and the dynamic characteristics between the data of the batch process are extracted by constructing the expansion matrix. Secondly, an improved combined value function is constructed to evaluate the structural similarity between different time series information; then, according to the evaluation requirements of dynamic structural similarity, the bottom-up search method is used for stage division, and MKPCA method is used for stage modeling. Finally, a new *combine* statistic is constructed to detect faults in each stage. In the simulation process of penicillin fermentation, the false alarm rate of the proposed algorithm is 3.40% when the control limit is 95%, and 7.98% when the control limit is 99%, compared with the comparison method, the false alarm rate is reduced by 2.12% and 1.26% respectively, which proves that the proposed method has better fault detection performance.

Keywords: multi-stage; batch process; combined value function; dynamic structural similarity

收稿日期: 2022-08-03 Received Date: 2022-08-03

* 基金项目: 国家重点研发计划项目(2020YFB1713600)、国家自然科学基金(62263021)、甘肃省科技计划资助(21YF5GA072, 21JR7RA206)、甘肃省教育厅产业支撑计划项目(2021CYZC-02)资助

0 引言

间歇过程也被称作批次过程^[1],目前已经广泛应用于食品生产、注塑成型、生物制药、精细化工和半导体制造等诸多行业当中^[2]。间歇过程往往会涉及到易燃易爆的物料或介质,并且现代生产设备之间的相互作用和耦合程度越来越强,一旦因为故障停机,则极有可能会对人们的安全和经济财产造成严重损伤^[3]。

多元统计方法是工业现场故障检测过程中重要的一类故障检测策略^[4]。主元分析法(principal component analysis, PCA)和偏最小二乘法(partial least squares, PLS)^[5]是常用的多变量统计方法,这一类方法通过降低原始数据的维度,然后从中提取出有用的特征信息,并建立相应的统计量进行实时过程监控。郭金玉等^[6]提出了多向主成分分析(multiway principal component analysis, MPCA)的方法,高学金等^[7]提出了多向偏最小二乘(multiway partial least squares, MPLS)的方法,但是在处理间歇过程的微小故障时,上述的这些方法效率都不太高。针对间歇过程的微小故障不易在线监测的问题,李慧霞等^[8]提出了递推主成分分析(recursive principal component analysis, RPCA)的过程监测方法,该方法利用主元递推算算法在线更新主元模型。为了更进一步提高检测效率,Apsemidis 等^[9]提出了多向核主成分分析(multiway kernel principal component analysis, MKPCA)的方法,将正常的三维数据按批次展开,然后使用 KPCA 方法捕捉数据的非线性特征。Vanhatalo 等^[10]提出一种动态核主元分析(dynamic kernel principal component analysis, DKPCA)故障检测方法,运用不可区分度准则将相关程度较小的变量进行删除,然后对得到的新数据构建增广矩阵,最后通过核主元分析方法进行建模。但 these 方法都忽略了间歇过程的多阶段特性。

针对多阶段特性所带来的检测效果不佳的问题,梁小凡等^[11]针对青霉素发酵过程所固有的多时段特性,提出了一种基于模糊 C 均值算法的分时段过程监控方法,但由于该算法受聚类中心的影响比较大,聚类中心点选择的不好也会影响时段划分的效果。张成等^[12]针对间歇过程故障检测中的断续以及数据不等长等问题,提出了一种模糊有序聚类算法进行阶段划分,但由于该方法是对单独批次数据采用聚类的方法,这就会导致阶段划分的精确度不高。

多阶段是许多间歇过程的固有特征,例如半导体的加工过程包括晶圆加工、氧化、光刻、刻蚀、薄膜沉积、互连、测试和封装 8 个步骤。精细化工的生产过程可分为底物预处理、微生物发酵阶段、稳定阶段 3 个部分。这些过程由于所处的阶段不同,从而导致运行状态也存在一

定的差异。为了更好地处理间歇过程多阶段特性所带来的监测效果不好的问题,邓晓刚等^[13]提出了一种基于多阶段多向核熵成分分析的故障检测方法,通过建立一种时序核熵主元关联度的矩阵进行阶段划分。张敏等^[14]利用 K 均值算法对数据进行阶段划分,将整个过程划分为若干子阶段。但是这些方法的阶段划分效果都不是很理想。于涛等^[15]提出了一种基于滑动窗口加权 MPCA 的方法,利用了相邻过程数据所蕴含信息的相关性,降低了过程扰动、观测噪声以及多阶段间过渡过程等不确定性因素对时段划分的影响,进而改善了间歇过程的监测效果。高学金等^[16]提出了一种将时间信息添加到数据矩阵中的阶段分割策略。但这种方法只是对工业生产数据的骤变比较敏感,而对数据的动态结构相似性的衡量效果并不理想。

综上所述,为了更好地衡量动态结构相似性,从而达到精确划分阶段的目的,使故障监测效果更为显著,本文提出了一种基于结合价值函数的多向核主成分分析(multi-phase kernel principal component analysis based on combined value function, CVF-MKPCA)方法进行阶段划分,该方法首先进行三维数据展开,构建增广矩阵来提取动态关系;然后,根据结合价值函数,衡量时间序列之间的结构相似性;最后采用自底到顶的搜索算法,进行间歇过程多阶段划分。

1 基本方法

1.1 基于描述方差变化的结合价值函数

设有一个多变量时间序列 $A = \{y_k \in W^n \mid 1 \leq k \leq N\}$ 是由时间点 t_1, \dots, t_N 标记的 N 个样本的有限集合, k 是采样点数, n 为变量数, N 为样本数。阶段划分是将全体时间序列集合分为 C 个互不相交的子序列时间间隔集合 $S_A^C = \{S_c(a_c, b_c) \mid 1 \leq c \leq C\}$, 其中 $a_c = b_{c-1} + 1, b_c = N, S_c(a_c, b_c) = \{a_c \leq k \leq b_c\}$ 为子序列, a_c 代表第 C 个子序列的起始时间点, b_c 代表第 C 个子序列的终止时间点。

通常,阶段划分的目标是从给定的时间序列当中找到同类段,在这种情况下,阶段划分的问题可以定义为受约束的聚类数据点应该根据其数据相似性进行分组,为了使目标形式化,应该细分多变量时间序列内同类段的结合价值函数。

对于时间序列的划分可以利用基于 T^2 、 Q 统计量构造的成本函数,如式(1)所示:

$$\begin{cases} \text{cost}_{T^2}(S_c(a_c, b_c)) = \frac{1}{(b_c - a_c + 1)} \sum_{k=a_c}^{b_c} T_{c,k}^2 \\ \text{cost}_Q(S_c(a_c, b_c)) = \frac{1}{(b_c - a_c + 1)} \sum_{k=a_c}^{b_c} Q_{c,k} \end{cases} \quad (1)$$

进而,基于式(1)定义了评估分割效果的结合成本函数^[17],如式(2)所示:

$$\begin{cases} g_{T^2}(S_A^C) = \sum_{c=1}^C ((b_c - a_c + 1/N)) \cdot \text{cost}_{T^2}(S_C(a_c, b_c)) \\ g_Q(S_A^C) = \sum_{c=1}^C ((b_c - a_c + 1/N)) \cdot \text{cost}_Q(S_C(a_c, b_c)) \end{cases} \quad (2)$$

其中, $T_{c,k}^2$ 和 $Q_{c,k}$ 可以按照式(3)和(4)进行计算, $\hat{\mathbf{g}}_k^T$ 是对 \mathbf{g}_k^T 利用主成分分析的方法进行投影操作所得到的数据, \mathbf{U}_C^T 是对 $S_C(a_c, b_c)$ 对应的数据矩阵分解得到的负载矩阵。

$$T_{c,k}^2 = \hat{\mathbf{g}}_k^T \hat{\mathbf{g}}_k = \mathbf{g}_k^T (\mathbf{U}_C \mathbf{U}_C^T) \mathbf{g}_k \quad (3)$$

$$Q_{c,k} = (\mathbf{g}_k - \hat{\mathbf{g}}_k)^T (\mathbf{g}_k - \hat{\mathbf{g}}_k) = \hat{\mathbf{g}}_k^T (I - \mathbf{U}_C \mathbf{U}_C^T) \mathbf{g}_k \quad (4)$$

结合式(3)、(4)和(1)就可以得到式(5),结合式(3)、(4)和(2)就可以得到式(6)。

$$\begin{cases} \text{cost}_{T^2}(S_C(a_c, b_c)) = \frac{1}{(b_c - a_c + 1)} \sum_{k=a_c}^{b_c} \|\hat{x}_k\|^2 \\ \text{cost}_Q(S_C(a_c, b_c)) = \frac{1}{(b_c - a_c + 1)} \sum_{k=a_c}^{b_c} \|x_k - \hat{x}_k\|^2 \end{cases} \quad (5)$$

$$\begin{cases} g_{T^2}(S_A^C) = (1/N) \cdot \sum_{c=1}^C \sum_{k=a_c}^{b_c} \|\hat{x}_k\|^2 \\ g_Q(S_A^C) = (1/N) \cdot \sum_{c=1}^C \sum_{k=a_c}^{b_c} \|x_k - \hat{x}_k\|^2 \end{cases} \quad (6)$$

其中, $\text{cost}_{T^2}(s_c(a_c, b_c))$ 是子序列对应数据经过主元分解之后的有解释方差, $\text{cost}_Q(s_c(a_c, b_c))$ 是子序列对应数据经过主元分解之后的无解释方差,而全局成本函数 $G_{T^2}(S_A^C)$ 定义了在分段子序列中使用主元分析方法所得到的平均有解释方差,全局成本函数 $G_Q(S_A^C)$ 定义了分段子序列中使用主元分析方法所得到的平均无解释方差。序列分割目标是寻找序列分割 S_A^C 使得 $G_{T^2}(S_A^C)$ 最大,换言之,也就是指 $G_Q(S_A^C)$ 最小^[17]。

1.2 多向核主成分分析

数据变量维数高是批处理过程的典型特点之一,在数据生成的三维矩阵 $\mathbf{X}(\mathbf{I} \times \mathbf{K} \times \mathbf{J})$ 中, \mathbf{I} 代表批次变量数, \mathbf{K} 代表采样点数, \mathbf{J} 代表选取的变量数^[18]。在对数据进行建模时,三维数据是无法直接建模的,必须先将其转换为二维数据在进行建模分析。MKPCA 方法是将三维数据矩阵沿时间轴的方向分别切割为沿批次和沿变量的数据块,每一数据块向右依次水平排列,得到三维数据矩阵 $\mathbf{X}(\mathbf{I} \times \mathbf{K} \times \mathbf{J})$ 相对应的二维数据矩阵 $\mathbf{X}(\mathbf{I} \times \mathbf{JK})$ ^[19]。

MKPCA 方法是在展开三维矩阵而形成的二维矩阵上进行主成分分析,在 PCA 分解之前,数据应该数据归

一化。数据经过归一化后为 $\mathbf{X}(\mathbf{I} \times \mathbf{JK})$, 则 MKPCA 模型可以表示为:

$$\mathbf{X} = \sum_{r=1}^R t_r p_r + \sum_{i=R+1}^{JK} t_i p_i = \mathbf{T}_r \mathbf{P}_r + \mathbf{E} \quad (7)$$

式中: \mathbf{T}_r 和 \mathbf{P}_r 分别是得分矩阵和残差矩阵。

2 基于改进价值函数的间歇过程监测

2.1 基于改进价值函数的多阶段划分

1) 改进的价值函数

本文在式(1)的基础上,构造出一种改进的结合价值函数 *combinecost*, 如式(8)所示。式(1)构造的结合价值函数是结合后的新子序列 $\text{cost}(S_c(a_c, b_c))$, 而式(8)所构造的结合价值函数是合并时间子序列前后的全局价值函数的改变量。

$$\begin{aligned} \text{combinecost}(c) &= ((b_{c+1} - a_c + 1)/N) \cdot \\ &\text{cost}(S(a_c, b_{c+1})) - ((b_{c+1} - a_{c+1} + 1)/N) \cdot \\ &\text{cost}(S_{c+1}(a_{c+1}, b_{c+1})) - ((b_c - a_c + 1)/N) \cdot \\ &\text{cost}(S_c(a_c, b_c)) \end{aligned} \quad (8)$$

其中, *combinecost*(*c*) 为结合子序列 $S_c(a_c, b_c)$ 和 $S_{c+1}(a_{c+1}, b_{c+1})$ 的结合价值函数, $S(a_c, b_{c+1})$ 是由 $S_c(a_c, b_c)$ 和 $S_{c+1}(a_{c+1}, b_{c+1})$ 合并得到的新的子序列。

选取了胡磊等^[17]采用的自底向上的划分策略求取最优解。自底向上算法需要构建结合价值函数来评估相邻的子序列之间对应的样本数据彼此间的结构相似度。接下来通过 5 个模型例子说明利用改进的结合价值函数评估数据样本结构差异的合理性,结合价值函数的表达图如图 1 所示。

对于一组二维数据,经过标准化处理之后,用圆圈型标注线 1 来表示,将该组样本数据绕原点顺时针分别旋转 0° 、 20° 、 40° 、 60° 、 90° 就可以得到一组全新的数据样本,用五角星型标注线 2 来表示,这个过程即模拟了两组样本数据的分布差异随旋转角度逐渐增大而增大,当每一次旋转的角度增加时,也就是样本数据分布相差越大的情况下,相同的主元数能够提取过程变化的资料就越少。

2) 基于改进价值函数的阶段划分

改进的基于价值函数的划分策略采用自底向上的搜索算法,该算法的阶段划分步骤如下:

(1) 创建初始子序列集合 $S_A^K = \{S_c(a_c, b_c) = \{c | 1 \leq c \leq K\}\}$, 也就是把每一个时刻作为一个子序列。根据确定的主元数目 *A* 和结合价值函数,分别计算每一个子序列的价值函数 $\text{cost}(S_c(a_c, b_c))$ 和相邻的子序列之间的结合价值函数 *combinecost*(*c*);

(2) 判断目前子序列分段数目是否大于目标的阶段

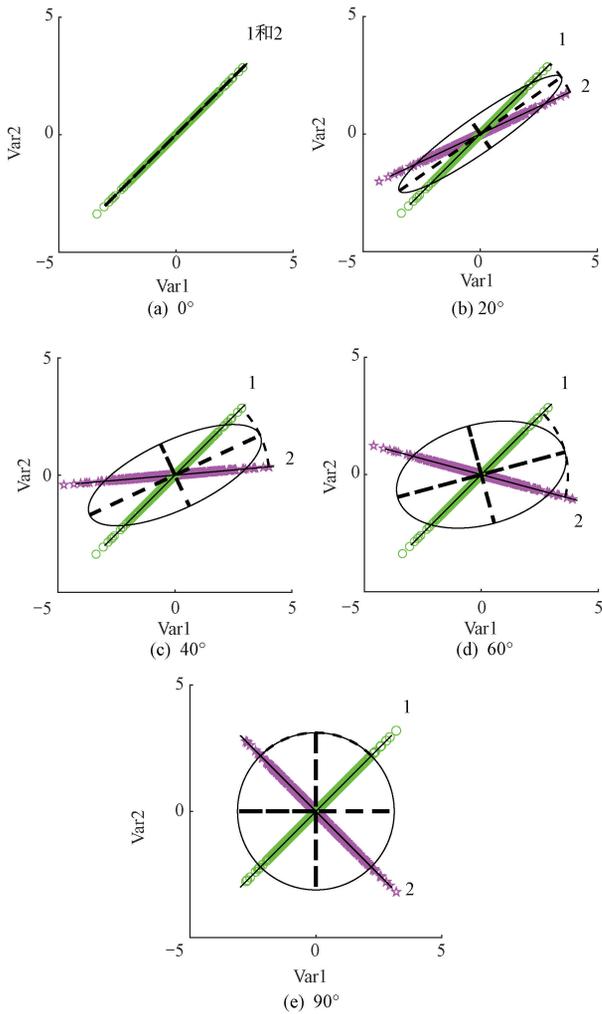


图 1 结合价值函数的表达图

Fig. 1 Expression diagram of combination value function

数目 C , 如果是, 则执行第(5)步; 否则, 执行第(3)步;

(3) 寻找结合价值函数 $combinecost(c)$ 对应的最小的两个子序列, 记为 S_x , 合并对应 S_x 和 S_{x+1} 的子序列;

(4) 删除 S_{x+1} , 更新合并后新的子序列的价值函数和相邻时间片的结合价值函数; 然后返回到第(2)步;

(5) 后续处理步骤: 由于实际的间歇过程同一阶段内物理化学反应过程通常需要保持一段时间, 对各阶段的数据长度带来物理约束, 对于不满足最小阶段长度条件的子序列, 将它分到与它相邻且合并代价最低的子序列之中。

2.2 基于 MKPCA 方法的统计建模

多向核主成分分析法实质是对主成分分析法的进一步推广, 更适用于提取数据特征。

设 $x_j \in R^m, j = 1, 2, \dots, N$ 表示一组均值为 0 的样本数据, 核函数通过特征映射 ϕ 达到点 x 到 F 的映射目的, 所获得的数据需要满足式(9)^[20]:

$$\sum_{i=1}^N \phi(x_i) = 0 \tag{9}$$

样本数据的协方差为:

$$C = \frac{1}{N} \sum_{j=1}^N \phi(x_j) \phi(x_j)^T \tag{10}$$

C 可通过特征值分解对角化为:

$$\lambda v = Cv \tag{11}$$

结合式(9)和(11)可以得到式(12):

$$Cv = \left(\frac{1}{N} \sum_{j=1}^N \phi(x_j) \phi(x_j)^T \right) v =$$

$$\frac{1}{N} \sum_{j=1}^N \langle \phi(x_j \cdot V) \rangle \phi(x_j) \tag{12}$$

所有 $\lambda \neq 0$ 都存在 $\phi(x_1), \phi(x_2), \dots, \phi(x_N)$, 可以线性表示为:

$$v = \sum_{i=1}^N \alpha_i \phi(x_i) \tag{13}$$

则有:

$$\frac{1}{N} \sum_{i=1}^N \alpha_i (\phi(x_v) \cdot \phi(x_w) \cdot \phi(x_w) \cdot \phi(x_i)) = \lambda \sum_{i=1}^N (\phi(x_v) \cdot \phi(x_i)) \tag{14}$$

其中, $v = 1, 2, \dots, N, \alpha_i$ 为常系数。

定义 $N \times N$ 维矩阵 K :

$$K_{i,v} = (\phi(x_i) \cdot \phi(x_v)) \tag{15}$$

化简式(14)可以得到式(16):

$$N\lambda K\alpha = K^2\alpha \tag{16}$$

求解式(16)得到特征值和特征向量。

数据样本 $\phi(x)$ 在特征空间 F 中 V^k 方向的投影是:

$$h_k(x) = \sum_{i=1}^N (\alpha_i)^k (\phi(x_i),) \phi(x) \tag{17}$$

将内积用核函数替换可以得到:

$$((v^k \cdot \phi(x))) = \sum_{i=1}^N (\alpha_i)^k K(x_i, x) \tag{18}$$

为了提取间歇过程变量间的动态关系, 本文采用时滞增广技术, 在时间片数据 X_k 的基础上生成时滞增广矩阵 $X_{k,d}$, 如式(19)所示。利用增广矩阵按照式(8)定义结合价值函数来计算动态结构的相似度, 用于衡量子序列动态内部结构的差异。

$$X_{k,d} = [X_{k+d} \quad X_{k+d-1} \cdots X_k] \tag{19}$$

对采集到的单个样本, 常采用式(20)计算 Hotelling's 的 T^2 统计量^[21]来描述主元空间的变化:

$$T^2 = \frac{R_c(N_c - 1)}{N_c(N_c - R_c)} F_{R_c, N_c - R_c; \alpha} \tag{20}$$

其中, $N_c = I(b_c - a_c + 1)$ 为模型建模样本个数, $F_{R_c, N_c - R_c; \alpha}$ 是自由度为 $R_c, N_c - R_c$, 显著性水平为 α 的 F 分布。

采用式(21)计算 Q 统计量来描述残差空间的变化:

$$Q = g_c \chi_{h_c, \alpha}^2 \quad (21)$$

其中, $\chi_{h_c, \alpha}^2$ 是自由度 h_c 的卡方分布的上 α 分位点, $g_c = v_c/2m_c$, m_c 是基于 Q 统计量的均值, v_c 是基于 Q 统计量的方差。

$$\begin{cases} Q \leq \delta^2 \\ T^2 \leq \chi_{1, \alpha}^2 \end{cases} \quad (22)$$

式中: δ^2 和 $\chi_{1, \alpha}^2$ 是 Q 和 T^2 的置信限。

为了更好地反映阶段划分的效果,定义了一种新的 *combine* 统计量可以同时监控 T^2 统计量和 Q 统计量对应的子空间的异常变化,就是将 T^2 统计量和 Q 统计量合并为一个综合 *combine* 统计量来监控间歇过程,采用这个 *combine* 统计量可以很快看出所采用方法进行阶段划分的效果,计算公式如式(23)所示:

$$combine = Q/\delta^2 + T^2/\chi_{1, \alpha}^2 \quad (23)$$

2.3 基于改进价值函数的多向核主成分分析方法多阶段间歇过程监测步骤

基于 CVF-MKPCA 算法的过程监测包括离线建模和在线监测两部分。

1) 离线建模

步骤(1)训练样本 $\mathbf{X}(I \times J \times K)$ 为正常工况下的若干批次数据, I 代表批次变量数, J 代表过程变量数, K 为采样点数^[22];将三维数据 $\mathbf{X}(I \times J \times K)$ 沿着批次方向展开,然后进行标准化处理得到二维数据矩阵 $\mathbf{X}(I \times JK)$,再沿变量方向展开成一系列的时间片矩阵 $\mathbf{X}(IK \times J)$,如图 2 所示。

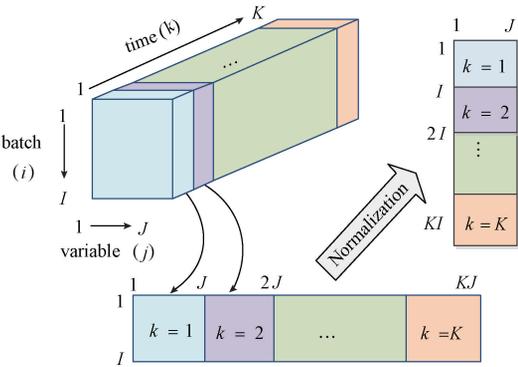


图 2 间歇过程展开图

Fig. 2 Batch process expansion diagram

步骤(2)增加时滞变量构成增广时间片矩阵 $\mathbf{X}_{k,d}$, 如果采样时间小于滞后时间,则重复构造时滞增广矩阵的过程。

步骤(3)对预处理得到的增广数据矩阵,采用自底向上算法,从最可能的时序近似值开始,合并每对相邻段的成本,最后迭代合并成本最低的时序对,直到满足一定的停止条件。

步骤(4)利用改进的结合价值函数 *combinecost* 进行数据相似度的评估。然后计算训练样本的 T^2 统计量和 Q 统计量及控制限。

2) 在线监测

步骤(1)将在线所获得的新的测试数据进行标准化,把标准化之后得到的结果记为 X_k ,然后把前 d 个时刻的数据组合成扩展数据矩阵 $\mathbf{X}_{k,d} = [X_{k+d} X_{k+d-1} \dots X_k]$;

步骤(2)采用基于改进价值函数的多向核主成分分析方法进行阶段划分之后,从划分结果得出当前时刻所处的阶段数目 C ;

步骤(3)按照阶段数目 C 对应的矩阵,得出相应的主元向量 t 和残差向量 e ;

步骤(4)按照式(20)和(21)计算出新的描述主元空间和残差空间变化的数据所对应的 Hotelling's T^2 统计量和 Q 统计量,按照式(23)计算出综合指标 *combine* 统计量。

步骤(5)比较 Hotelling's T^2 和 Q 是否大于控制线,如果两个统计量均未大于所设定的控制限,则判断无故障发生,处于正常阶段;反之,则判断有故障发生,处于故障阶段。

3 仿真实验

3.1 仿真平台介绍

青霉素是应用于临床医学的一种抗生素药品,在青霉素的发酵过程中,当一个过程变量发生变化时极有可能会引起其他变量随之发生变化。

青霉素的发酵过程中的代谢变化主要包括如下 3 个阶段:1)青霉素的菌体适应生长环境的过程;2)青霉素的产品生成过程;3)青霉素发酵细胞自溶过程。

本文采用 Pensim2.0 青霉素发酵仿真平台对所提出的算法进行验证^[22],该平台是美国 Illinois 理工学院西纳教授团队专门针对间歇过程所研发的仿真软件。青霉素的微生物发酵过程工艺图如图 3 所示。

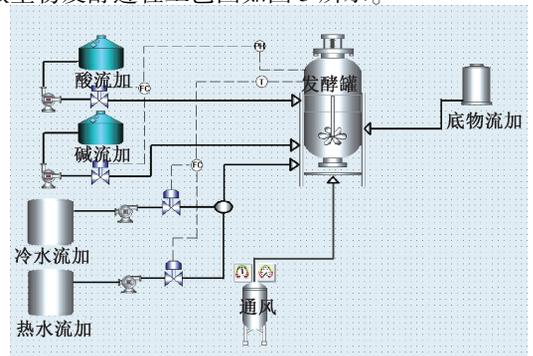


图 3 青霉素的发酵过程工艺图

Fig. 3 Process diagram of penicillin fermentation

青霉素发酵过程每个批次的反应时间为 400 h, 采样间隔为 1 h, 本文选择该过程的 10 个模拟变量进行过程监控。

监控变量表如表 1 所示^[23]。

表 1 监控变量表

Table 1 The monitoring variable table

变量序号	变量名
1	曝气速率
2	搅拌器速率
3	底物进料速率
4	底物进料温度
5	底物流浓度
6	溶解氧 (DO) 浓度
7	反应器体积
8	二氧化碳浓度
9	PH 值
10	发酵罐的温度

训练数据为 100 个批次的正常数据; 测试数据包括 100 个批次的正常运行数据, 以及 2 个批次的 2 种不同的故障数据。

故障 1 是在采样时间 100~223 h 增加 +10% 的阶跃信号作为故障信号。故障 2 是在采样时间 152~400 h 基质补料速度以 +1% 的斜率增加作为故障信号。

3.2 仿真实验验证与对比分析

1) 仿真实验验证

利用基于改进价值函数的间歇过程阶段划分方法, 确定系统阶数为 2, 阶段划分主元数为 1, 划分的阶段数为 10, 结合价值函数变化量随阶段数量的变化情况如图 4 所示。

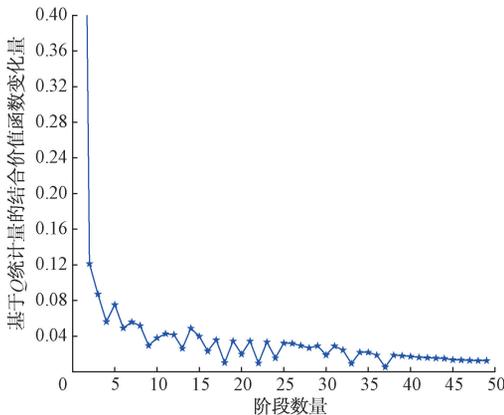


图 4 基于 Q 统计量的结合价值函数随阶段数量的变化图
Fig. 4 The change diagram of the combined value function with the number of stages based on Q statistics

CVF-MKPCA 方法阶段划分结果直观图如图 5 所示, 1~14 h 对应了菌体的反应迟滞期, 菌体在此时期对新的

环境进行适应; 15~47 h 和 48~53 h 对应菌体对数生长期, 这个时期菌体不断增加; 54~247 h 对应了青霉素的生成期, 这对应了青霉素发酵的平稳期; 而 248~400 h 对应了菌体衰亡期。

综上所述, 采用 CVF-MKPCA 方法进行阶段划分的结果与实际的微生物发酵过程有着明显的对应关系, 并且进行了更精确、更细致的划分, 表明了所提方法的优越性。

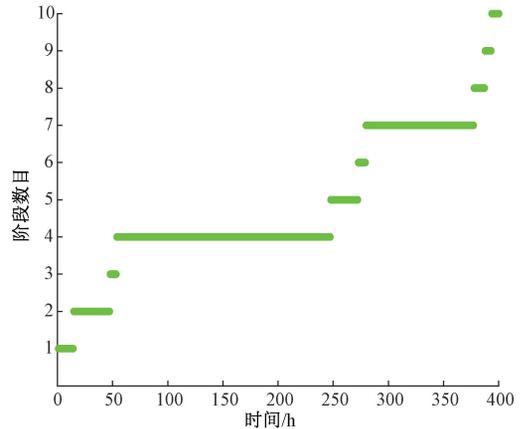


图 5 CVF-MKPCA 方法阶段划分结果直观图
Fig. 5 Visual diagram of stage division results of CVF-MKPCA method

SSPP 方法阶段划分直观图如图 6 所示, 可以看到采用 SSPP 方法进行阶段划分的效果远远不及 CVF-MKPCA 方法, SSPP 方法阶段划分不合理, 而且与实际的物理过程基本不存在对应关系。

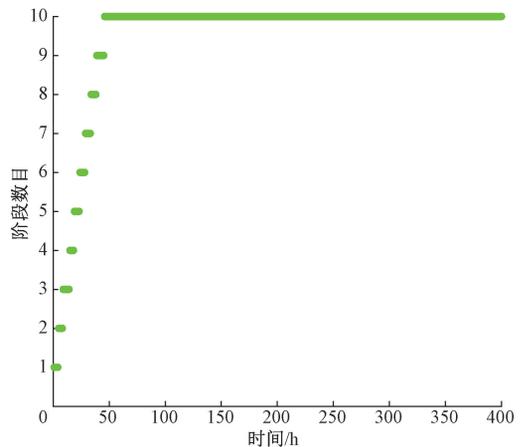


图 6 SSPP 方法阶段划分直观图
Fig. 6 Visual diagram of SSPP method phase division

MPPCA 方法阶段划分直观图如图 7 所示, 可以看到采用 MPPCA 方法进行阶段划分的结果。1~25 h 对应了菌体的反应迟滞期, 26~95 h, 96~113 h 对应菌体对数生

长期, 114~203 h 对应了青霉素的生成期, 204~234 h, 235~251 h, 252~313 h, 314~359 h, 360~400 h 对应了菌体衰亡期, 采用 MPPCA 方法进行阶段划分的结果与实际的微生物发酵过程有一定的对应关系, 但是将图 7 的 MPPCA 方法与图 5 的 CVF-MKPCA 方法进行对比, 图 5 的 CVF-MKPCA 方法进行阶段划分的结果与实际的微生物发酵过程更加对应。

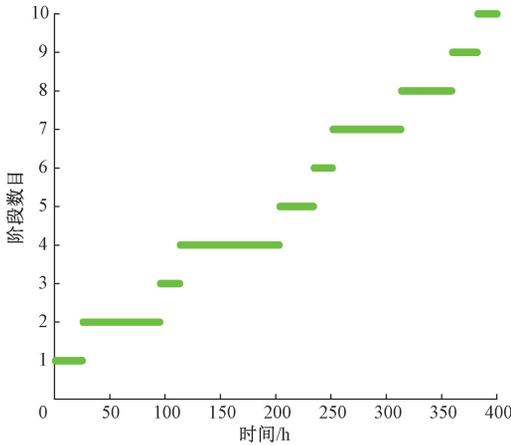


图 7 MPPCA 方法阶段划分直观图

Fig. 7 Visual diagram of MPPCA method phase division

由于青霉素的发酵仿真过程中可以引入两种不同类型的故障, 一种是阶跃信号的故障, 另一种是斜坡信号的故障。为了进一步验证 CVF-MKPCA 方法在线监控的高效性, 本文引入两种类型的故障分别检验 CVF-MKPCA 方法的效果, 其中, 图 8 是故障 1 的监测效果变化图, 故障 1 是在采样时间 100~223 h 增加 +10% 的阶跃信号作为故障信号。

从图 8 中可以清楚的看出, 在引入故障 1 之后, Q 统计量在 100 h 迅速检测到了故障, 并且在 223 h 时故障结束后 Q 统计量快速恢复到正常的控制限之下。

第 2 种故障类型是基质补料速度以 +1% 的斜率增加, 故障引入时间为 152 h, 结束时间为 400 h。故障 2 的监控效果图如图 9 所示, 在第 154 h 两个监控指标均超过控制限, 比较及时地监测到了故障的发生, 监控结果与故障实际发生情况一致。

2) 对比分析

图 10 是基于 Q 统计量的结合价值函数随阶段数目的变化曲线图。从图 10 中可以清楚的看出本文方法 (CVF-MKPCA) 的基于 Q 统计量的结合价值函数较小, SSPP 方法和 MPPCA 方法的基于 Q 统计量的结合价值函数比 CVF-MKPCA 方法大, 也就是采用 CVF-MKPCA 方法所能提取到间歇过程的动态信息更多, 从而阶段划分效果就越显著。

SSPP、MPPCA 和 CVF-MKPCA 算法的故障误报率结

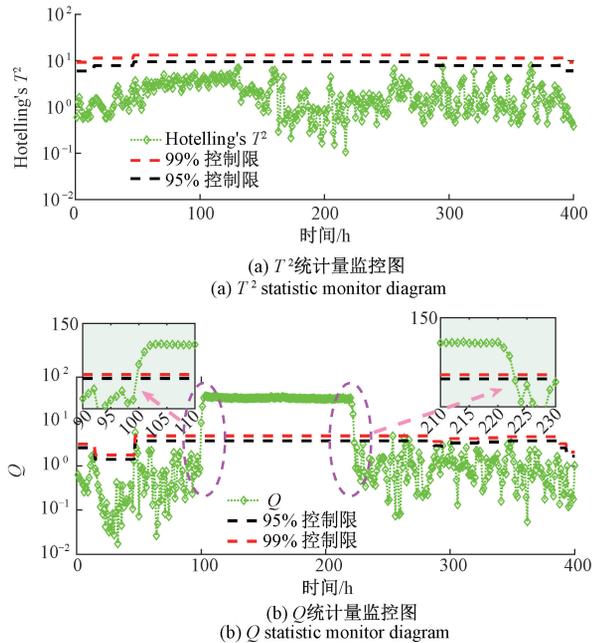


图 8 故障 1 的监测效果变化图

Fig. 8 Change diagram of fault 1 monitoring effect

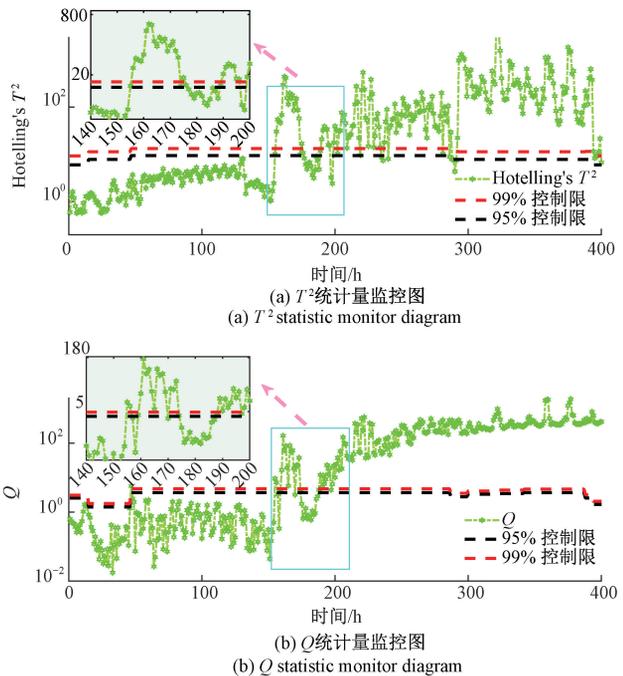


图 9 故障 2 的监测效果变化图

Fig. 9 Change diagram of fault 2 monitoring effect

果比较表如表 2 所示, 表 2 每一列分别给出 SSPP、MPPCA 和 CVF-MKPCA 方法对于不同统计指标的误报率, 可以清楚地看出, 对于 T^2 统计量, 不论显著性水平是 99% 还是 95%, 所提出的 CVF-MKPCA 方法误报率都低于对比方法 SSPP 和 MPPCA, 基于 Q 统计量, 显著性水

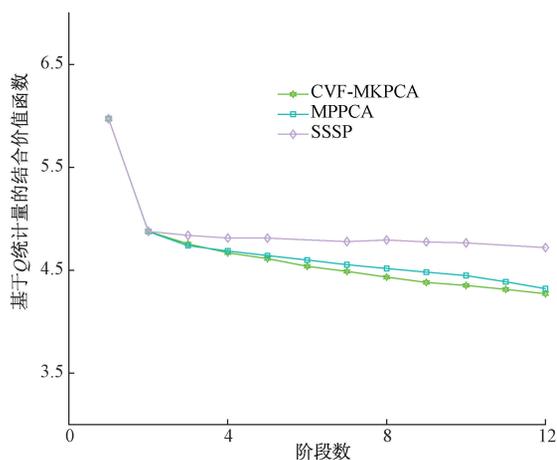


图 10 基于 Q 统计量的结合价值函数随阶段数目的变化曲线图

Fig. 10 The number of combined value functions with stages based on Q statistics the change curve diagram

平是 99% 时, CVF-MKPCA 方法虽然比 SSPP 和 MPPCA 方法稍高, 但是基于 *combine* 统计量 CVF-MKPCA 方法比 SSPP 方法和 MPPCA 方法的误报率更低。

表 2 故障误报率结果比较表

Table 2 Comparison table of fault false alarm rate results (%)

统计量	T^2		Q		<i>combine</i>	
	95	99	95	99	95	99
SSPP	3.22	5.96	3.49	5.41	5.52	9.24
MPPCA	2.35	5.07	3.89	5.74	5.28	8.87
CVF-MKPCA	0	3.75	3.40	6.10	3.40	7.98

改进的结合价值函数的多向核主成分分析方法利用了改进的结合价值函数进行间歇过程阶段划分, 更好的评估了数据间的动态结构相似性, 从而提高了故障检测率。表 3 是所提方法与其他方法的故障检测率结果比较表。从表 3 中可以清楚地看出, 基于 T^2 统计量和 Q 统计量, 不论显著性水平是 99% 还是 95%, 所提出的 CVF-MKPCA 方法故障检测率都高于对比方法 SSPP 和 MPPCA。

表 3 故障检测率结果比较表

Table 3 Fault detection rate results comparison table (%)

统计量	T^2		Q	
	95	99	95	99
SSPP	95.78	93.17	95.94	93.34
MPPCA	96.73	93.93	96.11	93.50
CVF-MKPCA	98.25	96.18	96.15	93.81

4 结论

由于间歇过程数据的动态性和多阶段特性同时存在所带来的过程检测准确度不高的问题, 本文提出了一种改进结合价值函数的多向核主成分分析故障检测方法, 先利用改进后的结合价值函数衡量数据结构相似性, 然后再采用自底向上的搜索算法进行阶段划分, 最后再采用 MKPCA 方法进行阶段建模。所提出的改进的结合价值函数的多阶段故障检测方法在青霉素发酵仿真平台上验证了其可靠性。

参考文献

[1] 赵春晖, 余万科, 高福荣. 非平稳间歇过程数据解析与状态监控-回顾与展望[J]. 自动化学报, 2020, 46(10): 2072-2091.
ZHAO CH H, YU W K, GAO F R. Non-stationary batch process data analysis and state monitoring-review and outlook[J]. Journal of Automation, 2020, 46(10): 2072-2091.

[2] 池荣虎, 侯忠生, 黄彪. 间歇过程最优迭代学习控制的发展: 从基于模型到数据驱动[J]. 自动化学报, 2017, 43(6): 917-932.
CHI R H, HOU ZH SH, HUANG B. The development of optimal iterative learning control for batch processes: From model-based to data-driven[J]. Journal of Automation, 2017, 43(6): 917-932.

[3] 高学金, 何紫鹤, 高慧慧, 等. 基于联合典型变量矩阵的多阶段发酵过程质量相关故障监测[J]. 化工学报, 2022, 73(3): 1300-1314.
GAO X J, HE Z H, GAO H H, et al. The quality-related fault monitoring of multi-stage fermentation process based on joint typical variable matrix[J]. Journal of Chemical Engineering, 2022, 73(3): 1300-1314.

[4] 常鹏, 王普, 高学金. 基于统计量模式分析的 T-KPLS 间歇过程故障监控[J]. 化工学报, 2018, 66(1): 265-271.
CHANG P, WANG P, GAO X J. Fault monitoring of T-KPLS batch process based on statistical model analysis[J]. Journal of Chemical Engineering, 2018, 66(1): 265-271.

[5] 谢彦红, 刘文静, 李元. 基于 NNDSVDD 的多模态工业过程故障检测[J]. 信息与控制, 2018, 47(5): 541-546.
XIE Y H, LIU W J, LI Y. Multimodal industrial process fault detection based on NNDSVDD[J]. Information and Control, 2018, 47(5): 541-546.

[6] 郭金玉, 齐蕾蕾, 李元. 基于 DMOLPP 的间歇过程在

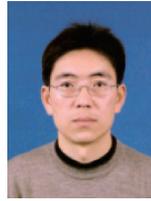
- 线故障检测[J]. 仪器仪表学报, 2015, 36(1): 134-142.
- GUO J Y, QI L L, LI Y. Online fault detection of batch process based on DMOLPP [J]. Chinese Journal of Scientific Instrument, 2015, 36(1): 134-142.
- [7] 高学金, 崔久莉, 齐咏生. 基于改进 MPLS 的间歇过程监测与故障诊断[J]. 计算机与应用化学, 2013, 30(8): 953-958.
- GAO X J, CUI J L, QI Y SH. Batch process monitoring and fault diagnosis based on improved MPLS [J]. Computer and Applied Chemistry, 2013, 30(8): 953-958.
- [8] 李慧霞, 林海军, 邵耿荣. 基于递推主元分析法的汽车衡称重传感器零点故障检测方法[J]. 电子测量与仪器学报, 2020, 34(1): 32-42.
- LI H X, LIN H J, SHAO G R. Zero point fault detection method of truck scale weighing sensor based on recursive principal component analysis [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(1): 32-42.
- [9] APSEMEDIS A, PSARAKIS S, MOGUERZA J. A review of machine learning kernel methods in statistical process monitoring [J]. Computers & Industrial Engineering, 2020, 142: 106376.
- [10] VANHATALO E, KULAHCI M, BERGQUIST B. On the structure of dynamic principal component analysis used in statistical process monitoring [J]. Chemometrics and Intelligent Laboratory Systems, 2017; 167: 1-11.
- [11] 梁小凡, 熊伟丽, 徐保国. 基于 FCM 的青霉素发酵分段统计建模及监控[J]. 计算机与应用化学, 2013, 30(12): 1427-1430.
- LIANG X F, XIONG W L, XU B G. Statistical modeling and monitoring of penicillin fermentation in different periods based on FCM [J]. Computer and Applied Chemistry, 2013, 30(12): 1427-1430.
- [12] 张成, 赵海涛, 孙韶媛. 基于模糊有序聚类算法的间歇过程故障检测[J]. 东华大学学报(自然科学版), 2018, 44(6): 940-947.
- ZHANG CH, ZHAO H T, SUN SH Y. Batch process fault detection based on fuzzy ordered clustering algorithm [J]. Journal of East China University (Natural Science Edition), 2018, 44(6): 940-947.
- [13] 邓晓刚, 张琛琛, 王磊. 基于多阶段多向核熵成分分析的间歇过程故障检测方法[J]. 化工学报, 2017, 68(5): 1961-1968.
- DENG X G, ZHANG CH CH, WANG L. Batch process fault detection method based on multi-stage and multi-directional kernel entropy component analysis [J]. Journal of Chemical Engineering, 2017, 68(5): 1961-1968.
- [14] 张敏, 王瑞琦. 基于 KECA-IGWO-KELM 的间歇过程故障诊断方法[J]. 信息与控制, 2021, 50(3): 321-328.
- ZHANG M, WANG R Q. Batch process fault diagnosis method based on KECA-IGWO-KELM [J]. Information and Control, 2021, 50(3): 321-328.
- [15] 于涛, 李和平, 王建林, 等. 基于滑动时间窗口加权 MPCA 的间歇过程监测方法[J]. 北京化工大学学报(自然科学版), 2015, 42(4): 112-119.
- YU T, LI H P, WANG J L, et al. Batch process monitoring method based on sliding time window weighted MPCA [J]. Journal of Beijing University of Chemical Technology (Natural Science Edition), 2015, 42(4): 112-119.
- [16] 高学金, 杨彦霞, 王普, 等. 基于扩展核熵负载矩阵的发酵过程故障监测[J]. 控制理论与应用, 2018, 35(6): 813-821.
- GAO X J, YANG Y X, WANG P, et al. Fault monitoring of fermentation process based on extended kernel entropy load matrix [J]. Control Theory and Application, 2018, 35(6): 813-821.
- [17] 胡磊, 刘强, 吴永建. 间歇过程动态潜结构阶段划分与在线监控[J]. 控制理论与应用, 2022, 39(2): 307-316.
- HU L, LIU Q, WU Y J. Dynamic latent structure stage division and online monitoring of batch processes [J]. Control Theory and Application, 2022, 39(2): 307-316.
- [18] 梁秀霞, 郑向博, 郑晓慧. 基于邻域保持嵌入算法的间歇过程故障检测[J]. 自动化与仪表, 2015, 30(10): 1-4.
- LIANG X X, ZHENG X B, ZHENG X H. Batch process fault detection based on neighborhood preserving embeddings algorithm [J]. Automation and Instrumentation, 2015, 30(10): 1-4.
- [19] 邵盟雅, 吕锋, 宋学君. 基于 K-means 最佳聚类的间歇过程故障诊断方法[J]. 控制工程, 2020, 27(9): 1642-1648.
- SHAO M Y, LYU F, SONG X J. A batch process fault diagnosis method based on K-means optimal clustering [J]. Control Engineering, 2020, 27(9): 1642-1648.
- [20] 朱路飞. 基于 GM-LSSVM-MKPCA 的 RBF 神经网络短期电力负荷预测方法研究 [D]. 长春: 吉林大学, 2020.
- ZHU L F. The short-term power load forecasting method of RBF neural network based on GM-LSSVM-MKPCA [D].

Changchun;Jilin University, 2020.

- [21] 赵小强, 牟森. 基于变量分块的 KDLV-DWSVDD 间歇过程故障检测算法研究[J]. 仪器仪表学报, 2021, 42(2): 244-256.
ZHAO X Q, MOU M. Research on KDLV-DWSVDD batch process fault detection algorithm based on variable block [J]. Chinese Journal of Scientific Instrument, 2021, 42(2): 244-256.
- [22] 王建林, 马琳钰, 邱科鹏. 基于 SVDD 的多时段间歇过程故障检测[J]. 仪器仪表学报, 2017, 38(11): 2752-2761.
WNAG J L, MA L Y, QIU K P. Fault detection of multi-period intermittent process based on SVDD[J]. Chinese Journal of Scientific Instrument, 2017, 38(11): 2752-2761.
- [23] 徐逸丰, 杨海麟, 孟繁松, 等. 基于 MW-KECA 与变量贡献 SVDD 的间歇过程故障检测系统[J]. 控制工程, 2022, 29(1): 143-151.
XU Y F, YANG H L, MENG F S, et al. Fault detection system for batch processes based on MW-KECA and variable contribution SVDD [J]. Control Engineering,

2022, 29(1): 143-151.

作者简介



赵小强 (通信作者), 2006 年于浙江大学获得博士学位, 现为兰州理工大学教授、博士生导师, 主要研究方向为故障诊断、图像处理、数据挖掘等。

E-mail: xqzhao@lut.edu.cn

Zhao Xiaoqiang (Corresponding author) received the Ph. D. degree from Zhejiang University in 2006. Now he is a professor and Ph. D. supervisor at Lanzhou University of Technology. His research interests include fault diagnosis, image processing and data mining etc.



徐蓉蓉, 2021 年于兰州工业学院获得学士学位, 现为兰州理工大学研究生, 主要研究方向为间歇过程故障诊断。

E-mail: 2313875156@qq.com

Xu Rongrong received the B. S. degree in 2021 from Lanzhou Institute of Technology. Now she is a M. Sc. candidate at Lanzhou University of Technology. Her main research interest includes batch process fault diagnosis.