

引用格式:梁秀霞,陈娇娇,严婷,等. 基于自适应动态时间规整(DTW)的GA-FCM多阶段间歇过程故障诊断[J]. 北京化工大学学报(自然科学版),2019,46(5):87-93.

LIANG XiuXia, CHEN JiaoJiao, YAN Ting, et al. Fault diagnosis of GA-FCM multi-stage batch processes based on adaptive dynamic time warping algorithm[J]. Journal of Beijing University of Chemical Technology (Natural Science), 2019,46(5):87-93.

基于自适应动态时间规整(DTW)的GA-FCM多阶段间歇过程故障诊断

梁秀霞 陈娇娇 严婷 周颖 张燕

(河北工业大学 人工智能与数据科学学院, 天津 300130)

摘要:多时段是间歇过程的固有特征,对间歇过程划分阶段可以提高故障诊断的精度。采用模糊C-均值聚类(FCM)算法划分阶段存在对初始聚类中心敏感、易于陷入局部极优值的问题。提出遗传算法与FCM算法相结合的方法(GA-FCM),用于克服FCM易于陷入局部极优值的问题,以达到全局最优。同时,针对间歇过程数据不等长问题,提出自适应动态时间规整(DTW)算法。随后,用GA-FCM方法完成阶段划分,再建立多向核主元分析(MKPCA)模型完成故障检测。最后将此算法应用于青霉素发酵过程,仿真结果验证了所提方法的可行性和有效性。

关键词:模糊C-均值聚类算法;遗传算法;动态时间规整;多向核主元分析;故障检测;间歇过程

中图分类号: TP277 **DOI:** 10.13543/j.bhxbzr.2019.05.013

引言

间歇生产是重要的工业生产过程,由于具有体积小、价值高的优点,广泛应用于发酵工程、半导体制造等批量生产中^[1-3]。近年来,间歇过程在工业化、集成化生产中所占的比例逐年升高,因此其安全性与可靠性也备受关注。

间歇过程的生产特点使得生产过程中会出现数据不等长问题,Nomikos等^[4]为解决该问题提出最短长度法,即将所有批次数据长度统一为众多批次中数据长度最短的那批数据的长度,该方法虽然简单,但是会造成数据的丢失,使数据之间的相关性降低;Cho等^[5]实时计算待测生产过程中已知的测量数据与历史过程数据相对应的那一部分距离,选择距离最小的值所对应的批次,将该批次的过程数据用于数据填充,但是该方法计算量过于复杂且单纯地把距离作为选取标准也不能代表实际生产过程中

的一些潜在关系和特点。

间歇过程还存在一个固有特性—多阶段特性,即随着生产过程的进行,控制目标和主导变量会随之发生改变^[6-8]。目前,多向核主元分析(MKPCA)方法在间歇过程中的应用较为常见^[9],但传统的MKPCA方法一般会使用整个批处理数据来构建模型,而忽略掉生产过程中的局部特征。针对多阶段特性,梁小凡等^[10]提出基于模糊C-均值算法的分时段过程监控算法实现间歇过程的阶段划分,但是模糊C-均值算法对初始聚类中心比较敏感,初始聚类中心选取不当会使阶段划分不准确,进而影响过程监测的准确性。张成等^[11]提出了一种模糊有序聚类算法,将模糊策略引入阶段划分中,但其只针对单独的批次数据进行聚类,最后才整合数据,准确度不高。

针对上述间歇过程研究方法的不足,本文提出自适应动态时间规整(DTW)算法以解决数据不等长问题,实现了 N 个批次数据时间长度一致的目标。针对间歇过程多阶段特性,采用遗传算法与模糊C-均值聚类(FCM)算法相结合的方法(GA-FCM)对间歇过程进行阶段划分,之后用MKPCA对每个子阶段分别建立模型并完成故障检测。

收稿日期:2019-04-29

基金项目:国家自然科学基金(61773151);河北省自然科学基金(F2018202279)

第一作者:女,1972年生,教授

E-mail: lxx68@163.com

1 多向核主元分析故障诊断模型

间歇生产的过程数据可以表示为三维矩阵 $\mathbf{X}(I \times J \times K)$, I 代表采样批次, J 代表过程变量, K 代表采样时刻。在建模时需要先将三维数据展开成二维数据, 展开方式主要有批次展开和变量展开^[12-14], 数据展开后 MKPCA 的建模过程与核主元分析(KPCA)相同。

KPCA 的基本思想是首先通过非线性映射将输入空间映射到特征空间, 然后提取该特征空间中的主要分量。设输入样本集 $\mathbf{X}_h = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, 其中 $\mathbf{x}_k \in \mathbf{R}^m$, N 为样本数。通过非线性映射 ϕ 将输入数据从输入空间投影到 KPCA 空间, 记为 $\phi(\mathbf{x}_k)$ 。定义核矩阵 $\mathbf{K} = [\mathbf{K}_{\beta\gamma}]$, 其中

$$\mathbf{K}_{\beta\gamma} = k(\mathbf{x}_\beta, \mathbf{x}_\gamma) = \langle \phi(\mathbf{x}_\beta), \phi(\mathbf{x}_\gamma) \rangle \quad (1)$$

引入核函数可以避免同时执行非线性映射和计算特征空间中两个向量内积的问题。对于任意测试向量 \mathbf{x} 的主要分量 \mathbf{t} 可按式(2)计算

$$\mathbf{t}_k(\mathbf{x}) = \sum_{\beta=1}^N \alpha_\beta^k \langle \phi(\mathbf{x}_\beta), \phi(\mathbf{x}) \rangle = \sum_{\beta=1}^N \alpha_\beta^k (\mathbf{x}_\beta, \mathbf{x}) \quad (2)$$

式中, $k=1, 2, \dots, p$, p 为保留主元数, α 为归一化后的核矩阵的特征向量。 T^2 和 SPE 统计量及控制限可参考文献[15]来计算。

2 自适应 DTW 算法

针对间歇生产过程中数据的不等长问题, 采用一种时间长度自适应 DTW 算法, 使采集的批次时间长度保持一致。本文采用欧氏距离计算两轨迹 \mathbf{A} 、 \mathbf{B} 之间的点-点距离, 从而可以获得 $a \times b$ 组局部距离。若设 $a \times b$ 中第 λ 个局部距离为 $d(\mu(\lambda), \eta(\lambda))$, 则

$$d(\mu(\lambda), \eta(\lambda)) = \{\mathbf{A}[\mu(\lambda), :] - \mathbf{B}[\eta(\lambda), :]\}^T \mathbf{W} \{\mathbf{A}[\mu(\lambda), :] - \mathbf{B}[\eta(\lambda), :]\} \quad (3)$$

式中, $\mu(\lambda)$ 、 $\eta(\lambda)$ 表示样本的时标, \mathbf{W} 是权矩阵, $\mathbf{W} = \mathbf{E}$, \mathbf{E} 是单位矩阵, $\mathbf{A}[\mu(\lambda), :]$ 表示 \mathbf{A} 中第 $\mu(\lambda)$ 个行向量, $\mathbf{B}[\eta(\lambda), :]$ 表示 \mathbf{B} 中第 $\eta(\lambda)$ 个行向量。在进行 DTW 算法同步化时求取的是两轨迹之间的最短距离, 在最短距离的基础上实现数据的处理。

用自适应 DTW 算法解决数据不等长问题时, 计算采集的 N 批实验数据平均时间长度 L , 选取 N 批里时间长度最接近 L 的批次 N_q 作为参考轨迹。

当 $L_{N_l} < L_{N_q}$ ($l \neq q$) 时先采用对称式 DTW 算法同步化再进行非对称式 DTW 算法同步化, $L_{N_l} > L_{N_q}$ ($l \neq q$) 时进行非对称式同步化, 得到同步化后的 N 个批次数据; 将 N 个批次数据进行平均化得到参考轨迹 R_{ref} , 平均化方式如式(4)

$$R_{\text{ref}} = \frac{1}{N} \sum_{l=1}^N N_l, l=1, 2, \dots, N \quad (4)$$

得到参考轨迹后, 将 N 批原始数据分别与 R_{ref} 进行对称式 DTW 算法或非对称式 DTW 算法同步化, 得到时间长度一致的 N 个批次数据。

将得到的 N 个批次的数据先按批次展开, 然后进行标准化处理, 将标准化后的数据再重新排列回 $\mathbf{X}(I \times J \times K)$ 的形式, 然后再按变量展开, 展开后沿时间轴方向将数据切割为 K 个时间片矩阵 $\mathbf{T}_k(I \times J)$, 计算每个时间片矩阵 \mathbf{T}_i 与其他时间片矩阵 \mathbf{T}_j 的相似度, 并将相似度矩阵 $\mathbf{S}(\mathbf{S} \in \mathbf{R}^{K \times K})$ 作为聚类算法的输入进行聚类分析。相似性度量的计算公式为

$$S_{ij} = \sum_{f=1}^I \sum_{z=1}^J (t_{i(fz)} - t_{j(fz)})^2 \quad (5)$$

3 GA-FCM 聚类算法的阶段划分

3.1 FCM 算法

FCM 算法是一种用 0~1 间的隶属度来确定每个样本点属于各个组的程度的模糊聚类方法^[16]。隶属度在归一化条件下满足

$$\sum_{g=1}^c u_{gh} = 1, \forall h=1, 2, \dots, n \quad (6)$$

FCM 算法是一种最小化价值函数的迭代优化算法, 价值函数定义为

$$J(\mathbf{U}, \mathbf{c}_1, \dots, \mathbf{c}_c) = \sum_{g=1}^c \sum_{h=1}^n u_{gh}^w d_{gh}^2 \quad (7)$$

式中, $u_{gh} \in (0, 1)$, \mathbf{c}_g 为组 g 的聚类中心, w 为模糊系数, $d_{gh} = \|\mathbf{c}_g - \mathbf{s}_h\|$ 为第 g 个聚类中心与第 h 个数据点间的欧氏距离。使式(7)达到最小的必要条件为

$$\mathbf{c}_g = \frac{\sum_{h=1}^n u_{gh}^w \mathbf{s}_h}{\sum_{h=1}^n u_{gh}^w} \quad (8)$$

$$u_{gh} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{gh}}{d_{kh}} \right)^{2/(w-1)}} \quad (9)$$

FCM 算法的缺点是对初始值较为敏感, 初始值如果选取不合适会使聚类效果变差, 因此本文使用

遗传算法对初始聚类中心进行参数寻优,改善 FCM 的聚类效果。

3.2 GA-FCM 算法

遗传算法是一种强大的随机算法,它通过选择目标函数将其与最优性概念联系起来,进而求得最优解。遗传算法是从代表问题可能解的一个种群开始的,并通过一些遗传操作(如选择、交叉和变异)有概率地修改种群,以寻求问题接近最优的解决方案。

3.2.1 种群初始化

S_{\min} 是相似性矩阵 S 每一维的最小值向量, S_{\max} 是相似性矩阵 S 每一维的最大值向量。首先通过随机的方式在 $[S_{\min}, S_{\max}]$ 区间内产生一代包含 H 个体的初始种群 M^1 , 该种群的计算方式为

$$M^1 = \begin{bmatrix} M_1^1 \\ M_2^1 \\ \vdots \\ M_O^1 \end{bmatrix} = \begin{bmatrix} rand(0,1) \times [M_{\max} - M_{\min}] + M_{\min} \\ rand(0,1) \times [M_{\max} - M_{\min}] + M_{\min} \\ \vdots \\ rand(0,1) \times [M_{\max} - M_{\min}] + M_{\min} \end{bmatrix}_{H \times 2400} \quad (10)$$

式中, M^1 为第一代群体, 是一个 H 行 1 列的向量组, 每一行是一个 1×2400 的向量, 代表一个个体, 由 2400 个染色体构成, 每个染色体是一个十进制编码的参数优化值, $rand(0,1)$ 则是 0,1 之间的随机数。

3.2.2 选择

选择操作采用轮盘赌法, 个体选择概率为

$$P(M_O^c) = Fit(M_O^c) / \sum_{o=1}^O Fit(M_O^c) \quad (11)$$

式中 $Fit(M_O^c) = 1/fit(M_O^c)$, $fit(M_O^c)$ 为个体适应度, $Fit(M_O^c)$ 为适应度倒数, $P(M_O^c)$ 为个体 M_O^c 在每次选择中被选中的概率, O 为每一代种群中所包含的个体总数。由于个体适应度越小则被选中的概率越大, 所以优化问题是求目标函数的最小值。在计算出个体被选择概率之后, 对其进行二进制编码, 则一个个体被扩充为 $1 \times (2400 \times D)$ 的 (0,1) 向量, D 为一个基因的长度。个体编码过程如图 1 所示。

3.2.3 适应度计算

首先对当代种群的个体进行解码, 将二进制数转换为十进制, 再将各个个体的聚类中心从一维矩阵转换为二维矩阵, 如图 2 所示。

适应度计算公式为

$$fit = \frac{D_1}{D_2} \quad (12)$$

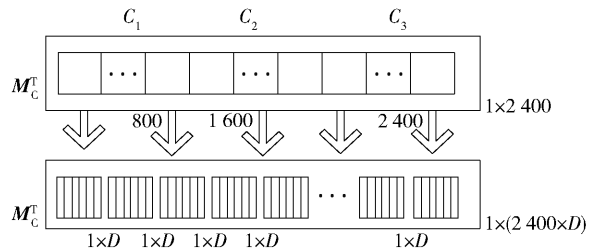


图 1 编码过程

Fig. 1 Coding process

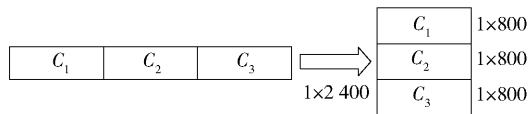


图 2 矩阵的转换

Fig. 2 Matrix conversion

式中, D_1 为类内距离, 表示同一模式点集内各样本本间的均方距离, 计算式为

$$D_1 = \sqrt{\frac{1}{c} \sum_{m=1}^c d^2(C_m - CT_m)} \quad (13)$$

式中, $d^2(C_m - CT_m) = \frac{1}{n_m} \sum_{m=1}^{n_m} d^2(C_m - n_{tm})$, n_{tm} 表示该集合中的所有特征, n_m 表示第 m 类样本集中包含的单位数目; c 为聚类个数所包含的特征集合; C_m 表示第 m 类样本集的聚类中心, 其中 $m = 1, 2, \dots, c$ 。

D_2 为类间距离, 表示模式类之间的距离, 计算式为

$$D_2(c) = \sqrt{\frac{1}{2c} \sum_{r=1}^c \sum_{e=1}^c (C_r - C_e)^2} \quad (14)$$

式中, C_r 表示第 r 类样本集的聚类中心, C_e 表示第 e 类样本集的聚类中心。

选取迭代过程中适应度最小的聚类中心作为遗传算法的解输出, 将输出的聚类中心作为 FCM 算法的初始值代入式(7)~(9), 进行 FCM 的聚类过程, 实现阶段划分。

改进 FCM 算法流程图如图 3 所示。

4 仿真结果

本文采用的间歇过程是青霉素生产的模拟补料分批发酵过程, 此生产过程具有非线性和多阶段的特征。根据工艺菌体生长周期, 青霉素发酵过程一般可以分为菌体生长、青霉素合成、菌体自溶 3 个阶段。PenSim2.0 是青霉素生产的模拟软件, 为检测青霉素的生产提供了标准平台, 本文实验数据均通

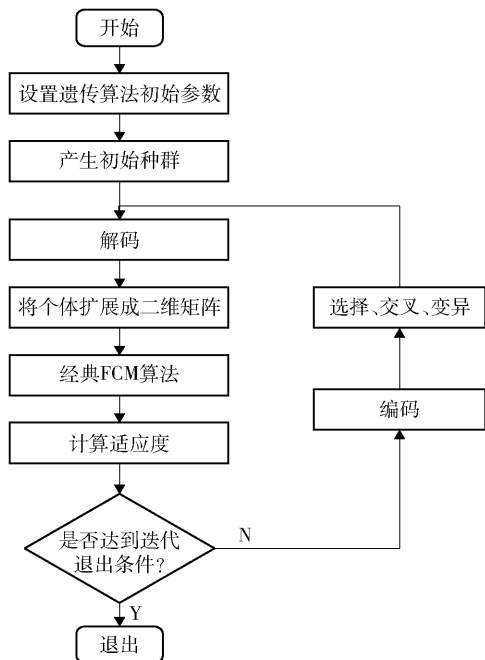


图3 改进 FCM 算法流程图

Fig. 3 Flow chart of GA-FCM algorithm

过此软件获得。

4.1 聚类有效性

聚类有效性可用类内距离与类间距离之比来衡量,即

$$F = \frac{D_1}{D_2} \quad (15)$$

D_1 数值越小,聚类效果越好; D_2 数值越大,聚类效果越好。将两个指标结合起来,则 F 值越小代表聚类效果越好。

经式(15)计算可得,本文所提算法 F 值为 1.6, FCM 算法 F 值为 2.3,可见本文所提算法聚类更为准确。

4.2 故障检测结果

本文选择了青霉素发酵过程中的 17 个变量进行研究,包括充气率、搅拌器功率、底物流加速率、底

物流温度、底物浓度、溶解氧浓度、生物质浓度、青霉素浓度、培养体积、二氧化碳浓度、pH、生物反应器温度、产生的热量、酸流率、碱流率、冷水流量、热水流量,模拟了 20 个发酵时间从 390 h 到 410 h 不等、采样间隔为 0.5 h 的正常参考批次,其中这些批次的平均时间长度为 400.2 h。选择发酵时间为 400 h 的这一批数据为 N_q 。最终采集的 20 个批次数据的时间长度经过自适应 DTW 算法后均调整到 400 h,实现了数据的等长。

分别通过 MKPCA 算法、FCM-MKPCA 算法和本文所提 GA-FCM-MKPCA 算法对间歇过程进行故障检测。

图 4 为本文所提算法的聚类结果。由图 4 可知,本文算法将间歇过程划分为 3 个阶段:1~188 是第一阶段;189~357 是第二阶段;358~800 是第三阶段。此阶段划分满足青霉素实际生产过程的特点。阶段划分完成后,对 3 个阶段分别建立 MKPCA 模型,进行故障检测。

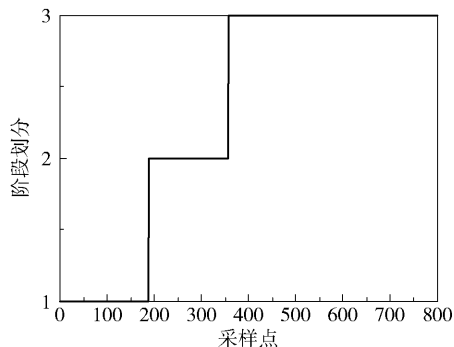


图4 GA-FCM 算法阶段划分

Fig. 4 Stage division result of GA-FCM algorithm

MKPCA 算法、FCM-MKPCA 算法和本文所提 GA-FCM-MKPCA 算法对正常批次的检测结果如图 5 所示。

由图 5 可知, MKPCA 算法 T^2 误报率为

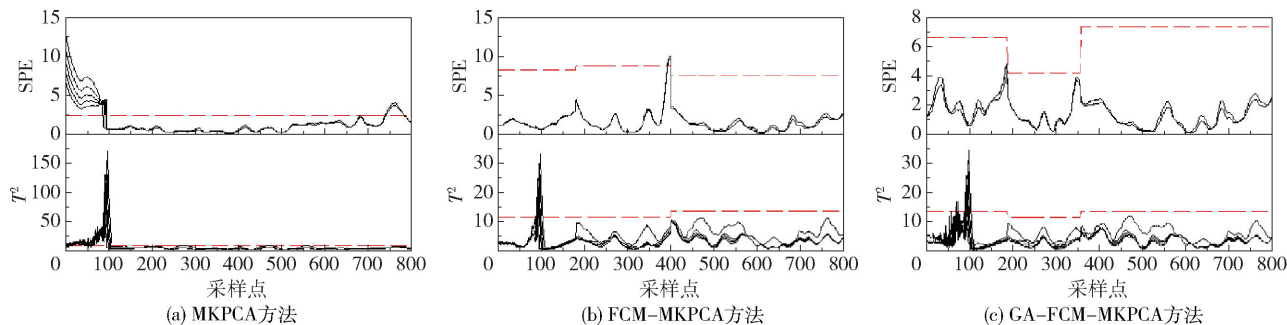


图5 正常批次仿真图

Fig. 5 Results of normal batch simulation

12.35%, FCM-MKPCA 算法 T^2 误报率为 1.95%, 而本文所提方法将 T^2 误报率降到了 0.9%; MKPCA 算法 SPE 误报率为 17.15%, FCM-MKPCA 算法 SPE 误报率为 0.88%, 本文所提算法 SPE 误报率为 0.13%。可见利用本文所提方法划分阶段后再建立模型会降低误报率。

本文还采集了两种故障数据进行检测:故障 1 对充气率引入故障信号,在 200 h 施加 +3% 阶跃信号,该信号持续至第 300 h,诊断结果如图 6 所示;故障 2 在 200 h 时对充气率施加 +0.2% 斜坡信号,该信号持续至第 300 h,诊断结果如图 7 所示。

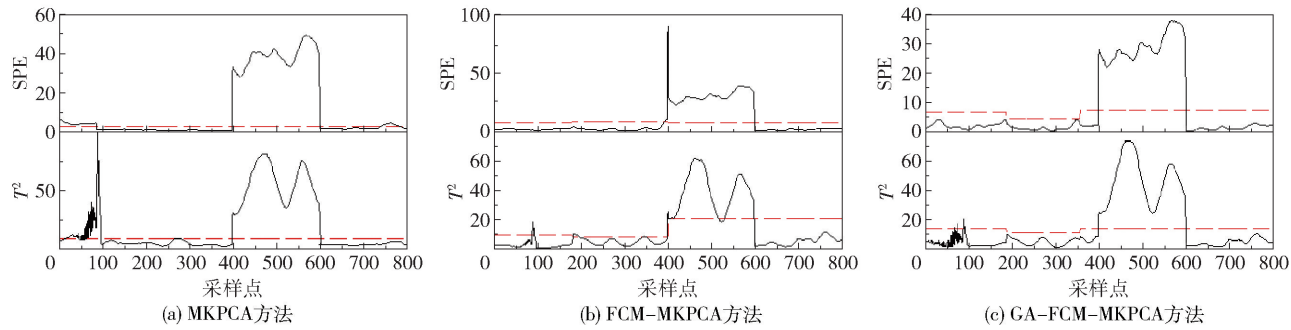


图 6 故障 1 检测图
Fig. 6 Diagnosis diagram of fault 1

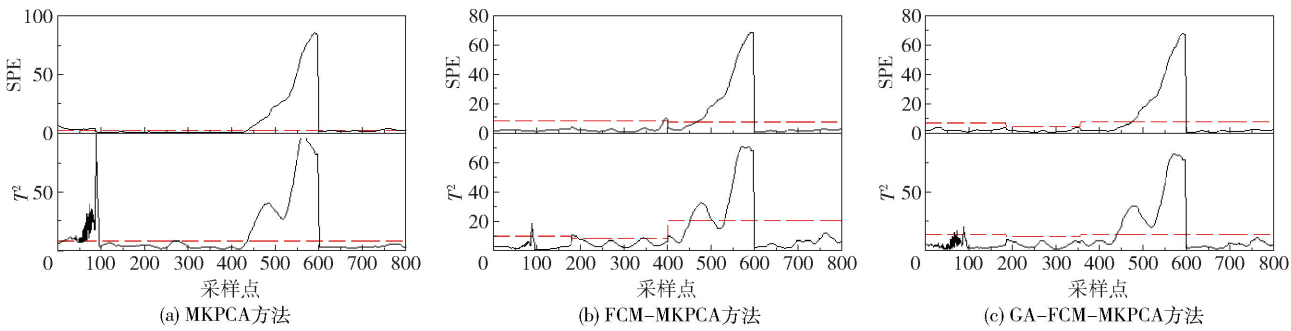


图 7 故障 2 检测图
Fig. 7 Diagnosis diagram of fault 2

从图 6 可知,FCM-MKPCA 算法 T^2 漏报率为 2.25%, MKPCA 方法和本文所提方法漏报率为 0, MKPCA 算法 T^2 误报率为 10.38%, FCM-MKPCA 算法 T^2 误报率为 3.63%, 本文所提方法 T^2 误报率为 0.88%; 3 种方法的 SPE 漏报率均为 0, 而对于 SPE 误报率, MKPCA 算法为 16.75%, FCM-MKPCA 算法为 1%, 本文所提算法降到了 0.1%。由此可知, 本文所提方法误报率大大降低, 故障诊断精度明显提高。

从图 7 可知, 对于 T^2 漏报率, MKPCA 算法为 5.00%, FCM-MKPCA 算法为 10.13%, 本文所提算法为 5.25%, 对于 T^2 误报率, MKPCA 算法为 10.25%, FCM-MKPCA 算法为 3.5%, 本文所提算法为 0.75%; 3 种方法得出的 SPE 漏报率基本一致, 而对于 SPE 误报率, MKPCA 算法为 16.63%, FCM-MKPCA 算法为 1.01%, 本文所提算法为 0。综上可

知, 本文所提方法的故障诊断准确度更高。

从图 5~7 可知, 50~100 采样点发生的误报比较多, 这是因为在此阶段菌体进入发酵期, 菌体浓度、溶解氧浓度、pH、碱补给等变量在此时段内发生波峰式明显变化, 并没有趋于稳定值, 虽然数据经过了 MKPCA 归一化处理, 但仍然会有生物过程的波动特性, 所以此时段误报最多。将青霉素发酵过程用本文方法分阶段后, 把这一过程划分到第一阶段, 并单独建立 MKPCA 的模型, 据此获得此阶段下合理的 T^2 、SPE 的受控限, 相比于整个批次获得的 T^2 、SPE 的受控限, 利用第一阶段获得的受控限进行该阶段的故障检测, 大大降低了误报率。

通过以上仿真实验的结果可知, 无论是对阶跃故障还是对斜坡故障, 本文所提方法的故障诊断精度都要比传统不分阶段的方法以及 FCM 分阶段的

方法高。

5 结论

(1) 针对间歇过程数据不等长问题, 提出了一种自适应 DTW 算法。自适应 DTW 算法发挥了对称式 DTW 算法与非对称式 DTW 算法在算法上的互通性与互补性, 将不等长数据变为等长数据, 实现了多个批次数据轨迹的时间长度同步化。

(2) 针对 FCM 算法在阶段划分中存在的对初始聚类中心敏感的问题, 提出遗传算法与 FCM 算法相结合来完成聚类过程。遗传算法采用的是整体搜索策略, 它从实际问题中抽离出来, 采用全局搜索方式, 通过选择、交叉、变异等操作完成搜索过程, 具有强大的全局搜索能力。遗传算法可以有效、快速地解决 FCM 算法对初始聚类中心敏感且易于陷入局部收敛的问题。

(3) 仿真实验的结果表明, GA-FCM 算法比单一 FCM 算法有更高的聚类有效性, 阶段划分更为准确, 故障诊断精度也明显提高。

参考文献:

- [1] YAN Z B, HUANG B L, YAO Y. Multivariate statistical process monitoring of batch-to-batch startups[J]. *AIChE Journal*, 2015, 61(11): 3719-3727.
- [2] 张成, 李元. 基于统计模量分析间歇过程故障检测方法研究[J]. *仪器仪表学报*, 2013, 34(9): 2103-2110.
ZHANG C, LI Y. Study on the fault-detection method in batch process based on statistical pattern analysis[J]. *Chinese Journal of Scientific Instrument*, 2013, 34(9): 2103-2110. (in Chinese)
- [3] 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. A method of batch process monitoring based on sliding-time-window weighted MPCA[J]. *Journal of Beijing University of Chemical Technology (Natural Science)*, 2015, 42(4): 112-119. (in Chinese)
- [4] NOMIKOS P, MACGREGOR J F. Multivariate SPC charts for monitoring batch processes[J]. *Technometrics*, 1995, 37(1): 41-59.
- [5] CHO H W, KIM K J. A method for predicting future observations in the monitoring of a batch process[J]. *Journal of Quality Technology*, 2003, 35(1): 59-69.
- [6] 王姝, 常玉清, 杨洁, 等. 时段划分的多向主元分析间歇过程监测及故障变量追溯[J]. *控制理论与应用*, 2011, 28(2): 149-156.
WANG S, CHANG Y Q, YANG J, et al. Multiway principle component analysis monitoring and fault variable detection based on substage separation for batch processes[J]. *Control Theory & Applications*, 2011, 28(2): 149-156. (in Chinese)
- [7] 于涛, 刘笑梅, 王建林, 等. 一种多阶段发酵过程的联合加权支持向量机建模方法[J]. *北京化工大学学报(自然科学版)*, 2016, 43(4): 102-108.
YU T, LIU X M, WANG J L, et al. A modeling approach for multi-stage fermentation processes based on jointly weighted support vector machine[J]. *Journal of Beijing University of Chemical Technology (Natural Science)*, 2016, 43(4): 102-108. (in Chinese)
- [8] 高学金, 黄梦丹, 齐咏生, 等. PDPSO 优化多阶段 AR-PCA 间歇过程监测方法[J]. *化工学报*, 2018, 69(9): 3914-3923.
GAO X J, HUANG M D, QI Y S, et al. Batch process monitoring using multiphase AR-PCA optimized with PDPSO[J]. *CIESC Journal*, 2018, 69(9): 3914-3923. (in Chinese)
- [9] 吕宁, 颜鲁齐, 白光远. 分段多向核主元分析的啤酒发酵过程故障检测[J]. *计算机科学*, 2016, 43(6A): 25-27, 33.
LV N, YAN L Q, BAI G Y. Fault detection for beer fermentation process based on segmentation multiway kernel principal component analysis[J]. *Computer Science*, 2016, 43(6A): 25-27, 33. (in Chinese)
- [10] 梁小凡, 熊伟丽, 徐保国. 基于 FCM 的青霉素发酵分时段统计建模及监控[J]. *计算机与应用化学*, 2013, 30(12): 1427-1430.
LIANG X F, XIONG W L, XU B G. Multi-phase statistical modeling and monitoring based improved FCM for penicillin fermentation process[J]. *Computers and Applied Chemistry*, 2013, 30(12): 1427-1430. (in Chinese)
- [11] 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 C, ZHAO H T, SUN S Y. Fault detection for batch processes based on fuzzy order clustering algorithm[J]. *Journal of Donghua University (Natural Science)*, 2018, 44(6): 940-947. (in Chinese)
- [12] NG Y S, SRINIVASAN R. An adjoined multi-model approach for monitoring batch and transient operations[J].

- Computers and Chemical Engineering, 2009, 33(4): 887-902.
- [13] 曹雪, 王建林, 邱科鹏, 等. 基于PARAFAC2的多时段间歇过程时段划分[J]. 北京化工大学学报(自然科学版), 2019, 46(2): 77-82.
- CAO X, WANG J L, QIU K P, et al. PARAFAC2-based phase partition of multiphase batch processes[J]. Journal of Beijing University of Chemical Technology (Natural Science), 2019, 46(2): 77-82. (in Chinese)
- [14] 王建林, 马琳钰, 邱科鹏, 等. 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.
- WANG J L, MA L Y, QIU K P, et al. Multi-phase batch processes fault detection based on support vector data description[J]. Chinese Journal of Scientific Instrument, 2017, 38(11): 2752-2761. (in Chinese)
- [15] 刘世成, 王海清, 李平. 基于多向核主元分析的青霉素生产过程在线监测[J]. 浙江大学学报(工学版), 2007, 41(2): 202-207.
- LIU S C, WANG H Q, LI P. On line monitoring of penicillin production process based on multiway kernel principal component analysis[J]. Journal of Zhejiang University (Engineering Science), 2007, 41(2): 202-207. (in Chinese)
- [16] KARI T, GAO W S. Power transformer fault diagnosis using FCM and improved PCA[J]. Journal of Engineering, 2017, 2017(14): 2605-2608.

Fault diagnosis of GA-FCM multi-stage batch processes based on adaptive dynamic time warping algorithm

LIANG XiuXia CHEN JiaoJiao YAN Ting ZHOU Ying ZHANG Yan

(School of Artificial Intelligence, Hebei University of Technology, Tianjin 300130, China)

Abstract: Multi-stage is an important characteristic of many batch processes. The division of batch processes could improve the accuracy of fault diagnosis. The fuzzy C-means clustering (FCM) algorithm is sensitive to initial cluster center and easy to fall into local excellent value, which may not search the global cluster center. The combination of genetic algorithm and FCM algorithm (GA-FCM) is proposed in this paper to overcome the problem that the fuzzy C-means clustering algorithm is easy to fall into local optimal value, which could achieve global optimization. In addition, an adaptive dynamic time warping (DTW) algorithm is presented for the batch process data with different time length. After that, the GA-FCM method is employed to accomplish the process of stage division and the multi-way kernel principal component analysis (MKPCA) model is established to accomplish the fault detection. Finally, the proposed method is utilized to the fermentation process of penicillin and the obtained results clearly demonstrate the power and advantages of the proposed method.

Key words: fuzzy C-means clustering (FCM) algorithm; genetic algorithm(GA); dynamic time warping (DTW); multi-way kernel principal component analysis (MKPCA); fault detection; batch process

(责任编辑:吴万玲)