

基于扩展核熵负载矩阵的发酵过程故障监测

高学金¹, 杨彦霞¹, 王 普¹, 李晓理^{1†}, 常 鹏¹, 齐咏生²

(1. 北京工业大学 信息学部, 北京 100124; 2. 内蒙古工业大学 电力学院, 内蒙古 呼和浩特 010051)

摘要: 为有效降低多阶段发酵过程硬分类缺陷而导致的误报和漏报率, 本文提出了一种基于扩展核熵负载矩阵的阶段划分策略. 首先, 将发酵过程的三维训练数据按批次方向展开成二维数据矩阵, 对每个时间片矩阵进行核熵成分分析(kernel entropy component analysis, KECA)得到其主元和负载矩阵, 根据所得主元个数实现操作阶段的第1步划分; 之后将时间片矩阵添加到核熵负载矩阵当中得到扩展核熵负载矩阵, 计算各扩展负载矩阵间的相似度, 并用模糊C-均值方法对其进行第二次阶段划分. 通过增加对体现生产过程改变的时间指标的考虑, 有效克服了硬化分的不足, 避免了跳变点错分的情况. 最终将整个生产操作过程划分为不同的稳定阶段和过渡阶段, 并在划分的每一阶段中分别建立KECA监测模型; 最后利用青霉素发酵仿真平台和大肠杆菌生产白介素-2数据进行实验. 实验结果表明该方法不但可以准确地对生产过程进行阶段划分、降低误报率, 而且可以使生产过程故障监测的时间大大提前.

关键词: 过程监测; 主元分析; 多阶段; 发酵过程

引用格式: 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. 控制理论与应用, 2018, 35(6): 813 – 821

中图分类号: TH165+.3 文献标识码: A

Fault monitoring of fermentation process based on extended kernel entropy load matrix

GAO Xue-jin¹, YANG Yan-xia¹, WANG Pu¹, LI Xiao-li^{1†}, CHANG Peng¹, QI Yong-sheng²

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;

2. School of Electric Power, Inner Mongolia University of Technology, Huhhot Inner Mongolia, 010051, China)

Abstract: Hard classification for multistage fermentation process and cause of the defects of false alarm and alarm failure, in order to effectively reduce the omission and the rate of false positives, this paper proposes a strategy based on extended nuclear entropy load matrix. First, the three-mention training data array of fermentation process is unfolded in batch ways, resulting in two-dimension forms. Then, kernel entropy component analysis (KECA) was done for each time slice matrix to obtain its load matrix. After that, time slice matrix was added to the nuclear load matrix of entropy, and the change of the nuclear load matrix of entropy was utilized to describe the changes of batch processes. The KECA monitoring model was established at each stage of the division after the stage of nuclear load matrix of entropy was determined by FCM algorithm. At last, the effectiveness and utility of the proposed method were validated through the simulation of fed-batch penicillin and E. coli production of interleukin-2. Results showed, the proposed method could not only divide the stage and reduce the false alarm precisely, but also detect the production difficulty more advance.

Key words: process monitoring; principal component analysis; multistage; fermentation process

Citation: GAO Xuejin, YANG Yanxia, WANG Pu, et al. Fault monitoring of fermentation process based on extended kernel entropy load matrix. *Control Theory & Applications*, 2018, 35(6): 813 – 821

1 引言(Introduction)

近年来, 间歇生产方式正逐渐超越连续生产方式成为市场的主流, 尤其在医学和生物制药领域. 基于数据驱动多元统计方法由于不必考虑复杂的过程

机理特性, 只需正常操作状态下的历史数据即可建立模型并进行在线故障监测, 在间歇过程领域取得了广泛应用. 尤其主成分分析(principal component analysis, PCA)^[1]方法和偏最小二乘(partial least square,

收稿日期: 2017-08-01; 录用日期: 2018-01-27.

†通信作者. E-mail: lixiaolibjut@bjut.edu.cn; Tel.: +86 10-67396716.

本文责任编辑: 周东华.

国家自然科学基金项目(61640312, 61473034, 61673053, 61174109), 北京市自然科学基金项目(4172007), 北京科技新星计划交叉学科合作项目(Z161100004916041)资助.

Supported by the National Natural Science Foundation of China (61174109, 61640312, 61473034, 61673053), the Beijing Natural Science Foundation (4172007) and the Beijing Science and Technology Nova Programs Interdisciplinary Projects (Z161100004916041).

PLS)^[2]方法以处理高维数据的独特优势越来越受到科研人员的青睐. 1994年Nomikos和Mac Grego首次提出多向主元分析(multiway PCA, MPCA)^[3]和多向偏最小二乘回归(multiway PLS, MPLS)^[4]并成功应用到间歇过程中, 掀起了基于多元统计技术的间歇过程建模、监测的热潮.

多阶段特性是间歇生产过程的一个固有特性, 不同阶段内有不同的关键过程变量和控制目标^[5-6], 且过程变量的均值和方差也有显著差异. 目前, 针对间歇过程的多阶段特性, 国内外专家做了大量研究工作. 如Lu等人^[7]利用PCA分解后的负载矩阵将整个生产过程划分为多个稳定操作阶段并在每一阶段分别建立模型进行监测取得了较好的效果, 但是这种方法属于硬分类的方法, 忽略了相邻两阶段间的过渡过程, 过渡过程虽不是生产过程的主要机理过程, 却也是普遍存在的过程行为. 在此基础上Zhao等人^[8]提出了基于K-means的间歇过程时段划分方法, 该方法引入模糊隶属度将其作为两个相邻稳定阶段的权重系数, 用两个相邻的稳定阶段来近似描述过渡阶段, 提高了模型的精度. 针对过渡阶段的过程检测, Qi等人^[9]提出了0~1模糊隶属度的概念, 提出在稳定阶段建立PCA模型, 过渡阶段建立KPCA模型, 该方法充分考虑了过渡阶段的非线性特性. Chang^[10]提出了基于信息熵的多阶段多向核熵成分分析(MKECA)方法, 通过在特征空间中依据熵结构信息计算时刻间矩阵的相似度来进行阶段划分, 与KPCA方法不同的是: KECA方法^[11]以信息熵大小作为选取主元的指标, 充分考虑了特征值和特征向量的变化, 取得了成功应用.

此外, 对于建模时数据非线性和软过渡的问题, 前人也有相关研究^[12-13]. Zhao等人^[12]用核映射的方法解决非线性, 在核映射后的高维空间中进行间歇过程的阶段划分和PCA建模得到主成分矩阵, 之后又进行ICA建模得到独立成分矩阵. 最后, 将主成分矩阵和独立成分矩阵进行综合得到一个新矩阵进行聚类分析, 取得了很好的效果. Zhao等人^[13]利用软过渡的方法先对第一个时刻过程数据进行PCA分析, 计算控制限; 之后不断叠加新时刻数据并计算相应的控制限, 通过相似性的比较来判断是否属于同一个阶段, 以此实现聚类. 但是, 以上方法多利用负载矩阵进行阶段划分, 忽略了跳变点存在的情况, 若负载矩阵中存在噪声或奇异值则势必会影响阶段划分的准确性, 从而影响模型的精确度造成误报或漏报.

针对以上所述问题, 本文提出一种基于扩展核熵负载矩阵的阶段划分方法. 首先对沿批次展开的二维数据矩阵进行核熵成分分析得到核熵负载矩阵和主元, 并以主元个数进行阶段的初步划分; 之后将时间片矩阵扩展到核熵负载矩阵中计算各扩展核熵负载矩阵间相似度, 再用模糊C-均值方法进行阶段的二次

软划分, 最终将整个生产过程划分为不同的稳定子时段和过渡子时段. 与前人工作相比, 本文方法主要有以下3个优点: 1) 在核特征空间中, 对原始数据进行KECA分析, 通过对输入空间熵值的贡献大小来选取主元, 不同类数据的主元聚集在不同的坐标轴附近. 因为所选主元携带了一定的角度信息, 使得原始数据的空间结构信息得到了更好的保留. 2) 数据特征的变化很大程度上直接反映了操作过程的变化, 本文利用反映过程特性的数据矩阵进行聚类, 得到的阶段划分结果更具代表性、更符合实际生产过程. 3) 将能反应过渡特性的时间变量充分考虑进来, 利用时间片矩阵和负载矩阵的综合变化来近似描述间歇生产的变化, 使得模型准确性大大提高.

2 发酵过程机理分析 (Analysis of fermentation process)

发酵过程的基本流程如图1所示, 主要包括: 1) 培养基的配制; 2) 发酵罐、培养基及其辅助设备的消毒灭菌; 3) 将培养好的菌株转接到发酵罐中; 4) 使接到发酵罐中的菌株在适宜条件下生长并生成相应的代谢产物; 5) 对产物进行分离和萃取以得到所需合格产品; 6) 处理发酵过程产生的废水和废料.

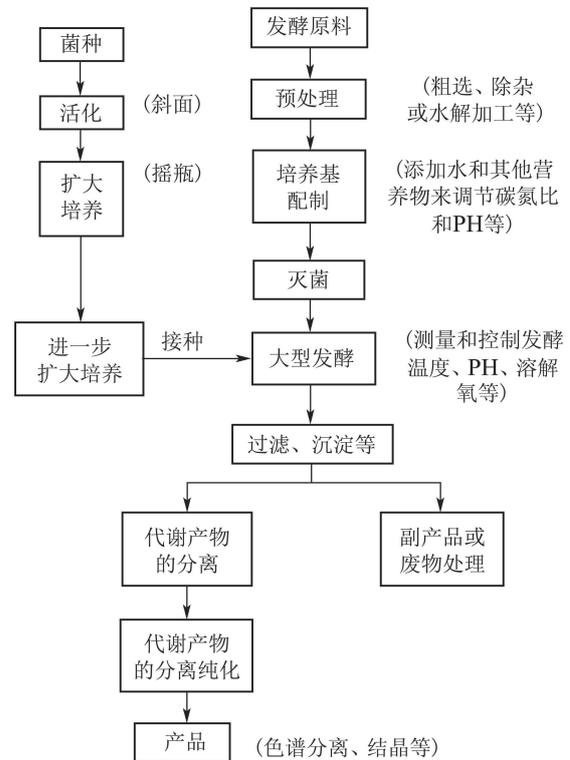


图1 发酵过程流程图

Fig. 1 Fermentation process flow chart

发酵过程作为间歇过程的一种典型, 非常复杂. 具有动态时变性、批次性、多阶段等特性. 发酵过程产品质量波动大, 故障不易及时发现. 一旦发现, 发酵过程已经不可逆转, 造成原材料的浪费. 因此, 确保发酵

过程产品效价的稳定非常重要. 准确并及时地对发酵过程处于“正常”状态, 还是处于“故障”状态进行判断和识别也是非常必要的^[14].

及时地测量发酵过程重要变量参数或者效价等, 实际上是及早发现过程异常最基本的手段. 但是, 这种方法存在一定的问题: 1) 需要不定时取样; 2) 需要昂贵的设备和有一定操作经验的工人, 成本较高; 3) 样品分析需要离心、稀释等多道工序, 存在很长的测量滞后. 通过获取在线的实际生产数据, 结合有效的数据处理方法, 从数据中挖掘关键信息和情报, 是实现发酵过程故障监测和早期预警的有效途径.

不同的过程特性产生了不同操作阶段, 阶段不等长导致了各个批次的不同长. 关于历史批次的数据不等长问题, 众多学者分别提出了多种策略. 其中比较典型的是最短长度法和最长长度法, 也统称为“截取法”^[15-16]. 这两种方式虽在一定程度上解决了批次不等长问题, 但都将数据的原始有效信息进行了删除, 造成了模型精度降低、监测结果不理想. 因此, 为尽可能多的保留原始信息、体现数据特征. 本文不进行任何的截取操作, 而是假定每个批次都是等长的, 以此建立监控模型, 对整个生产操作过程进行监测.

3 聚类方法分析(Method of clustering)

3.1 核熵成分分析(Kernel entropy component analysis)

2010年, Jenssen基于Renyi熵和Parzen窗密度估计两个概念提出了核熵成分分析(KECA)的方法^[11]. KECA方法的原理和KPCA非常相似, 都需要对核矩阵进行特征分解, 提取主成分. 不同的是, KECA不是以方差小来选择特征向量, 而是选择前 n 个对瑞利熵贡献最大的特征值所对应的特征向量. 详细实现过程如下:

$$H(p) = -\log V(p) = -\log \int p^2(x)dx, \quad (1)$$

式中: x 是样本, $p(x)$ 是样本的概率密度函数.

利用Parzen窗密度估计对 $V(p)$ 进行估计, 如式(2)所示:

$$\hat{p}(x) = \frac{1}{N} \sum_{x_i \in \mathbb{D}} k_{\sigma}(x, x_i), \quad (2)$$

令

$$k_{\sigma}(x, x_i) = \exp(-\|x - x_i\|/2\sigma^2). \quad (3)$$

对 $V(p)$ 进行估计, 可得到二次Renyi熵的估计如下:

$$\hat{H}(p) = -\log\left[\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N k_{\sqrt{2}\sigma}(x_j, x_i)\right], \quad (4)$$

由于单调性不变, 所以只考虑式中 $-\log$ 后面的部分得到

$$\hat{V}(p) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N k_{\sqrt{2}\sigma}(x_j, x_i) = \frac{1}{N^2} I^T K I, \quad (5)$$

式中: I 是 $N \times 1$ 的向量; K 是核矩阵. 将核矩阵 K 特征分解, $K = EDE^T$, 其中

$$D = \text{diag}\{\lambda_1, \dots, \lambda_N\},$$

$$E = (e_1, \dots, e_N),$$

详细求解过程参考文献[17-18]. 即

$$\hat{V}(p) = \frac{1}{N^2} \sum_{i=1}^N (\sqrt{\lambda_i} e_i^T I)^2. \quad (6)$$

从式中可以看出, 不同的特征值和特征向量对Renyi熵的贡献不同. KECA算法根据熵值贡献率由大到小依次选取前 P 个特征值及对应的特征向量. 得到投影后的主成分矩阵如下:

$$\phi_{\text{eca}} = D_i^{1/2} E_i^T. \quad (7)$$

3.2 模糊C-均值聚类方法(Fuzzy C-means)

模糊C-均值聚类算法的基本原理是根据数据间的相似度进行聚类, 通过数据点的多次迭代得到每个类的聚类中心及每个样本的隶属度值. 使同类样本间数据特征相似度值尽可能大, 不同类样本间数据特征相似度值尽可能小, 得到聚类结果. 将 n 个样本 $X = (x_1, x_2, \dots, x_n)$ 划分成 C 类. 模糊C-均值的最优函数表达式

$$E^{\text{FCM}} = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m d_{ij}^2, \quad (8)$$

式中: N 是样本数, m 是模糊隶属度的权重, C 是类的个数, $d_{ij} = \|x_j - v_i\|$ 代表第 j 个样本与第 i 个聚类中心 v_i 间的距离, u_{ij} 是 x_i 属于 v_i 的模糊程度.

模糊隶属度的取值范围为 $[0, 1]$, 每个样本的隶属度值相加之和为1. 即: $u_{ij} \in [0, 1]$, $\sum_{i=1}^C u_{ij} = 1, \forall j$ 使目标函数取得最小值条件为

$$E^{\overline{\text{FCM}}} = E^{\text{FCM}} + \sum_{j=1}^N \lambda^j \left(\sum_{i=1}^C u_{ij}^m - 1 \right) = \sum_{i=1}^C \sum_{j=1}^N u_{ij} d_{ij}^2 + \sum_{j=1}^N \lambda^j \left(\sum_{i=1}^C u_{ij} - 1 \right), \quad (9)$$

其中 λ^j 代表拉格朗日乘子. 对 u_{ij} 和 v_i 求导, 得到隶属度和聚类中心的迭代公式即

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}, \quad (10)$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}, \quad (11)$$

由式(10)和式(11)可知, 模糊C-均值经过多次迭代更新隶属度和聚类中心, 最终得到目标函数的最小值.

3.3 基于扩展核熵负载矩阵的阶段化分 (stage division based on extended kernel entropy load matrix)

由于测量的间歇过程数据中常常包含有噪声和奇异值,会导致数据在一些时刻发生突然变化,进而数据间的相关关系和核熵负载矩阵 P_k 也随之发生变化,造成阶段间的错误划分,即把当前阶段的数据错分到其它阶段中,在线监测时产生漏报和误报,甚至失去监测性能.考虑到间歇过程不同阶段间的巨大差异性首先体现在采样时刻上,为了解决阶段的误划分问题,本文将采样时刻 t_k 扩展到核熵负载矩阵当中,利用扩展核熵负载矩阵的变化来间接描述间歇过程的变化.扩展核熵负载矩阵为 $\hat{P}_k = [P_k \ t_k]$,欧式距离为

$$\|\hat{P}_i - \hat{P}_j\| = \sqrt{[P_i - P_j][P_i - P_j]^T + \|t_i - t_j\|^2} \quad (12)$$

由此可以看出,扩展核熵负载矩阵之间的欧式距离由两部分组成.一部分是负载矩阵间的欧式距离,而另一部分是采样时刻间 t_k 的欧式距离.如此,当不同阶段的负载矩阵间距离 $\|P_i - P_j\|^2$ 突然变小时,采样时刻间的距离 $\|t_i - t_j\|^2$ 也会作为一个约束使得两组数据被划分至不同的阶段.图2为本文所提阶段划分方法的原理图,具体实现过程如下:

传统的过程数据为三维数据形式 $X(I \times J \times K)$,其中 I, J, K 分别表示批次数、过程变量个数和采样时间^[18].为实现建模,需将三维的历史数据展开为二维.建模需求不同,故展开方式也不同,常用的展开方式有6种^[19].为更多的保留原始数据特征,从而保证所建模型的精确.本文假设建模所需的各批次操作时间都是等长的.

1) 将三维建模数据 X 沿批次方向切分为 K 个时间片矩阵 $X_k(I \times JK), k = 1, 2, \dots, K$.并进行标准化.

$$\tilde{x}_{k,j} = \frac{x_{k,j} - \bar{x}_{k,j}}{S_{k,j}} \quad (13)$$

其中: $x_{k,j}$ 为第 k 采样时刻的第 j 个过程变量, $\bar{x}_{k,j}$ 为第 k 采样时刻的第 j 个过程变量的平均值, $S_{k,j}$ 为第 k 采样时刻第 j 个过程变量的标准差, $j = 1, 2, \dots, J, k = 1, 2, \dots, K$.

2) 对每个时间片矩阵进行KECA分析,依据所得主元个数,实现生产过程的初步硬划分.

3) 将采样时间 t 扩展到核熵负载矩阵 $P_i(i = 1, 2, \dots, K)$ 中得到扩展核熵负载矩阵 $\hat{P}_i = [P_i \ t_i]$,本文定义负载矩阵 \hat{P}_i 和 \hat{P}_j 之间的相似度为

$$D = \text{diss}(\hat{P}_i, \hat{P}_j) = \frac{4}{J} \sum_{j=1}^J (\lambda_i^j - 0.5)^2 =$$

$$\frac{4}{J} \sum_{j=1}^J (\lambda_i^j - 0.5)^2, \quad (14)$$

根据对称性,若 \hat{P}_i 和 \hat{P}_j 差异较大,则 λ_i^j 接近1或0;否则应有 λ_i^j 接近0.5.

4) 将所得的各矩阵间的相似度值作为输入,用模糊C-均值方法进行阶段的二次划分.最终,将整个生产过程划分为3个稳定子时段和两个过渡子时段.

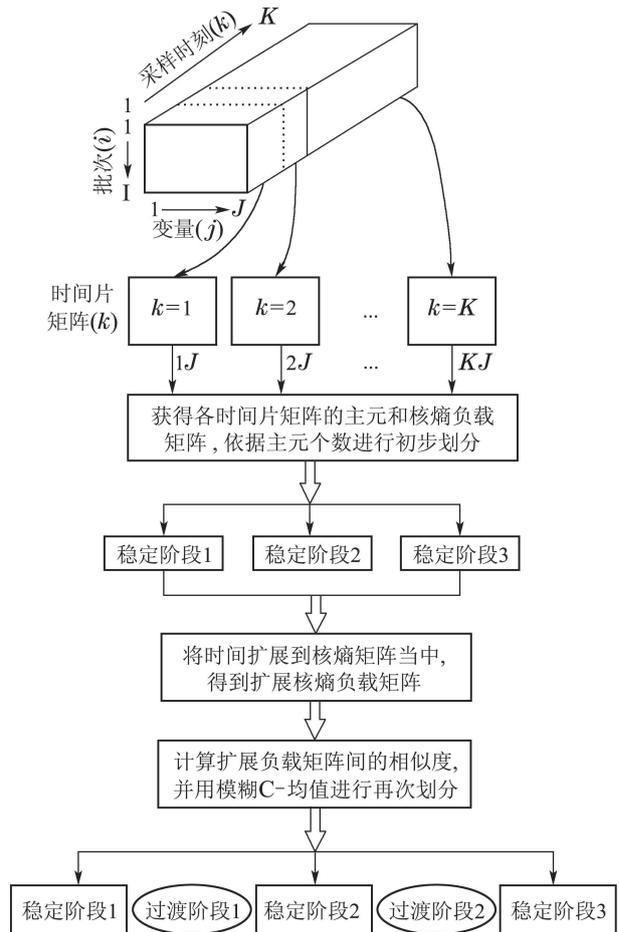


图2 基于扩展核熵负载矩阵的阶段划分

Fig. 2 Stage division based on extended kernel entropy load matrix

本文方法将反应时间这个重要过程变量考虑在内根据数据矩阵进行阶段的二次划分,较传统的硬化分结果更加精确;此外,在主元个数的选取上,本文选取4个主元即可达到原始数据在特征空间比值98%以上的信息,而传统的KPCA至少需要选取25个主元累计方差贡献率方能达到85%以上,主元个数的减少,直接降低了运算量.

4 基于扩展核熵负载矩阵的发酵过程故障检测 (Fault monitoring of fermentation process based on extended kernel entropy load matrix)

基于扩展核熵负载矩阵的发酵过程故障监测可分

为两部分: 离线建模、在线监测。

4.1 基于扩展核熵负载矩阵的发酵过程离线建模 (Off-line modeling of fermentation process based on extended kernel entropy load matrix)

阶段划分的目的是将具有高度相似特征的数据进行归类, 并建立统一模型。以局部模型代替整体模型可以大大提高模型的精度, 减小模型误差, 降低误报和漏报率。在第3.3节阶段划分结束后, 于每一阶段内分别建立离线监测模型, 计算出相应的控制限, 进而对生产过程实行在线监测。

1) 对阶段划分后的每一阶段数据 $X_{C_i}(I \times J \times K_{C_i})$ 沿批次方向展开为 $X_{C_i}(I \times JK_{C_i})$, 并进行标准化。其中: I 为批次数, J 为变量数, K_{C_i} 代表第 i 个阶段 C_i 所包含的时刻数。

2) 对标准化后的 $X_{C_i}(I \times JK_{C_i})$ 数据, 建立 KECA 监测模型。模型表达式为

$$\phi_{eca} = D_i^{1/2} E_i^T, \quad (15)$$

其中: $D = \text{diag}\{\lambda_1, \dots, \lambda_N\}$, $E = \{e_1, \dots, e_N\}$ 。

3) 计算样本的 T^2 , SPE 统计量及样本控制限。

4.2 基于扩展核熵负载矩阵的发酵过程在线监测 (On-line monitoring of fermentation process based on extended kernel entropy load matrix)

1) 获取新采样时刻的数据 $X_k(I \times J)$, $k = 1, 2, \dots, K$, 用离线建模所得的均值、方差进行标准化处理。

2) 由采样时刻判断当前所处阶段, 选择相应的监控模型, 计算此时此刻的 T^2 , SPE 统计量控制限。计算公式如下:

$$T_k^2 = t_{\text{new},k} S_C^{-1} t_{\text{new},k} \sim \frac{R(N^2 - 1)}{N(N - R)} F_{R, N-R, \alpha}, \quad (16)$$

$$\text{SPE}_k = e_{\text{new},k} e_{\text{new},k}^T \sim g_k \chi_{k,k,\alpha}^2, \quad (17)$$

$$e_{\text{new},k} = x_{\text{new},k} (I - P_k P_k^T), \quad (18)$$

$$g_k = \frac{V_k}{2m_k}, \quad h_k = \frac{2m_k^2}{V_k}, \quad (19)$$

其中: $t_{\text{new},k}$ 为新时刻 k 的得分向量, $S_C^{-1}(R \times R)$ 为 k 时刻 T_k 的协方差阵的逆阵, N 为阶段建模的样本点数目, α 为置信限, m_k 和 v_k 分别为建模时所得 SPE 的均值和方差。

3) 计算此时此刻的 T^2 , SPE 统计量, 若未超出控制限, 则说明生产过程正常; 否则, 则说明生产过程发生故障。

5 实验研究及仿真分析 (Experimental research and simulation analysis)

5.1 青霉素发酵仿真分析 (Simulation analysis of penicillin fermentation process.)

青霉素发酵过程是一种典型的多阶段生产过程。本文基于 Pensim 仿真平台^[20]对青霉素的发酵过程进行了实验仿真, 设定每批次发酵时间为 400 h, 采样时间间隔为 1 h, 10 个典型过程变量如表 1 所示。为了更接近实际生产状况, 所有的过程变量均加入了白噪声。本文选取 Pensim 产生的 30 个正常批次数据作为样本数据, 得到三维数据矩阵 $X(30 \times 10 \times 400)$ 。

表 1 建立模型所用变量

Table 1 Variables used in the monitoring of the benchmark model

变量编号	变量名称
X1	通风速率/(L·h ⁻¹)
X2	搅拌功率/W
X3	底物流加速率/(L·h ⁻¹)
X4	补料温度/K
X5	溶解氧浓度/(mole·L ⁻¹)
X6	排气二氧化碳浓度/(mole·L ⁻¹)
X7	PH值
X8	反应温度/K
X9	反应热/Cal
X10	冷水流加速率/(L·h ⁻¹)

采用本文所提方法, 整个生产过程依据 KECA 分析所得主元个数首先被分为明显的 3 个阶段, 分别为 (1 ~ 58) h, (59 ~ 201) h, (202 ~ 400) h。如图 3 所示。

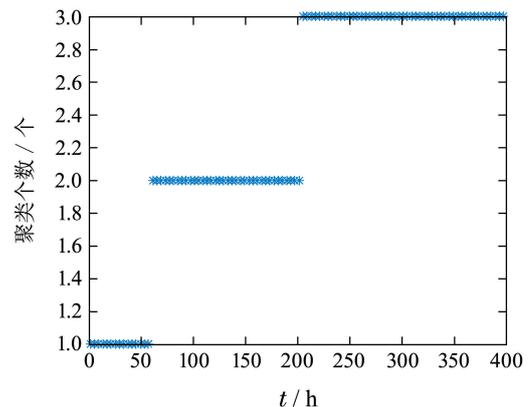


图 3 阶段划分结果

Fig. 3 Stage division result

将时间拓展到核熵负载矩阵当中, 计算扩展核熵负载矩阵间的相似度并以此为输入, 用模糊 C-均值方法进行再次划分, 最终划分为 (1~53) h, (79~185) h, (206~400) h 三个稳定阶段和 (54~78) h, (186~205) h 两个过渡阶段。如图 4 所示。

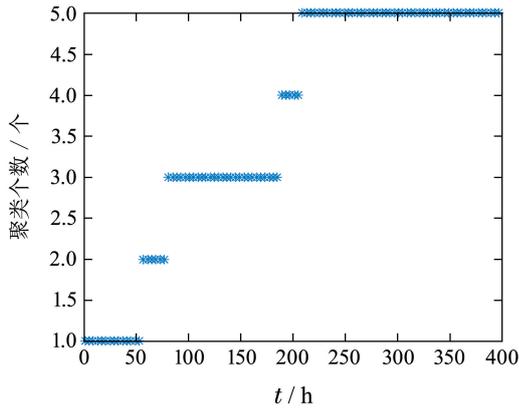


图4 阶段划分结果

Fig. 4 Stage division result

为验证所提方法在线监测时的有效性, 本文选取了3组测试样本, 一组为正常样本, 其余两组为故障样本. 如表2所示.

表2 故障设置情况

Table 2 Fault settings

故障种类	故障变量	故障类型	幅值/%	持续时间
1	通风速率	阶跃故障	10	200~400
2	底物流加速率	斜坡故障	10	200~400

选取多批正常和故障过程数据作为仿真训练数据, 将误报率和漏报率作为衡量本文方法有效性的指标. 由于篇幅限制, 此处只给出两种故障批次的监控图. 图5-7分别为KECA方法、多阶段MKPCA方法与本文所提方法对故障批次1的在线监测图, 此处控制限选取99%, 具体计算方法参考相关文献. 从图5和图6中我们可以看出在发酵过程的开始时刻 T^2 统计量均出现了误报警, 平均误报率大约13.25%. 从图7中可以看出整个生产过程不存在误报、漏报现象, 说明本文方法建立的模型精度更高些, 更贴近实际生产过程. 表3为三种方法在阶跃故障下误报与漏报率的对比结果.

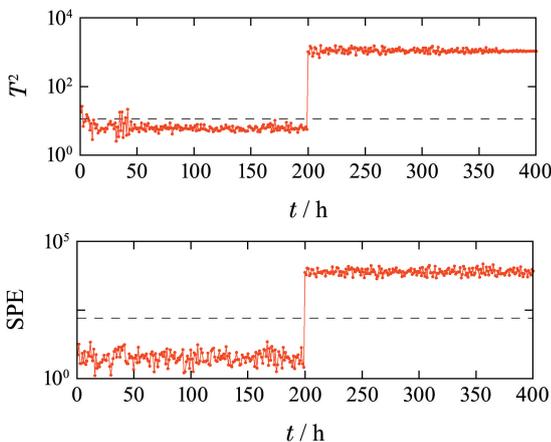


图5 MKECA 方法对故障批次1监测图

Fig. 5 The diagram of false batch 1 using the MKECA

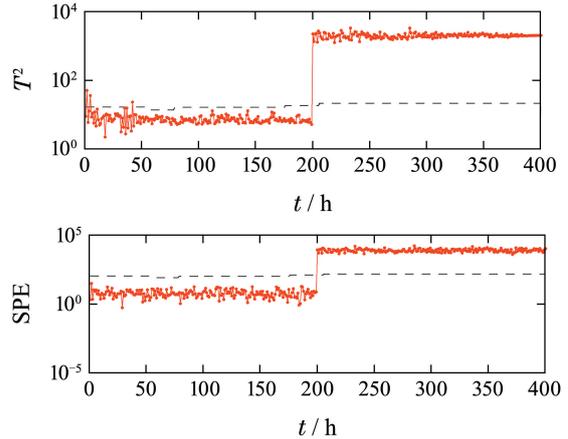


图6 多阶段MKPCA对故障批次1监测图

Fig. 6 The diagram of false batch 1 using the mul-stage MKPCA

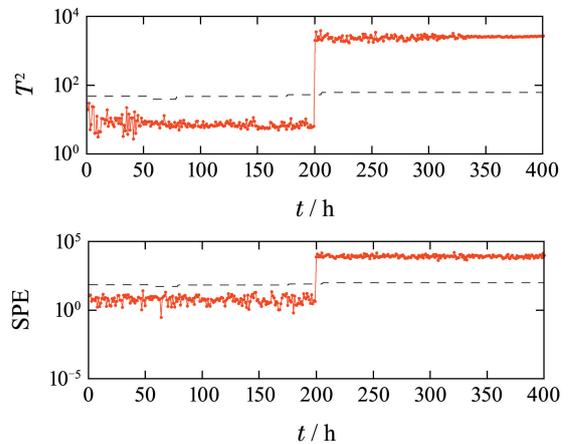


图7 本文方法对故障批次1的监测结果

Fig. 7 The diagram of false batch 1 using the proposed algorithm

表3 KECA方法、传统MKPCA与本文方法的误报率与漏报率

Table 3 Leaking alarm rate and False alarm rate of KECA, traditional MKPCA and the proposed algorithm

监控方法	T^2		SPE	
	漏报	误报	漏报	误报
MKECA方法	1.6%	14.7%	0.6%	11.3%
多阶段MKPCA	1.3%	10.9%	0	12.5%
本文方法	0	3.2%	0	5.2%

图8-10分别为 KECA 方法、多阶段 MKPCA 与本文方法对故障批次2的监测结果图. 从图8中可以看出在生产过程前期 T^2 , SPE均存在误报警现象, 都出现了跳变点控制限异常, 这和阶段划分建立局部模型有密不可分的关系. 与此相比, 多阶段MKPCA对整个生产过程进行阶段划分, 以多模型代替整体模型的方

法表现出了一定的优越性, 从图9中可以看到SPE统计量0误报, 很大程度上也消除了跳变点统计量异常的情况, 但 T^2 统计量前期仍有误报存在, 且对200时刻加入的10%斜坡故障在227 h才监测到, 灵敏度不高, 存在很大的延迟. 而本文方法表现出了极大优势, 由于离线建模时不单以方差贡献率大小来选取主元, 更多考虑了原始生产数据的结构信息保证了数据特征的相对完整, 所以聚类结果更为可靠. 将时间变量作为约束信息之一添加到负载矩阵当中, 以此建立的局部模型更加精确. 因此, 从监测图中可以看到本文所提方法在保证误报率、漏报率明显降低的基础上, 故障检测时刻也由227 h提前到了215 h, 提高了过程的监测性能.

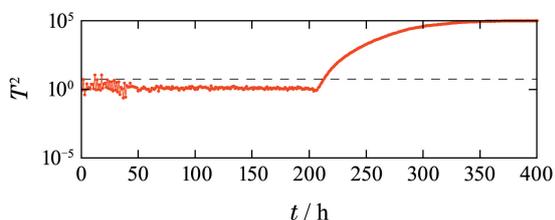


图 8 MKECA 方法对故障批次2监测图

Fig. 8 The diagram of false batch 2 using the MKECA

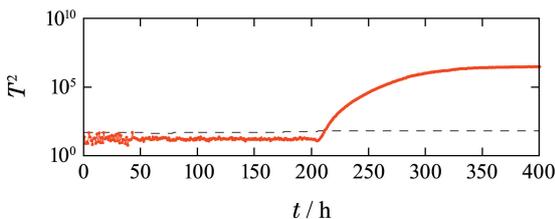


图 9 多阶段MKPCA对故障批次2监测图

Fig. 9 The diagram of false batch 2 using the mul-stage MKPCA

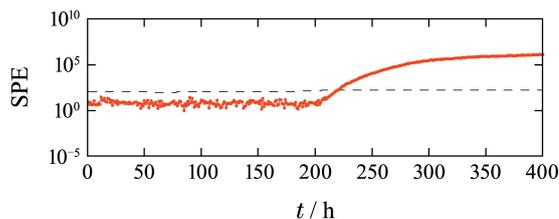
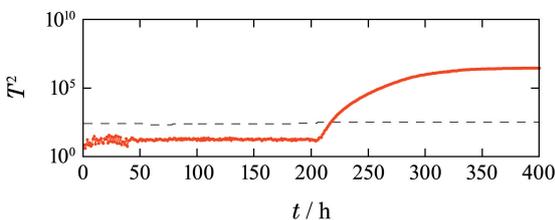


图 10 本文方法对故障批次2的监测结果

Fig. 10 The diagram of false batch 2 using the proposed algorithm

5.2 大肠杆菌发酵现场数据实验分析 (Experimental analysis of fermentation field data of *Escherichia*)

生物制药中利用重组大肠杆菌生产白介素-2的发酵过程是一个典型的间歇过程, 同样具有多阶段性等特点. 发酵周期大约6~7h.

选取北京某生物制药厂现场实际生产数据30批次, 采用以下7个变量作为本次实验的控制变量, 如表4所示.

表 4 大肠杆菌发酵过程变量

Table 4 Variables of *E. coli* fermentation process

变量编号	变量
X1	通气量/(L · min ⁻¹)
X2	温度/°C
X3	搅拌转速/(r · min ⁻¹)
X4	罐压/Bar
X5	溶解氧浓度/%
X6	罐内/P _H
X7	大肠杆菌浓度/(g · L ⁻¹)

现场的发酵时长为6.2 h, 采样时刻为5 min, 每批次75个采样时刻. 同仿真实验, 对选取的30个批次正常过程数据 $X(30 \times 7 \times 75)$ 进行离线建模, 正常批次数据和故障批次数据进行在线监测. 其中, 故障批次为 X_1 变量在40时刻时引入10%的阶跃故障, 直至反应结束.

图11为本文方法对大肠杆菌生产白介素-2的阶段划分结果, 5个阶段分别为稳定阶段(1~28) h、过渡阶段(29~36) h、稳定阶段(37~45) h、过渡阶段(46~58) h、稳定阶段(59~75) h. 该划分结果与实际发酵过程生产阶段存在一定偏差, 主要反应的是数据特征.

图12-13分别为多阶段MKPCA方法与本文所提方法对重组大肠杆菌生产白介素-2的正常批次监测结果图. 从图12中可以看到多阶段MKPCA存在严重的误报警, 这与实际发酵过程的环境密切相关. 更多未知噪声、变量间相关关系的突出表现, 都使得现场的实际监控结果比仿真结果要差一些. 本文考虑到生产过程的变化首先体现在采样时间上, 将时间扩展到

核熵负载矩阵当中,用负载矩阵和时间的综合变化代替负载矩阵的变化,一定程度上降低了未知噪声的影响,有效的减少了误报率.

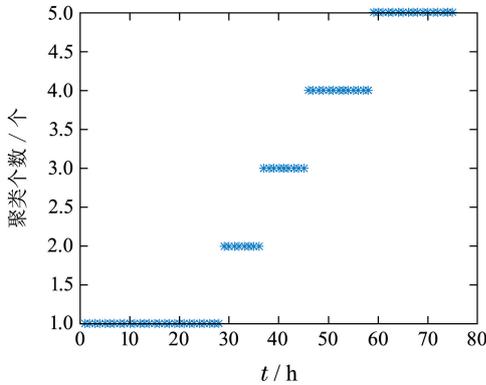


图 11 阶段划分结果

Fig. 11 Stage division result

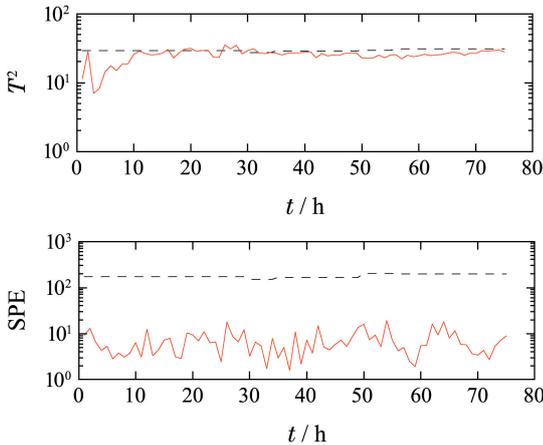


图 12 多阶段MKPCA 对正常批次监测图

Fig. 12 The diagram of normal batch using the mul-stage MKPCA algorithm

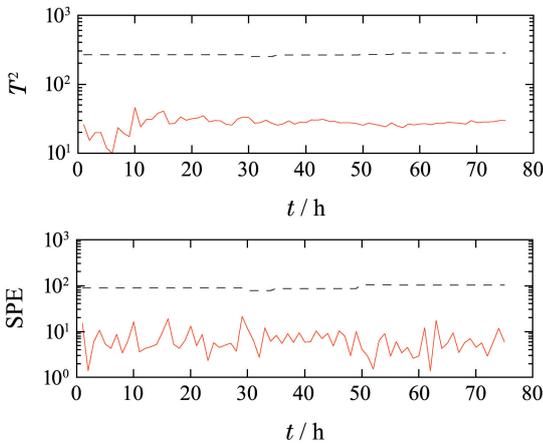


图 13 本文方法对正常批次监测图

Fig. 13 The diagram of normal batch using the proposed algorithm

图 14-15 为多阶段 MKPCA 方法与本文方法对故障批次的监测结果.从图 14 中可以看出,多阶段

MKPCA方法依然存在大量误报警,而本文所提方法没有出现误报的情况.

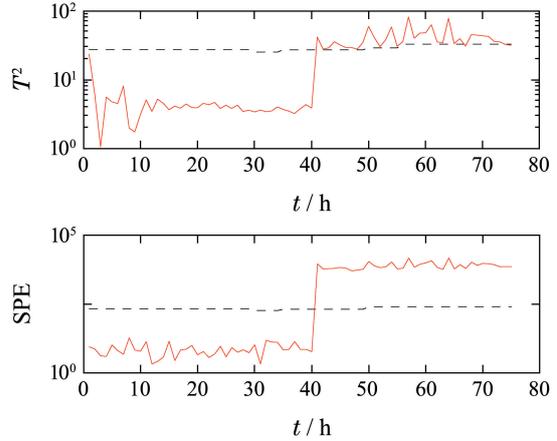


图 14 多阶段MKPCA对故障批次监测图

Fig. 14 The diagram of false batch using the mul-stage MKPCA

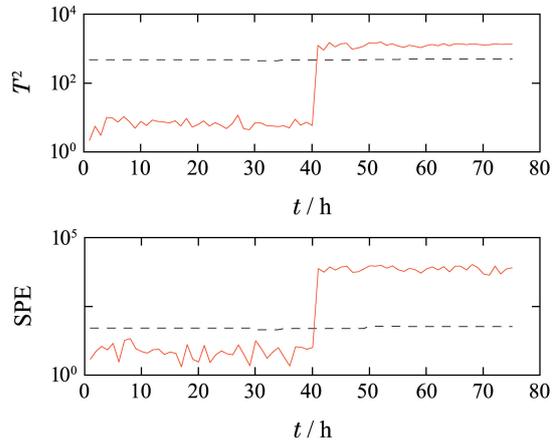


图 15 本文方法对故障批次监测图

Fig. 15 The diagram of false batch using the proposed algorithm

为进行进一步验证本文方法的有效性,对30个批次的过程数据分别用多阶段MKPCA和本文方法两种方法进行监测,最终将所有批次的漏报率、误报率进行加和并求取平均值,监测统计结果如表5所示.

表 5 过程监测统计结果

Table 5 Statistical results of process monitoring

监控方法	生产批次	T^2		SPE	
		漏报	误报/%	漏报	误报/%
多阶段KPCA	正常	0	11.6	0	5.8
	故障	9.2%	8.7	0	4.3
本文方法	正常	0	4.1	0	2.6
	故障	0	3.1	0	1.8

6 结论(Conclusions)

针对发酵过程多阶段的特性,本文提出一种基于扩展核熵负载矩阵的阶段划分策略,通过将时间添加

到核熵负载矩阵当中, 以扩展负载矩阵的变化来近似描述生产过程中的变化实现阶段划分, 在划分的阶段内建立监测模型进行故障监测, 同时以误报率、漏报率、监测到故障的时间作为性能评价标准。在青霉素发酵过程仿真数据和大肠杆菌生产白介素-2现场实验数据上分别进行了实验验证。实验结果表明, 本文所提方法能够显著降低生产过程的误报和漏报率, 在实际工业生产中有较强的应用价值。

参考文献(References):

- [1] JACKSON J E. *A User's Guide to Principal Components* [M]. New York: Wiley, 1991.
- [2] GE Z Q, SONG Z H, ZHAO L P, et al. Two-level PLS model for quality prediction of multiphase batch processes [J]. *Chemometrics & Intelligent Laboratory Systems*, 2014, 130(2): 29–36.
- [3] NOMIKOS P, MACGREGOR J F. Monitoring batch processes using multiway principal component analysis [J]. *Aiche Journal*, 1994, 40(8): 1361–1375.
- [4] NOMIKOS P, MACGREGOR J F. Multi-way partial least squares in monitoring batch processes [J]. *Chemometrics & Intelligent Laboratory Systems*, 1995, 30(1): 97–108.
- [5] ZHANG Ziyi, HU Yi, SHI Hongbo. Multi-stage batch process monitoring based on a clustering method [J]. *Journal of Chemical Industry*, 2013, 64(12): 4522–4528.
(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.)
- [6] TANG X C, LI Y, XIE Z. Phase division and process monitoring for multiphase batch process with transitions [J]. *Chemometrics & Intelligent Laboratory Systems*, 2015, 145: 72–83.
- [7] 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.
- [8] ZHAO C H, WANG F L. Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes [J]. *Journal of Process Control*, 2007, 17(9): 728–741.
- [9] QI Yongsheng, WANG Pu, GAO Xuejin. Fault detection and diagnosis of multiphase batch process based on kernel principal component analysis-principal component analysis [J]. *Control Theory & Applications*, 2012, 29(6): 69–79.
(齐咏生, 王普, 高学金. 基于核主元分析-主元分析的多阶段间歇过程故障监测与诊断 [J]. 控制理论与应用, 2012, 29(6): 69–79.)
- [10] CHANG Peng. Multi-stage separation and fault monitoring of microbial fermentation processes based on multi-way kernel entropy component analysis [J]. *Journal of Chemical Engineering of Chinese Universities*, 2015, 29(3): 650–656.
(常鹏. 基于多向核熵成分分析的微生物发酵过程多阶段划分及故障监测 [J]. 高校化学工程学报, 2015, 29(3): 650–656.)
- [11] JENSSEN R. Kernel entropy component analysis [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2010, 32(5): 847–860.
- [12] ZHAO C H, GAO F R, WANG F L. Nonlinear batch process monitoring using phase-based kernel independent component analysis-principal component analysis (KICA-PCA) [J]. *Industrial & Engineering Chemistry Research*, 2009, 48(20): 9163–9174.
- [13] 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.
- [14] XIA Lisha, YANG Yuying, FANG Huajing. Fault diagnosis performance improvement for chemical process based on Easy Ensemble method [J]. *Control Theory & Applications*, 2017, 34(1): 49–53.
夏丽莎, 杨玉英, 方华京. 基于 Easy Ensemble 的化工过程故障诊断性能改进 [J]. 控制理论与应用, 2017, 34(1): 49–53.)
- [15] ROTHWELL S R, MARTIN E B, MORRIS A J. Comparison of methods for dealing with uneven length batches [C] // *Proceedings of the 7th International Conference on Computer Applications in Biotechnology*. Osaka, Japan: Springer, 1998: 387–392.
- [16] KOURTI T. Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions [J]. *Journal of Chemometrics*, 2003, 17(1): 93–109.
- [17] JENSSEN R. Entropy-relevant dimensions in the kernel feature space: cluster-capturing dimensionality reduction [J]. *IEEE Signal Processing Magazine*, 2013, 30(4): 30–39.
- [18] LEE J M, YOO C K, LEE I B. Statistical process monitoring with independent component analysis [J]. *Journal of Process Control*, 2004, 14(5): 467–458.
- [19] WESTERHUIS J A, KOURTI T, MAC GREGOR J F. Comparing alternative approaches for multivariate statistical analysis of batch process data [J]. *Journal of Chemometrics*, 1999, 13(4): 397–413.
- [20] BIROL G, UNDEY C, CINAR A. A modular simulation package for fed-batch fermentation [J]. *Computers & Chemical Engineering*, 2002, 26(11): 1553–1565.

作者简介:

高学金 (1973–), 男, 副教授, 硕士生导师, 主要研究方向为复杂系统建模与控制、工业过程监控与故障诊断等, E-mail: gaouxuejin@bjut.edu.cn;

杨彦霞 (1990–), 女, 硕士研究生, 主要研究方向为多元统计过程控制与故障诊断、优化控制技术, E-mail: 18810299033@163.com;

王普 (1962–), 男, 教授, 博士生导师, 主要研究方向为工业过程控制与优化、复杂系统控制和计算机控制系统, E-mail: wangpu@163.com;

李晓理 (1971–), 男, 教授, 博士生导师, 主要研究方向为复杂流程工业建模与多模型控制, E-mail: lixiaolibjut@bjut.edu.cn;

常鹏 (1981–), 男, 讲师, 主要研究方向为间歇过程统计建模及故障诊断, E-mail: changpeng@bjut.edu.cn;

齐咏生 (1975–), 男, 教授, 硕士生导师, 研究方向为复杂工业过程建模与优化控制、故障检测与诊断, E-mail: qyslyt@163.com.