

基于高阶累积统计量的微生物发酵过程监测

常 鹏[†], 王 普, 高学金

(1. 北京工业大学 信息学部, 北京 100124; 2. 北京工业大学 教育部工程中心数字社区, 北京 100124;
3. 北京工业大学 计算智能与智能系统北京市重点实验室, 北京 100124)

摘 要: 传统多向核独立成分分析(MKICA)方法的实质是把基于独立成分分析(ICA)中的白化处理主元分析(PCA)替换为核主元分析(KPCA)后利用二阶统计量进行过程监控,并未利用过程数据的阶段特性和高阶累积量信息,为了解决此问题,提出高阶累积量分析(HCA)与多向核熵独立成份分析(MKECA)相结合的多向高阶累积量的核熵独立成分分析方法(HCA-MKEICA). 首先,采用核熵独立成份分析(KECA)对原始数据进行数据转换,解决数据的非线性;然后,在高维核熵空间利用HCA技术构建新的统计量用于过程监控;最后,将该方法应用于青霉素仿真平台和实际的工业过程并与MKICA方法进行对比,以验证所提出方法的有效性.

关键词: 间歇过程监控; MKICA; 高阶累积量; MKECA; MKEICA

中图分类号: TP273

文献标志码: A

Microbial fermentation process monitoring based on higher order cumulant statistics

CHANG Peng[†], WANG Pu, GAO Xue-jin

(1. Department of Information, Beijing University of Technology, Beijing 100124, China; 2. Ministry of Education Engineering Center Digital Community, Beijing University of Technology, Beijing 100124, China; 3. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China)

Abstract: The essence of the traditional multiway kernel independent component analysis(MKICA) method is that the independent component analysis(ICA) whitened principal component analysis(PCA) is replaced with KPCA by using second order statistics of the monitoring and controlling process, not by the stage characteristic of process data and higher-order cumulant information. To solve this problem, the high order cumulant analysis(HCA) and multiway kernel entropy independent component analysis(MKECA) are combined, and the analysis of high order cumulant multiway kernel entropy independent component analysis(HCA-MKEICA) method is proposed. Firstly, the kernel entropy independent component analysis(KECA) method is used for original data conversion to solve the problem of nonlinear. Then, in the high-dimensional kernel entropy space, the HCA technology is used to construct the new statistics for process monitoring. Finally, the proposed method is applied to the microbial fermentation process, and the comparison results with the traditional methods show that the proposed method can achieve a better detection, and verify its effectiveness.

Keywords: monitoring batch process; MKICA; hige-order cumulant anlaysis; MKECA; MKEICA

0 引 言

近些年来,基于独立成分分析(ICA)^[1-2]的过程监测方法逐渐发展起来. ICA^[2]最早来自盲源信号分离技术,它能够有效处理数据所具有的非高斯性问题. 为保证提取数据特征的完整性,使用高阶统计信息对数据进行特征提取,并保证所提取的成分信息之间相互独立,而PCA^[3-4]由于仅仅利用过程数据(均值

和方差)的低阶统计信息提取互不相关的变量之间的成分信息,所提取的成分之间是正交的但在统计上并不一定独立,并且低阶统计量在面对数据具有非高斯性等情况时,将不能提取数据的特征,即ICA提取的数据信息比PCA提取的数据信息更加全面,ICA提出的独立成分的统计意义强于PCA提取的主成分统计意义. 因此,在处理非高斯问题上,PCA将无能为力

收稿日期: 2016-10-08; 修回日期: 2017-02-27.

基金项目: 国家自然科学基金项目(61640312, 61364009, 61174109); 北京市自然科学基金项目(4172007).

作者简介: 常鹏(1981—),男,讲师,博士,从事复杂工业统计建模及故障诊断的研究;王普(1962—),男,教授,博士生导师,从事智能控制与智能信息处理等研究.

[†]通讯作者. E-mail: changpeng@bjut.edu.cn

力,ICA成为必然选择.但是,ICA是线性化方法,在进行过程数据的信息提取时会忽略过程数据中存在的非线性信息,而间歇过程的非线性又是其固有特性,为此ICA将不能完成提取间歇过程的特征.为解决此问题,文献[5-7]提出了核ICA方法,用以解决间歇过程的非线性、非高斯等问题.该算法的核心步骤包括:1)将原始数据用KPCA进行白化处理,解决数据的非线性,得到不相关的得分矩阵;2)将得分矩阵进行ICA分解.但是,以上KICA方法在进行过程监测时,所使用的监测统计量 I^2 和SPE是低阶统计量实质监测的均值和方差,不能监测高阶统计量的信息,即KICA是低阶统计量方法用于过程的数据提取,但是监测的是低阶统计量信息^[8],因此其不能利用数据的高阶统计量的信息.

为解决上述问题,研究一种基于KECA^[9-11]白化的多向核熵独立元分析(MKEICA)方法.该方法首先用批次与变量相结合的展开(AT)方法将生产过程的三维数据展开为二维数据,之后建立ICA模型用于过程特征的提取并构造监控统计量,实现过程监测.采用本文研究方法建模时有以下两点优势:一是不要各个批次之间的数据长度等长,并且用于在线应用监测时,无需对新批次的未来时刻测量值进行预估填充;二是在基于KECA方法作白化处理的基础上,建立MKEICA过程监测模型,专门针对实际工业过程的非线性、非高斯性共存的情况,更加接近实际间歇工业生产的情况,可以更加有效地对间歇生产过程进行监测,指导生产过程,避免生产故障的发生,进而保障生产过程的安全运行.最后将MKEICA、MKICA^[5-6]方法应用在青霉素仿真平台和实际的工业过程,对监控结果进行对比,从而验证本文方法的优越性.

1 基于HCA-MKEICA的过程监测策略

1.1 三阶累积量分析

三阶累积量分析(TCA)技术在数字信号处理领域得到了广泛应用^[8,12-14],三阶累积量的重要性质就是高斯过程的三阶累积量为零,对于非高斯分布的数据,三阶统计量包含着重要数据信息,可有效提取非高斯数据的数据信息.TCA已经广泛用于具有非高斯、非线性的问题图像识别领域^[12-14].假设任意一个零均值的高斯随机变量的二阶矩与二阶累积量相同并且均等于方差 σ^2 ,或者其奇数阶的矩恒为零但偶数阶的矩不为零;而三阶的矩及以上各阶矩的累积量恒等于零,即高阶累积量对高斯随机信号是无效的、屏蔽的.对于一个零均值变量 $y(i)$,其三阶累积量

定义如下^[8]:

$$\text{cum}_{3,y}(\tau_1, \tau_2) = E\{y(i)y(i+\tau_1)y(i+\tau_2)\}. \quad (1)$$

其中: E 表示表示期望; τ_1, τ_2 是滞后时间三阶累积量的样本估计,可由下式求得:

$$\text{cum}_{3,y}(\tau_1, \tau_2) = \frac{1}{N} \sum_{i=1}^N y(i)y(i+\tau_1)y(i+\tau_2). \quad (2)$$

1.2 离线建模

核方法^[5-11,15-16]的本质是利用核函数代替复杂的点积运算,进而将非线性的原始过程数据映射到高维核空间使其线性可分,从而把低维非线性问题转换为高维线性问题.对生产过程采集到的三维数据矩阵 X 进行预处理后,通过一个非线性函数 $\Phi(\cdot)$ 映射到高维特征空间中使其线性可分,具体形式如下:

$$\Phi=(\Phi(X),\Phi(\cdot)):R^{n \times m} \rightarrow F^{n \times n}, \quad (3)$$

则在特征空间 F 中 Φ 的协方差矩阵如下:

$$\sum^F = \frac{1}{n} \Phi \Phi^T. \quad (4)$$

通常情况下,非线性函数 $\Phi(\cdot)$ 的具体形式是未知的,可通过核变换来代替,具体形式如下:

$$K = \Phi \Phi^T. \quad (5)$$

将式(3)代入(4)可得 $\sum^F = \frac{1}{n} K$,举例如下式所示:

$$k_{ij} = \Phi(x_i)\Phi(x_j)^T = k(x_i, x_j). \quad (6)$$

其中: k_{ij} 是核矩阵 K 中的第 i 行、第 j 列的元素; x_i 和 x_j 分别是 X 的第 i 行和第 j 行; $k(\cdot, \cdot)$ 是一个核函数,通常使用高斯核函数^[7-11,15-16],如下式所示:

$$k(x, y) = \exp[-\|x - y\|^2/2\sigma^2]. \quad (7)$$

为了保证映射得到的 Φ 是标准化的 $\bar{\Phi}$,需要对核矩阵 K 进行标准化,即中心化-方差归一化处理,这与线性PCA对采样数据的标准化意义相同.首先,进行如下中心化处理 $\bar{K} = K - L_N K - K L_N + L_N K L_N$,其中 L_N 是系数为 $1/N$ 的 $N \times N$ 阶单位矩阵.之后,进行方差归一化处理,即

$$\bar{K}_{\text{scl}} = \bar{K}/\text{trace}(\bar{K})/(N-1), \quad (8)$$

其中 $\text{trace}(\cdot)$ 表示矩阵的迹.用 \bar{K}_{scl} 代替原有的 K 求解特征值和特征向量即可,具体如下式所示:

$$\lambda v = \bar{K} v. \quad (9)$$

可以获得对应于 \bar{K} 的 u 个较大特征值 $\lambda_1, \lambda_2, \dots, \lambda_u$ 和特征向量 v_1, v_2, \dots, v_u .本文采用相对特征值(RE)来确定 u 的个数,RE准则如下式所示:

$$\lambda_i/\text{sum}(\lambda) \geq 0.0001. \quad (10)$$

本文采用KECA白化矩阵,详见文献[10-11],这

里利用核熵值的累积贡献率进行核熵个数的选择, 累积核熵贡献率定义如下:

$$\sum_{i=1}^d \hat{V}(p) / \sum_{i=1}^n \hat{V}(p)_i \times 100\% \geq 85\%. \quad (11)$$

其中: $V = [v_1, v_2, \dots, v_u]$, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_u)$. PCA 的负载矩阵和白化矩阵分别为

$$H = (h_1, h_2, \dots, h_u) = \bar{\Phi}V\Lambda^{-1/2}, \quad (12)$$

$$Q = H\left(\frac{1}{n}\Lambda\right)^{-1/2} = \sqrt{n}\bar{\Phi}V\Lambda^{-1}\bar{k}^T. \quad (13)$$

KECA 得分矩阵为

$$\begin{aligned} z &= H^T\bar{\phi} = \Lambda^{-1/2}V^T\bar{\phi}^T\bar{\phi} = \\ &\Lambda^{-1/2}V^T[\bar{k}(x-1, x), \bar{k}(x_1, x), \dots, \bar{k}(x_n, x)]^T = \\ &\Lambda^{-1/2}V^T\bar{k}^T. \end{aligned} \quad (14)$$

其中: $\bar{\phi}$ 是 $\bar{\Phi}$ 的一列, \bar{k} 是 \bar{K} 的一行. 白化得分如下式所示:

$$\bar{z} = Q^T\bar{\phi} = \sqrt{n}\Lambda^{-1}V^T\bar{k}^T. \quad (15)$$

实际上 \bar{z} 是 z 的方差归一化结果, 即

$$z = \left(\frac{1}{n}\Lambda\right)^{1/2}\bar{z}. \quad (16)$$

对于一个新的待监测测量向量 x_{new} , 其对应的核熵向量为

$$\begin{aligned} K_{\text{new}} &= \\ &[k(x_1, x_{\text{new}}), k(x_2, x_{\text{new}}), \dots, k(x_n, x_{\text{new}})]. \end{aligned} \quad (17)$$

需要进行中心化和去量纲化, 即

$$\bar{K}_{\text{new}} = K_{\text{new}} - I_{\text{new}}K - K_{\text{new}}I_{\text{new}} + I_{\text{new}}KI_n, \quad (18)$$

$$\bar{K}_{\text{new}} = \frac{\bar{K}_{\text{new}}}{\text{trace}(\bar{K}_{\text{new}})/(N-1)}, \quad (19)$$

其中 I_{new} 为 $n \times n$ 的单位阵, 则新的 KECA 得分为

$$z_{\text{new}} = \Lambda^{-1/2}V^T\bar{K}_{\text{new}}^T, \quad (20)$$

新的白化得分为

$$\bar{z}_{\text{new}} = \sqrt{n}\Lambda^{-1}V^T\bar{K}_{\text{new}}^T. \quad (21)$$

对 \bar{z} 采用 ICA 算法, 即

$$s = B^T\bar{z}. \quad (22)$$

根据 s 的非高斯性大小 (可通过峰度系数表示)^[5-7] 选择主导独立成分 s_d , 与其对应的 B 矩阵的列组成矩阵 B_d , $s_d = B_d^T\bar{z}$. 定义新的监测统计量 HS 和 HE, 对于非高斯过程, 使用 ICA 提取独立成分, 将整个过程分为主导独立成分和模型预测误差两部分, 在采样 i 处, 第 p 个主导独立成分 s_d 的样本三阶累积量为

$$\text{hs}_d(i) = s_d(i)s_d(i-1)s_d(i-2) =$$

$$w_p\bar{K}(i)w_p\bar{K}(i-1)w_p\bar{K}(i-2). \quad (23)$$

其中: w_p 是解混矩阵 W_d 的第 p 行, $p = 1, 2, \dots, d$. 为了监测全部主导独立成分的三阶累积量, HCA 的第一个监测指标定义为

$$\begin{aligned} \text{HS}(i) &= \\ &\sum_{p=1}^d |hs_p(i)| \cdot [s_d(i)s_d(i-1)s_d(i-2)]^T \cdot \\ &[s_d(i)s_d(i-1)s_d(i-2)] = \\ &[w_pK(i)w_pK(i-1)w_pK(i-2)]^T \cdot \\ &[w_pK(i)w_pK(i-1)w_pK(i-2)]. \end{aligned} \quad (24)$$

在采样 i 处, 非高斯模型对第 q 个变量的预测误差的样本三阶累积量如下:

$$\begin{aligned} \text{he}_q(i) &= e_q(i)e_q(i-1)e_q(i-2) = \\ &l_qK(i)l_qK(i-1)l_qK(i-2). \end{aligned} \quad (25)$$

其中: l_q 是 L 的第 q 行, $q = 1, 2, \dots, m$. 为了监测所有预测误差的三阶累积量, HCA 的另一个监测指标定义如下:

$$\begin{aligned} \text{HE}(i) &= \\ &\sum_{q=1}^m |\text{he}_q(i)| \cdot [e_q(i)e_q(i-1)e_q(i-2)]^T \cdot \\ &[e_q(i)e_q(i-1)e_q(i-2)] = \\ &[l_qK(i)l_qK(i-1)l_qK(i-2)]^T \cdot \\ &[l_qK(i)l_qK(i-1)l_qK(i-2)]. \end{aligned} \quad (26)$$

HS 和 HE 监测统计量的控制限由核密度估计得出^[5-7,10-11].

1.3 在线建模

当在线监测时, 新时刻数据的监测统计量定义为

$$\begin{aligned} \text{hs}_{\text{new}}(i) &= s_{\text{new}}(i)s_{\text{new}}(i-1)s_{\text{new}}(i-2) = \\ &w_p\bar{K}_{\text{new}}(i)w_p\bar{K}_{\text{new}}(i-1)w_p\bar{K}_{\text{new}}(i-2), \end{aligned} \quad (27)$$

$$\begin{aligned} \text{HS}_{\text{new}}(i) &= \\ &\sum_{p=1}^d |hs_{\text{new}p}(i)| \cdot [s_{\text{new},d}(i)s_{\text{new},d}(i-1)s_{\text{new},d}(i-2)]^T \cdot \\ &[s_{\text{new},d}(i)s_{\text{new},d}(i-1)s_{\text{new},d}(i-2)] = \\ &[w_{\text{new},p}K_{\text{new}}(i)w_{\text{new},p}K_{\text{new}}(i-1)w_{\text{new},p}K_{\text{new}}(i-2)]^T \cdot \\ &[w_{\text{new},p}K_{\text{new}}(i)w_{\text{new},p}K_{\text{new}}(i-1)w_{\text{new},p}K_{\text{new}}(i-2)], \end{aligned} \quad (28)$$

$$\begin{aligned} \text{he}_{\text{new}}(i) &= e_{\text{new}}(i)e_{\text{new}}(i-1)e_{\text{new}}(i-2) = \\ &l_qK(i)l_qK(i-1)l_qK(i-2), \end{aligned} \quad (29)$$

$$\begin{aligned}
 HE_{new}(i) = & \sum_{q=1}^m |he_{new,q}(i)| \cdot \\
 & [e_{new,q}(i)e_{new,q}(i-1)e_{new,q}(i-2)]^T \cdot \\
 & [e_{new,q}(i)e_{new,q}(i-1)e_{new,q}(i-2)] = \\
 & [l_{new,q}K_{new}(i)l_{new,q}K_{new}(i-1)l_{new,q}K_{new}(i-2)]^T \cdot \\
 & [l_{new,q}K_{new}(i)l_{new,q}K_{new}(i-1)l_{new,q}K_{new}(i-2)].
 \end{aligned}
 \tag{30}$$

基于MKEICA统计建模及在线监测实施算法如图1所示。

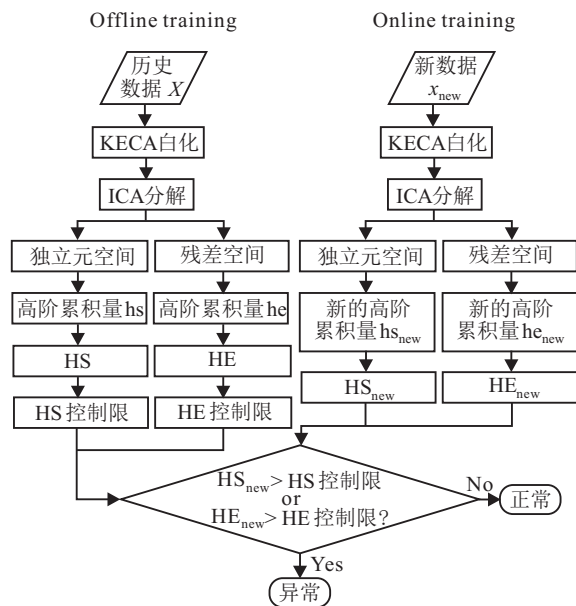


图1 过程监测流程

基于MKEICA统计建模及在线监测描述如下：该方法首先利用KECA代替传统KPCA作为MKICA数据的白化处理,使得白化后的数据矩阵可以更好地保持原始的数据结构,在白化得分矩阵里进行ICA分解,将核熵数据空间的数据分解为独立元子空间和残差子空间,在两个子空间内分别构建三阶累积监测统计量HS和HE,利用核密度估计两者的监测控制限用于过程的在线监测;将采集到新时刻的在线数据 x_{new} 进行标准化,将标准后的数据投影到核熵空间进行ICA分解为在线独立元空间和残差空间,在两个在线独立元空间和残差空间内计算监测统计量 HS_{new} 和 HE_{new} ,判断其是否超出监测控制限.如果新时刻的统计量没有超出监测控制限,则表明生产过程没有异常情况发生;如果新时刻的统计量任意一个或两个都超出监测控制限,则判断此时的生产出现异常,需要对其进行故障变量追溯.本文暂不考虑故障诊断的问题,只对生产过程是否出现异常进行监测.针对传统MKICA监测方法所构建的监测统计量

为二阶统计量这一缺陷^[8],提出三阶累积量的监测统计量用于过程监测,旨在克服传统统计量在进行故障监测时,存在较高误警和漏报的问题,用于改善故障监测的可靠性和灵敏度.

2 算法仿真应用对比

青霉素发酵过程是补料分批发酵过程,具有非线性和非高斯性的特点.这里实验的主要目的是验证下列观点:1)基于间歇过程的数据是非线性、非高斯性共存的,不是单一存在的;2)基于MKEICA的监测模型具备有效的故障监测能力.首先,基于仿真平台共产生40个长度不等的正常批次,建立初始模型参考数据库.本文将Pensim^[17]仿真平台产生的80批次的正常生产数据作为建模数据,每个批次的反应时间为400h,采样间隔为1h,根据工艺选取10个过程变量进行监测,如表1所示,故障类型如表2所示,监控结果如表3所示.

表1 建立模型所用变量

序号	变量名称	序号	变量名称
x_1	搅拌速率/(r/min)	x_7	碱流加速率/(g/L)
x_2	通风速率/(L/h)	x_8	酸流加速率/(g/L)
x_3	补料温度/K	x_9	温度/K
x_4	底物流加速率/(L/h)	x_{10}	冷水流速/(L/h)
x_5	溶氧浓度/(g/L)	x_{11}	菌体浓度/(g/L)
x_6	PH值	x_{12}	产物浓度/(g/L)

表2 仿真中用到的故障类型

序号	变量名称	故障类型	故障幅度	故障发生时间/h
1	搅拌速率	斜坡扰动	0.002	200-400
2	搅拌速率	斜坡扰动	0.0002	200-400
3	补料速率	斜坡扰动	0.003	50-400
4	通风速率	阶跃故障	0.005	200-250

表3 MKICA和MKEICA故障监测 h

故障类型	I^2 监测	HS 监测	SPE 监测	HE 监测
1	224	215	211	204
2	236	215	216	203
3	86	60	65	56
4	200	200	200	200

监测结果与讨论.首先引入表2中第1个故障,搅拌速率从200h直到反映结束,按照0.002的斜率增长,从表3的监测结果可知,MKICA的 I^2 和SPE两个监测统计量分别从224h和211h发现故障;MKEICA的HS和HE两个统计量分别从215h和204h发现故障.引入表2的第2个故障,从搅拌速率在200h时直到反映结束,按照0.0002的斜率增长,从表3的监测结果可以看出,MKICA的 I^2 统计量从236h超出控制限,SPE统计量在216h超出控制限;MKEICA的HS

和HE两个监测统计量分别在215h和203h超出控制限. 通过以上的监测结果对比发现, 这里的监测统计量 I^2 和SPE随着故障幅值的减小, 其发现故障的能力下降, 而HS和HE两个监测统计量并没有随着故障幅值的减小而降低捕获故障的能力. 这说明针对以上两个故障, 本文提出的MKEICA监测方法优于MKICA方法. 引入表2中的第3个故障, 通风速率从50h开始直到反映结束, 按照0.003的斜率增长, 从表3的监测结果可以看出, MKICA的 I^2 和SPE两个监测统计量分别从86h和65h超出控制限; MKEICA的HS和HE两个监测统计量分别在60h和56h超出控制限. 引入表2中第4个故障, 补料速率从在200h到250h时刻引入阶越变化, 增长幅值为0.005. 监测结果如表3所示, 故障一旦发生, MKICA和MKEICA的监测统计量指标均立刻超出控制限, 这是因为此类故障发生了大幅度突变, 两种监测模型在针对突变大幅值的故障时具有一样的故障识别能力. 综上所述, 在针对微小故障方面, MKEICA发现故障的能力优于传统MKICA方法; 在针对阶跃型监测方面, 其发现故障的能力等同于MKICA方法.

3 算法工业实际应用对比

实验的主要目的是验证: 1) 过程数据是非线性、非高斯性共存的, 不是单一存在的; 2) 基于MKEICA建立的监测模型具备有效的故障监测能力. 本节给出本文方法在北京亦庄某生物制药公司基因重组大肠杆菌外源蛋白表达制备白介素-2的发酵过程监测中的应用. 发酵过程采用Sartorius BIOSTAT BDL 15L发酵罐, 其中控制器通过蠕动泵调节补充培养液(葡萄糖、氨水、培养基)的速率, 并通过给定参数实现对通气量、搅拌转速、pH值、温度等的控制. 培养基包括酵母粉、无机盐等成分^[10-11]. 发酵过程的采样间隔为0.5h, 初始接种量为700ml. 选择6个主要过程变量综合表征菌体生长及外源蛋白表达的状况, 如表4所示.

表4 大肠杆菌发酵过程可检测变量

序号	变量
1	pH值
2	溶解氧浓度(DO)
3	搅拌转速/rpm
4	补葡萄糖量/ml
5	补培养基量/ml
6	通气量/(L·m-1)

取33个正常批次作为初始模型参考数据库, 得到不等长的三维数据矩阵 $X(33 \times 6 \times (38 - 40))$, 利用DTW^[12-13,18]技术使其变为等长 $X(33 \times 6 \times 39)$. 为

验证模型的有效性, 引入两种类型的故障如表5所示.

表5 仿真中的故障类型

序号	变量名称	故障类型	故障幅度	故障发生时间/h
1	搅拌速率	阶跃故障	0.1	15-39
2	补糖速率	斜坡扰动	0.05	15-39

监测结果与讨论. 针对表5中的故障1的监测结果如图2所示.

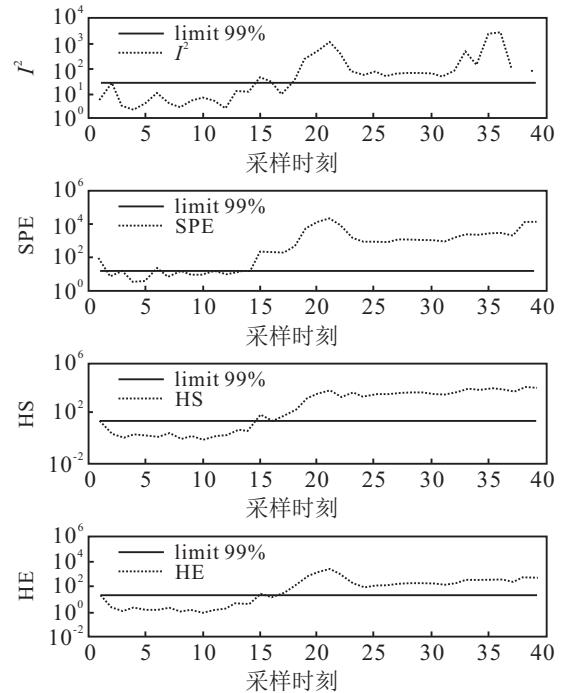


图2 针对故障1的监控结果对比

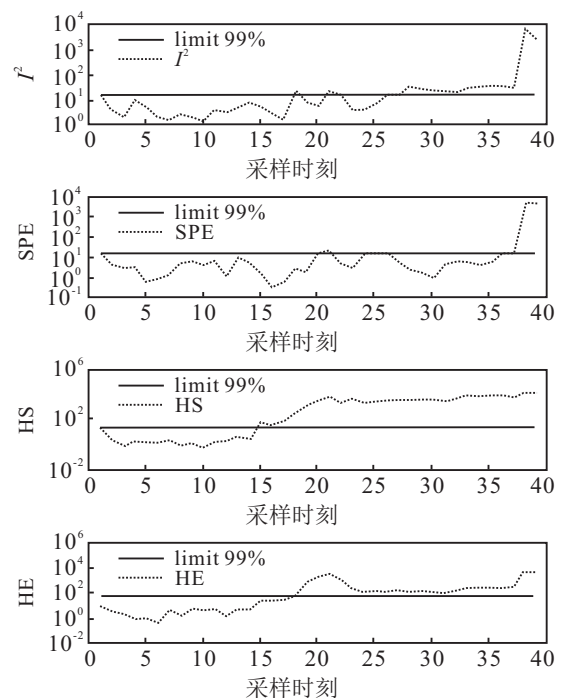


图3 针对故障2的监控结果对比

两种方法的监测统计量在故障发生的时刻第一时间都超出控制限, 但是从图2可以看出, I^2 监测图

存在故障的漏报警,漏报率为9%,SPE监测图存在故障的误报警,误报率为18%,而从图2可以看出,HS和HE监测图不存在故障的漏报警和误报警现象.针对表5中的故障2的监测结果如图3所示,MKICA的 I^2 监测统计量在27h才整体超出控制限,SPE监测统计量在36h附近超出监测控制限,基本失去了对此故障的识别能力.而MKEICA的HS和HE监测统计量体现了高效的故障监测能力,分别在16h和18h超出各自的控制限,并且不存在故障的误报警和漏报警.通过以上分析可知,MKEICA监测模型优于MKICA的监测模型.

4 结论

本文针对间歇过程的非线性和非高斯性问题,提出了基于MKEICA的过程监测方法.该方法首先利用KECA代替传统KPCA作为MKICA数据的白化处理,使得白化后的数据矩阵可以更好地保持原始的数据结构;其次针对传统MKICA监测方法所构建的监测统计量为二阶统计量的不足,提出了三阶累积量的监测统计量用于过程监测,旨在克服传统统计量在监测时存在较高误警和漏报的问题,改善了故障监测的可靠性和灵敏度.对青霉素仿真平台和工业发酵过程的应用表明,所提出的方法与传统MKICA方法相比,能有效减少系统出现的误警率和漏报率,为间歇发酵过程监测提供一种可行的解决方案,具有一定的实用价值.

参考文献(References)

- [1] Hyvarinen A. The fixed-point algorithm and maximum likelihood estimation for independent component analysis[J]. *Neural Processing Letters*, 1999, 10(1): 1-5.
- [2] Lee J M, Yoo C K, Lee I B. Statistical process monitoring with independent component analysis[J]. *J of Process Control*, 2004, 14(5): 467-485.
- [3] Nomikos P, MacGregor J F. Monitoring batch processes using multiway principal component analysis[J]. *J of American Institute of Chemical Engineers*, 1994, 40(8): 1361-1375.
- [4] Nomikos P, MacGregor J F. Multivariate SPC charts for monitoring batch process[J]. *Technometrics*, 1995, 37(1): 41-59.
- [5] Lee J M, Yoo C K, Lee I B. Fault detection of batch processes using multiway kernel principal component analysis[J]. *Computers and Chemical Engineering*, 2004, 28(9): 1837-1847.
- [6] Tian X, Zhang X, Deng X, et al. Multiway kernel independent component analysis based on feature samples for batch process monitoring[J]. *Neurocomputing*, 2009, 72(7): 1584-1596.
- [7] Cai L, Tian X, Zhang N. A kernel time structure independent component analysis method for nonlinear process monitoring[J]. *Chinese J of Chemical Engineering*, 2014, 22(11): 1243-1253.
- [8] Wang Y, Fan J, Yao Y. Online monitoring of multivariate processes using higher-order cumulants analysis[J]. *Industrial and Engineering Chemistry Research*, 2014, 53(11): 4328-4338.
- [9] Janssens R. Kernel entropy component analysis[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2009, 32(5): 847-860.
- [10] 常鹏,王普,高学金,等.基于核熵投影技术的多阶段间歇过程监测研究[J].*仪器仪表学报*, 2014, 35(7): 1654-1661.
(Chang P, Wang P, Gao X J. Research on batch process monitoring based on multi-stage kernel pattern entropy projection technology[J]. *Chinese J of Scientific Instrument*, 2014, 35(7): 1654-1661.)
- [11] 常鹏,王普,高学金.基于多向核熵成分分析的微生物发酵间歇过程监测研究[J].*高校化学工程学报*, 2015, 29(2): 395-399.
(Chang P, Wang P, Gao X J. Batch process monitoring for microbial fermentation based on multi-way kernel entropy component analysis[J]. *J of Chemical Engineering of Chinese Universities*, 2015, 29(2): 395-399.)
- [12] Giannakis G B, Mendel J M. Cumulant-based order determination of non-Gaussian ARMA models[J]. *IEEE Trans on Acoustics Speech and Signal Processing*, 1990, 38(8): 1411-1423.
- [13] Giannakis G B, Tsatsanis M K. Signal detection and classification using matched filtering and higher order statistics[J]. *IEEE Trans on Acoustics Speech and Signal Processing*, 1990, 38(7): 1284-1296.
- [14] Ling H, Zhiming H. Application of bispectral analysis in the nonlinear systems[J]. *Int Proc of Computer Science and Information Technology*, 2012, 46(1): 107-116.
- [15] Cho J H, Lee J M, Choi S W, et al. Fault identification for process monitoring using kernel principal component analysis[J]. *Chemical Engineering Science*, 2005, 60(1): 279-288.
- [16] Choi S W, Lee C, Lee J M, et al. Fault detection and identification of nonlinear processes based on kernel PCA[J]. *Chemometrics and Intelligent Laboratory Systems*, 2005, 75(1): 55-67.
- [17] Birol G, Undey C, Cinar A. A modular simulation package for fed-batch fermentation: Penicillin production[J]. *Computer and Chemical Engineering*, 2002, 26(11): 1553-1565.
- [18] Zhang Y, Zhang H. Fault detection for time-varying processes[J]. *IEEE Trans on Control Systems Technology*, 2014, 22(4): 1527-1535.

(责任编辑: 闫妍)