基于四阶累积分析的工业大肠杆菌制备过程故障诊断

常　鹏[1,2,3†], 乔俊飞[1,2], 张祥宇[1,3], 王　普[1,3]

(1. 北京工业大学 信息学部, 北京 100124; 2. 北京工业大学 计算智能和智能系统北京市重点实验室, 北京 100124;

3. 北京工业大学 数字社区教育部工程研究中心, 北京 100124)

摘要: 工业大肠杆菌制备过程具有非线性和非高斯性共存的特征, 导致难以对故障源进行有效定位, 针对这个问题, 提出一种基于多向核熵独立元分析(MKEICA)的过程监测方法; 同时针对传统低阶监控统计量(T^2, I^2和SPE)无法得到非高斯信息的不足提出了四阶累积监控统计量的方法; 其次通过对四阶累积监控量进行推导, 得到故障产生的原因. 最后将其应用在实际的工业过程并与多向核独立元分析(MKICA)监测模型进行对比验证该方法的可行性及有效性.

关键词: 多向核熵独立成分分析; 四阶累积分析; 多向核主成分分析; 多向核独立成分分析; 故障监测

引用格式: 常鹏, 乔俊飞, 张祥宇, 等. 基于四阶累积分析的工业大肠杆菌制备过程故障诊断. 控制理论与应用, 2020, 37(3): 667 – 675

DOI: 10.7641/CTA.2019.80361

Fourth order cumulant analysis based fault diagnosis of the preparation process of industrial Escherichia coli

CHANG Peng[1,2,3†], QIAO Jun-fei[1,2], ZHANG Xiang-yu[1,3], WANG Pu[1,3]

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

Abstract: In the process of industrial Escherichia coli preparation, process data has both nonlinear and non-Gaussian characteristics, making it difficult to locate fault sources effectively. Aiming at this problem, a modeling method based on multiway kernel entropy independent component analysis (MKEICA) is proposed. Furthermore, in order to overcome the insufficiency of traditional low-order monitoring statistics (T^2, I^2 and SPE) to obtain non-Gaussian information, a fourth-order cumulative monitoring statistic method was proposed. In the next place, through the derivation of the fourth order cumulative monitoring statistic, the cause of the fault was obtained. For industrial validation, the feasibility and superiority of the proposed monitoring method were demonstrated in the comparison with the multiway kernel independent component anlaysis (MKICA) monitoring model.

Key words: multiway kernel entropy independent component analysis; forth-order cumulant analysis; multiway kernel principal component analysis; multiway kernel independent component analysis; fault detection

Citation: CHANG Peng, QIAO Junfei, ZHANG Xiangyu, et al. Fourth order cumulant analysis based fault diagnosis of the preparation process of industrial Escherichia coli. *Control Theory & Applications*, 2020, 37(3): 667 – 675

1　引言

多元统计技术已经广泛应用到间歇过程的过程监测中, 因其只需要生产过程数据进行统计建模, 无需考虑间歇生产过程复杂的机理特性, 在工业界和学术界得到了卓有成效的应用, 其中针对非线性问题具有代表性的就是多向核主成分分析[1–4](multiway kernel principal component analysis, MKPCA), 多项核偏最小二乘方法[5–10](multiway kernel partial least square, MKPLS). 尽管MKPCA/MKPLS及其扩展方法在过程故障检测得到了广泛应用, 但其要达到好的监测的效果必须是假设生产过程属于高斯分布, 在进行过程特征提取时, 仅仅考虑低阶统计信息, 未考虑过程数据

收稿日期: 2018−05−17; 录用日期2019−07−15.
†通信作者. E-mail: changpeng@bjut.edu.cn; Tel: +86 10-67393897.
本文责任编委: 周东华.
国家自然科学基金项目(61364009, 61174109)资助.
Supported by the National Natural Science Foundation of China (61364009, 61174109).

的高阶统计量信息, 会造成特征提取的不充分[10–12]. 这是因为高斯过程的二阶以上统计信息为零, 而实际的间歇工业过程往往同时具有非线性和非高斯性, 用以上方法对非高斯过程进行监测势必导致监测模型存在大量的报警问题, 甚至导致失效性. 针对间歇生产过程数据普遍具有非线性和非高斯性共存的问题, 近些年来, 基于多向核独立元分析[10–18](multiway kernle independent component analysis, MKICA)的监测方法逐渐发展了起来. 其具体核心算法是将原始数据用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后进行ICA分解提取过程的独立成分. 但是KPCA在进行白化处理时其实质是在特征提取, 而其特征提取的依据是方差. 然而间歇过程具有明显的阶段簇结构特征[19], 以本文采用的对象, 工业重组大肠杆菌制备白介素–2为例, 其生产过程按照微生物机理分为无补料菌种培养阶段、菌种的补料快速生长阶段、诱导产物合成阶段, 具有明显的多阶段特性[20]. 为解决上述问题, 常鹏等人[20]提出了一种基于MKECA[21–22]白化的MKEICA方法. 该方法利用MKECA进行特征提取时能够保留数据的簇结构的阶段特征, 之后在核熵空间建立ICA监测模型用于生产过程的监控. 但是以上MKICA/MKEICA监测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, 只能监测生产过程的低阶统计信息, 不能监测生产过程的高阶统计的信息, 如生产过程中产生的故障具有明显的高阶统计信息时其将无能为力, 而实际的生产过程往往具有高阶统计信息[23–25]. 而文献[24–30]指出, 非高斯信息需要高阶矩(大于二阶矩)来分析, 并强调第四阶矩含有明显的非高斯信息. 为此本文将MKEICA[20]作为基础的建模监测模型, 构建四阶累积监控统计量用于过程监控, 当监测到故障时采用本文提出的故障诊断方法对故障变量进行追溯. 采用高阶累积分析建模时具有以下3点优势:

1) 高斯过程的高阶累积量为零, 而非高斯过程的高阶累积量不全为零. 因此, 高阶累积量可用于高斯噪声中非高斯信号的提取.

2) 高阶累积量不仅具有信号的幅值信息, 同时具有相位信息.

3) 高阶累积量可用于监测和描述系统的非线性.

综上本文提出的方法与传统MKICA方法同时用于工业大肠杆菌制备现场验证其算法的有效性, 当监测到生产过程出现故障时, 对故障变量进行追溯, 定位故障源, 验证该故障诊断方法的实用性.

2 四阶累积量分析

四阶累积量(forth-order cumulants analysis, FCA)是随机非高斯特性的参数矩估计[23–25], 对于一个零均值变量k维随机向量$X=[x_1\ \ x_2\ \ \cdots\ \ x_k]$, 若其联合概率密度函数为$f(x_1,x_2,\cdots,x_k)$, 则其第一特征函数定义为

$$\phi_x(\omega)=\mathrm{E}\{\mathrm{e}^{\mathrm{j}\omega^{\mathrm{T}}X}\}=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,x_2,\cdots,x_k)\mathrm{e}^{\mathrm{j}\omega^{\mathrm{T}}X}\mathrm{d}x_1\cdots\mathrm{d}x_2,\tag{1}$$

式中: $\omega=[\omega_1\ \ \omega_2\ \ \cdots\ \ \omega_k]$, $\mathrm{E}\{\cdot\}$表示数学期望, $\mathrm{j}=\sqrt{-1}$. X的第2特征函数又称为累积量生成函数定义为

$$\psi_x(\omega)=\ln\phi_x(\omega)=\ln\mathrm{E}\{\mathrm{e}^{\mathrm{j}\omega^{\mathrm{T}}X}\},$$

式中ln表示自然对数. 分别对ϕ_x和$\psi_x(\omega)$求关于ω的$r=r_1+r_2+\cdots+r_k$阶偏导, 同时令$[\omega_1=\omega_2=\cdots=\omega_k=0]$, 即可得到$X$的$r$阶矩$m_{r_1\cdots r_k}$与$r$阶累积量$c_{r_1\cdots r_k}$

$$\begin{aligned}&m_{r_1\cdots r_k}=\mathrm{mom}\{x_1,x_2,\cdots,x_k\}=\\&(-j)^r\frac{\partial^r\phi_{\mathrm{X}}(\omega_1,\omega_2,\cdots,\omega_k)}{\partial\omega_1^{r_1}\cdots\partial\omega_k^{r_k}}\Big|_{\omega_1=\omega_2=\cdots=\omega_k=0}=\\&\mathrm{E}\{x_1^{r_1},x_2^{r_2},\cdots,x_k^{r_k}\},\end{aligned}\tag{2}$$

$$\begin{aligned}&c_{r_1\cdots r_k}=\mathrm{cum}\{x_1,x_2,\cdots,x_k\}=\\&(-j)^r\frac{\partial^r\Psi_{\mathrm{X}}(\omega_1,\omega_2,\cdots,\omega_k)}{\partial\omega_1^{r_1}\cdots\partial\omega_k^{r_k}}\Big|_{\omega_1=\omega_2=\cdots=\omega_k=0},\end{aligned}\tag{3}$$

式中mom$\{\cdot\}$, cum$\{\cdot\}$分别是矩和累积量的符号. 实际应用中常取$r_1=r_2=\cdots=r_k=1$, 对于零均值的k阶平稳随机信号$x(h)$, 令$x_1=x(h)$, $x_2=x(t+h_1)$, $\cdots$, $x_k=x(t+h_{k-1})$, h为时间滞后. 则该信号的k阶矩和k阶累积量分别定义为

$$\begin{aligned}&m_{kx}(h_1,h_2,\cdots,h_{k-1})=\\&\mathrm{mom}\{x(h),x(t+h_1),\cdots,x(t+h_{k-1})\},\end{aligned}\tag{4}$$

$$\begin{aligned}&c_{kx}(h_1,h_2,\cdots,h_{k-1})=\\&\mathrm{cum}\{x(h),x(t+h_1),\cdots,x(t+h_{k-1})\},\end{aligned}\tag{5}$$

其中k阶矩$m_{kx}(h_1,h_2,\cdots,h_{k-1})$可由式(2)求得, 结果为

$$\begin{aligned}&m_{kx}(h_1,h_2,\cdots,h_{k-1})=\\&\mathrm{E}\{x(h),x(t+h_1),\cdots,x(t+h_{k-1})\}.\end{aligned}\tag{6}$$

k阶累积量$c_{kx}(h_1,h_2,\cdots,h_{k-1})$由矩与累积量转换关系求出, 首先引入符号集的概念, 对于k个随机变量的集合$\{x_1,x_2,\cdots,x_k\}$, 其符号集合为$I=\{1,2,\cdots,k\}$对I进行无连接的非空分割得到q个子集合记作I_p, 要求$I=\cup I_p$, I_p非空且是无序组合, 各I_p无交集. 累积量–矩转换公式(C–M公式)为

$$m_x(I)=\sum_{\bigcup_{p=1}^{q}I_p=I}\prod_{p=1}^{q}c_x(I_p).\tag{7}$$

对于均值为零的随机信号x_k, 其四阶累积量为

$$\begin{aligned}&c_{4x}(h_1,h_2,h_3)=\\&\text{cum}\{x(t)x(t+h_1)x(t+h_2)x(t+h_3)\}=\\&\text{E}\{x(t)x(t+h_1)x(t+h_2)x(t+h_3)\}-\\&R_x(h_1)R_x(h_3-h_2)-R_x(h_2)R_x(h_3-h_1)-\\&R_x(h_3)R_x(h_3-h_1),\end{aligned}\tag{8}$$

其中: E表示期望; R_x表示信号$x(t)$自相关系数; 令$h_1=1, h_2=2, h_3=3$是滞后时间. 本文的数据经ICA处理后具有独立性, 这里R_x恒为零, 故四阶累积量可改写为

$$\begin{aligned}&c_{4x}(h_1,h_2,h_3)=\\&\text{E}\{x(t)x(t+h_1)x(t+h_2)x(t+h_3)\}=\\&\text{E}\{x(t)x(t+1)x(t+2)x(t+3)\}.\end{aligned}\tag{9}$$

3　基于四阶累积量分析的过程故障诊断

3.1　四阶累积监测统计量的构建

核函数选择间歇过程监测领域最常见的径向基函数 (radial basis function, RBF) 核[1–23], 具体如下式所示:

$$\lambda v=\bar{K}v.\tag{10}$$

可以获得对应于$\bar{K}$的u个较大特征值$\lambda_1,\lambda_1,\cdots,\lambda_u$, 的特征向量$v_1,v_2,\cdots,v_u$. 根据相对特征值来确定$u$, 准则如下所示:

$$\frac{\lambda_i}{\text{sum}(\lambda)}\geqslant 0.0001.\tag{11}$$

本文利用核熵值的累积贡献率来进行核熵个数的选择, 累积核熵贡献率定义如下所示[20]:

$$\frac{\sum\limits_{i=1}^{d}\hat{V}(p)_i}{\sum\limits_{i=1}^{n}\hat{V}(p)_i}\times 100\%\geqslant 85\%.\tag{12}$$

由

$$V=[v_1\ \ v_2\ \ \cdots\ \ v_u],\ \varLambda=\text{diag}\{\lambda_1,\lambda_1,\cdots,\lambda_u\},$$
$$\hat{V}(p)=\frac{1}{N^2}\sum_{i=1}^{N}(\sqrt{\lambda_i}V_i^{\mathrm{T}}1)^2,$$

则MKECA负载矩阵为

$$H=(h_1,h_2,\cdots,h_u)=\bar{\varPhi}V\varLambda^{-\frac{1}{2}}.\tag{13}$$

白化矩阵为

$$Q=H(\frac{1}{n}\varLambda)^{-\frac{1}{2}}=\sqrt{n}\bar{\varPhi}V\varLambda^{-1}\bar{k}^{\mathrm{T}}.\tag{14}$$

MKECA得分为

$$\begin{aligned}&z=H^{\mathrm{T}}\bar{\phi}=\varLambda^{-\frac{1}{2}}V^{\mathrm{T}}\bar{\phi}^{\mathrm{T}}\bar{\phi}=\\&\varLambda^{-\frac{1}{2}}V^{\mathrm{T}}\big[\bar{k}(x_1,x)\ \ \bar{k}(x_2,x)\ \ \cdots\ \ \bar{k}(x_n,x)\big]^{\mathrm{T}}=\\&\varLambda^{-\frac{1}{2}}V^{\mathrm{T}}\bar{k}^{\mathrm{T}}.\end{aligned}\tag{15}$$

上式中$\bar{\phi}$是$\bar{\varPhi}$的一列, $\bar{k}$是$\bar{K}$的一行. 白化得分为

$$\bar{z}=Q^{\mathrm{T}}\bar{\phi}=\sqrt{n}\varLambda^{-1}V^{\mathrm{T}}\bar{k}^{\mathrm{T}}.\tag{16}$$

实际上$\bar{z}$与z的关系如下式所示:

$$z=(\frac{1}{n}\varLambda)^{\frac{1}{2}}\bar{z}.\tag{17}$$

对于一个新时刻的数据向量x_{new}, 其对应的核熵向量为

$$K_{\text{new}}=[k(x_1,x_{\text{new}})\ \ k(x_2,x_{\text{new}})\ \ \cdots\ \ k(x_n,x_{\text{new}})],\tag{18}$$

则新的KECA得分为

$$z_{\text{new}}=\varLambda^{-\frac{1}{2}}V^{\mathrm{T}}\bar{K}_{\text{new}}^{\mathrm{T}}.\tag{19}$$

新的白化得分为

$$\bar{z}_{\text{new}}=\sqrt{n}\varLambda^{-1}V^{\mathrm{T}}\bar{K}_{\text{new}}^{\mathrm{T}}.\tag{20}$$

对$\bar{z}$进行ICA分解, 如下式所示:

$$s=B^{\mathrm{T}}\bar{z}.\tag{21}$$

根据s的非高斯性大小这里采用独立元得分累积贡献最大的几个独立成分s_d, 与其对应的B矩阵的列组成矩阵B_d, 其独立成分为$s_d=B_d^{\mathrm{T}}\bar{z}$.

定义新的监测统计量HS和HE并计算其控制线, 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的样本四阶累积量为

$$\begin{aligned}&hs_d(i)=s_d(i)s_d(i{+}1)s_d(i{+}2)s_d(i{+}3)=\\&w_p\bar{K}(i)w_p\bar{K}(i{+}1)w_p\bar{K}(i{+}2)w_p\bar{K}(i{+}3),\end{aligned}\tag{22}$$

其中: w_p是解混矩阵W_d的第p个向量, $p=1,2,\cdots,d$. 为了监测全部主导独立成分的四阶累积量, 第1个监测指标定义为

$$\begin{aligned}&\text{HS}(i)=\\&\sum_{p=1}^{d}|hs_p(i)|[s_d(i)s_d(i{+}1)s_d(i{+}2)s_d(i{+}3)]^{\mathrm{T}}\times\\&[s_d(i)s_d(i{+}1)s_d(i{+}2)s_d(i{+}3)]=\\&[w_pK(i)w_pK(i{+}1)w_pK(i{+}2)w_pK(i{+}3)]^{\mathrm{T}}\times\\&[w_pK(i)w_pK(i{+}1)w_pK(i{+}2)w_pK(i{+}3)].\end{aligned}\tag{23}$$

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为

$$\begin{aligned}&he_q(i)=e_q(i)e_q(i{+}1)e_q(i{+}2)e_q(i{+}3)=\\&l_qK(i)l_qK(i{+}1)l_qK(i{+}2)l_qK(i{+}3),\end{aligned}\tag{24}$$

其中l_q是L的第q行, $q=1,2,\cdots,m$, HCA的第2个监测指标定义为

$$\begin{aligned}&\text{HE}(i)=\\&\sum_{q=1}^{m}|he_q(i)|[e_q(i)e_q(i{+}1)e_q(i{+}2)e_q(i{+}3))]^{\mathrm{T}}\times\\&[e_q(i)e_q(i{+}1)e_q(i{+}2)e_q(i{+}3)]=\end{aligned}$$

$$[l_qK(i)l_qK(i+1)l_qK(i+2)l_qK(i+3)]^{\mathrm{T}}\times [l_qK(i)l_qK(i+1)l_qK(i+2)l_qK(i+3)]. \tag{25}$$

至此, 离线阶段的模型建立完毕, 这里需要用核密度估计[26–28]分别计算监测统计量HS和HE的置信区间用于在线监测.

3.2 在线监测

对于在线监测, 新时刻数据的监测统计量定义如下式所示:

$$\begin{aligned}&hs_{\mathrm{new}}(i)=\\&s_{\mathrm{new}}(i)s_{\mathrm{new}}(i+1)s_{\mathrm{new}}(i+2)s_{\mathrm{new}}(i+3)=\\&w_p\bar{K}_{\mathrm{new}}(i)w_p\bar{K}_{\mathrm{new}}(i+1)w_p\times\\&\bar{K}_{\mathrm{new}}(i+2)w_p\bar{K}_{\mathrm{new}}(i+3),\end{aligned} \tag{26}$$

$$\begin{aligned}&\mathrm{HS}_{\mathrm{new}}(i)=\\&\sum_{p=1}^{d}|hs_{\mathrm{new}p}(i)|\times\\&[s_{\mathrm{new},d}(i)s_{\mathrm{new},d}(i+1)s_{\mathrm{new},d}(i+2)s_{\mathrm{new},d}(i+3)]^{\mathrm{T}}\times\\&[s_{\mathrm{new},d}(i)s_{\mathrm{new},d}(i+1)s_{\mathrm{new},d}(i+2)s_{\mathrm{new},d}(i+3)]=\\&A^{\mathrm{T}}A,\end{aligned} \tag{27}$$

其中:

$$\begin{aligned}A=[&w_{\mathrm{new},p}K_{\mathrm{new}}(i)w_{\mathrm{new},p}K_{\mathrm{new}}(i+1)\\&w_{\mathrm{new},p}K_{\mathrm{new}}(i+2)w_{\mathrm{new},p}K_{\mathrm{new}}(i+3)].\end{aligned}$$

$$\begin{aligned}&he_{\mathrm{new}}(i)=\\&e_{\mathrm{new}}(i)e_{\mathrm{new}}(i+1)e_{\mathrm{new}}(i+2)e_{\mathrm{new}}(i+3)=\\&l_qK(i)l_qK(i+1)l_qK(i+2)l_qK(i+3),\end{aligned} \tag{28}$$

$$\begin{aligned}&\mathrm{HE}_{\mathrm{new}}(i)=\\&\sum_{q=1}^{m}|he_{\mathrm{new},q}(i)|\times\\&[e_{\mathrm{new},q}(i)e_{\mathrm{new},q}(i+1)e_{\mathrm{new},q}(i+2)e_{\mathrm{new},q}(i+3))]^{\mathrm{T}}\times\\&[e_{\mathrm{new},q}(i)e_{\mathrm{new},q}(i+1)e_{\mathrm{new},q}(i+2)e_{\mathrm{new},q}(i+3)]=\\&B^{\mathrm{T}}B,\end{aligned} \tag{29}$$

其中:

$$\begin{aligned}B=[&l_{\mathrm{new},q}K_{\mathrm{new}}(i)l_{\mathrm{new},q}K_{\mathrm{new}}(i+1)l_{\mathrm{new},q}\\&K_{\mathrm{new}}(i+2)l_{\mathrm{new},q}K_{\mathrm{new}}(i+3)].\end{aligned}$$

3.3 故障诊断

采用RBF核函数计算核矩阵[1–25], 假设存在向量$v=[v_1\ \ v_2\ \ \cdots\ \ v_m]^{\mathrm{T}}, i=1,2,\cdots,m$, 核函数对于第$i$个变量$v_i$的偏导可用下式计算:

$$\begin{aligned}&\frac{\partial k(x_j,x_k)}{\partial v_i}=\frac{\partial k(v\cdot x_j,v\cdot x_k)}{\partial v_i}=\\&-\frac{1}{\sigma}(v_ix_{j,i}-v_ix_{k,i})^2k(v\cdot x_j,v\cdot x_k)=\\&-\frac{1}{\sigma}(v_ix_{j,i}-v_ix_{k,i})^2k(x_j,x_k)|_{v_i=1},\end{aligned} \tag{30}$$

其中$x_{j,i}$为第j个样本的第i个变量. 因此, 第i个变量对核向量第j个元素的贡献为求两个核函数乘积的偏导

$$\begin{aligned}&C\_k_{\mathrm{new},j}(i)=\\&\frac{\partial k(x_j,x_{\mathrm{new}})k(x_k,x_{\mathrm{new}})}{\partial v_i}=\\&-\frac{1}{\sigma}[(x_{j,i}-x_{\mathrm{new},j})^2+(x_{k,i}-x_{\mathrm{new},i})]\times\\&k(x_j,x_{\mathrm{new}})k(x_k,x_{\mathrm{new}}).\end{aligned} \tag{31}$$

对于在线监测, 根据变量对角贡献[11], 核向量的第j个元素对HS的贡献为

$$C\_\mathrm{HS}(k_{\mathrm{new},j})=nk_{\mathrm{new},j}^2\mathrm{HS}(j), \tag{32}$$

进一步, $hs_{\mathrm{new},p}$中第i个变量对HS的贡献为

$$C\_\mathrm{HS}(i)=n(C\_k_{\mathrm{new},j}(i)^2)\|hs_{\mathrm{new},p}(j)\|^2. \tag{33}$$

按照上述方法求出核向量第j个元素对HE的贡献. 可以采用变量贡献分析进行故障诊断. 对于一个监测指标, 变量对该指标的贡献由下式定义:

$$\mathrm{Index}=\sum_{j=1}^{m}C_j^{\mathrm{Index}}. \tag{34}$$

这里C_j^{Index}表示第j个变量对指标Index的贡献, 所有变量贡献之和应该等于该指标的值. 结合以上公式(23)(30)–(32), 对HS进行如下分解:

$$\begin{aligned}&\mathrm{HS}(i)=\\&\sum_{p=1}^{d}\Big|\frac{\{s_p(i)s_p(i+1)s_p(i+2)s_p(i+3)-mhs_p\}}{vhs_p}\Big|=\\&\sum_{p=1}^{d}\Big|\frac{\{w_px(i)s_p(i+1)s_p(i+2)s_p(i+3)-mhs_p\}}{vhs_p}\Big|=\\&\sum_{p=1}^{d}\mathrm{sgn}\{s_p(i)s_p(i+1)s_p(i+2)s_p(i+3)-mhs_p\}\times\\&\frac{\{w_px(i)s_p(i+1)s_p(i+2)s_p(i+3)-mhs_p\}}{vhs_p}=\\&\sum_{p=1}^{d}\mathrm{sgn}\{s_p(i)s_p(i+1)s_p(i+2)s_p(i+3)-mhs_p\}\times\\&\frac{\sum_{j=1}^{m}\{w_{p,j}x_j(i)s_p(i+1)s_p(i+2)s_p(i+3)-cmhs_{p,j}\}}{vhs_p}=\\&\sum_{j=1}^{m}\sum_{p=1}^{d}\mathrm{sgn}\{s_p(i)s_p(i+1)s_p(i+2)s_p(i+3)-cmhs_{p,j}\}\times\\&\frac{\sum_{j=1}^{m}\{w_{p,j}x_j(i)s_p(i+1)s_p(i+2)s_p(i+3)-cmhs_{p,j}\}}{vhs_p}=\\&\sum_{j=1}^{m}C_j^{\mathrm{HS}}(i),\end{aligned} \tag{35}$$

式中:

$$C_j^{\mathrm{HS}}(i) = \frac{\sum\limits_{p=1}^{d}\mathrm{sgn}\{s_p(i)s_p(i+1)s_p(i+2)s_p(i+3)-mhs_p\}\times \sum\limits_{j=1}^{m}\{w_{p,j}x_j(i)s_p(i+1)s_p(i+2)s_p(i+3)-cmhs_{p,j}\}}{vhs_p}$$

表示第j个变量对HS的贡献, $cmhs_{p,j}$为第j个变量对hsp的贡献为

$$cmhs_{p,j}= \mathrm{E}\{w_{p,j}x_j(i)s_p(i+1)s_p(i+2)s_p(i+3)\} = \frac{1}{n-3}\sum_{i=4}^{n}w_{p,j}x_j(i)s_p(i+1)s_p(i+2)s_p(i+3), \tag{36}$$

$$C\_\mathrm{HS}(k_{\mathrm{new},j}) = nk^2_{\mathrm{new},j}\mathrm{HE}(i), \tag{37}$$

进一步, $he_{\mathrm{new},p}$中第i个变量对HE的贡献为按照HS的分解方法

$$\begin{aligned}\mathrm{HE}(i) = &\sum_{q=1}^{m}\left|\frac{\{e_k(i)e_q(i+1)e_q(i+2)e_q(i+3)-mhe_q\}}{vhe_q}\right| = \\ &\sum_{q=1}^{m}\left|\frac{\{l_qx(i)l_q(i+1)l_q(i+2)l_q(i+3)-mhe_q\}}{vhe_q}\right| = \\ &\sum_{q=1}^{m}\mathrm{sgn}\{e_q(i)e_q(i+)e_q(i+2)e_q(i+3)-mhe_q\}\times \\ &\frac{\{l_{q,j}x_j(i)e_q(i+1)e_q(i+2)e_q(i+3)-cmhe_{q,j}\}}{vhe_q} = \\ &\sum_{q=1}^{m}\mathrm{sgn}\{e_q(i)e_q(i+1)e_q(i+2)e_q(i+3)-mhe_q\}\times \\ &\frac{\{l_{q,j}x(i)e_q(i+1)e_q(i+2)e_q(i+3)-cmhe_{q,j}\}}{vhe_q} = \\ &\sum_{j=1}^{m}\sum_{p=1}^{d}\mathrm{sgn}\{e_q(i)e_q(i+1)e_q(i+2)e_q(i+3)-mhe_q\}\times \\ &\frac{\{l_{q,j}x(i)e_q(i+1)e_q(i+2)e_q(i+3)-cmhe_{q,j}\}}{vhe_q} = \\ &\sum_{j=1}^{m}C_j^{\mathrm{HE}}(i),\end{aligned} \tag{38}$$

式中:

$$C_j^{\mathrm{HE}}(i) = \sum_{q=1}^{m}\mathrm{sgn}\{e_q(i)e_q(i+1)e_q(i+2)e_q(i+3)-mhe_q\}\times \frac{\{l_{q,j}x(i)e_q(i+1)e_q(i+2)e_q(i+3)-cmhe_{q,j}\}}{vhe_q}$$

表示第j个变量对HE的贡献. $cmhe_{q,j}$为第j个变量对he_q的贡献, 如下式所示:

$$cmhe_{q,j} = \mathrm{E}\{l_{q,j}x(i)s_p(i+1)s_p(i+2)s_p(i+3)\} = \frac{1}{n-2}\sum_{i=3}^{n}l_{q,j}x(i)s_p(i+1)s_p(i+2)s_p(i+3), \tag{39}$$

$$C\_\mathrm{HE}(k_{\mathrm{new},j}) = nk^2_{\mathrm{new},j}cmhe_{p,j}. \tag{40}$$

以上推导了基于核的四阶累积量监测统计量贡献值, 当过程监测模型发现有异常情况时, 可用此故障诊断方法对故障变量进行追溯.

4　算法实际工业验证

这里实验的主要目的是证明下列观点: 1) 基于阶段的过程数据是非线性、非高斯性、多阶段共存的, 不是单一存在的; 2) 基于四阶累积监测统计量的故障监测方法具备有效的故障监测能力; 3) 基于四阶累积监测统计量的故障诊断方法具备故障定位的能力. 本文实验的数据取自北京某微生物制药公司的基因重组大肠杆菌外源蛋白表达制备白介素–2. 生产发酵过程采用Sartorius BIOSTAT BDL 15L发酵罐, 如图1所示. 选择6个主要过程变量来综合表征菌体生长及外源蛋白表达的状况, 如表1所示. 选取35个正常批次作为离线建模数据. 对过程变量数据进行正态检验结果如图2所示, 如果数据服从正态分布, 则图2中数据点击图中蓝色 ‘+’ 应该近似为一条直线, 即图中红线附近, 变量1至变量6都具有非高斯特性. 此外, 为了验证模型的有效性, 引入两种类型的故障如表2所示.

表 1　大肠杆菌发酵过程可检测变量

Table 1　The measuring parameters in Escherichia coli fermentation processes

序号	变量
1	pH值
2	溶解氧浓度(DO)/%
3	温度(temperature)/°C
4	搅拌转速(agitator speed)/($\mathrm{r}\cdot\mathrm{min}^{-1}$)
5	补葡萄糖量(glucose feed rate)/ml
6	补培养基量(culture medium feed rate)/ml

表 2　工厂过程故障类型

Table 2　Fault types introduced in process

序号	变量名称	故障类型	幅度/%	时间/h
1	搅拌功率	阶跃变化	2	15∼39
2	补糖速率	斜坡扰动	10	15∼39

4.1　监测结果与讨论

针对表2中的故障1的监测结果如图3所示, 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3(a)可以看出, I^2监测图在正常阶段存在故障的错误报警现象, 如采样点2附近、采样点15附近, SPE监测图存在故障的误报警, 如采样点1到3附近, 采样点6附近、采样点11附近, 从图3(b)可以看出本文所提监控方法的统计量HS和HE监测图不存在故

障的漏报警和误报警现象, 表现出可靠的过程监测能力, 当监测到生产过程存在异常情况时, 采用本文所提的故障诊断方法对其进行故障诊断, 这里选择25 h进行故障诊断.

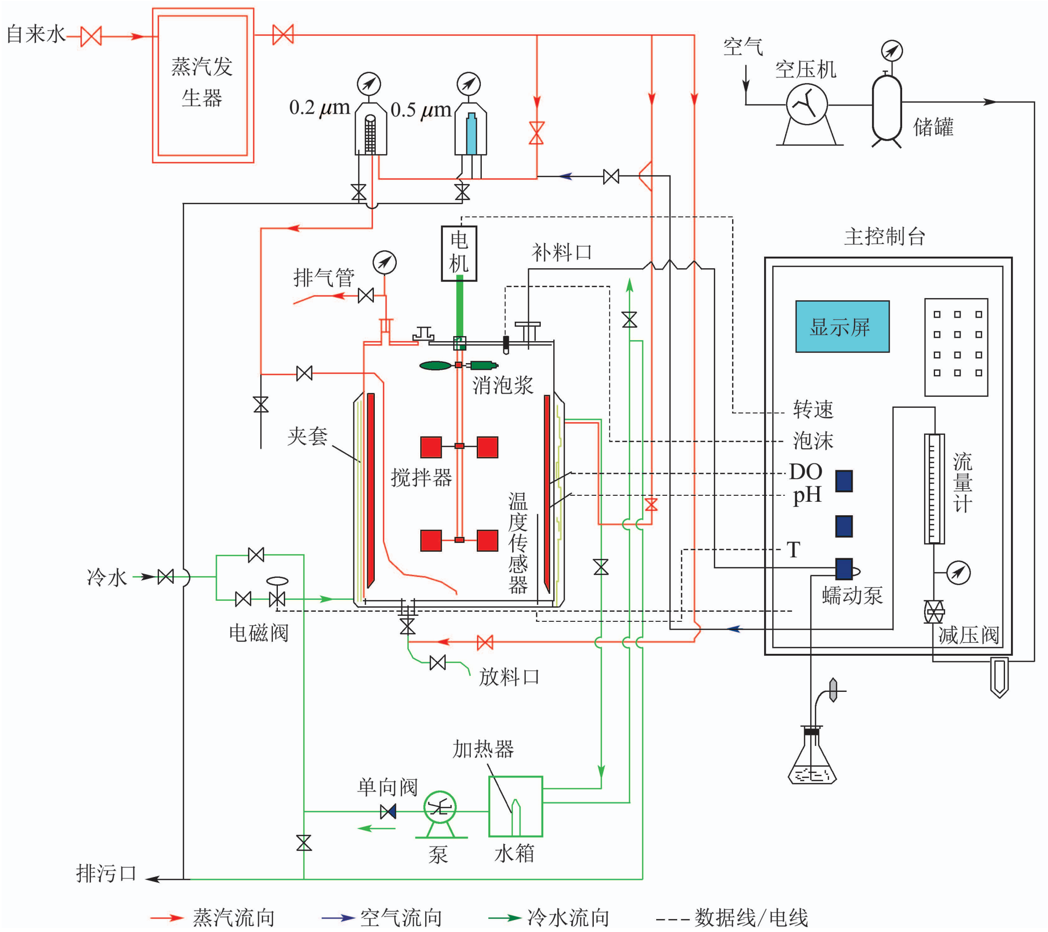
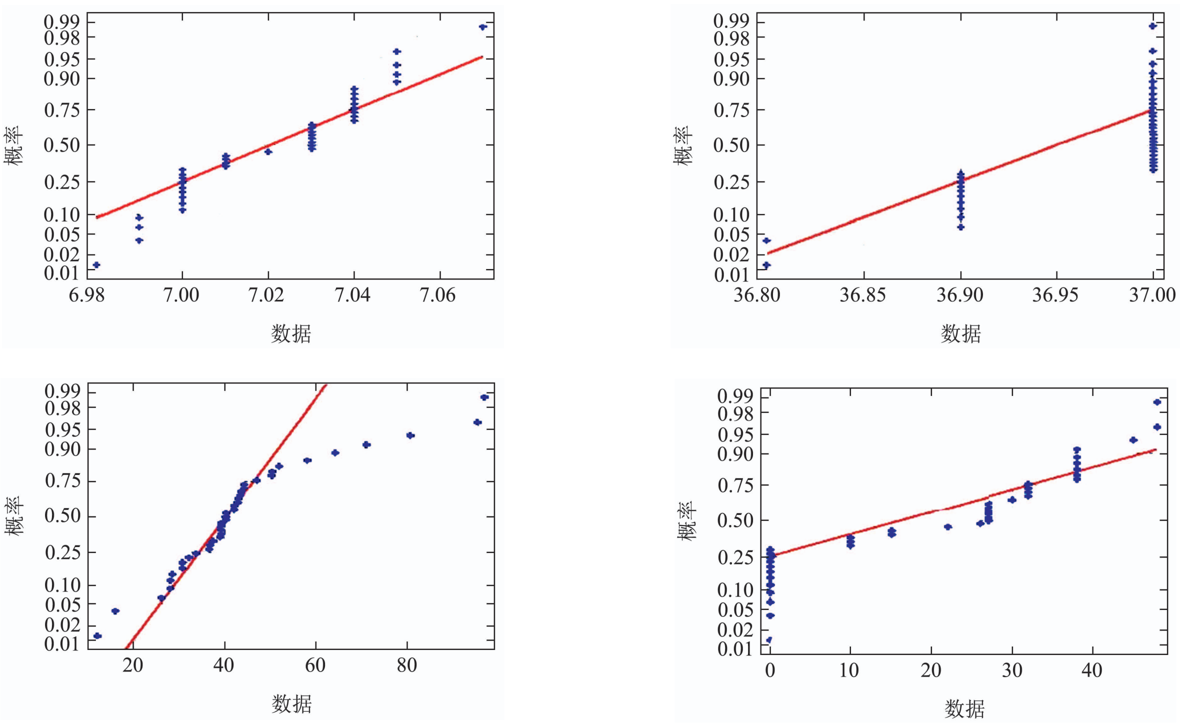


图 1 大肠杆菌发酵制备系统原理示意图

Fig. 1 Schematic of fermentation equipment and control system



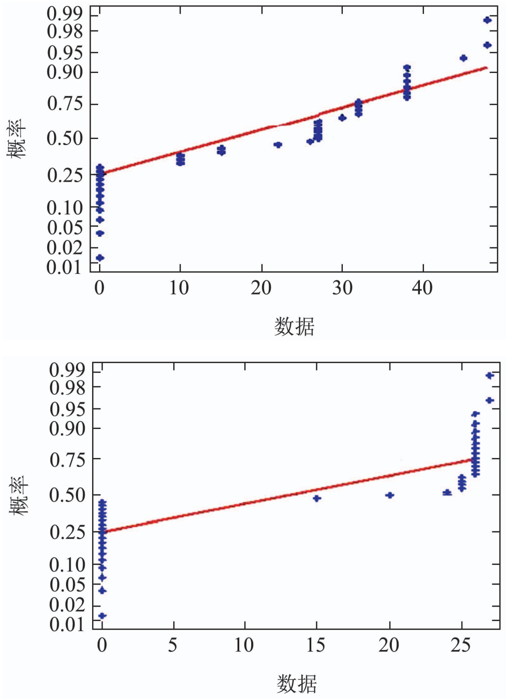
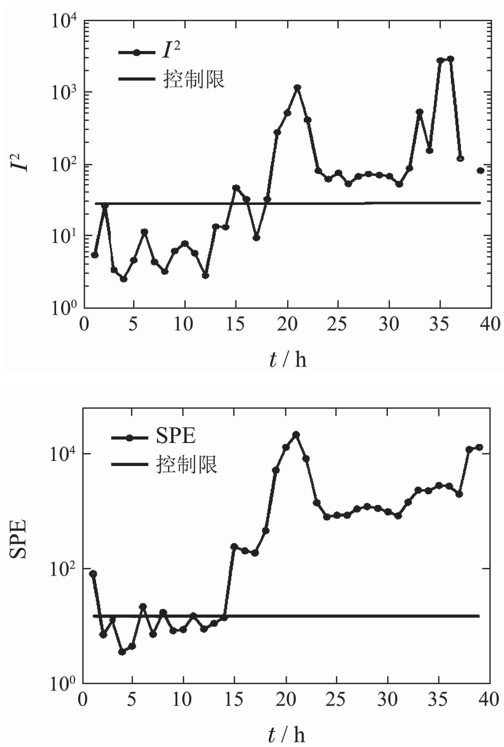


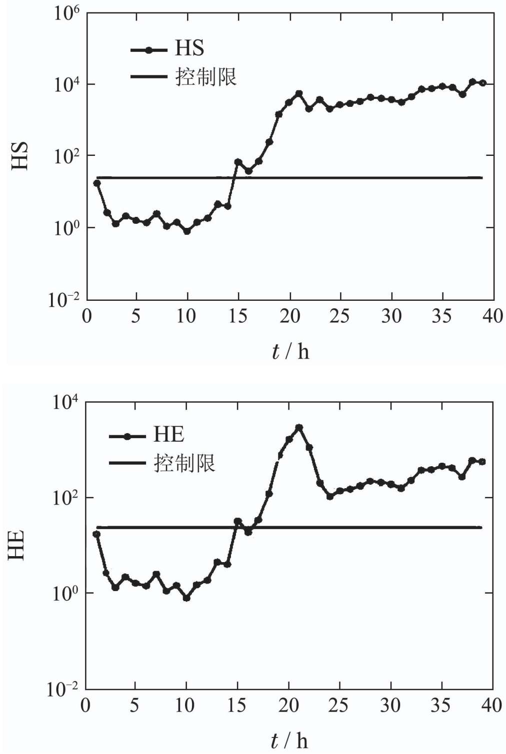
图 2　6个过程变量正态分布检验

Fig. 2　6 process variables of normal distribution test

针对表2中的故障2的监测结果如图4所示, MKICA的I^2监测图在28采样点附近超出控制限, SPE监测统计量在36采样点附近超出监测控制限, 基本失去了对此故障的发现能力. 而本文所提监控方法的统计量HS在16采样点附近超出控制限, HE监测图在18采样点附近超出控制限, 体现了高效的故障监测识别能力, 并且不存在故障的误报警和漏报警.



(a)　MKICA方法

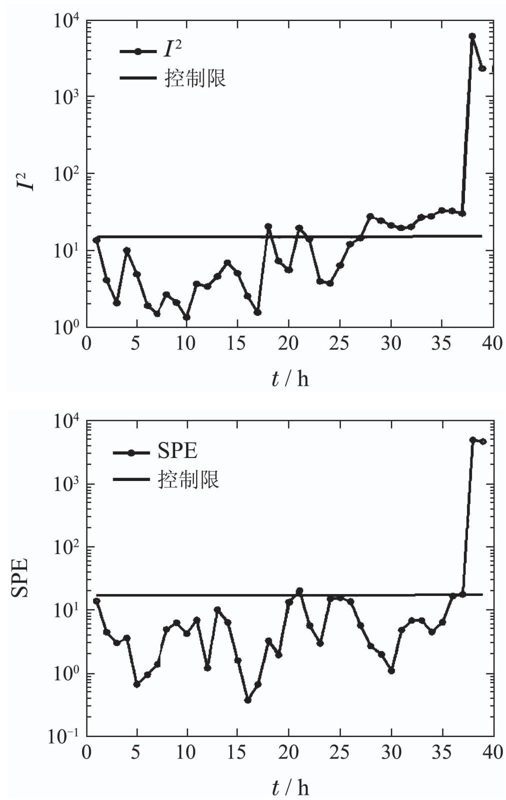


(b)　FCA方法

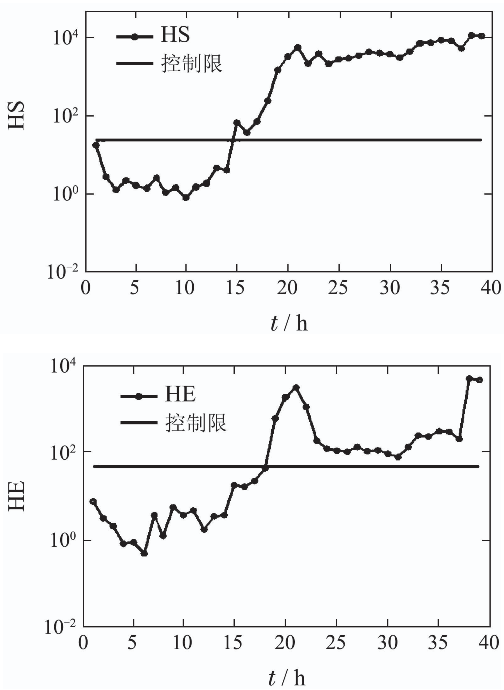
图 3　比较两种方法在工业发酵过程故障批次1的监测结果

Fig. 3　Monitoring results using two methods for industry fault batch 1

采用本文故障诊断方法, 选择发现故障后的稳定采样点对其进行故障诊断, 选择了第26个采样点.



(a)　MKICA方法

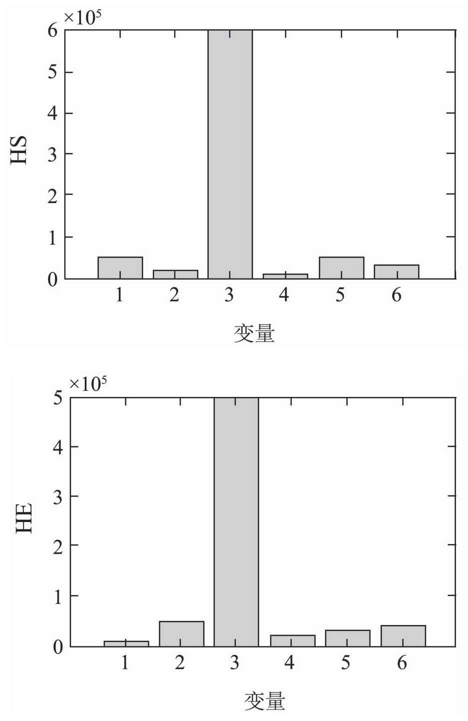


(b) FCA方法

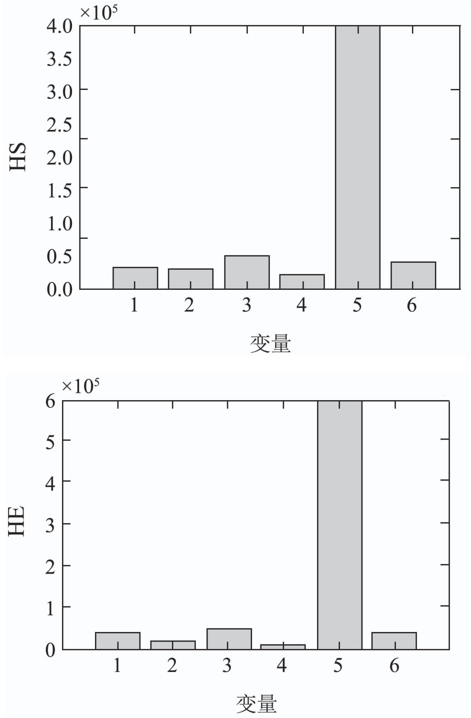
图 4 比较两种方法在工业发酵过程故障批次2的监测结果

Fig. 4 Monitoring results using two methods for industry fault batch 2

针对故障1的故障诊断如图5(a)所示, 变量3为故障变量, 结合表1–2进行分析可知,可以准确识别故障源. 如图5(b)所示, 变量5对监测统计量的贡献最大, 结合由表1–2进行分析可知, 该方法可以准确识别该故障.



(a) 针对故障1的故障诊断



(b) 针对故障2的故障诊断

图 5 故障诊断

Fig. 5 Fault diagnosis

通过以上分析, 得出FCA监测模型优于MKICA的监测模型.

5 结论

针对工业重组大肠杆菌外源蛋白表达制备白介素-2的数据同时具有非线性、多阶段、非高斯性的特点, 同时为了解决传统监测模型难于进行故障监测以及故障源定位的难题, 提出基于FCA的过程监测方法. 该方法首先利用MKEICA代替传统MKICA监测模型在进行数据特征提取时可以更好保持原始的数据的簇结构阶段信息, 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, 提出了四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报和漏报的问题, 此外, 本文还推导了基于FCA方法的故障源定位. 通过对工业制备大肠杆菌发酵过程的实际应用, 表明本文方法与传统MKICA方法相比, 不仅能有效提高监控系统发现故障的能力, 并且还可以利用本文提出的故障诊断方法准确识别故障源, 为间歇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:

[1] LEE J M, YOO C K, LEE I B. Fault detection of batch processes using multiway kernel principal component analysis. *Computers & Chemical Engineering*, 2004, 28(9): 1837 – 1847.

[2] YOO C K, VILLEZ K, LEE I B, et al. Multivariate nonlinear statistical process control of a sequencing batch reactor. *Journal of Chemical Engineering of Japan*, 2006, 39(1): 43 – 51.

[3] YOO C K, LEE I B, VANROLLEGHEM P A. On-line adaptive and nonlinear process monitoring of a pilot-scale sequencing batch reactor. *Environmental Monitoring and Assessment*, 2006, 119(1/3): 349 – 366.

[4] JIA M, CHU F, WANG F, et al. On-line batch process monitoring using batch dynamic kernel principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2010, 101(2): 110 – 122.

[5] CHANG Pei, WANG Pu, GAO Xuejin. Batch process monitoring and quality prediction based on statistics pattern analysis and MKPLS. *Chinese Journal of Scientific Instrument*, 2014, 35(6): 1409 – 1416.
(常鹏, 王普, 高学金. 基于统计模式分析与MKPLS的批处理过程监控及质量预测. 仪器仪表学报, 2014, 35(6): 1409 – 1416.)

[6] CHANG Pei, WANG Pu, GAO Xuejin. T–PLs intermittent process fault monitoring based on statistical analysis. *Journal of Chemical Industry*, 2015, 66(1): 265 – 271.
(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. 化工学报, 2015, 66(1): 265 – 271.)

[7] GAO Y, KONG X, HU C, et al. Multivariate data modeling using modified kernel partial least squares. *Chemical Engineering Research and Design*, 2015, 94(1): 466 – 474.

[8] HU Y, MA H, SHI H. Enhanced batch process monitoring using just-in-time-learning based kernel partial least squares. *Chemometrics and Intelligent Laboratory Systems*, 2013, 123(1): 15 – 27.

[9] MORI J, YU J. Quality relevant nonlinear batch process performance monitoring using a kernel based multiway non-Gaussian latent subspace projection approach. *Journal of Process Control*, 2014, 24(1): 57 – 71.

[10] ZHANG Xiaoling, TIAN Xuemin. Study on the method of batch process monitoring based on nonlinear multidirectional ICA. *Journal of System Simulation*, 2009, 21(11): 3365 – 3369.
(张晓玲, 田学民. 基于非线性多向ICA的间歇过程监控方法研究. 系统仿真学报, 2009, 21(11): 3365 – 3369.)

[11] ZHANG Y, QIN S J. Fault detection of nonlinear processes using multiway kernel independent component analysis. *Industrial & Engineering Chemistry Research*, 2007, 46(23): 7780 – 7787.

[12] RASHID M M, YU J. Nonlinear and non-Gaussian dynamic batch process monitoring using a new multiway kernel independent component analysis and multidimensional mutual information based dissimilarity approach. *Industrial & Engineering Chemistry Research*, 2012, 51(33): 10910 – 10920.

[13] FAN J, QIN S J, WANG Y. Online monitoring of nonlinear multivariate industrial processes using filtering KICA–PCA. *Control Engineering Practice*, 2014, 22(1): 205 – 216.

[14] ZHAO C, GAO F, WANG F. Nonlinear batch process monitoring using phase based kernel-independent component analysis principal component analysis. *Industrial & Engineering Chemistry Research*, 2009, 48(20): 9163 – 9174.

[15] LEE J M, CHANG Y. LEE B. Fault detection of batch process using multiway kernel principal component analysis. *Computers and Chemical Engineering*, 2004, 28(9): 1837 – 1847.

[16] CAI Lianfang, TIAN Xuemin, ZHANG Ni. A kernel time structure independent component analysis method for nonlinear process monitoring. *Chinese Journal of Chemical Engineering*, 2014, 22(11): 1243 – 1253.
(蔡连芳, 田学民, 张妮. 一种用于非线性过程监测的核时间结构独立成分分析方法. 中国化学工程学报, 2014, 22(11): 1243 – 1253.)

[17] CAI L, TIAN X, CHEN S. A process monitoring method based on nosiy independent component anlaysis. *Neurocomputing*, 2014, 127(3): 231 – 246.

[18] ZHANG Y W, AN J Y, ZHANG H L. Monitoring of time varying process using kernel independent compoent analysis. *Chemical Engineering Science*, 2013, 88(4): 23 – 32.

[19] ROBERT J. Kernel entropy component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(5): 847 – 860.

[20] CHANG P, QIAO J F, WANG P. The monitoring of non-gaussian and non-linear coexistence intermittent process based on MKECA. *Journal of Chemical Industry*, 2018, 69(3): 1200 – 1206.
(常鹏, 乔俊飞, 王普. 基于MKECA的非高斯性和非线性共存的间歇过程监测. 化工学报, 2018, 69(3): 1200 – 1206.)

[21] JIANG Q, YAN X, LV Z, et al. Fault detection in nonlinear chemical processes based on kernel entropy component analysis and angular structure. *Korean Journal of Chemical Engineering*, 2013, 30(6): 1181 – 1186.

[22] CHANG Pei, WANG Pu, GAO Xuejin. Monitoring of microbial fermentation batch process based on the analysis of polynuclear entropy components. *College Chemical Engineering Journal*, 2015, 29(2): 395 – 399.
(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. 高校化学工程学报, 2015, 29(2): 395 – 399.)

[23] PORAT B, FRIEDLANDER B. Direction finding algorithms based on high-order statistics. *IEEE Transactions on Signal Processing*, 1991, 39(9): 2016 – 2024.

[24] CAI L, THORNHILL N F, PAL B C. Multivariate detection of power system disturbances based on fourth order moment and singular value decomposition. *IEEE Transactions on Power Systems*, 2017, 32(6): 4289 – 4297.

[25] CARDOSO J F. High-order contrasts for independent component analysis. *Neural Computation*, 1999, 11(1): 157 – 192.

[26] PORAT B, FRIEDLANDER B. Direction finding algorithms based on high-order statistics. *IEEE Transactions on Signal Processing*, 1991, 39(9): 2016 – 2024.

[27] MUSTAPHA H, DIMITRAKOPOULOS R. Geologically enhanced simulation of complex mineral deposits through high-order spatial cumulants. *Advances in Applied Strategic Mine Planning*. Cham: Springer, 2018.

[28] SHEATHER S J, JONES M C. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, 1991, 53(3): 683 – 690.

[29] YANG C, DURAISWAMI R, GUMEROV N A, et al. Improved fast gauss transform and efficient kernel density estimation. *Proceedings of 2003 IEEE International Conference on Computer Vision*. Nice, France: IEEE, 2003: 664 – 671.

[30] TERRELL G R, SCOTT D W. Variable kernel density estimation. *Annals of Statistics*, 1992, 20(3): 1236 – 1265.

作者简介:

常 鹏 讲师, 博士, 目前研究方向为间歇过程统计建模、过程监测、故障诊断, E-mail: changpeng@bjut.edu.cn;

乔俊飞 教授, 博士生导师, 目前研究方向为污水处理过程知识自动化, E-mail: junfeiqiao@bjut.edu.cn;

张祥宇 硕士研究生, 目前研究方向为间歇过程统计建模、过程监测、故障诊断, E-mail: zhangxiangyu@bjut.edu.cn;

王 普 教授, 博士生导师, 目前研究方向为复杂工业过程建模与优化、复杂系统控制, E-mail: wangpu@bjut.edu.cn.