非线性和非高斯性共存的序批次反应处理过程故障诊断

常 鹏[1,2,3†], 乔俊飞[1,2], 王 普[1,3], 高学金[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: 序批式反应器(SBR)的处理过程的数据具有非高斯分布和高度非线性的特点, 传统特征提取方法在进行特征提取时仅仅考虑信息最大化而忽略数据的簇结构信息导致数据特征提取的不完整. 由于多向核熵成分分析是一种新的监测方法, 在监测过程中的应用表明能够克服传统监测方法的缺陷, 减少误报警率. 因此本文结合多向核熵成分分析的的优势, 提出多向核熵独立成分分析方法用于SBR过程监测及故障诊断. 首先, 将三维SBR过程数据利用一种新的数据展开技术变为二维数据; 其次, 利用核熵成分分析将展开后的二维数据映射到高维空间用独立成分分析进行独立成分提取; 最后提出一种基于多向核熵独立成分分析的故障诊断方法进行故障诊断. 将该方法和传统方法应用于80升的SBR处理过程的监测结果表明, 本文提出的方法优于传统的多向独立成分分析方法.

关键词: 序批式反应器; 多向核熵独立成分; 故障检测; 故障诊断

引用格式: 常鹏, 乔俊飞, 王普, 等. 非线性和非高斯性共存的序批次反应处理过程故障诊断. 控制理论与应用, 2019, 36(5): 728 – 736

DOI: 10.7641/CTA.2018.70657

Fault diagnosis of sequential batch reaction process with nonlinear and non-Gaussian coexistence

CHANG Peng[1,2,3†], QIAO Jun-fei[1,2], WANG Pu[1,3], GAO Xue-jin[3]

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

Abstract: The data of sequencing batch reactor (SBR) has characteristics of non-Gaussian distribution and high nonlinearity, In order to solve the problem that SBR process monitoring algorithm can only maximize the use of data information and ignore the information in the structure of data cluster, a new multi-way kernel entropy component analysis (MKECA) method is proposed. It also address the shortcomings of the traditional monitoring method in omission failure rate. A novel contribution analysis scheme named bar plot is developed for MKEICA to diagnose faults. The proposed MKEICA method consist of three steps: 1) the three-dimensional data of SBR is unfolded into two-dimensional by a new data expanding method. 2) kernel entropy principal component analysis (KEPCA) is adopted to map the two-dimensional data into a high dimensional feature space and use independent component analysis (ICA) to extract independent components (ICs) in feature space. 3) in the stage of online monitoring,bar plot is used to identify the variables causing the fault. This method is successfully applied to an 80 L lab-scale SBR, and the experimental results demonstrate that, comparing with traditional MKICA, the proposed MKEICA method exhibits better performance in fault detection and diagnose.

Key words: sequencing batch reactor; multi-way kernel entropy independent component; fault detection; fault diagnosis

Citation: CHANG Peng, QIAO Junfei, WANG Pu, et al. Fault diagnosis of sequential batch reaction process with nonlinear and non-Gaussian coexistence. *Control Theory & Applications*, 2019, 36(5): 728 – 736

1 Introduction

As a flexible and low-cost process, sequencing batch reactors (SBRs) are commonly used for biological waste water treatment. The SBR's process is normally operated in a series of predefined phases: fill, react, settle, draw, and idle[1]. The advantage of the SBR's process is that its single-tank designs and the flexibility enable it to meet different treatment objectives[2]. In consideration of the increasingly stringent regulations of effluent quality, the on-line monitoring of SBR's processes shows its great significance in enhancing process performance by detecting disturbances resulting to abnormal process operation at an early stage. In recent years, several methods based on multivariate statisti-

Received: 12 September 2017; accepted 22 May 2018.
†Corresponding author. E-mail: changpeng@bjut.edu.cn; Tel.: +86 10-67391865.
Recommended by Associate Editor: WANG Wei.
Supported by the National Natural Science Foundation of China (61364009, 61174109) and the Beijing Postdoctoral Research Foundation.

cal analysis have been developed for online monitoring of the SBR's process. However, the SBR's process is highly non-linear and non-Gaussian from the perspective of nonlinear biological reaction kinetics[3]. Thus, real-time online monitoring of the SBR's process is a giant challenge.

More recently, fault detection based on independent component analysis (ICA) has become a hot topic[4–15]due to ICA's assumption that latent variables do not contain a Gaussian distribution, which is consider as an extension of the principal component analysis(PCA). The method exploits higer-order statistical information to extract mutually independent latent variables called independent components (ICs) from the non-Gaussian process. Therefore, ICA is specially suitable for fault detection on non-Gaussian process. In spite of attractive advantages of ICA, it is naturally a linear statistical method and can only be adopted to separate linearly mixed ICs. But when process data are nonlinear, ICA would be unable to recover ICs from process data, leading to insufficient feature extraction. Kernel ICA (KICA)[16–24] is emerged to come up with the nonlinear problems which is essentially kernel PCA plus ICA. Normal operating data whitened by KPCA (kernel PCA) and ICA is performed to search for ICs. In light of superior capability of KICA in tackling with nonlinear behavior of process. Lately, fault detection based on MKICA has been focused, the multiway kernel ICA (MKICA) methods uses higher order statistics like negentropy or mutual information and extracts more information to obtain the independent components. Zhang ect.[16] utilized KICA to extract some dominant nonlinear independent component from nonlinear batches and proposed their usage for process monitoring. Tian ect.[17] proposed to use KICA based on feature samples as to reduce computational cost. Cai ect.[22] put forward a modified KICA method to monitor nonlinear time-varying processes. The main part of KICA, commonly used for linear ICA, adopts the maximum non-Gaussian criterion for the optimization objective of IC feature extraction. If there're two or more ICs in Gaussian distribution, the criterion cannot effectively distinguish their distribution characteristics. On a more practical level, Fan ect.[23] applied filter KICA–PCA for nonlinear batch process monitoring, Zhao ect.[24] combined MKICA and MKPCA to improve the performance of fault detection in nonlinear cases.

KPCA based data whiten solely focusing on maximizing data variance and ignore the phase process data information. Robert Jenssen ect.[25–26], proposed a new Kernel entropy component analysis method for data transformation and dimensionality reduction, it based on kernel entropy to extract data information. the stage of treatment of bleaching dealing with damage characteristics of original data can guarantee the angle structure data information. The KECA algorithm realizes data transformation and dimensionality by projecting the input space to the KPCA axes, and its largest difference from KPCA is to select the principal element[15] by the contribution of the entropy value of the input space. Some scholars have applied KECA on process monitoring[27–30], and have proven the useful in chemical processes modeling. Several extension have been proposed for feature selection[25–26] class dependent feature extraction. And Semi supervised learning as well. In this article, we use a novel spectral data transformation method, which is fundamentally different from other spectral methods in two very important ways as follows:

1) KECA algorithm project input space onto KPCA axes to realize data transformation and dimension reduction. The principal component selected by KECA has a certain angle structure.

2) The linear ICA is extended to the nonlinear domain.

The remainder of the paper is organized as follows: In Section 2, MKECA and MKEICA is introduced; Section 3 process falut monitoring using MKEICA for fault diagnosis is developed; Finally, Section 4 concludes this article.

2 Preliminary

2.1 Kernel entropy component analysis

The Renyi quadratic entropy[25–26] is given by

$$H(p) = -\log \int p^2(x)\mathrm{d}x, \tag{1}$$

where $p(x)$ is the probability density function generating the data set, or sample, $D = [x_1 \ \ x_2 \ \ \cdots \ \ x_n]$ Since the logarithm is a monotonic function，we may concentrate on the quantity:

$$V(p) = -\log \int p^2(x)\mathrm{d}x. \tag{2}$$

In oder to estimate $V(p)$ and $H(p)$, we may invoke a Parzen window density estimation described as

$$\hat{p}(x) = \frac{1}{N}\sum_{x_t \in D} k_\sigma(x, x_i), \tag{3}$$

where $k_\sigma(x, x_i)$ is the so-called Parzen window, or kernl, centred at x_i and with a width governed by the parameter σ. Moreover, the Renyi entropy estimator may be expressed in terms of the eigenvalues and eigenvectors of the kernel matrix, which may be eigen decomposed as $K = EDE^{\mathrm{T}}$, where as before, D is a diagonal matrix storing the eigenvalues $\lambda_1, \cdots, \lambda_N$ and E is a matrix with the corresponding eigenvectors $e_1, \cdots, e_N$ as columns. We then have been rewriting:

$$\hat{V}(p) = (\frac{1}{N^2}\sum_{i=1}^{N}\sqrt{\lambda_i}e_i^{\mathrm{T}}1). \tag{4}$$

Each term in Eq. (2) will contribute to the entropy estimate. This means that certain eigenvalues and eigenvec-

tors will contribute more to the entropy estimate than others since the terms depend on different eigenvalues and eigenvectors. The eigenvalues and eigenvectors selected are the first d largest contribution to the entropy estimate in KECA so that the cumulative contribution rate of the selected Renyi entropy reaches 85% of all the Renyi entropy. Note that kernel PCA performs dimensionality reduction by selecting eigenvalues and eigenvectors solely based on the size of the eigenvalues, and the resulting transformation may be based on uninformative eigenvectors from an entropy perspective.

2.2 Kernel entropy independent component analysis

Assuming that the observation vectors $x_1, x_2, \cdots, x_n$ are non-linearly mapped into the feature space F, the covariance matrix[12] of the mapped data can be expressed by

$$C^{\mathrm{F}} = \frac{1}{n}\Phi\Phi^{\mathrm{T}}. \tag{5}$$

Assume for the time being that the mapped data are centred. Define $Q = [\Phi(x_1)\ \ \Phi(x_2)\ \ \cdots\ \ \Phi(x_n)]$. Define the the Gram matrix $\boldsymbol{K} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$.

Whose elements are defined by the given kernel function $\boldsymbol{k_{ij}}$ as follows:

$$\boldsymbol{k_{ij}} = \boldsymbol{\Phi}(\boldsymbol{x_i})\boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{x_j}) = \boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j}). \tag{6}$$

The polynomial, sigmoid and radial basis kernels are widely for $\boldsymbol{k_{ij}}$. In this paper, the radial basis kernel is used:

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) = \exp[-\|x - y\|^2/2\sigma^2], \tag{7}$$

where σ is the Gaussian kernel width. The application of the eigen decompoistion of the Gram matrix leads to $K = V\Lambda V^{\mathrm{T}}$, where $V \in K^{n\times d}$ is the matrix containing the orghonormal eigenvectors of K and $\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_d\}$ is the diagonal matrix consisting of the eigenvalues of K. The eigenvalues and eigenvectors selected are the first d largest to the variance in KECA so that the cumulative contribution rate of the selected reaches 85% of all the variances.

$$\sum_{i=1}^{d}\hat{V}(p)_i / \sum_{i=1}^{n}\hat{V}(p)_i \times 100\% \geqslant 85\%. \tag{8}$$

Thus the expression of C^{F} in the case of $K = Q^{\mathrm{T}}Q$ is as follows:

$$C^{\mathrm{F}} = (QV\Lambda^{-1/2})\frac{\Lambda}{n}(QV\Lambda^{-1/2})^{-1}. \tag{9}$$

As mentioned previously, the mapped data in the feature space F should be centred before application of the KECA projection. But it is difficult to centre the data in F because the nonlinear map is unknown. However, a slight modification of notation circumvents this difficulty, and one can centre the matrix K and the vector k_{new} respectively, as follows[1–12]:

$$\bar{K} = K - L_{\mathrm{N}}K - KL_{\mathrm{N}} + L_{\mathrm{N}}KL_{\mathrm{N}}, \tag{10}$$

$$\bar{K}_{\mathrm{new}} = K_{\mathrm{new}} - L_{\mathrm{new}}K - K_{\mathrm{new}}L_{\mathrm{N}} + L_{\mathrm{new}}KL_n \tag{11}$$

with

$$L_n = 1/n\begin{bmatrix}1 & \cdots & 1\\ \vdots & \ddots & \vdots\\ 1 & \cdots & 1\end{bmatrix} \in z^{n\times n},$$

$$L_{\mathrm{new}} = 1/n[1\ \ \cdots 1] \in z^n.$$

The whitened score is

$$z = \sqrt{n}\Lambda^{-1}V^{\mathrm{T}}K^{\mathrm{T}}. \tag{12}$$

The next task is to implement the ICA algorithm:

$$\hat{s} = W \cdot \bar{z}. \tag{13}$$

From $E\{\hat{s}\hat{s}^{\mathrm{T}}\} = WE\{\bar{z}\bar{z}^{\mathrm{T}}\}W^{\mathrm{T}} = WW^{\mathrm{T}} = I$, we get the conclusion that W is an orthogonal matrix which can of course be obtained by the fast ICA algorithm[12]. Assume that the most dominant independent components $\hat{s}_m$ are extracted, and denote the corresponding de-mixing matrix as W_m. Then

$$\hat{s}_m = W_m \cdot \bar{z}, \tag{14}$$

where the independent components s have the maximized statistical independency in terms of entropy among each other. Given a new sample vector x_{new}, and its corresponding independent component can be expressed as

$$\hat{s}_{\mathrm{new}} = W_m \cdot \bar{z}_{\mathrm{new}}. \tag{15}$$

Furthermore, the I^2 and SPE statistics can be defined process monitoring as follows:

$$I^2 = \hat{s}_m^{\mathrm{T}}\hat{s}_m \cdot \sum_{p=1}^{m}\hat{s}_p^{\mathrm{T}}\hat{s}_p, \tag{16}$$

$$\mathrm{SPE} = e^{\mathrm{T}}e \cdot \sum_{q=1}^{m} e_q^{\mathrm{T}}e_q, \tag{17}$$

where $e = (L - A_mW_m)\bar{K}$, with I denotes an identify matrix A_m is the inverse matrix W_m. Given a new sample vector x_{new}, its corresponding and SPE_{new} statistics can be defined process monitoring as follows:

$$I^2_{\mathrm{new}} = \hat{s}^{\mathrm{T}}_{\mathrm{new}}\hat{s}_{\mathrm{new}} \cdot \sum_{p=1}^{m}\hat{s}^{\mathrm{T}}_{\mathrm{new},p}\hat{s}_{\mathrm{new},p}, \tag{18}$$

$$\mathrm{SPE}_{\mathrm{new}} = e^{\mathrm{T}}_{\mathrm{new}}e_{\mathrm{new}} \cdot \sum_{q=1}^{m} e^{\mathrm{T}}_{\mathrm{new},q}e_{\mathrm{new},q}, \tag{19}$$

where $e_{\mathrm{new}} = (L_{\mathrm{new}} - A_mW_m)\bar{K}_{\mathrm{new}}$, the confidence limits for the two statistics above can be estimated through kernel density estimation strategy[10]. Conventional statistical process monitoring methods like ICA lack the ability to monitor non-Gaussian processes for three main reasons. Firstly, the ICA method only uses lower-order statistics, such as variance, to monitor processes. Variance is a lower-order statistic, and hence it is only sensitive to amplitude instead of sensitive to

phase. However, phase and amplitude are equally important in monitoring non-Gaussian processes. Secondly, the ICA method only utilizes higher-order statistics to extracting the ICs rather than monitor the process.

3 Process monitoring based on KEICA

3.1 Three dimensional data of the SBR

As a typical batch process, the data of SBR is a three-dimensional array $X(I \times J \times K)$, where I, J, and K are numbers of batches, variables, and samples, respectively. Since the multivariate statistical analysis methods can only model for two-dimensional data, first of all, SBR's three-dimensional data should be transformed into two-dimensional. The original method is to unfold along batch[12], which means to unfold the three-dimensional array to a matrix, and then normalize the matrix according to Equations (20)–(22). The disadvantage of this method is that the future value of the current batch must be estimated for on-line monitoring.

$$\tilde{x}_{i,kj} = \frac{x_{i,kj} - \bar{x}_{kj}}{s_{kj}}, \tag{20}$$

$$\bar{x}_{kj} = \frac{1}{I}\sum_{i=1}^{I} \tilde{\tilde{x}}_{i,kj}, \tag{21}$$

$$s_{kj} = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I} \left(\tilde{\tilde{x}}_{i,kj} - \bar{x}_{kj}\right)^2}. \tag{22}$$

To avoid the drawback of unfolding along batch, Lee et al.[14] proposed a new unfolding method, which firstly unfolds the array $X(I \times J \times K)$ to a matrix $X(I \times JK)$, and normalizes the matrix according to Equations (20)–(22), and then unfolds the normalized matrix to $\bar{X}(KI \times J)$ along variable.

3.2 Fault diagnose

Once the fault is detected by the MKICA method, the next step is to find out the root cause variables. Fault reconstruction and contribution plots are two commonly used fault diagnosis methods, but both method need a number of historical fault, which limits its application. Therefore, this paper considers a new contribution analysis scheme, bar graph, which can be conveniently applied to kernel methods without historical failure data.

The centroid matrix $C \in \mathbb{R}^{K \times J}$ of $\bar{X}(KI \times J)$ which is obtained in section "Processing of SBR's three-dimensional data" can be calculated by the following equation:

$$\bar{x}_{kj} = \frac{1}{I}\sum_{i=1}^{I} \tilde{\tilde{x}}_{i,kj}. \tag{23}$$

Assuming that x_k the test data at sample time k, The steps to find a fault source using the bar plot method are introduced as follows. First of all, the row vector B ($B \in \mathbb{R}^J$) at k sample time is take from C and let $j = 1$. Repeat the following 1)–5) steps until $j > J$.

1) Let $\bar{B} = B$.

2) The value of jth variable in $\bar{B}$ is replaced by the value of jth variable in x_k.

3) The new data $\bar{B}$ is monitored by MKEICA again, and we can obtain the new $I^2(j)$ and SPE(j) statistics.

4) Let the new obtained $I^2(j)$ and $SPE(j)$ subtract the control limits at k sample time respectively, and the results can be expressed as $D_{I^2}(j)$ and $D_{\mathrm{SPE}}(j)$.

5) Let $j = j + 1$.

When the Jth variables of $D_{I^2}(j)$ and $D_{\mathrm{SPE}}(j)$ ($D_{I^2} \in \mathbb{R}^J$, $D_{\mathrm{SPE}} \in \mathbb{R}^J$) are achieved according to the above steps, we can judge whether the $D_{I^2}(j)$ and $D_{\mathrm{SPE}}(j)$ statistics are greater than zero respectively. If the results are greater than zero, the Jth variable can be regarded as a fault variable; or else, it will be regarded as a normal one.

3.2.1 Off-line modeling

1) Acquire an operating normal data. Unfold $\underline{X}(I \times J \times K)$ to $X(I \times KJ)$ and normalize the matrix according to Equation (1), then unfold the normalized matrix to $\bar{X}(KI \times J)$ along variable.

2) Obtain the centroid matrix C according to Equation (23). Calculate the gram Kernel matrix G according to Equation (4). Then, center and normalize G according to Equations (11) and (12).

3) Apply eigenvalue decomposition to $\tilde{G}_{\mathrm{scl}}$ according to Equation (8). Select d positive eigenvalues according to Equation (10).

4) Extract the independent component matrix S by the ICA model which is obtained by Equation (11), and then p primary ICs are chosen to construct primary component space S_p.

5) Construct two monitoring statistics I^2 and SPE according to Equations (19) and (20) respectively. Carry out the kernel density estimation method to compute the control limits of I^2 and SPE.

3.2.2 On-line monitoring

1) For the k time data, $x_k(1 \times J)$, normalize it and achieve the kernel vector. Obtain the new independent component s_k and the new residual e_k according to Equation (19) and Equation (20). Calculate the new ${I^2}_k$ and ${SPE}_k$ statistics and judge them whether surpass their control limits. If either of them surpasses the control limit, execute Step 4; else, execute Step 4.

2) Take out the row vector $B(B \in \mathbb{R}^J)$ at the k sample time from C and let $j = 1$. let $\bar{B} = B$, replace the value of jth variable in $\bar{B}$ by the value of jth

variable in x_k; monitoring the replaced new data $\bar{B}$ by MKEICA again, then we obtain the new monitoring statistics of I^2 and SPE.

3) Let the new acquired I^2 and SPE subtract their control limits at the k samples, and the results can be expressed as the $D_{I^2}(j)$ and $D_{\rm SPE}(j)$.

4) Let $j = j + 1$ jump to Setp 2; until $j \geqslant J$. Let $k = k + 1$; if $k \leqslant K$, go to Step 1; or else finish.

3.3 Case study

3.3.1 Lab-scale plant SBR

The SBR data used in this article comes from a pilot-scale SBR system. A fill-and-draw SBR with an 80-liter working volume is operated in a 6 h cycle mode and each cycle consists of fill/anaerobic (60 min), aerobic (150 min), anoxic (60 min), re-aerobic (30 min) and settling/draw (60 min) phases. Temperature (T), PH, oxidation reduction potential (ORP), the first derivative of ORP (dORP/dt), dissolved oxygen (DO), six electrodes for conductivity, weight (W) in Table 1 are connected to the individual sensors to check the status of the SBR. A set of on-line measurements is obtained every 3 minutes (120 times points per cycle). The experiment lasted for 20 days, and 80 batches of data are collected among which two batches fault data are included. Fault 1 is a single fault. Aeration system should have been stopped when the process moved to the third stage (anoxic) from the second stage (aerobic). But in fact, the aeration system was broken，causing to abnormal rise of DO. This fault lasted from 211 min (sample number 71) to 290 min (sample number 90). Fault 2 is a multiple fault, which is artificially added as a sensor fault on the basis of Fault 1. The value of PH sensor no longer changes since 131 min (sample number 44) until the end of the reaction, namely the value remains at 130 min. As biological reactions in the settling and drawing phases were assumed to be negligible, the first 100 sampling time instants were only used. 60 batches of normal data are selected to build model, and the two fault batches described above are utilized to simulate monitoring on-line. Traditional MKICA and our proposed method are experimented respectively.

Table 1 List of variables used for the monitoring of process

Symbol	Variables
x_1	Temperature (T)
x_2	Ph
x_3	Oxidation reduction potential (ORP)
x_4	The first derivative of ORP ((dORP/dt))
x_5	Dissolved oxygen (DO)
x_6	Six electrodes for conductivity
x_7	weight (W)

3.3.2 Analysis of the results

60 batches of normal operating data $\underline{X}(60 \times 7 \times 100)$ are used to build MKICA and MKEICA monitoring model, and Kernel Density Estimation method is utilized to compute the control limits with 99% confidence limits, and then two faulty batches data are monitored respectively. Three variables 1, 3, 5 are selected for non-Gaussian test. The results show that the processing data deviate from the red line, indicating that the data have non Gaussian characteristics from Fig. 1.

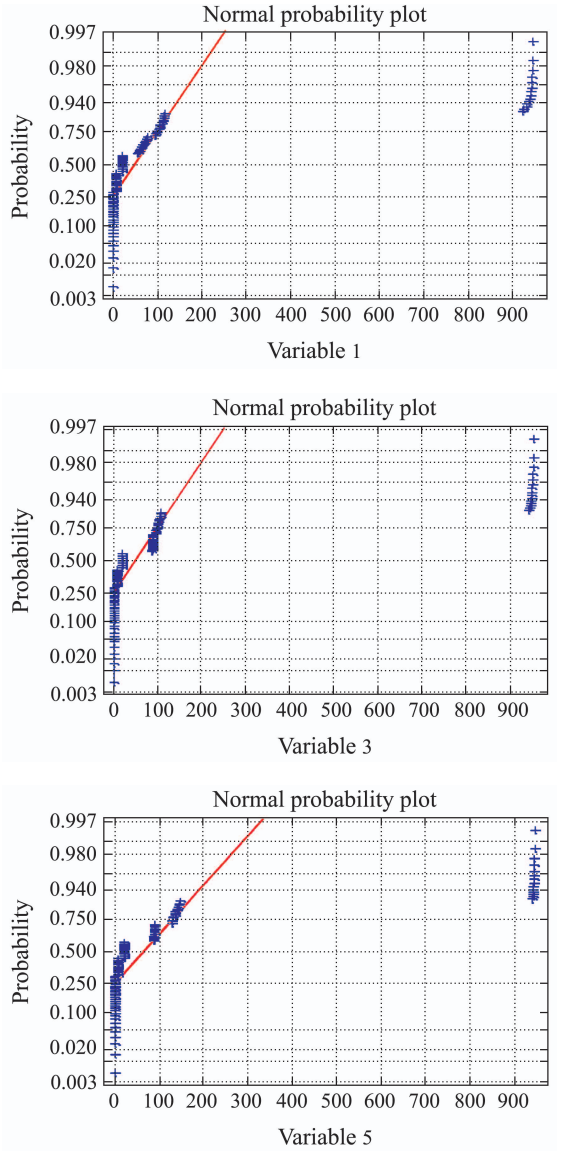

Fig. 1 1, 3, 5 normal test of process variables

Figure 2 shows the online monitoring results of MKICA and MKEICA for the fault 1 batch. The fault lasted from sample number 70th to 90th. Although there is a false alarm point at about 70th sample in the chart of MKICA and MKEICA, overall, both methods' charts show fine fault detection performance. Nevertheless, the SPE chart of MKICA occurs a number of leakage alarms between sample number 75th and 87th. The fault detection performance is much less than the SPE chart of MKICA.

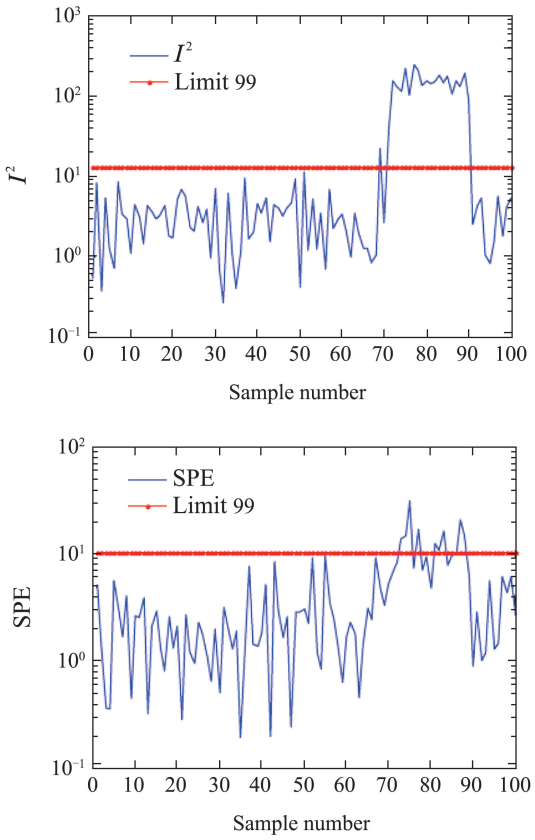

Fig. 2(a) Monitoring charts for MKICA

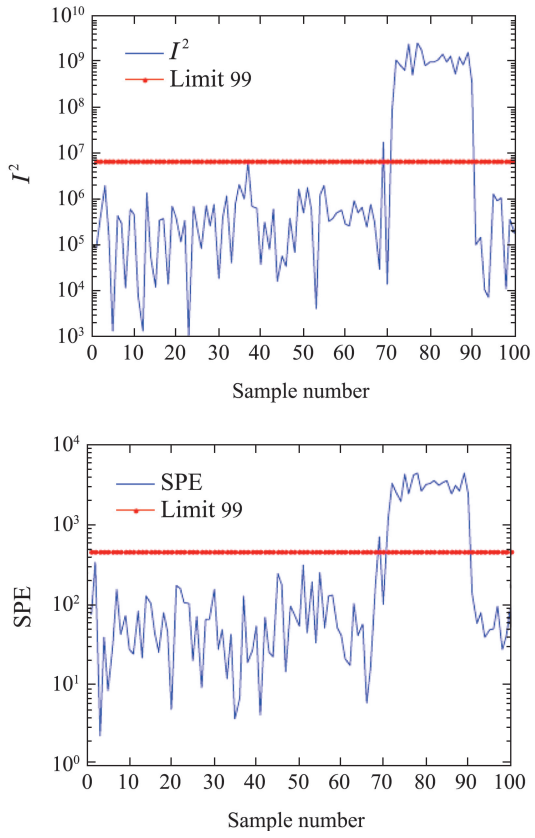

Fig. 2(b) Monitoring charts for MKEICA

Figure 3 shows the online monitoring results of the and SPE charts of MKICA and MKEICA for the fault 2 batch. Fault 2 is a multiple fault, which is caused by two variables. There into, the first PH sensor fault appeared in the sample number 44th. This PH sensor fault is a simulation of the value that no longer change, thus the fault magnitude shows less at the initial stage, and it is unable to detect the fault yet. Both and SPE charts of M-KICA start to detect the occurrence of failure at sample number 48th, and then they give an alarm till the end of the process. Whereas the chart of MKICA detects the occurrence of failure at sample number 49th. Even worse, some leak alarms occur at sample number 50th, 55th and 97th. In addition, the SPE chart of MKICA detects the occurrence of failure starting at sample number 48th, but leaks a point at sample number 49th. In summary, in consideration of the nonlinear problem of the process variables, MKEICA shows better performance than MKICA both at the timeliness of fault detection aspects and at the occurrence of false alarm and alarm failure aspects. Once a fault is detected, fault diagnose comes next. The traditional contribution plot considers the variable having the biggest contribution to be faulty source while our proposed bar plot method considers all variables greater than zero as fault variables.

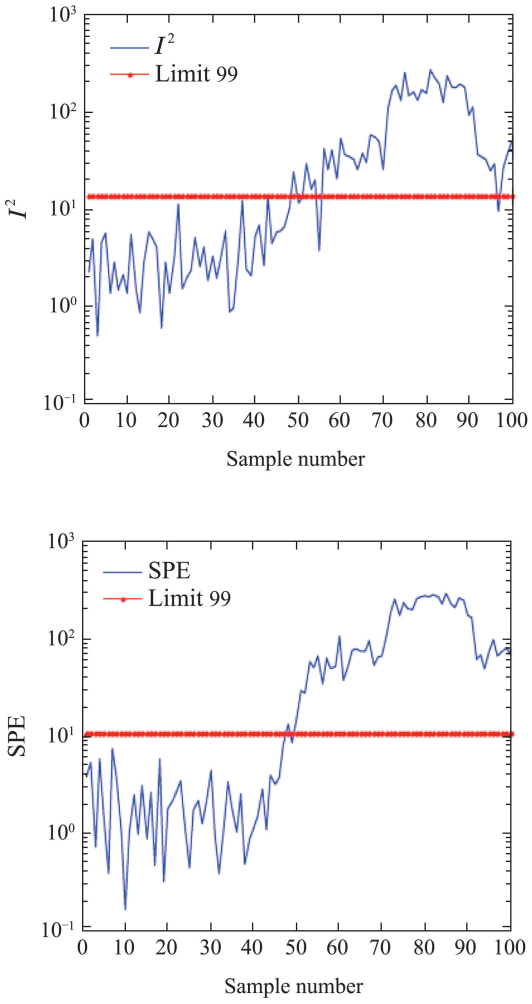

Fig. 3(a) Monitoring charts for MKICA

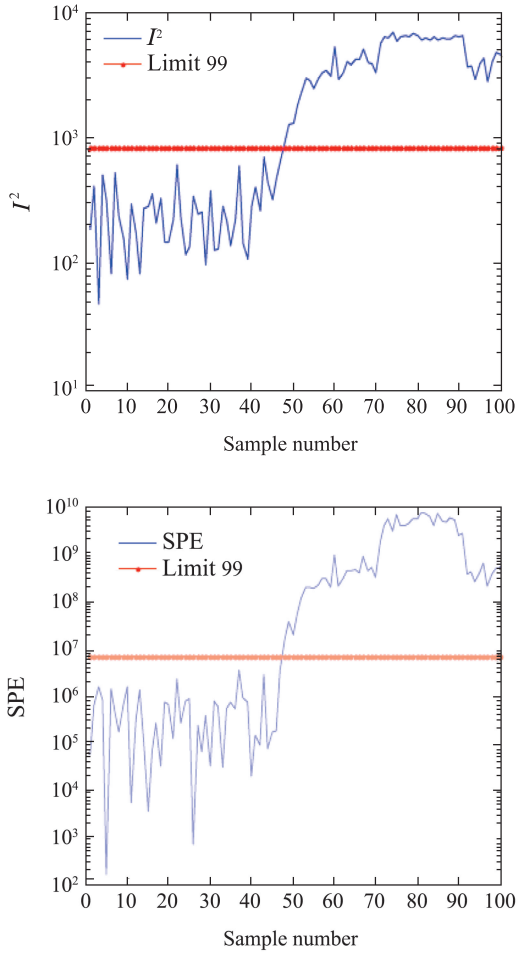

Fig. 3(b) Monitoring charts for MKEICA

Figure 4 is the contribution plots of MKEICA for fault 1 at sample number 76th.

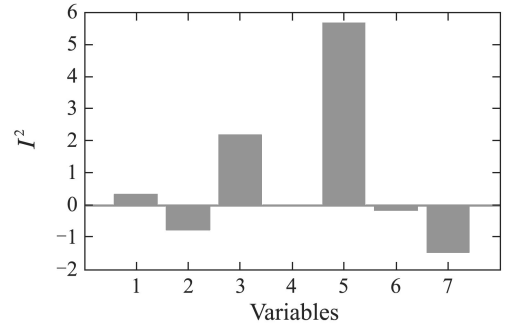

Fig. 4(a) Contribution plots of I^2

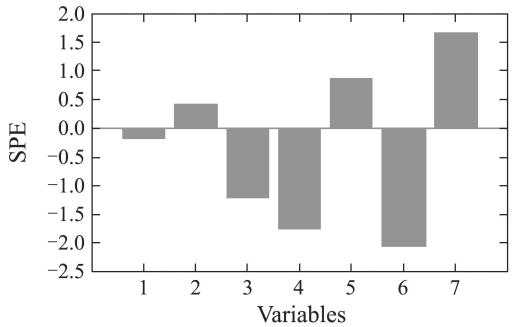

Fig. 4(b) Contribution plots of SPE

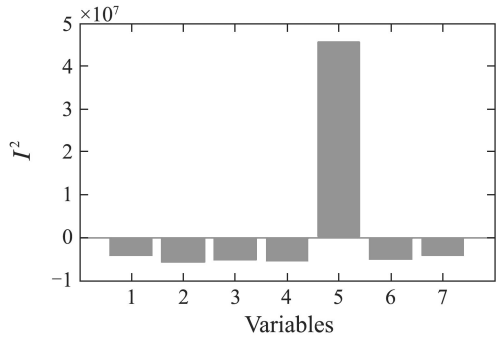

Fig. 4(c) New bar plots of I^2 for fault 1 at sample 75th

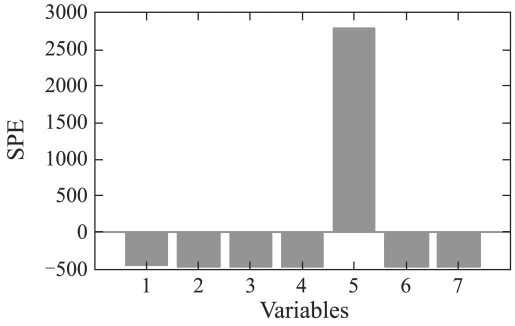

Fig. 4(d) New bar plots of SPE for fault 1 at sample 75th

Figure 4(a) indicates that variable 5 (DO) is the fault variable, which is right. However, Fig. 4(b) indicates that variable 7 (W) is the fault variable, which is false. Fig. 4 is the new bar plots of MKEICA for Fault 1 at sample number 75th. Both Fig. 4(c) and Fig. 4(d) indicate that variable 5 (DO) is the fault variable, which is right.

Figure 5 is fault diagnosis charts of MKEICA for Fault 2 at sample number 82th. Fault 2 is a multiple fault at this sample number. From Fig. 5(c) and Fig. 5(d), we can see clearly that the values of variable 2 (Ph) and variable 5 (DO) are greater than zero, so they are fault variables. Our proposed method can completely and accurately identify the two faulty sources while the traditional contribution plot normally considers the variable having the biggest contribution to be faulty source. Fig. 5(a) and Fig. 5(b) can only identify variable 2 (Ph) as the fault variable, but neglect another fault variable, namely variable 5 (DO).

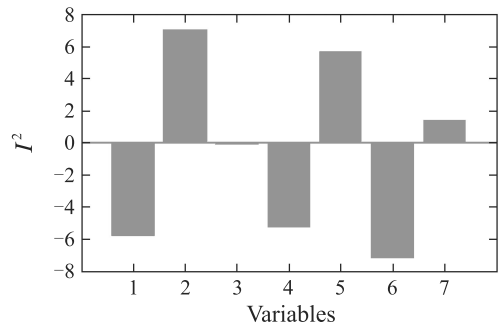

Fig. 5(a) Contribution plots of I^2

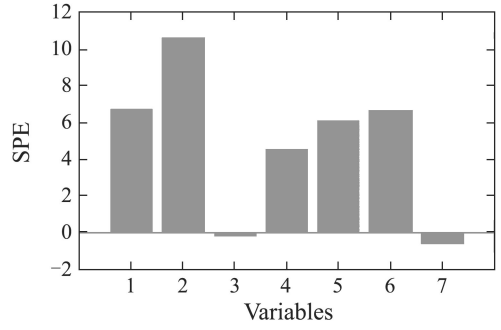

Fig. 5(b) Contribution plots of SPE

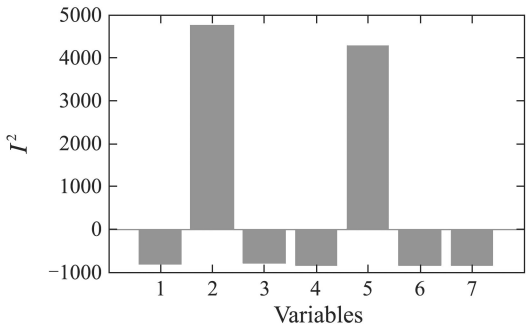

Fig. 5(c) New bar plots of I^2 for Fault 1 at sample 85th

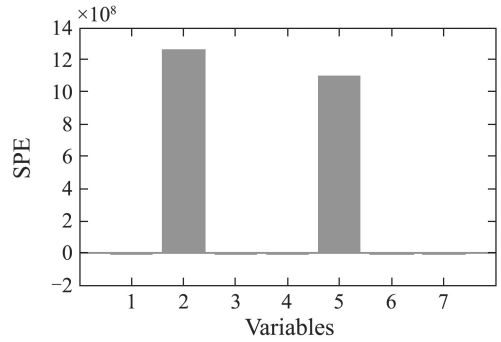

Fig. 5(d) New bar plots of SPE for Fault 1 at sample 85th

4 Conclusions

In this article, we proposed a novel monitoring strategy based on MKEICA for the online detection fault of batch process. The data mapped into feature space become redundant and linear data introduce error when the kernel trick is used. In addition, in the training process of kernel principal component analysis (KPCA), the eigenvalues and eigenvectors selected are the first largest contribution to entropy estimate in KECA so that the cumulative contribution rate of the selected Renyi entropy reaches 85% of all the Renyi entropy. Unlike other kernel feature extraction methods, the top eigenvalues and eigenvectors of the kernel matrix are not necessarily chosen. Data are interestingly mapped with a distinct angular structure, which is exploited to derive a new angle-based spectral clustering algorithm based on the mapped data. The method takes advantage of higer-order statistics, which are more sensitive to faults. Consequently, the proposed approach can effectively capture the nonlinear non-gaussian relationship among the process variables and its application to process monitoring shows better performance. The case study on the Penicillin fermentation process demonstrates the method of MKEICA is considerably more effective than conventional method of MKICA in fault detection and diagnosis. With further development, the proposed method will be useful to nonlinear non-gaussian SBR's process monitoring.

References:

[1] MARSILI-LIBELLI S, SPAGNI A, SUSINI R. Intelligent monitoring system for long-term control of sequencing batch reactors. *Water Science and Technology*, 2008, 57(3): 431 – 436.

[2] KYOO Y C, INBEUM L, VANROLLEGHEM A. Multivariate nonlinear statistical process control of a sequencing batch reactor. *Journal of Chemical Engineering of Japan*, 2006, 39(1): 43 – 51.

[3] LEE J M, QIN S J, LEE I B. Fault detection of nonlinear processes using kernel independent component analysis. *The Canadian Journal of Chemical Engineering*, 2007, 85(4): 526 – 536.

[4] GE Z, SONG Z, GAO F. Review of recent research on data-based process monitoring. *Industrial and Engineering Chemistry Research*, 2013, 52(10): 3543 – 3562.

[5] VENKATASUBRAMANIAN V, RENGASWAMY R, KAVURI S N. A review of process fault detection and diagnosis, Part II: quanlitative models and search strategies. *Computers and Chemical Engineering*, 2003, 27(3): 327 – 346.

[6] KYOO Y C, INBEUM L, VANROLLEGHEM A. Multivariate nonlinear statistical process control of a sequencing batch reactor. *Journal of Chemical Engineering of Japan*, 2006, 39(1): 43 – 51.

[7] YAO Y, GAO F. A survey on multistage/multiphase statistical modeling methods for batch processes. *Annual Reviews in Control*, 2009, 33(2): 172 – 183.

[8] KYOO Y C, INBEUM L, VANROLLEGHEM A. Multivariate nonlinear statistical process control of a sequencing batch reactor. *Journal of Chemical Engineering of Japan*, 2006, 39(1): 43 – 51.

[9] GUO J Y, YUAN W Q. Palmprint recognition based on kernel principal component analysis and fisher linear discriminant. *Journal of Optoelectronics Laser*, 2008, 23(2): 354 – 358.

[10] RASHID M M, YU J. Hidden markov model based adaptive independent component analysis approach for complex chemical process monitoring and fault detection. *Industrial and Engineering Chemistry Research*, 2012, 51(15): 5506 – 5514.

[11] GUO J Y, YUAN W Q. Palmprint recognition based on kernel principal component analysis and fisher linear discriminant. *Journal of Optoelectronics Laser*, 2008, 23(2): 354 – 358.

[12] NOMIKOS P, MACGREGOR J F. Monitoring batch processes using multiway principal component analysis. *Aiche Journal*, 1994, 40(8): 1361 – 1375.

[13] KOURTI T, MACGREGOR J F. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics and Intelligent Laboratory Systems*, 1995, 28(1): 3 – 21.

[14] LEE J M, YOO C K, SANG W C, et al. Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 2004, 59(1): 223 – 234.

[15] JIA M, XU H, LIU X, et al. The optimization of the kind and parameters of kernel function in KPCA for process monitoring. *Computers and Chemical Engineering*, 2012, 46(22): 94 – 104.

[16] CHANG K Y, LEE J M, VANROLLEGHEM P A, et al. On-line monitoring of batch processes using multiway independent component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2004, 71(2): 151 – 163.

[17] WANG J, ZHANG Y, CAO H, et al. Dimension reduction method of independent component analysis for process monitoring based on minimum mean square error. *Journal of Process Control*, 2012, 22(2): 477 – 487.

[18] TIAN X, ZHANG X, DENG X, et al. Multiway kernel independent component analysis based on feature samples for batch process monitoring. *Neurocomputing*, 2009, 72(7/8/9): 1584 – 1596.

[19] FAN J, WANG Y. Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis. *Information Sciences*, 2014, 259(3): 369 – 379.

[20] ZHANG Y, AN J, ZHANG H. Monitoring of time-varying processes using kernel independent component analysis. *Chemical Engineering Science*, 2013, 88(2): 23 – 32.

[21] CAI L F, TIAN X M, ZHANG N. A kernel time structure independent component analysis method for nonlinear process monitoring. *Chinese Journal of Chemical Engineering*, 2014, 22(11/12): 1243 – 1253.

[22] CAI L F, TIAN X M, ZHANG N. A kernel time structure independent component analysis method for nonlinear process monitoring. *Chinese Journal of Chemical Engineering*, 2014, 22(11/12): 1243 – 1253.

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

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

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

[26] GOMEZ-CHOVA L, JENSSEN R, CAMPS-VALLS G. Kernel entropy component analysis for remote sensing image clustering. *IEEE Geoscience and Remote Sensing Letters*, 2012, 9(2): 312 – 316.

[27] CHANG P, QIAO J F, WANG P, et al. Monitoring non-Gaussian and non-linear batch process based on multi-way kernel entropy component analysis. *Journal of Chemical Industry*, 2018, 69(3): 1200 – 1206.

[28] ZHANG H, QI Y, WANG L, et al. Fault detection and diagnosis of chemical process using enhanced KECA. *Chemometrics and Intelligent Laboratory Systems*, 2017, 161(2): 61 – 69.

[29] CHANG P, WANG P, GAO X J. Microbial pharmaceutical process monitoring based on projection kernel entropy component. *Information and Control*, 2014, 43(4): 490 – 494.

[30] CHANG P, WANG P, GAO X J. Research on microbial fermentation process monitoring based on higher order cumulant statistics. *Control and Decision*, 2017, 32(12): 2273 – 2278.

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;

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;

高学金　副教授, 硕士生导师, 目前研究方向为基于数据驱动的工业过程建模, E-mail: gaoxuejin@bjut.edu.cn.