Chemical Engineering Science 64 (2009) 801 – 811

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM

Yingwei Zhang[a,b,*]

[a]Key Laboratory of Integrated Automation of Process Industry, Ministry of Education, Northeastern University, Shenyang, Liaoning 110004, PR China
[b]Department of Chemical Engineering, The University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Article history:
Received 18 June 2007
Received in revised form 7 October 2008
Accepted 11 October 2008
Available online 1 November 2008

Keywords:
Kernel principal component analysis (KPCA)
Kernel independent component analysis (KICA)
Support vector machine (SVM)
Nonlinear process monitoring
Fault detection and diagnosis

ABSTRACT

In this paper, some drawbacks of original kernel independent component analysis (KICA) and support vector machine (SVM) algorithms are analyzed for the purpose of multivariate statistical process monitoring (MSPM). When the measured variables follow non-Gaussian distribution, KICA provides more meaningful knowledge by extracting higher-order statistics compared with PCA and kernel principal component analysis (KPCA). However, in real industrial processes, process variables are complex and are not absolutely Gaussian or non-Gaussian distributed. Any single technique is not sufficient to extract the hidden information. Hence, both KICA (non-Gaussion part) and KPCA (Gaussion part) are used for fault detection in this paper, which combine the advantages of KPCA and KICA to develop a nonlinear dynamic approach to detect fault online compared to other nonlinear approaches. Because SVM is available for classifying faults, it is used to diagnose fault in this paper.
For above mentioned kernel methods, the calculation of eigenvectors and support vectors will be time consuming when the sample number becomes large. Hence, some dissimilar data are analyzed in the input and feature space.
The proposed approach is applied to the fault detection and diagnosis in the Tennessee Eastman process. Application of the proposed approach indicates that proposed method effectively captures the nonlinear dynamics in the process variables.

1. Introduction

Multivariate statistical process monitoring (MSPM) methods have been widely applied for chemical, semiconductor, food and biology process monitoring in the last decade (Hyvärinen, 1998; Hyvärinen, 1999; Hyvärinen et al., 2001; Knirsch et al., 1999; Krzanowski, 1979; Lee, 1998; MacGregor and Kourti, 1995; Zuo et al., 2005). Numerous review books and papers have been given by many researchers (Qin, 2003; Lee et al., 2004; Wang, 1999). Multivariate statistical approaches including principal component analysis (PCA) and partial least square (PLS) have been used in chemical industry for process monitoring, where the statistical analysis models are developed implicitly based on Gaussian-distribution assumption. Recently, several MSPM methods based on independent component analysis (ICA) have been proposed (Kano et al., 2003, 2004; Lee et al., 2003, 2004a, 2005; Xia et al., 2005; Albazzaz and Wang, 2004). The independent components were extracted to reduce data dimension of the monitored variables (Li and Wang, 2002). The monitoring performance was improved by finding several independent variables as linear combinations of measured variables that drive a process (Kano et al., 2003). ICA performs poorly due to the assumptions that a continuous process is operated in a particular steady state and that variables are normally distributed. To solve this problem, external analysis was given to distinguish faults from normal changes in operating conditions (Kano et al., 2004). Also, a dynamic monitoring method has been proposed to monitor dynamic process based on dynamic ICA (Lee et al., 2004b). Batch process monitoring problem was investigated using multi-way ICA (Yoo et al., 2004). Multi-resolution spectral ICA method is proposed to detect and isolate the sources of multiple oscillations in a chemical process, which is distinctive to other ICA based monitoring approaches (Xia et al., 2005).

The initial idea of combined ICA and PCA was firstly introduced in an attempt to combine the advantages of both algorithms for a practical continuous process case, where the combined ICA and PCA algorithm was utilized to analyze the prototype of original process trajectory, and constructed fixed model structures, assuming that the operation status were invariant (Kano et al., 2002). However, above mentioned MSPM methods have a strong linearity assumption, greatly limiting their application.

* Corresponding author at: Department of Chemical Engineering, The University of Texas at Austin, Austin, TX 78712, USA. Tel.: +1 512 471 4417; fax: +1 512 471 7060.
E-mail address: zhangyi@che.utexas.edu (Y. Zhang).

0009-2509/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ces.2008.10.012

To solve the nonlinear problem of observer data, some nonlinear approaches have been developed (Zhang and Qin, 2008; Hiden et al., 1999; Chiang et al., 2001; Tan and Mavrovouniotis, 1995; Geng and Zhu, 2005; Chen and Liao, 2002; Jia et al., 1998; Kramer, 1991; Kulkarni et al., 2004; Dong and McAvoy, 1996). Among the existing nonlinear methods, kernel-based techniques have been successfully developed for tackling the nonlinear problem in recent years (Baudat and Anouar, 2000; Choi et al., 2005; Christianini and Shawe-Taylor, 2000; Cortes and Vapnik, 1995; Kim et al., 2005; Lee et al., 2004; Rosipal and Trejo, 2001; Yang et al., 2005). They have attracted wide attentions, including support vector machine (SVM), kernel principal component analysis (KPCA), kernel PLS and kernel FDA. The basic idea is that the mapped data are analyzed using conventional linear statistical analysis techniques in feature space, which is equivalent to nonlinear analysis in original input space. Kernel independent component analysis (KICA) is a new nonlinear extension of whitened KPCA plus ICA (Mika et al., 1999; Romdhani et al., 1999; Schölkopf et al., 1998; Choi et al., 2005; Cho et al., 2005). When the measured variables follow non-Gaussian distribution, KICA provides more meaningful knowledge by extracting higher-order statistics compared with KPCA. However, in real industrial processes, process variables are complex and are not absolutely Gaussian or non-Gaussian distributed. Any single technique is not sufficient to extract the hidden information. Hence, both KICA (non-Gaussion part) and KPCA (Gaussion part) are used for fault detection in this paper, which combine the advantages of KPCA and KICA to develop a nonlinear dynamic approach to detect fault online compared to other nonlinear approaches. In addition to the modeling applications, SVM classifier is used to distinguish the vertical and horizontal two-phase flow regimes in a pipe so as to predict the transition region between the flow regimes (Trafalis et al., 2005). A decision support system is designed by combining a soft margin SVM classifier with decision tree (Jemwa and Aldrich, 2005). A hybrid approach that combines the self-organizing map and SVM for wafer bin map classification is proposed (Li et al., 2007). Application of SVM for fault diagnosis in the chemical industry now is still limited, although its increasing growth in the near future is foreseeable. The neighborhood score multi-class SVM has been successfully used for roller bearing's fault diagnosis (Sugumaran et al., 2008). A fault diagnosis scheme was proposed by determining the values of unknown features and by regenerating completely described samples to diagnose the system based on SVM classifiers (Ren et al., 2008). The performances of FDA, SVM and proximal SVM for fault diagnosis in the Tennessee Eastman (TE) process are compared (Chiang et al., 2004). It was found that it is important to properly select input variables to train the classifiers. In this paper, the kernel-transformed principal components from KPCA or independent components from KICA will be introduced as the inputs of SVM to solve the difficult problem of variable selection in SVM. For the above mentioned kernel methods, the calculation of eigenvectors and support vectors will be time consuming when the sample number becomes large. Hence, some dissimilar data are analyzed in the input and feature space.

The remainder of the paper is organized as follows. The ICA algorithm is introduced, in Section 2. Then KPCA, KICA and SVM algorithms are introduced in Section 3, respectively. In Section 4, two similarity factors are defined and dissimilar data points are isolated from original data and fault detection and diagnosis approaches based on KICA plus KPCA plus LS-SVM are investigated. The illustrative example is given to demonstrate the effectiveness of the proposed methods in Section 5. Finally, conclusions are drawn.

2. ICA algorithm

ICA is a recently developed method in which the goal is to find the statistically independent non-Gaussian hidden factors, or as independent as possible which constitute the observed variables through linear combination. Such a representation seems to capture the essential structure of the measurement data in many applications, including feature extraction and signal separation. There are further promising applications for ICA since it works in terms of higher order statistics. Suppose that d measured variables $x_1, x_2, \ldots, x_d$ can be expressed as linear combinations of m ($\leqslant d$) unknown independent components $s_1, s_2, \ldots, s_m$. The independent components and the measured variables have zero mean. If we denote the random column vectors as $\mathbf{x} = [x_1, x_2, \ldots, x_d]^{\mathrm{T}}$ and $\mathbf{s} = [s_1, s_2, \ldots, s_m]^{\mathrm{T}}$, the relationship between them is given by

$$\mathbf{x} = \mathbf{As} \tag{1}$$

where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m] \in R^{d\times m}$ is the unknown mixing matrix. The basic problem of ICA is to estimate both the mixing matrix $\mathbf{A}$ and the independent components $\mathbf{s}$ from only the observed data without any related prior knowledge, terms as blind separation. Therefore, the objective of ICA is to calculate a separating matrix so that components of the resconstructed data matrix become independent. This solution is equivalent to finding a demixing matrix $\mathbf{W}$ whose form is such that the elements of the reconstructed vector $\hat{\mathbf{s}}$, given as

$$\hat{\mathbf{s}} = \mathbf{Wx} \tag{2}$$

become as independent of each other as possible. The independent component vectors, should be as statistically independent of each other as possible, which is the most different point from PCA.

Based on the approximate form for the negentropy, Hyvärinen introduced a very simple and highly efficient fixed-point algorithm for ICA, calculated over sphered zero-mean vectors $\mathbf{z}$ (Hyvärinen and Oja, 2000). The algorithm, known as FastICA, calculates one column of the independent matrix and allows the identification of one independent component; the corresponding IC can then be found. Then we can obtain $\hat{\mathbf{s}}$ and demixing matrix $\mathbf{W}$. More details about the ICA algorithm, can be found in Hyvärinen and Oja (2000).

3. Kernel algorithms—KPCA, KICA and SVM

Kernel-based algorithm can be efficiently implemented in high-dimensional feature spaces by the use of integral operator and nonlinear kernel functions. Compared with other nonlinear methods, the main advantage of kernel algorithm is that it does not involve any nonlinear optimization. For example, KPCA method requires only the solution of an eigenvalue problem. In fact, it essentially requires only standard linear algebra, and can handle a wide range of nonlinearities due to its ability to employ different kernels. The goal of KPCA is to map the input space into a feature space via nonlinear mapping and then to extract whitened principal components in that feature space such that their covariance structure is an identity matrix (Schölkopf et al., 1998). That is, we can avoid performing the nonlinear mappings and computing inner products in the feature space by introducing a *kernel function* of form $k(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{\Phi}(\boldsymbol{x}), \boldsymbol{\Phi}(\boldsymbol{y})\rangle$ (Schölkopf et al., 1998; Romdhani et al., 1999). Some of the most widely used kernel functions are: *radial basis kernel*: $k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-(\|\boldsymbol{x} - \boldsymbol{y}\|^2)/c)$; *polynomial kernel*: $k(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{x}, \boldsymbol{y}\rangle^r$; *sigmoid kernel*: $k(\boldsymbol{x}, \boldsymbol{y}) = \tanh(\beta_0\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \beta_1)$, where c, r, β_0 and β_1 have to be specified. The polynomial kernel and radial basis kernel always satisfy Mercer's theorem, whereas the sigmoid kernel satisfies it only for certain parameter values (Haykin, 1999). In practice, the Gaussian kernel function is the most commonly used (Silverman, 1986). The specific choice of kernel function implicitly defines the form of mapping and the feature space, which, thus, actually determines how well the nonlinearity of a system can be captured. The prior knowledge of nonlinear characteristics could be useful to help select a proper kernel function. Up to now, how to choose the ideal kernel for a given nonlinear process has been an

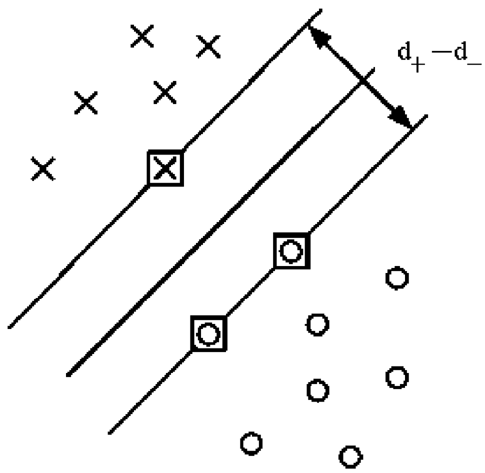

Fig. 1. SVM diagram.

open problem. No deterministic selection standard has been developed. Furthermore, once the kernel is set, proper kernel parameters should be set. However, there is also no theoretical framework to specify the optimal values of kernel parameters. Generally, both kernel functions and their parameters are set by trail and error. First, one initialization is empirically chosen; then the changing trend around the initial value is investigated; and finally the candidate which shows best performance is obtained. In the latter SVM simulation section, the effects of different kernel parameters are small since selection of input variables is avoided.

KICA is an emerging nonlinear feature extraction technique that formulates ICA in the kernel-induced feature space (Kocsor and Tóth, 2004; Yang et al., 2005; Lee et al., 2006). The basic idea of KICA is to nonlinearly map the data into a feature space using KPCA and to extract useful information to further perform ICA in the KPCA feature space. The basis KICA algorithm is based on the formalism presented in Yang et al. (2005). More details, please see Zhang and Qin, 2007.

The support vector machine (SVM) is a learning algorithm using a firm statistical learning theory approach. In particular,

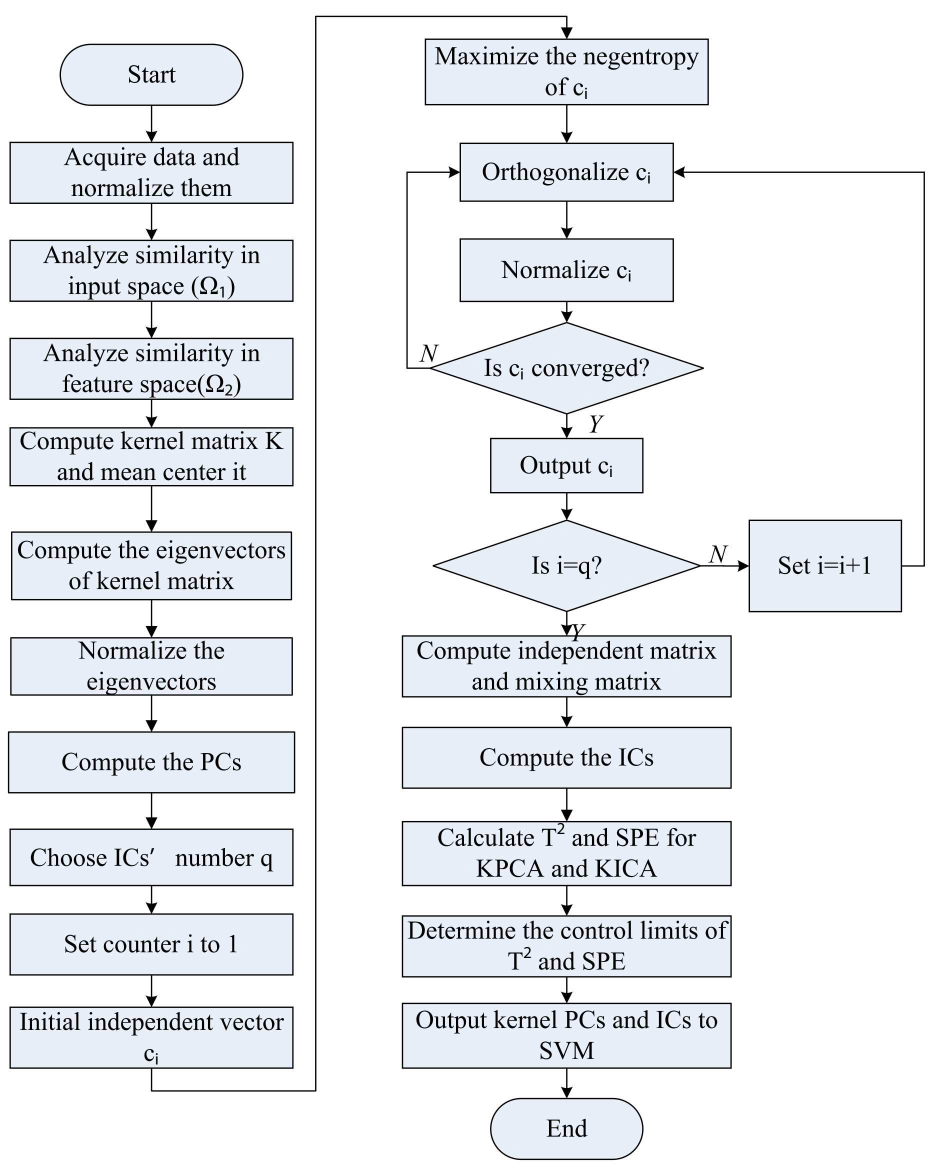

Fig. 2. Fault detection flow diagram.

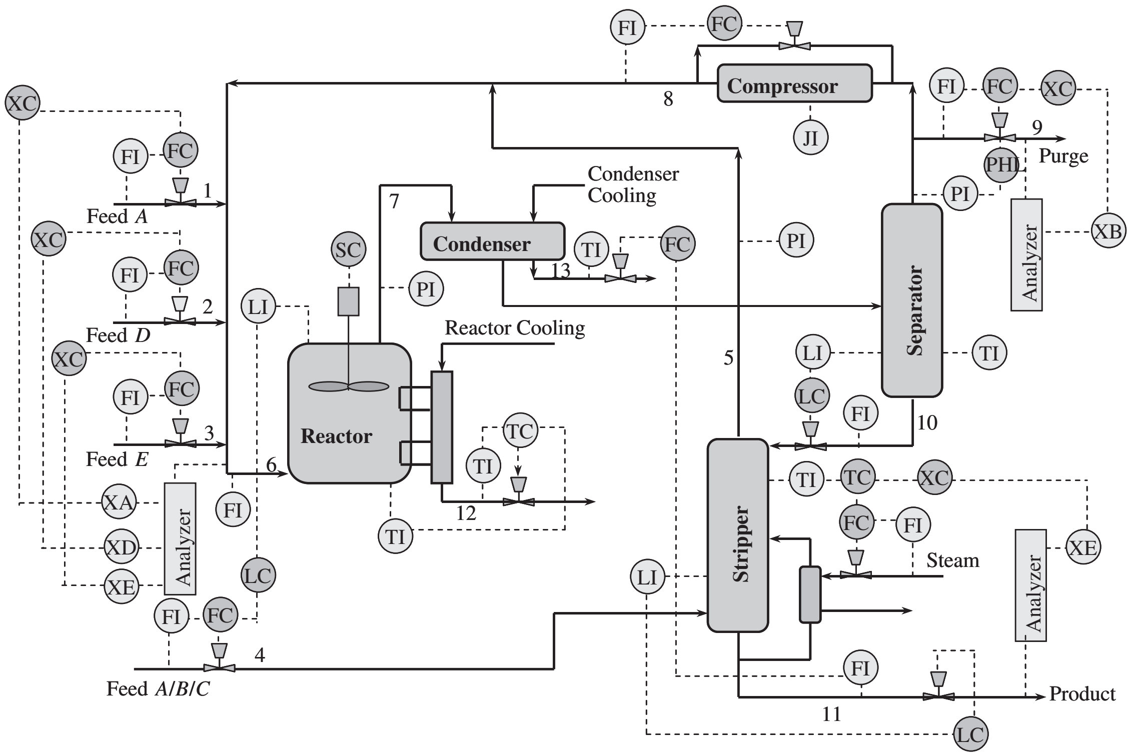

Fig. 3. Process layout of the Tennessee Eastman process.

Table 1
Monitored variables in the Tennessee Eastman process

No.	Variables	No.	Variables
1	*A* feed (stream 1)	18	Stripper temperature
2	*D* feed (stream 2)	19	Stripper steam Flow
3	*E* feed (stream 3)	20	Compressor work
4	Total feed (stream 4)	21	Reactor cooling water outlet temperature
5	Recycle flow (stream 8)	22	Separator cooling water outlet temperature
6	Reactor feed rate (stream 6)	23	*D* feed flow valve (stream 2)
7	Reactor pressure	24	*E* feed flow valve (stream 3)
8	Reactor level	25	*A* feed flow valve (stream 1)
9	Reactor temperature	26	Total feed flow valve (stream 4)
10	Purge rate (stream 9)	27	Compressor recycle valve
11	Product separator temperature	28	Purge valve (stream 9)
12	Product separator level	29	Separator pot liquid flow valve (stream 10)
13	Product separator pressure	30	Stripper liquid product flow valve (stream 11)
14	Product separator underflow (stream 10)	31	Stripper steam valve
15	Stripper level	32	Reactor cooling water flow
16	Stripper pressure	33	Condenser cooling water flow
17	Stripper underflow (stream 11)		

they showed that the problem of function estimation (encountered in pattern recognition, regression and density-estimation tasks) can be conducted in a general statistical framework that minimizes an upper bound on the expected loss using observed data (Schölkopf et al., 1999; Cortes and Vapnik, 1995).

The SVM algorithm finds the decision function as a linear combination of certain points of the training set (called *support vectors*), which condense the information contained in the training set. The decision boundary, which is linear in some possibly high-dimensional feature space, is obtained from solving a quadratic programming problem that depends on a regularization parameter. The plane separating the classes is called the *separating hyperplane*. It is possible to define many such hyperplanes that can classify the two classes. Defining $d_+(d_-)$ as the shortest distance to a separating hyperplane, shown in Fig. 1, the *margin* of a separating hyperplane is then $(d_+ - d_-)$. It can be shown that the shortest distance and the weight vector specifying the direction of the hyperplane are related through $d_+(d_-) = \mathbf{1}/\|\mathbf{w}\||$. Thus, maximizing the margin is equivalent to minimizing the term $\|\mathbf{w}\|/2$. The size of the allowable errors can be controlled by introducing a regularizing term $C\sum_{i=i}^{n}\xi_i^2$, where C is a constant and ξ_i is a slack variable. The support vector algorithm searches for an optimal separating hyperplane that simultaneously maximizes the margin and minimizes the errors.

Table 2
Process faults for the Tennessee Eastman process

No.	Description	Type
1	*A/C* feed ratio, *B* composition constant (stream 4)	Step
2	*B* composition, *A/C* ratio constant (stream 4)	Step
3	*D* feed temperature (stream 2)	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	*A* feed loss (stream 1)	Step
7	C header pressure loss-reduced availability (stream 4)	Step
8	*A*, *B*, *C* feed composition (stream 4)	Random variation
9	*D* feed temperature (stream 2)	Random variation
10	*C* feed temperature (stream 4)	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16	Unknown	
17	Unknown	
18	Unknown	
19	Unknown	
20	Unknown	
21	The valve for stream 4 was fixed at the steady state position	Constant position

Table 3
Fault detection rates of each method in the Tennessee Eastman process

Faults	PCA		Modified ICA		KPCA plus KICA	
	T^2	SPE	T^2	SPE	T^2	SPE
1	99	100	100	100	100	100
2	98	96	98	98	99	99
3	2	1	1	1	6	8
4	6	99	65	96	83	100
5	23	20	24	23	29	30
6	99	100	100	100	100	100
7	42	100	100	100	100	100
8	97	89	97	98	97	98
9	1	1	1	2	6	6
10	30	18	70	67	80	77
11	22	72	43	66	81	78
12	97	90	98	97	98	98
13	93	95	95	94	95	97
14	81	100	100	100	100	100
15	1	2	1	2	8	6
16	13	16	76	73	77	74
17	74	93	87	94	95	95
18	89	90	90	90	91	90
19	0	29	26	29	75	70
20	32	45	70	66	72	69

4. Similarity analysis and fault diagnosis

Kernel theory has found increasing numbers of applications in the nonlinear processes. The observed data are analyzed by exploring the distance and angle dependency of samples in the input space and feature space. Only angle dependency is not enough for similar definition (Zhang and Qin, 2008; Zhao et al., 2009). The similarity concept based on the distance and angle dependency of samples is introduced to define similarity factor between new data and the note set in feature space.

The note set at time i is denoted as $N_i = \{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_{n_i}\}$, $i = 1, \ldots, t_1$, and satisfying $n_i < t_1$. The subspace spanned by the vectors in set N_i is denoted as F_i. In the beginning of learning, there is only two data point, that is $N_1 = \{\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2\} = \{\mathbf{x}_1, \mathbf{x}_2\}$. The new data point will be added into the set according to their dependencies on the space F_i using the following proposed similarity factor:

$$S_i = \frac{1}{2} e^{\|\mathbf{x}_{\text{new}} - \tilde{\mathbf{x}}_i\|} + \frac{1}{2}\left(\frac{(\mathbf{x}_{\text{new}} - \tilde{\mathbf{x}}_i)^{\mathrm{T}}(\mathbf{x}_{\text{new}} - \tilde{\mathbf{x}}_{i-1})}{\|\mathbf{x}_{\text{new}} - \tilde{\mathbf{x}}_i\|_2 \|\mathbf{x}_{\text{new}} - \tilde{\mathbf{x}}_{i-1}\|_2}\right)^2 \tag{3}$$

where $\mathbf{x}_{\text{new}}$ is the new data, $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_{i-1}$ are continuous data points in N_{i-1}. If $S_i < \sqrt{\gamma_0}$, where γ_0 is a predefined small value satisfying $0 \leqslant \sqrt{\gamma_0} \leqslant 1$, the new data will be introduced to constitute the expanded node set; otherwise, this data point will be isolated and rejected without the node set expansion, that is $N_i = N_{i-1}$. After the similarity analysis in the input space, the number the retained data point is denoted n_1. And we obtain the datasets Ω_1.

For the first data subsets Ω_1, mapping the input space into feature space, similarity factor in feature space is defined as follows:

$$S_f = \frac{1}{2}\|\boldsymbol{\Phi}(\mathbf{x}_{\text{new}}) - \tilde{\boldsymbol{\Phi}}(\mathbf{x}_i)\|_2 + \frac{1}{2}\left(\frac{(\boldsymbol{\Phi}(\mathbf{x}_{\text{new}}) - \tilde{\boldsymbol{\Phi}}(\mathbf{x}_i))^{\mathrm{T}}(\boldsymbol{\Phi}(\mathbf{x}_{\text{new}}) - \tilde{\boldsymbol{\Phi}}(\mathbf{x}_{i-1}))}{\|\boldsymbol{\Phi}(\mathbf{x}_{\text{new}}) - \tilde{\boldsymbol{\Phi}}(\mathbf{x}_i)\|_2 \|\boldsymbol{\Phi}(\mathbf{x}_{\text{new}}) - \boldsymbol{\Phi}(\mathbf{x}_{i-1})\|_2}\right)^2 \tag{4}$$

If $S_f < \sqrt{\gamma_1}$, where γ_1 is a predefined small value satisfying $0 \leqslant \sqrt{\gamma_1} \leqslant 1$, the new data will be introduced to constitute the expanded node set; otherwise, this data point will be rejected without the node set expansion. After the similarity analysis in the feature space, the number the retained data point is denoted n_2. And we obtain the datasets Ω_2.

Fault detection and diagnosis process is shown in Fig. 2. In PCA and KPCA analysis, the changes in the variable relation according to the PCs and nPCs reflect the changes in underlying process behavior. Likewise, in KICA analysis, the changes in the variable relation according to the ICs reflect the changes in underlying process behavior. Both KPCA (Gaussion part) and KICA (non-Gaussion part) are used for fault detection in this paper. For the computation approaches of T^2 and SPE and their control limit, please see Lee et al. (2006) and Zhang and Qin (2007).Once fault is detected, the kernel PCs or kernel ICs are as inputs of SVM for fault diagnosis.

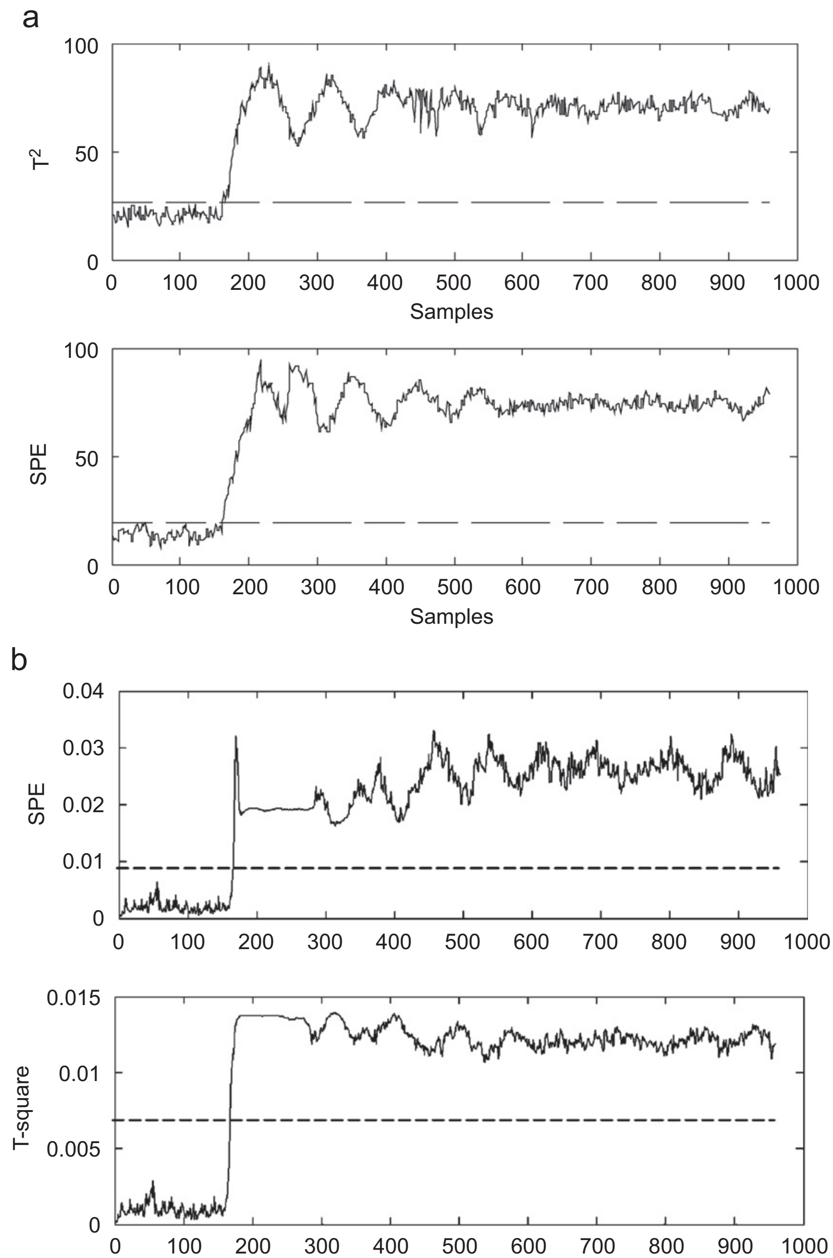

Fig. 4. Monitoring results of the Tennessee Eastman process in the case of fault 1 based on (a) KPCA and (b) KICA.

5. Simulation result

5.1. Tennessee Eastman process

In this section, the proposed method is applied to the Tennessee Eastman process simulation data and is compared with KPCA and original KICA monitoring results. The Tennessee Eastman process is a complex nonlinear process, which was created by Eastman Chemical Company to provide a realistic industrial process for evaluating process control and monitoring methods. The test process is based on a simulation of an actual industrial process where the components, kinetics, and operating conditions have been modified for proprietary reasons. The control structure is shown schematically in Fig. 3. There are five major unit operations in the process: a reactor, a condenser, a recycle compressor, a separator, and a stripper; and it contains eight components: A, B, C, D, E, F, G and H. The four reactants A, C, D and E and the inert B are fed to the reactor where the products G and H are formed and a byproduct F is also produced. The process has 22 continuous process measurements, 12 manipulated variables, and 19 composition measurements sampled less frequently. The details on the process description are well explained in Chiang et al. (2001). A total of 52 variables are used for monitoring in this study. Those main variables are listed in Table 1. We excluded all composition measurements since they are hard to measure on-line in practice. A sampling interval of 3 min was used to collect the simulated data for the training and testing sets. A set of programmed faults (fault 1–21) is listed in Table 2. For each fault, two sets of data were generated. The training data were used to build the models and the testing data were used to validate the model. The training data sets for each fault are composed of 480 observations. The testing data sets for each fault are composed of 960 observations. All faults in the test data set were introduced from sample 160. The data are generated by

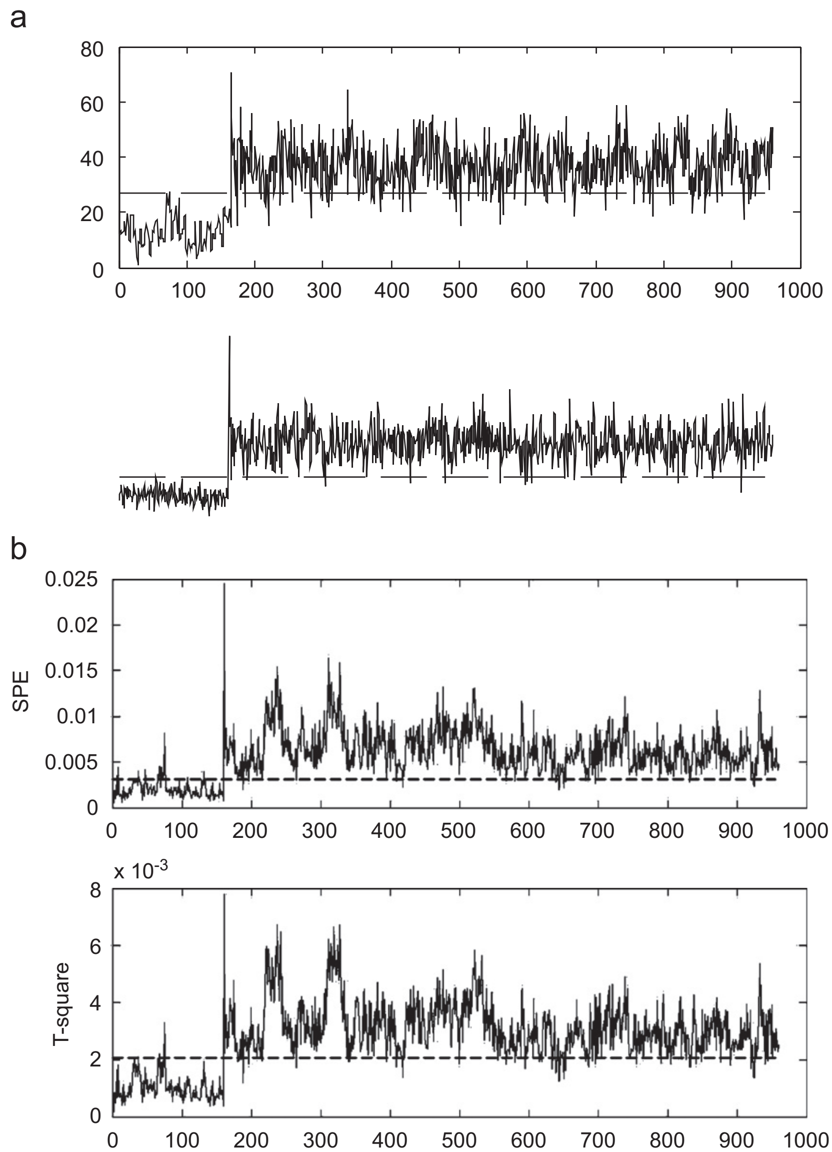

Fig. 5. Monitoring results of the Tennessee Eastman process in the case of fault 4 based on (a) KPCA and (b) KICA.

Chiang et al. (2001) and can be downloaded from http://brahms.scs.uiuc.edu.

All the data were auto-scaled prior using the mean and standard deviation of each variable to the application of the PCA, ICA, KPCA, KICA, i.e., the data were mean centered and variance scaled. For the data obtained after the fault occurrence, the percentage of the samples outside the 95% control limits is calculated in each simulation and termed as the detection rate. The 95% control limits are expressed as a dotted line in those figures. RBF kernel is used for the present work. And the RBF kernel is selected. And $\sigma = 0.5$. For the KPCA and KICA, 30 whitened vectors are selected where the corresponding relative eigenvalues $\lambda_i/sum(\lambda_i)$ in feature space are larger than 0.0001. Among 30 whitened vectors, 12 nPCs and nICs are selected from average eigenvalue criterion, respectively. Fault detection rates are calculated and tabulated in Table 3. Obviously, for most cases, the fault detection rates based on KPCA plus KICA are higher than PCA and modified ICA. For fault 1, the monitoring results based on KPCA and KICA are shown in Fig. 4(a) and (b), respectively. In the case of fault 1, the A/C feed ratio, B composition constant of stream 4 is changed step by step. In the training procedure, normal training samples are needed for KPCA, and both the normal training samples and the faulty training samples are needed for SVM. For this fault, KPCA and KICA can detect all the faults from sample 160. In the case of fault 4, the reactor cooling water inlet temperature is changed step by step. KPCA can detect the fault from about sample 160 (Fig. 5(a)). However, despite the presence of the fault, there are some samples (samples 300, 420 and 740) below the 95% control limit of SPE chart, giving the process operator an incorrect picture of the process status. The KICA monitoring charts show that SPE statistics successfully detect the fault at sample 300, 420 and 740 (Fig. 5(b)). In the case of fault 5, KICA monitoring charts can successfully detect the fault at samples 430–440. However, KPCA monitoring chars cannot detect it (Fig. 6(a) and (b)). When the measured variables follow non-Gaussian distribution, KICA provides more meaningful knowledge by extracting higher-order statistics compared with KPCA. In the case of fault 11, the reactor cooling water inlet temperature is randomly changed. From Fig. 7(a), although KICA detect the fault from about sample 160, there are some samples below the 95% control limit (samples 490–500). From Fig. 7(b), KPCA is able to detect the fault at samples 490–500, which are more efficiently than KICA

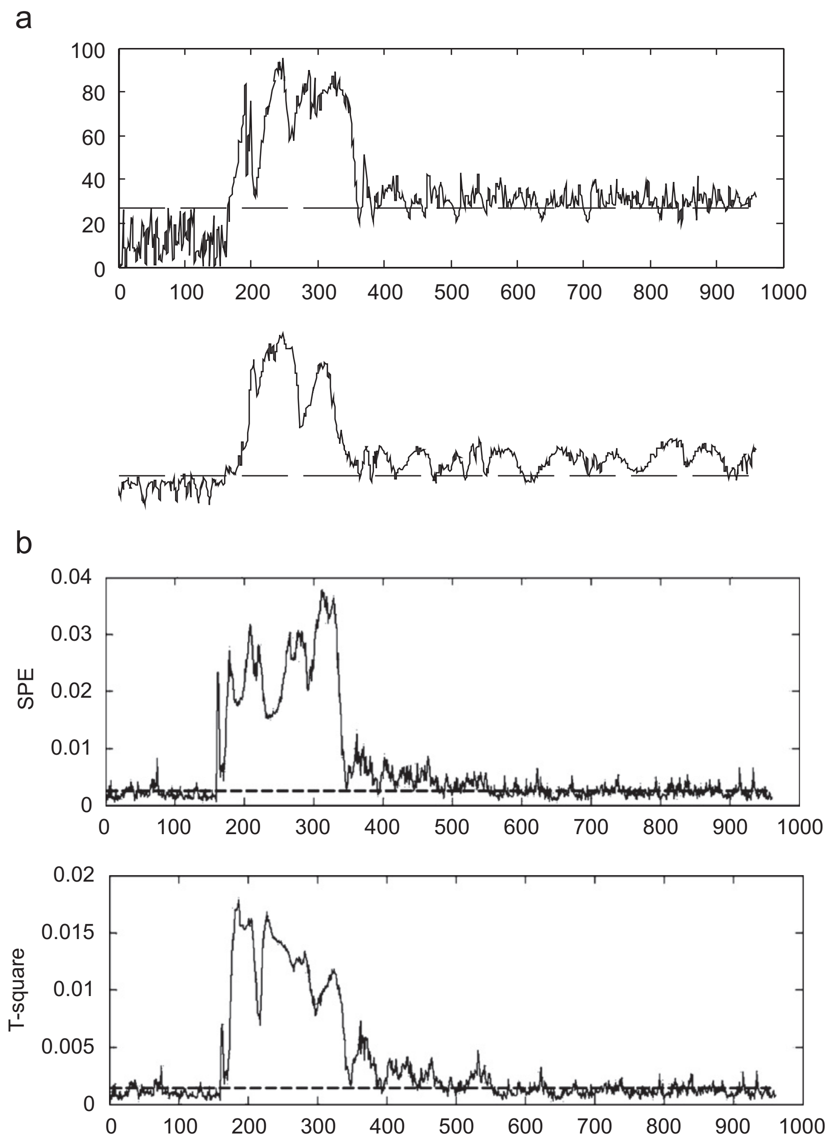

Fig. 6. Monitoring results of the Tennessee Eastman process in the case of fault 5 based on (a) KPCA and (b) KICA.

without false alarms. In contrast to the KICA monitoring, the KPCA monitoring charts are available for Gaussian distributed processes. Hence, both KICA (non-Gaussion part) and KPCA (Gaussion part) are used for fault detection in this paper, which combine the advantages of KPCA and KICA to develop a nonlinear dynamic approach to detect fault online compared to other nonlinear approaches. The hypothesis is that KPCA and KICA based MSPC charts can be used together, which should improve the capability of detecting the faults.

From Table 3, it should be noted that the overall detection rate of the KPCA plus KICA is higher than that of PCA and modified ICA. The missed detection rates for the faults 3, 9 and 15 are very high for all the fault detection statistics. No observed change in the mean, the variance and the higher order variance can be detected. It conjectured that any statistic will result in high missed detection rates for those fault, in other words, the faults 3, 9 and 15 are unobservable from the test data.

In the following section, the object is to verify and exemplify the proposed approach for fault diagnosis. The kernel-transformed nPCs from KPCA will be directly introduced as the inputs of SVM to diagnose fault. Selection of the input variable is thus avoided. The three faults are: fault 4 (reactor cooling water inlet temperature), fault 9 (*D* feed temperature (stream 2)) and fault 11 (reactor cooling water inlet temperature), listed in Table 2. However, only variables 9 and 51 are important for faults 4 and 11. Fault 4 datasets overlap with fault 11. From Figs. 8–10, and Table 4, after similarity analysis, the classification rates of nPCs plus SVM are higher than that of PC plus SVM. The classification rates of nICs plus SVM are higher than the classification rate of PCs plus SVM. And the training and test times of the KPCA plus SVM with similarity analysis are less than that of original KPCA plus SVM. In this simulation example, the effects of different kernel parameters are small since selection of input variables is avoided.

6. Conclusions

In real industrial processes, process variables are complex and are not absolutely Gaussian or non-Gaussian distributed. Any single technique is not sufficient to extract the hidden information. Hence, both KICA (non-Gaussion part) and KPCA (Gaussion part) are used for fault detection in this paper, which combine the advantages of

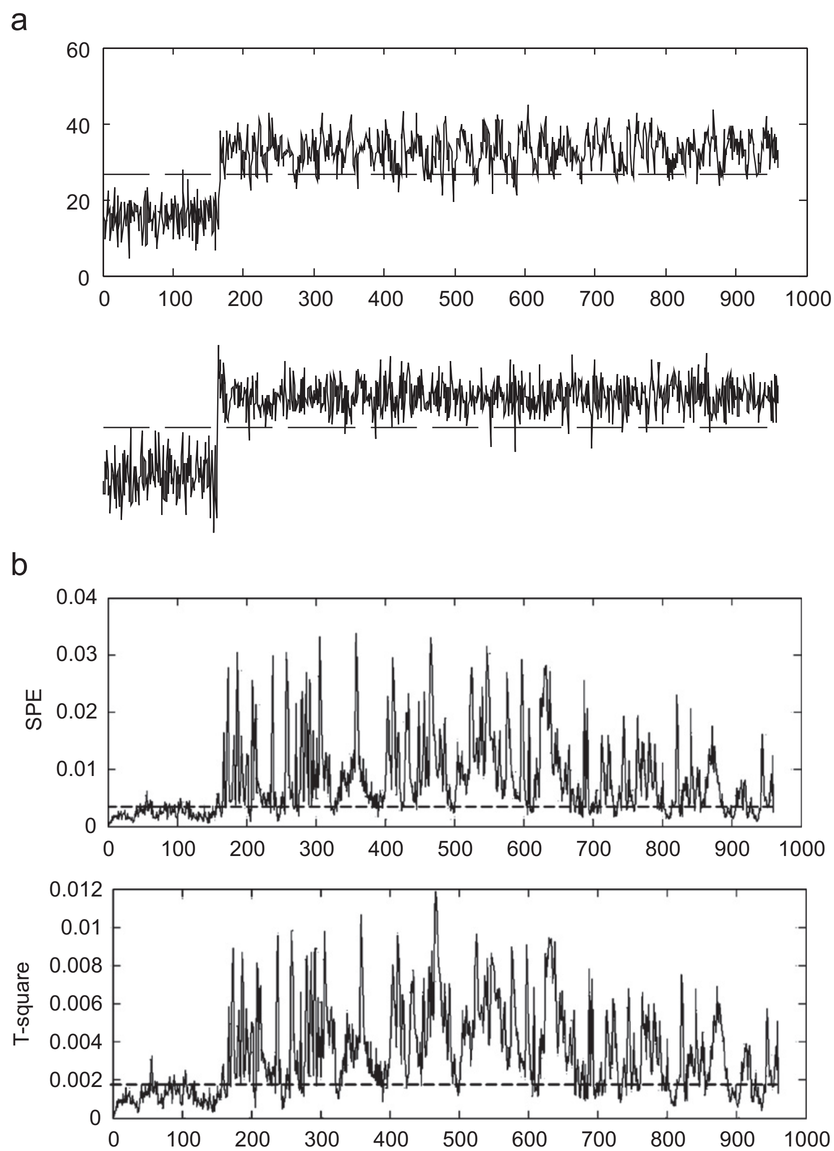

Fig. 7. Monitoring results of the Tennessee Eastman process in the case of fault 11 based on (a) KPCA and (b) KICA.

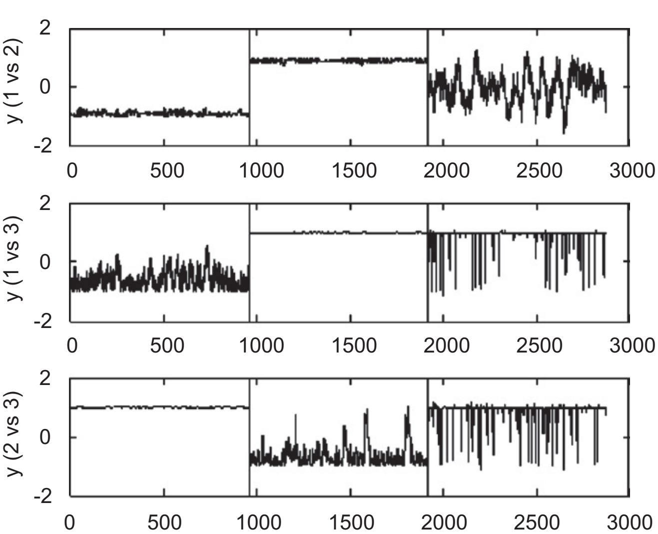

Fig. 8. Classification results for testing data using the PCA–SVM classifiers.

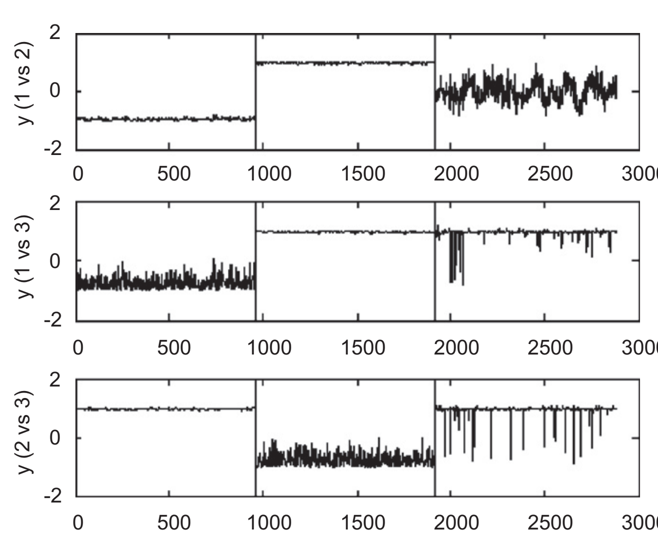

Fig. 9. Classification results for testing data using the KPCA–SVM classifiers.

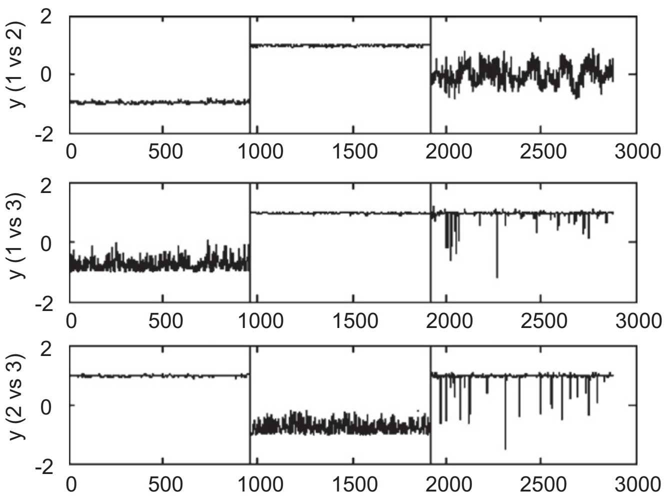

Fig. 10. Classification results for testing data using the KICA–SVM classifiers.

Table 4
Comparation of three classification methods

	Feature selection time (s)	Training time (s)	Classification time (s)	Misclassification (training) (%)	Misclassification (test) (%)
PCA–SVM	2.820	1.230	0.622	0.8	5.9
KPCA–SVM	1.760	0.547	0.322	0	2.3
KICA–SVM	1.240	0.562	0.320	0	1.6

KPCA and KICA. Because SVM is available for classifying faults, SVM is used to diagnose fault in this paper. For KPCA, KICA and SVM, the calculation of eigenvectors and support vectors will be difficult when the sample number becomes large. Hence, some similar data are not retained after new similarity factors of the observed data are defined and similar characteristics are analyzed in the input and feature space.

The proposed approach is applied to the fault detection and diagnosis in the Tennessee Eastman process. Applications of the proposed approach indicate that proposed method effectively compensate each other in the process monitoring.

References

Albazzaz, H., Wang, X.Z., 2004. Statistical process control charts for batch operations based on independent component analysis. Industrial and Engineering Chemistry Research 43 (21), 6731–6741.

Baudat, G., Anouar, F., 2000. Generalized discriminant analysis using a kernel approach. Neural Computation 12 (10), 2385–2404.

Chen, J.H., Liao, C.-M., 2002. Dynamic process fault monitoring based on neural network and PCA. Journal of Process Control 12, 277–289.

Chiang, L.H., Russell, F.L., Braatz, R.D., 2001. Fault Detection and Diagnosis in Industrial Systems. Springer, London.

Chiang, L.H., Kotanchek, M.E., Kordon, A.K., 2004. Fault diagnosis based on Fisher discriminant analysis and support vector machines. Computers and Chemical Engineering 28, 1389–1401.

Cho, J.-H., Lee, J.-M., Choi, S.W., Lee, D., Lee, I.-B., 2005. Fault identification for process monitoring using kernel principal component analysis. Chemical Engineering Science 60, 279–288.

Choi, S.W., Lee, C., Lee, J.-M., Park, J.H., Lee, I.-B., 2005. Fault detection and identification of nonlinear processes based on KPCA. Chemometrics and Intelligent Laboratory Systems 75, 55–67.

Christianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, UK.

Cortes, C., Vapnik, V., 1995. Support vector networks. Machine Learning 20 (3), 273–297.

Dong, D., McAvoy, T.J., 1996. Nonlinear principal component analysis based on principal curves and neural networks. Computers and Chemical Engineering 20, 65–78.

Geng, Z.Q., Zhu, Q.X., 2005. Multiscale nonlinear principal component analysis (NLPCA) and its application for chemical process monitoring. Industrial and Engineering Chemistry Research 44, 3585–3593.

Haykin, S., 1999. Neural Networks. Prentice-Hall, Englewood Cliffs, NJ.

Hiden, H.G., Willis, M.J., Tham, M.T., Montague, G.A., 1999. Non-linear principal components analysis using genetic programming. Computers and Chemical Engineering 23, 413–425.

Hyvärinen, A., 1998. New approximations of differential entropy for independent component analysis and projection pursuit. Advances in Neural Information Processing Systems 10, 273–279.

Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks 10, 626–634.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Networks 13 (4–5), 411–430.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. Wiley, New York.

Jemwa, G.T., Aldrich, C., 2005. Improving process operations using support vector machines and decision trees. A.I.Ch.E. Journal 51, 526–543.

Jia, F., Martin, E.B., Morris, A.J., 1998. Non-linear principal component analysis for process fault detection. Computers and Chemical Engineering 22, S851–S854.

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I., Ohno, H., 2002. Combination of independent component analysis and principal component analysis for multivariate statistical process control. In: Proceedings of International Symposium on Design, Operation and Control of Chemical Plants, Taipei, Taiwan, p. 319.

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I., Ohno, H., 2003. Monitoring independent components for fault detection. A.I.Ch.E. Journal 49 (4), 969–976.

Kano, M., Hasebe, S., Hashimoto, I., Ohno, H., 2004. Evolution of multivariate statistical process control: independent component analysis and external analysis. Computers & Chemical Engineering 28 (6–7), 1157–1166.

Kim, K., Lee, J.-M., Lee, I.-B., 2005. A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction. Chemometrics and Intelligent Laboratory Systems 79 (1/2), 22–30.

Knirsch, J.C., MJuller, P., RJatsch, K.-R., 1999. Input space versus feature space in kernel-based methods. IEEE Transactions on Neural Networks 10 (5), 1000–1016.

Kocsor, A., Tóth, L., 2004. Kernel based feature extraction with a speech technology application. IEEE Transactions on Signal Processing 52 (8), 2250–2263.

Kramer, M.A., 1991. Non-linear principal component analysis using autoassociative neural networks. A.I.Ch.E. Journal 37, 233–243.

Krzanowski, W.J., 1979. Between-groups comparison of principal components. Journal of the American Statistical Association 74, 703–707.

Kulkarni, S.G., Chaudhary, A.K., Nandi, S., Tambe, S.S., Kulkarni, B.D., 2004. Modeling and monitoring of batch processes using principal component analysis (PCA) assisted generalized regression neural networks (GRNN). Biochemical Engineering Journal 18, 193–210.

Lee, T., 1998. Independent Component Analysis: Theory and Applications. Kluwer Academic Publishers, Boston.

Lee, J.M., Yoo, C.K., Lee, I.B., 2003. New monitoring technique with ICA algorithm in wastewater treatment process. Water Science and Technology 47 (12), 49–56.

Lee, G., Han, C.H., Yoon, E.S., 2004. Multiple-fault diagnosis of the Tennessee Eastman process based on system decomposition and dynamic PLS. Industrial and Engineering Chemistry Research 43, 8037–8048.

Lee, J.M., Yoo, C.K., Lee, I.B., 2004a. Statistical process monitoring with independent component analysis. Journal of Process Control 14, 467–485.

Lee, J.-M., Yoo, C.K., Lee, I.B., 2004b. Statistical monitoring of dynamic processes based on dynamic independent component analysis. Chemical Engineering Science 59, 2995–3006.

Lee, J.M., Qin, S.J., Lee, I.B., 2006. Fault detection and diagnosis of multivariate processes based on modified independent component analysis. A.I.Ch.E. Journal 52, 3501–3514.

Li, R.F., Wang, X.Z., 2002. Dimension reduction of process dynamic trends using independent component analysis. Computers & Chemical Engineering 26, 467–473.

MacGregor, J.F., Kourti, T., 1995. Statistical process control of multivariate processes. Control Engineering Practice 3, 403–414.

Mika, S., Schölkopf, B., Smola, A.J., MJuller, K.-R., Scholz, M., RJatsch, G., 1999. KPCA and de-noising in feature spaces. Advances in Neural Information Processing Systems 11, 536–542.

Qin, S.J., 2003. Statistical process monitoring: basics and beyond. Journal of Chemometrics 17, 480–502.

Ren, N., Liang, J., Gu, B., Han, H., 2008. Fault diagnosis strategy for incompletely described samples and its application to refrigeration system. Mechanical Systems and Signal Processing 22 (2), 436–450.

Romdhani, S., Gong, S., Psarrou, A., 1999. A multi-view nonlinear active shape model using KPCA. In: Proceedings of BMVC, Nottingham, UK, pp. 483–492.

Rosipal, R., Trejo, LJ., 2001. Kernel partial least squares regression in reproducing kernel Hilbert space. Journal of Machine Learning Research 2, 97–123.

Schölkopf, B., Smola, A.J., MJuller, K., 1998. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10 (5), 1299–1399.

Schölkopf, B., Burges, C.J.C., et al., 1999. Advances in Kernel Methods—Support Vector Learning. MIT Press, England.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Hapman & Hall, UK.

Sugumaran, V., Sabareesh, G.R., Ramachandran, K.I., 2008. Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. Expert Systems with Applications 34 (4), 3090–3098.

Tan, S., Mavrovouniotis, M.L., 1995. Reducing data dimensionality through optimizing neural networks inputs. A.I.Ch.E. Journal 41, 1471–1480.

Trafalis, T.B., Oladunni, O., Papavassiliou, D.V., 2005. Two-phase flow regime identification with a multiclassification support vector machine (SVM) model. Industrial & Engineering Chemistry Research 44, 4414–4426.

Wang, X.Z., 1999. Data mining and knowledge discovery for process monitoring and control. Springer, London.

Xia, C.M., Howell, J., Thornhill, N.F., 2005. Detecting and isolating plant-wide oscillations via spectral independent component analysis. Automatica 41, 067–2075.

Yang, J., Frangi, A.F., Yang, J.-Y., Zhang, D., Jin, Z., 2005. KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2), 230–244.

Yoo, C.K., Lee, J.-M., Vanrolleghem, P.A., Lee, I.B., 2004. On-line monitoring of batch processes using multiway independent component analysis. Chemometrics and Intelligent Laboratory Systems 71, 151–163.

Zhang, Y., Qin, S.J., 2007. Fault detection of nonlinear processes using multiway kernel independent analysis. Industrial & Engineering Chemistry Research 46, 7780–7787.

Zhang, Y., Qin, S.J., 2008. Improved nonlinear fault detection technique and statistical analysis. A.I.Ch.E. Journal 54 (12), 3207–3220.

Zhao, C., Wang, F., Zhang, Y., 2009. Nonlinear process monitoring based on kernel dissimilarity analysis. Control Engineering Practice 17 (1), 221–230.

Zuo, M.J., Lin, J., Fan, X.F., 2005. Feature separation using ICA for a one-dimensional time series and its application in fault detection. Journal of Sound and Vibration 287, 614–624.