Switched Fault Diagnosis Approach for Industrial Processes based on Hidden Markov Model

Lin Wang, Chunjie Yang*, Youxian Sun, Yijun Pan, and Ruqiao An
State Key Laboratory of Industrial Control Technology, Institute of Industrial processes Control, Zhejiang University, Hangzhou, Zhejiang, China

E-mail: cjyang@iipc.zju.edu.cn

Abstract. Traditional fault diagnosis methods based on hidden Markov model (HMM) use a unified method for feature extraction, such as principal component analysis (PCA), kernel principal component analysis (KPCA) and independent component analysis (ICA). However, every method has its own limitations. For example, PCA cannot extract nonlinear relationships among process variables. So it is inappropriate to extract all features of variables by only one method, especially when data characteristics are very complex. This article proposes a switched feature extraction procedure using PCA and KPCA based on nonlinearity measure. By the proposed method, we are able to choose the most suitable feature extraction method, which could improve the accuracy of fault diagnosis. A simulation from the Tennessee Eastman (TE) process demonstrates that the proposed approach is superior to the traditional one based on HMM and could achieve more accurate classification of various process faults.

1. Introduction

With the growth of complexity of industrial processes, effective monitoring and diagnosis play a significant role in ensuring operational safety of chemical equipment, maintaining product quality, optimizing product profit, and improving environmental sustainability [1].

In process industry, more and more data have been collected with a wide range utilization of sensors. However, it will lead to low efficiency if original data are stored and processed directly. Besides, they do not need all the features of objects in some particular applications. This is because many features cannot reflect the nature of the objects. Sometimes redundant features will even become obstacles to the subsequent processing. Based on the views above, feature extraction is needed to gain essential characteristics of data. It will reduce the interference information like noise and highlight the useful information. Feature extraction is the key step in fault diagnosis. Because of this, the accuracy of the fault diagnosis will be improved by enhancing the effectiveness of feature extraction.

There are a lot of traditional feature extraction methods, such as PCA、KPCA and ICA [2]. PCA can deal with linear relationships among variables effectively, but it cannot extract the nonlinear relationships, which are very common in the process industry [3]. KPCA was proposed to deal with it. It could extract nonlinear relationships among variables. Because of this, it is widely utilized in image recognition, feature extraction, fault detection and other fields [4]. However, there is no effective way to select the kernel function and parameters of KPCA at present. If KPCA was used to deal with process variables with linear relationships, not only the choice of kernel function and its parameters is cumbersome and time-consuming, but also it will greatly decrease accuracy of fault diagnosis with the inappropriate parameters. So it is inappropriate to use KPCA for feature extraction blindly. So far,

when KPCA is utilized, nonlinear relationships among variables are a prior condition in most researches. That is to say, it is necessary to analyse the nonlinear relationships among process variables. This article proposes a switched feature extraction method to deal with the proposed problem. As we all know, a better feature extraction would lead to a better classification. So the proposed method has a great research value. A nonlinearity measure proposed by Kruger to analyse the nonlinearity among variables is used to choose the appropriate method for feature extraction [5]. The extracted features will be used as the observation sequences of HMM, which is used as a classifier.

HMM is a valuable pattern recognition tool for sequential data. Its application for industrial processes monitoring is gradually increasing. It is an intelligent diagnosis technology with some unique characteristics. It is a doubly stochastic process which establishes a statistical model of time series by Markov chains and random functions [6]. So HMM has a strong ability to set up model of the dynamic process and classify sequential patterns. Due to the characteristics above, HMM has a good application prospect in the field of fault diagnosis and detection in process industry. So we choose it to clarify the efficiency of the proposed method.

In this article, the proposed method properly takes the characteristics of process data into consideration. By analysing data characteristics, we could overcome the insufficient of the traditional fault diagnosis methods which sometimes extract features of process variables by inappropriate method. By this way, we could improve the accuracy of the fault diagnosis.

The remainder of this article is organized as follows. In section II, PCA, KPCA and HMM are briefly described. In section III, switched fault diagnosis approach for industrial processes based on HMM is developed in detail. In section IV, TE process is used to illustrate the effectiveness of the proposed method. Section V gives the conclusions.

2. Preliminaries

2.1. Principal component analysis

PCA is a linear dimension reduction technique. It transforms high-dimensional correlated variables into low-dimensional uncorrelated ones. PCA determines a series of orthogonal vectors which are called loading vectors. Consider a data matrix $X \in R^{n \times m}$, for n observations with m measurement variables. The loading vectors could be derived by Singular Value Decomposition (SVD) as follows:

$$\frac{1}{\sqrt{n-1}} X = U \Sigma V^{T} \tag{1}$$

where $U \in R^{n \times n}$ and $V \in R^{m \times m}$ are both unitary matrices. The matrix $\Sigma \in R^{n \times m}$ contains real and decreasing singular values ($\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m$) along the main diagonal, with other off-diagonal elements to be zero. Loading vectors are the orthogonal columns of the matrix V and the variance of the data set projecting along the kth column of V is equal to σ_k^2. $\sigma_k^2 (i = 1, 2, \cdots, m)$ is arranged in decreasing order to determine the principal components. The first a PCs are selected to build the PCA model, so score vectors for kth sample can be represented by loading matrix P and sample vectors x_k.

$$t_k = P^T x_k \tag{2}$$

2.2. Kernel principal component analysis

PCA performs well when the relationships among variables are linear, but it can't deal with nonlinearity among variables. KPCA projects observation matrix to a high dimensional feature space F by using kernel function φ and then implements PCA in F, which can be expressed as

$$\Phi: R^m \rightarrow F \tag{3}$$

Assume mapping data for each variable have been scaled to zero mean, just $\sum_{k=1}^{n} \Phi(x_k) = 0$, and n denotes the data size. The corresponding covariance matrix can be defined as

$$S^F = \frac{1}{n} \sum_{k=1}^{n} \Phi(x_k) \Phi(x_k)^T \tag{4}$$

In essence, KPCA has realized the nonlinear transformation between data space and feature space. x_i and x_j are sample points in data space and the mapping function is Φ. The fundamental of KPCA is inner product transformation.

$$K_{ij} = \left\langle \Phi(x_i), \Phi(x_j) \right\rangle = K(x_i, x_j) \tag{5}$$

Data points $\Phi(x_k), (k = 1, \cdots, n)$ need to be processed in the feature space F, the process is as follows

$$\tilde{K} = K - 1_n K - K 1_n + 1_n K 1_n \tag{6}$$

where

$$1_n = \frac{1}{n} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \in R^{n \times n}$$

By eigenvalue decomposition to $\tilde{K}$ as shown

$$\lambda \alpha = \tilde{K} \alpha \tag{7}$$

We can obtain orthogonal eigenvectors $\alpha_1, \alpha_2, \cdots, \alpha_n$ and corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The dimension reduction can be obtained by mapping $\Phi(x)$ to the eigenvectors v_k in F, where $(k = 1, 2, \cdots, a)$, which can be expressed as

$$t_k = \left\langle v_k, \Phi(x) \right\rangle = \sum_{i=1}^{n} \alpha_i^k \left\langle \Phi(x_i), \Phi(x) \right\rangle \tag{8}$$

In the KPCA method, we just need to calculate matrix K to deal with the problem of eigenvalues by ignoring the specific nonlinear mapping relation Φ. In this article, we choose radial basis kernel, that is

$$K(x, y) = exp(-\frac{\|x - y\|^2}{\sigma}) \tag{9}$$

with $\sigma = rm$, where r is a constant to be selected and m is the dimension of the input space.

2.3. Hidden Markov model
Generally speaking, HMM is a valuable pattern recognition tool for sequential data. Its application for pro-cess monitoring is gradually increasing. A continuous HMM (CHMM) is used since the TE data are continuous. It usually can be described by the following parameters:

- N: the number of hid den states, the states are expressed as $S = \left\{ S_1, S_2, \cdots, S_N \right\}$

- $B = \left\{ b_j(O), j = 1, 2, \cdots N \right\}$: observation probability matrix

  where $b_j(O) = \sum_{m=1}^{M} C_{jm} H[O, \mu_{jm}, U_{jm}], \quad 1 \leq j \leq N$

- $A = \{a_{ij}\}$: state transition probability matrix

$$a_{ij} = P(q_{t+1} = S_j / q_t = S_i), 1 \le i, j \le N$$

where $\sum_{j=1}^{N} a_{i,j} = 1, \quad 1 \le i \le N$

- O is the observation vector. C_{jm} is a mixed coefficient of the mth mixing element of the hidden state j. H is Gaussian mixture probability density. μ_{jm} is mean vector of the mth mixing element of the hidden state S_j. U_{jm} is the covariance matrix of the mth mixing element of the hidden state S_j.

where $\sum_{m=1}^{M} C_{jm} = 1, \quad 1 \le j \le N$

$$C_{jm} \ge 0, \quad 1 \le j \le N, \quad 1 \le m \le M$$

- $\pi = \{\pi_i\}$:the initial state probability matrix

$$\pi_i = P(q_1 = S_i), 1 \le i \le N$$

where $\sum_{i=1}^{N} \pi_i = 1$

3. Switched fault diagnosis approach based on HMM

Industrial processes are usually sophisticated, so the process data have various characteristics, such as non-Gaussianity, dynamic correlations, nonlinear relations and uncertainties in the process measurement, which makes feature extraction significant [3]. Only one feature extraction method would be used in traditional fault diagnosis approach based on HMM, such as PCA [7]. But data have their own characteristics in different operation conditions, which makes it unable to perform well by using a unified feature extraction method. Here, we just consider linear and nonlinear relationships among process variables. This article proposes a switched feature extraction method based on PCA and KPCA. Nonlinearity measure is used to analyse the nonlinear relationships among process variables and choose the appropriate method for feature extraction. The method will be introduced below.

3.1. Nonlinear measure
The original data can be divided into several regions based on a priori knowledge of the process, and the accuracy boundaries are determined according to the confidence intervals of the correlation matrix of one region. By using the principle of cross-validation [8], we can deal with the problem of choosing region, which helps us to determine the accuracy boundaries. The relationships among variables are decided by comparing the residuals in each region with their corresponding accuracy boundaries. Specific steps are shown as follows.

Suppose a data matrix $Z \in R^{n \times m}$ having m variables and n samples is divided into l regions, each of which includes $\tilde{n} = n / l$ observations.

The correlation matrix for the original data is defined as:

$$S_{zz} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mm} \end{bmatrix} \tag{10}$$

The element in matrix is expressed as:

$$s_{ij} = \frac{1}{\tilde{n}-1}\sum_{k=1}^{\tilde{n}} \frac{z_{ki}-\tilde{z}_i}{\sigma_i}\frac{z_{kj}-\tilde{z}_j}{\sigma_j} \tag{11}$$

where $\tilde{z}_i$, $\tilde{z}_i$ are the mean values and σ_i, σ_j are the standard deviations of ith, jth variables.

Since the variance and mean values of variables follow χ^2 and t distribution, respectively. The confidence intervals of the mean value and variance of ith variable, $_zCONF_\alpha^{(i)}$, for a confidence level

Table 1. Calculation of confidence limits for mean value.

Step	Description	Equation
1	Determine solution of c_i for t distribution	$c_i = f_1^{-1}(\frac{1+\alpha}{2})$
2	Calculate mean $\tilde{z}_i$ and variance s_i	$\tilde{z}_i = \frac{1}{\tilde{n}}\sum_{k=1}^{\tilde{n}} z_{ki}, s_i = \frac{1}{\tilde{n}-1}\sum_{k=1}^{\tilde{n}}(z_{ki}-\tilde{z}_i)^2$
3	Calculate μ_i	$\mu_i = \frac{s_i c_i}{\sqrt{\tilde{n}}}$
4	Define confidence limit	$_zCONF_\alpha^{(i)}\{\tilde{z}_i - \mu_i \leq \hat{z}_i \leq \tilde{z}_i + \mu_i\}$

Table 2. Calculation of confidence limits for variance.

Step	Description	Equation
1	Determine solution of c_{1i} and c_{2i} for χ^2 distribution	$c_{1i} = f_2^{-1}(\frac{1-\alpha}{2}), c_{2i} = f_2^{-1}(\frac{1+\alpha}{2})$
2	Calculate $(\tilde{n}-1)s_i$	
3	Calculate μ_{1i} and μ_{2i}	$\mu_{1i} = \frac{(\tilde{n}-1)s_i}{c_{1i}}, \mu_{2i} = \frac{(\tilde{n}-1)s_i}{c_{2i}}$
4	Define confidence limit	$_sCONF_\alpha^{(i)}\{\mu_{2i} \leq \hat{s}_i \leq \mu_{1i}\}$

of $\alpha = 95\%$ or $\alpha = 99\%$ can be obtained as expressed in Table 1 and Table 2, respectively. According to them, we can obtain the upper and lower threshold for each element of matrix S_{zz}. It can be shown as:

$$S_{zz} = \begin{bmatrix} s_{11_L} \leq s_{11} \leq s_{11_U} & s_{12_L} \leq s_{12} \leq s_{12_U} & \cdots & s_{1m_L} \leq s_{1m} \leq s_{1m_U} \\ s_{21_L} \leq s_{21} \leq s_{21_U} & s_{22_L} \leq s_{22} \leq s_{22_U} & \cdots & s_{2m_L} \leq s_{2m} \leq s_{2m_U} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1_L} \leq s_{m1} \leq s_{m1_U} & s_{m2_L} \leq s_{m2} \leq s_{m2_U} & \cdots & s_{mm_L} \leq s_{mm} \leq s_{mm_U} \end{bmatrix} \tag{12}$$

where the subscripts U and L represent the upper and lower limit. The more details of calculation are presented in a paper authored by Kruger [5]. The formula above can be simplified as follows

$$S_{zz_L} \leq S_{zz} \leq S_{zz_U} \tag{13}$$

The accuracy boundaries can be determined according to the confidence interval of correlation matrix. In terms of frobenius norm, the sum of the standard deviation σ is equal to the sum of the discarded eigenvalues of a PCA model.

$$\sigma = \sum_{j=1}^{m} \sigma_j = \frac{1}{\tilde{n}-1} \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{m} e_{ij}^2 = \sum_{k=a+1}^{m} \lambda_k \tag{14}$$

Because the elements in correlation matrix S_{zz} decide the eigenvalues $\lambda_{a+1}, \cdots, \lambda_m$. The estimates of accuracy boundaries can be converted into the following optimization problem.

$$\lambda_{k_{MAX}} = \underset{\Delta S_{zz_{MAX}}}{argmax} \lambda_k (S_{ZZ} + \Delta S_{ZZ_{MAX}})$$
$$\lambda_{k_{MIN}} = \underset{\Delta S_{zz_{MIN}}}{argmin} \lambda_k (S_{ZZ} + \Delta S_{ZZ_{MIN}}) \tag{15}$$

where $\Delta S_{ZZ_{MAX}}$ and $\Delta S_{ZZ_{MIN}}$ are volatilities of S_{ZZ} to determine $\lambda_{k_{MAX}}$ and $\lambda_{k_{MIN}}$. Finally, we get the accuracy boundaries as follows

$$\sigma_{MAX} = \sum_{k=a+1}^{m} \lambda_{k_{MAX}}, \sigma_{MIN} = \sum_{k=a+1}^{m} \lambda_{k_{MIN}} \tag{16}$$

If all the residual variances fall inside the accuracy boundaries, the process is considered to be linear and then we choose PCA for feature extraction, vice versa.

3.2. Fault diagnosis approach

In this article, HMM is integrated with the switched feature extraction method for fault diagnosis. The method mentioned above contains three procedures: nonlinearity measure, feature extraction and fault classification.

Firstly, measure the nonlinearity among original process variables for every mode. The original data are divided into several regions and the accuracy boundaries are determined. If all the residual variances lie inside the accuracy boundaries for each region, PCA is much more appropriate than KPCA for feature extraction and vice versa.

Secondly, the data from every mode are processed respectively by the feature extraction method above selectively. For example, if PCA was selected, the PCs extracted are utilized as observation sequences of HMM.

Thirdly, set up a model library of HMMs by Baum-Welch algorithm, including one HMM for normal operating condition and other HMMs for abnormal conditions. When an unknown fault needs to be classified, it is processed by the suitable method for feature extraction. $P(O|\lambda_i)(i=1,\cdots,N)$ would be calculated, where λ_1 represents HMM for normal operating condition, $\lambda_2,\cdots,\lambda_N$ represent HMMs for all abnormal conditions, and O represents observation sequence. $P(O|\lambda_i)$ represents the probability of the appearance of the observation sequence o with given λ_i. The maximum $P(O|\lambda_i)$ is gained by calculating all the probability, which shows that fault $i\,(i=1,\cdots,N)$ occurred. A flow diagram of the proposed approach is shown in Fig.1.

4. Application of examples

TE process was created to provide a realistic industrial process to test the performance of various monitoring and diagnosis approaches [9]. This process consists of five major units: reactor, condenser, separator, compressor and stripper. It has 41 measured variables (22 continuous process variables and 19 composition variables) and 12 manipulated variables. It includes 21 preset faults which are denoted
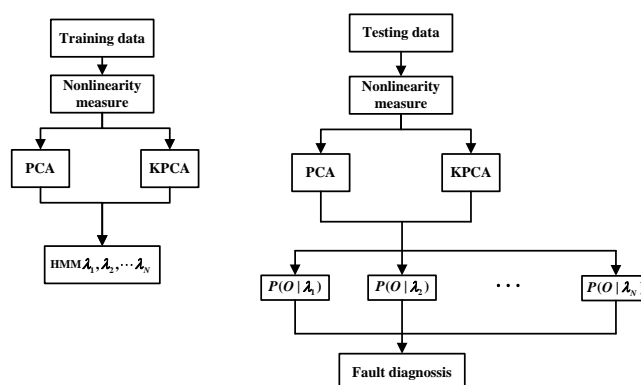
Figure 1. Flow diagram of the proposed approach.

as $\mathrm{IDV}(i)\,(i=1,\cdots,21)$.

In this article, training set consists of three modes, R_1, R_2 and R_3. R_1 represents the normal operating condition, whereas R_2 and R_3 represent two abnormal conditions, IDV(1) and IDV(7). There are 33 variables that have been selected: 22 process measurements and 11 manipulated variables. The agitation speed is not involved because it is not manipulated in simulation process. In addition, 19 composition measurements are excluded [10]. More details are shown in Table 3.

Table 3. Training set.

Pattern	Fault Number	Variable	Type	Time(minute)
R_1	Normal	------	------	480
R_2	IDV(1)	A/C feed ratio, B composition constant	step	480
R_3	IDV(7)	C header pressure loss-reduced availability	step	480

Table 4. Nonlinearity measure applied to the process using three sets.

Region	UCL99	UCL95	LCL99	LCL95	Region1	Region2	Region3
1	5.9241	5.8930	5.1061	4.8307	**6.0042**	**3.9945**	**3.5391**
2	5.8769	5.8439	5.0108	4.7111	**11.6015**	5.8740	**6.4194**
3	12.3295	11.1726	10.6114	9.8594	**17.4527**	12.1941	12.1379

Firstly, we analyse the nonlinear relationships among variables in the training set. R_1 is divided into three regions of 160 samples each. The results of nonlinearity measure of R_1 for each region are shown in Table 4. The accurate boundaries and residuals of each region are presented. The bold figures represent the residuals fall outside of the accurate boundaries. Fig.2 is a graphical representation of the case where the accuracy boundaries were obtained from the first region. According to it, the discarded eigenvalues of PCA model for all data sets fall outside the accuracy boundary. It shows the strong nonlinear relationships among variables. So it is necessary to use KPCA for feature extraction. The other modes are also processed in this way. Because of length limitations of paper, the results of the other faults are not presented here.
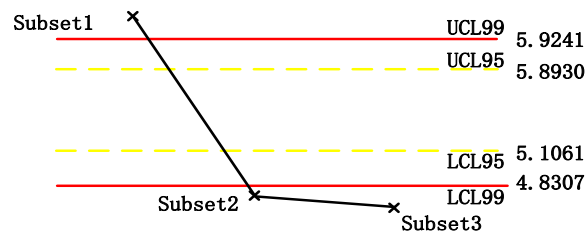
Figure 2. Graphical representation of nonlinearity measure for accuracy bounds for first region.

Secondly, there are three HMMs which need to be trained in this article, one for normal operating condition and the other two for corresponding faults. For each training mode R_i (i=1, 2, 3), the extracted features are used as observation sequences to train parameters λ_i (i =1, 2, 3) of HMM by
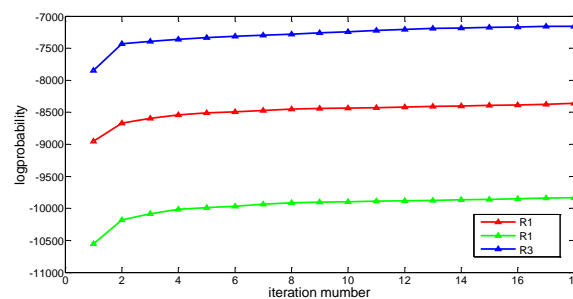

Figure 3. Iteration process for training three HMMs.

Table 5. Testing set.

Pattern	Fault Number	Fault Type	Fault occurs from(minute)
T_1	Normal	------	------
T_2	IDV(1)	step	160
T_3	IDV(7)	step	160

Baum-Welch method. The results are shown in Fig.3. It is clear that three training processes are quite different from each other, which indicates the modes of operation present distinct characteristics.

There are also three modes in the testing set, T_1, T_2 and T_3. T_1 represents the normal operating condition, whereas T_2 and T_3 represent two abnormal conditions, IDV (1) and IDV (7). All faults in the testing set are same with the faults of the training set, and each begins with normal operation and introduces a fault after 160 minutes. The more details are shown in Table 5.

The classifying results of three modes in the testing set are shown in Fig. 4, Fig. 5, Fig. 6. The ordinates represent $\log P(O|\lambda_i)(i = 1, 2, 3)$, the log probabilities of the appearance of the observation sequence sequence O with given λ_i in the model library, and abscissa represents time. From all figures, we can see that different faults can be identified correctly by the proposed fault diagnosis method. In Fig. 4, we can see that all the method can identify the fault correctly, but it is important to note that $P(O|\lambda_2)$ and $P(O|\lambda_3)$ are close in (b) and (c). This means that R_2 and R_3 are not distinguished well in the traditional method. In Fig. 5, either HMM based on PCA or KPCA cannot identify the fault accurately, while the switched method has a better effect. According to the results shown in Fig. 6, the proposed method performs better although HMM based on PCA could identify the fault.
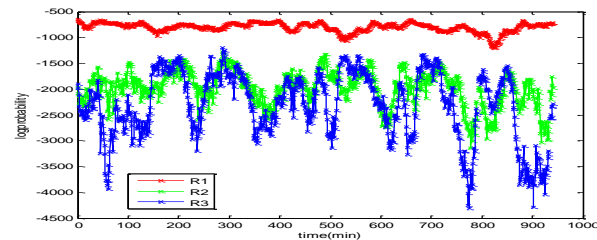
5. Conclusion

Switched fault diagnosis for industrial processes based on hidden Markov model is proposed. Generally speaking, the swiched fault diagnosis approach can enhance the diagnostic capacity of TE process because the method properly takes the characteristics of process data into consideration. By analysing characteristic of data, we overcome the insufficient of the traditional methods which sometimes extracts data features by inappropriate method. By implementing the simulation on TE, we are able to get the results which exhibit that the proposed diagnosis approach possesses a superior performance when it is compared to the traditional diagnosis based on HMM.

This study shows the superiority of the switched fault diagnosis approach for industrial processes based on HMM. However, there are some issues to be further addressed. The effectiveness of the proposed method is demonstrated by using the simulated process data. Future work can implement the proposed method with the real world industrial data. In this article we just consider linearity and non-linearity by using PCA and KPCA. In the future, we can take more characteristic of data into consider-ration, such as Gaussianity and non-Gaussianity by enhancing the accuracy of fault diagnosis for different modes.
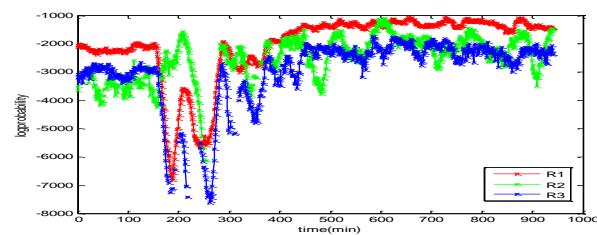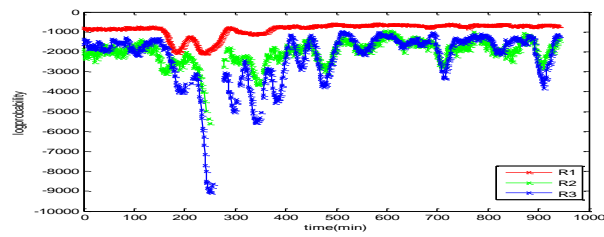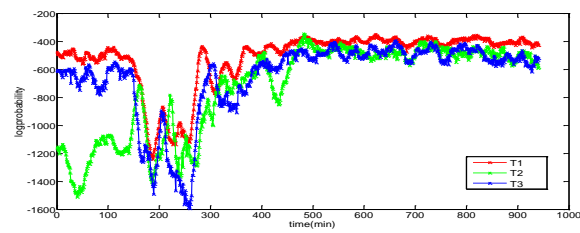

(a)

(b)

(c)

Figure 4. Diagnostic result for T1: (a) result based on switched method (b) result based on KPCA (c) result based on PCA.
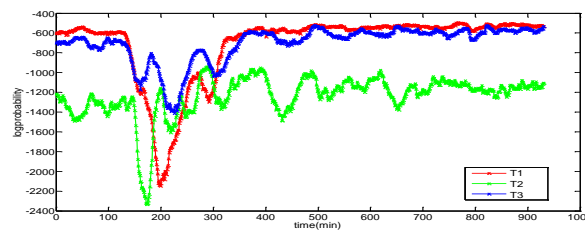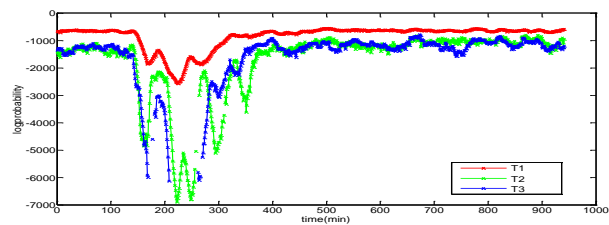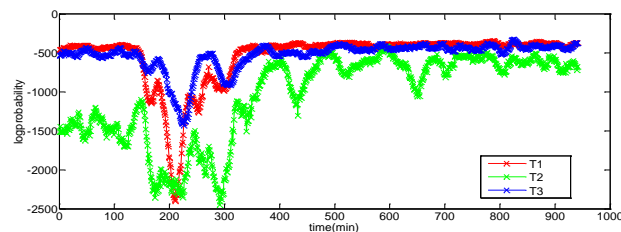

(a)

(b)

(c)

Figure 5. Diagnostic result for T2: (a) result based on switched method (b) result based on KPCA (c) result based on PCA.

(a)

(b)

(c)

Figure 6. Diagnostic result for T3: (a) result based on switched method (b) result based on KPCA (c) result based on PCA.

References

[1]   M. R. Rashid and Y. Jie, Hidden Markov model based adaptive independent component analysis approach for complex chemical process monitoring and fault detection, *Ind. Eng. Chem. Res*, vol. 51, pp. 5506–5514, Mar. 2012

[2]   L. J. Cao, K. S. Chua, W. K. Chong, H. P. Lee and Q. M. Gu, A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine, *Neurocomputing*, vol. 55, pp. 321–336, Mar. 2003

[3]   Z. Q. Ge, Z. H. Song and F. R. Gao, Review of recent research on data-based process monitoring, *Ind. Eng. Chem. Res*, vol. 52, pp. 3544–3562, Feb. 2013.

[4]   C. Y. Cheng, C. C. Hsu and M. C. Chen, Adaptive kernel principal component analysis (KPCA) for monitoring small disturbances of nonlinear processes, *Ind. Eng. Chem. Res*, vol. 46, pp. 2054–2262, Jan. 2010.

[5]   U. Kruger, D. Antory, J. Hahn, G. W. Irwin and G. M. Cullough, Introduction of a nonlinearity measure for principal component models, *Computers & Chem. Eng*, vol. 29, pp. 2355–2362, July 2005.

[6]   L. R. Rashid, A Tutorial on the hidden Markov models and selected applications in the speech recognition, *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[7]   S. Y. Zhou, J. M. Zhang and S. Q. Wang, Fault diagnosis in industrial processes using principal component analysis and hidden Markov model, in *Conf. Rec. 2004 American Control Conference*, pp. 5680–5685.

[8]   M. Stone. Cross-validatory choice and assessment of statistical prediction (with discussion) , *Journal of the Royal Statistical Society, Series B Methodologyical*, vol. 36, pp. 111–133.

[9]   A. Singhai and D. E. Seborg, Evaluation of a pattern matching method for the Tennessee Eastman challenge process, *Journal of Process Control*, vol. 16, pp. 601–613, Mar. 2006

[10] Z. Q. Ge and Z. H. Song, Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors, *Ind. Eng. Chem. Res*, vol. 46, pp. 2054–2063, Mar. 2007