A Hybrid Fault Detection and Diagnosis System Based on KPCA and DDAG

Qiang Gao[1,2], Guojing Wang[2], and Xiaopeng Hao[2]

[1] School of Electrical Engineering & Automation of Tianjin University, 300072, China
[2] Tianjin Key Laboratory for Control Theory & Applications in Complicated Systems, Tianjin University of Technology, Tianjin, 300384, China
{gaoanbei,Chaoyue91}@163.com,88279093@qq.com

Abstract. In order to improve the capability of the fault diagnosis, this paper introduces the Decision Directed Acyclic Graph (DDAG) algorithm and establishes a new detection and diagnosis system combing the DDAG with the Kernel Principal Component Analyses (KPCA) method. The hybrid system uses KPCA and DDAG to detect and identify the fault. A specific description of the principles and procedures about how to use KPCA method and DDAG is given. The new detection and diagnosis system has an excellent performance in the fault detection and diagnosis of the Tennessee-Eastman (TE) process. This paper gives a new way to research the fault detection and diagnosis in industrial nonlinear system.

Keywords: Fault diagnosis, KPCA, DDAG, TE, nonlinear system.

1 Introduction

With the rapid development of mass production and new technology, automatic control systems for modern industrial processes have been expanding to increasing complexity and larger scale. Its timely and accurate diagnosis can reduce the chemical production downtime, increase production safety. In terms of efficiency point of view or perspective of safety, fault diagnosis has the necessity and urgency. For complex industrial processes, variables often have a nonlinear relationship. To solve this problem, B.Schölkopf and others proposed a nonlinear process of detection of kernel PCA (KPCA [1-3]), with high practical value.

DDAG [4,5], a new learning machine based on Statistical Theory, is a Multi-category Support Vector Machines (M-SVM).It is a new learning method based on a limited sample of learning theory. Taking advantage of the original data, it can be effective for fault identification. In this paper, KPCA and DDAG are combined to enhance the fault detection and diagnostic capability.

2 KPCA

2.1 The Principle of KPCA

Through the nonlinear mapping φ , firstly the original input space ($x_1, x_2, \cdots, x_n \in R^m$, where n is the number of samples, m is the dimension of

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 549–555.
springerlink.com © Springer-Verlag Berlin Heidelberg 2012

measured variables) is mapped to a high-dimensional feature space F (as is $\varphi : R^m \to F$). Then the nonlinear problem of input space is transformed into the linear problem of feature space. The mapping x_i will be denoted by $\varphi(x_i) = \varphi_b$, and the covariance matrix of the feature space F can be expressed as:

$$C^F = \frac{1}{N} \sum_{i=1}^{N} \varphi_i \varphi_i^T \tag{1}$$

Set λ as the eigenvalue of the matrix C^F and v as the eigenvector, then:

$$\lambda v = C^F v \tag{2}$$

Eigenvector v can be obtained by mapping the sample of the feature space:

$$v = \sum_{i=1}^{N} \alpha_i \varphi_i \tag{3}$$

From equation (1), the v corresponding to the maximum value of λ is the first principal component of the feature space and the v corresponding to the minimum is the last main principal, then $\lambda v = C^F v$ is equivalent to:

$$\lambda < \varphi_k, v > = < \varphi_k, C^F v > , \varphi_k = \varphi(x_k) , k = 1, \cdots, N \tag{4}$$

$< x, y >$ is the dot product of x x and y . Combing with equation (3) and (4), we can get:

$$\lambda \sum_{i=1}^{N} \alpha_i < \varphi_k, \varphi_i > = \frac{1}{N} \sum_{i=1}^{N} \alpha_i < \varphi_k, \sum_{j=1}^{N} \varphi_i > < \varphi_p \varphi_i > \tag{5}$$

Define the matrix $K \in R^{N \times N}$, and set that $[K]_{ij} = k_{ij} = < \varphi_{ij} \varphi_j >$, so by equation (3) we can get:

$$\lambda N \alpha = K \alpha , \alpha = [\alpha_1, ..., \alpha_k]^T \tag{6}$$

Before the analysis of principal component in the feature space F , you should make standardized. Substitution matrix K with the following equation:

$$\hat{K} = K - I_N K - K I_N + I_N K I_N \tag{7}$$

In the equation (7), I_N is equal to the multiplication of $1/N$ and a matrix $E = R^{N \times N}$. So the analysis of principal component in the feature space is equivalent to the eigenvalue of equation (6). Combined with equation (6) and (3), eigenvectors v of matrix C^F can be obtained by the eigenvector α of the matrix K , and satisfies:

$$\langle v_k, v_k \rangle = 1 \,,\; k = 1, \cdots, p \,. \tag{8}$$

And the p is the principal number. Then it should calculate the principal component by calculating the projection of the mapping data on the eigenvectors v_k.

$$t_k < v_k, \varphi(x) >= \sum_{i=1}^{N} \alpha_i^k < \varphi(x_i), \varphi(x) >= \sum_{i=1}^{N} \alpha_i^k k(x_i, x) \tag{9}$$

To solve the eigenvalue problem of equation (6), and calculate the main principal vector of feature space from the input space directly by using (9), we introduce the kernel function of the form of dot product in the feature space, namely $k(x, y) =< \varphi(x), \varphi(y) >$, to avoid a direct calculation of non-linear mapping.

2.2 Fault Detection Strategies of KPCA

KPCA-based fault detection method uses the T^2 and SPE statistics to detect faults in the feature space. T^2 statistic is the square of the standard principal component vector and is defined that:

$$T^2 = [t_1, ..., t_p] \Lambda^{-1} [t_1, ..., t_p]^T \tag{10}$$

In the equation (10), t_i can be get by equation (9), Λ^{-1} is the eigenvalue of principal composition.

The statistical control limit of T^2 can be obtained by F-distribution:

$$T_{\lim}^2 = \frac{p(N-1)}{N(N-p)} F_\alpha (p, N-p) \tag{11}$$

In the equation above: p is the number of the principal component, N is the number of samples.

Squared prediction error (SPE) is a measure of changes in external data of the model. It is defined that:

$$SPE = \left\| \varphi(x) - \varphi(y) \right\|^2 = \sum_{i=1}^{N} t_i^2 - \sum_{i=1}^{p} t_i^2 \tag{12}$$

The control limits of the SPE statistic is:

$$SPE_{\lim} = g X_h^2 \tag{13}$$

In the above equation: g and h are the constant coefficients of the mean and variance about the SPE.

Application of KPCA for fault diagnosis of industrial processes, firstly, pretreats the collected data from normal condition, and carries on kernel principal component analysis to establish the kernel principal component model which reflects normal operation. Secondly calculate the T^2 and SPE statistic to determine the control limits, and then calculate the observed sample statistics. Excess of control limits stands for system may have faults at the time. [6]

3 Applied Research

3.1 Tennessee Eastman Process

TE (Tennessee Eastman) process is proposed by Downs and Vogel of the U.S. Eastman Chemical Company in 1990. It is a real chemical process model for the development, research and evaluation of process control technology and monitoring methods. Many foreign scholars and experts use it as a data source to research the control algorithm, optimization and fault diagnosis. [7, 8]

TE process has five major unit operations: the reactor, the product condenser, a vapor-liquid separator, a recycle compressor cycle and a product stripper. There are four reactions and two kinds of products. The process has 12 manipulated variables and 41 measurements (including 22 continuous measurement variables and 19 composition measurements). TE process control structure shown in Fig. 1.

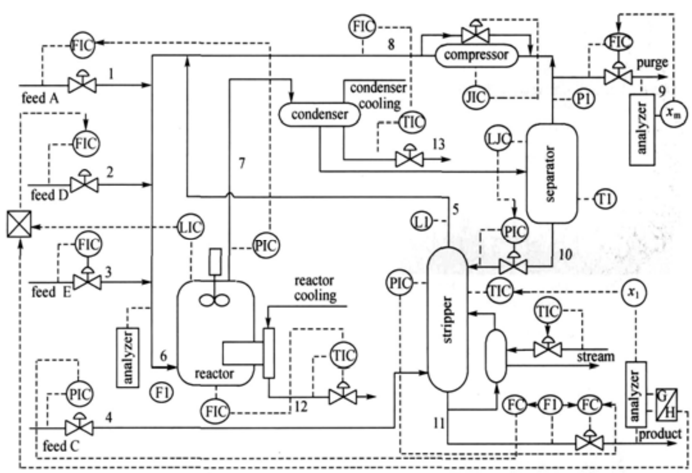

Fig. 1. Tennessee - Eastman process control structure

3.2 Simulation Results and Analysis

KPCA algorithm is applied to the fault detection of TE process. In the simulation, the kernel function uses Gaussian kernel: $k(x,y)=\exp(-\|x-y\|^2/\sigma)$, kernel function σ is 500, and monitor the SPE statistic. Figures 2 to 5 are the SPE charts of the fault 5, 7,8,12.

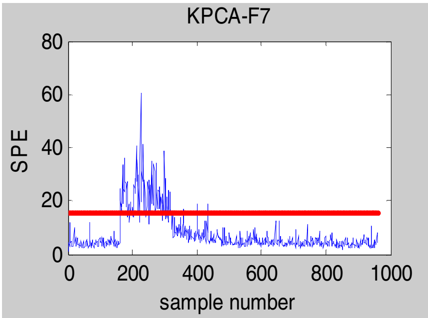

Fig. 2. Detection of fault 5

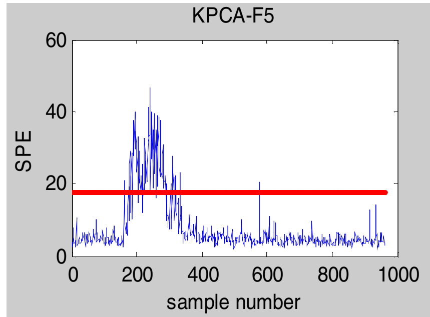

Fig. 3. Detection of fault 7

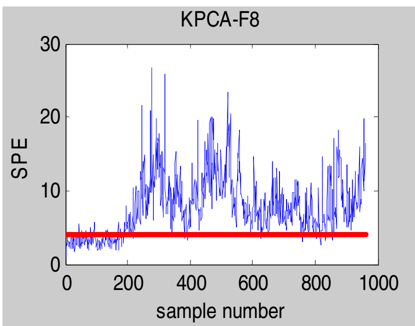

Fig. 4. Detection of fault 8

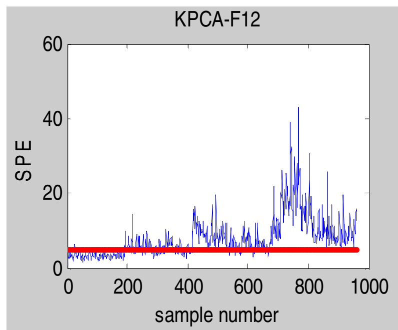

Fig. 5. Detection of fault 12

Fault 5, 7 belong to a step change and fault 8 and 12 belong to random changes in the system fault. When the fault occurs (after the 160 data points), SPE statistics produced significant changes. The result of KPCA in the fault diagnosis can clearly show whether the fault occurred. The use of KPCA method for chemical process fault detection can achieve better results, but it is Lack of capacity in the identification of fault types.

4 Improved Research

4.1 DDAG

DDAG is a new learning framework for support vector machine. [9] As a more complete multi-class classification algorithm, it can solve the identification of fault type excellently.

DDAG is proposed to solve the phenomenon of refusal and misclassification of “One-Against-One "SVM. In the training phase, the sub-classification of DDAG is like the sub-classification of the “One-Against-One” SVM, and need construct the surface of each classification between two. For a K-class classification problem, $k(k-1)/2$ classifiers are needed. In the testing phase, the approach constructs an acyclic graph with two value-oriented and a root, which have $k(k-1)/2$ leaf nodes and k internal nodes. Each internal node corresponds to a binary support vector machine classifier, and each leaf node corresponds to a class mark. When an unknown sample needed to be the classified, it starts from the top of the root and reaches a leaf node until the bottom, which is the type of unknown sample. Node of the next layer is selected According to the classification of root node. [10]

Before the Fault diagnosis of DDAG, code known fault in accordance with binary numbers, and give the corresponding explanation. To Tennessee - Eastman process, for example, a known fault libraries can be established. According to T^2 and SPE charts, selecting the fault data as DDAG training input vectors and selecting binary number as the target vector, Stable DDAG fault diagnosis model can be obtained after training. When we enter a sample of fault, the type of fault can be drawn.

4.2 Hybrid diagnostic system of KPCA-DDAG

Because of the deficiency of the KPCA method in fault identification and diagnosis, DDAG is introduced to establish hybrid diagnostic system of KPCA-DDAG. DDAG fault classification method is precisely to make up for lack of KPCA. Hybrid diagnostic system of KPCA-DDAG is divided into two parts, KPCA detection and DDAG fault Identification. Flow Chart of Hybrid fault diagnosis system is shown in Figure 6.

For the hybrid diagnostic system, process data is firstly transported to the detection module of KPCA for fault detection. If no fault is detected, the next set of data will be transported to the detection module. When T^2 and SPE statistic exceed its threshold, it indicates there is a fault. Then the program enters the DDAG classification model, type of the fault will be identified out.

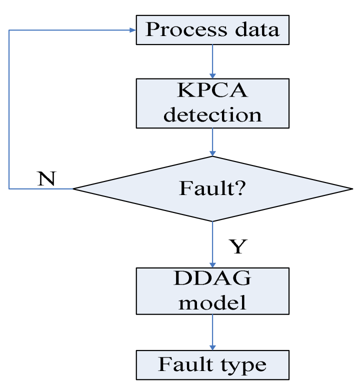

Fig. 6. Flow Chart of Hybrid fault diagnosis system

5 Conclusion

In this paper, the DDAG and KPCA have been combined to improve the fault diagnosis system. It uses DDAG classification model to identify the type of fault when there is a fault detected by KPCA. The new way is given to enhance the capability of fault detection and diagnosis by the establishment of the new hybrid diagnostic system. And it shows excellent performance in the fault detection and diagnosis of Tennessee - Eastman process.

References

1. Schölkopf, B., Smola, A., Müller, K.-R.: Kernel principal component analysis. In: Advances in Kernel Methods-Support Vector Learning, pp. 196–201. MIT Press, Cambridge (1999)
2. Kim, K.I., Park, S.H., Kim, H.J.: Kernel principal component analysis for texture classification. IEEE SPL 8(2), 39–41 (2001)
3. Lee, J.M., Yoo, C., Choi, S.W., et al.: Nonlinear process monitoring using kernel principal component analysis. Chemical Engineering Science 59(1), 223–234 (2004)
4. Platt, J.C., Cristianini, N., Shawe-Taylor, J.: Large margin DAGs for multiclass classification. Advances in Neural Information Processing Systems 12, 547–553 (2000)
5. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines. University Press, Cambridge (2000)
6. Choi, S.W., Lee, C., Lee, J.M., et al.: Fault detection and identification of nonlinear processes based on kernel PCA. Chemometrics and Intelligent Laboratory Systems 75(1), 55–67 (2005)
7. Downs, J.J., Vogel, E.F.: A plant-wide industrial process control problem. Computers and Chemical Engineering 17(3), 245–255 (1993)
8. Lyman, P.R.: Plant-Wide Control Structures for the Tennessee Eastman Process, Lehigh University (1992)
9. Hsu, C., Lin, C.: A Comparison of Methods for Multiclass Support Vector Machines. IEEE Trans. on Neural Networks 13(2), 415–425 (2002)
10. Chapelle, O., Vapnik, V., Bousquet, O., Mukherjee, S.: Choosing Multiple Parameters for Support Vector Machines. Machine Learning 46(1-3) (2002)