

# KPCA-based Visual Fault Diagnosis for Nonlinear Industrial Process

Jiahui Yu<sup>1</sup> and Hongwei Gao<sup>1</sup> and Zhaojie Ju<sup>2,3</sup>

<sup>1</sup>College of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China

<sup>2</sup>State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

<sup>3</sup>University of Portsmouth, Portsmouth PO1 3HE, U.K.  
yujiahui77@163.com

**Abstract.** With the increasingly large-scale, continuous, and complicated chemical process, it is particularly important to ensure the stability and safety of the production process. However, in past studies, the accuracy of fault diagnosis and the degree of system visualization are still insufficient. Here, in order to solve these problems, a visual fault diagnosis system based on LabVIEW and Matlab is designed. First, the system uses LabVIEW interface design, applying Matlab to compile the algorithm program, which makes the system has a powerful data calculation and processing functions, as well as a clear visual interface, the system design also optimizes the communication interface. Second, the typical chemical production process TE (Tennessee Eastman) process is the subject of systematic testing. Additionally, because most of the industrial processes are non-linear, the fault diagnosis method based on Kernel Principal Component Analysis (KPCA) is used in the system design, and the implementation process of this method is elaborated. Finally, the system achieves the functions of TE process data acquisition, data preprocessing, and fault diagnosis lamps. A large number of simulation results verify the effectiveness of the proposed method. The system has entered the stage of laboratory application and provides a good application platform for the research of fault diagnosis of complex systems such as chemical process control.

**Keywords:** Fault Diagnosis; TE Process; KPCA; Visualization System.

## 1 Introduction

The process control method with industrial process as the main research object has the characteristics of systematic, complexity and diversity. In order to reduce accidents in the process and reduce losses to a certain extent, a set of fault diagnosis systems need to be designed to further determine what kind of faults have occurred, where and how far the faults occurred, and the causes.

Fault diagnosis is usually through the information collected by the measuring equipment in the diagnostic system to analyze and judge the status of the process, and to find

the time when the fault occurred and the location and mode of the fault. Isermann clearly defines the type, location, magnitude and time of failure as the definition of fault diagnosis [1]. The United States Tennessee Eastman Chemical Co, Ltd. put forward a typical platform for research and development of process industry fault diagnosis in FORTRAN source code. Later, Bathelt A et al. performed the simulation of the process on the Matlab, and continued to develop [2].

The principal component analysis (PCA) method is currently widely used in process analysis and monitoring [3]. In 1933, Hotelling proposed Principal Component Analysis (PCA), which is a multivariate statistical method that converts multiple indicators into several comprehensive indicators according to certain rules and actual needs without losing information. It is mainly used to eliminate multivariate multicollinearity to achieve dimensionality reduction [4–5]. However, it is a linear method that does not apply to nonlinear processes. Aiming at the nonlinear characteristics of modern industrial processes, various nonlinear monitoring methods have emerged, and kernel principal component analysis is one of them [6]. It considers the nonlinear characteristics of the process, can effectively extract the nonlinear information of the data, and has no complicated nonlinear calculation in the implementation process, and has fewer adjustable parameters, so it is widely used in feature extraction, face recognition, image processing and faults diagnosis and other fields [7–9]. In the KPCA method, the correct choice of the kernel function largely determines the quality of the system's nonlinear feature extraction. Therefore, the choice of the type of kernel function and its parameters has a great impact on the results of fault detection [10].

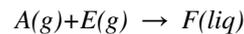
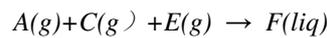
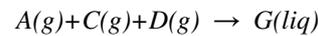
All kernel functions can be divided into two types: global kernel function with strong generalization ability, weak learning ability, and local kernel function with strong learning ability and weak generalization ability. Using only one kernel function has certain limitations when analyzing the performance of the system. It can be known from the nature of the kernel function that a linear combination of any two kernel functions can form a new kernel function [11]. Therefore, this paper combines a Gaussian radial basis kernel function with strong learning ability and a polynomial kernel function with strong generalization ability to construct a mixed kernel function that has both advantages. The main contributions of this paper are summarized as follows.

After analyzing the characteristics of common kernel functions, this paper mixes the global kernel function with the local kernel function and improves a kernel function. The improved kernel function has powerful learning ability and generalization ability. The hybrid kernel function is applied to the KPCA method to achieve the goal of improving the traditional KPCA method. The improved KPCA fault detection method is applied to typical nonlinear processes and TE processes, and the fault diagnosis method of this KPCA is introduced. The principle and operation process of the KPCA are described in detail, and the fault detection method and the application in the normal chemical process, and the program is constructed and tested to verify the effectiveness of the improved method. The algorithm program compiles and visualizes the interface design, which makes the system have powerful data calculation and processing functions, as well as a clear visual interface. Typical nonlinear process and TE process monitoring results show that this method has higher accuracy in fault detection.

The rest of this paper is organized as follows: Chapter 2 reviews the relevant content of the TE process. The third chapter introduces the process of the proposed method. The fourth chapter gives the analysis and comparison of various experimental results. The fifth part summarizes the work of this paper.

## 2 TE process

The Tennessee-Eastman Process (TEP) is a realistic industrial process and a method of fault detection and diagnosis in process control evaluation [12–13]. It contains five operating units: reactor, condenser, recycle compressor, separator and stripper. Including eight components: A, B, C, D, E, F, G, H enter the reactor gas ACDE and inert component B, will form liquids G and H in the reactor, the various reactions of the reactor are [14–15]:



The designed fault diagnosis system performed 22 simulations on the data in the training set, and each simulation process was different [16]. Each simulation time is 25 hours. When the simulation starts, normal system operation is performed. When the simulation time reached 1h, a fault was introduced. For each simulation, N=500 observations were generated. Of the 500 observations, the last 480 observations were faulty. The data in the tester will be subjected to 22 different simulations. In the 22 simulations, one of them was operating in the normal state, that is, fault 0: the other could operate in a stable state, namely fault 21. The remaining 20 simulations, from fault 1 to fault 20, were run under different fault conditions, each for a duration of 48 hours. When the simulation started, normal system operation was performed. When the simulation time reaches 8h, the fault is reintroduced. For each simulation, N=960 observations were generated. Among them, the first 160 observations are normal, and the last 800 observations are faulty.

## 3 Improved Fault Diagnosis Algorithm Based on KPCA

### 3.1 The basic principle of KPCA

The kernel principal component analysis method is an objective weighting method. It fully utilizes all the information of the original variables, determines the weight coefficient of the principal component according to the variance contribution rate, and can comprehensively reflect the importance degree of each variable, so that the comprehensive evaluation result More effective [17–18].

4

Assume that the training sample is  $x_1, x_2, \dots, x_M$ . The input space is  $\{x_i\}$ . The basic idea of the KPCA method is to map the input space to some high-dimensional space (often called the feature space) in some implicit way, and to implement PCA in the feature space [19]. Assuming the corresponding mapping is  $\Phi$ , Its definition is as follows

$$x \mapsto \xi = \Phi(x) \quad (1)$$

The kernel function maps implicitly implemented points to F through mapping, so that the data in the mapped feature space satisfies the concentration condition, which is

$$\sum_{\mu=1}^M \Phi(x_{\mu}) = 0 \quad (2)$$

Then the covariance matrix in the feature space is:

$$C = \frac{1}{M} \sum_{\mu=1}^M \Phi(x_{\mu}) \Phi(x_{\mu})^T \quad (3)$$

Finding the eigenvalues  $\lambda \geq 0$  and eigenvectors of  $C$

$$V \in F \setminus \{0\}, \quad CV = \lambda V \quad (4)$$

That is

$$(\Phi(x_v) \cdot CV) = \lambda (\Phi(x_v) \cdot V) \quad (5)$$

Considering that all eigenvectors can be expressed as a linear expansion of, ie

$$v = \sum_{i=1}^M \alpha_i \Phi(x_i) \quad (6)$$

That is

$$\frac{1}{M} \sum_{\mu=1}^M \alpha_{\mu} \left( \sum_{w=1}^M (\Phi(x_v) \cdot \Phi(x_w) \Phi(x_w) \Phi(x_{\mu})) \right) = \lambda \sum_{\mu=1}^M (\Phi(x_v) \cdot \Phi(x_{\mu})) \quad (7)$$

Among them  $v=1, 2, \dots, M$ , defining  $M \times M$  dimensional matrix K

$$K_{\mu\nu} := (\Phi(x_{\mu}) \cdot \Phi(x_{\nu})) \quad (8)$$

Then the expression (7) can be simplified as

$$M \lambda K \alpha = K^2 \alpha \quad (9)$$

Obviously satisfied

5

$$M\lambda\alpha = K\alpha \quad (10)$$

Solving (10) can get eigenvalues and eigenvectors. The projection of the test sample in the feature vector space can be expressed as

$$(\nu^k \cdot \Phi(x)) = \sum_{i=1}^M (\alpha_i)^k (\Phi(x_i), \Phi(x)) \quad (11)$$

Replace the inner product with a kernel function.

$$(\nu^k \cdot \Phi(x)) = \sum_{i=1}^M (\alpha_i)^k K(x_i, x) \quad (12)$$

When (2) is not established, adjustments need to be made

$$\begin{aligned} \Phi(x_\mu) &\rightarrow \Phi(x_\mu) - \frac{1}{M} \sum_{v=1}^M \Phi(x_v) \\ \mu &= 1, \dots, M \end{aligned} \quad (13)$$

Then the nuclear matrix can be modified to

$$\begin{aligned} K_{\mu\nu} &\rightarrow K_{\mu\nu} \\ &- \frac{1}{M} \left( \sum_{w=1}^M K_{\mu w} + \sum_{w=1}^M K_{w\nu} \right) + \frac{1}{M^2} \sum_{w,\tau=1}^M K_{w\tau} \end{aligned} \quad (14)$$

### 3.2 Algorithm implementation process

According to the principle of KPCA, the following process is obtained. Writing a batch of data of  $n$  obtained index ( $m$  samples for each index) into a  $(m \times n)$  dimensional data matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}_o$$

Calculate the parameters in the Gaussian radial kernel function by calculating the above kernel matrix, and then calculate the kernel matrix  $K$  by equation (8).  $KL$  is obtained from (14) modified kernel matrix. The Jacobi iterative method is used to calculate the characteristic value  $\lambda_1, \dots, \lambda_n$  of  $KL$ , that is, the corresponding feature vectors  $v_1, \dots, v_n$ . The eigenvalues are sorted in descending order (by sorting) to get feature vectors  $\lambda_1' > \dots > \lambda_n'$ , and adjusted the feature vector accordingly  $v_1', \dots, v_n'$ .  $\alpha_1, \dots, \alpha_n$  orthonormalized feature vectors by the Schmidt normalization method.

Calculate the cumulative contribution rate  $B_1, \dots, B_n$  of the eigenvalues. According to the given extraction efficiency  $p$ , if  $B_i \geq p$ , extract  $\alpha_1, \dots, \alpha_i$  main components.

6

Calculate the projection  $Y = KL \cdot \alpha$  of the modified kernel matrix  $X$  on the extracted feature vector, where  $\alpha = (\alpha_1, \dots, \alpha_i)$ .

The data after KPCA dimensionality reduction is projection  $Y$ .

## 4 System design and testing

### 4.1 System design process

In the process control modeling and simulation, Matlab is often used, but in order to compensate for the lack of interface development and acquisition of signals, LabVIEW software is always used [20]. In order to make the system versatile, the operating system for fault diagnosis operation is Windows 7 (64bit), and its development tools are Matlab/Simulink (2016a) and LabVIEW2012.

In this paper, our system has excellent interface display features that can perfectly display the TE process on the host computer. System design uses a combination of SIT and Simulink compiler, a good implementation of the interface display function, while solving the problem of the communication interface. The system combines Matlab with powerful data processing and LabVIEW with high-efficiency image processing. The entire system includes the functions of data acquisition, data preprocessing, and fault diagnosis of the TE process.

### 4.2 System test

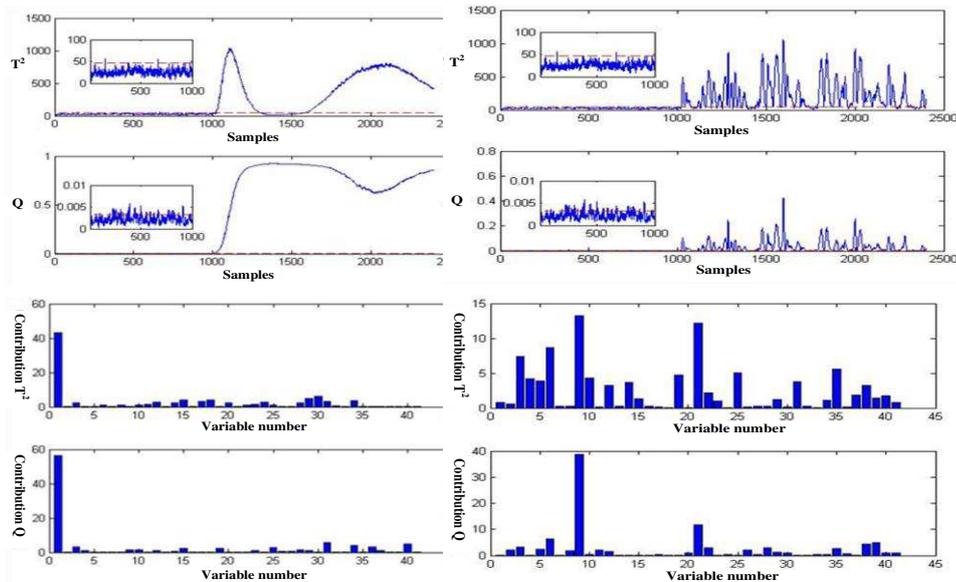
In the process of processing data, it is not necessary to manually input each set of data. It only needs to be imported by computer according to the address of data storage. In this way, data processing can become accurate, rapid, and reduced. The computational cost greatly improves our efficiency, enables real-time monitoring of chemical processes and facilitates the detection of failures at any time [21].

Figure 1 shows the login interface of the system. Enter the correct account password to log in to the system.



**Fig. 1.** login screen

After logging in to the system, enter the system function interface, where you can choose to implement the corresponding functions of the system, as shown in Figure 2.

**Fig. 2.** Experiment platform interface**Fig. 3.** Fault 2 Simulation**Fig. 4.** Fault 11 Simulation

Before the simulation, the program was written for the entire process, and then the program written and able to run correctly was run on LabVIEW. Theoretically, it was clearly seen in the figure that the fault was added at the 1000th sampling point. Causes the trend of the polyline to fluctuate. This is because the man-in-the-middle joins the fault. Therefore, in the normal chemical production process, if the image above is obtained after running the program, it is not difficult to know the time when the failure occurred, so that it can be used to prevent the occurrence of failures. The following can select several representative faults to be added to the TE process and simulate them.

8

The specific simulation results and simulation images are as follows. The  $T^2$  statistic mainly describes the change of the data in the principal element space and reflects the main operating characteristics of the industrial process. The  $Q$  statistic is built in the residual space of the monitoring model to summarize the sum of the errors caused by the modeling.

Enter the fault 2 in the process, input the fault 2 data into the program, run the program, you will get the simulation image of Figure 3. According to the image, the fault occurred after the 1000th data, and the fault was mainly caused by fault 1.

Add fault 11 to the process, enter the data of fault 11 into the program, and run the program, as shown in Figure 4. According to the image, the fault appears after the 1000th data, and the fault is mainly caused by the fault 9 and the fault 21. Adding fault 14 to the process, input the data of fault 14 to the program, and run the program, as shown in Figure 5. From the image, it can be seen that the fault occurred after the 1000th data, and the fault is mainly caused by the fault 28 and the fault 34.

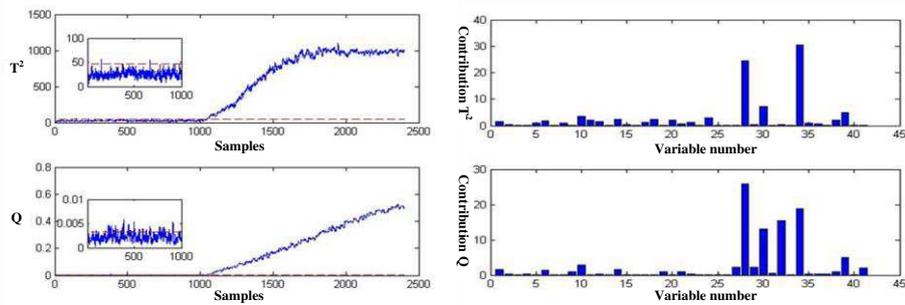


Fig. 5. Fault 14 Simulation

## 5 Conclusion

Through the joint use of both useful platforms, fault diagnosis of the TE process is realized. The fault diagnosis system based on LabVIEW and Matlab not only has powerful data processing functions, but also has a perfect visual interface. The improved KPCA algorithm used in this system solves the problem that nonlinear engineering is difficult to reduce dimensions.

In this paper, the improved KPCA method is used to detect faults in nonlinear processes. The method uses a hybrid kernel function to perform high-dimensional mapping on data. Since the mixed kernel function has the advantages of both types of kernel functions, the mapped data has strong generalization ability. The principal element contribution rate and the fault detection rate are used as the dual indicators of the hybrid kernel function parameter optimization, so that the mixed kernel function to the optimal parameters is superior to the single kernel function in feature extraction and fault detection. The online detection of typical nonlinear processes and TE process data shows that the improved KPCA method is more effective.

The system has a clear and visible interface and has been systematically tested to verify its stability. It opens up a new way of thinking in terms of fault diagnosis and model prediction.

## References

1. Downs Vogel E F. A Plant-wide Industrial Process Control Problem [J]. *Computer & Chemical Engineering*, 1993, 17(3): 245 ~ 255.
2. Bathelt A, Ricker L N, Jelali M. Revision of the Tennessee Eastman Process Mode [C]. *International Symposium on Advanced Control of Chemical Processes*. British Columbia, Canada: IFAC, 2015: 309 ~ 314.
3. Shan Yan, Li Juanling, Sun Dujing. Research on Parallel Data Flow Based on Principal Component Analysis[J]. *Journal of Nanjing University of Posts and Telecommunications*. 2015.10, Vol.35, No.5:100-102.
4. Zhang Yongfang, Yu Le. Research on parallel data mining algorithm based on Hadoop platform[D]. *Anhui University of Science and Technology*. 2016.6, 5-25.
5. WANG P L, YAN W J. New monitoring method for dynamic non-Gaussian process[J]. *Chinese Journal of Scientific Instrument*, 2009, 30(3): 471-476.
6. SCHOLKOPF B, SMOLA A, MULLER K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. *Neural Computation*, 1998, 10(5): 1299-1319.
7. Zhao Jianhua. Research on face recognition based on combined kernel function KPCA[J]. *Computer Engineering and Design*, 2014, 35(2): 631-635.
8. Wan Jiaqiang, Wang Yue, Liu Yu. Improved KPCA feature extraction of classification data[J]. *Computer Engineering and Design*, 2010, 31(18): 4085-4092.
9. Liu Chunyan, Yu Chunmei. TE Fault Diagnosis Based on Improved Kernel Principal Component Analysis[J]. *Computer Measurement & Control*, 2016, 24(10): 36-41.
10. Li Xue. Research on Statistical Modeling and Monitoring Method of Intermittent Process Based on Mixed Kernel Function [D]. *Shenyang: Northeastern University*, 2014.
11. YUAN Zhe, SHI Huai-tao. Fault Diagnosis Method Based on Step-by-Step Dynamic Kernel Principal Component Analysis[J]. *Shenyang Jianzhu University Journal (Natural Science Edition)*, 2013, 29(6): 1092-109
12. [12] Li J C, Hu N Q, Qin G J, et al. Optimizing method for fault diagnosis strategy based on bayesian networks[J]. *Control and Decision*, 2003, 18(5): 568-572.
13. Downs J J, Vogel E F. Plant-wide industrial process control problem[J]. *Computers and Chemical Engineering*, 1993, Vol.17, No.3: 245-55.
14. CUI Guanxun, LI Liang, WANG Keke, ZOU Guanglei, ZOU Hang. Research and improvement of Apriori algorithm in association rule mining[J]. *Journal of Computer Applications*. 2010, Vol.30, No.11:2952-2954.
15. Zhao C H, Wang F L, Lu N Y. Stage-based soft-transition multiple PCA modeling and online monitoring strategy for batch processes [J]. *Journal of Process Control*, 2007, 17(9): 728—741.
16. XU Xian-Zhen, XIE Lei, WANG Shu-Qing. Multi-operation process monitoring based on PCA hybrid model[J]. *CIESC Journal*, 2011, 62(3): 743-752.
17. Shenzhen Guotaian Education Technology Co., Ltd., Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences - Guotaian Financial Big Data Research Center. *Big Data Analysis: R Foundation and Application [M]*. Tsinghua University Press. 2016.3, 120-180
18. Yu Ying, Wang Bin, Zhang Liming. A Fast PCA Algorithm for Data Learning[J]. *Pattern Recognition and Artificial Intelligence PR&AI*. 2009.8, Vol.20, No.4:568-569.
19. Zhu J L, Ge Z Q, Song Z H. Multimode process data modeling: A Dirichlet process mixture model based Bayesian robust factor analyzer approach [J]. *Chemometrics and Intelligent Laboratory Systems*, 2015, 142: 231-244.

10

20. Liu Qiang, Chai Tianyou, Zhao Lijie. Overview of industrial process monitoring and fault diagnosis based on data and knowledge[D]. Control and Decision, 2010,25(6):801-807.
21. Duan L, Xie M, Bai T, et al. A new support vector data description method for machinery fault diagnosis with unbalanced datasets[J]. Expert Systems with Applications, 2016,64:239-246.