International Core Journal of Engineering Volume 7 Issue 1, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0059

Research on TE Process Fault Diagnosis based on PCA

Jinpeng Wang
School of Shanghai, Maritime University, Shanghai 201306, China.

Abstract

With the great improvement of automation and complexity of chemical industry process, more and more fault factors cause process control. Therefore, monitoring and timely diagnosis of fault of chemical equipment is very important for safe and efficient production of chemical process.Principal component analysis algorithm can make use of the normal working data of the system for modeling, and then make fault judgment according to the input working data. It solves the problem of non-universal detection rules caused by different systems and provides convenient detection methods, so the application scope of the algorithm is greatly expanded. In this paper, the method of principai Component Analysis (PCA) is applied to the Tennesse-Eastman process. The simulation results verify that the method can effectively identify whether the system is in normal working condition or fault state, and it is an effective method for the analysis and diagnosis of system faults.

Keywords

PCA, Fault diagnosis, TE, The chemical industry.

1. Introduction

1.1 Industrial background and significance of fault diagnosis

With the increasing quality requirements of chemical products, the degree of automation and complexity of chemical industrial processes are also greatly improved [1]. As the carrier of chemical industry process, the abnormal state of chemical equipment will affect the quality of products. At the same time, the chemical industry is a high-risk industry. If production equipment or instruments fail and cannot be eliminated in a timely and effective manner, it will not only reduce the qualified rate of product quality, but also cause safety accidents, such as leakage, fire and environmental pollution, which will seriously endanger the health and life safety of personnel [2]. Therefore, it is very important to monitor the chemical equipment and diagnose the fault in time for the safe and efficient production of chemical process.

By monitoring the abnormal status in complex systems, subsystems and production components, fault diagnosis can accurately locate fault points and timely correct the system, so as to ensure the stability, reliability and safety of the industrial system and achieve the purpose of improving production efficiency, product quality and production safety [3]. In the process of chemical industry, fault diagnosis can effectively diagnose the parameters such as real-time monitoring and off-line detection of chemical equipment, media composition and residual wall thickness of equipment, provide fault types and causes in time, guide the maintenance and replacement of equipment, so as to ensure production safety and product quality.

1.2 Methods of fault diagnosis

For modern chemical processes, there are three widely used fault diagnosis methods[4], which are model-based, signal-process-based and data-driven. Among them, model-based methods are difficult to establish an accurate mathematical model due to the uncertainties, nonlinearity, time variability

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0059

and other characteristics in the chemical process [5]. The method based on signal processing is dependent on prior knowledge and cannot be widely used in the modern chemical process where prior knowledge is difficult to obtain. The data-driven method does not rely on process knowledge, but only needs to analyze and extract the process information contained in the process data to represent the running state of the process [6]. If the key process information in the data is extracted effectively, the running state of the process can be accurately described to achieve reliable process monitoring.

For process monitoring, many scholars use Tennessee-Eastman (TE) process to verify the proposed algorithm.Russel E L et al. [7] used Canonical Viriate Analysis to detect the faults of TE process. Kano M et al. adopted the sliding window principal element analysis method to detect the changes in the relationship between variables by monitoring the minimum eigenvalues of subspace transformation matrices. In addition, other monitoring algorithms also include Fisher Discriminant Analysis (FDA) proposed by Chiang L H et al.Kano M et al. compared the performance of several different monitoring algorithms in TE process monitoring. Thanks to modern chemical production process is complicated, in the system variables, such as the number of sensors, need test data is very huge, test has certain difficulty, and different production system is not the same, has its own particularity, shall not apply to the same kind of detection rules for testing, and principal component analysis (principai component analysis, PCA) solved the problem. Principal component analysis algorithm is to use the normal working data of the system for modeling, and then make fault judgment according to the input working data. It solves the problem of non-universal detection rules caused by different systems, greatly expands the application scope of the algorithm, and provides a convenient detection method.

1.3 Tennessee-Eastman Process Overview

Tennessee-Eastman Process is a realistic chemical process model proposed by Downs and Vogel of Eastman Chemical Company in the United States for the development, research and evaluation of process control technologies and monitoring methods. Many scholars and experts at home and abroad refer to it as a data source for control, optimization and fault diagnosis research. This paper presents an application example of PCA based TE process fault diagnosis, which demonstrates the effectiveness of PCA method.By calculating each original variables in the PCA monitoring statistic T^2 and the contribution rate of the SPE, according to the difference in the level of each variable contribution to monitoring statistics, success will be some came out of the system fault information feedback, if there is a job the size of the data is beyond the control limit, shows that the system has failure occurs in the sampling points, to facilitate inspection notice to the staff.

2. The basic principle of PCA algorithm

2.1 PCA algorithm idea

The basic idea of PCA algorithm is to project the high-dimensional process data into an orthogonal low-dimensional subspace and retain the main process information.Geometrically, the coordinate system composed of samples is rotated to the new coordinate space through some linear combination, and the new coordinate axis represents the direction with the maximum variance.

2.2 PCA algorithm implementation steps

Suppose $x \in R^m$ represents a measurement sample containing m sensors, each sensor has n independent samples, and the measurement data matrix $X \in R^{n \times m}$ is constructed, in which each column represents a measurement variable and each row represents a sample.

(1) The data matrix was decomposed by covariance, and the number of principal elements was selected

The covariance matrix of X is $S \approx \frac{X^T \cdot X}{n-1}$, eigenvalue decomposition is carried out, and sorted in descending order according to the magnitude of eigenvalues, as follows:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0059

$$S \approx \frac{X^T \cdot X}{n-1} = V \cdot \Lambda \cdot V^T = [P \cdot \bar{P}] \cdot \Lambda \cdot [P \cdot \bar{P}]^T \quad (1)$$

Where, Λ is a diagonal matrix, which is also the eigenvalue matrix of S, and the elements on its diagonal satisfy $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$.V is the eigenvector matrix of S, with dimension m x m, P is the first A column of V, containing information of all pivot entries, and $\bar{P}$ is the remaining m-A column, containing information of non-pivot entries.

(2) Decompose the original data to obtain the principal element subspace and the residual subspace

After the eigenvalue decomposition of X, X can be decomposed as follows:

$$X = \hat{X} + E = T \cdot P^T + E \quad (2)$$

Where, $\hat{X} = T \cdot P^T$ is called the primary subspace; $E = X - \hat{X}$, is called the residual subspace; $T_{n\times A} = X_{n\times m} \cdot P_{m\times A}$,called the score matrix; $P_{m\times A}$ is called the load matrix and is composed of the first A eigenvectors of S.

(3) Two indicators or criteria for fault detection

The x mentioned below is the newly collected sample with dimension mx1.

The SPE statistic

The SPE index measures the change of the projection of the sample vector in the residual space.

$$SPE = \left\| (I - P \cdot P^T) \cdot x \right\|^2 \leq \delta_\alpha^2 \quad (3)$$

Where, δ_α^2 represents the control limit of α with confidence.

δ_α^2 commonly used calculation formula is as follows:

$$\delta_\alpha^2 = \theta_1 \left(\frac{C_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{1/h_0} \quad (4)$$

Where, $\theta_i = \sum_{j=A+1}^{m} \lambda_j^i, i = 1,2,3, h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_1^2}$, λ_j^i is the eigenvalue of the covariance matrix of X, and C_α is the threshold value of the standard normal distribution under the confidence degree of α.

T^2 statistic

The Hotelling's T^2 Astatistic measures the change of the sample vector in the pivot space.

$$T^2 = x^T P \cdot \Lambda^{-1} \cdot P^T x \leq T_\alpha^2 \quad (5)$$

Where, $\Lambda = diag\{\lambda_1, \lambda_2, ..., \lambda_A\}$, T_α^2 A is the control limit of confidence α.

The common calculation method of control limit is as follows:

$$T_\alpha^2 = \frac{A(n^2 - 1)}{n(n-1)} \cdot F_{A,n-A;\alpha} \quad (6)$$

Where, $F_{A,n-A;\alpha}$ is the distribution value of F with A and n-A degrees of freedom and confidence of α.

2.3 Contribution

The x mentioned below is the newly collected sample with dimension mx1.P is the load matrix.

The most commonly used contribution graph statistics are the SPE and T^2.

The contribution diagram based on SPE is defined as follows:

$$Cont_i^{SPE} = (\xi_i^T \tilde{C} \cdot x)^2, i = 1, ..., m \quad (7)$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0059

Where, $Cont_i^{SPE}$ represents the contribution value of each variable to SPE statistics, $\tilde{C} = I - PP^T$, and ξ_i represent the ith column of identity matrix I.

The contribution diagram based on T^2 is defined as follows:

$$Cont_i^{T^2} = x^T \cdot D \cdot \xi^i \xi_i^T \cdot x \tag{8}$$

Where, $D = P^T \Lambda^{-1} P$, The rest are defined as SPE contribution diagrams.When a fault is detected, the larger variables in the contribution diagram are considered to be the ones likely to cause the fault.But someone with a process background is needed to determine the ultimate cause of the failure.

According to the analysis, T2 statistics detect the information of variables significantly related to principal elements, while SPE statistics detect the residual information of the detected data. Therefore, as long as we separate the variables related to principal elements in SPE statistics, we can better cooperate with T2 statistics to detect system faults.

3. PCA based TE process simulation study

3.1 TE process

The Tennessee-Eastman Process, a Benchmark Process proposed by Downs and Vogel, is a model based on an actual chemical Process. It has been widely used as a Benchmark Process for control and monitoring studies. The process consists of five main units (reactor, condenser, compressor, separator and stripper). TE process is the production of two products from four reactants, as well as an inert product and a by-product, with a total of eight compon.

Table 1 TE failure summary

Variable symbol	Process variables	type
IDV(0)	Normal operation	There is no
IDV(1)	A/C feeding ratio, B composition unchanged	step
IDV(2)	Component B, A/C feed ratio unchanged	step
IDV(3)	The inlet temperature of D	step
IDV(4)	Inlet temperature of reactor cooling water	step
IDV(5)	Inlet temperature of condenser cooling water	step
IDV(6)	A Feed loss	step
IDV(7)	C There is pressure loss - reduced availability	step
IDV(8)	A, B, C Feed ingredients	A random variable
IDV(9)	D of the feeding temperature	A random variable
IDV(10)	C of the feed temperature	A random variable
IDV(11)	Inlet temperature of reactor cooling water	A random variable
IDV(12)	Inlet temperature of condenser cooling water	A random variable
IDV(13)	Reaction dynamic	A random variable
IDV(14)	Reactor cooling water valve	Stick
IDV(15)	Condenser cooling water valve	Stick
IDV(16)	unknown	unknown
IDV(17)	unknown	unknown
IDV(18)	unknown	unknown
IDV(19)	unknown	unknown
IDV(20)	unknown	unknown

3.2 TE process fault classification

The TE process includes 41 measurement variables and 11 control variables.TE process is a large sample of complex non-linear chemical system, it includes 21 kinds of predefined faults, represent the step, random variation, slow drift, viscous and constant position, such as fault type, each state includes training and test, visible Tennessee Eastman process data is very large, the connection

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0059

between the parts is very tight, it is because of the Tennessee Eastman the representation of the system is very strong, so the Tennessee Eastman process can be act as a benchmark evaluation process control and monitoring method of industrial process.The data provided by this system is used to verify whether the pca method can accurately detect the fault. The fault types are shown in Table 1.

3.3 Analysis of simulation results

PCA algorithm is applied to TE process fault diagnosis. Data samples of TE process are used as input of fault feature information, and all data sets are smoothed and normalized.For the control limit, the confidence of both T^2 statistic and SPE statistic is 99%.

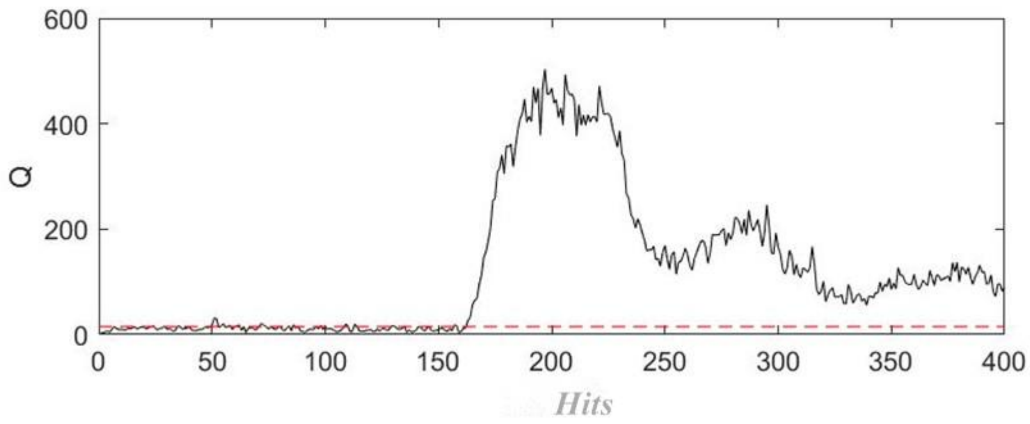

Fig 1 Changes of principal component analysis statistics

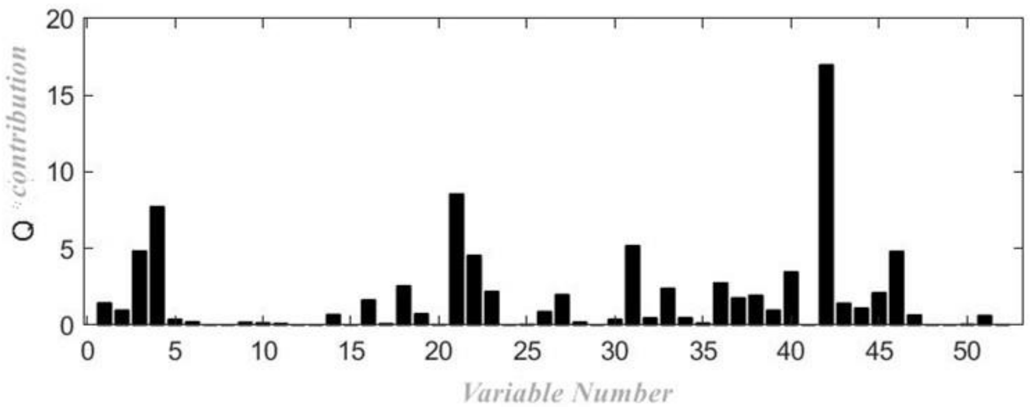

Fig 2 SPE contribution rate diagram

The diagnosis results of fault data using PCA method are shown in Figure 1 and Figure 2.Figure 1 shows that when the monitoring statistic SPE exceeds the control limit, it indicates that a fault has occurred. The straight line represents the control limit. Figure 1 shows that a fault has occurred from the 160th data. After the fault is detected, the SPE contribution diagram (FIG. 2) is further used for fault identification, and the fault source is identified according to the different degree of contribution of each variable to the monitoring statistics. The experimental results show that the principal component analysis method can be used to detect and identify faults effectively.

4. Summary

The data of process control fault detection is huge, and different production systems have their own particularity, so they are not applicable to the same detection law. However, principal component analysis solves this problem. The principal component analysis algorithm USES the system's own normal working data for modeling, and then makes fault judgment based on the input working data, thus solving the problem of non-universal detection rules caused by different systems. This paper

ISSN: 2414-1895 DOI: 10.6919/ICJE.202101_7(1).0059

proposes a TE process fault detection method based on PCA, first by using the method of cumulative variance contribution rate to select principal component, by calculating each original variables in the PCA monitoring statistic T^2 and the contribution rate of the SPE, according to the difference in the level of each variable contribution to monitoring statistics, can identify the fault source, thereby effectively solving the problem of fault identification. The simulation study of TE (Tennessee Eastman) process shows that the proposed method is effective and feasible.

References

[1] Dong, J., Zhang, K., Huang, Y., Li, G., Peng, K.. Adaptive total PLS based quality relevant process monitoring with application to the Tennessee Eastman process. Neuro computing, 2015,154,77–85.

[2] Chetouani Y. Design of a multi-model observer-based estimator for Fault Detection and Isolation (FDI) strategy: application to a chemical reactor[J]. Brazilian Journal of Chemical Engineering, 2008, 25(4): 777-788.

[3] Hajihosseini P, Anzehaee M M, Behnam B. Fault detection and isolation in the challenging Tennessee Eastman process by using image processing techniques[J]. ISA transactions, 2018.

[4] Isermann, R. Fault-diagnosis applications: model-based condition monitoring: actuators drives, machinery, plants, sensors, and fault-tolerant systems. Springer Science & Business Media, 2011.

[5] Tidriri K, Chatti N, Verron S, et al. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges[J]. Annual Reviews in Control, 2016, 42: 63-81.

[6] Ge Z. Review on data-driven modeling and monitoring for plant-wide industrial processes[J]. Chemometrics and Intelligent Laboratory Systems, 2017,171:16-25.

[7] Russel E L, Chiang L H, Braatz R D .Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis[J] .Chemometrics and intel.lab.sysr., 2000, 51:81-93.