

Research Article

Fault Detection and Diagnosis in Process Data Using Support Vector Machines

Fang Wu,¹ Shen Yin,¹ and Hamid Reza Karimi²

¹ Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Heilongjiang 150001, China

² Department of Engineering, Faculty of Engineering and Science, The University of Agder, 4898 Grimstad, Norway

Correspondence should be addressed to Shen Yin; shen.yin2011@gmail.com

Received 15 January 2014; Accepted 27 January 2014; Published 2 March 2014

Academic Editor: Weichao Sun

Copyright © 2014 Fang Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For the complex industrial process, it has become increasingly challenging to effectively diagnose complicated faults. In this paper, a combined measure of the original Support Vector Machine (SVM) and Principal Component Analysis (PCA) is provided to carry out the fault classification, and compare its result with what is based on SVM-RFE (Recursive Feature Elimination) method. RFE is used for feature extraction, and PCA is utilized to project the original data onto a lower dimensional space. PCA T^2 , SPE statistics, and original SVM are proposed to detect the faults. Some common faults of the Tennessee Eastman Process (TEP) are analyzed in terms of the practical system and reflections of the dataset. PCA-SVM and SVM-RFE can effectively detect and diagnose these common faults. In RFE algorithm, all variables are decreasingly ordered according to their contributions. The classification accuracy rate is improved by choosing a reasonable number of features.

1. Introduction

Now, for the complex industrial production systems, fault diagnosis and prediction play an extremely important role. Fault diagnosis is to identify the abnormal circumstances of a system [1]. To find the fault type and to determine the cause of the fault as soon as possible have a vital significance [2]. It can not only reduce the pecuniary loss and avoid the waste of resources but also ensure the safety of a production. In recent years, numerous multivariate statistical process control (MSPC) methods have been developed and have gotten industry's attention [3, 4]. There are many MSPC tools, for example, principal components analysis (PCA) [5], dynamic principal components analysis (DPCA), correspondence analysis (CA) [6], canonical variate analysis (CVA) [7], kernel independent component analysis (KICA) [8], and modified independent component analysis (MICA) [9] which have been used in the actual industry process [10, 11]. Support vector machine (SVM) is a form of machine learning technology, which has a robust mathematical foundation and has many advantages in solving various classification and regression problems. The generalization ability of SVM is also great. So, more and more people start to research on the SVM.

RFE arithmetic is to calculate sorting coefficient based on the weight vector w , which is got from the SVM model training process, and then make the feature variables in descending order according to the sorting coefficient. The study in this paper is based on the Tennessee Eastman (TE) process, which has been extensively applied in process control and statistical process monitoring [12, 13]. Datasets of this process have been used in a variety of machine learning researches.

In this paper, the PCA and RFE algorithm is first introduced. The fault detection methods by using PCA T^2 , SPE statistic, and original SVM are then outlined followed by the SVM-RFE method to find out the most relevant variables of these common faults. The pertinent variables of these common faults are shown after that. Finally, PCA-SVM and SVM-RFE methods are utilized for the classification of several kinds of faults.

2. Related Words

2.1. Support Vector Machine. SVM was firstly proposed by Cortes and Vapnik in 1995 [14] and developed from the statistical learning. It is closely associated with two theories. The first is VC dimension theory, which equals the complexity

of a problem. That is to say, the higher the VC dimension is, the weaker its generalization ability becomes. The second is structural risk minimization theory, which SVM method pursues, meaning to minimize the empirical risk and confidence at the same time. SVM technology has many advantages in solving the problems of which the size of the dataset is relatively small or the samples are nonlinear or the dataset has a high dimension. SVM algorithm firstly maps the low-dimensional samples onto a higher feature space. In this way, the samples which cannot be divided into a low-dimensional space will become linearly separable. A maximum separating hyperplane will be constructed in the higher dimensional space. SVM solves the linear nonseparable case by using slack variables and error penalty. Since the separating plane is constructed based on the support vectors, SVM is a good solution to solve the problem of high dimension. Kernel function is introduced to replace the nonlinear mapping, which avoids numerous unresolved problems.

Assume that the input vectors have two categories. The labels of the first category are $l_i = +1$, and the labels of the second category are $l_i = -1$. m_i represents an input vector. The input samples can be written as follows:

$$D_i = (m_i, l_i). \quad (1)$$

The classification hyperplane to separate given data is as follows:

$$y = w^T m + d = 0, \quad (2)$$

where w represented the weight vector and d represented the constant

$$l_i y = l_i (w^T m_i + d) \geq 1. \quad (3)$$

Taking into account the samples that cannot be classified, the slack variable is ζ and the penalty variable is P . The function to get optimal hyperplane could be written as follows:

$$\min \frac{1}{2} \|w\|^2 + P \sum_{i=1}^M \zeta_i, \quad (4)$$

$$l_i (w^T m_i + d) \geq 1 - \zeta_i,$$

where ζ_i reflects the distance between m_i and the margin, $\zeta_i \geq 0$. m_i is the wrong classified samples. The calculation can be simplified into another problem, which is as follows:

$$\min F = \frac{1}{2} \|w\|^2 - \sum_{i=1}^M \alpha_i l_i (w m_i + d). \quad (5)$$

In order to get the value of w and d which lead (5) to be minimum, the equation at the extreme point can be written as

$$\frac{\partial F}{\partial w} = 0, \quad \frac{\partial F}{\partial d} = 0, \quad (6)$$

$$w = \sum_{i=1}^M \alpha_i l_i m_i, \quad \sum_{i=1}^M \alpha_i l_i = 0. \quad (7)$$

According to (5) and (7), the decision function is developed into a dual quadratic optimization problem, which will be

$$y = \text{sign} \left(\sum_{i,j=1}^M \alpha_i l_i (m_i m_j) + d \right). \quad (8)$$

The kernel function plays an important role in solving the nonlinear problems. The low-dimensional vectors are projected onto a high-dimensional space according to nonlinear function $\Phi(m) = (\phi_1(m), \dots, \phi_i(m))$. The linear decision function is replaced as

$$y = \text{sign} \left(\sum_{i,j=1}^M \alpha_i l_i (\Phi^T(m_i) \cdot \Phi(m_j)) + d \right). \quad (9)$$

The kernel function could be utilized for solving the computational problem. The modality of Φ does not need to be known, when the kernel functions are applied. The decision can be written as

$$y = \text{sign} \left(\sum_{i,j=1}^M \alpha_i l_i K(m_i, m_j) + d \right). \quad (10)$$

There are three ways to utilize SVM method for multiclass classification: 1-v-r SVMs, 1-v-1 SVMs, and H-SVMs.

For 1-v-r SVMs classification method, each type of samples is treated as a class for each time and the other remaining samples belong to the other class. After the SVM models are trained in this way, these k types of samples will construct k classification models. A testing sample will be put into a class in which the sample has the maximal classification function value. For 1-v-1 SVMs classification method, since an SVM classification model is trained for every two kinds of samples, k categories of training samples should train $k(k-1)/2$ SVM classification models. Then, every classification model will be used on a testing sample. The kind of category that the testing sample belongs to depends on the number of votes, which are cast by the classification models. For H-SVMs, all categories are divided into two subsets. After testing by SVM classifier, the subset which the sample belongs to is divided into another two categories. In this way, the right category is finally determined [15, 16].

Because of the rigorous mathematical foundation, support vector machine reflects many unique advantages. It has been widely used in various researches. In the age of big data, machine learning methods and support vector machine technology will have a larger development in the future.

2.2. Principal Components Analysis. PCA was proposed by Pearson in 1901. PCA aims at decorrelating the industrial process data, which is highly correlated and multivariate, and reducing the dimension of the data. In order to reduce the dimension and keep more information of the original data at the same time, a low-dimensional data model needs to be established and the variance of the data reaches the maximum value through projection. After dimensional reduction, the noise and redundant information in the original higher

dimensional space can be eliminated. Because of this, the error can be reduced and the recognition accuracy will be improved. The internal structural characteristics of the data itself can also be found by dimension reduction method. Finally, to realize the purpose of improving the calculation speed, PCA statistics are used as metrics to determine whether the samples are faulty or not. If the metric of the testing data outstrips the cordon, the samples will be treated as faulty [17, 18].

3. Algorithm

3.1. Recursive Feature Elimination Algorithm. RFE arithmetic is to calculate the sorting coefficient at first according to the weight vector w , which is got from the SVM model training process, then take out the feature variable which has minimal sorting coefficient in each iteration, and get the descending order of feature variables finally. The classical SVM-RFE is based on the linear kernel function, while, in nonlinear case, the RBF is used as the kernel [19].

In each iteration, the feature variable with minimal sorting coefficient will be removed. The new sorting coefficient will be obtained by using SVM to train the rest of the feature variables. By executing this process iteratively, a feature ranking list is got. According to this ranking list, the relative degree for each feature with the category will be known. A plurality of nested feature subsets to the SVM model is defined. The quality of these subsets can be assessed by the classification accuracy. So, the optimal feature subset can be obtained. To TE process, such a feature ranking list is got for each type of fault. The first feature in the list is taken as the most relevant one to be analysed [20].

3.2. PCA-SVM Algorithm. PCA aims at decorrelating the industrial process data, which is highly correlated and multivariate, and reducing the dimension of the data [21].

First of all, to get a data matrix X_{nm} under normal conditions, the rows of the matrix represent n specimens and columns represent m variables. This matrix is autoscaled firstly. The singular value decomposition of this matrix is as follows:

$$\frac{1}{\sqrt{n-1}}X = U \sum N^T. \quad (11)$$

The singular value decomposition of this matrix can be replaced by the eigenvalue decomposition of the sample's covariance matrix R . Because the sample matrix has been autoscaled, the eigenvalue decomposition is as follows:

$$R = \frac{1}{n-1}X^T X = \widehat{N}N^T. \quad (12)$$

Eigenvalues are ranked depending on their values in descending order. The feature vectors in N are ranked corresponding to their eigenvalues. Selecting the first a linearly independent eigenvectors $\widehat{N} = [n_1, \dots, n_a]$ to constitute the principal component space \widehat{S} , the rest of the vectors are $\widetilde{N} = [n_{a+1}, \dots, n_m]$ form the residual space \widetilde{S} .

X is decomposed as follows:

$$X = MN^T = \widehat{M}\widehat{N}^T + E. \quad (13)$$

$\widehat{M}\widehat{N}^T = m_1n_1^T + m_2n_2^T + \dots + m_a n_a^T$ represents the score matrix. E is an error matrix which is caused by noise and contains only a little useful information. It can be removed with little losses. So, X can be written as follows:

$$X \approx m_1n_1^T + m_2n_2^T + \dots + m_a n_a^T. \quad (14)$$

After establishing the principal component model, the undetected data x can be written as follows:

$$x = \widehat{x} + \widetilde{x} = \widehat{M}\widehat{N}^T + \widetilde{M}\widetilde{N}^T = x\widehat{N}\widehat{N}^T + x(I - \widehat{N}\widehat{N}^T). \quad (15)$$

The vectors \widehat{x} and \widetilde{x} represent the data that is filtered by the principal component space \widehat{S} and the residual space \widetilde{S} .

When PCA-SVM method is used to classify the types of fault, the dataset should be autoscaled at first. X is the dataset that has been autoscaled

$$X' = \frac{X}{\text{std}(X)} - \text{mean}(\text{std}(X)). \quad (16)$$

The functions $\text{std}(\cdot)$ and $\text{mean}(\cdot)$ are used to calculate the standard deviation and mean value, respectively. Both the training data and the testing data should be autoscaled. X_{tr} is the autoscaled training data and X_{te} is the autoscaled testing data.

The PCA decomposition is done to the normal sample data X_0 . The number of principal components is set as a . The control limit of PCA T^2 statistic is determined by F distribution as follows:

$$T_\alpha^2 = \frac{a(n-1)}{n-a} F_{a, n-a, \alpha}. \quad (17)$$

The control limit of PCA SPE statistic is calculated as follows:

$$\text{SPE}_\alpha = \theta_1 \left[\frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0}, \quad (18)$$

where $\theta_i = \sum_{j=a+1}^m \lambda_j^i$, $i = 1, 2, 3$, and $h_0 = 1 - 2\theta_1\theta_3/3\theta_2^2$.

The autoscaled fault samples X_{tr} and X_{te} are mapping onto the space constructed by the normal samples. Dimension reduction method is used for these fault samples according to (15). The PCA T^2 statistics for the training and testing samples are calculated as follows:

$$T_i^2 = x_i \widehat{N} \Lambda_a^{-1} \widehat{N}^T x_i^T. \quad (19)$$

The PCA SPE statistics for the training and testing samples are calculated as follows:

$$\text{SPE}_i = x_i (I - \widehat{N}\widehat{N}^T) x_i. \quad (20)$$

If the calculated T^2 and SPE values exceed the threshold, the sample can be categorized as faulty.

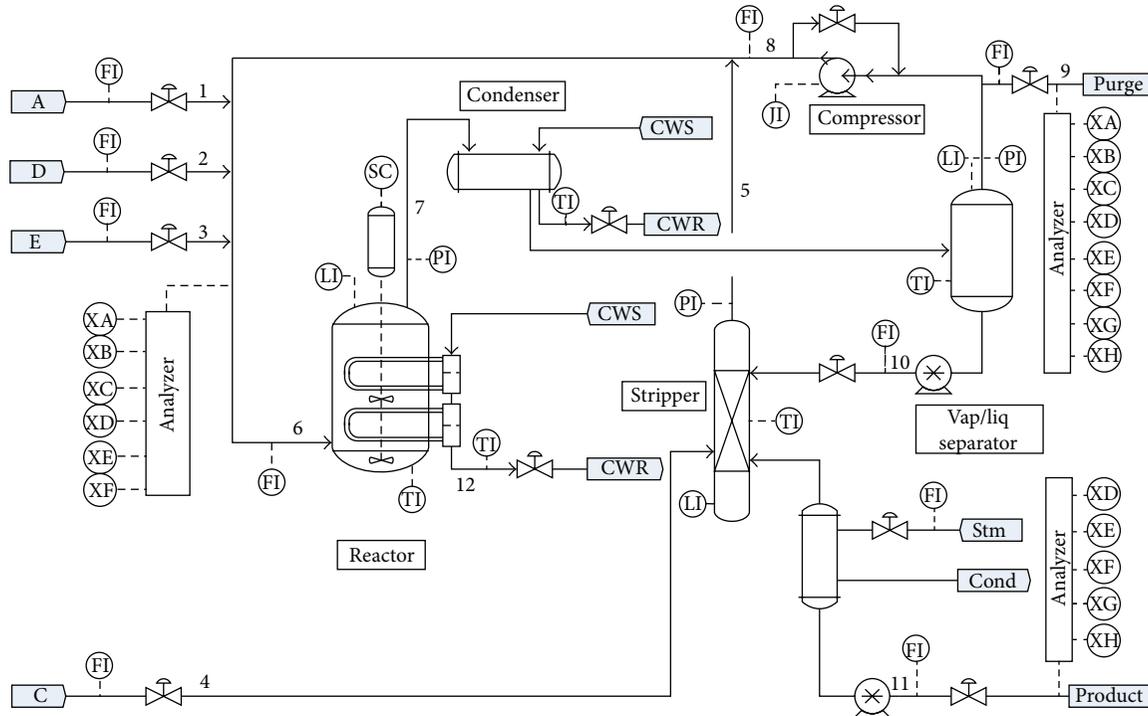


FIGURE 1: The Tennessee Eastman process.

4. Tennessee Eastman Process

The TE process was created by the Eastman Chemical Company to provide a realistic industrial process for evaluating process control and monitoring methods. It was firstly proposed by Downs and Vogel on the AIChE (American Institute of Chemical Engineers) [22]. It has been extensively used in the aspects of control and optimization, process control, fault diagnosis, statistical process monitoring, data driven, and so on. The dataset has become a common data source for these research directions [23].

As shown in Figure 1, the TE plant mainly consists of five units. There are eight components. A, C, D, and E are gaseous materials. B is an inert substance. F is a liquid reaction by-product. G and H are liquid reaction products. The reaction of H has a lower energy than that of G. G has a higher sensitivity to temperature.

Due to the presence of a catalyst, the gaseous reactants will become liquid product when it enters into the reactor. This catalyst is permanent and soluble in liquid. The reactor has an internal condenser. It is used to remove the heat generated by the reaction. Along with the component which is not completely reacted, the product leaves the reactor in the form of steam. The product comes to the gas-liquid separator through a condenser. Part of the steam is condensed and transported to the stripper. The stream mainly contains A. C is treated as the stripping stream. The remaining unreacted components are separated from the bottom of the stripper. Inert matter and the by-product are mainly out in the form of gas from the gas-liquid separator [24].

The whole TE process simulation has 21 kinds of preprogrammed faults, and it consists of 22 continuous measured

variables, 19 composition measurements, and 12 manipulated variables. Both continuous measured variables and composition measurements are belonging to measured variables.

5. Fault Diagnosis

The dataset of the TEP can be got from the internet. All the variables are taken into account to analyse [24]. Select faults 1, 4, 5, 7, and 11 to research. The training dataset contains 500 samples which are obtained from the normal condition. The thresholds of the PCA T^2 and SPE statistics are calculated by this dataset. The testing dataset, using the PCA method to detect fault, contains 960 samples of each fault. The testing dataset to detect fault by SVM contains last 800 samples of each fault. Select 480 samples of each fault, respectively. Treat these samples as the training dataset of the SVM-PCA and SVM-RFE algorithm. There are 960 samples of each fault in the dataset. The first 160 samples which are obtained during the first 8 hours are normal. So, the last 800 samples are chosen from the original dataset of each fault as the testing dataset of the SVM-PCA and SVM-RFE algorithm.

By SVM-RFE algorithm, the feature ranking lists for each fault can be got, from which the most relevant variables will be known, as shown in Table 1. According to the relevant variables, the reasons of each fault could be analysed.

SVM method is also used to determine whether a testing sample is faulty or not. The model has been trained by the training dataset. The category of the testing samples will be directly predicted by the model. The sample, which predicted label is 0, is normal. Otherwise, it is a fault one.

PCA T^2 and SPE statistics are used as the metrics for fault detection. The threshold can be calculated by the normal

TABLE 1: The most relevant variable for each fault.

Fault	Variable
Fault 1	19
Fault 4	51
Fault 5	3
Fault 7	45
Fault 11	51

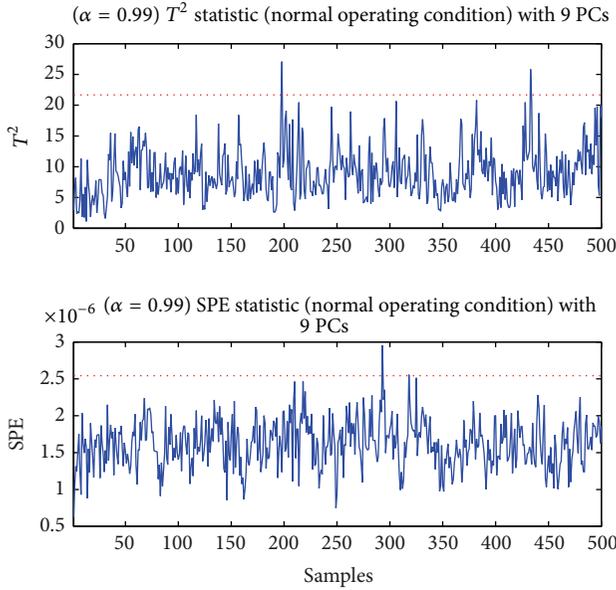


FIGURE 2: The PCA statistics under normal conditions.

samples. If the value of a testing sample exceeds either one of the threshold, this sample will be treated as faulty.

Depending on the normal dataset of the TE process, the threshold of PCA T^2 and SPE statistics will be got. As shown in Figure 2, the data which exceeds the red dotted line can be considered as a fault data. The normal dataset is under the threshold as a whole except some abnormal individual points. The threshold of T^2 statistic is 21.666. The threshold of SPE statistic is $2.5437 * (10^{-6})$.

5.1. *Fault 1.* As shown in Figure 3, the value of variable 19 has a step change, which means that the ratio of A/C feeds takes a step change. But, under normal circumstances, variable 19 does not have such a change, which results in the occurrence of the fault. The training samples also have a step change, so SVM model can learn this change through the training samples, which results in that the fault samples can be detected.

The changes of the flow rates and compositions of stream 6 can affect the variable in the reactor. The rate of flow in stream 4 will arrive at a steady-state value, which is lower than the normal operating conditions. Since many variables are varied apparently, it will be easy to detect these faults.

As shown in Figure 4, SVM method is used to detect the fault 1. The accuracy rate of detecting the fault 1 is 92.6%. Some points fail to be effectively recognised. The main reason why some data cannot be correctly distinguished is that the

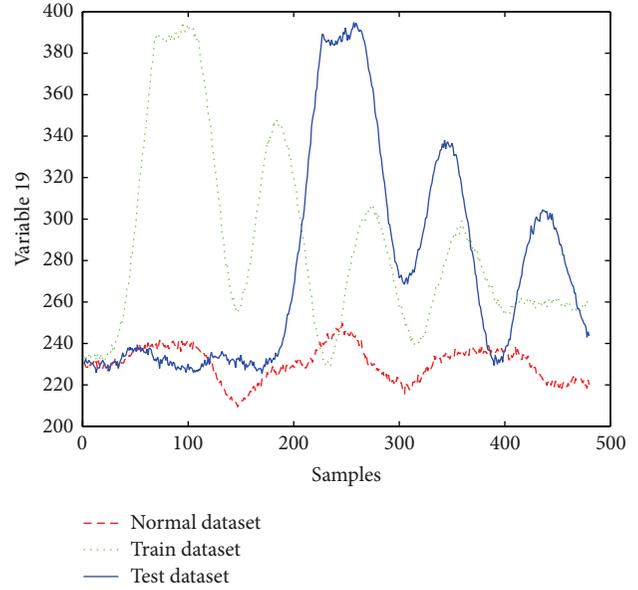


FIGURE 3: TE: plot of variable 19 in fault 1.

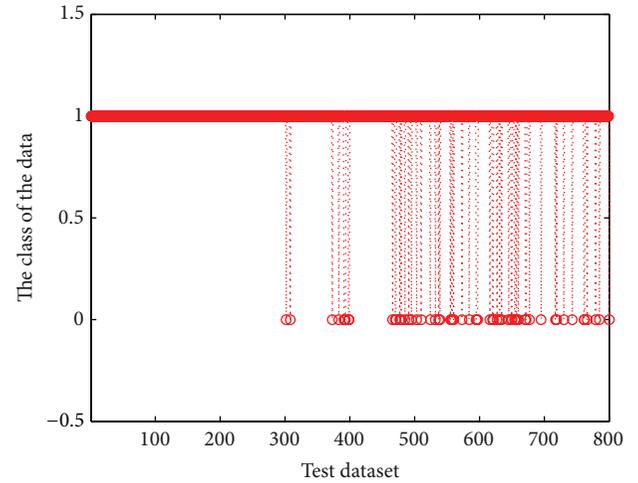


FIGURE 4: The SVM for fault detection for fault 1.

feature extraction is not efficient and reasonable, so that the characteristics of the fault are not able to be identified very well by SVM classification. Compared with the PCA method, the detection rate by SVM is lower.

Figure 5 shows the PCA-based statistics for fault detection. The dotted line in the figure is the T^2 and SPE threshold. The first eight hours before the fault is produced are under normal operating conditions. Therefore, all the data are below the predetermined threshold. The accuracy rate of fault detection is 99.5% by using this method. The PCA statistic is more sensitive than the SVM in the detection of fault 1. To a fault which results in the variable changes obviously, many methods do well in the fault detection.

5.2. *Fault 4.* Significant impact of fault 4 is the temperature step changes of the reactor cooling water inlet. When the fault

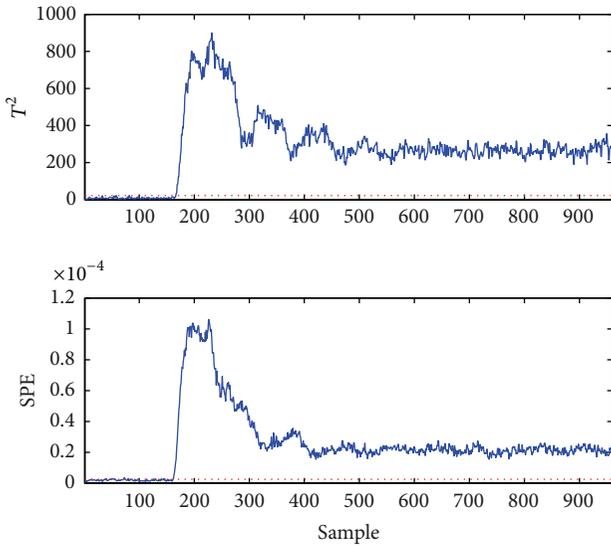


FIGURE 5: The PCA statistics for fault detection for fault 1.

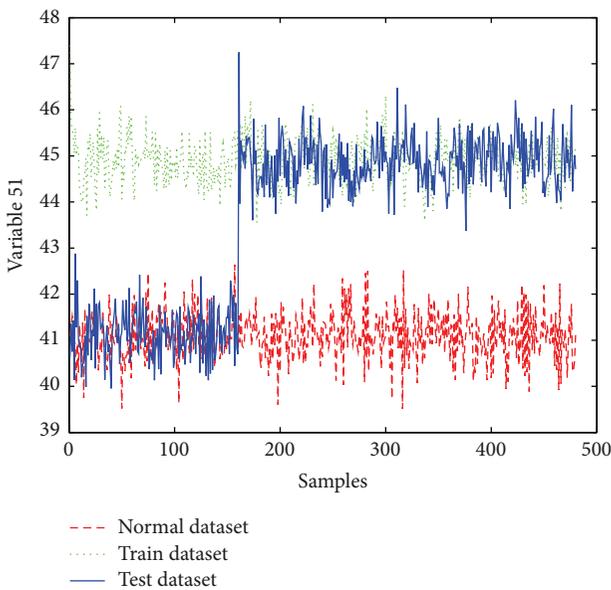


FIGURE 6: TE: plot of variable 51 in fault 4.

happens, that temperature increases quickly, but the standard deviation of the other variables is similar to that under normal conditions. This enhances the difficulty of fault detection.

As shown in Figure 6, it can be seen that the value of variable 51 in the testing dataset has an obvious step change. The changed value of variable 51 was maintained between 44 and 46, while the normal dataset does not have a step change, and its normal value is between 40 and 42. For the training samples of fault 4, the value ranges between 44 and 46 and is different from the normal values. This character helps SVM classification model distinguish fault samples correctly.

As shown in Figure 7, SVM method is used to classify the fault 4 data from the normal samples. To the majority of the fault data, SVM can correctly classify them. The recognition

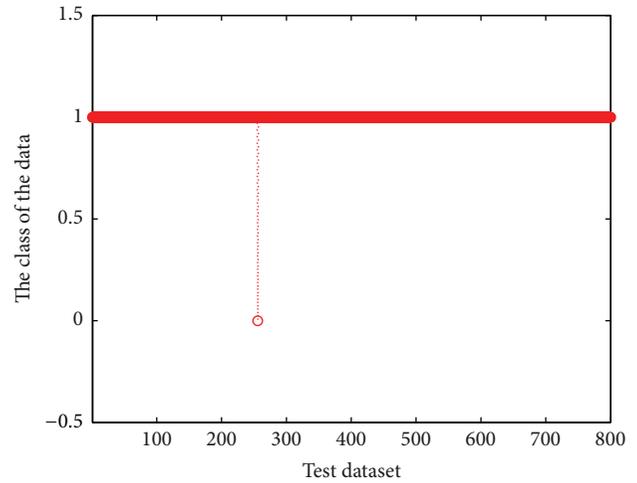


FIGURE 7: The SVM for fault detection for fault 4.

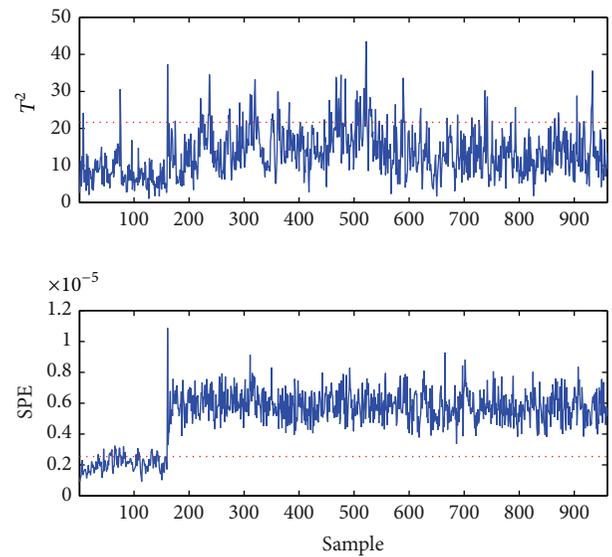


FIGURE 8: The PCA statistics for fault detection for fault 4.

rate of fault 4 is 99.8%. The detection rate is similar to that obtained from PCA method.

As shown in Figure 8, the PCA-based statistics are utilized for fault 4 detection; the dotted line represents the threshold of T^2 and SPE. After 160 sets of normal samples, the value of statistics appears a step variety. This change corresponds to the step change of that temperature. After the fault occurs, the SPE statistics are higher than the threshold as a whole and the T^2 statistics are partly higher than its threshold. The fault can be detected easily from this figure. The detection rate of fault 4 by PCA is almost 100%.

5.3. *Fault 11.* The reason of fault 11 is the same as fault 4. As shown in Figure 9, the fault causes a greater oscillation for the temperature of the cooling water. The vibration amplitude is much larger than the vibration under normal conditions. Under normal circumstances, the range of the variable's value

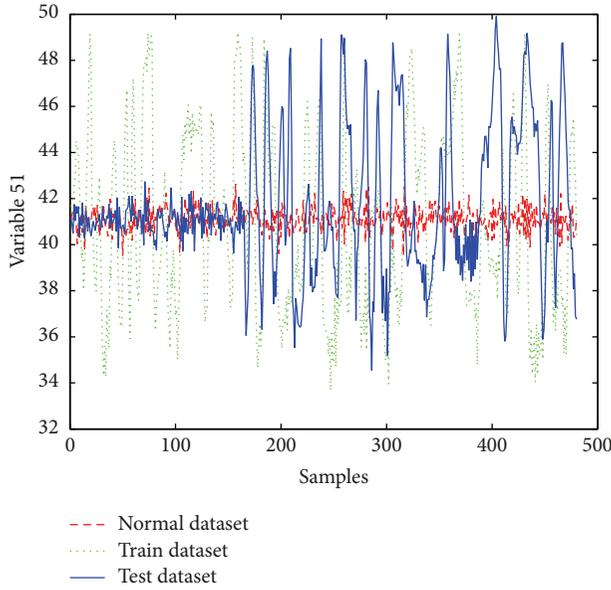


FIGURE 9: TE: plot of variable 51 in fault 11.

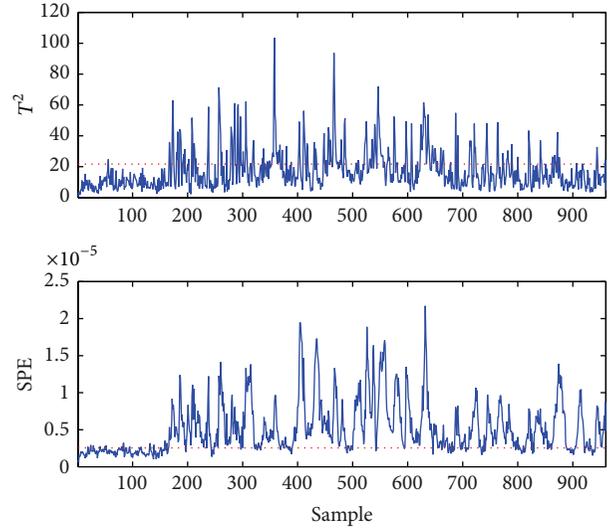


FIGURE 11: The PCA statistics for fault detection for fault 11.

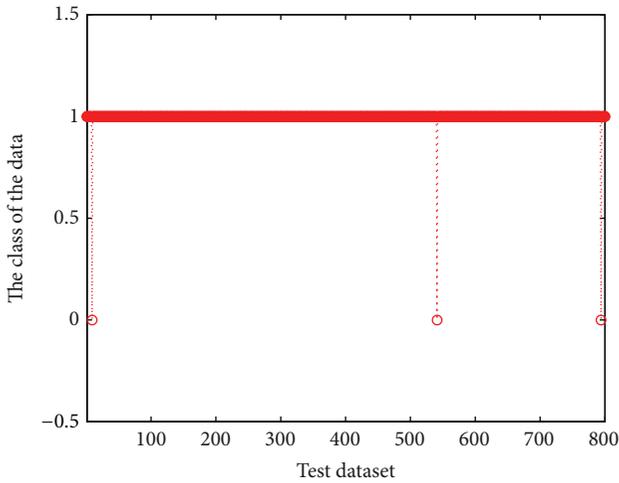


FIGURE 10: The SVM for fault detection for fault 11.

is between 40 and 42, while the variable's value is between 35 and 50 when a fault occurs. The other variables are retained at the level of normal operating conditions.

As shown in Figure 10, the accuracy rate of fault 11 by SVM method is 99.6%. The sample whose label is 0 is treated as normal data, while the testing dataset consists of 800 faulty samples. These points are put into a wrong class. Data preprocessing step can be further increased to improve the accuracy rate of fault detection by SVM.

As is shown in Figure 11, it is sensitive for PCA-based statistics to detect the samples of fault 11. After the first 160 samples, the T^2 and SPE statistics of the samples have a larger fluctuation when compared to that under normal conditions. Most of the data are beyond the red line, and the recognition rate of this fault is 87%. The detection rate is lower than that of SVM method.

According to the simulation results, when compared with the original SVM classification method, the PCA-based method reduces the complexity of the classification model. Different detection methods are suitable for diverse faults; for instance, the SVM method is more appropriate in the detection of fault 11 than the PCA method. The method of SVM on the fault detection could be further strengthened by adding data preprocessing and choosing an appropriate feature extraction way to increase the detection accuracy.

5.4. Fault Detection. In the SVM-RFE algorithm, the features have been ranked in a list by the calculated sorting coefficient, from which the most relevant variables and the relevance ranking to corresponding faults can be got. In order to get the maximum accuracy rate, different numbers of features are selected to train the classification model. The classification accuracy rates by simulation for different numbers of features are as follows.

The first feature in the ranking list is chosen as the most relevant one to the fault to analyze. The feature ranking in the front row may not make SVM classifier to obtain the best classification performance individually. By combining multiple features together, the classifier achieves optimal performance. Therefore, SVM-RFE algorithm can select the best complementary feature combination. As shown in Table 2, if the dimension is reduced to an improper degree, useful information will be reduced, and that will lead to the descent of classification accuracy rate and the faults cannot be diagnosed very well.

In the PCA-SVM algorithm, the autoscaled data should be preprocessed. Then, the PCA method is used to take feature extraction and dimensionality reduction. After the faulty samples are projected onto the principal component space which is constructed by the normal samples, both the dimension and the amount of computation are decreased. The principal features of each fault are extracted.

TABLE 2: TE process: fault detection rate for training and testing data by SVM-RFE.

The number of variables	Training data	Testing data
5	94.3	94.8
10	94.9	94.5
15	94.6	93.6
20	94.9	93.4
25	95.3	93.4
30	97.3	95.5
35	97.5	95.2
40	97.6	94.8
52	97.6	94.7

TABLE 3: TE process: fault detection rate for testing data.

Fault	SVM	SVM-RFE	PCA-SVM
Fault 1	99	99	99
Fault 4	98	99	99
Fault 5	99	100	99
Fault 7	100	100	100
Fault 11	75	79	73
5 Faults	95	96	94

As shown in Table 3, three different methods have been used to distinguish the samples of the 5 faults. The classification accuracy rates for each fault and the entire testing dataset can be got from this table. From the classification results, for the faults 1, 4, 5, and 7, these three methods can achieve high classification accuracy and only some special samples in the testing data are wrongly classified. For the fault 11, which has strong sequence correlation, three kinds of methods are not able to classify well, probably since these three methods cannot well consider the deviation caused by the serial correlation. The overall result says that the RFE feature extraction method can improve the accuracy rate by selecting an appropriate feature vector number, when compared with the original SVM algorithm. Though the effect of PCA-SVM method is similar to the original SVM method in detecting these types of faults, it greatly improves the computing speed.

6. Conclusion

Depending on the method of PCA-SVM and SVM-RFE, these common faults in the TE process are effectively diagnosed. These two methods have been compared in terms of the classification accuracy rate. The results are relatively ideal through the simulations of these algorithms. Since the dimension of the dataset obtained from the TE process is only 52, it is not reliable to reduce the dimension of the TE dataset to rapidly improve the classification accuracy simply by SVM-RFE algorithm. If the selected number of feature is not appropriate, useful information will be reduced while the dimension is decreased, which will lead to the disadvantage of fault diagnosis.

In addition, the threshold of PCA statistics is got by the normal samples obtained from the TE process. It can be used

to test whether a sample is faulty or normal, which achieves very ideal detection effect. The original SVM method is also used to detect the faults. The effect is pretty good. Different detection methods are suitable for diverse faults. According to the method of RFE, the most relevant variables to each fault have been found. These variables have been utilized to analyse the reason why these faults occur. The methods in this paper have a few shortcomings. How to quickly and accurately distinguish all the faults in the TE process by SVM is the aspect which needs to be studied.

Conflict of Interests

There is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors acknowledge the support of China Postdoctoral Science Foundation Grant no. 2012M520738 and Heilongjiang Postdoctoral Fund no. LBH-Z12092.

References

- [1] S. Yin, G. Wang, and H. Karimi, "Data-driven design of robust fault detection system for wind turbines," *Mechatronics*, vol. 11, no. 9, 2013.
- [2] S. Yin, X. Yang, and H. Karimi, "Data-driven adaptive observer for fault diagnosis," *Mathematical Problems in Engineering*, vol. 2012, p. 21, 2012.
- [3] H. Li, J. Yu, C. Hilton, and H. Liu, "Adaptive sliding mode control for nonlinear active suspension vehicle systems using TS fuzzy Approach," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3328–3338, 2013.
- [4] H. Li, X. Jing, and H. R. Karimi, "Output-feedback based H-infinity control for active suspension systems with control delay," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 1, pp. 436–446, 2014.
- [5] J. V. Kresta, J. F. MacGregor, and T. E. Marlin, "Multivariate statistical monitoring of process operating performance," *Canadian Journal of Chemical Engineering*, vol. 69, no. 1, pp. 35–47, 1991.
- [6] K. P. Detroja, R. D. Gudi, and S. C. Patwardhan, "Plant-wide detection and diagnosis using correspondence analysis," *Control Engineering Practice*, vol. 15, no. 12, pp. 1468–1483, 2007.
- [7] E. L. Russell, L. H. Chiang, and R. D. Braatz, "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 81–93, 2000.
- [8] J.-M. Lee, S. Joe Qin, and I.-B. Lee, "Fault detection of nonlinear processes using kernel independent component analysis," *Canadian Journal of Chemical Engineering*, vol. 85, no. 4, pp. 526–536, 2007.
- [9] J.-M. Lee, S. J. Qin, and I.-B. Lee, "Fault detection and diagnosis based on modified independent component analysis," *AIChE Journal*, vol. 52, no. 10, pp. 3501–3514, 2006.
- [10] Q. Zhou, P. Shi, J. Lu, and S. Xu, "Adaptive output-feedback fuzzy tracking control for a class of nonlinear systems," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 972–982, 2011.

- [11] Q. Zhou, P. Shi, S. Xu, and H. Li, "Observer-based adaptive neural network control for nonlinear stochastic systems with time-delay," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 1, pp. 71–80, 2013.
- [12] S. Yin, S. Ding, A. Haghani, and H. Hao, "Data-driven monitoring for stochastic systems and its application on batch process," *International Journal of Systems Science*, vol. 44, no. 7, pp. 1366–1376, 2013.
- [13] S. Yin, H. Luo, and S. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 2402–2411, 2014.
- [14] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1999.
- [15] S.-F. Yuan and F.-L. Chu, "Support vector machines-based fault diagnosis for turbo-pump rotor," *Mechanical Systems and Signal Processing*, vol. 20, no. 4, pp. 939–952, 2006.
- [16] S. Abe, *Support Vector Machines For Pattern Classification*, Springer, New York, NY, USA, 2005.
- [17] S. Yin, S. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.
- [18] P. Nomikos and J. F. MacGregor, "Monitoring batch processes using multiway principal component analysis," *AIChE Journal*, vol. 40, no. 8, pp. 1361–1373, 1994.
- [19] S. S. Keerthi and C.-J. Lin, "Asymptotic behaviors of support vector machines with gaussian kernel," *Neural Computation*, vol. 15, no. 7, pp. 1667–1689, 2003.
- [20] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.
- [21] S. J. Qin, "Statistical process monitoring: basics and beyond," *Journal of Chemometrics*, vol. 17, no. 8–9, pp. 480–502, 2003.
- [22] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers and Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993.
- [23] P. R. Lyman and C. Georgakis, "Plant-wide control of the Tennessee Eastman problem," *Computers and Chemical Engineering*, vol. 19, no. 3, pp. 321–331, 1995.
- [24] L. Chiang, E. Russell, and R. Braatz, *Fault Detection and Diagnosis in Industrial Systems*, Springer, London, UK, 2001.