

Research Article

Nonlinear Fault Separation for Redundancy Process Variables Based on FNN in MKFDA Subspace

Ying-ying Su,^{1,2} Shan Liang,¹ Jing-zhe Li,³ Xiao-gang Deng,⁴ Tai-fu Li,²
and Cheng Zeng¹

¹ College of Automation, Chongqing University, Chongqing 400044, China

² School of Electrical and Information Engineering, Chongqing University of Science and Technology, Chongqing 401331, China

³ School of Safety Engineering, Chongqing University of Science and Technology, Chongqing 401331, China

⁴ College of Mechanical and Power Engineering, Chongqing University of Science and Technology, Chongqing 401331, China

Correspondence should be addressed to Shan Liang; lightsun@cqu.edu.cn

Received 19 July 2013; Revised 22 October 2013; Accepted 23 October 2013; Published 4 February 2014

Academic Editor: Xianxia Zhang

Copyright © 2014 Ying-ying Su 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.

Nonlinear faults are difficultly separated for amounts of redundancy process variables in process industry. This paper introduces an improved kernel fisher distinguish analysis method (KFDA). All the original process variables with faults are firstly optimally classified in multi-KFDA (MKFDA) subspace to obtain fisher criterion values. Multikernel is used to consider different distributions for variables. Then each variable is eliminated once from original sets, and new projection is computed with the same MKFDA direction. From this, differences between new Fisher criterion values and the original ones are tested. If it changed obviously, the effect of eliminated variable should be much important on faults called false nearest neighbors (FNN). The same test is applied to the remaining variables in turn. Two nonlinear faults crossed in Tennessee Eastman process are separated with lower observation variables for further study. Results show that the method in the paper can eliminate redundant and irrelevant nonlinear process variables as well as enhancing the accuracy of classification.

1. Introduction

With developments of modern process industry, multivariate monitor from sensors has showed their multicollinearity, nonlinear correlative coupling, time delay, and redundancy. It makes complexity increasing with exponent to fault separation and diagnosis, called “Curse of Dimension” [1, 2]. On the other hand, right ratio of fault classification decreases with multivariate and redundancy process variables. Therefore, many attentions have been paid on two points of view that are variable selection and dimension reduction [3, 4].

Among the study of variable selection, the existed methods can be broadly classified into three categories: random search techniques, measure-based method, and intelligent computation. In random search, each process variable is directly deleted or involved in the classification model one time in turn to search the most suitable input sets under a certain criterion, such as forward selection, backward selection, and stepwise that are simple and easily realized methods [5].

While it was studied by Masion and Gunst [6] that these methods would result in mistaken results, variable set appears multicollinearity. Measure-based method appears to select variable with computing relevancy among all variables, as well as that between variables and labels. The variables with highest similar characteristic will be gathered in one kind. According to different definition, K-L information measure, minimum description length, and mutual information are used [7–9]. Intelligent computation deepens to solve nonlinear variable selection problem, such as neural network that is once used to nonlinear model, while its selection criterion is uncertain [10].

Dimension reduction is different from variable selection, which mainly depends on transformation and information extraction of original variable matrix. It projects original variables with a certain mapping to a new subspace and extracts information in lower dimension, such as principal component analysis (PCA) [11] and partial least squares (PLS) [12].

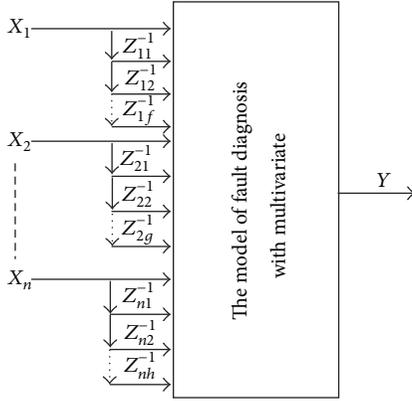


FIGURE 1: The fault diagnosis with multivariate.

Original variables with linear-relative process variables are linearly projected according to the maximum direction of covariance matrix. Meanwhile, the maximum original information can be kept as most as possible. Contribution chart method is the way to calculate contribution of each variable to certain fault with T^2 statics and SPE [13, 14] for PCA. The above linear methods have been extended to nonlinear ones after kernel method presented [15–20], such as kernel principal component analysis (KPCA), kernel partial least squares (KPLS), and kernel fisher discriminant analysis (KFDA). Kernel method converts a linear classification learning algorithm into nonlinear one, by mapping the original observations into a higher-dimensional space. So that linear classifier in the new space equals to a nonlinear classifier in the original space.

However, nonlinear information projected to the new feature space has higher dimension, and data matrix has lost their original physical meaning in original sample space. If we separated nonlinear faults crossed together in original space, the dimension of classifier with kernel method would become huge, while right ratio would decrease with redundancy and multicollinearity variables.

The objective of this paper is to deepen dimension reduction method for the above problems with measure method in variable selection called MKFDA-FNN. Nonlinear process variables are projected in higher-dimension space with MKFDA. Discriminant vector and its corresponding feature vector with maximum separation are computed to cluster original variables with highest similarity. With embed-dimension increasing, false nearest neighbors (FNN) with high similarity are able to be removed in turn. Thus, nonlinear redundancy and multicollinearity process variables can be removed from input sets to nonlinear classifier. Finally, we give an actual fault separation problem in classical chemical process Tennessee Eastman (TE) to further study.

2. Problem Description

In fault separation problem presented above, it equals to screen original process variables related to certain faults as most as possible. Multivariate data matrix considered initially with normal and fault information is described in Figure 1, where X_1, X_2, \dots, X_n are process variables with n -dimension,

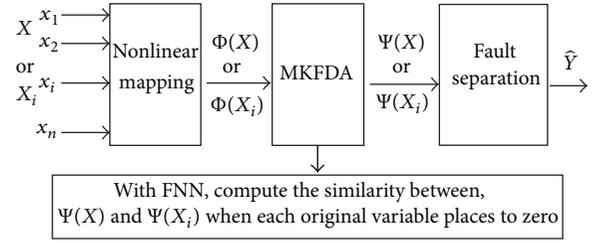


FIGURE 2: Nonlinear fault diagnosis with redundancy process variables based on FNN in MKFDA subspace.

$Z_{11}^{-1}, Z_{12}^{-1}, \dots, Z_{1f}^{-1}$ present time-delay variables of X_1 at different sample time, $Z_{21}^{-1}, Z_{22}^{-1}, \dots, Z_{2g}^{-1}$ present time-delay variables of X_2 , and $Z_{n1}^{-1}, Z_{n2}^{-1}, \dots, Z_{nh}^{-1}$ present time-delay variables of X_n at different sample-time. In this way, original data matrix is composed of n -dimension process/control variables and their delay variables in

$$X_{(f+g+h) \times l}(k) = \begin{bmatrix} x_1(k-1) & x_1(k-2) & \cdots & x_1(k-l) \\ x_1(k-2) & x_1(k-3) & \cdots & x_1(k-l-1) \\ \vdots & \vdots & \vdots & \vdots \\ x_1(k-f) & x_1(k-f-1) & \cdots & x_1(k-f-l+1) \\ x_2(k-1) & x_2(k-2) & \cdots & x_2(k-l) \\ x_2(k-2) & x_2(k-3) & \cdots & x_2(k-l-1) \\ \vdots & \vdots & \vdots & \vdots \\ x_2(k-g) & x_2(k-g-1) & \cdots & x_2(k-g-l+1) \\ \vdots & \vdots & \vdots & \vdots \\ x_n(k-1) & x_n(k-2) & \cdots & x_n(k-l) \\ x_n(k-2) & x_n(k-3) & \cdots & x_n(k-l-1) \\ \vdots & \vdots & \vdots & \vdots \\ x_n(k-h) & x_n(k-h-1) & \cdots & x_n(k-h-l+1) \end{bmatrix}, \quad (1)$$

where f, g , and h present maximum delay order of process/control variables x_1, x_2 , and x_n , k presents current sample time, and l is sample length.

3. Multivariate Fault Separation Based on MKFDA-FNN

To fault separation problem with nonlinear redundancy process/control variables, an approach is proposed in Figure 2. Correlated nonlinear variables are firstly projected to a higher-dimension MKFDA subspace. Furthermore, in order to find fairly useful variables, the importance of each input is measured in subspace with distance measure inspired by FNN. Accordingly, redundant variables are recognized. It makes separation of faults crossed together easily.

3.1. False Nearest Neighbors. FNN is the feature selection method on the basis of phase space reconstruction (PSR) in high-dimension data space [21]. With embed-dimension

increasing, movement locus becomes open, and false nearest neighbors with high similarity are able to be removed in turn. It restores the locus of chaos. Its algorithm is as follows.

In d -dimension phase space including original variables and their time delay, each phase vector $x(i) = \{x(i), x(i + \tau), \dots, x(i + (d - 1)\tau)\}$ has one nearest neighbors $x^{NN}(i)$. Their 2-norm distance is

$$R_{d+1}^2(i) = R_d^2(i) + \|x(i + \tau d) - x^{NN}(i + \tau d)\|^2. \quad (2)$$

When d -dimension is increased to $d + 1$, the above phase vector is changed as new one, noted as $R_{d+1}(i)$ in

$$R_d(i) = \|x(i) - x^{NN}(i)\|. \quad (3)$$

If $R_{d+1}(i)$ was much bigger than $R_d(i)$, it means the projection of two nonneighbor phase vector from higher dimension to lower one. So the two neighbors are the false nearest neighbors.

Note that

$$a_1(i, d) = \frac{\|x(i + \tau d) - x^{NN}(i + \tau d)\|}{R_d(i)}. \quad (4)$$

If $a_1(i, d)$ is larger than R_τ , $x^{NN}(i)$ should be fault nearest neighbor of $x(i)$. Threshold R_τ is determined between interval (10, 50). Once there appeared noise in process data, the following judge criterion should be involved. If $R_{d+1}(i)/R_A \geq 2$, $x^{NN}(i)$ should be nearest fault neighbor of $x(i)$, where R_A is

$$R_A = \frac{1}{N} \sum_{i=1}^N [x(i) - \bar{x}], \quad \bar{x} = \frac{1}{N} \sum_{i=1}^N x(i). \quad (5)$$

The distance measure between vectors can explain the similarity of false nearest neighbors factually in (6). Assume that there was a data space Q with d -dimension variable, and one sample vector is $A = (q_1, q_2, \dots, q_i, \dots, q_d)$. We set variable q_i as zero, standing for vector A without variable q_i that is noted as $B = (q_1, q_2, \dots, 0, \dots, q_d)$ in Figure 3.

The similarity between A and B is

$$\delta = \|A - B\|^2. \quad (6)$$

If distance measure is small, it shows that vectors A and B have highly similarity. That is, the removed variable q_i makes little impact on nonlinear pattern, and process variable q_i has low interpreting ability. Otherwise, if it was much bigger, it reveals that B much differs from A . Process variable q_i is important to interpreting of nonlinear pattern. B is false nearest neighbors of A .

3.2. Kernel Fisher Discriminant Analysis. KFDA is most useful to nonlinear classification problems [22]. Nonlinear discriminant vector in original space is extracted to linear optimal discriminate vector in high-dimension feature space H with conventional fisher discriminant analysis (FDA). Since dimension of H is much higher, it is hard to directly confirm nonlinear mapping function from original space to the feature space. Reproducing kernel-based method widely

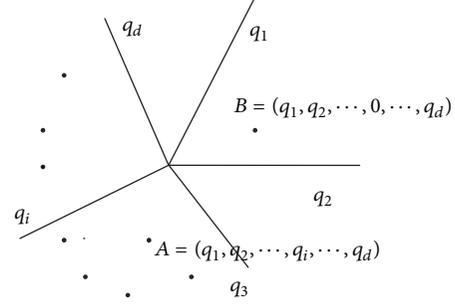


FIGURE 3: Data space Q with d -dimension variable.

developed in machine learning (ML) can achieve this goal. Nonlinear mapping is indirectly found according to $k(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y})$ in Gram-space [23], where $\Phi: \mathbf{R}^d \rightarrow H$.

Conventional kernel function can be selected as follows [6].

- (i) Polynomial kernel function $K(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + c)^d$, $d = 1, 2, \dots, N$, c is constant.
- (ii) Gaussian kernel function $K(\mathbf{x}, \mathbf{x}') = \exp(-(\|\mathbf{x} - \mathbf{x}'\|/2\sigma^2))$, δ is the parameter of breadth.
- (iii) Sigmoid kernel function $K(\mathbf{x}, \mathbf{x}') = \tanh(f\langle \mathbf{x}, \mathbf{x}' \rangle + \theta)$.

Assume that original sample set was $X = \{x_1, x_2, \dots, x_N\}$ with d -dimension and N -samples, where X_i is the sample of i th type, $N_i = |X_i|$, and $i = 1, 2$. There exists nonlinear mapping function $\Phi: \mathbf{R}^d \rightarrow H$. It transforms nonlinear original sample space \mathbf{R}^d to linear classification in high-dimension data space H ; that is, $\Phi(\mathbf{x}) \in H$, $\mathbf{x} \in \mathbf{R}^d$. In space H , distance scatter of intraclass and classes with training data is \mathbf{S}_ω and \mathbf{S}_b in (7) and (9), respectively,

$$\mathbf{S}_\omega = \mathbf{S}_1 + \mathbf{S}_2, \quad (7)$$

$$\mathbf{S}_i = \sum_{\mathbf{x} \in X_i} (\Phi(\mathbf{x}) - \mathbf{m}_i)(\Phi(\mathbf{x}) - \mathbf{m}_i)^T, \quad i = 1, 2, \quad (8)$$

$$\mathbf{S}_b = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T, \quad (9)$$

where \mathbf{m}_i is the mean of i th type in feature space. KFDA is to find a projection direction \mathbf{w} , which meets the following two properties: (1) data that has similar characteristic should be gathered together as most as possible; (2) the ones with different characteristic should be gathered as far as possible. So a key is to search projection direction \mathbf{w}^* and its corresponding discriminant function $g(\mathbf{x}) = (\mathbf{w}^*)^T \mathbf{x} - y_0$. Similarly with linear FDA, the optimal projection direction \mathbf{w}^* is to search vector \mathbf{w} , which maximizes fisher criterion function (10), where \mathbf{w}^* is optimal projection direction:

$$J_H(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_\omega \mathbf{w}}. \quad (10)$$

Since dimension of feature space H is usually high and Φ is indirect mapping function, discriminant vector is hard to compute directly. Thus, each solution \mathbf{w} is expressed as linear

TABLE I: Steps designed in this paper.

Inputs	$T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}, \mathbf{x}_i \in \mathbf{R}^d, y_i \in \{+1, -1\}, i = 1, 2, \dots, N$
Step 1	Initiate $\lambda \in [0, 1]$ and compute $\mathbf{u}_i, i = 1, 2$
Step 2	Select suitable multikernel function
Step 3	Compute the kernel mean vector between two kinds with $\mathbf{k}_{w_i} = \sum_{j=1}^{N_i} (\mathbf{K}_{x_j^i} - \mathbf{u}_i) (\mathbf{K}_{x_j^i} - \mathbf{u}_i)^T$
Step 4	Compute the kernel scatter matrix of intraclass $\mathbf{k}_w = \lambda \mathbf{k}_{w_1} + (1 - \lambda) \mathbf{k}_{w_2}$
Step 5	Compute $\boldsymbol{\alpha}^* = \mathbf{k}_w^{-1} (\mathbf{u}_1 - \mathbf{u}_2), \mathbf{y}_0 = \boldsymbol{\alpha}^T (\mathbf{u}_1 - \mathbf{u}_2)$
Step 6	Get the optimal solution of (16)
Step 7	Place the inspected process variable as zero in original samples
Step 8	Project the new samples into the feature space
Step 9	Compute the contribution of one variable at one time with FNN in MKFDA
Step 10	Repeat the above course for the remaining variables
Outputs	The distance measure $\boldsymbol{\delta} = [\delta_1, \delta_2, \dots, \delta_n]$ of each original variable is obtained

combination of samples in (11), according to kernel-based method,

$$\mathbf{w} = \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_i) = \boldsymbol{\psi} \boldsymbol{\alpha}, \quad (11)$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$.

Moreover, nonlinear transformation function $\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_N)$ of samples can be projected to feature space H with direction \mathbf{w} in

$$\begin{aligned} \mathbf{w}^T \Phi(\mathbf{x}_i) &= \boldsymbol{\alpha}^T \boldsymbol{\psi}^T \Phi(\mathbf{x}_i) \\ &= \boldsymbol{\alpha}^T (\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_N))^T \Phi(\mathbf{x}_i), \\ &= \boldsymbol{\alpha}^T (k(\mathbf{x}_1, \mathbf{x}_i), k(\mathbf{x}_2, \mathbf{x}_i), \dots, k(\mathbf{x}_N, \mathbf{x}_i)), \\ &= \boldsymbol{\alpha}^T \mathbf{k}_{x_i}, \end{aligned} \quad (12)$$

From (11), for all $\mathbf{x} \in \mathbf{R}^d$, assume that $\mathbf{k}_x = (k(\mathbf{x}_1, \mathbf{x}), k(\mathbf{x}_2, \mathbf{x}), \dots, k(\mathbf{x}_N, \mathbf{x}))^T$ and projection of mean vector \mathbf{m}_i with direction \mathbf{w}^* in feature space H is

$$\mathbf{w}^* \mathbf{m}_i = \boldsymbol{\alpha}^T \boldsymbol{\psi}^T \left(\frac{1}{N_i} \sum_{j=1}^{N_i} \Phi(\mathbf{x}_j^i) \right) = \boldsymbol{\alpha}^T \mathbf{u}_i, \quad (13)$$

where $\mathbf{u}_i = ((1/N_i) \sum_{j=1}^{N_i} \Phi(\mathbf{x}_1) \Phi(\mathbf{x}_j^i), (1/N_i) \sum_{j=1}^{N_i} \Phi(\mathbf{x}_2) \Phi(\mathbf{x}_j^i), \dots)^T$.

From (12) and (13), we have

$$\begin{aligned} \mathbf{w}^T \mathbf{S}_b \mathbf{w} &= \boldsymbol{\alpha}^T \mathbf{k}_b \boldsymbol{\alpha}, \\ \mathbf{w}^T \mathbf{S}_w \mathbf{w} &= \boldsymbol{\alpha}^T \mathbf{k}_w \boldsymbol{\alpha}, \end{aligned} \quad (14)$$

where $\mathbf{k}_b = (\mathbf{u}_1 - \mathbf{u}_2)(\mathbf{u}_1 - \mathbf{u}_2)^T, \mathbf{k}_w = \mathbf{k}_{w_1} + \mathbf{k}_{w_2}, \mathbf{k}_{w_i} = \sum_{j=1}^{N_i} (\mathbf{k}_{x_j^i} - \mathbf{u}_i)(\mathbf{k}_{x_j^i} - \mathbf{u}_i)^T$.

Since fisher criterion function is optimal solution of (15), vector \mathbf{w} can be resolved as $\boldsymbol{\alpha}$ in the following fisher criterion (16) [24]:

$$J_H(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}, \quad (15)$$

$$J_H(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^T \mathbf{k}_b \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \mathbf{k}_w \boldsymbol{\alpha}}. \quad (16)$$

Furthermore, the solution of optimal vector $\boldsymbol{\alpha}^*$ and \mathbf{y}_0 can be solved [25] with

$$\begin{aligned} \boldsymbol{\alpha}^* &= \mathbf{k}_w^{-1} (\mathbf{u}_1 - \mathbf{u}_2), \\ \mathbf{y}_0 &= \frac{(\mathbf{w}^*)^T \mathbf{m}_1 + (\mathbf{w}^*)^T \mathbf{m}_2}{2} = \frac{(\boldsymbol{\alpha}^*)^T (\mathbf{u}_1 + \mathbf{u}_2)}{2}. \end{aligned} \quad (17)$$

Thus, the corresponding function of kernel fisher discriminant function is obtained as

$$g(\mathbf{x}) = (\boldsymbol{\alpha}^*)^T \mathbf{k}_x - \mathbf{y}_0. \quad (18)$$

3.3. Multikernel Fisher Discriminant Analysis. From Section 3.2, the solution of maximizing (15) equals to the solution of maximizing (16). Assume that $\boldsymbol{\alpha}^* = \mathbf{k}_w^{-1} (\mathbf{u}_1 - \mathbf{u}_2)$ is optimal solution to classification effect, whereas $\boldsymbol{\alpha}^*$ is both determined by kernel scatter matrix \mathbf{k}_w and difference of kernel mean vector $(\mathbf{u}_1 - \mathbf{u}_2)$. In the condition of independent and identically distributed, kernel mean of samples is independent with number of samples. It indicates difference of kernel mean vector $(\mathbf{u}_1 - \mathbf{u}_2)$ doing nothing with the unbalance of samples. So $\boldsymbol{\alpha}^*$ is only determined by kernel scatter matrix \mathbf{k}_w for intraclass. If distribution of different variables differed, it should result in the contributions $\mathbf{k}_{w_1}, \mathbf{k}_{w_2}, \dots, \mathbf{k}_w$ not in the similar interval. Besides that the solution of $\boldsymbol{\alpha}^*$ is not the optimal one. Hence, in order to avoid the influence of different distribution for samples, we presented multi-kernel fisher discriminant analysis method. It advances the kernel criterion function $\mathbf{k}_w = (\mathbf{k}_{w_1} + \mathbf{k}_{w_2})$ into

$$\mathbf{k}_w = \lambda \mathbf{k}_{w_1} + (1 - \lambda) \mathbf{k}_{w_2}, \quad (19)$$

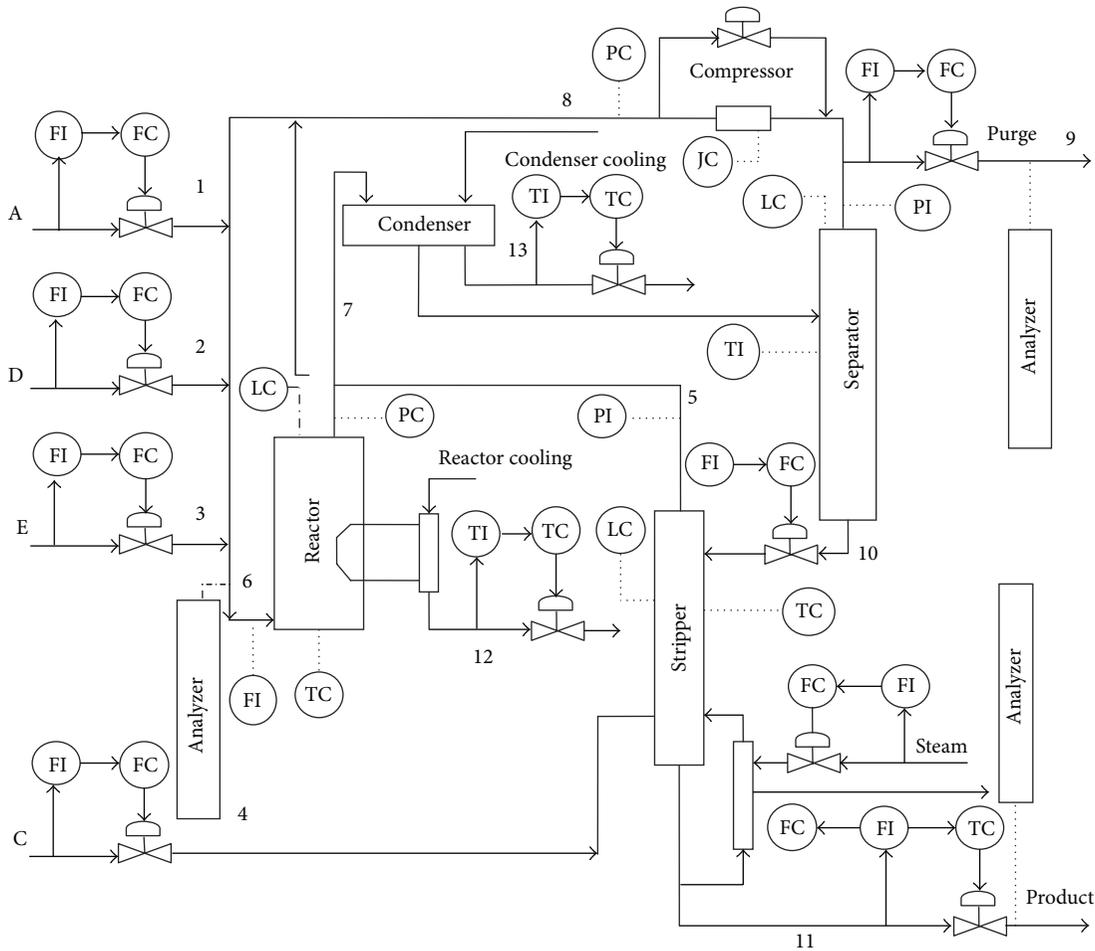


FIGURE 4: The technological process of Tennessee Eastman.

where $\lambda (\lambda \in [0, 1])$ is the adjustable MATLAB parameter and \mathbf{k}_{w_1} and \mathbf{k}_{w_2} are the kernel matrix computed with each suitable kernel function from Section 3.2 (i)/(ii)/(iii).

In this way, the influence with different sample distributions is considered with the suitable kernel function.

The above algorithm in this paper can be chiefly described in Table 1. In this way, the contribution of each original process variable q_i to the certain fault is measured.

4. Fault Separation of Tennessee Eastman with Redundancy Variables

4.1. Tennessee Eastman Chemical Process. Tennessee Eastman (TE) is a classical chemical process created by Eastman Chemical Company in 1993 [26]. Its technological process is shown in Figure 4. There are four reactants (A, C, D, and E) and two products (G, H). Besides that, there is one inert material B and byproduct F.

In TE process, the dynamic TE model is composed of five major units: a reactor, a separator, a stripper, a condenser, and a compressor. Each unit can be expressed with some equations, in all of 148 algebraic equations and 30 differential equations. So it becomes one of the most complex models and is widely used to test study algorithm with control, system

monitor, fault diagnosis, and so forth. Here, we take Tennessee Eastman as the study object to measure its fault separation ability with our method.

4.2. Nonlinear Fault Separation of Redundancy Variables. In TE process, there are 41 observed variables and 12 manipulated variables from controller, some of which are nonlinear redundancy variables. Moreover, there are 20 types of classical fault in TE process shown in Table 2. Since Fault9 and Fault11 are nonlinear overlapped together shown in Figure 5, we take their fault separation as the study goal, meanwhile, 53 process variables must be screened for their multicollinearity and nonlinear redundancy. Process data of TE is simulated at one-minute sampling time in MATLAB software from Downs [27]. All the measurements have Gaussian noise. A total of 1000 samples are collected for training, where 800 data are collected for Fault9 and 200 for Fault11. In addition, 835 samples are applied to test separation validity with 644 for Fault9 and 171 for Fault11.

4.3. Results and Discussion. If we distinguished Fault9 and Fault11, there are 53 variables to be considered in all. Therefore, we compute the contribution of 53 variables with mentioned method to see the importance of each process variables

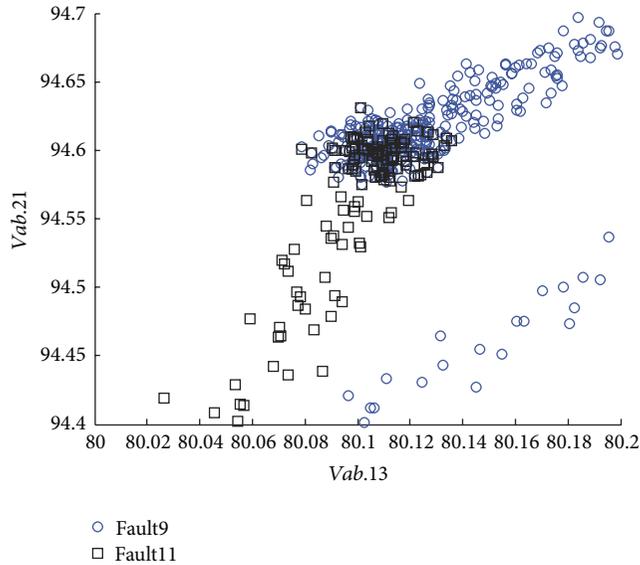


FIGURE 5: The distribution of Fault9 and Fault11, described in 2-dimension diagram with x -axis of process variable $Vab.13$ and y -axis of process variable $Vab.21$.

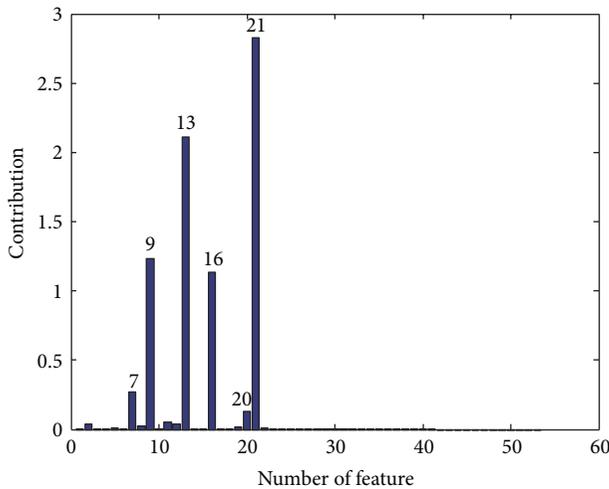
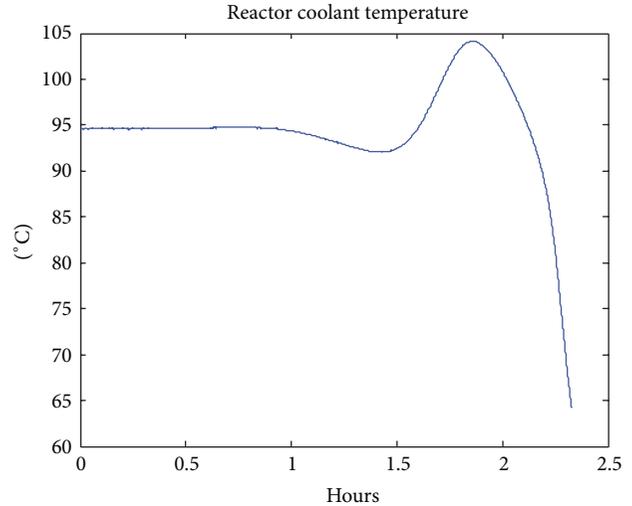


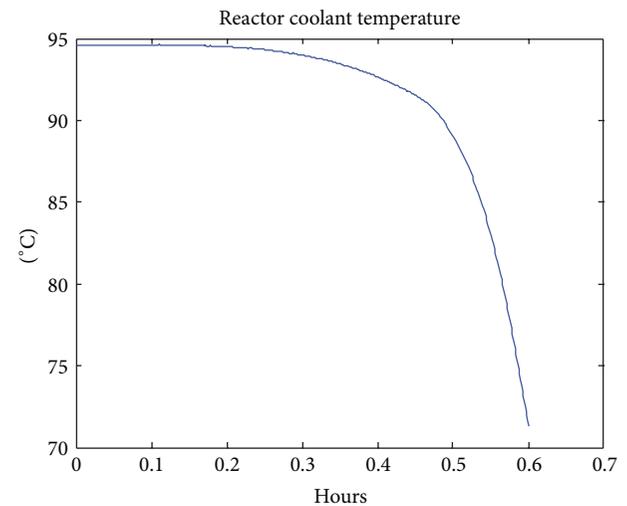
FIGURE 6: Contribution of all the 53 process variables to distinguish Fault9 and Fault11.

on faults. Multikernel function is selected as Gaussian kernel and polynomial kernel, each comprised of 50%. The contributions of each variable to the faults are computed with steps in Section 3.3 that is shown in Figure 6 and Table 3. From large to small, the proper importance of all the 53 process variables is reordered as $\{Vab.21, Vab.13, Vab.9, Vab.16, Vab.7, Vab.20, Vab.11, Vab.2, Vab.12, Vab.8, Vab.19, Vab.5, Vab.22, Vab.6, Vab.3, Vab.18, Vab.14, Vab.15, Vab.17, Vab.10, Vab.41, Vab.40, Vab.27, Vab.23, Vab.29, Vab.31, Vab.26, Vab.33, Vab.25, Vab.32, Vab.4, Vab.24, Vab.30, Vab.35, Vab.34, Vab.37, Vab.36, Vab.28, Vab.39, Vab.38, Vab.1, Vab.53, Vab.52, Vab.51, Vab.50, Vab.49, Vab.48, Vab.47, Vab.46, Vab.45, Vab.44, Vab.43, Vab.42\}$.

In the Following, the curves of the first two important $Vab.21$ and $Vab.13$ in TE process are given in Figures 7(a) and 7(b) and Figures 8(a) and 8(b), respectively. It expresses the



(a) Process $Vab.21$ (reactor coolant temperature) in Fault9



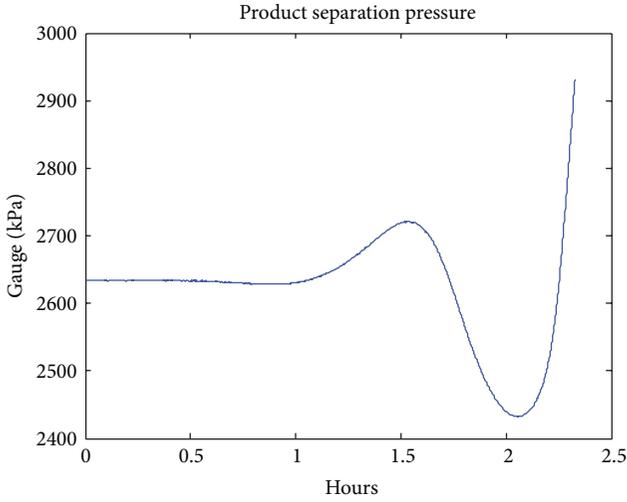
(b) Process $Vab.21$ (reactor coolant temperature) in Fault11

FIGURE 7: The changing of process $Vab.21$ in actual TE.

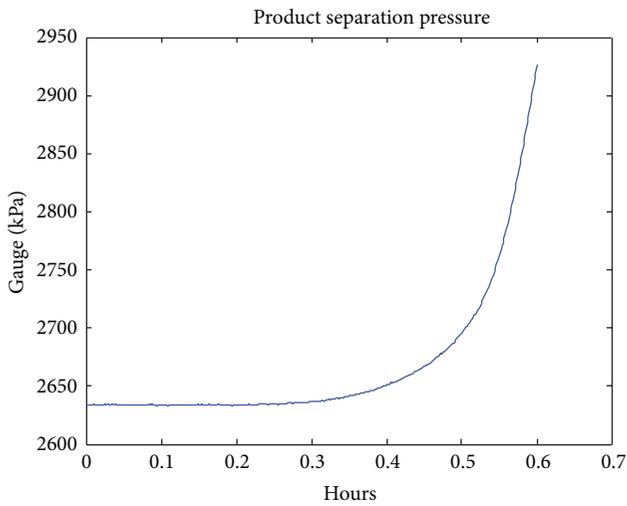
strong variation of process variables $Vab.21$ and $Vab.13$, actually.

According to the sequence of each process variable, the different feature sets are constructed as $\{Vab.21\}$, $\{Vab.21, Vab.13\}$, $\{Vab.21, Vab.13, Vab.9\}$, and so on. Nonlinear pattern classification of Fault9 and Fault11 is tested with support vector machine (SVM), which is widely used in pattern recognition. The parameters of SVM are optimized with cross-validation $c = 2035$ and $g = 1024$. With the above variable sets, the accuracy of fault separation between Fault9 and Fault11 is successively tested. The results are shown in Figure 9 and Table 4. It reveals that the separation accuracy becomes lower when the considered variables increase.

From the above results, we conclude that (1) if all the 53 process variables were used to separate Fault9 and Fault11, right ratio is merely 72.12%. It indicates that not all of the variables are directly related to certain fault. Some redundancy or irrelevant variables may decrease the classification accuracy



(a) Process *Vab.13* (product separation pressure) in Fault9



(b) Process *Vab.13* (product separation pressure) in Fault11

FIGURE 8: The changing of process *Vab.13* in actual TE.

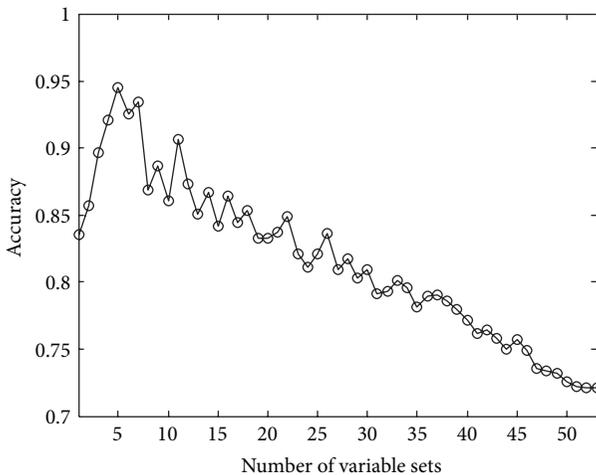


FIGURE 9: The accuracy with different feature sets to indentify Fault9 and Fault11 with testing data.

TABLE 2: State distribution in TE process.

Fault	Disturbance	Type
1	<i>A/C</i> feed ratio, <i>B</i> composition constant	Step
2	<i>B</i> composition, <i>A/C</i> ratio constant	Step
3	<i>D</i> feed temperature	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	<i>A</i> feed loss	Step
7	<i>C</i> header pressure loss-reduced availability	Step
8	<i>A, B, C</i> feed composition	Random
9	<i>D</i> feed temperature	Random
10	<i>C</i> feed temperature	Random
11	Reactor cooling water inlet temperature	Random
12	Condenser cooling water inlet temperature	Random
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16–20	Unknown	Unknown

TABLE 3: The contributions of 53 process variables to fault separation.

Process variable	Contribution	Process variable	Contribution
<i>Vab.21</i>	2.8273	<i>Vab.17</i>	0.0000
<i>Vab.13</i>	2.1145	<i>Vab.10</i>	0.0000
<i>Vab.9</i>	1.2318	<i>Vab.41</i>	0.0000
<i>Vab.16</i>	1.1313	<i>Vab.40</i>	0.0000
<i>Vab.7</i>	0.2687	<i>Vab.27</i>	0.0000
<i>Vab.20</i>	0.1319	<i>Vab.23</i>	0.0000
<i>Vab.11</i>	0.0522	<i>Vab.29</i>	0.0000
<i>Vab.2</i>	0.0355	<i>Vab.31</i>	0.0000
<i>Vab.12</i>	0.0259	<i>Vab.26</i>	0.0000
<i>Vab.8</i>	0.0191	<i>Vab.33</i>	0.0000
<i>Vab.19</i>	0.0092	<i>Vab.25</i>	0.0000
<i>Vab.5</i>	0.0012	<i>Vab.32</i>	0.0000
<i>Vab.22</i>	0.0006	<i>Vab.4</i>	0.0000
<i>Vab.6</i>	0.0004	<i>Vab.24</i>	0.0000
<i>Vab.3</i>	0.0002	<i>Vab.30</i>	0.0000
<i>Vab.18</i>	0.0000	<i>Vab.35</i>	0.0000
<i>Vab.14</i>	0.0000	⋮	⋮
<i>Vab.15</i>	0.0000	<i>Vab.42</i>	0.0000

and must be eliminated. (2) If the feature were selected as the first five process variables {*Vab.21, Vab.13, Vab.9, Vab.16, Vab.7*}, the accuracy increases to the highest as 94.55%. It means that the above five process variables are key to the fault separation. (3) If the model should be simplified at most, the process variable {*Vab.21*} is the best feature variable. We can recognize Fault9 and Fault11 according to the process changing of *Vab.21*.

On the other hand, Fault9 stands for the random disturbance to *D* feed temperature. Fault11 is random disturbance to reactor cooling water inlet temperature. While {*Vab.21*,

TABLE 4: The accuracy with different feature sets with testing data.

Feature set	Combination of variables	Accuracy
Set ₁	Vab.21	83.521%
Set ₂	Vab.21, Vab.13	85.731%
Set ₃	Vab.21, Vab.13, Vab.9	89.652%
Set ₄	Vab.21, Vab.13, Vab.9, Vab.16	92.123%
Set ₅	Vab.21, Vab.13, Vab.9, Vab.16, Vab.7	94.547%
Set ₆	Vab.21, Vab.13, Vab.9, Vab.16, Vab.7, Vab.20	92.532%
⋮	⋮	⋮
Set ₅₃	Vab.21, Vab.13, Vab.9, Vab.16, Vab.7, Vab.20, ... Vab.42	72.1199

Vab.13, Vab.9, Vab.16, Vab.7} are the reactor coolant temperature, product separation pressure, reactor temperature, stripper pressure, reactor pressure, respectively, it is easy to see that the five selected variables are fairly relative to Fault9 and Fault11. The simulation results keep pace with the reality.

5. Conclusions

Nonlinear redundancy and multicollinearity variables can decrease the accuracy in classifier that must be eliminated. For the problem, FNN in MKFDA subspace is studied in the paper. Nonlinear variables are projected to a new linear higher dimension subspace with single-kernel fisher discriminant analysis to get optimal classification with the intra-class nearest and inter-class farthest as most as possible. Furthermore, conventional single-kernel KFDA is expanded to multikernel method to solve the influence of each process variable with different distribution function. In order to reduce the higher dimension emerging in multi-KFDA subspace, FNN is composed to recognize the importance of each process variables on faults. According to simulation results in TE process, original variables are reduced to 5 in this paper, and the accuracy of tested right ratio reaches to 94.55% compared with tested right ratio 72.12% in the classifier between Fault9 and Fault11.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research is supported by National Natural Science Foundation of China (no. 51376204, no. 51075418, no. 50905194, and no. 61174015), the Natural Science Foundation Project (no. CQCSTC2012jjA40026, no. CQCSTC2012jjA90011), and the Research Foundation of Chongqing University of Sci. and Tec. (no. CK2011Z01, no. CK2011B04).

References

- [1] D. H. Lim, S. H. Lee, and M. G. Na, "Smart soft-sensing for the feedwater flowrate at PWRs using a GMDH algorithm," *IEEE Transactions on Nuclear Science*, vol. 57, no. 1, pp. 340–347, 2010.
- [2] M. J. Brusco and D. Steinley, "Exact and approximate algorithms for variable selection in linear discriminant analysis," *Computational Statistics and Data Analysis*, vol. 55, no. 1, pp. 123–131, 2011.
- [3] F. A. Michelsen, B. F. Lund, and I. J. Halvorsen, "Selection of optimal, controlled variables for the TEALARC LNG process," *Industrial and Engineering Chemistry Research*, vol. 49, no. 18, pp. 8624–8632, 2010.
- [4] F. Cipollini and G. M. Gallo, "Automated variable selection in vector multiplicative error models," *Computational Statistics and Data Analysis*, vol. 54, no. 11, pp. 2470–2486, 2010.
- [5] A. J. Miller, *Subset Selection in Regression*, Chapman and Hall, London, UK, 2002.
- [6] R. L. Mason and R. F. Gunst, *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*, John Wiley & Sons, Hoboken, NJ, USA, 2004.
- [7] Y. Yang and J. O. Pederson, "A Comparative study on feature selection in text categorization," in *Proceedings of the 14th International Conference on Machine Learning*, pp. 412–420, 1997.
- [8] K. Kira and L. A. Rendell, "The Feature selection problem: traditional methods and a new algorithm," in *Proceedings of the 9th National Conference on Artificial Intelligence (AAAI '92)*, pp. 129–134, July 1992.
- [9] B. Pfahringer, "Compression-based feature subset selection," in *Proceedings of the Workshop on Data Engineering for Inductive Learning (IJCAI '95)*, pp. 101–106, 1995.
- [10] J. C. Isaac, *Kernel methods and component analysis for pattern recognition [Ph.D. thesis]*, 2007.
- [11] J. C. Huang, J. S. Zhao, W. Sun, and Y. K. Ding, "PCA-based early fault diagnosis of solid waste incinerator," *Chemical Industry and Engineering Progress*, vol. 25, no. 12, pp. 1489–1492, 2006.
- [12] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [13] J. D. Wu, P. H. Chiang, Y. W. Chang, and Y. J. Shiao, "An expert system for fault diagnosis in internal combustion engines using probability neural network," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2704–2713, 2008.
- [14] D. F. Wang, S. J. Wang, and J. He, "Maintaining and fault removing on hydraulic system of CAK6140," *Machinery Design and Manufacture*, vol. 7, pp. 161–162, 2010.
- [15] J. H. Li and P. L. Cui, "Improved kernel fisher discriminant analysis for fault diagnosis," *Expert Systems With Applications*, vol. 36, no. 2, pp. 1423–1432, 2009.
- [16] M. Journée, Y. Nesterov, P. Richtárik, and R. Sepulchre, "Generalized power method for sparse principal component analysis," *Journal of Machine Learning Research*, vol. 11, pp. 517–553, 2010.
- [17] K. Kim, J.-M. Lee, and I.-B. Lee, "A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction," *Chemometrics and Intelligent Laboratory Systems*, vol. 79, no. 1-2, pp. 22–30, 2005.
- [18] R. Jenssen, "Kernel entropy component analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 847–860, 2010.
- [19] N. Otopal, "Restricted kernel canonical correlation analysis," *Linear Algebra and Its Applications*, vol. 437, no. 1, pp. 1–13, 2012.

- [20] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [21] H. Y. Wang and Z. H. Sheng, "Choice of the parameters for the phase space reconstruction of chaotic time series," *Journal of Southeast University*, vol. 30, no. 5, pp. 113–117, 2000.
- [22] Z. B. Zhu and Z. H. Song, "A novel fault diagnosis system using pattern classification on kernel FDA subspace," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6895–6905, 2011.
- [23] S. Mika, G. Ratsch, J. Weston, B. Schölkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Proceedings of the 9th IEEE Workshop on Neural Networks for Signal Processing (NNSP'99)*, pp. 41–48, August 1999.
- [24] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [25] J. Y. Gan and Y. W. Zhang, "Generalized kernel fisher optimal discriminant in pattern recognition," *Pattern Recognition and Artificial Intelligence*, vol. 15, no. 4, pp. 429–434, 2002.
- [26] 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.
- [27] N. Lawrence Kicker, "Tennessee eastman," 2013, <http://www.cheme.washington.edu/facresearch/faculty/ricker.html>.