Fault Diagnosis Based on Improved Kernel Fisher Discriminant Analysis

Zhengwei Li
School of Computer Science and Technology. China University of Mining and Technology
Xuzhou, China
Email:zwli@cumt.edu.cn

Zhengwei Li, Guojun Tan,Yuan Li
School of Information and Electrical Engineering, China University of Mining and Technology
Xuzhou, China
Email: {zwli,gjtan, Liyuan}@cumt.edu.cn

***Abstract*—There are two fundamental problems of the Kernel Fisher Discriminant Analysis (KFDA) for nonlinear fault diagnosis. The first one is the classification performance of KFDA between the normal data and fault data degenerates as long as overlapping samples exist. The second one is that the computational cost of kernel matrix becomes large when the training sample number increases. Aiming at the two major problems, in this paper, an improved fault diagnosis method based on KFDA(IKFDA) is proposed. There are two aspects are improved in the method. Firstly, the variable weighting vector was incorprated into KFDA which can improve the discriminant performance. Secondly, when the training sample number becomes large, a feature vector selection scheme based on a geometrical consideration is given to reduce the computational complexity of KFDA for fault diagnosis. Finally, Gaussian mixture model (GMM) is applied for fault isolation and diagnosis on the KFDA subspace. Experimental results show that the proposed method outperforms traditional kernel principal component analysis (KPCA) and general KDA algorithms.**

***Index Terms*—kernel fisher discriminant analysis, fault diagnosis, variable weighting, feature vector selection, gaussian mixture model**

I. INTRODUCTION

Since the fault diagnosis problem can be considered as a multi-class classification problem, pattern recognition methods with good generalization and accurate performances have been proposed in recent years. Choi et al.[1] proposed a fault detection and isolation methodology based on principal component analysis–Gaussian mixture model and discriminant analysis–Gaussian mixture model. Fisher discriminant analysis (FDA) has been proved to outperform PCA in discriminating different classes, in the aspect that PCA aims at reconstruction instead of classification, while FDA seeks directions that are optimal for discrimination[2]. However, FDA is a linear method. In order to handle the nonlinear problem of process data, kernel FDA (KFDA) is proposed by Mika et al.[3]. KFDA performs a nonlinear discriminant through kernel feature space mapping before FDA method is used. Yang et al.[4] made an in-depth analysis on the KFDA algorithm, and reformulated it as a two-step procedure: kernel principal component analysis (KPCA) plus FDA. Recently, KFDA has been proved superior to PCA and FDA in fault diagnosis, which makes it a promising way for process monitoring[5,6]. The basic idea of the kernel trick is that input data are mapped into a kernel feature space by a nonlinear mapping function and then these mapped data are analyzed.

However, the general KFDA method has some shortcomings for fault diagnosis. Firstly, the conventional KFDA views the same contribution of each variable to the classification and all variables are used in a same level so that the data sets are masked with irrelevant information. As a result, the classification performance of KFDA for fault diagnosis degenerates when the samples of the normal data and the fault data are overlapped[7]. Focusing on the multi-classification where data are overlapped, the paper proposes a variable-weighted schema into the general KFDA. Secondly, in the training stage of KFDA, it requires to store and manipulate the kernel matrix, the size of which is the square of the sample number. When the sample number becomes large, the eigen-decomposition and the matrix inversion calculation will be time-consuming, and then reducing the calculation time is very important. In this paper, a feature vector selection scheme based on a geometrical consideration[8] is given to reduce the computational complexity of KFDA when the number of training samples becomes large.

This paper is organized as follows. In Section 2, KFDA is explained. In Section 3, Improved KFDA is proposed. Fault diagnosis results of the above schemes

Project for major transformation of scientific and technological achievements from Jiangsu province (BA2008029).

Corresponding author: Zhengwei Li, Email: zwli@cumt.edu.cn

doi:10.4304/jsw.7.12.2657-2662

are given by simulations in Section 4. Finally in Section 5, the main points are summarized.

II. KFDA

The basic idea of KFDA is to solve the problem of linear FDA in an implicit feature space F. However, it is difficult to do so directly because the dimension h of the feature space F can be arbitrarily large or even infinite. In implementation, the implicit feature vector in F does not need to be computed explicitly, while it is just done by computing the inner product of two vectors in F with a kernel function.

Let the dimensionality of original sample feature space be m and the number of sample classes be C, the total original sample $X=\{X_1,X_2,...,X_C\}$, the j^{th} (j = 1, 2,. . . ,C) class X_j contains N_j samples, namely $X_j=\{X_1^j,X_2^j,...,X_{N_j}^j\}$. *Here,* $X_1^j,X_2^j,...,X_{N_j}^j\in R^m$ is used to denote the N_j training samples(column vectors) of class j for KFDA learning. N is the total number of original training samples, and then $N=\sum_{j=1}^{C}N_j$.

By the nonlinear mapping ϕ, the measured inputs are extended into the hyper-dimensional feature space as follows

$$\phi: x\in R^m \to \phi(x)\in F^h \tag{1}$$

The mapping of sample x_i is simply noted as $\phi(x_i)=\phi_i$, the total mapped sample set and the j^{th} mapped class are given by

$$\phi(X)=\{\phi(X_1),\phi(X_2),...,\phi(X_C)\}$$
$$\phi(X_j)=\{\phi(X_1^j),\phi(X_2^j),...,\phi(X_{N_j}^j)\}$$

The mean of the mapped sample class $\phi(X_j)$ is given by $m_j=(1/N_j)\sum_{i=1}^{N_j}\phi(x_i)$, and the global mean of the total mapped samples is given by $m=(1/N_j)\sum_{j=1}^{C}\sum_{i=1}^{N_j}\phi(x_i)$. The within-class scatter matrix S_W in F and between-class scatter matrix S_B in F are defined as

$$S_W=\frac{1}{N}\sum_{j=1}^{C}\sum_{I=1}^{N_j}(\phi(x_i^j-m_j)(\phi(x_i^j-m_j)^T \tag{2}$$

$$S_B=\frac{1}{N}\sum_{j=1}^{C}N_j(m_j-m)(m_j-m)^T \tag{3}$$

Performing FDA in F means maximizing the between class scatter matrix S_B and minimizing the within-class scatter matrix S_W. This is equivalent to maximizing the following function

$$J(w)=\arg\max_{w}\frac{|w^TS_Bw|}{|w^TS_Ww|} \tag{4}$$

The problem of KFDA is converted into finding the leading eigenvectors of $S_W^{-1}S_B$. Here, the dimension of $S_W^{-1}S_B$ can be infinite, and it cannot be calculated directly. Since any solution $w\in F$ must lie in the span of all the samples in F, there exists coefficients $\alpha=\{\alpha_i,i=1,2,...,n\}$, such that

$$w=\sum_{i=1}^{n}\alpha_i\phi_i \tag{5}$$

Combine with (5), we can write

$$w^TS_Bw=\alpha^TK_B\alpha \tag{6}$$

$$w^TS_Ww=\alpha^TK_W\alpha \tag{7}$$

Here, K_B and K_w are in the form of matrix {k(x, y)}, the kernel matrix of the samples (x, y), and k(x, y) is the kernel function. Where

$$\begin{cases}K_B=\dfrac{1}{C(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{C}(q_i-q_j)(q_i-q_j)^T\\ q_i=\left(\dfrac{1}{N_i}\sum_{j=1}^{N_i}k(x_1,x_j),\dfrac{1}{N_i}\sum_{j=1}^{N_i}k(x_2,x_j),...,\dfrac{1}{N_i}\sum_{j=1}^{N_i}k(x_N,x_j)\right)T\end{cases} \tag{8}$$

$$\begin{cases}K_W=\dfrac{1}{C}\sum_{i=1}^{C}\dfrac{1}{N_i}\sum_{j=1}^{N_i}(p_j-p_i)(p_j-p_i)^T\\ p_j=\left(k(x_1,x_j),k(x_2,x_j),...,k(x_N,x_j)\right)T\end{cases} \tag{9}$$

So the solution of Eq. (4) can be obtained by maximizing

$$J(\alpha)=\arg\max_{\alpha}\frac{|\alpha^TK_B\alpha|}{|\alpha^TK_W\alpha|} \tag{10}$$

Then, the problem of KFDA is converted into finding the leading eigenvectors of $K_W^{-1}K_B$. Let column vectors β_i (i = 1,2, . . .,N) be the eigenvectors of $K_W^{-1}K_B$. To a new columnvector sample x_{new}, the mapping to the feature space is $\phi(x_{new})$. The projection of x_{new} onto the eigenvectors β_i=($\beta_{i1},\beta_{i2},\ldots,\beta_{iN}$)(i = 1, 2,. . .,$N$) is t = ($t_1,t_2,\ldots,t_N$)^T, and it is also called KFDA-transformed feature vector

$$t_i=(w\cdot\phi(x_{new}))=\sum_{j=1}^{N}\beta_{ij}k(x_{new},x_j),\quad i=1,2,...,N \tag{11}$$

When KFDA is used for feature extraction, a problem arises that the matrix K_W cannot be guaranteed to be nonsingular. Several techniques have been proposed to handle this problem for numerical computation. In this paper, when the matrix K_W is singular, it is replaced with $K_W+\mu I$, where μ is a very small constant and I is an identity matrix[3,9,10].

III. IMPROVED KFDA(IKFDA)

A. Variable Weighting Vector

To fault identification, the variable weighting determines key variables responsible for the fault from the datasets masked with much irrelevant information by maximizing separation between the normal and each fault data sets. The variable weighting maximizes separation between the normal and each fault data. By making full use of the normal data information, the weight vector of each fault can be obtained. After fault data are weighted by the corresponding weight vectors, KFDA is performed on these weighted fault data, which offers important supplemental classification information to KFDA.

Fault diagnosis is often characteristic of large scale and nonlinear behavior. When all variables are used in a same level, the data sets are masked with irrelevant information, which results in the classification performance degenerating. Correctly representing the corresponding variable's contribution to a special fault, the weight vector is helpful to extract discriminative features from overlapping fault data which effectively improves the multi-classification performance of KFDA.

FDA is a well-known linear technique for reducing dimensions and pattern classification. It determines a set of Fisher optimal discriminant vectors that maximize the scatter between the classes while minimizing the scatter within each class. Different from traditional selection methods where the deleted variables and the selected variables are essentially weighted with discrete values: 0 and 1, respectively, the variable weighting is to weight the variables with continuous non-negative values[10]. Pair-wise FDA is performed on normal data and each class of fault data to gain the Fisher optimal discriminant vector, here named the fault direction. Each fault direction associated with a special fault optimally separates the fault data from normal data. Taking into account nonlinear characteristics of most industrial processes, we investigate the nonlinear pair-wise variable weighting. The weight vector of each fault maximizes the distance between the normal and each class of fault data.

The concept of kernel target alignment was proposed by Cristianini et al. to evaluate the similarity between two kernel matrices. Since kernel-based learning methods are based around kernel matrix, its properties reflect the relative positions of the points in the feature space. For the two-class (the normal class and the fault class classification problem $x \in x_{0,f} \to \{-1,1\}$, consider the kernel target matrix yy^T, where $y = [-f_{n_0}, f_{n_0}]^T$ then the kernel target alignment is given by

$$A = \frac{< K, yy^T >_F}{\sqrt{< K, K >_F < yy^T, yy^T >_F}} = \frac{y^T K y}{n\|K\|_F} \tag{12}$$

Thus, Eq. (12) can be rewritten by

$$\max_{w_f} A(w_f) = \frac{y^T K_{w_f} y}{n\left\|K_{w_f}\right\|_F} \qquad s.t.\ \mathrm{w_f}(i) \ge 0 \quad \mathrm{i} = 1,...,\mathrm{m} \tag{13}$$

To obtain the weight vector w_f, the width of Gaussian kernel σ first requires to be determined using cross validation with pair-wise KFDA on $x_{0,f}$ by minimizing the total classification error rate, and then the optimization problem, i.e. Eq. (13), should be solved. The process is repeated until all *c* fault classes are analyzed and all weight vectors $(w_1..., w_f, ...w_c)$ are obtained[11-14].

Consider Gaussian kernel function, any element of the kernel matrix K_{w_f} of $\tilde{x}_i$ can be given by

$$K_{w_f}(i,j) = \frac{\left\|\tilde{x}_i^{0,f} - \tilde{x}_i^{0,f}\right\|}{2\sigma^2} = \frac{\left\|diag(w_f)x_i^{0,f} - diag(w_f)x_i^{0,f}\right\|}{2\sigma^2} \quad \mathrm{i,\ j} = 1,...(\mathrm{n}_0 + n_f) \tag{14}$$

where $x^{0,f}$ or $\tilde{x}^{0,f}$ is the ith row of $x^{0,f}$ or $\tilde{x}^{0,f}$.

Due to depending on only the kernel matrix K_{W_F}, the kernel target alignment is selected as the variable weighting criteria. Due to depending on only the kernel matrix K_{W_F}, the kernel target alignment is selected as the variable weighting criteria. Thus, the weight vector can be obtained by solving the following optimal problem. Therefore, the Rayleigh quotient is selected as the variable weighting criteria. Then the variable weighting becomes the following optimal problem:

$$\max_{w_f} \int^{\phi}(w_f) = \frac{\alpha;^T (K_{W_F} W K_{W_F})\alpha}{\alpha^T (K_{W_F} K_{W_F})\alpha} \qquad s.t.\quad \mathrm{w_f}(i) \ge 0\ \ \mathrm{i} = 1,...,\mathrm{m} \tag{15}$$

In the above equation, Rayleigh quotient depends on not only the kernel matrix but also the optimal discriminant vector K_{W_f} such that KFDA should be re-performed to obtain the optimal discriminant vector α during the optimization procedure, which would be computationally expensive. Instead of the Rayleigh quotient in KFDA, the kernel target alignment is selected as the variable weighting criteria.

B. Feature Vector Selection

In this paper, a preprocessing scheme called feature vector selection (FVS) is adopted to reduce the computational complexity of KFDA whereas preserve the geometrical structure of the whole data in F.

In Eq. (5), all the training samples in F, ϕ_i (i = 1, 2,…,n), are used to represent eigenvector w. In practice, the dimensionality of the subspace spanned by ϕ_i is just equal to the rank of kernel matrix K, and the rank of

K is often less than n, that is $rank(K) <$ n. The FVS scheme is look for vectors that are sufficient to express all of the data in F as a linear combination of those selected vectors in F. Suppose a basis of the feature vectors, ϕ_{b_i} (i=1, 2, . . ., $rank(K)$; $b_i \in [1,n]$) is known, then Eq. (5) can been rewritten as follows:

$$w = \sum_{i=1}^{rank(K)} \alpha_{bi}\phi_{bi} \tag{16}$$

Since $rank(K)$ <<n, this will largely reduce the computational complexity. In order to obtain such a basis of the feature vectors in F, a geometrical consideration based method[15] is used in our research. The idea is to look for a subset of the samples whose mappings in F are sufficient to express all of the data in F as a linear combination of them.

Assume that $S = \{x_{S_1}, x_{S_2}, ..., x_{S_{LS}}\}$ is a selected sample set, where LS is the number of selected vectors. The estimation of the mapping of any vector x_i is regarded as a linear combination of the samples in S, and it can be written as follows:

$$\hat{\phi}_i = \phi_S \cdot \tau_i \tag{17}$$

Now we aim to find the coefficient vector v_i, so as to make the estimated mapping ϕ_i as close to the real mapping $\hat{\phi}_i$ as possible. This can be achieved by

$$\delta_i = \frac{\| \hat{\phi}_i - \phi_i \|^2}{\| \phi_i \|^2} \tag{18}$$

Putting the derivatives with respect to v_i to zero, and rewriting it in matrix form, obtains

$$\min(\delta_i) = 1 - \frac{K_{Si}^T K_{SS}^{-1} K_{Si}}{k_{ii}} \tag{19}$$

where $k_{ii} = k(x_i, x_i)$, $k_{SS} = (k(x_{S_p}, x_{S_q}),$ $1 \le p \le L_S$, $1 \le q \le L_S)$ is a square matrix of dot products of the selected vectors, and $k_{Si} = (k(x_{S_p}, x_i),\ 1 \le p \le L_S)$ is the vector of dot product between x_i and the selected set S.

Letting Eq. (20) satisfy all the samples, we can get

$$\max_S J_S = \frac{1}{n} \sum_{x_i} \frac{K_{Si}^T K_{SS}^T K_{Si}}{k_{ii}} \tag{20}$$

The solution of the problem can be obtained by an iterative process, and the process stops when K_{SS} is no longer invertible or a predefined number of selected vectors is reached [16-18].

IV. IKFDA AND GMM FOR FAULT DIAGNOSIS

After calculate the optimal discriminant vectors in *F*, the IKFDA discriminant vectors of the *p* types of pattern data can be obtained. Assume $z_i\{z_{i,1},...z_{i,k},...z_{i,m_i}\}$, *i* = 1,. . . ,*p* are the first-two dimension feature vectors of the IKFDA discriminant score vectors, which are the output of IKFDA based feature extraction and input of Gaussian mixture model (GMM) based pattern classification.

In IKFDA-GMM, for the p patterns for fault isolation, we can have a set of GMMs. The density estimation of class i is made by mixture of Gaussians density estimate. The GMM of class i can be written as:

$$p_{i,GMM} = \sum_{j=1}^{M} w_j^i p_j^i(z_i; u_j, \sum_j) \tag{21}$$

$$\mathrm{p}_j^i(z_i; u_j, \sum_j) = \frac{1}{(2\pi)\left|\sum_j\right|^{1/2}} \exp\{-\frac{1}{2}(z_i - u_j)^T \sum_j^{-1}(z_i - u_j)\} \tag{22}$$

where M is the number of mixtures, w_j^i j =1,. . . ,*M*, are the mixture weights with the constraint that $\sum_{j=1}^{M} w_j^i = 1$, $\mathrm{p}_j^i(z_i; u_j, \sum_j)$ means the normal distribution with mean and covariance matrix $\sum_j$ of pattern *i*.

The optimal number of mixtures should be considered. In order to determine the number properly, both the precise of the model and modeling time should be taken into consideration. As the number of the local models increases, the data distribution is described more correctly, at the same time, the number of estimated parameters and the time of modeling are increasing.

The parameters of the GMM w_j^i , u_j and $\sum_j$ can be estimated by the expectation-maximization algorithm, which yields a maximum likelihood estimation.

$$u_j^i = \frac{\sum_{k-1}^{m} p_j^i(z_{i,k}) z_{i,k}}{\sum_{k-1}^{m} p_j^i(z_{i,k})}$$
$$\sum_j^i = \frac{\sum_{k-1}^{m} p_j^i(z_{i,k}) z_{i,k}}{\sum_{k-1}^{m} p_j^i(z_{i,k})} \tag{23}$$

The classification performance can be evaluated by the test dataset, and the class label of test data can be calculated through:

$$Classify(z_i) = \arg\max_{i-1}^{p} P_{i,GMM} \tag{24}$$

The procedure of the fault diagnosis system is summarized below.

(1) Real-time unclassified process sample imports.

(2) Project to IKFDA first-two dimension feature subspace.

(3) Classify the pattern by GMM.

(4) Fault isolation and diagnosis.

V. APPLICATIONS

In this section, Tennessee Eastman (TE) process description and simulation design are presented firstly,

and then, simulation results and discussion of slight and serious imbalance problems are presented.

A. Tennessee Eastman (TE) Process

Tennessee Eastman (TE) Industrial Challenge Problem is designed to provide a realistic industrial process by the Eastman Chemical Company. The process consists of five major unit operations: a product condenser, a recycle compressor, a reactor, a product stripper, and a vapor–liquid separator. As a well-known benchmark simulation, it has been widely used to compare and evaluate the performance of various monitoring approaches.

There are 41 measured variables and 12 manipulated variables in the TE process, in this paper, the selected 33 monitoring variables include 22 measured variables and 11 manipulated variables shown in table I. There are 22 process patterns (1 normal operating condition and 21 faulty conditions) in the TE process. In order to design the simulation study reasonably, in this paper, the TE process simulator is used to generate the normal and five classes of fault data, i.e. fault 4, 8, 13, 14 and 19. These five faults covering all the fault types in TE process are divided into two simulation studies.

Each pattern combination includes the normal pattern and three different faults. These selected process patterns for simulation study are listed in Table II. For each pattern, there are 960 observations, and all faults are introduced in the process at sample 160. The simulation data are separated into two parts: the training and testing dataset. The training and testing data amount of these two cases are also listed in Table I.

B. Simulation Results and Discussion

In both of the two cases, we select radial basis kernel function in KPCA and choose the width of Gaussian kernel as 500m, where m is the dimension of the inputs. KPCA based feature extraction is applied with 85% variation of eigenvalue, and the confidence limit in the Gaussian distribution is 98%.

Firstly, IKFDA based first-three dimension projections of the four patterns (normal, fault 3, 7 and 14) are presented. According to the first–third dimension projection, fault 3 has overlapped with fault 7. As the fact that the rare classes have less impact on accuracy than the common classes, and the pattern determination is apt to the majority one.

Fault diagnosis results are listed in Table III, we can find that 28% of fault 7 data are misclassified as the normal pattern by general KFDA. Such high rate error will endanger the process, even results in serious accidents. Considering the data amount of each pattern, normal pattern is the majority class. The average diagnosis rates of general KFDA are listed in Table III.

Comparing the performance of IKFDA-GMM with general KFDA and PCA, the diagnosis rate of normal pattern reaches to 99%, but the diagnosis performances of fault 7 and 14 are even a litter poorer. The diagnosis performances of all the three fault patterns are improved compared those of general KFDA and KPCA. The diagnosis rate of normal pattern descends form 99% to 66%. For example, the rate of misclassifying fault 7 as normal pattern falls from 47% to 2%. It is obvious that the IKFDA is instable. Comparing to general KFDA, the

TABLE I.
MONITORED VARIABLES IN TE PROCESS

Number	Measured variables	No	Measured variables	No	Manipulated Variables
1	A feed	11	Product separator temperature	21	D feed flow valve
2	D feed	12	Product separator level	22	E feed flow valve
3	E feed	13	Product separator pressure	23	A feed flow valve
4	Total feed	14	Product separator underflow	24	Total feed flow7 valve
5	Recycle flow	15	Stripper level	25	Compressor recycle valve
6	Reactor feed rate	16	Stripper pressure	26	Purge valve
7	Reactor pressure	17	Stripper underflow	27	Separator pot liquid flow7 valve
8	Reactor level	18	Stripper temperature	28	Stripper liquid product flow valve
9	Reactor temperature	19	Stripper steam flow	29	Stripper steam valve
10	Purge rate	20	Compressor work	30	Reactor cooling water flow

TABLE II.
SELECTED PROCESS PATTERNS FOR SIMULATION

Pattern	Fault description	Type	Number of training data
Normal			400(300)
Fault 3	Reactor cooling water inlet temperature	Step	200(300)
Fault 7	A,B,C feed composition	Random variation	250(300)
Fault 13	Reactor cooling water valve	Sticking	300(300)

TABLE III.
FAULT DIAGNOSIS RESULTS(%)

Methods	Normal	Fault 3	Fault 7	Fault 14
KPCA	66	71	53	62
KFDA	79	76	68	70
IKFDA-GMM	99	98	98	97

performances of the proposed IKFDA approaches are better in some cases.

VI. CONCLUSIONS

In recent years, KFDA has been utilized directly for nonlinear process fault diagnosis, and it has been proven to outperform conventional FDA method. This paper focuses on the improvement of KFDA for fault diagnosis from two aspects, which provides effective tools for fault diagnosis of nonlinear multivariate process.

Firstly, the classification performance of KFDA may degenerate as long as overlapping samples exist. The nonlinear variable weighting finds out the weight vector of each fault by maximizing the variable weighting criteria. Each weight vector maximizes separation between the normal data and each class of fault data. By weighting fault data with the corresponding weight vector, the proposed method extracts discriminative features more effectively than the traditional KFDA from overlapping fault data.

Secondly, a feature vector selection scheme based on a geometrical consideration is adopted for sample vector selection before KFDA calculation. Simulations conducted on TE process have shown that, IKFDA based on feature vector selection has nearly the same fault recognition rates as KFDA method. Moreover, IKFDA based on feature vector selection method can reduce the computational complexity significantly, especially when the training sample set is very large.

Finally, Gaussian mixture model (GMM) is applied for fault isolation and diagnosis on the KFDA subspace. Experimental results show that the proposed method outperforms traditional kernel principal component analysis (KPCA) and general KDA algorithms.

ACKNOWLEDGMENT

This work was supported by the fund project for major transformation of scientific and technological achievements from Jiangsu province(BA2008029).

REFERENCES

[1] Choi, S.W., Park, J.H. and Lee, I.B.. “Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis”. *Computers and Chemical Engineering*, 2004, 28(8) , pp. 1377-1387.

[2] Chiang, L. H., Russell, E. L., & Braatz, R. D.. “Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis”. *Chemometrics and Intelligent Laboratory Systems*, vol.50, pp. 243-252, April 2000.

[3] Mika,s., “Fisher discriminant analysis with kernels”, *In Proceedings of IEEE Neural Networks for Singal Processing Workshop Madison*, WI,USA,1999, pp. 41-51.

[4] Yang, J., Jin, Z., Yang, J.Y., Zhang, D. and Frangi, A.F., “Essence of kernel Fisher discriminant: KPCA plus LDA.” *Pattern Recognition*, 2004,37(10) , pp. 2097-2100.

[5] Cho, H.W., “Nonlinear feature extraction and classification of multivariate process data in kernel feature space”. *Expert Systems with Applications*, 2007,32(2) , pp. 534-542.

[6] Cho, H.W.. “Identification of contributing variables using kernel based discriminant modeling and reconstruction”. *Expert Systems with Applications*, 2007,33(2), pp. 274-285.

[7] Baudat,G., & Anouar, F.. “Kernel-based methods and function approximation”. *In Proceedings of International Conference on Neural Networks*, Washington, DC,2001, pp. 1244-1249.

[8] Cui, P. L., Li, J. H., & Wang, G. Z.. “Improved kernel principal component analysis for fault detection”. *Expert Systems with Applications*,2008,34(2), pp. 1210-1219.

[9] Louw, N., Steel, S.J.. “A review of kernel Fisher discriminant analysis for statistical classification”. *South African Statist. J.* 2005(39), pp. 1-21.

[10] Dong-Sheng Cao,Yi-Zeng Liang,Qing-Song Xu, “Exploring nonlinear relationships in chemical data using kernel-based methods”. *Chemometrics and Intelligent Laboratory Systems* 2011(101), pp. 106-115.

[11] Zhi-Bo Zhu, Zhi-Huan Song. “A novel fault diagnosis system using pattern classification on kernel FDA subspace”.*Expert Systems with Applications* 2011(38) , pp. 6895-6905.

[12] Achmad Widodo, Bo-Suk Yang. “Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors”. *Expert Systems with Applications* 2007(33), pp. 241-250

[13] Guang Dai, Dit-YanYeung, Yun-Tao Qian. “Face recognition using a kernel fractional-step discriminant analysis algorithm”. *Pattern Recognition* 2007(40):229-243.

[14] Zhi-Bo Zhu, Zhi-Huan Song. “Fault diagnosis based on imbalance modified kernel Fisher discriminant analysis”. *Chemical Engineering Research and Design* 2010(88), pp. 936–951.

[15] Ji-Dong Shaoa, Gang Ronga Jong Min Leeb. “Learning a data-dependent kernel function for KPCA-based nonlinear process monitoring”. *Chemical Engineering Research and Design*, 2009(87) pp. 1471–1480.

[16] Hyun-Woo Cho. “Nonlinear feature extraction and classification of multivariate process data in kernel feature space”. *Expert Systems with Applications*, 2007(32), pp. 534-542.

[17] Cheong Hee Park, Haesun Park. “Nonlinear feature extraction based on centroids and kernel functions”. *Pattern Recognition*, 2004(37) pp. 801-810.

[18] N. Louw, S.J. Steel. “Variable selection in kernel Fisher discriminant analysis by means of recursive feature elimination”. *Computational Statistics & Data Analysis* 2006(51), pp. 2043-2055.

Zhengwei Li, My research interests mainly include machine learning and fault diagnosis. Currently, I am an associate professor at the School of Computer Science and Engineering, China University of Mining and Technology. I received my B.Sc. and M.Sc. degrees in computer science all from Department of Computer Science & Technology, China University of Mining and Technology in 1999 and 2005 respectively.