A Novel Integrated SVM for Fault Diagnosis Using KPCA and GA

Jinning LI
Harbin Institute of Technology, No.92 Xidazhijie Street, Harbin city, Heilongjiang Province, China.
Email: 1150410107@stu.hit.edu.cn

Abstract. Fault diagnosis has been more and more significant in modern factories to ensure the proper functionality of the manufacturing process and to improve product quality and efficiency. A novel integrated SVM model is developed in this paper. Since KPCA can implicitly mapped the data into a higher nonlinear feature space, it is applied in our model to extract key information of original data sets for further training of SVMs. SVM is the main part to detect and identify different kinds of faults, and GA is utilized to optimize the parameters of SVM to avoid over fitting or under fitting problems caused by wrong parameters. Compared with other fault diagnosis methods applied on Tennessee Eastman (TE) process benchmark, we can safely conclude that the proposed model can achieve good predictive accuracy in a relatively short time based on our experiment results.

1. Introduction
High-performance industrial processes in modern factories, such as chemical production and manufacturing, are closely related to safety issues. Thus, fault diagnosis of those complex systems can be crucial to ensure the proper functionality of the whole process and to improve product quality and efficiency. Fault detection, fault separation and fault classification are three important components of fault diagnosis [1]. An algorithm with higher accuracy and less processing time is always needed in fault diagnosis. Many researchers have been dedicated to this field. Among all statistically-based monitoring technique, the most popular one is principle components analysis (PCA). Fault diagnosis systems based on PCA have been proved to be valid and have been applied to complex industrial systems successfully. It is employed to extract key information from a higher dimension and reduce the original data to a lower dimension which also has major trends as before [2]. There are also many other fault diagnosis methods, for example, Partial least squares (PLS), Support Vector Machine (SVM), Independent Component Analysis (ICA), neural network (NN) and decision tree.

However, because the main part of PCA is a linear projection of the given data set, it can only extract linear structure and fails to identify nonlinear structures in many general data sets from real industrial process. Therefore, many researchers make efforts to develop the nonlinear extension of PCA. The so-called kernel-PCA is proposed by Scholkopf et al. (1999) [3]. It can implicitly transform the data to a space with higher dimension and in exactly this higher dimension space, we apply PCA to the data set. In this way, researchers can deal with nonlinear structures and carry out fault diagnosis successfully. Cho et al. [4] then proposed an identification strategy by calculating statistic sensitivities of every test variable in terms of T² and Q. Bin Shams [5] developed a new fault identification procedure using KPCA. He utilized the power series approximation in the process of fault diagnosis.
Besides KPCA, support vector machine (SVM) can also effectively perform fault diagnosis on industrial big data according to Khan [6]. SVM is a margin-based classifier. It can learn from a relatively small sample and then get good generalization capabilities, and thus this kind of method is frequently used in real world classification [7]. Taeshik Shon [8] combined the benefits of supervised and unsupervised learning to form an enhanced SVM, which can detect and classify novel attacks in network as a general framework. Horng [9] combined a hierarchical clustering algorithm and simple SVM to reduce the training time with an improved accuracy of SVM classification result. Eskin [10] built an unsupervised framework which was applied in clustering method, k-nearest neighbour and SVM without any supervision.

Nevertheless, there are also some limitations of standard SVM, such as the parameters selection in SVM can affect the performance of the classification, and it will generate plenty of support vectors because of RBF. Moreover, training time can also be longer. Thus, an efficient method to search appropriate parameters is needed to refine standard SVM. Genetic Algorithm (GA) is a computational model which can search for optimum solution by imitating natural selection in theory of evolution and principles of genetics of processes of biological evolution. Thus, GA can be a powerful tool for parameters selection in SVM.

To cut back training time and the huge number of support vectors generated by standard SVM when detecting fault in Tennessee Eastman process, we develop a new method using KPCA to abstract the key structure of raw data, using GA to optimize the parameters in SVM and using SVM to detect and classify different fault types.

2. KPCA-SVM-GA Model

2.1. Kernel Principle Component Analysis

KPCA is a bridge that connects linear structures and nonlinear structures. It can extract nonlinear structure of a given data set because it can implicitly map the data set into a high-dimension nonlinear feature space. The advantage of KPCA is that it can reduce its compute largely, because the calculations of inner product in nonlinear feature space in a high dimension are converted to calculations of kernel function in the original space.

According to Mercer theorem, each given symmetric kernel function $\kappa(x_i, x_j)$ is an inner product of a high-dimension feature space if and only if for all $g(x) \neq 0$, $\int g(x)^2 dx < \infty$, s.t. $\int \kappa(x,y)g(x)g(y)dx dy \geq 0$. Let $\kappa(x_i, x_j)$ be the kernel function, and thus we can find a Hilbert space $H_f$ and a function $\Phi : R^m \rightarrow H_f$, s.t.

$$\kappa(x_i, x_j) = \sum_{n=1}^{f} \Phi_n(x_i) \cdot \Phi_n(x_j) \quad (1)$$

We consider $\chi \in R^{m \times n}$ as the input space and $x_1, x_2, x_3 ... x_m$ are m training samples.

We consider an input data matrix $\chi \in R^{n \times m}$ being mapped to a higher dimension feature space $H$ by $x_i \rightarrow \Phi_i = \varphi(x_i) : R^m \rightarrow H^f$. In a matrix form, the data which is mapped into a higher-dimension space $\Phi \in R^{n \times h}$ where h is the dimension of the feature space and $h >> m$. The data in the feature space should satisfy centralized condition. The covariance matrix in the mapped feature space is given as

$$C = \frac{1}{m} \sum_{i=1}^{n} \varphi(x_i)\varphi(x_i)^T \quad (2)$$

We evaluate eigenvalues $\lambda \geq 0$ and eigenvectors $\nu$ of matrix $C$, thus,

$$(\varphi(x_i) \cdot Cv) = \lambda(\varphi(x_i) \cdot \nu) \quad (3)$$

All eigenvectors are the linear span of $\varphi(x_1), \varphi(x_2), ..., \varphi(x_m)$, namely,

$$\nu = \sum_{i=1}^{m} \alpha_i \varphi(x_i) \quad (4)$$
Thus, 
\[ \frac{1}{m} \sum_{\mu=1}^{m} \alpha_\mu \left( \sum_{\omega=1}^{m} (\varphi(x_\mu) \cdot \varphi(x_\omega)) \varphi(x_\mu) \right) = \lambda \sum_{\mu=1}^{m} (\varphi(x_\mu) \cdot \varphi(x_\mu)) \] (5)

Where \( \nu = 1, 2, ..., m \). We define a matrix \( K \in \mathbb{R}^{m \times m} \), where \( K_{\mu\nu} = \varphi(x_\mu) \cdot \varphi(x_\nu) \). Thus, (5) can be simplified to

\[ M\alpha = \Lambda \alpha \] (6)

Eigenvalues and eigenvectors of \( C \) can be solved from equation (6). We can use the kernel function to substitute the inner product in the feature space, thus

\[ (\nu^k \cdot \varphi(x)) = \sum_{i=1}^{m} (\alpha_i) K(x_i, x) \] (7)

Then if the equation is not established, we need to centralize the kernel function

\[ K_{\mu\nu} \rightarrow K_{\mu\nu} - \frac{1}{m} \left( \sum_{\omega=1}^{m} K_{\mu\omega} + \sum_{\omega=1}^{m} K_{\omega\nu} \right) + \frac{1}{m^2} \sum_{\omega, \tau=1}^{m} K_{\omega\tau} \] (8)

Thus, by introducing kernel function to our algorithm, we avoid the explicitly calculation of the mapping to feature space.

### 2.2. Support Vector Machine

Suppose we add labels to each given training simple \( x_1, x_2, x_3, ..., x_m \), thus the training set is \( \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\} \in \mathcal{X} \). After using KPCA to extract the main structure from the original space, we can express the training data set as \( \{(t_1, y_1), (t_2, y_2), ..., (t_p, y_p)\} \in \mathcal{T}, t \in \mathbb{R}^d, d < n \). Our basic idea is to formulate a function \( f(t) \) with a deviation no more than \( \varepsilon \) from the given labels \( y_i \) corresponding to training data. Sometimes we need to allow for some errors, thus Cortes and Vapnik [11] introduced slack variables \( \xi_i, \xi_i^* \) to solve the problem caused by uncompromised constraints during the process of optimization. Thus, equation (9) can be obtained.

\[ \text{minimize} \quad \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \] (9)

subject to

\[ y_i - \langle w, t_i \rangle - b < \varepsilon + \xi_i \]

\[ \langle w, t_i \rangle + b - y_i < \varepsilon + \xi_i^* \]

\[ \xi_i, \xi_i^* \geq 0 \]

where \( C > 0 \) is a constant. The existence of \( C \) can regulate the degree to which the deviations more than \( \varepsilon \) of \( f \) can be omitted. The corresponding \( \varepsilon \)-insensitive loss function \( |\xi|_\varepsilon \) is manifested by

\[ |\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \] (10)

### 2.3. Proposed Model for Fault Identification

For fault identification problems, we can sort them into classification problems. Procedures of the presented KPCA-SVM-GA method are shown in figure 1. Thus, our model uses KPCA to extract main structure of raw data sets first for the further training of SVMs. KPCA can implicitly transform the data to a higher dimension nonlinear space and in exactly this space, we can employ linear PCA. It is a powerful tool to extract the feature and can reduce the training and testing time. We use GA to optimize parameters in SVM to avoid over fitting or under fitting problems caused by wrong parameters in SVM. The procedure of performing GA is introduced in Table 1. For the proposed SVM method of different kinds of faults, we use four SVMs for identification as shown in figure 2. The first one is to sort out fault 1 from other faults by labelling fault 1 as No.1 and other faults as No.2. The second is used to classify fault 2 from other faults by labelling fault 2 as No.1 and other faults as No.2. SVM3 aims to identify fault 6 out of fault 3 and fault 4, and SVM4 is used to distinguish fault 3 and...
fault 4. Thus, in the experiment of our proposed method, the training and testing time will subsequently shorten from fault 1, fault 2, fault 6, to fault 3 and fault 4. The training time for fault 3 and fault 4 will be the same because they are sorted by the very same SVM.

**Figure 1.** Procedures of the proposed method

**Figure 2.** Procedures of SVM classification

**Table 1.** Main body of GA

**Input:** Training and Testing Set of SVM, Set of parameters (PopSize, IterMax, ChromoLength, Possibility of Crossover: PC, Possibility of Mutation: PM)

**Output:** Optimized Parameters of SVM (OP)

1. $\text{PS} \leftarrow$ initialize the population with Popsize randomly with individual of length of ChromoLength
2. for $\text{iter} = 1$ to IterMax do
3.  $\text{OV} \leftarrow$ use the training and testing set to obtain accuracy of individuals of PS by dividing each chromosome into two parts, one for PC and one for PG
4.  $\text{FAS} \leftarrow$ calculate fitness value of PS based on OV
5.  $\text{PS}_{\text{new}} \leftarrow$ operate selection, crossover and mutation based on PS, PC, PM, FAS, then update PS with the solutions obtained by the operation
6.  $\text{OP} \leftarrow$ find the best fitness value and the corresponding individual, the first half of the best chromosome is parameter c, the second half is parameter g
7. end for
3. Experimental Results and Discussion

3.1. Experiment Description

Tennessee Eastman (TE) process is a commonly used benchmark data set for fault diagnosis to evaluate the results. The entire process is the production of two products from five reactants. The unit are composed of five operations, namely reactor, condenser, stripper, separator and compressor [12]. TE process consists of 52 variables and 21 types of faults. Downs and Vogel [13]-[14] made TE data set to contain one training set and one testing set for each type of fault respectively. Each training set is obtained by 25h of operations with a sampling time of 3min. One hour after the beginning, the fault is introduced. The first 20 normal samples are thrown away except for the normal training set, thus the training set for normal operation has 500 samples in 52 dimensions and each of other training sets for the 21 types of fault has 480 samples in 52 dimensions. Testing data sets contain 960 samples obtained by 48h operations with 160 normal samples and 800 fault samples.

We selected training and testing data from fault 1, fault 2, fault 3, fault 4 and fault 6 as samples, and compare our KPCA-SVM-GA model with PCA-SVM-GA model and single-SVM model in terms of fault identification accuracy, correlation coefficient, training time and testing time to evaluate our method. Fault identification accuracy (FIA) is the rate that a given fault sample is correctly sorted into its set:

\[
FIA = \frac{\text{Number of fault samples correctly sorted}}{\text{Total number of fault samples}} \times 100\%
\]  

Correlation coefficient (cc) measures the correlation between predicted results and real output. Cc ranges from -1 to 1, where 1 represents that predicted results are identical to real output and -1 represents that those predictions are carried out randomly.

\[
cc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}}
\]

where TP represents that the given fault sample is correctly sorted, FP represents that samples of other faults are sorted in the given fault set, FN indicates that the given fault sample is wrongly sorted into other set and TN represents samples of other faults are correctly sorted.

3.2. Experiments of KPCA-SVM-GA model

The following experiments aim to evaluate the effectiveness of our KPCA-SVM-GA model on TE process by comparing the proposed method with KPCA-SVM, PCA-SVM-GA, single-SVM model. We selected data from fault 1, fault 2, fault 3, fault 4, fault 6 to carry out the experiment. First, we use the training data sets of those five faults to form a new training set, and then use KPCA to extract the main structure of those data and reduce the dimension of the data set. The principle components are selected by calculating the accumulate contribution of the variance and the threshold is set to be 90%, thus

\[
\sum_{i=1}^{p} \lambda_i \geq 90\%
\]

where \(n\) is the dimension of the original data set and \(p\) is the dimension after extraction of the main structure. Second, we use GA to optimize the coefficients of SVM (c and g) on behalf of the sorted accuracy of the training data set. Third, we use the obtained training data set and optimized coefficients to train the SVM model and identify each testing samples of those five kinds of faults and record the indicators introduced above.

The kernel function in KPCA is chosen to be radial basis function and \(\sigma\) is decided to be 40000 based on experience. In GA model, from many experiments, the parameters are chosen as follows: population size 50, chromosome length 20 (10 for c and 10 for g), possibility of crossover 0.9, possibility of mutation 0.01, iteration 100 times. After 10 experiments, we obtained \(c = 8.011, g = 0.198\) for SVM model in KPCA-SVM-GA. Parameters in other SVM models are selected randomly. The experiment result is shown in Table 2.
As is shown in Table 2, SVM model trained after the extraction of main structure on TE process has obviously higher fault identification accuracy and lower training and testing time than those of single SVM model, because the dimension reduction can provide more useful information of the given data set while reducing the number of samples for training and testing. We can also see that feature extraction using KPCA is more efficient than using PCA, because KPCA can implicitly analyze the given data set in a higher dimension feature space by using kernel function which means it can extract nonlinear information of the data set. However, PCA is only a linear method for feature extraction, thus KPCA has a better performance on fault identification accuracy than PCA. It is also shown in Table 2 that the fault identification accuracy and cc of fault 4 is much lower than that of fault 1. Fault 1 is caused by a sudden increase of A/C feed and it is obvious enough for fault identification. Fault 4 is caused by a step increase of cooling inlet temperature, but under the close loop control of the system, the changes of all variables remain in the range of 2%. Thus, it is more difficult to identify fault 4 than other faults.

This experiment shows that KPCA can effectively reduce the training and testing time by feature extraction compared to other methods and can improve the identification accuracy by implicitly analyse nonlinear information of the given data set. Thus, this kind of KPCA-SVM-GA method is reliable for fault identification of TE process.

4. Conclusion

In this paper, we presented a KPCA-SVM-GA model for fault identification and evaluate our model on TE process. The result of experiments on five given faults in TE process has shown that KPCA-SVM-GA model is reliable for fault identification. The performance of SVM trained by data after feature extraction is better than SVM with original data, and the performance of KPCA using RBF for feature extraction is better than SVM with PCA. KPCA is applied to reduce the dimension and extract nonlinear structure of the given data set. It is proved that the reduction of dimension by RBF kernel function can obviously reduce the training and testing time in the subsequent sorting carried out by SVM models. Besides, feature extraction by kernel function can also improve the performance of fault identification. SVM is proposed to classify the different kinds of faults and GA is introduced to optimize parameters of SVM to improve the performance of identification and avoid the over fitting or under fitting problems of SVM caused by incorrect parameters.
As for the future work, we plan to carry out research on the optimization of kernel function used in KPCA to extract main structure more efficiently, and finding a new reliable way to optimize parameters in SVM.

References
[1] Z. Gao, C. Cecati and Steven X. Ding 2015 A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis with Model-Based and Signal-Based Approaches IEEE Trans. Ind. Electron 6 p. 3757
[2] Z. Gao, C. Cecati and Steven X. Ding 2015 A survey of fault diagnosis and fault-tolerant techniques Part II: Fault diagnosis with knowledge-based and hybrid/active approaches IEEE Trans. Ind. Electron 6 p. 3757
[3] Nonlinear Component Analysis as A Kernel Eigenvalue Problem 1998 United States North America
[4] Cho, J.H. Lee, J.M., Choi, S.W., Lee, D., Lee, I.B. 2005 Fault identification for process monitoring using kernel principal component analysis Chemical Engineering Science 60 pp. 279-88
[5] Mohamed A. Bin Shams 2011 Fault Identification using Kernel Principle Component Analysis Proceedings of the 18th World Congress The International Federation of Automatic Control 4320-25
[6] L. Khan, M. Awad, B. Thuraisingham 2007 A new intrusion detection system using support vector machines and hierarchical clustering Int. J. Very Data Bases 16 507–21
[7] A. Majid, A. Khan, A.M. Mirza 2006 Combining support vector machines using genetic programming Int. J. Hybrid Intell. Syst. 3 (2) 109–25
[8] Taeshik Shon, Yongdae Kim, Cheolwon Lee and Jongsub Moon 2005 A machine learning framework for network anomaly detection using SVM and GA Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop pp. 176-83
[9] S.J. Horng, M.Y. Su, Y.H. Chen, T.W. Kao, et al. 2011 A novel intrusion detection system based on hierarchical clustering and support vector machines Expert Syst. Appl. 38 306–13
[10] E. Eskin 2000 Anomaly detection over noisy data using learned probability distributions, in: Proceedings of the International Conference on Machine Learning pp. 255-62
[11] C. Cortes and V. Vapnik 1995 Support vector networks M. Learning 20 pp. 273-97
[12] X.B. He, Y.P. Yang, Y.H. Yang 2008 Fault diagnosis based on variable-weighted kernel Fisher discriminant analysis Chemom. Intell. Lab. Syst. 93 (1) pp. 27–33.
[13] J.J. Downs, E.F. Vogel 1993 A plant-wide industrial process control problem Compute. Chem. Eng. 17 (3) pp. 245–55
[14] P.R. Lyman, C. Georgakis 1995 Plant-wide control of the Tennessee Eastman problem Compute. Chem. Eng. 19 (3) pp. 321–31