Neurocomputing 174 (2016) 906–911

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process

Xin Gao, Jian Hou *

College of Engineering, Bohai University, Jinzhou 121013, China

ARTICLE INFO

Article history:
Received 1 August 2015
Received in revised form
22 September 2015
Accepted 5 October 2015
Communicated by Yang Tang
Available online 23 October 2015

Keywords:
Support vector machine (SVM)
Fault diagnosis
Tennessee Eastman (TE) process
Principal component analysis (PCA)
Gridsearch (GS)
Genetic algorithm (GA)

ABSTRACT

In modern industry, fault diagnosis and process supervision are very important in detecting machinery failures and keeping the stability of production systems. In this paper, a multi-class support vector machine (SVM) based process supervision and fault diagnosis scheme is proposed to predict the status of the Tennessee Eastman (TE) Process. After preprocessing the collected data, principal component analysis (PCA) is firstly used to reduce the feature dimension. Then, to increase prediction accuracy and reduce computation load, the optimization of SVM parameters is accomplished with the grid search (GS) method, which generates comparable classification accuracy to genetic algorithm (GA) and particle swarm optimization (PSO) while being more efficient than the latter two algorithms. Finally, to demonstrate the effectiveness of the proposed SVM integrated GS-PCA fault diagnosis approach, a comparison is made with other related fault diagnosis methods.

1. Introduction

Nowadays, the modern industrial process is usually complicated, and various uncertainty and nonlinearity are involved. In this case, a seemingly small fault may cause unimaginable consequences. Therefore, the importance of fault detection and diagnosis in the industry can not be ignored any more. As fault diagnosis is a potential approach to improve productivity, increase production process utilization and reduce the maintenance costs, it has attracted much attention in recent years. Some of the newly proposed fault detection and diagnosis methods can be found in [1–6]. Most of the existing fault diagnosis schemes rely on the availability of process analytical models and only a few papers, e.g., [1,7,8], are devoted to data-driven fault detection and isolation approaches. Data driven techniques can be quite effective in the case that the analytical process models are not available. Especially, for the large-scale plants, whose physical models are generally difficult to be established, data driven approaches offer an effective alternative solution for process monitoring. Therefore, the production process monitoring and fault diagnosis schemes with data based fault identifying techniques have prospered and begun to be applied in the industry to protect the production system from disastrous accidents and keep stability [9,34].

With the availability of statistical learning theory [10–13] and convenient process signals collecting techniques, data based techniques have been developed quickly, and some works in this area include [8,14,15,7]. Furthermore, machine learning techniques have been applied widely in many domains ranging from image processing, face and object categorization, information and image retrieval, etc. Several machine learning algorithms, e.g., artificial neural networks (ANN), principal component analysis (PCA), fuzzy expert system, lazy learning, random forest, have been applied in industrial production systems to monitor machine condition and detect process faults [16,17]. For example, PCA was used to analyze product quality for a pilot plant [18] and in [19] a key performance indicator prediction scheme was designed to be used in a hot strip mill. However, Support vector machine (SVM) is rarely used in machine condition monitoring and fault diagnosis. Sometimes, it is used in combination with some other technologies, e.g., wavelet package transform, genetic algorithm (GA) and fractal dimension [20–22]. And it is usually expertise-oriented and problem-oriented [23].

In machine learning, SVM is an important classification approach. SVM is designed based on the Vapnik–Chervonenkis theory [24,25] and exhibits superior generalization ability in practice. It is able to handle the classification problem with finite (or small) samples as well as large feature space [23,26]. Since in the manufacture field the number of faulty samples is relatively small and signal data can be easily collected, SVM has been introduced to conduct the machine condition monitoring and fault diagnosis.

* Corresponding author.
E-mail addresses: xin.gao1991@qq.com (X. Gao), dr.houjian@gmail.com (J. Hou).

http://dx.doi.org/10.1016/j.neucom.2015.10.018
0925-2312/

In fault diagnosis, data analysis technique is one of the most preferable and reliable methods. Generally, a diagnosis scheme is composed of three stages. The first stage is signal data acquisition with appropriate sensors. Then data processing including normalization, dimension reduction and statistical feature extraction are executed in the next step. Some feature selection techniques including Genetic algorithms (GAs), PCA and Particle swam optimization technique (PSO), are commonly used to reduce the dimension of the recorded data set as well as save computation load [9]. In the final step, the operation status of the monitored process is predicted by maintenance personnel manually or by some automated computational intelligence methods such as ANN and SVM. SVM is preferred here as it is a representative nonlinear classifier and possesses superior generalization ability [27]. These nice properties make it extensively used in the field of classification analysis, such as process monitoring and fault diagnosis [28]. In this paper, after methods comparison, an improved SVM based industrial process faults diagnosis scheme is proposed. Fault diagnosis experiments are conducted with the TE process model, and the results show that the proposed SVM integrated GS-PCA fault diagnosis scheme performs better in both classification accuracy and computation efficiency.

The remainder of this paper is organized as follows. The SVM classification algorithm, the PCA algorithm, and three optimization algorithms, i.e., GA, PSO and GS, are briefly introduced in the next section. In Section 3, simulations are conducted to predict the status of the Tennessee Eastman process and the results validate the effectiveness of the proposed approach. Finally, Section 4 summarizes the conclusions.

2. Methods

2.1. Support vector machine (SVM)

SVM is a popular classification technique widely used in various fields. In the following we briefly introduce the theoretical foundation of SVM. Generally speaking, a classification task involves training data and testing data. The training data is used to train a prediction model, which is then used to predict the labels of testing data. In industrial applications, the collected historical data usually serve as the training data. Every sample in the training set consists of two parts, i.e., the observed variables (called attributes) and the corresponding class labels. Since this paper focuses on the industrial fault diagnosis, the class labels here refer to the statuses of the industry process, i.e., normal status or the faulty statuses (needed to be specific to the serial number of the faults). Then, the training data are used to built a prediction model. The testing data generally contain only the attributes. The missing class labels will be predicted by feeding the testing data to the prediction model. In this way, statuses of the process can be identified based on the collected data.

The key of SVM is to find out linear hyperplanes to separate the data in different classes with maximum margins. Taking the binary classification problem for example, the distance between two classes of training data should be as large as possible. The separating hyperplane should have maximal distances with the closest data points belonging to the both sides and it is located in the middle of the margin. The nearest points to the hyperplane are called support vectors (SVs), and they are representative data points because they contain almost all the information needed to determine the classifier.

In training set, an instance is denoted as $(x_i, y_i), i = 1, \ldots, m$ where $x_i \in R^n$ and $y \in \{1, -1\}$. m is the number of the observed samples while n represents the number of attributes. The hyperplane is denoted by $w^T x_i + b = 0$, and the SVM classification problem is transformed to finding out optimal parameters w^T and b to maximize the distance between hyperplanes $w^T x_i + b = 1$ and $w^T x_i + b = -1$. According to [25], the problem translates to the following optimization problem:

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i \tag{1}$$

$$s.t.\ y_i(w^T x_i + b) \geq 1 - \xi_i, \tag{2}$$

$$\xi_i \geq 0, \tag{3}$$

In Eq. (1), the parameter C is a user-specified penalty parameter of error term. The error term is represented by the parameter ξ_i.

In some cases, the original input data can not be linearly separated. Then the mapping function ϕ is introduced to map the input data into a new feature space with higher dimension, where the input data become linearly separable. Using the corresponding mapping function, the kernel function is denoted as

$$K(x_i, x_j) = <\phi(x_i), \phi(x_j)> \tag{4}$$

Usually the kernel function is used more extensively than the mapping function, as it is less expensive to calculate. It enables the optimization problem Eq. (1) to be solved in the input space itself instead of the higher dimensional mapping space [25]. Table 1 presents four commonly used kernels.

In this paper, we adopt the Gaussian radial basis function (RBF) in the SVM classifier. Generally speaking, the RBF kernel function is usually the first option, especially when the data set has large numbers of instances. The main reason is as follows. Firstly, by adjusting the parameters, the RBF kernel can perform similarly with sigmoid kernel and linear kernel. Secondly, the RBF kernel involves only penalty error parameter C and kernel parameter γ, and has relatively small numerical difficulty [29].

2.2. Dimension reduction with PCA

It is very important to scale the attribute values in the training data and testing data into a small range before they are put into the SVM prediction model. On one hand, it can avoid the large values dominating the small ones. On the other hand, scaling attribute values into a small range will avoid dimension disaster and decreasing calculation load [35]. In this paper, each attribute value will be linearly scaled to the range between 0 and 1 to avoid the numerical difficulty caused by large attribute values. What is more, the training set and testing set are scaled at the same time.

As a commonly used dimension reduction algorithm, PCA extracts a set of uncorrelated variables and store them in features of smaller dimensions. Due to its efficiency and simplicity in dealing with the data sets with high dimensions, PCA has been widely applied in various fields in practice, especially for data compression. The procedure of standard PCA approach to reduce the dimension of data is briefly presented as follows.

The original data set with N samples denoted as $Z^T = [z_1, \ldots, z_N] \in R^{m \times N}$, where m is the number of measurement signals in each sample. Then, singular value decomposition (SVD) is applied to the covariance estimation matrix of the observed data

Table 1
Formulation of kernel functions.

linear	$rx_i^T \cdot x_j$
polynomial	$(\gamma x_i^T \cdot x_j + r)^d$
radial basis function (RBF)	$e^{-\gamma\|x_i - x_j\|^2}$
sigmoid	$\tanh(\gamma x_i^T \cdot x_j + r)$

γ, rand d are parameters.

set Z^T.

$$\Sigma = \frac{1}{N-1} Z^T Z = P \Lambda P^T, \tag{5}$$

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m), \lambda_1 \geq \cdots \geq \lambda_m > 0, \tag{6}$$

$$\xi = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{m} \lambda_j}. \tag{7}$$

where $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of the covariance matrix Σ. Then the top k largest eigenvalues are selected and their accumulated contribution can be calculated using the certain criteria in Eq. (7). In this paper, the number of principal components (PCs) k is determined to ensure that the accumulated contribution is just over 90%. Then, Λ, P are divided into two parts as

$$\Lambda = \begin{bmatrix} \Lambda pc & 0 \\ 0 & \Lambda res \end{bmatrix}, \quad \Lambda_{pc} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k), \tag{8}$$

$$\Lambda_{res} = \mathrm{diag}(\lambda_{k+1}, \ldots, \lambda_m), \tag{9}$$

$$P = [P_{pc}, P_{res}], \quad P_{pc} \in R^{m \times k}, \quad P_{res} = R^{m \times (m-k)}. \tag{10}$$

Finally, the original data matrix Z^T is decomposed as the sum of $Z_{pc}{}^T$ and E_p, where $Z_{pc}{}^T$ is the dimension-reduced data set and E_p is the residual matrix [1].

$$Z^T = Z_{pc}^T + E_p, \tag{11}$$

$$Z_{pc}^T = P_{pc}^T * Z^T. \tag{12}$$

2.3. Cross-validation and GS

For SVM based classifier with RBF kernel function, there are two parameters, i.e., C and γ. These two parameters have significant influence on the prediction accuracy of the classifier. Therefore, some algorithms should be applied to identify the optimal parameters.

Cross-validation is one of the commonly used validating methods validation methods to evaluate the accuracy of a classifier [36]. Taking k-fold cross-validation for example, it refers to the procedure that at first the original training set is separated into k equal-size subsets, and then each subset is tested while the other k–1 subsets serve as training set. In this way, each sample in the original training set will be tested, and the percentage of correctly predicted samples is regarded as the accuracy of the classifier.

Grid search (GS) is a simple and straightforward method to find out the best parameter values for the SVM classifier. Since the two parameters C and γ are independent, the GS process can be conducted in parallel. Specifically, a set of candidates are firstly selected for both γ and C. Then each pair of γ and C is evaluated by cross-validation and the pair with the highest accuracy is determine as the optimal parameters. In this paper, based on the experiments of previous researchers [29], the two candidates of C and γ are selected in a exponentially increasing way.

2.4. Genetic algorithm (GA)

Genetic algorithms (GAs) are a series of evolutionary computation techniques and have been applied in several domains, e.g., neural networks, optimization design, and expert systems [30]. In dealing a specific problem, the GA firstly encodes the potential solutions into individual chromosomes. Then these chromosomes constitute a search space, where a set of chromosomes are randomly chosen to form the initial population. The fitness value of each chromosome is estimated by means of an objective function and the chromosomes with higher fitness values are regarded as better ones. The genetic algorithm operates selection, mutation and crossover to produce a new generation of chromosomes whose qualities (fitness values) are expected to be greater than their elder generation. When the termination criterion is satisfied, the evolutionary process is terminated and the chromosome with the largest fitness in the latest generation is selected as the optimal solution. The flow chart of GA is shown in Fig. 1.

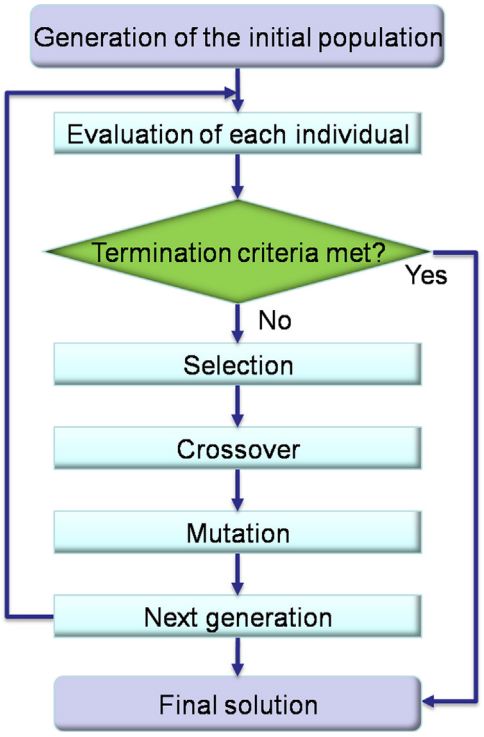

Fig. 1. The flow chart of GA.

2.5. Particle swarm optimization (PSO)

Particle Swarm Optimization (PSO) is inspired by the random actions performed collectively by a group of birds to seek food. PSO is an optimizer and the best solution is found out by imitating the swarm seeking the expected best particle. Since PSO has no crossover and mutation, it can be implemented more easily, intelligently, and faster. PSO is initialized with m particles randomly chosen from the D-dimension search space. $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD})(i = 1, \ldots, m)$ denotes the location of the ith particle. The fitness value of a particle is estimated by an objective function, and particles with higher fitness are regarded as better ones. The best position of the ith particle is denoted as $P_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$, and the best position in all m particles is denoted as $P_g = (p_{g1}, p_{g2}, \ldots, p_{gD})$. The pace of the ith particle moving to another position is symbolized as $V_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$. The particles are moved according to the following equations:

$$V_i(k+1) = wV_i(k) + c_1 r_1 (P_i - X_i(k)) + c_2 r_2 (P_g - X_i(k)) \tag{13}$$

$$X_i(k+1) = X_i(k) + V_i(k+1). \tag{14}$$

where k is the number of iteration, and w is inertia weight. If w is chosen appropriately, the number of iteration required can be small. $c1$ and $c2$ are acceleration constants. $r1$ and $r2$ are defined randomly between 0 and 1. $[V_{\min}, V_{\max}]$ are designated vectors, and V_i is restricted between them. According to Eq. (14) the new position of ith particle can be calculated. If the fitness value of P_g reaches the designated value or the number of iterations reaches

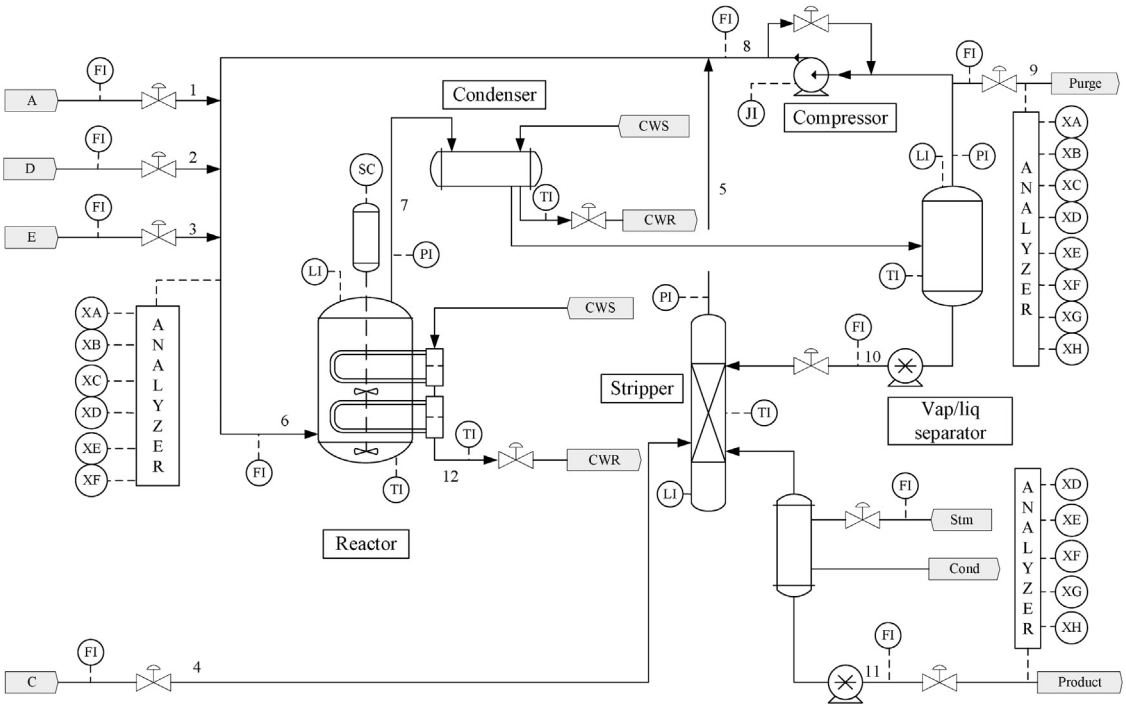

Fig. 2. A process flowsheet of the TEP.

the allowed maximum number, the process of iteration will be terminated and the optimal solution is found.

3. Case study

3.1. Tennessee Eastman process

The Tennessee Eastman process benchmark is a plant-wide simulation platform built on the base of a practical chemical process [31]. As shown in Fig. 2, this process benchmark is composed of 5 major parts, i.e., a reactor, a condenser, a vapor-liquid separator, a recycle compressor, and a product stripper. Besides, it contains a total of 8 components, including 2 products, a by-product, 4 inputs, and an inert, which are denoted by A, B, C, D, E, F, G, and H. More details of this benchmark can be found in [1,32]. This process is suitable to evaluate process monitoring schemes and control strategies. It is programmed to simulate the normal production process as well as 21 different faulty processes caused by inputting some known disturbances. The details of the faults used in this paper are presented in Table 2. In this simulation process, there are 52 monitoring variables, including 11 manipulated variables and 41 measured variables. Once a faulty condition is taken on, all variables in the process will be effected and some values may change. Besides, all instances are sampled and recorded at the interval of 3 min. The data set used in this paper is the same as the one given in [33], which is widely used in process monitoring and fault diagnosis field. In the training set, as for the 22 different operating conditions, each faulty condition has 480 observation samples while the normal set has 500 observation samples. On the other hand, the testing data sets are obtained after 48 h continuous operation, so for 22 kinds of condition each testing data set has 960 faulty samples. Since the faults are introduced into the simulation process after 8 h normal operation, in each faulty testing data, the first 160 samples are in normal status [1].

Table 2
Descriptions of the faults used in this paper.

Fault number	Fault description
IDV0	Normal
IDV1	A/C feed ratio, B composition constant
IDV2	B composition, A/C ratio constant
IDV4	Reactor cooling water inlet temperature
IDV5	Condenser cooling water inlet temperature

3.2. Simulation results and discussion

In this section, the collected data sets from TE process model are used to evaluate the performance of fault diagnosis using SVM classifiers. Since the parameters involved in SVM classifiers have significant influence on the prediction accuracy, we need a method to determine the appropriate values of these parameters. As we have introduce three optimization methods for this purpose, i.e., GS, GA and PSO, which are really commonly used in recent studies, in the following we compare these methods and use the best-performing one to identify the optimal parameters for SVM classifiers. The training data considered in this paper contains the training data sets under normal condition, fault 1 and fault 2. And testing data from fault 1 serve as the testing data. All the data are scaled to zero mean and standard deviation one, and the feature dimension are reduced by PCA. In this paper, prediction accuracy are evaluated by

$$\text{Accuracy} = \frac{\text{the quantity of testing data correctly predicated}}{\text{the quantity of all testing data}} \times 100\% \quad (15)$$

The prediction results with the three optimization methods to choose parameters for SVM classifiers are presented in Table 3.

In order to report the results in more details, the prediction results with the three parameters optimization procedures are shown in Fig. 3. The horizontal axis (x-axis) is on behalf of the 960 testing samples, in which the first 160 instances are under normal condition and the remaining 800 instances are sampled after

Table 3
The detailed information with three optimization methods.

The optimization method	Average accuracy (%)	Average computing time (sec)
GS	99.0583	116.52
GA	99.2708	2063.60
PSO	99.0625	908.45

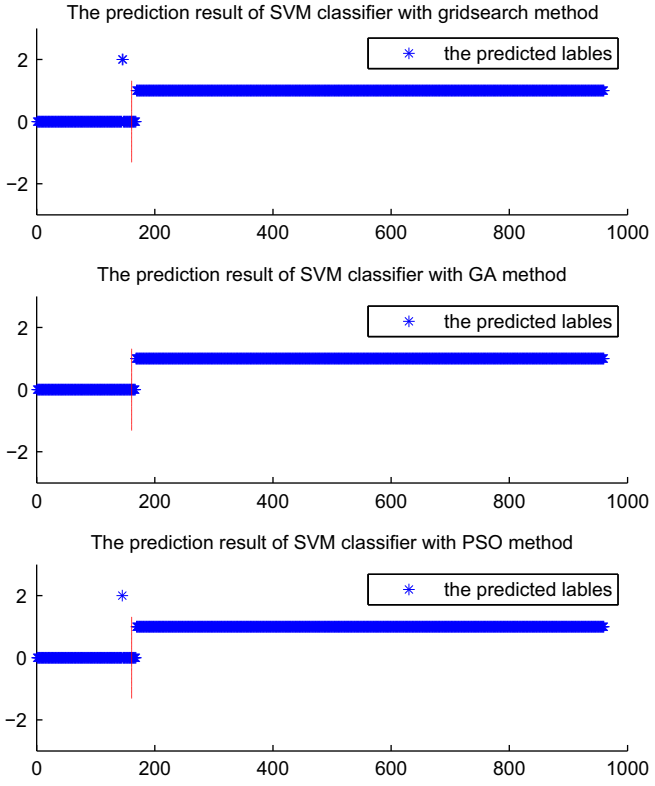

Fig. 3. The prediction results of SVM classifier on the testing data of Fault 1 with three different parameters optimization procedures. The top plot is the classification result using GS technique, the middle one is with GA algorithm and the last plot uses PSO algorithm. The vertical lines in the 3 plots represent the moment that Fault 1 is introduced. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

faults are added into the TE process. The vertical axis (y-axis) displays the predicted class label of each sample. The samples which are predicted to be from formal condition are labeled as '0', while the samples predicted from faulty condition are labeled as '1'. Fig. 3 shows that the SVM based classifiers have high prediction accuracy and the predicted results are correct in most cases.

From Table 3 and Fig. 3, it can be seen that the SVM classifiers with optimal hyper parameters could classify the testing data accurately. The prediction accuracy with the three parameter optimization methods are quite similar to each other. However, the one with the GS method costs the least time. Therefore, in the next part, the GS method is chosen to optimize the parameters of the SVM classifiers.

In this section, the proposed SVM integrated GS-PCA fault diagnosis scheme is tested with the TE process model. Firstly, all the data sets (both training sets and testing sets) from the TE process are scaled to zero mean and standard deviation one and then feature dimension reduction is conducted by PCA. Next, the training sets are used with the GS method to identify the optimal hyper parameters for SVM classifier to enhance its predictive ability. Finally, a testing data set is fed into the SVM classifier to

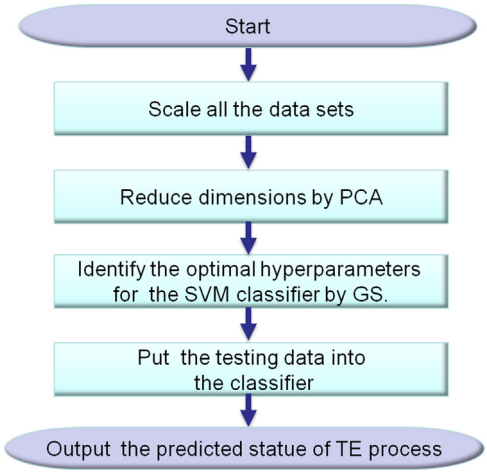

Fig. 4. The computation steps of the proposed SVM integrated GS-PCA fault diagnosis scheme.

Table 4
Detailed information of the classifiers

Classifier	Data processing	Parameters optimized
1	–	–
2	√	–
3	–	√
4	√	√

Data processing refers that data set are scaled to zero mean and standard deviation one and also be dimensionally reduced by PCA. Parameters optimized means the optimal hyper parameters for SVM classifier are identified with the training sets using the GS method.

predict the statue of TE process. In this part, 5 types of training sets, i.e., normal station, IDV1, IDV2, IDV4, and IDV5, are used to build the classifier model and the testing data are of fault 1. The procedure of the SVM based fault diagnosis process is shown in Fig. 4. In order to illustrate the superiority of this SVM integrated GS-PCA fault diagnosis scheme, some other schemes are also tested in fault diagnosis of the TE process for comparison, and the results are reported in Table 5.

In Table 5, Average Running Time refers to the prediction time of classifiers excluding the parameters optimization time. Classifier 1, 2, 3 and 4 are built based on different predictive schemes. The detailed information of these 4 classifiers is presented in Table 4. Classifier 1 is built purely based on SVM algorithm without parameters optimization or data handling, and classifier 2 is based on SVM algorithm with data processing but without parameters optimization, while classifier 3 is with parameters optimized without data processing. In the last, classifier 4 is with tuned parameters and data handled based on the improved SVM integrated GS-PCA fault diagnosis approach.

From Table 5 it is evident that SVM based classifiers has strong predictive abilities in fault prediction on the TE process benchmark. Besides, the proposed diagnosis scheme obtained the highest prediction accuracy with the least running time. There are two facts worth mentioning. First, SVM hyper parameters are optimized by the GS algorithm. Therefore, the predictive ability of the classifier is enhanced evidently. Besides, the training data and testing data have been preprocessed, and feature dimension is reduced using the PCA algorithm. These processing saves time in identifying faults and makes it possible to take measures in time. This merit will be more obvious if the scheme is used with large amounts of data. In conclusion, the proposed SVM scheme is suitable to be applied in the industrial process to diagnose faults.

Table 5
Performance comparison of classifiers

Classifier	Average accuracy (%)	Average running time (sec)
1	91.87	3.88
2	96.45	2.84
3	91.87	3.28
4	96.77	1.35

4. Conclusion

In this paper we present an improved SVM based fault diagnosis method. The original feature data are firstly scaled and normalized, and then PCA is applied to reduce the feature dimension. In the next step, we compare the performance of GS, GA and PSO in optimizing the SVM parameters. As a result, we find that GS leads to comparable classification accuracy to the other two methods while being more efficient than the latter. In simulations with the Tennessee Eastman Process, we compare our GS-PCA based SVM with other SVM based fault diagnosis methods. Simulation results indicate that our method shows advantage over others in both classification accuracy and computation efficiency.

References

[1] S. Yin, S.X. Ding, A. Haghani, H. Hao, P. Zhang, A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process, J. Process Control 22 (9) (2012) 1567–1581.
[2] L. Zhang, E.K. Boukas, L. Baron, H.R. Karimi, Fault detection for discrete-time Markov jump linear systems with partially known transition probabilities, Int. J. Control 83 (8) (2010) 1564–1572.
[3] H. Li, H. Gao, P. Shi, X. Zhao, Fault-tolerant control of Markovian jump stochastic systems via the augmented sliding mode observer approach, Automatica 50 (7) (2014) 1825–1834.
[4] S. Laghrouche, J. Liu, F.S. Ahmed, M. Harmouche, M. Wack, Adaptive second-order sliding mode observer-based fault reconstruction for PEM fuel cell air-feed system, IEEE Trans. Control Syst. Technol. 23 (2015) 1098–1109.
[5] W. Sun, H. Pan, J. Yu, H. Gao, Reliability control for uncertain half-car active suspension systems with possible actuator faults, IET Control Theory Appl. 8 (9) (2014) 746–754.
[6] S. Yin, H. Luo, S.X. Ding, Real-time implementation of fault-tolerant control systems with performance optimization, IEEE Trans. Ind. Electron. 61 (5) (2014) 2402–2411.
[7] S. Yin, G. Wang, H.R. Karimi, Data-driven design of robust fault detection system for wind turbines, Mechatronics 24 (4) (2014) 298–306.
[8] Z. Wu, P. Shi, H. Su, J. Chu, Exponential stabilization for sampled-data neural-network-based control systems, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 2180–2190.
[9] C. Rajeswari, B. Sathiyabhama, S. Devendiran, K. Manivannan, Bearing fault diagnosis using wavelet packet transform, hybrid PSO and support vector machine, Procedia Eng. 97 (2014) 1772–1783.
[10] J. Liu, Observer-based higher order sliding mode control of power factor in three-phase AC/DC converter for hybrid electric vehicle applications, Int. J. Control 87 (6) (2014) 1117–1130.
[11] J. Liu, S. Laghrouche, M. Harmouche, M. Wack, Adaptive-gain second order sliding mode observer design for switching power converters, Control Eng. Pract. 30 (9) (2013) 124–131.
[12] X. Yang, H. Gao, Multiple model approach to linear parameter varying time-delay system identification with EM algorithm, J. Frankl. Inst. 351 (2) (2014) 5565–5581.
[13] X. Yang, Y. Lu, Z. Yan, Robust global identification of linear parameter varying systems with generalised expectation–maximisation algorithm, IET Control Theory Appl. 9 (7) (2015) 1103–1110.
[14] Z. Wu, P. Shi, H. Su, J. Chu, Sampled-data exponential synchronization of complex dynamical networks with time-varying coupling delay, IEEE Trans. Neural Netw. Learn. Syst. 24 (8) (2013) 1177–1187.
[15] J. Qiu, G. Feng, H. Gao, Observer-based piecewise affine output feedback controller synthesis of continuous-time t-s fuzzy affine dynamic systems using quantized measurements, IEEE Trans. Fuzzy Syst. 20 (6) (2012) 1046–1062.
[16] A. Kulkarni, V.K. Jayaraman, B.D. Kulkarni, Support vector classification with parameter tuning assisted by agent-based technique, Comput. Chem. Eng. 28 (3) (2004) 311–318.
[17] L.H. Chiang, M.E. Kotanchek, A.K. Kordon, Fault diagnosis based on fisher discriminant analysis and support vector machines, Comput. Chem. Eng. 28 (8) (2004) 1389–1401.
[18] Y. Zhang, Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM, Chem. Eng. Sci. 64 (5) (2009) 801–811.
[19] S.X. Ding, S. Yin, K. Peng, H. Hao, B. Shen, A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip mill, IEEE Trans. Ind. Inform. 9 (4) (2013) 2239–2247.
[20] Q. Hu, Z. He, Z. Zhang, Y. Zi, Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble, Mech. Syst. Signal Process. 21 (2) (2007) 688–705.
[21] S. Fei, X. Zhang, Fault diagnosis of power transformer based on support vector machine with genetic algorithm, Expert Syst. Appl. 36 (8) (2009) 11352–11357.
[22] J. Yang, Y. Zhang, Y. Zhu, Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension, Mech. Syst. Signal Process. 21 (5) (2007) 2012–2024.
[23] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Signal Process. 21 (6) (2007) 2560–2574.
[24] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM, New York, 1992, pp. 144–152.
[25] S.R. Gunn, Support vector machines for classification and regression, ISIS Technical Report, 14, 1998.
[26] S. Mahadevan, S.L. Shah, Fault detection and diagnosis in process data using one-class support vector machines, J. Process Control 19 (10) (2009) 1627–1639.
[27] B. Scholkopf, A.J. Smola, Learning with Kernels, Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, New York, 2002.
[28] Y. Liu, C. Xu, J. Shi, Tennessee Eastman process monitoring based on support vector data description, In: 2012 Second International Conference on Intelligent System Design and Engineering Application (ISDEA), IEEE, Sanya, 2012, pp. 553–555.
[29] C.J. Lin, A practical guide to support vector classification, Department of Computer Science & Information Engineering National Taiwan University 67 (5) (2003) 1–14.
[30] K.F. Man, K.S. Tang, S. Kwong, Genetic Algorithms: Concepts and Designs, Avec disquette, 1, Springer Science & Business Media, London, 1999.
[31] P.R. Lyman, C. Georgakis, Plant-wide control of the Tennessee Eastman problem, Comput. Chem. Eng. 19 (3) (1995) 321–331.
[32] J.J. Downs, E.F. Vogel, A plant-wide industrial-process control problem, Comput. Chem. Eng. 17 (1993) 245–255.
[33] L.H. Chiang, R.D. Braatz, E.L. Russell, Fault Detection and Diagnosis in Industrial Systems, Springer Science & Business Media, London, 2001.
[34] X. Xie, S. Yin, H. Gao, O. Kaynak, Asymptotic stability and stabilisation of uncertain delta operator systems with time-varying delays, IET Control Theory & Applications 7 (8) (2013) 1071–1078.
[35] G. Wang, S. Yin, Quality-related fault detection approach based on orthogonal signal correction and modified PLS, IEEE Transactions on Industrial Informatics 11 (2) (2015) 398–405.
[36] S. Yin, X. Zhu, O. Kaynak, Improved PLS focused on key performance indictor related fault diagnosis, IEEE Transactions on Industrial Electronics 62 (3) (2015) 1651–1658.

Xin Gao received the B.E. degree in automation from North China Electricity Power University, Baoding, China, in 2013. Now, she is pursuing the M.E. degree in control theory and control engineering in Bohai University. Her research interests include data-driven fault diagnosis and their application to large-scale industrial processes.

Jian Hou received his Ph.D. degree in 2007 from Harbin Institute of Technology, China. From 2007 to 2010, he worked as a postdoctoral researcher in National University of Singapore, Singapore and University of Venice, Italy. He joined Bohai University, China, in 2012, where he is currently an associate professor. His research interests include pattern recognition, machine learning and computer vision.