Neurocomputing 167 (2015) 260–279

Contents lists available at ScienceDirect

Neurocomputing

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

Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization

XiaoLi Zhang [a,b], Wei Chen [c], BaoJian Wang [a], XueFeng Chen [a,*]

[a] State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, PR China
[b] Key Laboratory of Road Construction Technology and Equipment, Ministry of Education, Chang'an University, Xi'an 710064, PR China
[c] Air Force Engineering University, Xi'an 710038, PR China

ARTICLE INFO

Article history:
Received 9 October 2014
Received in revised form
12 April 2015
Accepted 27 April 2015
Communicated by Hongli Dong
Available online 7 May 2015

Keywords:
Intelligent fault diagnosis
Support vector machine
Ant colony algorithm
Rotating machinery
Rotor
Roller bearing

ABSTRACT

The failure of rotating machinery can result in fatal damage and economic loss since rotating machinery plays an important role in the modern manufacturing industry. The development of a reliable and efficient intelligent fault diagnosis approach is an ongoing attempt. Support vector machine (SVM) is a widely used machine learning method in intelligent fault diagnosis. But finding out good features that can discriminate different fault conditions and optimizing parameters for support vector machine can be regarded as the most two important problems that can highly affect the final diagnosis accuracy of support vector machine. Until now, the two issues of feature selection and parameter optimization are usually treated separately, weakening the effects of both efforts. Therefore, an ant colony algorithm for synchronous feature selection and parameter optimization for support vector machine in intelligent fault diagnosis of rotating machinery is presented. Comparing with other methods, the advantages of the proposed method are evaluated on an experiment of rotor system and an engineering application of locomotive roller bearings, which proves it can attain much better results.

1. Introduction

Rotating machinery covers a broad range of machines and plays an important role in industry. Effectively diagnosing the incipient faults of rotating machinery as early as possible can improve the reliability of the machinery. More importantly, it can effectively avoid catastrophic damage and huge economic loss. Therefore, the fault diagnosis of rotating machinery is of great significance in industry and has attracted more and more attention. Roller bearings [1], gears [2], rotors [3], turbo pumps [4], power generators [5] and other rotating machinery are being researched. Among them, the vibration based analysis method is the most commonly used fault diagnosis technique [6], which is feasible to detect the vibration signal changes caused by fault components with time-domain analysis method, frequency-domain analysis method, time-frequency analysis method and so on. However, it is usually very difficult to extract the fault features of the measured vibration signals in the case of much noise and complicated multiple faults with different types and severity. Moreover, these traditional methods usually require professional skills and rich experience, which are hard for non-experts to use. Consequently, there is a great demand for intelligent fault diagnosis technique that can produce a reliable, fast and automated recognition and prediction results. With the development of machinery fault diagnosis aiming at the intelligent and automated direction, it has been undergoing a transformation from a manual monitoring approach based on experts to an intelligent one that are capable of learning from the cause-symptom relationship and then make machinery health status decision automatically [7]. The intelligent fault diagnosis problem is treated as a pattern recognition/classification problem based on training fault patterns from samples. Until present, various intelligent fault diagnosis methods such as expert system [8,9], neural network [10,11], fuzzy logic [12,13], rough set [14,15], and their hybrid method [7,16,17] have been successfully applied.

Although the neural network and other conventional artificial intelligent techniques have been widely used in machinery fault diagnosis, they demand sufficient samples and have limitations on generalization of results in models that can over-fit the samples for the reason that their theoretical basis is empirical risk minimization principle [4,18,19]. The reason why support vector machine has excellent performance is that it is based on statistical learning theory and has specialties for a small number of samples [20,21]. Since the machinery fault samples are often very scarce, support vector machine exhibits superiority in machinery fault

* Corresponding author. Tel./fax: +86 29 82663689.
E-mail address: chenxf@mail.xjtu.edu.cn (X. Chen).

http://dx.doi.org/10.1016/j.neucom.2015.04.069
0925-2312/

diagnosis due to its high accuracy and good generalization for a smaller number of fault samples [19]. Its superiority is testified in contrast with many other intelligent methods in turbo-pump [4], roller bearing [22], power transformer [23,24], and so on. But finding out good features that can discriminate different fault conditions and optimizing parameters for support vector machine can be regarded as the most two important problems that can highly affect the final diagnosis accuracy of support vector machine. The motivation for feature selection is three-fold: improve generalization error, determine the relevant features, reduce the dimensionality of the input space [25]. It is also testified that support vector machine usually depends on several parameters [25], one of which (denoted C) controls the tradeoff between margin maximization and error minimization. Other parameters called kernel parameters implicitly define the non-linear mapping from input space to high-dimensional feature space. As the performance of support vector machine will be weakened if these parameters are not properly chosen, it is an indispensable step to optimize the parameters of support vector machine for a good performance in handling a learning task.

At present, some feature selection methods with genetic algorithm [26], principal component analysis [27] and their hybrid method [28] are proposed. And also, there are many methods proposed for parameter optimization of support vector machine, such as genetic algorithm [23,29], immune algorithm [30] and ACO-based algorithm [31]. It is known that the efforts of feature selection and parameter optimization will be weakened if the two issues are treated separately without considering their jointly effect. So it faces the conundrum of neglecting one or the other, which restrict the generalization performance of support vector machine. Therefore, it is demand to synchronously deal with the feature selection and parameter optimization problem so as to obtain a subset of representative features associated with the conditions of machinery components and synchronously find the optimal parameters corresponding to the structure of support vector machine which is trained by the representative sample features.

Ant colony algorithm was introduced by M. Dorigo and his colleagues as a nature-inspired method for the solution of combinatorial optimization problems in the early 1990s [32]. From then on, researchers have successfully applied ant colony algorithm to many optimization problems, such as multi-objective resource allocation [33], vehicle routing [34], dynamic continuous optimization [35], global optimum function [36]. Ant colony algorithm is easy to realize, which only involves basic mathematic operation. The most important is the parallelism and distributional characteristics which ensures the capability of processing difficult combination optimization problems. Since it is demand to deal with the feature selection and parameter optimization problem synchronously to manifest the good generalization performance of support vector machine in machinery fault diagnosis. Ant colony algorithm is particularly attractive for feature selection and parameter optimization because there is no reliable heuristic available for finding the optimal feature subset and optimal parameters. It is expected that the ants can synchronously discover good feature and parameter combinations as they proceed through their search space. Therefore, an ant colony algorithm for synchronous feature selection and parameter optimization of support vector machine in intelligent fault diagnosis is firstly presented.

The organization of this paper is as follows. The basic theory of ant colony algorithm and support vector machine is briefly reviewed in Section 2. The proposed ant colony algorithm for synchronous feature selection and parameter optimization of support vector machine is explained in Section 3. In Section 4, the intelligent fault diagnosis method including data acquisition, feature extraction, and pattern recognition with the proposed ant colony algorithm for synchronous feature selection and parameter optimization of support vector machine is described. In Section 5, the experiment on rotor system is conducted as a benchmark to verify the proposed method in contrast to seven different methods. In Section 6, the proposed method is applied in the locomotive roller bearings to diagnose multiple faults including compound fault and some faults with different severity. Conclusion and discussion are conducted in Section 7.

2. Theory background

2.1. Ant colony algorithm

Introduced in the early 1990s by M. Dorigo, ant colony algorithm aims at finding approximate solutions of optimization problems by artificial ants and their indirect communication via synthetic pheromones [37]. These ants deposit pheromone on the ground in order to mark some favorable path that should be followed by other members of the colony. Ant colony optimization exploits a similar mechanism for solving optimization problems [38]. Ant colony algorithm is a stochastic search procedure. The central component is the pheromone model, which is used to probabilistically sample the search space [39]. The pheromone model can be derived from a model of the tackled combinatorial optimization problem, which is defined as follows:

A model $P=(S,\Omega,f)$ of ant colony algorithm consists of

- A search (or solution) space S defined over a finite set of discrete decision variables and a set Ω of constraints among the variables;
- An objective function $f: S\rightarrow \mathscr{R}^+$ to be minimized.

The search space S is defined as follows: given a set of n discrete variables X_i with values $v_i^j \in D_i=\left\{v_i^1, v_i^2, \cdots, v_i^{|D_i|}\right\}$, $i=1,\cdots,n$. A variable instantiation, that is, the assignment of a value v_i^j to a variable X_i, is denoted by $X_i=v_i^j$. A feasible solution $s\in S$ is a complete assignment (i.e., an assignment in which each decision variable has a domain value assigned) that satisfies the constraints. If the set of constraints Ω is empty, then each decision variable can take any value from its domain independently of the values of the other decision variables. In this case we call P an unconstrained problem model, otherwise a constrained problem model. A feasible solution $s^*\in S$ is called a globally optimal solution (or global optimum), if $f(s^*)\le f(s), \forall s\in S$. The set of globally optimal solutions is denoted by $S^*\in S$. One has to find a solution $s^*\in S^*$ to solve the problem.

2.2. Support vector machine

Support vector machine, a powerful machine learning methods for classification and regression problems of small samples and high dimensions, was initially presented by Vapnik in the last decade of the 20th century based on statistical learning theory and structural risk minimization principle [20,40]. Two different classification strategies of support vector machine (typical support vector machine for binary classification and one-against one support vector machine for multi-class classification) are briefly introduced as follows.

2.2.1. Typical support vector machine for binary classification

Given a sample set $ST=\left\{(\boldsymbol{x}_i,y_i)\,|\,\boldsymbol{x}_i\in H, y_i\in\{\pm 1\}, i=1,\cdots,l\right\}$, where $\boldsymbol{x}_i$ are the input vectors and y_i are the labels of $\boldsymbol{x}_i$, H is the feature set, l is the number of the training samples, an optimal hyper-plane is computed in a feature space to construct support vector machine. A classifier which generalizes well is then found by controlling both the classifier capacity (via $\|\boldsymbol{w}\|$) and the number of training errors in the following form

$$\text{Minimize } T(\boldsymbol{w},\boldsymbol{\xi}) = \langle \boldsymbol{w}\cdot\boldsymbol{w}\rangle + C\sum_{i=1}^{l}\xi_i \tag{1}$$

$$\text{Subject to}\begin{cases} y_i\cdot(\langle \boldsymbol{w}\cdot\mathbf{x}_i\rangle + b) \ge 1-\xi_i, & i=1,\cdots,l \\ \xi_i \ge 0, & i=1,\cdots,l \\ C>0 \end{cases} \tag{2}$$

where ξ_i is slack variable and C is a penalty constant. The data to be classified is mapped into a high dimensional feature space, where $\Phi(\boldsymbol{x}_i)$ is the high dimensional transformation of $\boldsymbol{x}_i$. Mercer kernel returns a dot product of the feature space mappings of original data points, stated as $K(\boldsymbol{x}_i,\boldsymbol{x}_j) = \left(\Phi^T(\boldsymbol{x}_i)\cdot\Phi(\boldsymbol{x}_j)\right)$. By incorporating kernels and rewriting it in Lagrange multipliers, the dual quadratic optimization problem is given by

$$\text{Maximize } W(\boldsymbol{\alpha}) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{3}$$

$$\text{Subject to}\begin{cases} \sum_{i=1}^{l}\alpha_i y_i = 0, & i=1,\cdots,l \\ 0\le\alpha_i\le C, & i=1,\cdots,l \end{cases} \tag{4}$$

where α_i is Lagrange multiplier.

Thus, by solving the dual optimization problem, the decision function is

$$f(\boldsymbol{x}) = sign\left(\sum_{i=1}^{l} y_i\alpha_i\cdot K(\boldsymbol{x},\boldsymbol{x}_i) + b\right) \tag{5}$$

The bias value b can be computed by exploiting the fact that for all support vectors $\boldsymbol{x}_i$ with $\alpha_i < C$, the slack variable ξ_i is zero, which follows from the Karush–Kuhn–Tucker complementarity conditions, and hence

$$\sum_{j=1}^{l} y_j\alpha_j\cdot K(\boldsymbol{x}_i\cdot\boldsymbol{x}_j) + b = y_i \tag{6}$$

In this study, radial basis function (RBF) kernel is adopted due to its good properties and universal usage, which is shown as follows:

$$K(\boldsymbol{x}_i,\boldsymbol{x}_j) = \exp\left(\left(-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2\right)/2\sigma^2\right) \tag{7}$$

2.2.2. One-against-one support vector machine for multi-class classification

Given a training sample set in the input space

$$ST = \{(\boldsymbol{x}_t, y_t) \mid \boldsymbol{x}_t\in H, y_t\in\{1,2,\cdots,k\}, t=1,\cdots,l\} \tag{8}$$

where $\boldsymbol{x}_t$ are the input vector, y_t are the corresponding labels of $\boldsymbol{x}_t$, l is the number of the training samples, and k is the number of the different classes. A multi-class classification strategy of support

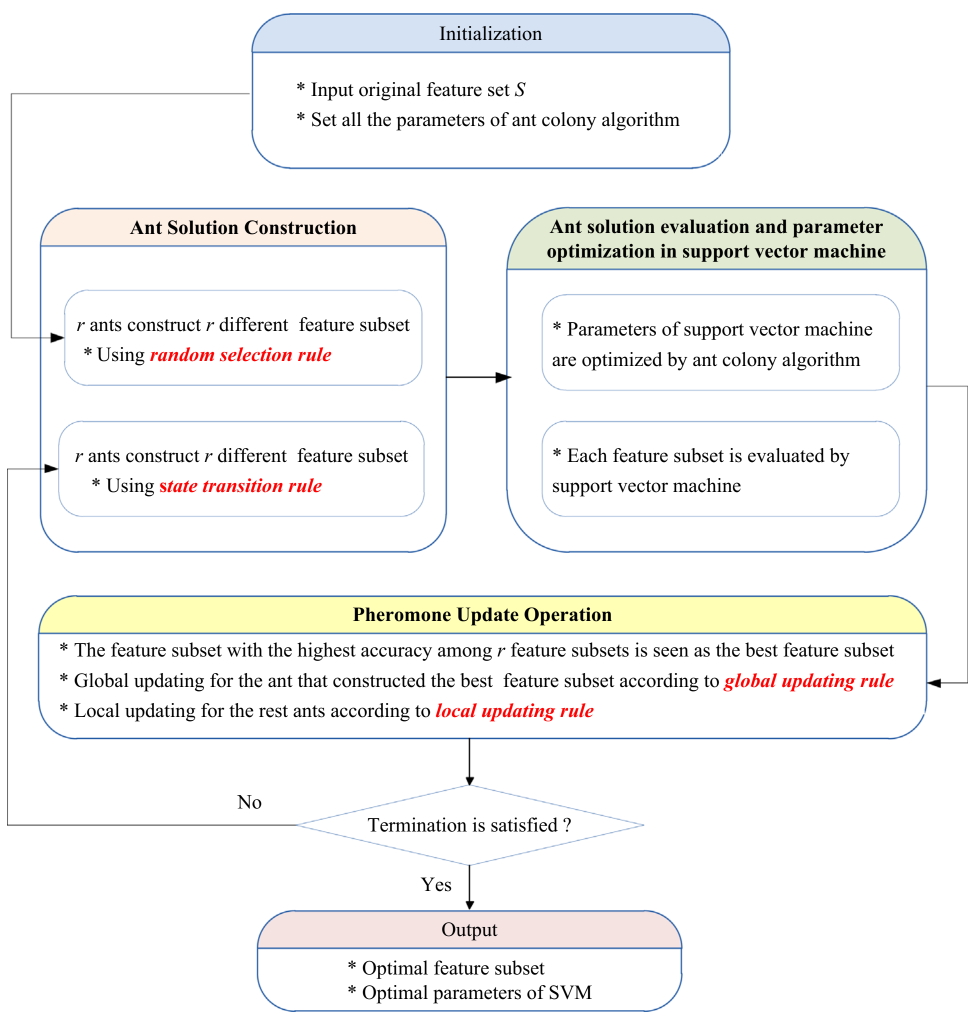

Fig. 1. Flow chart of ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine.

vector machine which is called one-against-one support vector machine constructs $k(k-1)/2$ binary classifiers to recognize the different classes. For training data from the *ith* and *jth* classes, the following binary classification problem can be solved:

$$\text{Minimize}\tfrac{1}{2}(\boldsymbol{w}^{ij})^T\boldsymbol{w}^{ij}+C\sum_t \xi_t^{ij}(\boldsymbol{w}^{ij})^T \tag{9}$$

$$\text{Subject to}\begin{array}{ll}(\boldsymbol{w}^{ij})^T\Phi(\boldsymbol{x}_t)+b^{ij}\geq 1-\xi_t^{ij} & \text{if } y_t=i,\\ (\boldsymbol{w}^{ij})^T\Phi(\boldsymbol{x}_t)+b^{ij}\leq 1-\xi_t^{ij} & \text{if } y_t=j,\\ \xi_t^{ij}\geq 0. & \end{array} \tag{10}$$

where ξ_t^{ij} is a slack variable, C is a penalty constant. By incorporating kernels and rewriting it in Lagrange multipliers, the above binary classification problem can be transformed into the dual quadratic optimization problem and finally forms the same decision function as in formula (5).

$$f^{ij}(\boldsymbol{x})=sign\left(\sum_{t=1}^{l} y_t\alpha^{ij}\cdot K(\boldsymbol{x},\boldsymbol{x}_t)+b^{ij}\right) \tag{11}$$

After all the $k(k-1)/2$ classifiers are constructed, the classification decision of the one-against-one support vector machine is made using the following strategy: if $f^{ij}(\boldsymbol{x})$ says sample $\boldsymbol{x}$ is in the *ith* class, then the vote for the *ith* class is added by one. Otherwise, the *jth* is increased by one. After being tested with the $k(k-1)/2$ classifiers respectively, $\boldsymbol{x}$ belongs to the class which has the maximal votes.

3. Ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine

Based on the predominance of ant colony algorithm, an ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine is proposed as a pattern recognition technology. The central components of the proposed method are the ant solution construction part, ant solution evaluation part and the pheromone update operation part, which utilize the heuristic information of artificial ants to guide the search of an optimal feature subset and optimal parameters for support vector machine so as to obtain the best generalization performance. An overview of the proposed ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine is presented in Fig. 1, which is worked as follows.

3.1. Initialization

The original feature set is inputted and the relevant parameters of ant colony algorithm are initialized. The number of ants initialized depends upon the number of features given by the problem. After the first step of initialization, the next step of ant solution construction is started.

3.2. Ant solution construction

Each ant initialized in the first step will select a subset of features from the original set of N features with the random selection rule, which is explained in Section 3.2.1. Later, ants select features based on the state transition rule, which is explained in Section 3.2.2. Thus, each ant selected particular features and forms different feature subsets s_1, s_2, ⋯, s_r (r is the number of ants). Each feature subset has n_1, n_2, ⋯, n_r features. This part is the second step of the ant colony algorithm. The main content of the ant solution construction part is the random selection rule and state transition rule, which are explained as follows.

3.2.1. Random selection rule

The role of each ant is to build solution subset of features. Initially, the pheromone level of each feature is the same, thus there is a uniform distribution on all these features. Therefore, all of the ants stochastically select features and construct their solution subset.

3.2.2. State transition rule

Except the circumstance described in Section 3.2.1 that the ants selected features with random selection rule at the initial of the algorithm proceeded, ants build solutions applying a probabilistic decision policy called state transition rule to move through adjacent states. In the state transition rule, an ant chooses a feature as follows:

$$s=\arg\max\{\tau(u)\} \tag{12}$$

where $\tau(u)$ is the pheromone level at state u, so the state transition rule of formula (12) favors the selection of features which are associated with high amount of pheromone level. High amount of pheromone level is according with high accuracy which is explained in the following Section 3.4.

3.3. Ant solution evaluation and parameter optimization in support vector machine

Each selected feature subset obtained in Section 3.2 is input to support vector machine for evaluation. Since the regularization constant C controls the tradeoff between margin maximization and error minimization, and the kernel parameters such as the bandwidth σ of the radial basis function (RBF) kernel implicitly define the non-linear mapping from input space to high-dimensional feature space, which exert considerable influence on the performance of support vector machine. It is an indispensable step to synchronously optimize the parameters of support vector machine by ant colony algorithm.

Firstly, the two parameters (C,σ) of support vector machine are meshed into N grids and the grid interval of each parameter is calculated by formula (13).

$$h_j=\left(v_j^{upper}-v_j^{lower}\right)/N,(j=1,2,\cdots,m) \tag{13}$$

where v_j^{upper} and v_j^{lower} respectively denote the upper and lower limit of a parameter, m is the number of parameters of support vector machine. This paper sets $v_j^{upper}=2^6$, $v_j^{lower}=2^{-5}$, $j=1,2$, $N=10$. Every grid is equidistant on each parameter interval. And each grid node denotes a parameter combination.

Since there is a uniform distribution on these grid nodes in the beginning, every of the ants stochastically select a parameter combination as a starting point. Then the training process of support vector machine with the selected parameters starts and the test error E is calculated according to formula (14).

$$E=\tfrac{1}{p}\sum_{i=1}^{p}\Psi\left(-y_i'f(\mathbf{x}_i')\right) \tag{14}$$

where $\boldsymbol{x}_i'$ is a testing sample, y_i' is the label of the testing sample, p is the number of the testing samples, Ψ is a step function, $\Psi(x)=1$ when $x>0$and $\Psi(x)=0$ else, f is the decision function of support vector machine.

Once all the ants have completed their selection, the artificial pheromone in formula (15) is applied to all the nodes ant selected.

$$\varphi_{ij}^{new}=(1-\kappa)\varphi_{ij}^{old}+(\lambda/e^E) \tag{15}$$

where E is the value of function in formula (14), κ is an evaporation coefficient, and λ is pheromone intensity. The node of the parameter combination with the fewest test error E in formula (14) is rewarded with more pheromone so as to make it more attractive for the future ants to select.

After artificial pheromone is applied to all the nodes selected by ants, the corresponding subscript of the node with maximal pheromone quality is found out, and the parameter scope is diminished according to formulas (16) and (17)

$$v_j^{lower} \leftarrow v_j^{lower} + (m_j - \Delta) * h_j \tag{16}$$

$$v_j^{upper} \leftarrow v_j^{upper} + (m_j + \Delta) * h_j \tag{17}$$

where Δ is a coefficient, m_j is the corresponding subscript of the node with maximal pheromone quality, v_j^{upper} is the upper limit of a parameter and v_j^{lower} is the lower limit of a parameter in current iteration. From then on, the ants search parameters in the neighborhood of the node with maximal pheromone until to the current iteration.

The ant solution evaluation steps are repeated until the grid interval h_j is less than predefined precision ε. In order to get a tradeoff between the accuracy and computation complexity, the precision ε is predetermined as 0.01 in the following part. At this time, the optimal parameters are obtained as follows:

$$v_j^* = \left(v_j^{lower} + v_j^{upper}\right)/2, j = 1, \cdots, m \tag{18}$$

where v_j^* denotes the obtained optimal parameter values.

3.4. Pheromone update operation

Once all the ants have completed their tasks of ant solution construction and ant solution evaluation in the above steps, the pheromone update operation is conducted. The main content of the pheromone update operation includes a global updating rule and a local updating rule, which are explained as follows.

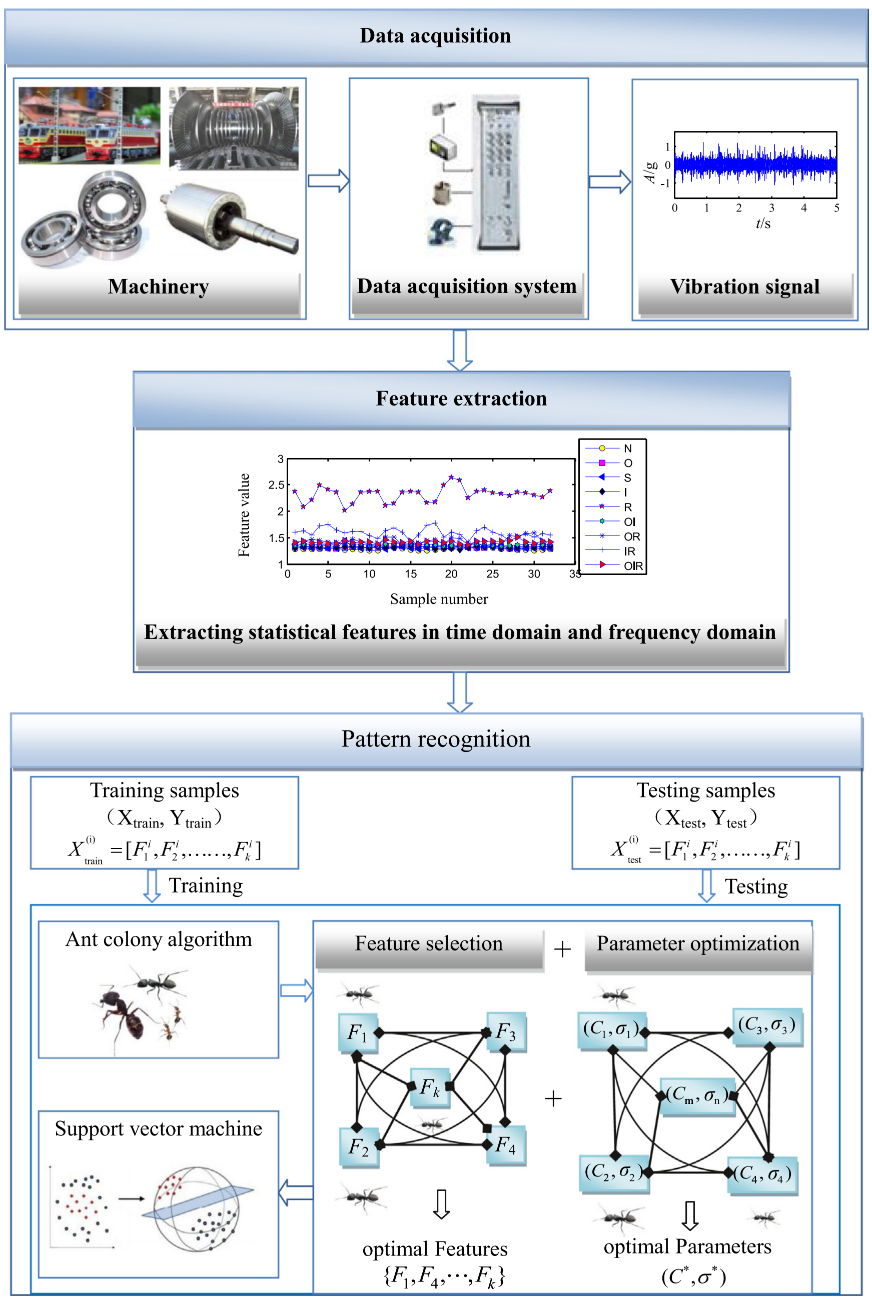

Fig. 2. Illustration of Intelligent fault diagnosis with ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine.

Table 1
The frequency statistical features.

$F_7 = \frac{\sum_{k=1}^{K} s(k)}{K}$	$F_8 = \frac{\sum_{k=1}^{K} (s(k)-F_1)^2}{K-1}$	$F_9 = \frac{\sum_{k=1}^{K} (s(k)-F_1)^3}{K(\sqrt{F_2})^3}$
$F_{10} = \frac{\sum_{k=1}^{K} (s(k)-F_1)^4}{Kp_2^2}$	$F_{11} = \frac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)}$	$F_{12} = \sqrt{\frac{\sum_{k=1}^{K} (f_k-F_5)^2 s(k)}{K}}$
$F_{13} = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} s(k)}}$	$F_{14} = \sqrt{\frac{\sum_{k=1}^{K} f_k^4 s(k)}{\sum_{k=1}^{K} f_k^2 s(k)}}$	$F_{15} = \frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sqrt{\sum_{k=1}^{K} s(k) \sum_{k=1}^{K} f_k^4 s(k)}}$
$F_{16} = \frac{F_6}{F_5}$	$F_{17} = \frac{\sum_{k=1}^{K} (f_k-F_5)^3 s(k)}{KF_6^3}$	$F_{18} = \frac{\sum_{k=1}^{K} (f_k-F_5)^4 s(k)}{KF_6^4}$
$F_{19} = \frac{\sum_{k=1}^{K} (f_k-F_5)^{1/2} s(k)}{K\sqrt{F_6}}$	where $s(k)$ is a spectrum for $k=1,2,\ldots,K$, K represents the number of spectrum lines; f_k denotes the frequency value of the kth spectrum line.	

3.4.1. Global updating rule

The purpose of the global updating rule is to encourage the ants that produced best feature subset with highest accuracy in the current iteration. Therefore, the pheromone level of these features which are selected as a component of the best feature subset will be incremented and thus make them more attractive for ants to select in the future. Global updating rule is performed only after all the ants have developed their respective solutions. The pheromone level is incremented by applying the global updating rule

$$\tau(k+1) = (1-\rho) * \tau(k) + Q * T_{\max} \tag{19}$$

where ρ is an evaporation coefficient, and Q is pheromone intensity. $T_{\max}$ is the highest accuracy of the solution among all the solution constructed by each individual ant.

Supposing there is a testing dataset $V' = \{(\boldsymbol{x}'_i, y'_i) \mid \boldsymbol{x}'_i \in s_r, y'_i \in Y, i = 1, \cdots, q\}$, Y is the label set, $s_r = (e_1, e_2, \cdots, e_k)$ is the feature subset selected by the ant r, q is the number of samples in the testing dataset, then the accuracy of the solution built by the ant r is

$$T^r_{ant} = 1 - \frac{1}{q} \sum_{i=1}^{q} \Psi\left(-y'_i f(x'_i)\right) \tag{20}$$

where Ψ is a step function, $\Psi(x) = 1$ when $x > 0$ and $\Psi(x) = 0$ else. f is the decision function of support vector machine.

The highest accuracy of the global best solution is

$$T_{\max} = \max\{T^r_{ant}\} \tag{21}$$

3.4.2. Local updating rule

The objective of local updating rule is to decrease the pheromone level of other features that were selected by ants which did not produce a good solution, and maintain the pheromone level of features that have not been exploited by ants until to present. This local updating rule not only makes the irrelevant features less desirable, but also helps ants to select those features which have never been selected previously. The local updating rule is as follows:

$$\tau(k+1) = (1-\alpha_0) * \tau(k) + \alpha_0 * \tau_0 \tag{22}$$

where $\alpha_0 (0 < \alpha_0 < 1)$ is called local pheromone update parameter and τ_0 is the initial pheromone level at the beginning of the algorithm.

By using the global updating rule and local updating rule, the pheromone level of each feature which is the component of the best feature subset in the current iteration is increased and the pheromone level of the feature which was selected but did not produce the best solution is decreased, while the pheromone level of the rest features that have not be selected maintain the same. This is good for ants to select optimal features and so as to construct the best feature subset in the future.

3.5. The termination requirement

The termination requirement of the ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine is that either the pattern recognition accuracy with the best feature subset in the current iteration hits 100% or all the features are selected by each individual ant. The above steps are repeated until the termination requirement is satisfied. Finally, the optimal feature subset and the corresponding optimal parameters of support vector machine are synchronously output and the final results are obtained.

4. Intelligent fault diagnosis with ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine

The intelligent fault diagnosis with ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine as a pattern recognition approach shown in Fig. 2 is mainly consist of data acquisition, feature extraction and pattern recognition, which is explained as follows.

4.1. Data acquisition

The intelligent fault diagnosis of rotating machinery starts with data acquisition to collect the machinery health information. Vibration signal acquisition is the most commonly used method which is realized by sensors. For the example given in the paper, the vibration signals are obtained by sensors and data acquisition system for intelligent fault diagnosis of rotating machinery.

4.2. Feature extraction

It is desirable that features extracted from the sensory signal are sensitive to machinery faults and robust to the varying machinery operating conditions and background noise with inexpensive computations. So there has been a lot of signal processing approach to obtain desirable features for machinery fault diagnosis, among which the Fast Fourier Transform (FFT) is one of widely used and well-established methods. When a fault occurs, new frequency components may appear and a change of the convergence of frequency spectrum may take place [41]. As a result, the amplitude and the distribution of frequency spectrum are likely to exhibit different characteristics in frequency-domain. Scalar indicators extracted from the time domain provide

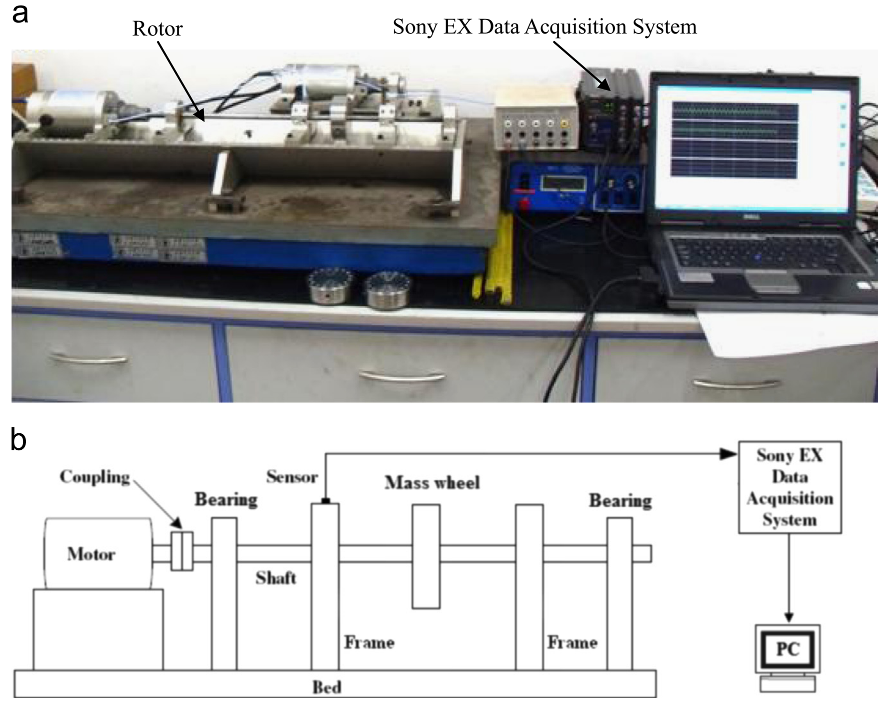

Fig. 3. Test bench of rotor system (a) experiment picture and (b) schematic illustration.

Table 2
Description of fault conditions in rotor system.

Condition description	Abbreviation of condition type	Label
Mass unbalance	U	1
Oil whirl	W	2
Rotor radial rub	R	3
Shaft crack	S	4
Compound faults of mass unbalance and rotor radial rub	C	5
Normal	N	6

information on the defects and the evolution of their values indicates the level of aggravation of defects. The attitude towards feature extraction for intelligent fault diagnosis emphasizes on including many easily acquired features without much expertise. Therefore, 19 statistical features of the acquired raw signals and the corresponding FFT spectrums are adopted to characterize the machinery condition. The statistical feature F_1–F_6 shown in formula (23)–(28) are time domain features, which are usually called shape factor, crest factor, impulse factor, margin factor, kurtosis factor and skewness factor. The parameters F_7–F_{19} are frequency domain features, which are listed in Table 1. Feature F_7 indicates the vibration energy in frequency domain. Feature F_8-F_{10}, F_{12} and $F_{16}-F_{19}$ describe the convergence of frequency spectrum power. Feature $F_{11}, F_{13}-F_{15}$ show the position change of main frequencies [42]. Each of the vibration signals collected in the following experiments is processed respectively to get the 19 features. By extracting features from the acquired vibration signals of machines in normal and faulty conditions, it offers very important information for pattern recognition.

(1) Shape factor :
$$F_1 = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N} x(n)^2}}{\frac{1}{N}\sum_{n=1}^{N} |x(n)|} \tag{23}$$

where $x(n)$ represents a signal series for $n=1,2,\ldots,N$, N denotes the number of data points.

(2) Crest factor :
$$F_2 = \frac{\max|x(n)|}{\sqrt{\frac{1}{N}\sum_{n=1}^{N} x(n)^2}} \tag{24}$$

(3) Impulse factor :
$$F_3 = \frac{\max|x(n)|}{\frac{1}{N}\sum_{n=1}^{N} |x(n)|} \tag{25}$$

(4) Margin factor :
$$F_4 = \frac{\max|x(n)|}{\left(\frac{1}{N}\sum_{n=1}^{N} \sqrt{|x(n)|}\right)^2} \tag{26}$$

(5) Kurtosis factor :
$$F_5 = \frac{\frac{1}{N}\sum_{n=1}^{N} (x(n)-\bar{x})^4}{\left(\sqrt{\frac{1}{N}\sum_{n=1}^{N} (x(n)-\bar{x})^2}\right)^4} \tag{27}$$

where $\bar{x}$ is the mean value of the signal series $x(n)$, $\bar{x} = 1/N\sum_{n=1}^{N} x(n)$, $n=1,2,\ldots,N$, , N denotes the number of data points.

(6) Skewness factor :
$$F_6 = \frac{\frac{1}{N}\sum_{n=1}^{N} (x(n)-\bar{x})^3}{\left(\sqrt{\frac{1}{N}\sum_{n=1}^{N} (x(n)-\bar{x})^2}\right)^3} \tag{28}$$

4.3. Pattern recognition

In the intelligent fault diagnosis system, support vector machine is widely used as a pattern recognition approach to indicate whether the machinery is in fault running condition and identifying the type and the severity of the fault occurred. The proposed ant colony algorithm is adopted for synchronous

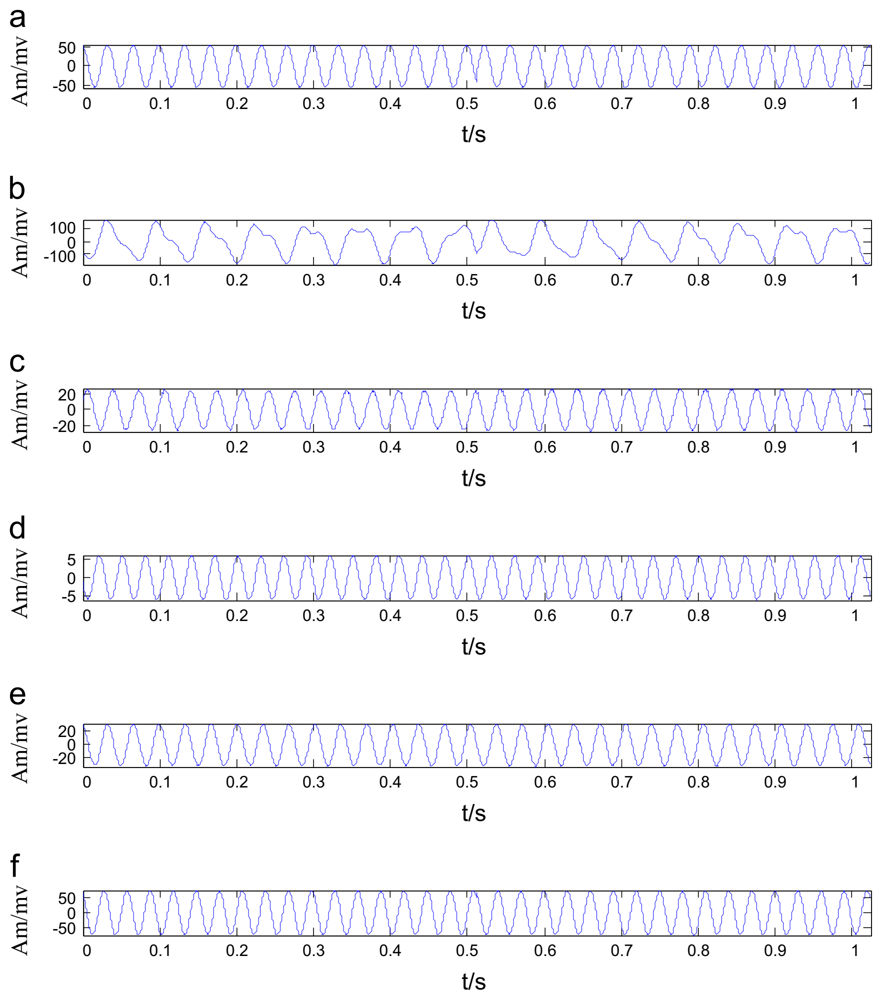

Fig. 4. Waveforms in time domain of six rotor vibration signals in different conditions (a) mass unbalance, (b) oil whirl, (c) rotor radial rub, (d) f shaft crack, (e) compound faults of mass unbalance and rotor radial rub, and (f) normal condition.

feature selection and parameter optimization in support vector machine.

5. Experiment on the intelligent fault diagnosis of rotor system

Rotor system is the most important component in the rotating machines, so an experiment is conducted on a widely used rotor system test bench as a benchmark by simulating different faults occurring in rotor system so as to test the effectiveness of the proposed method in contrast to other different fault diagnosis methods.

5.1. Experimental system description

The test bench of rotor system is shown in Fig. 3, which is used to simulate different fault conditions of rotor system. Fig. 2(a) is a picture of the experimental system, while Fig. 2(b) is a schematic illustration. The rotor system mainly consists of rotor system test bed (comprising a motor, two sliding bearings, a shaft and a rotor mass wheel, a motor speed controller, a signal conditioner), sensors and Sony EX data acquisition system. Eddy current sensors are installed towards radial direction of the shaft on the frame with sampling frequency of 2000 Hz as shown in Fig. 2. In the experiment, the diameter of the shaft is 10 mm and the length of the shaft is 560 mm. The rotor mass wheel is 800 g and its diameter is 75 mm. The rotor system experiment is conducted under six different running conditions including mass unbalance (0.5 g weight unbalance mass), oil whirl, slight rotor radial rub, shaft crack (the crack depth is 0.5 mm), compound fault of mass unbalance (0.5 g weight) and slight rotor radial rub, and normal condition (without any fault or defect), which are listed in Table 2.

5.2. Experimental result

The waveforms and the corresponding FFT spectrums of the vibration signals in six conditions are respectively plotted in Figs. 4 and 5. From the figures, it is seen that the waveforms in time domain and the spectrums in frequency domain reflect a little abnormal characteristics in fault state, such as the amplitude of the vibration signal under various fault state and normal condition is different, and very little waveform and spectrum changes is appeared. But it is not enough to accurately reveal the fault feature of each state. It is possible that an advanced signal processing approach is demanded to extract fault features or an intelligent fault diagnosis method is required to deal with the problem. As the latter, support vector machine is considered as a pattern recognition method to classify each state. Each of the six vibration signals corresponding to six faulty conditions is divided into 32 samples of 1024 points. Nineteen features in both time domain and frequency domain as described in Section 3.2 are extracted from each signal sample at first to reveal the vibration characteristics of the rotor system in different running conditions,

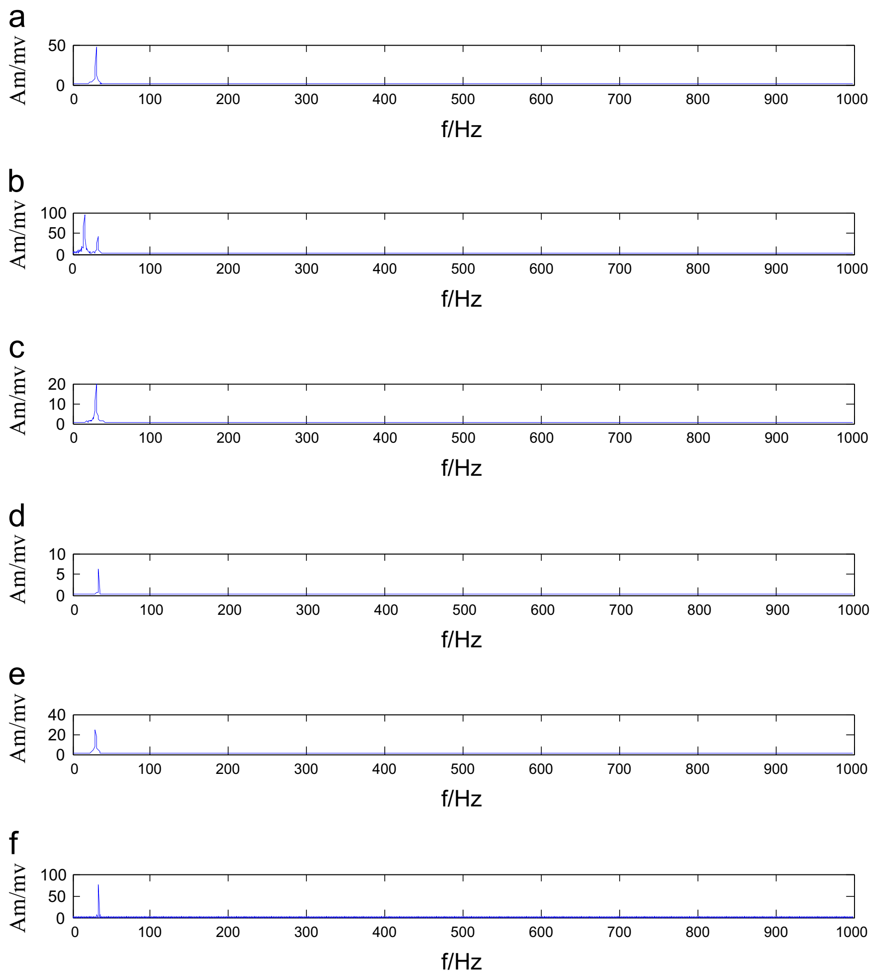

Fig. 5. Spectrums in frequency domain of six vibration signals in different condition (a) mass unbalance, (b) oil whirl, (c) rotor radial rub, (d) f shaft crack, (e) compound faults of mass unbalance and rotor radial rub, and (f) normal condition.

which are respectively plotted in Fig. 6. From Fig. 6(a) to (s), feature F_1–F_{19} of the 32 vibration signal samples under six different running conditions are respectively revealed, which is more or less irregular. For example, the feature characteristic of the 32 vibration signal samples under oil whirl condition is changed unsteadily and rapid fluctuated, except for the feature F_{14}, F_{17}, and F_{18} belonging to Fig. 6(n), (q), and (r) respectively. The feature characteristic of the 32 vibration signal samples under other five conditions (unbalance, radial rub, shaft crack, compound faults of mass unbalance and rotor radial rub, and normal condition) are nearly close to the same, such as the feature F_1, F_3, F_4, F_5, F_6, F_9, F_{10}, and F_{18} in Fig. 6(a), (c), (d), (e), (f), (i), (j) and (r) respectively, according to which it is hard to recognize the different condition of the rotor system. So it is demand to correctly recognize the patterns using some intelligent methods. Each condition of the first 16 samples are used for training support vector machine and the rest 16 samples are used for testing. One-against-one support vector machine described in Section 2.2.2 is used as the base learner machine for multi-class classification, and then the proposed ant colony algorithm for synchronous feature selection and parameter optimization is applied.

In order to test the effectiveness of the proposed method, it is compared with two methods based on back-propagation neural network (Method 1, Method 2) described in Ref. [43], two methods based on C4.5 decision tree (Method 3, Method 4) proposed in Ref. [43], and three different methods based on support vector machine (Method 5, Method 6, Method 7), which are listed in Table 3. Method 5 is the normal support vector machine which is none of feature selection and parameter optimization, so all of the extracted 19 features are input to support vector machine and the kernel parameters of support vector machine (C, σ) are set optionally since there is no prior knowledge to guide for proper parameters at first. Method 6 denotes the support vector machine with ant colony algorithm for parameter optimization which does not include feature selection, so all the extracted 19 features are wholly input to support vector machine and the ranges of kernel parameters of support vector machine (C, σ) are set as $\left[2^{-5}, 2^{6}\right]$ for optimization. Method 7 is the support vector machine with ant colony algorithm for feature selection which does not contain parameter optimization, and the kernel parameters of support vector machine (C, σ) are set as the optimal parameters obtained by Method 6. Method 8 denotes the proposed ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine, in which the ranges of kernel parameters of support vector machine (C, σ) are set as same as Method 6.

Method 1–Method 4 is realized by Sun and his colleagues for fault diagnosis under six rotor running conditions (mass unbalance, oil whirl, rotor radial rub, shaft crack, compound faults of mass unbalance and rotor radial rub, normal condition), in which the experiment were conducted as the same as the rotor experiment conducted in this paper. Seven statistical features (peak to peak, shape factor, crest factor, impulse factor, margin factor, kurtosis factor and skewness factor) of signals in time domain

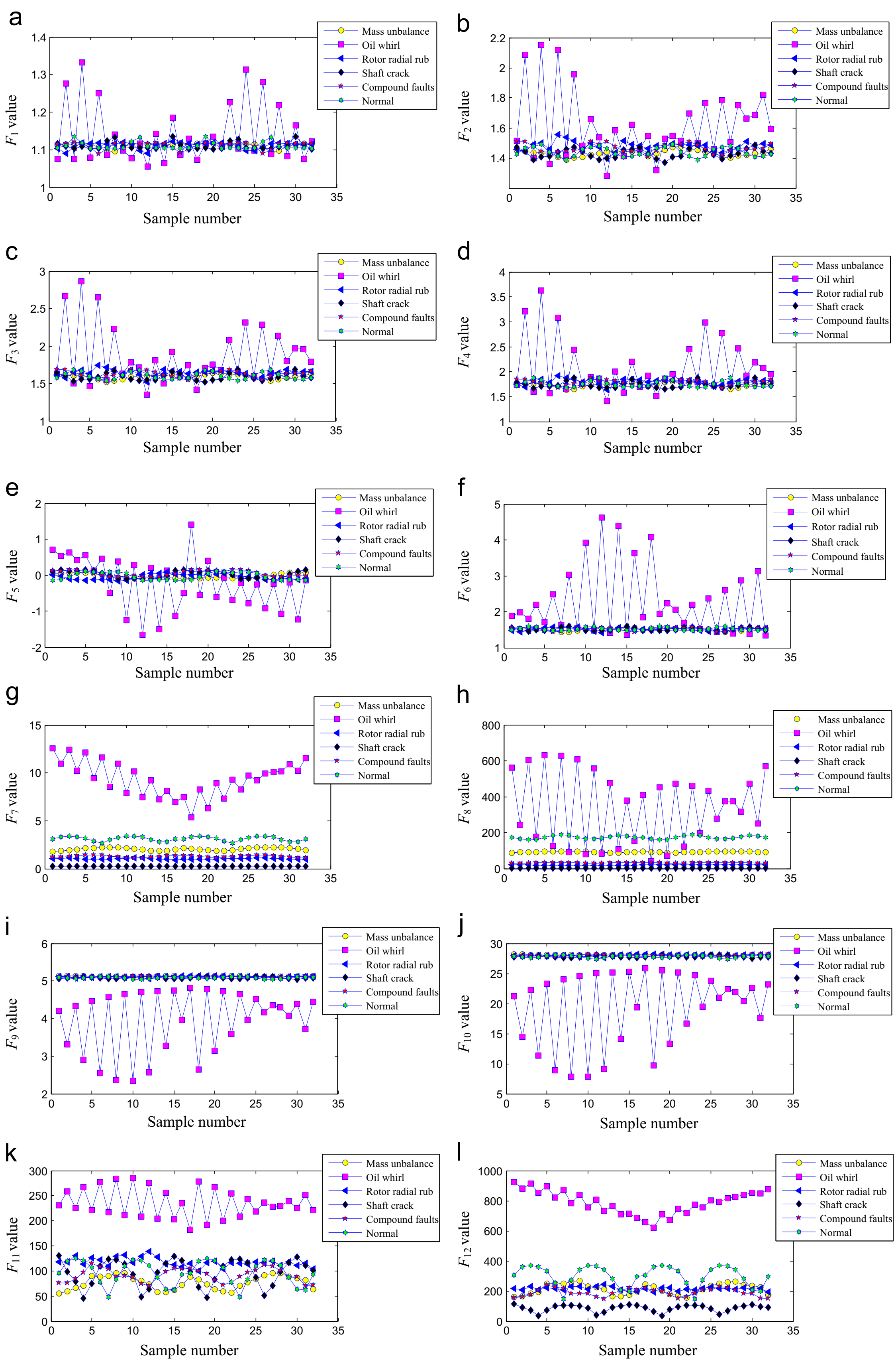

Fig. 6. Nineteen statistical features (F_1–F_{19}) distribution of the vibration signal in six running conditions of the rotor system. (a) feature F_1, (b) feature F_2, (c) feature F_3, (d) feature F_4, (e) feature F_5, (f) feature F_6, (g) feature F_7, (h) feature F_8, (i) feature F_9, (j) feature F_{10}, (k) feature F_{11}, (l) feature F_{12}, (m) feature F_{13}, (n) feature F_{14}, (o) feature F_{15}, (p) feature F_{16}, (q) feature F_{17}, (r) feature F_{18}, and (s) feature F_{19}.

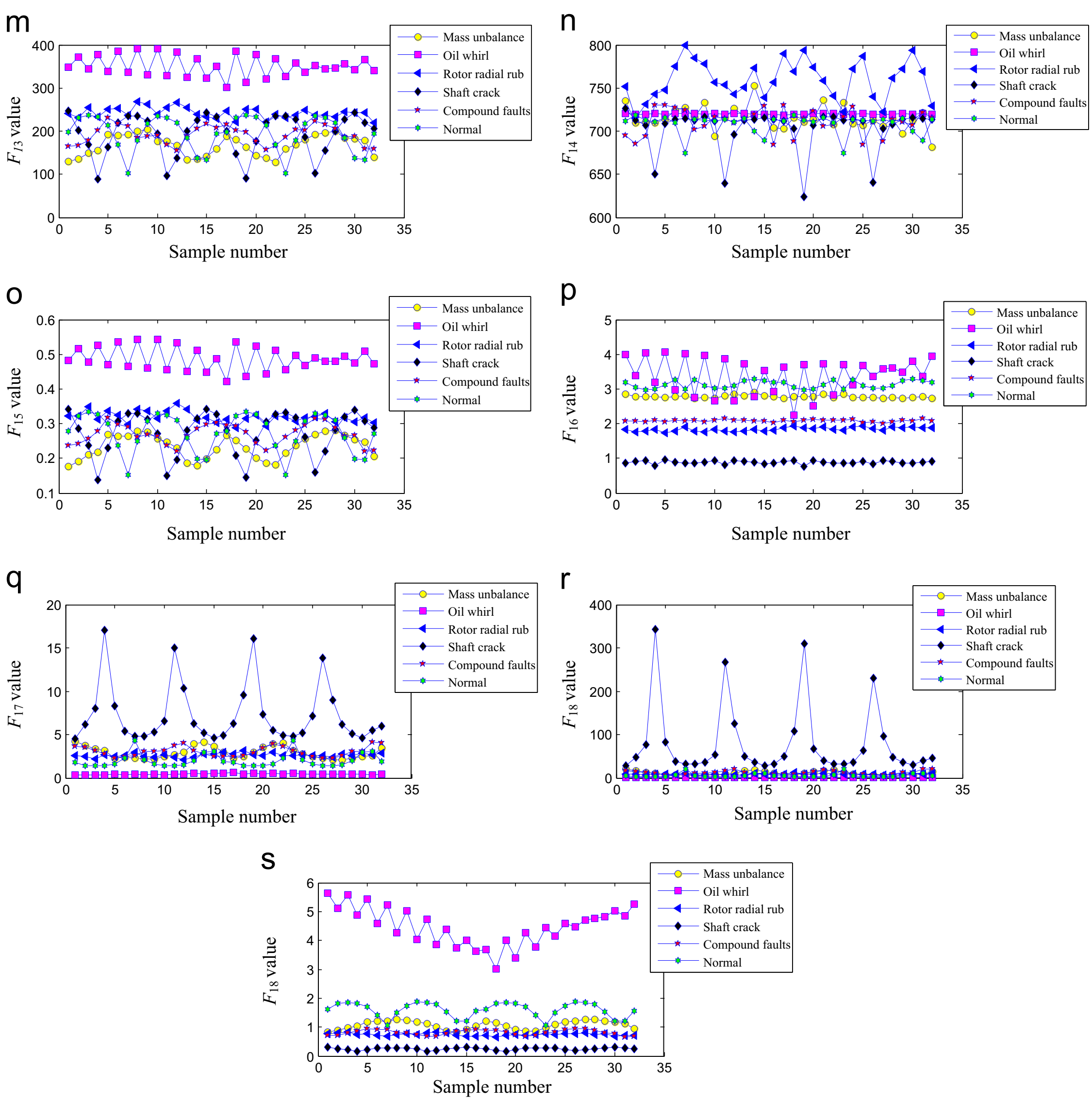

Fig. 6. (*continued*)

Table 3
Description of different fault diagnosis methods.

Abbreviation	Fault diagnosis method
Method 1	Back-propagation neural network without feature selection
Method 2	Back-propagation neural network with feature selection
Method 3	C4.5 decision tree without feature selection
Method 4	C4.5 decision tree with feature selection
Method 5	SVM without feature selection and parameter optimization
Method 6	SVM with ant colony algorithm for parameter optimization
Method 7	SVM with ant colony algorithm for feature selection
Method 8	SVM with ant colony algorithm for feature selection and parameter optimization

under various conditions and eleven frequency features were extracted. And then backward propagation neural networks and C4.5 decision tree were respectively used to realize the fault intelligent diagnosis, the experimental results of which are shown in Table 4.

Comparison results of the different fault diagnosis methods described above on rotor system are shown in Table 4. It is seen that the back-propagation neural network based methods (Method 1 and Method 2) attains the worst diagnosis results in all of the eight methods. The normal support vector machine which is none of feature selection and parameter optimization (Method 5), support vector machine with ant colony algorithm for parameter optimization (Method 6), and support vector machine with ant colony algorithm for feature selection (Method 7) have better diagnosis performance of 97.92% accuracy, which cannot perfectly diagnose the normal state and perform worse than the methods based on C4.5 decision tree (Method 3 and Method 4). In Method 3 and Method 4, Fault diagnostic error occurred in recognizing mass unbalance condition, compound faults of mass unbalance and rotor radial rub, since the fault characteristic of

Table 4
Comparison of different intelligent fault diagnosis methods on rotor system.

Name	Optimal features	Optimal parameters (C, σ)	Accuracy of each fault (%)						Average accuracy (%)
			U	W	R	S	C	N	
Method 1 [43]	–	–	100	85	100	90	95	100	95
Method 2 [43]	–	–	100	95	100	85	95	100	95.8
Method 3 [43]	–	–	100	100	95	100	95	100	98.3
Method 4 [43]	–	–	95	100	100	100	95	100	98.3
Method 5	–	–	100	100	100	100	100	87.5	97.92
Method 6	–	63.49, 37.64	100	100	100	100	100	87.5	97.92
Method 7	$F_1, F_3, F_5, F_7, F_9, F_{14}, F_{17}, F_{19}$	–	100	100	100	100	100	87.5	97.92
Method 8	$F_1, F_3, F_6, F_9, F_{10}–F_{14}, F_{19}$	50.69, 0.39	100	100	100	100	100	100	**100**

The bold in the table means that Method 8 is the proposed method and 100 is the result obtained by Method 8.

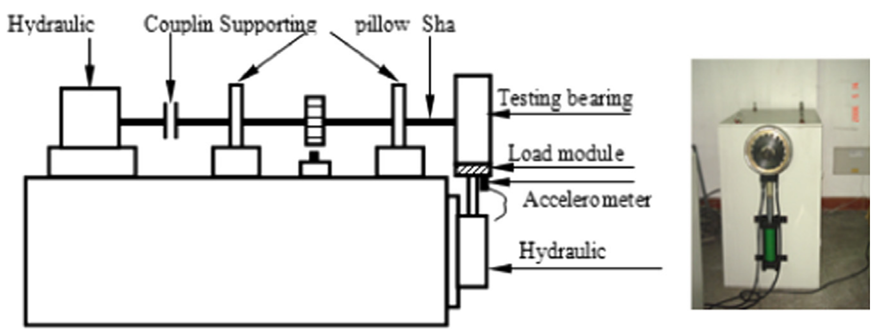

Fig. 7. Test bench of the locomotive roller bearing.

Table 5
Description of fault conditions of locomotive roller bearings.

Condition description	Abbreviation of condition type	Label
Normal	N	1
Slight fault in outer race	O	2
Serious fault in outer race	S	3
Inner race fault	I	4
Roller fault	R	5
Compound faults in outer race and inner race	OI	6
Compound faults in outer race and rollers	OR	7
Compound faults in inner race and rollers	IR	8
Compound faults in outer race, inner race and rollers	OIR	9

mass unbalance is contained in compound faults and thus the difficulty of correctly recognizing the two fault patterns are increased. It means that the performance of support vector machine is jointly affected by the parameters and sample features. Either the optimal features are selected or the parameters are optimized, the generalization of support vector machine cannot achieve optimal and the fault diagnosis accuracy is not very well. Therefore, it is necessary to achieve the matched optimal features and optimal parameters synchronously by an effective optimization algorithm so as to improve the generalization of support vector machine

With the proposed ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine (Method 8), $F_1, F_3, F_6, F_9, F_{10}–F_{14}, F_{19}$ are selected as the optimal features and the parameters are optimized as $C = 50.69$, $\sigma = 0.39$, which attains the highest diagnosis accuracy of 100%. So it can be concluded that the proposed ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine as a pattern recognition approach for intelligent fault diagnosis can improve the diagnosis performance and attains optimal diagnosis result since the synchronously achieved optimal features and parameters are well matching.

6. Application in the fault diagnosis of locomotive roller bearings

Intelligent diagnosis of locomotive roller bearings is necessary in industry, as unexpected failure of locomotive is a serious damage and manual checking of a number of roller bearings may take an unacceptably long time, which consequently results in great loss. Therefore, it is important to detect the existence and severity of roller bearings.

6.1. Experimental system description

The sketch map of the experimental system is described in Fig. 7. The test bench contains a hydraulic motor, two supporting pillow blocks (mounting with normal bearing), test bearings which are loaded on the outer race by a hydraulic cylinder, a hydraulic radial load application system, and a tachometer for shaft speed measurement. 608A11-type ICP accelerometers are mounted on the load module near the outer race of the test bearing, which measure the vibration data of the test bearings with sampling frequency of 6400 Hz.

The vibration signals are acquired respectively in the nine states of locomotive roller beatings including normal state, slight fault in outer

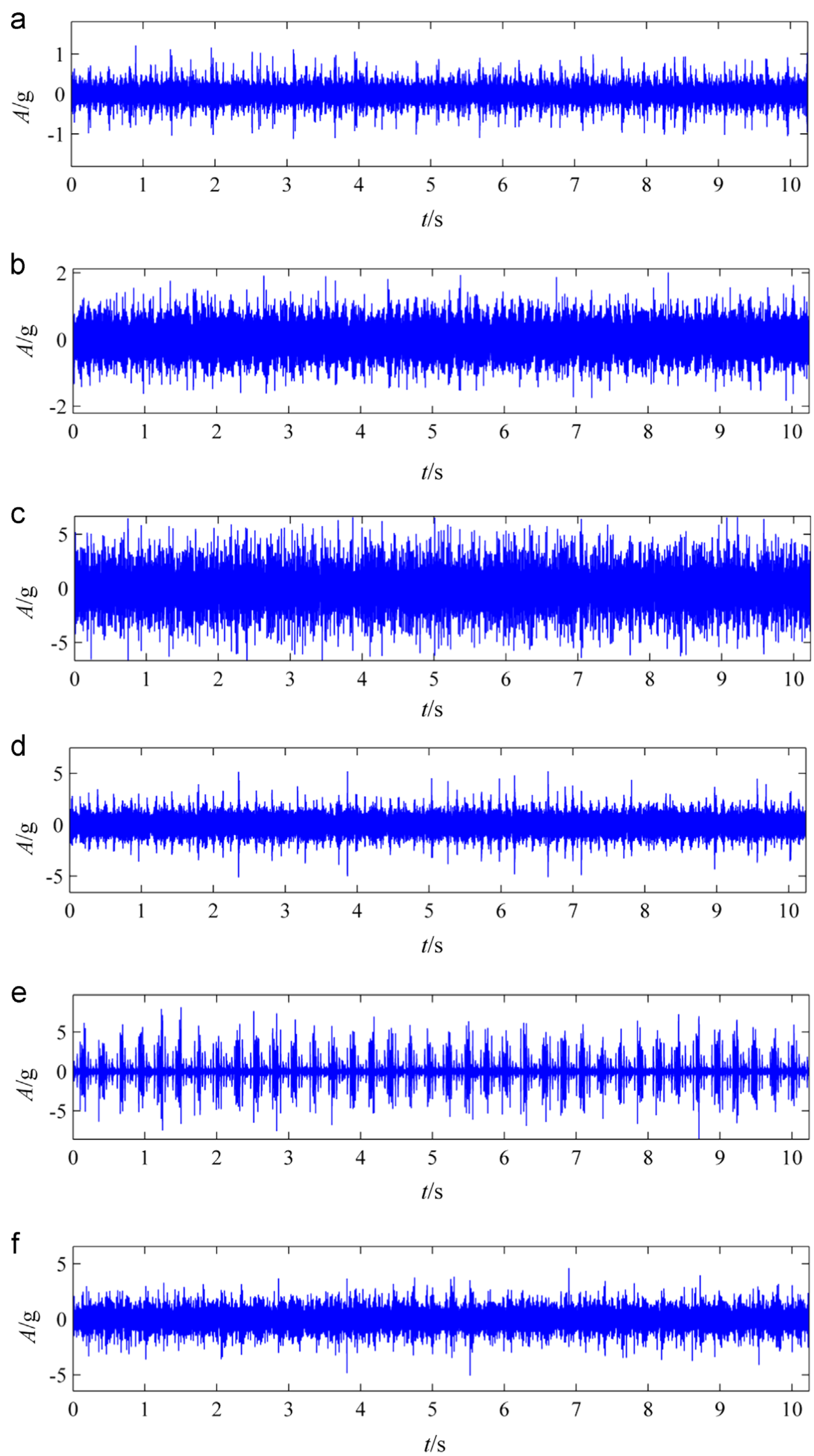

Fig. 8. Vibration signals corresponding to nine faulty conditions (a) normal condition, (b) slight fault in outer race, (c) serious fault in outer race, (d) inner race fault, (e) roller fault, (f) compound faults in outer race and inner race, (g) compound faults in outer race and rollers condition, (h) compound faults in inner race and rollers, and (i) compound faults in outer race, inner race and rollers.

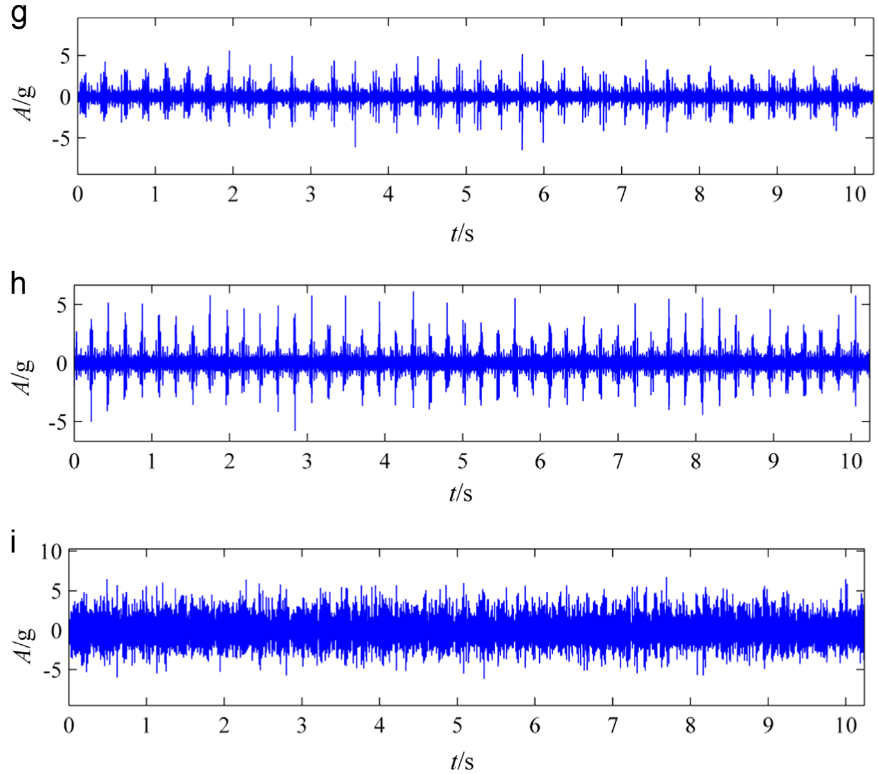

Fig. 8. (*continued*)

race, serious fault in outer race, inner race fault, roller fault, compound faults in outer race and inner race, compound faults in outer race and rollers, compound faults in inner race and rollers, compound faults in outer race and inner race and rollers, which are described in Table 5.

6.2. Experimental result

The vibration waveforms of the nine states in time domain are shown in Fig. 8, the characteristics are at least as described below:

(1) The vibration waveforms of the locomotive roller bearings in nine states are disturbed by much noise, and the fault features are submerged at different levels. The amplitudes of the vibration signal waveform in normal condition is smaller than that in other eight fault conditions.
(2) When a fault occurs in the outer race of the roller bearing, the rolling element passing through the damage position of the outer race will cause vibration and impulse. Therefore, some impulse occurs in Fig. 8(b) and (c). In comparison with the two figures, it is found that the vibration amplitude of the slight fault in the outer race as shown in Fig. 8(b) is smaller than that of serious fault in the outer race as shown in Fig. 8(c).
(3) When a fault occurs in the inner race of the roller bearing, some impulses occours in the waveform of the acquired vibration signal, which is as shown in Fig. 8(d).
(4) When a fault occurs in the rolling element, some impulses occours and the phenomenon of amplitude modulation happened in Fig. 8(e) since the surfaces of inner race and outer races are respectively contacted once by the rolling element when the rolling element rotates once.
(5) When compound faults occurs in outer race, inner race and rolling element, the waveforms of the acquired vibration signal have impulses of varying degrees. Moreover, the phenomenon of amplitude modulation occurs in the acquired vibration signal waveforms of the compound faults of outer race and rollers, compound faults of inner race and rollers, which are shown in Fig. 8(f)–(i).

The FFT spectrums of the corresponding vibration signals in nine states are shown in Fig. 9. It can be seen that the spectrums of the vibration signals in frequency-domain are submerged by much noise interference and the fault characteristics are ambiguity, so it is hard to recognize different fault types according to the spectrums in frequency-domain and the waveforms in the time-domain. It is possible that an advanced signal processing approach is demanded to extract fault features or an intelligent fault diagnosis method is required to deal with the problem. As the latter, 19 features in both time domain and frequency domain as described in Section 3.2 are extracted at first. Each of the nine vibration signals corresponding to nine faulty conditions is divided into 32 samples of 2048 points. The 19 statistical features of the 32 samples in time domain and frequency domain are respectively plotted in Fig. 10. From Fig. 10(a)–(s), feature F_1–F_{19} of the 32 vibration signal samples under nine different running conditions are respectively revealed, which are changed unsteadily and rapid fluctuated. The values of some features are nearly overlap, according to which it is hard to recognize the different condition of the roller bearings. So it is demand to correctly recognize the patterns by some intelligent methods.

Each of the nine signals corresponding to nine faulty conditions is divided into 32 samples of 2048 points. Each condition of the first 16 samples are used for training support vector machine and the rest 16 samples are used for testing. One-against-one support vector machine described in Section 2.2.2 is used as the base learner machine for multi-class classification, and then the proposed ant colony algorithm for synchronous feature selection and parameter optimization is applied. In order to test the effectiveness of the proposed method, it is compared with two methods based on support vector machine (SVM) to diagnose the faults in locomotive roller bearings, which are listed in Table 6.

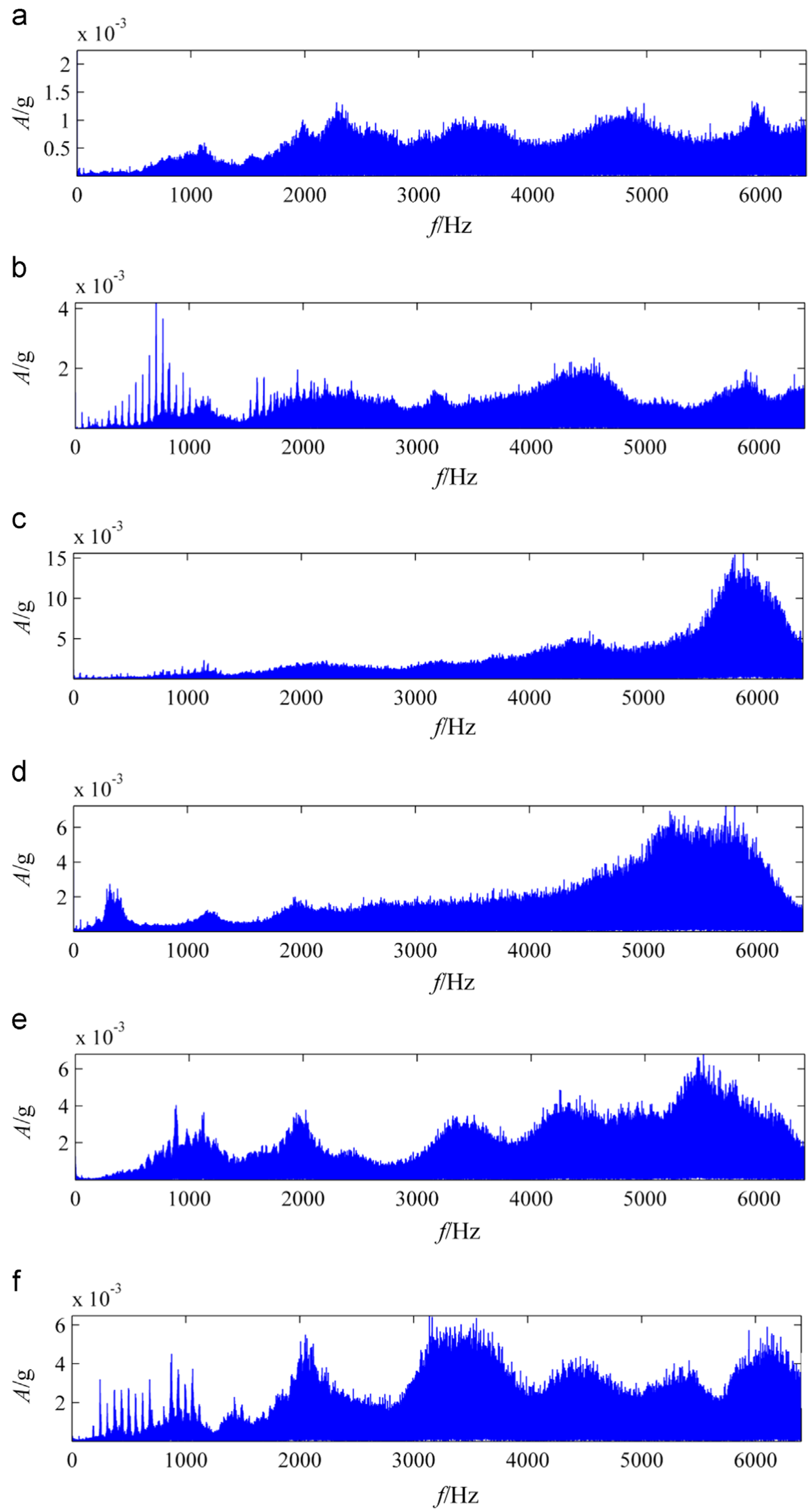

Fig. 9. Vibration spectrums of nine faulty conditions (a) normal condition, (b) slight fault in outer race, (c) serious fault in outer race, (d) inner race fault, (e) roller fault, (f) compound faults in outer race and inner race, (g) compound faults in outer race and rollers condition, (h) compound faults in inner race and rollers, and (i) compound faults in outer race, inner race and rollers.

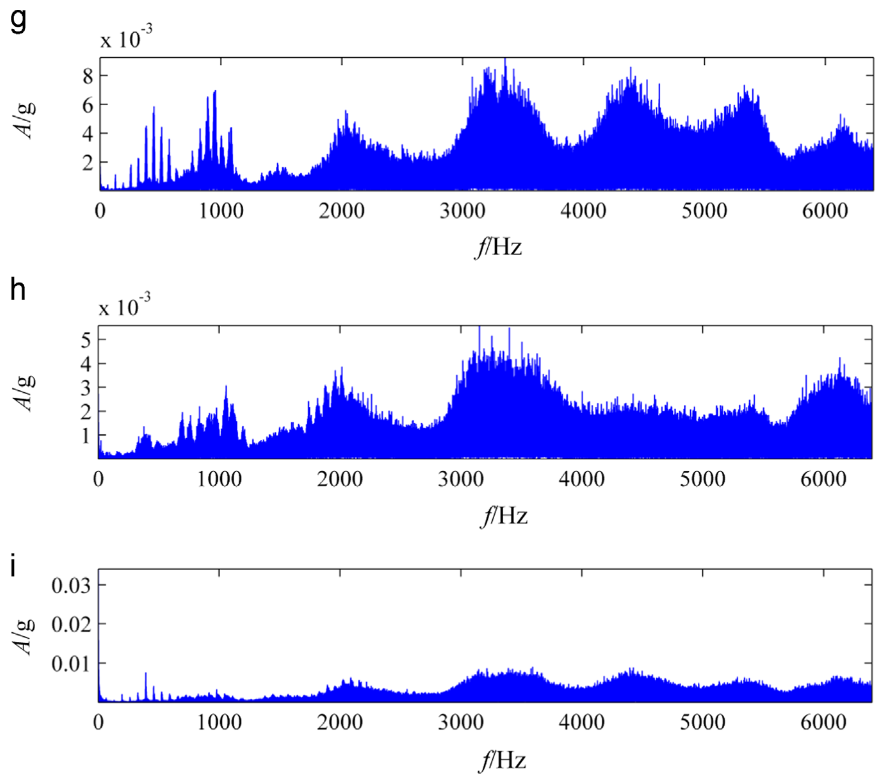

Fig. 9. (*continued*)

The intelligent diagnosis results by three different methods are shown in Table 7. Support vector machine with ant colony algorithm for parameter optimization (Method 1) attains the optimal parameters $C=57.60$ and $\sigma=64.26$ in the range of $C,\sigma \in [2^{-10}, 2^{10}]$, which achieves the average accuracy of 89.58% in recognizing the nine states of locomotive roller bearings. Although support vector machine with ant colony algorithm for feature selection (Method 2) uses the optimal parameters obtained in Method 1 and the features $F_2, F_4, F_6, F_{13}, F_{14}, F_{19}$ are selected as the optimal features, the average classification accuracy of the nine faults is also 89.58% since the selected features by Method 2 are not matched with the optimized parameters solved by Method 1. The same results of the Method 1 and Method 2 demonstrates that the performance of support vector machine is jointly affected by the parameters and sample features. So it is a demand to achieve the matched optimal features and optimal parameters synchronously by an effective optimization algorithm so as to comprehensively improve the generalization of support vector machine.

The proposed ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine (method 3) improves the diagnosis performance and attains desirable result of 95.83% accuracy with selected features F_1–F_{18} and optimal parameters ($C=1.02$, $\sigma=0.04$). As shown in Table 7, the recognition of the normal state, inner race fault state and roller fault state hit 100% accuracy in all the three methods, which means that the diagnosis performance of three methods based on support vector machine are the same in recognizing simple fault states. But the proposed ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine (Method 3) improves the recognition accuracy of serious fault in outer race, compound faults in outer and inner races, compound faults in outer race and rollers, compound faults in inner race and rollers, compound faults in outer and inner races and rollers, which testifies that the generalization performance of support vector machine is further improved by selecting optimal features and synchronously attaining optimal parameters matched with the optimal features at a time. So the proposed method attain the highest diagnosis accuracy than other two methods.

7. Discussion and conclusion

7.1. Conclusion

In this paper, an ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine is proposed as a pattern recognition approach for intelligent fault diagnosis of rotating machinery. The experiments of rotor system and the locomotive roller bearings show that the proposed method can effectively diagnose multi-class faults including complicated compound faults. Comparing with other pattern recognition methods, the proposed ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine attains the highest accuracy of fault diagnosis results.

7.2. Discussion

The feature extraction step of the proposed method is accomplished by extracting statistical features from raw vibration signals and the corresponding FFT spectrums, because it is easy for non-expert to attain without much experience. The fault diagnosis result may be further improved if advanced feature extraction methods are adopted.

Comparing with other methods, the proposed ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine improves the fault diagnosis result and attains desirable accuracy, which testifies that good diagnosis result not only depends on extracting proper features but also depends on finding optimal parameters.

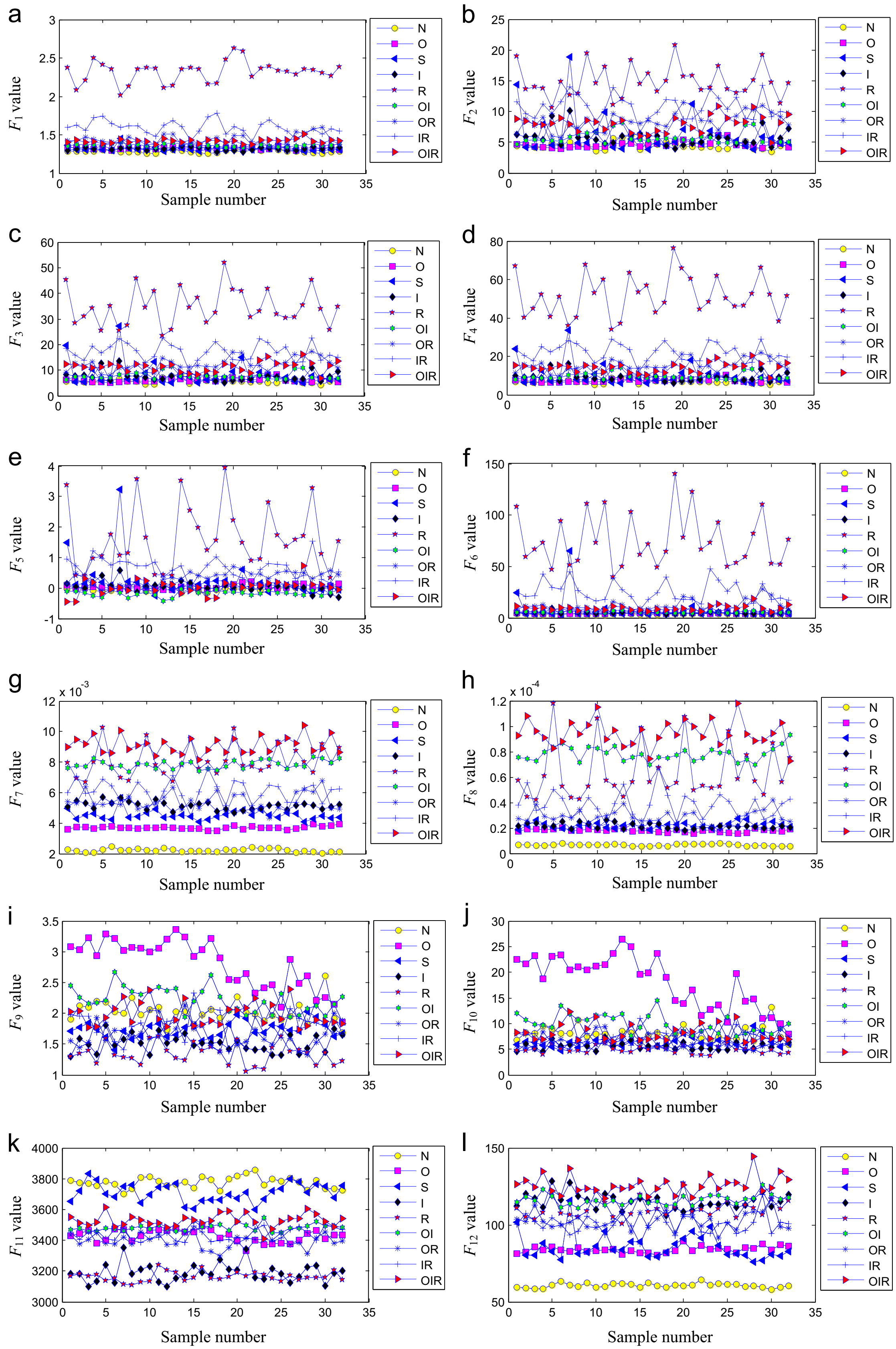

Fig. 10. Nineteen statistical features (F_1–F_{19}) distribution of the vibration signal in nine states of the locomotive roller beatings. (a) feature F_1, (b) feature F_2, (c) feature F_3, (d) feature F_4, (e) feature F_5, (f) feature F_6, (g) feature F_7, (h) feature F_8, (i) feature F_9, (j) feature F_{10}, (k) feature F_{11}, (l) feature F_{12}, (m) feature F_{13}, (n) feature F_{14}, (o) feature F_{15}, (p) feature F_{16}, (q) feature F_{17}, (r) feature F_{18}, and (s) feature F_{19}.

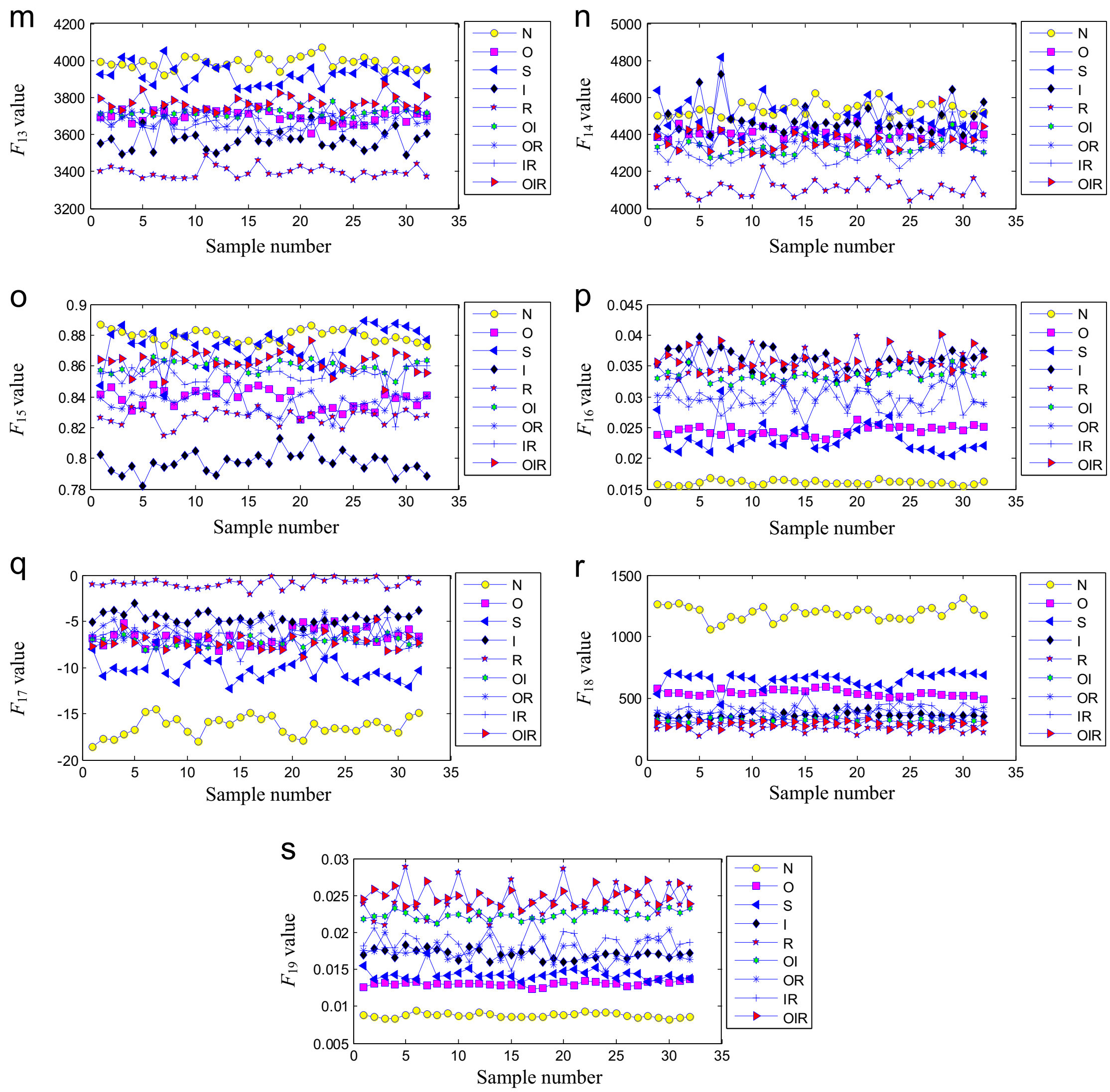

Fig. 10. (*continued*)

Table 6
Description of different fault diagnosis methods.

Abbreviation	Fault diagnosis method
Method 1	SVM with ant colony algorithm for parameter optimization
Method 2	SVM with ant colony algorithm for feature selection
Method 3	SVM with ant colony algorithm for feature selection and parameter optimization

Table 7
Intelligent fault diagnosis results of locomotive roller bearings.

Name	Optimal features	Optimal parameters (C, σ)	Accuracy of each fault (%)									Average Accuracy (%)
			N	O	S	I	R	OI	OR	IR	OIR	
Method 1	–	57.60, 64.26	100	100	87.5	100	100	75	75	75	93.75	89.58
Method 2	$F_2, F_4, F_6, F_{13}, F_{14}, F_{19}$	–	100	100	87.5	100	100	75	75	75	93.75	89.58
Method 3	F_1–F_{18}	1.02, 0.04	100	93.75	93.75	100	100	87.5	93.75	93.75	100	95.83

The proposed ant colony algorithm for synchronous feature selection and parameter optimization in support vector machine is proved to be effective in intelligent fault diagnosis of rotating machinery by the experiment of rotor system and locomotive roller bearings. The proposed method can also be applied to other machinery to achieve success.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant No. 51405028, and 51175509), the National Key Basic Research Program of China (Grant No. 2015CB057400), and the Fundamental Research Funds for the Central Universities of China, Chang' an University (Grant No. 2013G1502046).

References

[1] D.H. Kwak, D.H. Lee, J.H. Ahn, et al., Fault detection of roller-bearings using signal processing and optimization algorithms, Sensors 14 (1) (2014) 283–298.
[2] D.P. Jena, S. Sahoo, S.N. Panigrahi, Gear fault diagnosis using active noise cancellation and adaptive wavelet transform, Measurement 47 (2014) 356–372.
[3] P. Shi, Z. Chen, Y. Vagapov, et al., A new diagnosis of broken rotor bar fault extent in three phase squirrel cage induction motor, Mech. Syst. Signal Process. 42 (1–2) (2014) 388–403.
[4] S.F. Yuan, F.L. Chu, Support vector machines-based fault diagnosis for turbo-pump rotor, Mech. Syst. Signal Process. 20 (4) (2006) 939–952.
[5] A. Tantawy, X. Koutsoukos, G. Biswas, Aircraft power generators: hybrid modeling and simulation for fault detection, IEEE Trans. Aerosp. Electron. Syst. 48 (1) (2012) 552–571.
[6] A.K.S. Jardine, D.M. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Signal Process. 20 (7) (2006) 1483–1510.
[7] Y. Lei, Z.J. He, Y.Y. Zi, Application of a novel hybrid intelligent method to compound fault diagnosis of locomotive roller bearings, J. Vib. Acoust. ASME 130 (3) (2008).
[8] D.Y. Ma, Y.C. Liang, X.S. Zhao, et al., Multi-BP expert system for fault diagnosis of power system, Eng. Appl. Artif. Intell. 26 (3) (2013) 937–944.
[9] P. Jayaswal, S.N. Verma, A.K. Wadhwani, Development of EBP-artificial neural network expert system for rolling element bearing fault diagnosis, J. Vib. Control 17 (8) (2011) 1131–1148.
[10] Y. Shatnawi, M. Al-khassaweneh, Fault diagnosis in internal combustion engines using extension neural network, IEEE Trans. Ind. Electron. 61 (3) (2014) 1434–1443.
[11] S.S. Tayarani-Bathaie, Z.N.S. Vanini, K. Khorasani, Dynamic neural network-based fault diagnosis of gas turbine engines, Neurocomputing 125 (2014) 153–165.
[12] J.D. Wu, C.C. Hsu, Fault gear identification using vibration signal with discrete wavelet transform technique and fuzzy-logic inference, Expert. Syst. Appl. 36 (2) (2009) 3785–3794.
[13] D.P. Winston, M. Saravanan, Single parameter fault identification technique for DC motor through wavelet analysis and fuzzy logic, J. Electr. Eng. Technol. 8 (5) (2013) 1049–1055.
[14] V. Muralidharan, V. Sugumaran, Rough set based rule learning and fuzzy classification of wavelet features for fault diagnosis of monoblock centrifugal pump, Measurement 46 (9) (2013) 3057–3063.
[15] N.R. Sakthivel, V. Sugumaran, B.B. Nair, Comparison of decision tree-fuzzy and rough set-fuzzy methods for fault categorization of mono-block centrifugal pump, Mech. Syst. Signal Process. 24 (6) (2010) 1887–1906.
[16] Q. Liao, X.B. Li, B. Huang, Hybrid fault-feature extraction of rolling element bearing via customized-lifting multi-wavelet packet transform, Proc. Inst. Mech. Eng. Part C–J. Mech. Eng. Sci. 228 (12) (2014) 2204–2216.
[17] M. Seera, C.P. Lim, Online motor fault detection and diagnosis using a hybrid FMM-CART model, IEEE Trans. Neural Netw. Learn. Syst. 25 (4) (2014) 806–812.
[18] V. Sugumaran, G.R. Sabareesh, K.I. Ramachandran, Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine, Expert. Syst. Appl. 34 (4) (2008) 3090–3098.
[19] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Signal Process. 21 (6) (2007) 2560–2574.
[20] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1999.
[21] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machine: and Other Kernel Based Learning Methods, Cambridge University Press, Cambridge, 2000.
[22] Y. Yang, D.J. Yu, J.S. Cheng, A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM, Measurement 40 (9–10) (2007) 943–950.
[23] S.W. Fei, X.B. Zhang, Fault diagnosis of power transformer based on support vector machine with genetic algorithm, Expert. Syst. Appl. 36 (8) (2009) 11352–11357.
[24] G. Lv, H.Z. Cheng, H.B. Zhai, et al., Fault diagnosis of power transformer based on multi-layer SVM classifier, Electr. Power Syst. Res. 75 (1) (2005) 9–15.
[25] O. Chapelle, V. Vapnik, O. Bousquet, et al., Choosing multiple parameters for support vector machines, Mach. Learn. 46 (1–3) (2002) 131–159.
[26] I. Ahmad, M. Hussain, A. Alghamdi, et al., Enhancing SVM performance in intrusion detection using optimal feature subset selection based on genetic principal components, Neural Comput. Appl. 24 (7–8) (2014) 1671–1682.
[27] H. Lian, On feature selection with principal component analysis for one-class SVM, Pattern Recognit. Lett. 33 (9) (2012) 1027–1031.
[28] F. Abdat, M. Amouroux, Y. Guermeur, et al., Hybrid feature selection and SVM-based classification for mouse skin precancerous stages diagnosis from bimodal spectroscopy, Opt. Express 20 (1) (2012) 228–244.
[29] F.F. Chen, B.P. Tang, R.X. Chen, A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm, Measurement 46 (1) (2013) 220–232.
[30] I. Aydin, M. Karakose, E. Akin, A multi-objective artificial immune algorithm for parameter optimization in support vector machine, Appl. Soft Comput. 11 (1) (2011) 120–129.
[31] X.L. Zhang, X.F. Chen, Z.J. He, An ACO-based algorithm for parameter optimization of support vector machines, Expert. Syst. Appl. 37 (9) (2010) 6618–6628.
[32] M. Dorigo, Optimization, Learning and Natural Algorithms, Dipartimento di Elettronica, Politecnico di Milano, Italy, 1992 (in Italian).
[33] S.K. Chaharsooghi, A.H.M. Kermani, An effective ant colony optimization algorithm (ACO) for multi-objective resource allocation problem (MORAP), Appl. Math. Comput. 200 (1) (2008) 167–177.
[34] J.F. Tang, Y.Y. Ma, J. Guan, et al., A max–min ant system for the split delivery weighted vehicle routing problem, Expert. Syst. Appl. 40 (18) (2013) 7468–7477.
[35] W. Tfaili, P. Siarry, A new charged ant colony algorithm for continuous dynamic optimization, Appl. Math. Comput. 197 (2) (2008) 604–613.
[36] M.D. Toksari, Ant colony optimization for finding the global minimum, Appl. Math. Comput. 176 (1) (2006) 308–316.
[37] R.J. Mullen, D. Monekosso, S. Barman, et al., A review of ant algorithms, Expert. Syst. Appl. 36 (6) (2009) 9608–9617.
[38] M. Dorigo, M. Birattari, T. Stutzle, Ant colony optimization – artificial ants as a computational intelligence technique, IEEE Comput. Intell. Mag. 1 (4) (2006) 28–39.
[39] M. Dorigo, C. Blum, Ant colony optimization theory: a survey, Theor. Comput. Sci. 344 (2–3) (2005) 243–278.
[40] V.N. Vapnik, Statistical Learning Theory, Wiley & Sons, Inc., New York, 1998.
[41] H.K. Jiang, Z.J. He, C.D. Duan, et al., Gearbox fault diagnosis using adaptive redundant lifting scheme, Mech. Syst. Signal Process. 20 (8) (2006) 1992–2006.
[42] Y. Lei, Z. He, Y. Zi, et al., Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAS, Mech. Syst. Signal Process. 21 (5) (2007) 2280–2294.
[43] W.X. Sun, J. Chen, J.Q. Li, Decision tree and PCA-based fault diagnosis of rotating machinery, Mech. Syst. Signal Process. 21 (3) (2007) 1300–1317.

XiaoLi Zhang received the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, Xi'an, China, 2011. Currently, she is a lecturer at the Department of Mechanical and Electronic Engineering, School of Construction Machinery, Chang'an University, Xi'an, China. She is also a post-doctor at the School of Mechanical Engineering and the State Key Laboratory for Manufacturing Systems Engineering in Xi'an Jiaotong University. Her research interests include artificial intelligence, swarm intelligence algorithm, machinery condition monitoring and fault diagnosis.

Wei Chen received the Master degree from Air Force Engineering University, Xi'an, China, 1997. Currently, he is an vice professor at Aeronautics and Astronautics Engineering College, Air Force Engineering University, Xi'an, China. His research interests include structural strength, vibration and reliability.

BaoJian Wang received the Master degree in mechanical engineering from Northwestern Polytechnical University, Xi'an, China, 2008. Currently, he is an Engineer at the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China. His research interests include testing technology, machinery condition monitoring and fault diagnosis.

XueFeng Chen received the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, 2004. He is a Professor of mechanical engineering department and the State Key Laboratory for Manufacturing Systems Engineering with Xi'an Jiaotong University. His current research interests include finite-element method, machinery condition monitoring and prognostics, aero-engine fault diagnosis and wind turbine system monitoring. Dr. Chen was a recipient of the National Excellent Doctoral Dissertation of China in 2007, the Second Awards of Technology Invention of China in 2009, the China National Funds for Distinguished Young Scientists in 2012 and a chief scientist of the National Key Basic Research Program of China (973 Program) in 2015.