www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

AN INTELLIGENT FEATURE SELECTION AND CLASSIFICATION METHOD BASED ON HYBRID ABC–SVM

Jie Li[1], Qiuwen Zhang*[, 1], Zhang Yongzhi[2], Li Chang[3], Xiao Jian[4]

[1]School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, PR China

[2]PowerChina Hubei Electric Engineering Corporation, Wuhan, Hubei 430040, PR China

[3]Bureau of Science and Technology of Changsha Municipality, Changsha, Hunan 410013, PR China

[4]State Grid Hunan Electric Power Corporation Research Institute, Changsha, Hunan 430007 PR China

[*] Corresponding author: qwzhang@hust.edu.cn, Lijie_linda@126.com

Submitted: July 17, 2016 ***Accepted: Oct.16*** ***Published: Dec.1, 2016***

Abstract- This paper presents a new approach to feature selection and classifcation based on support vector machine and hybrid artificial bee colony. The approach consists of two stages. At the first stage, this paper presented a hybrid artificial bee colony-based classifier model that combines artificial bee colony to improve classification accuracy with the most superior model parameter and features were selected from the original feature set. The classification accuracy and the feature subset provided by the SVM classifier are both considered to update the food source. Finally, the most superior features and optimal model parameter are fed into SVM to identify different class. The testing results verify the effectiveness of the method in extracting feature subset and pattern classification

Index terms*:* Hybrid Artificial Bee Colony; Pattern classification; Feature Selection; Parameter Optimization

I. INTRODUCTION

Pattern recognition is an import problem which could be used in many parts, such as the fault diagnosis, face recognition. A general framework for pattern classification is described as a sequential process involving two main steps: feature selection, and pattern classification.

Many approaches are developed to build classifier, including Bayesian network, neural networks, rough set theory and support vector machines (SVM). Support vector machine (SVM) proposed by Vapnik in 1998 is a relatively new machine learning method based on the statistical learning theory (SLT) [1] and has won great success in pattern classification [2－6].Compared with other similar learning method, SVM shows excellent performance for pattern recognition with small samples.

However, to build a SVM-based classifier model, model parameter optimization is an important issue; due to appropriate model parameter setting can improve the SVM classification accuracy [7]. The parameters that should be optimized include the penalty parameter *C* and the kernel function type and parameters. Nevertheless, the appropriate SVM parameters are very difficult to select due to the lack of the corresponding theoretical basis.

In addition to parameter optimization, feature subset selection is also an important issue. Feature subset selection can show lots of advantage to pattern recognition algorithms [2], such as reducing the measurement cost and storage requirements, handling with the degradation of the classification performance due to the number of training sample sets, reducing training and utilization time, and facilitating data visualization and data understanding. For the sake of above, it is necessary to select the optimal features to identify class information. Feature selection algorithms broadly fall into two categories: filter models and wrapper models [8]. Filter models generally make use of statistical properties of features and are independent of a given learning algorithm. In comparison, wrapper models select the feature subsets according to the accuracy on the training data and are implemented by first defining the learning algorithm. Since a wrapper's search for the best feature subset is guided by prediction accuracy, the results are generally more promising than results based on filters.

Since the feature subset selection and model parameter optimization both have a heavy impact on the classification accuracy and there is an internal relation between feature and model parameter. So, obtaining the optimal feature subset and model parameters must occur simultaneously. In previous literature, some optimization techniques such as the genetic algorithm (GA) [9], simulated annealing (SA) [10], ant colony optimization, (ACO) [11] and particle swarm optimization (PSO) [12] were employed to tune the parameters of SVMs and optimize the input feature subset. But these methods are easy to trap into local optimum and could not get the global solution.

Artificial Bee Colony (ABC) algorithm proposed by Karaboga is based on the food searching behavior of swarm of honey bees [13]. And it has been compared the performance with other algorithms [14] and results show that the performance of the ABC is better than or similar to those of other population-based algorithms with the advantage of employing fewer control parameters. However, the original version of ABC algorithm is only able to optimize continuous problems and would be not appropriate to select optimal feature subset directly. In this paper, we propose a Hybrid ABC (HABC) to solve the problem of feature subset and model parameters optimizing simultaneously. HABC contain two type of ABC: the real-valued type and binary-valued type. The real-valued type is used to optimize the best SVM model parameters, and the binary-valued is used to search the optimal feature subset.

In view of the above analysis, a new intelligent classifier method is developed in this paper. The SVM is employed to build classifier. And HABC is integrated with SVM to identify the optimal classification outcomes.

This paper is organized as follows. Section 1 introduced the multi-class SVM classifier. Section 2 introduces HABC to optimize model parameters and feature subset of SVM and the proposed classifier method. Section 3 details the experimental results and contains a discussion there of results obtained by the proposed method are compared with results of other methods. Finally, a brief conclusion is offered in Section 4.

II. SUPPORT VECTOR MACHINE AND CLASSIFIER

a. Support Vector Machine

Support vector machine (SVM) was derived from the work of Vapnik [1]. The principles of SVM stem from statistical learning theory. By using the information of limited samples, SVM searches for a compromise between the model complexity and learning ability to obtain good generalization ability. SVM has been applied to many field and provided better generalization ability than the conventional methods. In this section, some basic conceptions for SVM are introduced which is also depicted in Figure 1.

Let a training inputs xi (i=1, 2… m) belong to Class 1 or Class 2 and the associated labels yi=1 for Class 1 and yi =-1 for Class2. The data points will be correctly classified by

$$f(x,a) = (w \cdot x) + b \qquad (1)$$

where w is a weight vector orthogonal to the decision surface, b is an offset term.

The original formulation of SVM algorithm seeks a linear decision surface that separates the two opposite classes with a maximal margin, whose solution is found by minimize the constraint optimization problem:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \qquad (2)$$

subject to:

$$\begin{cases} y_i(w \cdot \varphi(x_i)) + b \geq 1 - \xi_i & i = 1,...l \\ \xi_i \geq 0, \quad i = 1,...l \end{cases} \qquad (3)$$

where the coefficient C is a penalty factor, which implements a trade-off between empirical risk and confidence interval. The coefficient εi is a slack factor which is solves a soft margin to solve the linearly inseparable problems.

This optimization problem can be transformed into its corresponding dual problem:

$$L(\alpha) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}a_i a_j y_i y_j (x_i \cdot x_j) \qquad (4)$$

subject to:

$$\begin{cases} a_i, a_j \geq 0, i = 1, \cdots, l \\ \sum_{i=l}^{l} y_i a_i = 0 \end{cases} \qquad (5)$$

where ai are Lagrange multipliers, and only a small subset of ai usually tend to be greater than zero. The respective training vectors having nonzero ai are called support vectors, these vectors are the closest to the optimal hyper-plane geometrically.

The decision function can be got as follows:

$$f(x) = \mathrm{sgn}\left(\sum_{x_i \in SVs} a_i y_i (x_i \cdot x) + b\right) \qquad (6)$$

In practice, most of the problems are linearly inseparable, even though soft margin SVM is adopted. Thus, the input data is mapped into a high dimensional feature space, in which the data are sparse and possibly more separable. Suppose the mapping function is Φ, then the inner product $(x_i \cdot x_j)$ in Eq. (17) can be replaced by $\Phi(x_i)\cdot\Phi(x_j)$. In SVM, a kernel function $K(x_i, x_j)$ is employed to instead Φ(xi)·Φ(xj). Consequently, the decision function becomes:

$$f(x) = \mathrm{sgn}\left(\sum_{x_i \in SVs} a_i y_i K(x_i, x) + b\right) \qquad (7)$$

The generally used kernel functions are reported in Tab.2. One of the most popular kernel functions is radial basis function (RBF) which is employed in this paper.

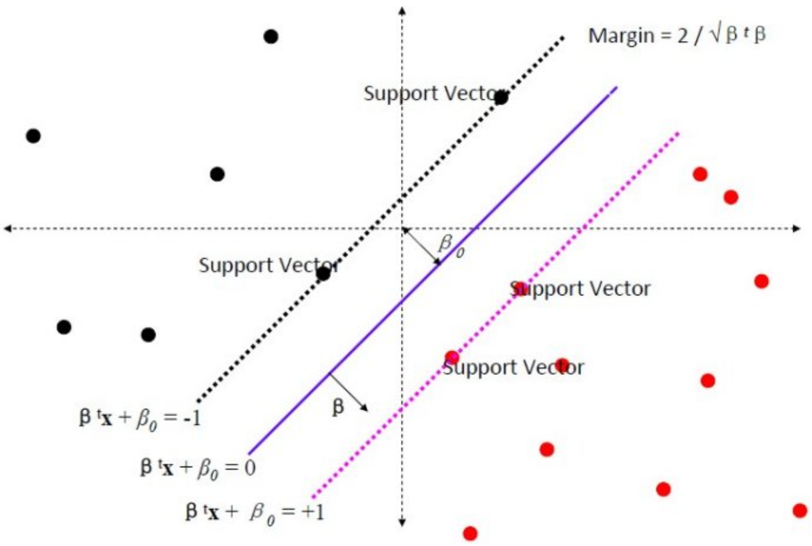

Figure.1 the schematic diagram of SVM

Table 1: the kernel functions

Kernel	Formula
Linear	$K(x,x')=x^T x'$
Polynomial	$K(x,x')=(x^T x'+1)^d$
RBF	$K(x,x')=\exp(-\gamma\|x-x'\|^2)$
Sigmoid	$K(x,x')=\tanh(\gamma x^T x'+\gamma)^d$

b. Multi-class SVM classifier

In the real world, most of fault diagnosis problems are multi-class. Since SVM is essentially binary classifiers, a critical step is the aggregation of the predictions from the ensemble of binary classifiers into a final classification. Several methods have been proposed for extending binary SVM to multi-class such as one-against-one (OAO), one-against-rest (OAR). OAR method is made up with k-SVM models where k is the number of classes. The i-th SVM is trained with all the examples in the i-th class with positive labels, and all the other examples with negative labels. OAO method constructs k (k-1)/2 classifiers, each of which is trained on data from two classes and get the final classification result by using voting strategy. The OAO method outperforms other methods for solving multi-class problems applying the SVM [31]. So, in this paper, we employed the OAO method.

c. The effect of SVM parameter

When the SVM were employed to data mining and pattern classification, the type of kernel function、the parameter of kernel function and penalty factor C should be chosen and tuning. The width parameter σ of the RBF kernel could balance the empirical risk and the classification hyper-plane complexity when the data samples are mapped to high dimensional feature space. And the fine tuning penalty factor C could improve the fitting ability and generalization ability of SVM. To discuss the effects of punishment factor C and width parameter on the performance of SVM, segment dataset of UCI is used to prove how the correct rate of cross validation are effected by the combination of the parameters of the support vector machine.

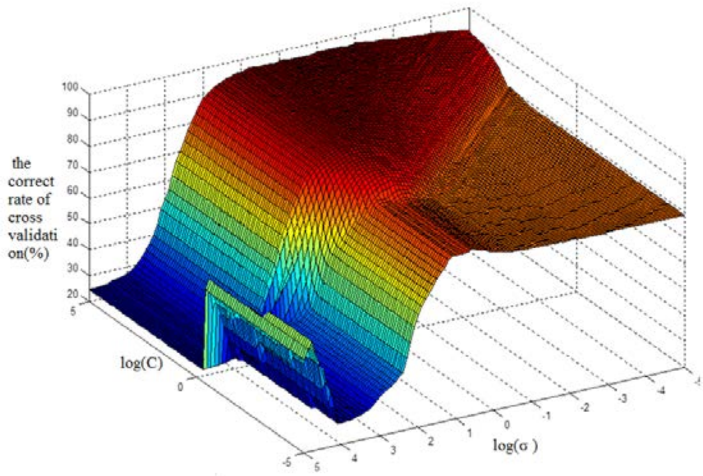

Figure. 2 The effect of SVM’s parameter

As is shown in the Figure 2, a multi peak function mapping relationship exists among the penalty factor of SVM , the parameters of kernel function and classification accuray, and the traditional optimization algorithm for SVM parameters optimization are easy to fall into local optimum and unable to obtain the optimal SVM classification model. Therefore, the better algorithm should be proposed in this paper.

III. HYBIRD ARTIFICIAL BEE COLONY AND THE PROPOSED DIAGNOSIS MODEL

a. Reviews of artificial bee colony

Artificial Bee Colony (ABC) algorithm is a novel heuristic approach introduced by Karaboga for numerical function optimization problem by simulating the foraging behaviors of honey bee swarm. In ABC algorithm, the potential solution is represented as the position of a food source and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. The frameworks of ABC algorithm [17, 18] can be described in Figure 3. The employed bee stage is the mutation mechanism. Given a solution xi to be updated which denotes the i-th solution in the population, and let vi = xi. In the employed bee stage, a new candidate solution is firstly given by the following solution search equation:

$$x_{new(j)} = x_{ij} + (x_{ij} - x_{kj}) \times r \tag{8}$$

where x_{ij} (or v_{ij}) denotes the *j*-th element of x_i (or v_i), and *j* is a random index. x_k denotes another solution selected randomly from the population. And r_{ij} is a uniform random number in [-1, 1]. Then, a greedy selection is done between x_i and v_i, which completes the update process.

The onlooker stage is the solutions selection mechanism using the roulette strategy like GA. The main distinction between the employed bee stage and the onlooker stage is that every solution in the employed bee stage involves the update process, while only the selected solutions have the opportunity to update in the onlooker stage.

Finally, the scout stage disposes an inactive solution which does not change over a certain number of generations and replace it by a new randomly generated solution.

Initialization

Repeat

Employed bee stage: Perform an update process for each solution in the solution population.

Onlooker stage: Randomly select solutions depending on their fitness values, then perform the same update process for each selected solution

Scout stage: Select one of the most inactive solutions, then replace it by a new randomly generated solution.

Until (conditions are satisfied)

Figure 3. The Framework of ABC

b. Hybrid artificial bee colony

(1) Initialization of the population

HABC is designed to construct the optimal SVM diagnosis model by optimizing the feature subset and the SVM model parameters; a food source represents a combination of the feature subset and the SVM parameters, C and γ, based on the radial basis function (RBF) kernel of the SVM classifier.

Firstly, the number of food sources (*SN*), number of trials and the maximum cycle number (*MCN*) should be set. The *SN* is equal to the number of the employed bees or onlooker bees.

The number of trials is the number after which a food source is assumed to be abandoned (limit) which is set to 1 as same as the original ABC algorithm.

Secondly, HABC generates a randomly distributed initial population $P(C=0)$ of SN solutions (food source positions). Aiming at the feature selection and parameter optimization of SVM The real-type part of food source is generated as follows:

$$p_{ij} = LB_j + (UB_j - LB_j) \times r \tag{9}$$

The binary-type part of food source is generated as follows:

$$f_{ij}(t+1) = \begin{cases} 1 & if: r < 0.5) \\ 0 & if: r \geq 0.5) \end{cases} \tag{10}$$

The food source could be expressed as equation (22):

$$x_{ij} = (p_1, \ldots, p_m, f_1, \ldots f_n,) \tag{11}$$

where p_i denotes the candidate solutions of SVM model parameters which is coded by real-type. f_j denotes the value of feature mask—"1'' represents that feature j is selected and ''0'' represents that feature j is not selected, which is coded by discrete-type. x_{ij} denotes the ith food source in the population of the jth dimension, LB and UB denotes the lower bound and upper bound of the jth dimension, respectively and r is a uniformly distributed real number in [0,1].

(2) Employed bee phase

In this phase, each employed bee xij generates a candidate food source x_{new} in the neighborhood of its present position as same as the original ABC. To determine the position of food source as 0 or 1, the binary-type part of food source changed in employed bee phase, onlooker bee phase or scout bee phase using the following formula:

$$x_{ij}(t+1) = \begin{cases} 1 & if: r() < L(x_{ij}(t+1)) \\ 0 & otherwise \end{cases} \tag{12}$$

$$L(x_{ij}(t+1)) = \frac{X_i^{k+1} - X_{\min}}{X_{\max} - X_{\min}} \tag{13}$$

where $L(\cdot)$ is a linear function with an output value belongs to [0,1]; r is the uniform random number in [0, 1], and $[X_{min}, X_{max}]$ is a predefined range for gaining the probability value with $L(\cdot)$ function. x_{ij} is a binary value mapping from the associated position vector X_i through function $L(\cdot)$

Once x_{new} is obtained, a greedy selection mechanism is employed between x_i and x_{new}, that is, x_{new} will replace x_i and become a new member of the population if the fitness of x_{new} is equal to or better than that of x_i.

In this paper, fitness function is consist of two parts—classification accuracy and the number of selected features. Thus, for the food source with a high fitness value has high classification accuracy and an efficient features subset. The objective fitness function is followed as equation:

$$fitness_i = w_A \times cv_i + w_F \times \left[1 - \frac{\left(\sum_{j=1}^{n_F} f_i\right)}{n_F}\right] \qquad (14)$$

where w_A and w_F are two predefined weights for the classification accuracy and the selected feature respectively. And n_F is the total number of features. cv_i denotes the k -fold cross validation accuracy.

(3) Onlooker bee phase

An onlooker bee evaluates the nectar information taken from all the employed bees and selects a food source x_i depending on its probability value p_i calculated by the following expression:

$$p_i = \frac{f_i}{\sum_{i=1}^{SN} f_i} \qquad (15)$$

where f_i is the nectar amount (i.e., the fitness value) of the i-th food source x_i. Obviously, the food source with the higher nectar amount can be selected much possibly.

Once the onlooker has selected its food source x_i, it produces a modification on x_i by using Eq. (12). As in the case of the employed bees, if the modified food source has a better or equal nectar amount than x_i, the modified food source will replace x_i and become a new member in the population.

(4) Scout bee phase

In scout bee phase, if a food source x_i cannot be further improved through a predetermined number of trials limit, the food source is assumed to be abandoned, and the corresponding employed bee becomes a scout. The scout produces a food source as the same way as the initialization of the population.

c. The proposed classification model

An intelligent classification method is proposed in this paper. The feature subset and the kernel parameters of SVM are dynamically optimized by implementing the HABC evolutionary process and then the SVM model performs the classification task using the optimal parameter values and feature subset. The classification steps are described as follows:

Step1: Extract the feature subset from the original data.Step2: Adopt SVM with HABC to optimize parameters and feature subset using the one against others strategy and to solve the multi-class classification problem.

Step3: The final output of HABC-SVM could be considered as the global parameters and feature and carried out to train fault classification model.

Step4: Classify the unknown fault samples using the trained SVM model with the optimal features.

The flow chart is shown in Figure 4

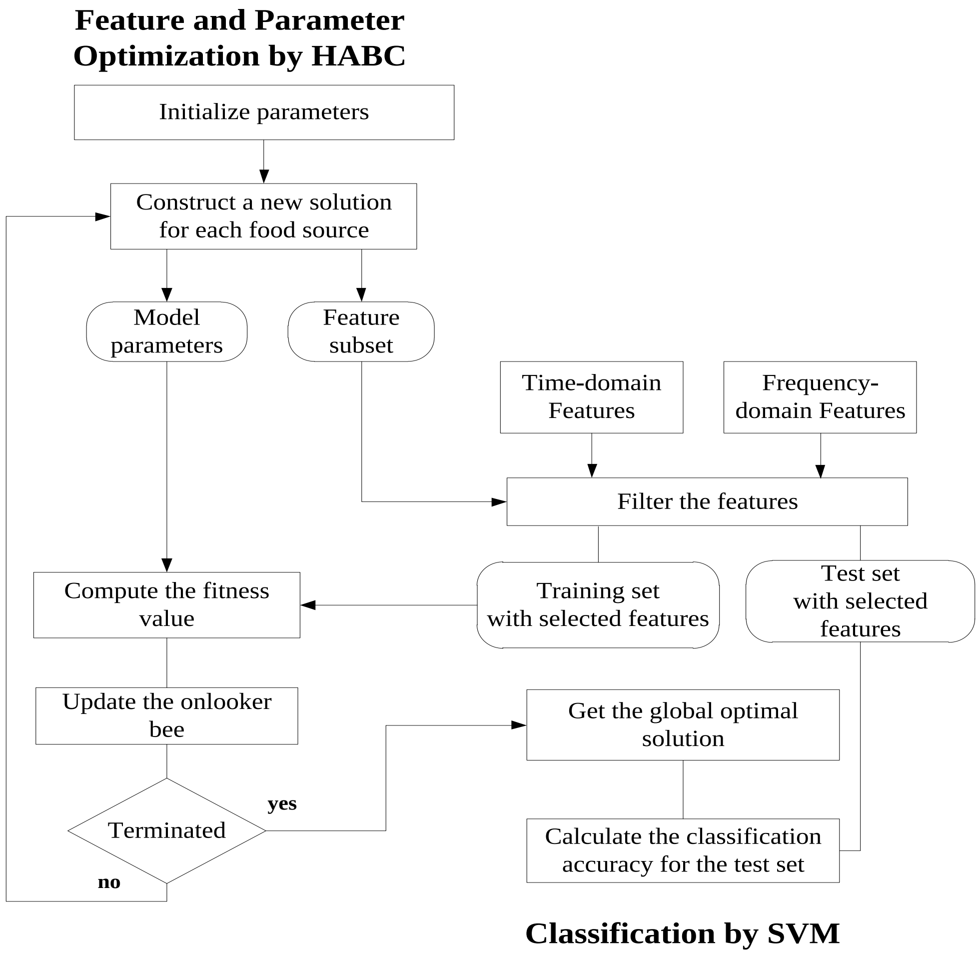

Figure 4. The flow chart of proposed method

IV. EXPERIMENTAL RESULTS AND ANALYSIS

To demonstrate the effectiveness of the proposed intelligent classification method in this paper, we applied it to fault classification of rolling element bearings. All of the experiments were implemented in Pentium IV 2.8 GHz CPU with 1G RAM using VC++ 2010 compiler and Matlab2010a simulation software based on Windows XP operating system. The LibSVM (version3.12) originally designed by [18] is used as the SVM classifier.

a. Experimental Description

The vibration signals of rolling element bearings are provided by the CWRU bearing data center as shown in Fig. 5[19]. The smallest fault diameter has been selected for this study with shaft rotational speed of 1730 rpm (with no motor load condition), and for each of the four operating conditions, 40 vibration signals with the length of 2048 points were collected. Thus, this is a four-class classification problem to identify different fault categories. The original signal waveform and frequency form can be described in Figure 6.

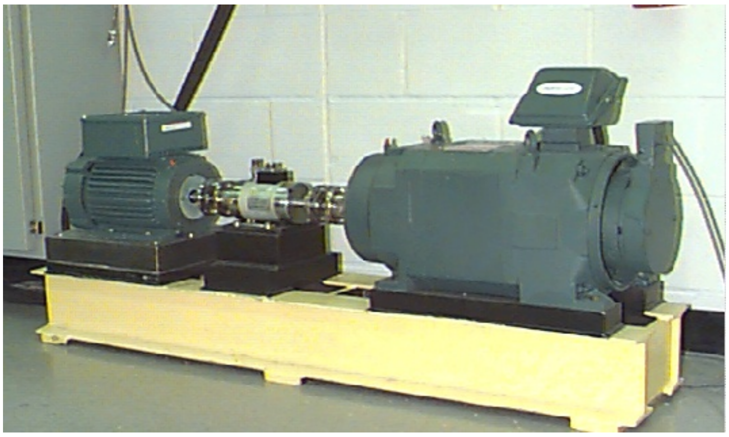

Figure 5. Experimental test stand.

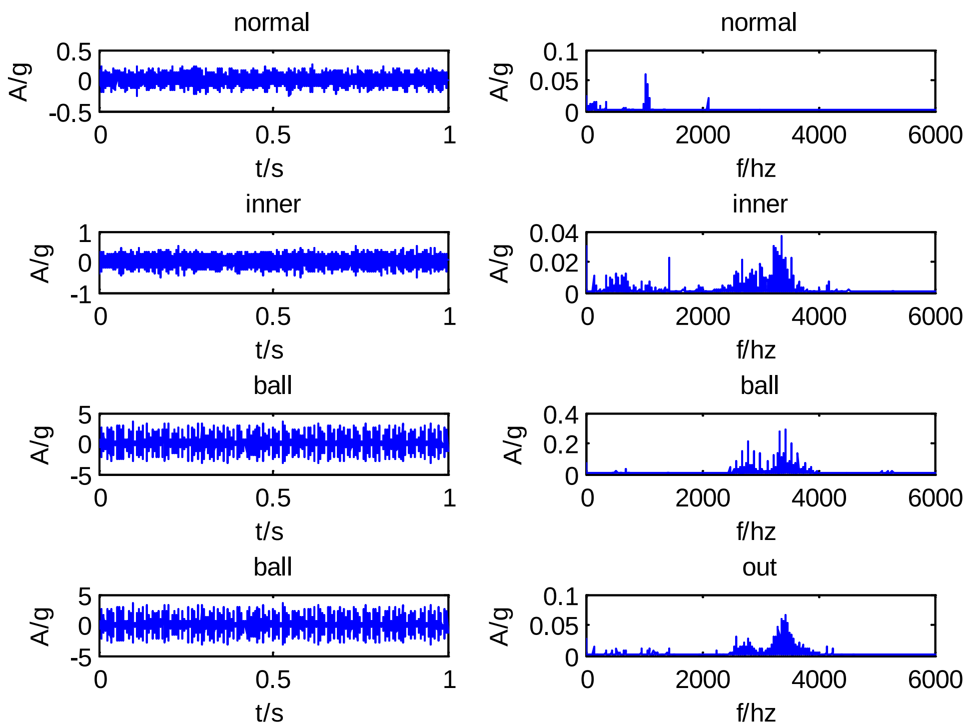

Figure 6. Original signal waveform and frequency form

b. Feature definition

When the faults occur in rotating machinery, the raw signal may change; it is also showed in some PFs modulated by raw signal using LMD. So amplitude and distribution, as well as the frequency spectrum and its distribution of these PFs, may be different from those under normal condition.

In this work, we can utilize the LMD to decompose the vibration signal, and extracted time domain and frequency domain statistical characteristics of first several PFs which containing almost all valid information were selected. And the features parameters exact from the PFs are definite in Tab.2. The parameters above are effective and practical in fault diagnosis of rotating machinery due to their relative sensitivity to early faults, and robustness to various loads and speeds. Ten parameters (TF1–TF10) are time-domain statistical characteristics, and the other ten parameters (FF1–FF10) are frequency-domain statistical characteristics. Parameter TF1–TF6 may reflect the vibration amplitude and energy in time domain. Parameter TF7–TF10 may represent the time series distribution of the signal in time domain. Parameter FF1–FF4 may indicate the vibration energy in frequency domain. Parameter FF1 may describe vibration energy in frequency domain. Parameter FF7–FF10 may describe the convergence of the spectrum power. Parameter FF2–FF6 may show the position change of the main frequencies.

Table2 the feature parameter

Time-Domain		Frequency-Domain	
Feature	Equation	Feature	Equation
Mean	$TF_1 = \frac{1}{N}\sum_{n=1}^{N} x(n)$	Mean frequency	$FF_1 = \frac{\sum_{k=1}^{K} s(k)}{K}$
Standard deviation	$TF_2 = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[x(n) - TF_1\right]^2}$	Frequency centre	$FF_2 = \frac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)}$
Root mean square	$TF_3 = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x^2(n)}$	Root mean square frequency	$FF_3 = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} s(k)}}$
Peak	$TF_4 = \max\lvert x(n)\rvert$	Standard deviation frequency	$FF_4 = \sqrt{\frac{\sum_{k=1}^{K} (f_k - FF_2)^2 s(k)}{K}}$
Skewness	$TF_5 = \frac{N}{(N-1)(N-2)}\sum_{n=1}^{N}(\frac{x(n) - TF_1}{TF_2})^3$	Mean frequency that cross the mean of time-domain signal	$FF_5 = \sqrt{\frac{\sum_{k=1}^{K} f_k^4 s(k)}{\sum_{k=1}^{K} f_k^2 s(k)}}$
Kurtosis	$TF_6 = \left\{\frac{N(N+1)}{(N-1)(N-2)(N-3)}\right\}\sum_{n=1}^{N}\left(\frac{(x(n) - TF_1)}{TF_2}\right)^4 - \frac{3(N-1)^2}{(N-2)(N-3)}$	Stabilization factor	$FF_6 = \sqrt{\frac{\sum_{k=1}^{K} f_k^2 s(k)}{\sum_{k=1}^{K} s(k)\sum_{k=1}^{K} f_k^4 s(k)}}$
Crest factor	$TF_7 = \frac{TF_4}{TF_3}$	Coefficient of variability	$FF_7 = \frac{FF_4}{FF_1}$
Clearance factor	$TF_8 = \frac{TF_4}{\left(\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert\right)^2}$	Frequency-domain Skewness	$FF_8 = \frac{\sum_{k=1}^{K} (f_k - FF_1)^3 s(k)}{KFF_4^3}$
Shape factor	$TF_9 = \frac{TF_3}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$	Frequency-domain kurtosis	$FF_9 = \frac{\sum_{k=1}^{K} (f_k - FF_5)^4 s(k)}{KFF_6^4}$
Impulse factor	$TF_{10} = \frac{TF_4}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$	Root-mean-square ratio	$FF_{10} = \frac{\sum_{k=1}^{K} (f_k - FF_5)^{1/2} s(k)}{K\sqrt{FF_6}}$
where $x(n)$ is a signal series for n = 1,2,. . .,N, N is the number of data points		where $s(k)$ is a spectrum for k = 1,2,. . .,K, K is the number of spectrum lines; f_k is the frequency value of the k-th spectrum line	

b. Performance comparison with applying PSO、GA algorithm

After the feature extraction, the 40 samples of each condition are split into two sets randomly: 20 samples for training and 20 samples for testing. Applying the proposed method to fault diagnosis of the bearings, the feature set with 60 feature values and the SVM parameter is optimized simultaneously by using HABC-SVM and 10 runs were conducted by 10-fold cross-validation with the rolling bearing dataset to obtain a more reliable result. The results obtained by the proposed HABC–SVM were compared with those of traditional GA-SVM and PSO-SVM. For the method HABC-SVM, the HABC algorithm is directly used to search the optimal parameters (C, γ) in [10-1~103] and [10-1~103] for PSO-SVM and GA-SVM are the same as that of HABC-SVM.

In order to verify the effect of feature selection of HABC-SVM, a principal component analysis (PCA), have also been applied to analyze the selected feature subset and original feature. The comparison results are plotted in Fig.7~8 and we can get the conclusion that the selected feature subset has the better diverted and can be distinguish the four type condition.

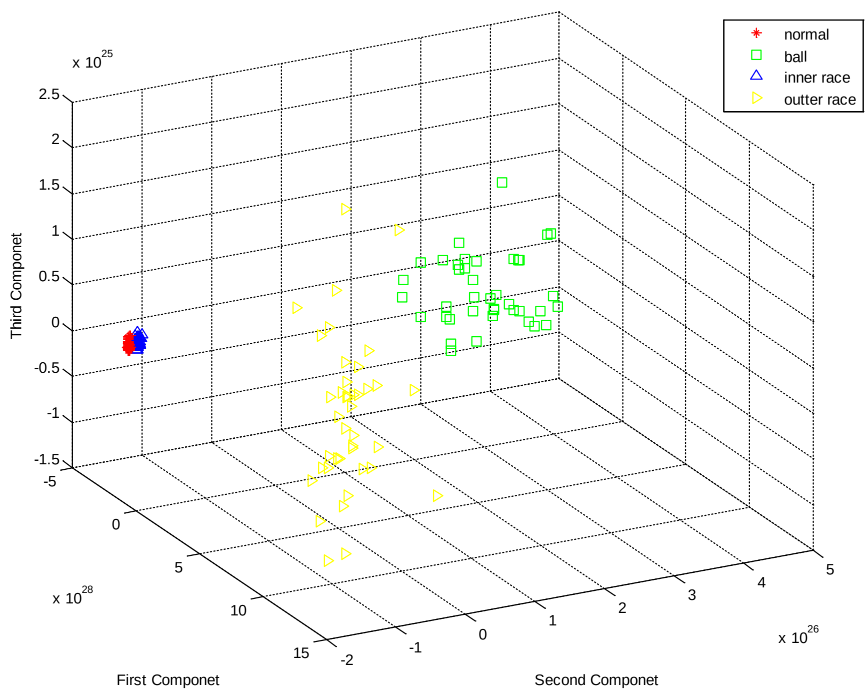

Figure 7. the principal component analysis with all features

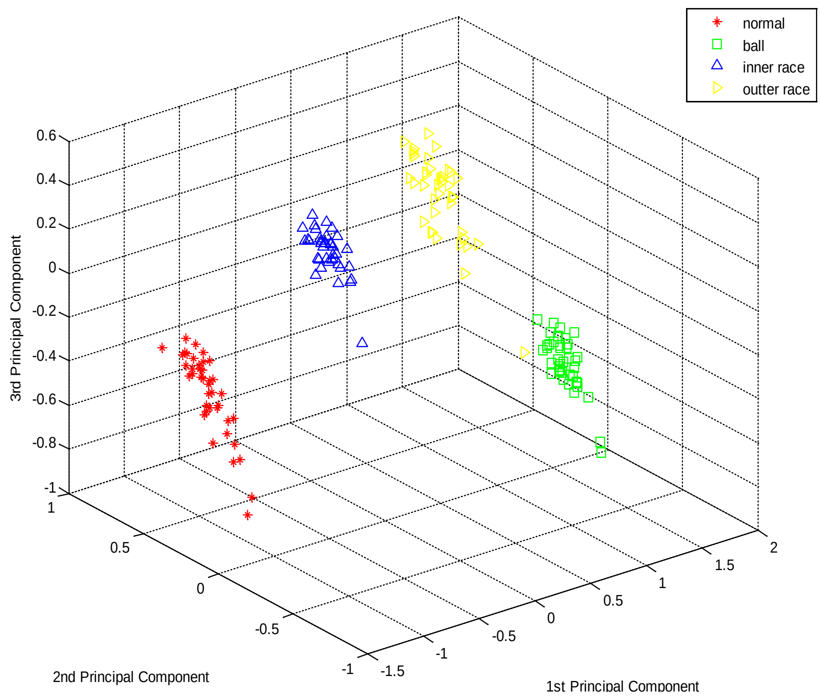

Figure 8. the principal component analysis with selected features

Table 3 presents the average accuracy of the training set. The cross-validation rate was 95.00% for the PSO–SVM, 89.19% for the GA–SVM, and 99.97% for the HABC–SVM. And the classification accuracy of the HABC-SVMs proves to be superior to that of the SVM with the optimal parameters and feature selected by GA、PSO.

Table 3: the computational results of different algorithm

	C	g	Number of selected feature	Cross-validation rate (%)	Testing accuracy rates	SV number
HABC-SVM	6.31	6.49	18	99.97	98.75%	93
PSO-SVM	79.56	18.183	9	95.00	91.25%	86
GA-SVM	236.97	508.09	10	89.19	85.00%	80

After the training, the testing samples are utilized to evaluate the generalization performance of the proposed method. The testing results are also depicted in Table 3 and Figure 9. These results suggest that the proposed method HABC-SVM yields the best testing rate in these methods.

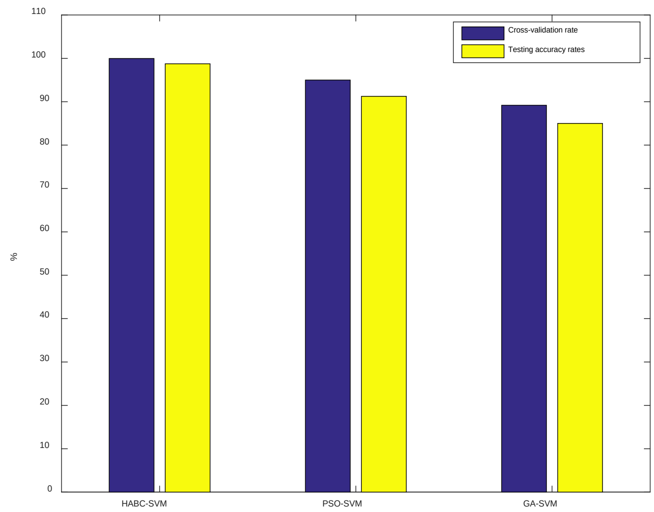

Figure 9. The performance comparison between HABC and other algorithm

V. CONLUSION

In this paper, an intelligent classification method is proposed in this paper. In order to improve classification accuracy with the most superior features are selected from the original feature set, this paper presents a hybrid ABC-based classifier model that combines artificial bee colony (ABC) and support vector machines (SVM) to improve classification accuracy with the most superior features are selected from the original feature set. The proposed method in this paper is applied to pattern classification of rolling element bearings fault and the experimental results indicate using the most superior features, the proposed approach enables the detection of abnormalities pattern and at the same time identification of the category and severity of pattern with a high accuracy.

Acknowledgements

This study is supported by the Science and Technology Department "soft science" key project of Hunan Province in China (2013ZK2038) , the National Natural Science Foundation of China (41672263), the Natural Science Foundation of Hubei Province in China (2015CFA134), andEducational Department scientific research project of Hunan Province in China(13C515).

REFERENCES

[1]Vapnik, Vladimir N. The Nature of Statistical Learning Theory [M]// The nature of statistical learning theory /. Springer, 2000:988 - 999.

[2] Zhang X, Zhou J, Wang C, et al. Multi-class support vector machine optimized by inter-cluster distance and self-adaptive deferential evolution. Applied Mathematics & Computation, 2012, 218(9):4973-4987.

[3]Shen Z, Chen X, Zhang X, et al. A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM Measurement, 2012, 45(1):30-40.

[4]Chen Z, Li J, Wei L. A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue . Artificial Intelligence in Medicine, 2007, 41(2):161-175.

[5]Widodo A, Yang B S. Wavelet support vector machine for induction machine fault diagnosis based on transient current signal . Expert Systems with Applications, 2008, 35(1-2):307-316.

[6]Xiang X, Zhou J, An X, et al. Fault diagnosis based on Walsh transforms and support vector machine . Mechanical Systems & Signal Processing, 2008, 22(7):1685-1693.

[7]Gao X, Liu F. Robust visual tracking based on support vector machine and weighted sampling method . International Journal on Smart Sensing & Intelligent Systems, 2015, 8(1):255-271.

[8]Zhang K, Li Y, Scarf P, et al. Feature selection for high-dimensional machinery fault diagnosis data using multiple models and Radial Basis Function networks Neurocomputing, 2011, 74(17):2941-2952.

[9] Anne E, Alip B, Stephen B, et al. Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm-support vector machines: HGASVM. Expert Systems with Applications, 2009, 36(2):1391-1402.

[10]Lin S W, Lee Z J, Chen S C, et al. Parameter determination of support vector machine and feature selection using simulated annealing approach Applied Soft Computing, 2008, 8(4):1505-1512.

[11]Huang C L. ACO-based hybrid classification system with feature subset selection and model parameters optimization Neurocomputing, 2009, 73(1–3):438-448.

[12] Widiyanto D, Purnomo D M J, Jati G, et al. Modification of particle swarm optimization by reforming global best term to accelerate the searching of odor sources. International Journal on Smart Sensing & Intelligent Systems, 2016, 9(3):1412-1430.

[13]Karaboga D, Basturk B. On the performance of artificial bee colony (ABC) algorithm Applied Soft Computing, 2008, 8(1):687-697.

[14] Karaboga D, Akay B. A comparative study of Artificial Bee Colony algorithm Applied Mathematics & Computation, 2009, 214(1):108-132.

[15]Lei Y, He Z, Zi Y. A new approach to intelligent fault diagnosis of rotating machinery Expert Systems with Applications, 2008, 35(4):1593-1600.

[16] Hsu C W, Lin C J. Errata to "A comparison of methods for multiclass support vector machines".. IEEE Transactions on Neural Networks, 2002, 13(4):415-425.
[17] Xu C, Duan H, Liu F. Chaotic artificial bee colony approach to Uninhabited Combat Air Vehicle (UCAV) path planning . Aerospace Science & Technology, 2010, 14(8):535-541.
[18]C.-W. Hsu, C.-C. Chang, C.-J. Lin. “A Practical Guide to Support Vector Classification”, <http://www.csie.ntu.edu.tw/_cjlin/libsvm/>, 2009.
[19]K. A. Loparo, “Bearings Vibration Data Set”, Case Western Reserve University [EB/OL]<http://www.eecs.case.edu/laboratory/bearing/welcome_over view.htm>