Random forests classifier for machine fault diagnosis

Bo-Suk Yang[1,*], Xiao Di[1] and Tian Han[2]

[1]*School of Mechanical Engineering, Pukyong National University, San 100, Yongdang-dong, Nam-gu, Busan 608-739, South Korea*
[2]*School of Mechanical Engineering, University of Science and Technology Beijing,*
30 Xueyuan Road, Haidian District, 100083, Beijing, China

Abstract

This paper investigates the possibilities of applying the random forests algorithm (RF) in machine fault diagnosis, and proposes a hybrid method combined with genetic algorithm to improve the classification accuracy. The proposed method is based on RF, a novel ensemble classifier which builds a number of decision trees to improve the single tree classifier. Although there are several existing techniques for faults diagnosis, the application research on RF is meaningful and necessary because of its fast execution speed, the characteristics of tree classifier, and high performance in machine faults diagnosis. The proposed method is demonstrated by a case study on induction motor fault diagnosis. Experimental results indicate the validity and reliability of RF-based diagnosis method.

Keywords: Random forests algorithm; Genetic algorithm; Machine learning; Fault diagnosis; Rotating machinery

---

1. Introduction

Rotating machinery plays an important role in modern industry, and is equipped in many crucial departments. A potential problem is that its breakdown will result in a huge loss. Therefore, fault diagnosis of machines is gaining importance in industry because of its capability to increase reliability and to decrease possible loss of production. Efficient and accurate fault classification is critical to machinery operating in normal condition.

In recent years, a number of classifiers have been reported in machine fault diagnosis such as artificial neural networks (ANNs), support vector machines (SVMs), rule-based induction, case-based reasoning, etc. The proposed techniques and their extended research increase the intelligence, precision and applicability in the diagnosis domain. While the passion for developing fault diagnosis methods is increasing, a number of obstacles still exist:

- Different kinds of faults may result in a certain symptom.
- Because of the background noise, some faults are difficult to distinguish in the machine.
- There are a number of subassemblies with rotating machinery and a high level internal interaction between these subassemblies such as bearings, rotor, etc.

Hence, the machine fault diagnosis method which is employed to make hypotheses should be powerful enough to classify the malfunctions in a correct way. The critical issue in classification is how to integrate the classification power to achieve higher classification accuracy. To do that, improving the capability of diagnosis is the main motivation to inspire researchers synchronizing existent technologies and exploring new theories.

Ensemble classification methods train several classifiers and combine the decision of a set of classifiers by weighted or unweighted voting process to classify unknown examples. An ensemble classifier is generally found to be more accurate than any of the individual classifiers making up the ensemble [1]. The

*Corresponding author. Tel.: +82 51 629 6152, Fax.: +82 51 629 6150
E-mail address: bsyang@pknu.ac.kr

most widely used ensemble methods are *bagging* (an acronym of *bootstrap aggregating*) [2] and *boosting* [3]. Bagging is based on training many classifiers on bootstrap samples from the training set, which has been shown to reduce the variance of the classification [4]. Boosting was introduced by Schapire [5] as a method for boosting the performance of a weak learning algorithm. Boosting uses iterative re-training, where the incorrectly classified samples are given increased weighting as the iteration progresses. Therefore, it generally reduces both the variance and the bias of the classification and has been shown to be a very accurate classification method [6]. However, it has various drawbacks: it is very slow, it can over-train and it is sensitive to noise [7].

Random forests algorithm (RF), which was introduced by Breiman [8], is a general term for ensemble methods using tree-type classifiers. RF builds a large number of decision trees [9, 10] out of a sub-dataset from a unique original training set by using bagging, which is a meta-algorithm to improve classification, and regression models according to stability and classification accuracy. Bagging reduces variance and helps to avoid over-fitting synchronously. This procedure extracts cases randomly from original training data set and the bootstrap sets are used to construct each of the decision trees in the RF. Each tree classifier is named a component predictor. The RF makes decisions by counting the votes of component predictors on each class and then selecting the winner class in terms of the number of votes to it [11].

RF has been employed in various fields such as land cover [12], drug discovery [13] and geographic data [4, 14]. The possibility of using RF in machine fault diagnosis application is considered [15]. RF provides good performance in applications in these fields and can be a competitor for rotating machinery fault diagnosis, because of these distinctive features [8]:

- It is unexcelled in accuracy among current algorithms.
- It runs efficiently on large data bases.
- It can estimate the importance of each variable in the classification.
- It has methods for estimating missing data and maintains accuracy when a large proportion of the data are missing.
- It computes proximities between pairs of cases that can be used in clustering, locating outliers, or five interesting views of the data.

- It generates an internal unbiased estimate of the generalization error as the forest building progresses.

In this paper, we confirm the possibilities of using RF in machine fault diagnosis and propose an optimized RF method combined with genetic algorithm (GA) to improve the classification accuracy. To increase the diagnosis accuracy, we acquire the data of three-direction vibration signals as the original inputs of the system. And a number of feature parameters in time and frequency domains and regression coefficients are calculated to extract helpful information and remove the background noise of the data. Then RF diagnosis system detects certain faulty type based on these features. It is an effective approach to promote the capability of the diagnosis system [16]. Experimental results show the optimized RF-based method achieves a very high accuracy by combining RF with GA.

2. Random forest

RF, which is derived from decision tree classifier is an assembled method; it grows trees using CART (classification and regression trees) methodology to maximum size and without pruning. Fig. 1 shows the construction of an RF. This section gives a brief description of the building blocks for CART-based RF (RF-CART) explored in this paper.

2.1 CART methodology

CART [17] grows classification and regression trees to predict continuous dependent variables (regression) and categorical predictor variables (classification). There are four basic steps in CART method-
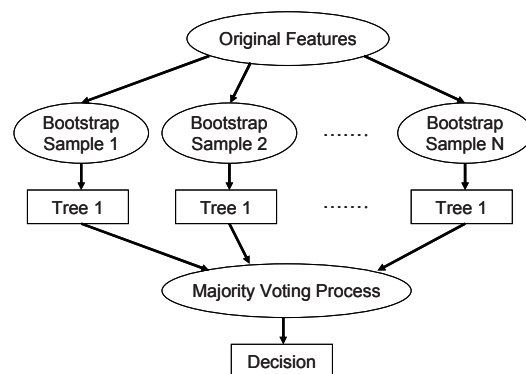

Fig. 1. Construction of random forest.

ology. At the first step, a tree is built by using recursive splitting of nodes. Each terminal node is assigned to a certain class, judged by class probability distribution of the dependent variable at terminal node. The assignment of a predicted class to each node occurs whether or not that node is subsequently split into child nodes. The second step consists of stopping the tree growing process. The final two steps, tree pruning and optimal tree selection, are ignored because RF grows trees freely without any pruning process.

2.1.1 Tree building

The tree building process begins with departing the root node into binary nodes by a very simple question of the form, $\mathbf{x} \leq d$? Here $\mathbf{x}$ is variables in the data set and d is a real number. Initially, all observations are located in the root node. CART implements a computer-intensive algorithm that searches for the best split at all possible split points for each variable. The methodology which CART uses for building trees is known as binary recursive partitioning. Adopting the Gini diversity index as a splitting rule, the tree building process is as follows:

Step 1: CART splits the first variable at all of its possible split points, at all of the values the variable assumes in the sample. At each possible split point of a variable, the sample splits into binary or two child nodes. Cases with a "yes" response to the question posed are sent to the left node and those with "no" responses are sent to the right node.

Step 2: CART then applies its goodness-of-split criteria to each split point and evaluates the reduction in impurity that is achieved by using the formula:

$$\Delta i(s,t) = i(t) - p_L \left[i(t_L) \right] - p_R \left[i(t_R) \right] \quad (1)$$

where s is a particular split, p_L is the proportion of the observations at node t which go into the left child node t_L, p_R is for the right node t_R similar with p_L. $i(t_L)$ and $i(t_R)$ are impurity of left and right nodes respectively.

Step 3: CART selects the best split of the variable as that split for which the reduction in impurity is highest. Three steps above are repeated for each of the remaining variables at the root node.

Step 4: CART then ranks all of the best splits on each variable according to the reduction in impurity achieved by each split and selects the variable and its split point that most reduces the impurity of the root or parent node.

Step 5: CART then assigns classes to these nodes according to the rule that minimizes mis-classification costs. CART has a built-in algorithm that takes into account user-defined variable misclassification costs during the splitting process. The default is unit or equal misclassification costs.

Because the CART procedure is recursive, steps 1 - 5 are repeatedly applied to each non-terminal child node at each successive stage.

2.1.2 Stopping tree building

CART stops the splitting process when:
- There is only one observation in each of the child nodes;
- All observations within each child node have the identical distribution of predictor variables, making splitting impossible.
- The user sets an external limit on the number of levels in the maximal tree previously.

Standing by these steps, a CART algorithm-based decision tree without pruning and optimizing will be built.

2.2 Random forests algorithm (RF)

RF can improve classification accuracy resulting from growing an ensemble of trees and making them vote for the most promising class. A convenient method to build the ensembles is by random vectors which are generated via random selection procedure from integrated training set. The constituent in this method is that we prepare k random vectors, Θ_k, which are independent of the past random vectors Θ_1, Θ_2, Θ_3, …, Θ_{k-1} but with the same distribution to build the trees among the RF. The corresponding individual classifier is noted by $C(\mathbf{X}, \Theta_k)$. For example, in the bagging processing the random vector Θ as the N observations randomly draws out from entire training data proportionally where N is the number of observations of training data. And then they vote for the most popular class. Breiman names these procedures as RFs. A definition drawn from original paper is available here [8].

Definition 1 An RF is a classifier consisting of a collection of tree structured classifiers $\{C(\mathbf{X}, \Theta_k), k = 1, …\}$ where the Θ_k is independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input $\mathbf{X}$.

2.2.1 Two randomized procedures in RF tree building

As mentioned above, RF enhances the classification accuracy compared with decision tree classifier significantly. It is the reason that RF applies two randomized procedures when it builds trees. Each tree is built as follows. Firstly, assume that the number of cases in the training set is N and the number of variables in the classifier is M. Select the number of input variables that will be used to determine the decision at a node of the tree. This number, m, should be much less than $M(m << M)$. Secondly, choose a training set by choosing N samples from the training set with replacement. And then, for each node of the tree randomly select m of the M variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set. Finally, each tree is fully grown and not pruned.

Two distinctive randomized procedures exist among the four steps below. That is, RF extracts a fixed quantity from a training set randomly, or names it bagging process [2]. Each base classifier in the ensemble is trained on a bootstrap from the entirety of available data. However, each of these bootstrap replicates tends to leave out roughly one-third of the sample. Each classifier in the ensemble is thus trained on roughly two-thirds of the original data. Consequently, each element in the sample of size n trains roughly $(2/3)k$ of all classifiers in the ensemble so that it can be used to validate the remaining $k/3$ classifiers (Fig. 2) where n is the number of training data, k is the total number of single tree classifier. This part of data is named *out-of-bag data* to get an unbiased estimate of the test set error of an individual tree. The rest of the data is used to build the single tree classifier.
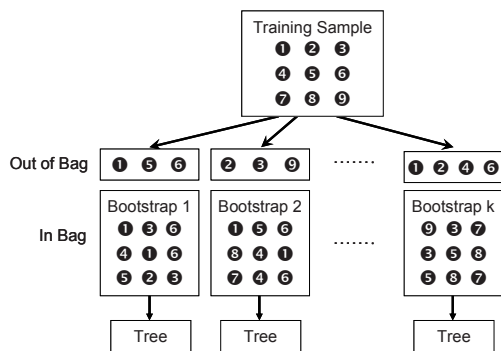
After the bagging processing, the other randomized procedure is appeared in node splitting during the tree classifier being built. Different from normal CART-like decision tree splitting algorithm, CART within RF algorithm searches only in n variables which are small number and drawn at random from all M variables instead of entire variables.

The research of Breiman states why these two randomized procedures make classification accuracy increase effectively: Improvement will occur for unstable procedures where a small change in training set can result in large change between component classifiers and classifier trained by entire training set. In RF, whatever the bagging processing or the random selection of variables to split the node both make difference in individual tree and forests. Therefore, these two sources of randomness are most important features of RF.

2.2.2 Convergence of RF

RF adopts an ensemble of decision trees and determines the categorical classes by majority vote algorithm. Thus, a serious consideration of over-fitting is necessary for testing RF performance. Normally an over-fitting will occur where learning is performed for too long or where training examples are rare; the learner may be limited in very specific random features of the training data that has no causal relation with the target function. But RF can avoid the overfitting completely [2]. To affirm this point, we define a margin function first.

Given an ensemble of a series of classifiers $C_1(\mathbf{X})$, $C_2(\mathbf{X})$, …, $C_k(\mathbf{X})$, and with the training set drawn at random from the distribution of the random vector Y, $\mathbf{X}$, define the margin function as

$$mg(\mathbf{X}, Y) = av_k I\left(C_k(\mathbf{X}) = Y\right) - \max_{j \neq Y} av_k I\left(C_k(\mathbf{X}) = j\right) \tag{2}$$

where $\mathbf{X}$ is input metric, av_k is the average number of votes at $\mathbf{X}$, Y for the corresponding class and $I(\cdot)$ is the indicator function. The margin measures the extent to which the average number of votes at $\mathbf{X}$, Y for the right class exceeds the average vote for any other class. The larger the margin, the more confidence in the classification.

According to this function, the generalization error is given by:

Fig. 2. Schematic of bagging using the decision tree as the base classifier.

$$PE^* = P_{\mathbf{X},Y}\left(mg\left(\mathbf{X},Y\right)<0\right) \tag{3}$$

where $P_{\mathbf{X},Y}$ indicates the probability which is over the $\mathbf{X}$, Y space.

Theorem 1 As the number of trees increases, for almost surely all sequences Θ_1, PE^* converges to:

$$P_{\mathbf{X},Y}\left(P_\Theta\left(C(\mathbf{X},\Theta)=Y\right)-\max_{j\neq Y}P_\Theta\left(C(\mathbf{X},\Theta)=j\right)<0\right) \tag{4}$$

Theorem 1 is proved with the strong law of large numbers and the tree structure. It indicates that it is unnecessary for RF to employ common anti-overfitting methods, for instance, cross-validation, early stopping, etc. RF does not overfit when more trees are added; meanwhile it results in a limiting value of the generalization error. This is another important feature of RF beside the two randomized procedures mentioned above.

2.2.3 Accuracy of RF depending on strength and correlation

In the previous section, the anti-overfitting characteristic of RF was proved, but we are more concerned about its accuracy. According to the analysis built in references, an upper bound of RF can be derived for the generalization error in terms of two parameters that are measures of how accurate the individual classifiers are and of the dependence between them. These also lead to an in-depth view of how RF works. Firstly, we define a margin function and raw margin function for RF.

The margin function for a RF is:

$$mr\left(\mathbf{X},Y\right)=P_\Theta\left(C(\mathbf{X},\Theta)=Y\right)-\max_{j\neq Y}P_\Theta\left(C(\mathbf{X},\Theta)=j\right) \tag{5}$$

The raw margin function is:

$$rmg\left(\Theta,\mathbf{X},Y\right)=I\left(C(\mathbf{X},\Theta)=Y\right)-I\left(C(\mathbf{X},\Theta)=\hat{j}(\mathbf{X},Y)\right) \tag{6}$$

Distinctively, $mr(\mathbf{X}, Y)$ is the expectation of $rmg(\Theta, \mathbf{X}, Y)$ with respect to Θ. And the strength of the number of individual classifiers $C(\mathbf{X}, \Theta)$ is:

$$S = E_{\mathbf{X},Y}\,mr\left(\mathbf{X},Y\right)$$

where $E_{\mathbf{X},Y}$ is the expected value of margin function over $\mathbf{X}$, Y space.

Then we compute the variance of margin function:

$$\operatorname{var}\left(mr\right)=\bar{\rho}\left(E_\Theta sd\left(\Theta\right)\right)^2\leq\bar{\rho}E_\Theta\operatorname{var}\left(\Theta\right) \tag{7}$$

Write

$$E_\Theta\operatorname{var}\left(\Theta\right)\leq E_\Theta\left(E_{\mathbf{X},Y}rmg\left(\Theta,\mathbf{X},Y\right)\right)^2-S^2\leq1-S^2 \tag{8}$$

where $\bar{\rho}$ is the mean value of the correlation, $sd\left(\cdot\right)$ is the standard deviation of $rmg\left(\Theta, \mathbf{X}, Y\right)$.

Considering functions (7), (8) and Chebychev inequality, theorem 2 can be concluded.

Theorem 2 An upper bound for the generalization error is given by

$$PE^*\leq\bar{\rho}\left(1-S^2\right)/S^2 \tag{9}$$

Although the bound is likely to be loose, it fulfills the same suggestive function for RF as VC-type bounds do for other types of classifiers. It shows that the two ingredients involved in the generalization error for RFs are the strength of the individual classifiers in the forest, and the correlation between them in terms of the raw margin functions. There is a conclusion drawn from this upper bound: the smaller this ration is, the better performance RF provides.

2.3 Genetic algorithm

RF can be strengthened by a standard genetic algorithm (GAs) [18] in this paper. GA is a simulation of evolution where the rule of survival of the fittest is applied to a population of individuals, or it can be considered as a parallel search procedure that simulates the evolutionary process by applying genetic operators. Compared with other search algorithms, GA has been well-known for its superior performance. And the most powerful feature of GAs is its great simplicity. They do not need too much code and no differentiability or continuity requirements to be satisfied. The usual GA flowchart (Fig. 3) and steps are shown as follows:

Step 1: Coding: generate an initial population (usually a randomly string)

Step 2: Fitness evaluation: apply some function or formula to the individuals to get the fitness of each individual.

Step 3: Selection: according to the fitness, individuals are selected to be the parents of next generation.

Step 4: Crossover: it is used to create two child individuals from the parent which pass the selection successfully via exchanging their chromosomes.

Step 5: Mutation: it assigns a new value to a randomly chosen gene and is controlled by a mutation probability.

Step 6: Repeat steps 3 to 5 until the evolved result satisfies the termination criteria, or a certain fixed number of generations are achieved.

The function of GA is to evaluate the best parameters of RF. Fitness is the criterion which indicates the capacity of each individual. In RF the diagnosis accuracy rate value is assigned to fitness which represents the performance of certain parameters. After generating the initial population, fitness values are calculated and assigned to individuals which include two key parameters of RF. The GA proceeds to the next generation through three genetic operators: selection, crossover, and mutation.

Selection is the most important part of GA. This operator impacts on the trend of GA and makes GA's running time shorten. It picks up the excellent parents to reproduce the individuals within the limitation. The normalization probability for individuals to be selected, N_p is described as the following equation:

$$N_p(i) = \frac{B_s(i)}{1 - \left(1 - B_s(i)\right)^{N_g(i)}} \tag{10}$$

where i is an individual, N_g is the number of generation. B_s is probability of selecting best individual from the current population.

The selection probability of each individual is:

$$P_s(i) = N(i)\left(1 - B_s(i)\right)^{I(i)} \tag{11}$$

where $I(i)$ is the sorted index of individuals according to the fitness.

The selection probability stands for the opportunity of individuals to be chosen as parents of the next generation. The new individuals are reproduced by the survivals from selection by crossover and mutation procedure.

3. Experiments and motor faults classification

3.1 Data acquisition and feature calculation

The experiments are designed to simulate six universal categories of induction motor faults, which are broken rotor bar, bowed rotor, bearing outer race fault, rotor unbalance, adjustable eccentricity motor (misalignment) and phase unbalance [19]. First, four motor faults are shown in Fig. 4 as an example. The load of the motors can be changed by adjusting the blade pitch angle or the number of the blades. The platform of these experiments consists of six 0.5 kW, 60 Hz, 4-pole induction motors, pulleys, belt, shaft and fan with changeable blade pitch angle. Detailed information is listed in Table 1.
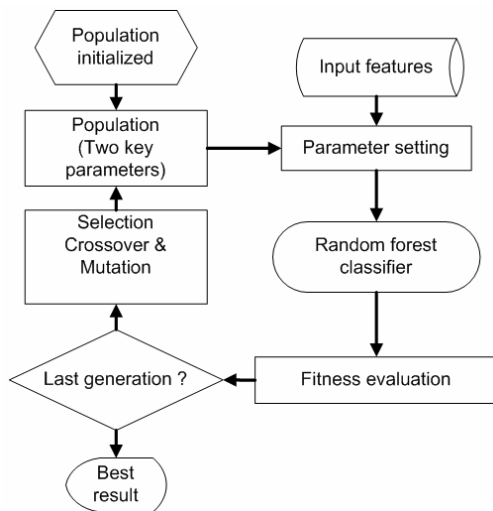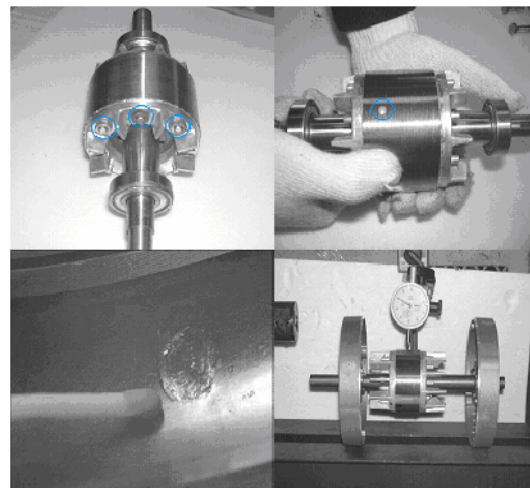

Fig. 3. Flowchart of genetic algorithm.

Fig. 4. Faults on induction motor.

Table 1. Fault category of induction motors.

Fault condition	Fault description	Others
Broken rotor bar	Number of broken bar: 12 ea	Total number of 34 bars
Bowed rotor	Max. bowed shaft deflection: 0.075 mm	Air-gap: 0.25 mm
Faulty bearing	A spalling on outer raceway	#6203
Rotor unbalance	Unbalance mass on the rotor	8.4 g
Eccentricity	Parallel and angular misalignments	Adjusting the bearing pedestal
Phase unbalance	Add resistance on one phase	8.4%

Table 2. Representation of input features parameters for RF.

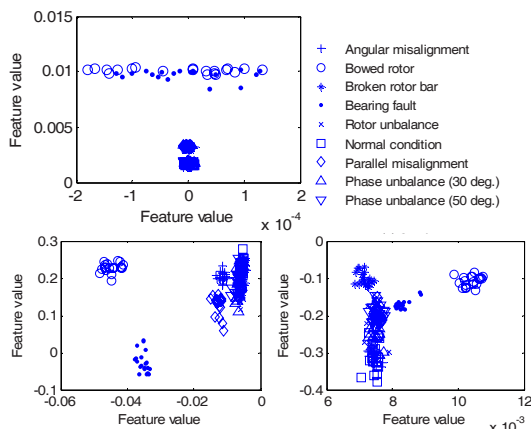| Time domain	Frequency domain	Auto regression
Mean RMS Shape factor Skewness Kurtosis Crest factor Entropy error Entropy estimation Histogram lower Histogram upper	Root mean square frequency Frequency center Root variance frequency	AR coefficients $(a1 \sim a8)$

Fig. 5. Connection between features and fault categories.

Three AC current probes and another three accelerometers were used to measure the stator current of three phase power supply and vibration signals of horizontal, vertical and axial directions for evaluating the RF-based fault diagnosis system.

After measuring the raw data, a preprocessing is implemented on the data to obtain the most important features for the RF-based diagnosis methodology.

Table 3. Information of class and samples.

Class No.	Class	Training samples	Test samples
1	Angular misalignment	20	10
2	Bowed rotor	20	10
3	Broken rotor bar	20	10
4	Bearing outer race fault	20	10
5	Mechanical unbalance	20	10
6	Normal condition	20	10
7	Parallel misalignment	20	10
8	Phase unbalance (30°)	20	10
9	Phase unbalance (50°)	20-	10
	Total samples	180	90

Finally, there are 63 features left for the next procedure, induction motor fault diagnosis by using RF. The description of feature parameters is shown in Table 2.

Fig. 5 marks eight conditions of a faulty motor and one normal condition via selecting two features randomly for three times. The purpose is to explain the relation between features and fault categories. It can be seen that although selected features are calculated for standing difference out among diverse fault categories, overlap cannot be avoided; in some samples it is even serious based on two randomly selected features. Generally, some statistical techniques such as principle component analysis (PCA), kernel PCA and linear discriminant analysis, are employed to compress the data again by reducing the number of dimensions of the data. But for RF-based system, it is wise to do diagnosis without such techniques for two reasons. First, information of data has been extracted when features are calculated. If dimension-reducing technique is employed, it may cause over-compressing problem of data. Second, according to Breiman's test, RF always gives good performance when the data scale is large [8]. Therefore, RF is more sensitive to the over-compressing problem. Thus, RF finished the diagnostic task without any feature extraction methods.

3.2 Fault diagnosis results and discussion

In this section, RF was tested on the induction motor fault data. Number and faulty sorts of training and testing data are shown in Table 3. The experimental results for RF-based method are given in Table 4. Confusion matrices for the training data in RF are given by Tables 5 and 6, which indicate the accuracies of each fault class for training and testing data

Table 4. Fault diagnosis accuracies based on RF.

No. of trees	Split variables	Test set accuracy (%)
200	1	88.89
	5	71.11
	8	81.22
500	1	94.44
	5	75.56
	8	83.33
1200	1	95.56
	5	72.23
	8	82.34
2000	1	93.33
	5	73.34
	8	83.33
5000	1	92.23
	5	72.23
	8	78.89
10000	1	92.25
	5	74.44
	8	77.78

Table 5. Accuracy of each fault class for training data with 907 trees and selecting 1 variable every split.

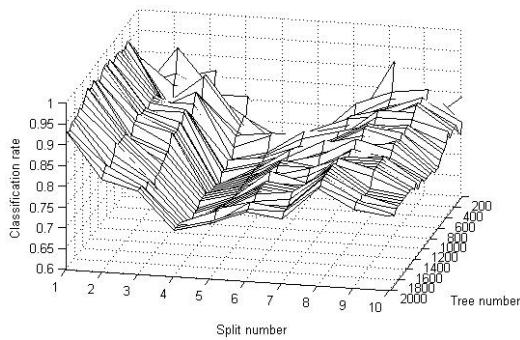| Class No.	1	2	3	4	5	6	7	8	9	Accuracy (%)
1	20	0	0	0	0	0	0	0	0	100
2	0	20	0	0	0	0	0	0	0	100
3	0	0	20	0	0	0	0	0	0	100
4	0	0	0	20	0	0	0	0	0	100
5	0	0	0	0	20	0	0	0	0	100
6	0	0	0	0	0	20	0	0	0	100
7	0	0	0	0	0	0	20	0	0	100
8	0	0	0	0	0	0	0	20	0	100
9	0	0	0	0	0	0	0	0	20	100

Fig. 6. Classification rate against random split number and tree number.

with 907 trees, with selecting 1 variable every split.

Fig. 5 shows the classification rate according to the experiment which represents three characteristics of RF very clearly. First, to compare with the number of

Table 6. Accuracy of each fault class for test data with 907 trees and selecting 1 variable every split.

Class No.	1	2	3	4	5	6	7	8	9	Accuracy (%)
1	10	0	0	0	0	0	0	0	0	100
2	0	10	0	0	0	0	0	0	0	100
3	0	0	10	0	0	0	0	0	0	100
4	0	0	0	10	0	0	0	0	0	100
5	0	0	0	0	9	1	0	0	0	90
6	0	0	0	0	0	10	0	0	0	100
7	0	0	0	0	0	0	10	0	0	100
8	0	0	0	0	0	0	0	10	0	100
9	0	0	0	0	0	0	0	0	10	100

component classification trees, the parameter, random split number at each node, is more sensitive to the classification accuracy. Hence a prudential searching procedure is necessary to find the best split variables number by an experimental way. Second, if the split variables number is decided, the sum of individual tree classifiers should achieve an appropriate quantity to get a better performance. Lastly , when we increase trees until a high number, for example 5000 or 10000, there is no over-fitting occurring but a little undulating exists.

Table 6 indicates that incorrect diagnosis of RF-based methodology often occurs at a certain fault category. So we can apply some assistant diagnosis method which is a function in that specific kind of fault to improve the diagnosis precision.

In general, the normal RF has achieved satisfactory fault diagnosis accuracy. But it should be noticed that two parameters, the number of trees and random split number, which greatly affect the classification result are set manually. It means accuracy of normal RF depends on a researcher's experience. This situation exists at almost all the applications of RF. So that applying the GA to do the parameter optimization is necessary. The effect of this cooperation is proved by using the same data. According to the previous research, in order to reduce execution time of a GA program and find the optimized point synchronously, the number of trees and random split number are limited in the range from 500 to 1500 and from 1 to 10, respectively.

Fig. 6 shows the trace information of every generation. Fitness adopts the classification accuracy of the test data set. The solid line is the best fitness value and the other one is mean fitness value of each generation. The risen and convergent trend of mean fitness value indicates that GA well cooperates with RF-

based methodology on the motor fault diagnosis, and the best fitness value lays out the optimization point, which is 907 trees and 1 random split created by 9th generation. The classification accuracy at this point achieves 98.89%, 3.33% higher than the best value of normal RF. It means GA can clearly enhance the capability of RF.

As shown in Table 7, we have also investigated the accuracy of other classifiers, adaptive resonance theory-Kohonen neural network (ART-KNN) [16], SVM [20] and CART classification tree [17]. Classification accuracies were obtained as 95.56% for RF only and 98.89% for RF optimized by GA (RFOGA), while the classification accuracies were 86.67% for ART-KNN, 87.15% for SVM and 77.78% for CART. It can be seen from the results that RFOGA and RF only achieves higher classification rates than ART-KNN, SVM and CART. RF has greatly increased the capability of tree classification method, to 95.56% from 77.78% for CART. This comparative result means all we have done is significant and further research is important and necessary.

4. Conclusions

The purpose of this paper is to confirm the possi-

Table 7. Overall classification accuracy of each classifier.

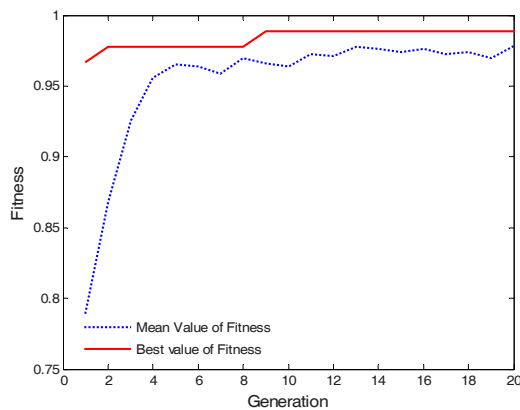| Classifier	Overall accuracy (%)
ART-KNN	86.67
SVM	87.15
CART	77.78
RF only	95.56
RFOGA	98.89

Fig. 7. Optimization trace within 40th generation.

bilities of using the random forests algorithm (RF) in machine fault diagnosis and propose a hybrid method combined with genetic algorithm to improve the classification accuracy. The proposed method is based on RF, a novel ensemble classifier which builds a large number of decision trees to improve the single tree classifier. Although there are several existing techniques for faults diagnosis, the research on RF is meaningful and necessary because of its fast execution speed, the characteristic of tree classifier, and high performance in machine fault diagnosis. Evaluation of the RF-based method has been demonstrated by a case study on induction motor fault diagnosis. Experimental results indicate the validity and reliability of the RF-based fault diagnosis method. In this paper, the RF and optimized RF-based fault diagnosis method of rotating machinery were investigated. The performance of two methods was proved by the fault diagnosis test of an induction motor. The optimized approach attains a high accuracy rate of diagnosis, 98.89%. The comparison result also shows that the optimized RF-based method is competitive with other classification method.

The extended research will focus on two parts. First part is to improve this hybrid method RFOGA: GA is not only for the parameter optimization, it also can be used to select the best combination of sub-classification trees from the forest to get a more accurate result. In the second part, we will decrease the redundancy of the RF and try other optimization algorithm or more effective voting principle. The extended research will focus on decreasing the redundancy of the RF and try other optimization algorithm or more effective voting principle.

References

[1] T. G. Dietterich, Ensemble Learning, The Handbook of Brain Theory and Neural Networks, M.A. Arbib (ED.), The MIT Press, Cambridge, MA, (2002).

[2] L. Breiman, Bagging Predictors, *Machine Learning*, 26 (2) (1996) 123-140.

[3] E. Bauer and R. Kohavi, An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants, *Machine Learning*, 36 (1999) 105-139.

[4] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, Random Forests for Land Cover Classification, *Pattern Recognition Letter*, 27 (4) (2006)

294-300.

[5] R. E. Schapire, The Strength of Weak Learnability, *Machine Learning*, 5 (2) (1990) 197-227.

[6] S. L. Joelsson, J. A. Benediktsson and J. R. Sveinsson, Random Forest Classifiers for Hyperspectral Data, *IEEE Publications*, (2005) 160-163.

[7] G. J. Briem, J. A. Benediktsson and J. R. Sveinsson, Multiple Classifiers Applied to Multisource Remote Sensing Data, *IEEE Trans. Geoscience and Remote Sensing*, 40 (2002) 2291-2299.

[8] L. Breiman, Random Forests, *Machine Learning*, 45 (1) (2001) 5-32.

[9] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, CA, (1986).

[10] B. S. Yang, C. H. Park and H. J. Kim, An Efficient Method of Vibration Diagnostics For Rotating Machinery Using a Decision Tree, *International Journal of Rotating Machinery*, 6 (1) (2000) 19-27.

[11] L. Breiman, Random Forest User Notes, Statistics Department, University of California, Berkeley, ftp://ftp.stat.berkeley.edu/pub/users/breiman/notes_on_random_forests_v2.pdf, (2006).

[12] M. Pal, Random Forest Classifier for Remote Sensing Classification, *International Journal of Remote Sensing*, 26 (1) (2005) 217-222.

[13] K. Remlinger, Introduction and Application of Random Forest on High Throughput Screening Data from Drug Discovery, *Proc. of Workshop for the SAMSI Program on Data Mining and Machine Learning*, (2003) 1-6.

[14] J. S. Ham, Y. Chen, M. M. Crawford and J. Ghosh, Investigation of the Random Forest Framework for Classification of Hyperspectral Data, *IEEE Trans. Geoscience and Remote Sensing*, 43 (3) (2005) 492-501.

[15] X. Di, T. Han and B. S. Yang, Application of Random Forest Algorithm in Machine Fault Diagnosis, *Inaugural World Congress on Engineering Asset Management*, Gold Coast, Australia, (2006).

[16] B. S. Yang, T. Han and J. L. An, ART-KOHONEN Neural Network for Fault Diagnosis of Rotating Machinery, *Mechanical Systems and Signal Processing*, 18 (3) (2004) 645-657.

[17] L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, Classification and Regression Trees, Wadsworth, Belmont, (1984).

[18] G. E. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning, Addison Wesley, New York, (1989).

[19] B. S. Yang and K. J. Kim, Application of Data Fusion and Dempster-Shafer Theory in Fault Diagnostics of Induction Motors, *Mechanical Systems and Signal Processing*, 20 (2) (2006) 403-420.

[20] A. Widodo, T. Han and B. S. Yang, Combination of Independent Component Analysis and Multiclass Support Vector Machines for Fault Diagnosis of Induction Motors, *Expert Systems with Applications*, 32 (2) (2007) 299-312.