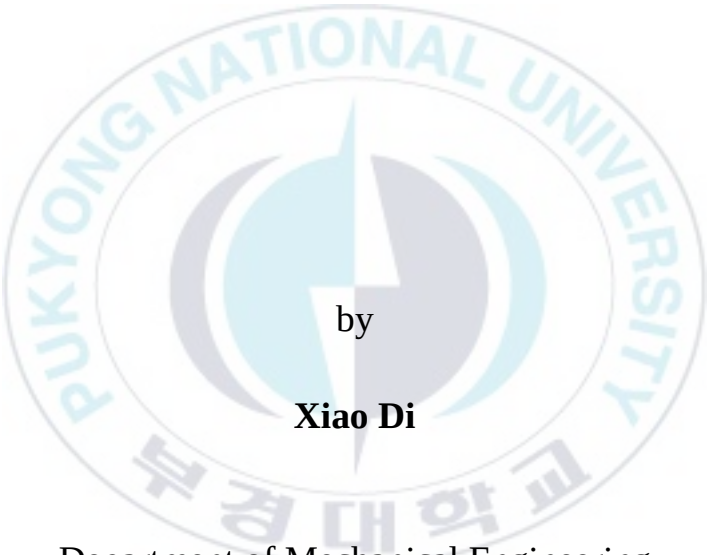
Thesis for the Degree of Master of Engineering

Random Forest based Faults Diagnosis Algorithm and Application to Induction Motor



by

Xiao Di

Department of Mechanical Engineering
The Graduate School

Pukyong National University

February 2007

Random Forest based Faults Diagnosis Algorithm and Application to Induction Motor

Random Forest 기반의 결함진단 알고리듬과 유도전동기에서의 응용

Advisor: Su-Joo Lee

by

Xiao Di

A thesis submitted in partial fulfillment of the requirements
for the degree of

Master of Engineering

in the Department of Mechanical Engineering
The Graduate School
Pukyong National University

February 2007

Random Forest based Faults Diagnosis Algorithm and Application to Induction Motor

A thesis

by

Xiao Di

Approved as to style and content by:

_____

Chairman

_____ _____

Member Member

February, 2007

List of Figures

List of Tables

Random Forest based Faults Diagnosis Algorithm and Application on Induction Motor

Xiao-Di

Department of Mechanical Engineering,

The Graduate School

Abstract

In this thesis, ensemble theory is represented as a powerful and effective methodology. This theory plays the role as tache between the Classification and Regression Tree (CART) and machine fault diagnosis theory. This combination shows its highlight on the induction motor faults diagnosis which is name Random Forest Algorithm.

This is a methodology by which rotating machinery faults can be diagnosed. The proposed method is based on random forests algorithm (RF), a novel assemble classifier which builds a large amount of decision trees to improve on the single tree classifier. Although there are several existed techniques for faults diagnosis, such as artificial neural network, support vector machines etc, the research on RF is meaningful and necessary because of its fast executed speed, the characteristic of tree classifier, and high performance in machine faults diagnosis. Evaluation of the RF based method has been demonstrated by a case study of

induction motors faults diagnosis. Experiment results indicate the validity and reliability of RF based fault diagnosis methodology. Furthermore, an optimized form of RF is also provided in this paper. We employ the genetic algorithm to strengthen RF, and valid this optimized RF algorithm's enhanced performance by the same experiment data. It is the evidence that RF based diagnosis methodology can touch more accurate outcome by combining with other optimization method.

I. Introduction

1.1 Background

1.1.1 Significance of faults diagnosis

Along with the application of new technologies on the modern equipment, the structure and function of advanced equipments are becoming more complicated and comprehensive, their automaticity is going higher too. Thus there are many unavoidable factors which cause various malfunctions existed on the machinery. These malfunctions will result in serious accidents bringing on great loss in economy and human lives. In additional, faults of machinery which is located in vital department may cause incredible losing. Hence it is such an exigent issue to ensure the equipments worked under normal condition and accidents will not be happened.

The security and reliability of modern machinery are depended on two aspects. one is to guarantee design and quality of the machinery in accordance with the guild line. Besides equipment fixing, running, managing, maintenance and diagnosis should be made appropriately and correctly.

It is important that machinery malfunctions (faults) diagnosis can produce the great benefit, there is many reports stating the advantage of machinery faults diagnosis all over the world:

(1) On the view of manufacturer, implement of faults diagnosis system will decrease accident occurring rate, therefore the rate of profit against investment will arrive at a high stage.

The method which is taken by Perdrul power plant to estimate the benefit from diagnosis program in USA can be taken as an evidence. The capability of Perdurl is $100\times10^4 kW$, electricity charge is one hundred million dollar. Stopping production loss is 150 thousand dollar per day. There are 50 important sections need to be monitored, the all investment is more than200 thousand dollar.

Monitoring charge just costs only 15 thousand dollar every year. According to reliable calculation, the breakdown will occur 14 times per year. After adopting diagnosis technology, 50% of the accidents can be inspected, and half of that 50% is detected by monitor and diagnosis system, 20% of all is pseudo alarm, every accident need 3 days to repair in average. Finally, diagnosis system can save the money B is:

$$\begin{aligned} B &= 0.5\times 0.5\times 14\times 3\times 15\times (1-0.2) \\ &= 1260000\$ \end{aligned}$$

Diagnosis cost:

$$\begin{aligned} A &= (20/10 dep/year)+1.5 \\ &= 35000\$/year \end{aligned}$$

Then economic profit coefficient C is:

$$C = \frac{A}{B} = \frac{1260000}{35000} = 36$$

Thus it can be seen that the profit of applying diagnosis system is 36 times of the investment for it.

(2) Employing malfunction diagnosis system can prolong maintenance period, decrease the breakdown time of equipment. And it is also the foundation of setting down an appropriate maintenance strategy which may promote the profit greatly.

For example, the capability of a power plant is $100\times 10^4 kW$, generating $2400\times 10^4 kW$ per day, production value up to hundreds thousand dollar everyday. If it is possible to prolong the time of maintenance cycle, such as shorten 10 days one year, the corresponding benefit can touch millions dollar.

(3) The charge of maintenance for equipments is a huge amount of money, but applying diagnosis system can depress this charge to bottom.

For example, the revenue of USA is 750 billion US dollar in 1980, but almost 30% is put into the equipment servicing. According to the analysis by expertise, one third of the fee for equipment servicing, 75 billion dollars, is wasted because

of improper maintenance method, i.e. lack of condition monitoring and malfunction diagnosis. Thus it can be seen that the investment on the diagnosis system will bring great benefit.

1.1.2 Objective of faults diagnosis

It is important to know what the purpose of diagnosis system as well, which are:

- Fault diagnosis system can detect the malfunction precisely, and as soon as possible. It can prevent and avoid the machine broken down, enhance the reliability, security and efficiency of equipment, thereby this system reduce the loss by machine fault under the lowest point.
- Fault diagnosis system makes use of the capability of equipment maximally. A proper designed monitoring and diagnosis program extend the live cycle of equipment, so that the cost of product is down at same time.
- By applying condition monitoring, malfunction analysis, performance estimation…etc, important information of machine reconstruction, optimization, product processing rationalization are gathered to improve the hole product line.

All in all, machine fault diagnosis not only ensures the equipment run in normal state, but also obtains great benefit both in economy and society.

1.1.3 Mission of machinery fault diagnosis

The responsibility of machinery fault diagnosis is to monitor the machine on-line. And it estimates its running condition. It also diagnoses and eliminates the faults. Finally, it directs the strategy of management and maintenance of equipment.

1.1.3.1 Condition monitoring

The task of condition monitoring is to monitor machine working state, including adopting multifarious detection, measure, monitoring, analysis and

distinguish method. By Combining data from history and actuality of machine system and considering environmental factor, working state of machine system is evaluated. Then it judges machine condition is normal or abnormal by certain rules, and record and display this condition. It will give an alarm, if the condition is abnormal. So technician response to this problem will be solved as possible to prevent the machine broken down occurred. At last, condition monitoring provides important information and basic data for fault analysis, performance estimation, correct and safe operation on equipment.

Usually, the condition of equipment can be divided into three instances which are normal condition, abnormal condition and failure condition. Normal condition means there is no fault in the machine system, or fault exists but under the permitted level. Abnormal condition means the fault of equipment deteriorates and impacts on other connected components. The performance of equipment is declining, but still can keep working. When equipment is in abnormal condition, it should be running under monitoring system. Failure condition means the performance of equipment is dropping quickly, and can not satisfy basic need. In addition, failure condition can be separated to three phases: Early fault stage that fault exists and just has the trend to go worse. Normal functional fault stage which the equipment is running on top of the lowest limitation. Ruinous failure stage which equipment is broken down and waiting for fixing and instantaneous failure stage caused by some unexpected reasons. There are several alarm sign response to different condition of machine. Usually it is represented via different colors of indicator light. Green means machine is running under normal condition, yellow means there is a warning of the failure, red means breakdown could be occurred.

Furthermore, in order to find the causation of the failure out after the event, information of e failure is recorded, including storage function of the signal of the ruinous failure,

1.1.3.2 Fault diagnosis

Fault diagnosis bases on information gathered by condition monitoring. Then it will be integrated with characteristic and parameter of the construction and environmental factors. After combining the log file of certain equipment which consists of run-time record, failure and maintenance history data, failure which will happen in future is predicted.

Different fault location and category may cause the degradation of equipment and the running condition in different ways. So another task of fault diagnosis is to decide the type and position of the fault via condition and signal of the equipment when the fault occurs at one or more then one component. Because the amount of measured signal is huge, it is necessary to calculate the features from raw signal data to simplify decision-making work and enhance the successful rate of diagnosis. The variety of raw data caused by only one kind of fault is named the symptom. To determine what component is broken and which category of fault is are the procedure of fault diagnosis. Thereby, the essential of fault diagnosis is a kind of status identification problem.

The most difficult thing met in fault diagnosis is that the relation between faults and symptoms. It is not simple one standing for one, but more complicated. One type of fault may be expressed by several symptoms. Similarly one symptom could be the phenomena of a number of faults. Such as, rotor unbalance causes increasing of mechanical vibration. Frequency component of operation speed can express change in vibration signal clearly, so it is the main symptom. But synchronously increasing of frequency component of operation speed is not only

for the rotor unbalance. There are many other faults may be result in that symptom. That is the reason why correct diagnosis is hard to reach. Therefore fault diagnosis is a procedure of reduplicate experiment: Firstly, base on the diagnosis knowledge to extract the symptoms, and then put it into diagnosis system. Purpose is to find out countermeasure, do adjustment and experiment on the equipment. Even sometimes machine is operated till it is down to repair it. At last turn the machine on and check its working condition. If it still abnormal means we need more information to do the diagnosis, so do the whole procedure again till the equipment back to normal condition.

1.1.3.3 Directing the maintenance and management strategy of equipments

The management and maintenance strategy of equipments comes through three phases: from Run-to-Breakdown Maintenance, to Time-based Preventive Maintenance, untill now Condition-based Maintenance. Time-based Preventive Maintenance can prevent the accident occurred. But disadvantage of this method is it often causes the lack of or over maintenance. Condition-based Maintenance is more scientific and reasonable maintenance strategy. But the implement of Condition-based Maintenance is depended on condition monitoring and faults diagnosis system working effectively. It is also why this technique is attached importance to the all over the world. With developing and implement of faults diagnosis technique, management and maintenance of equipment will be up to a higher level. At one time, the live-cycle of equipment will be prolonged farther, and the malignant accidents will be minimized, the economy will go faster and healthier.

1.2 Definition, Contents and Basic Methodologies of Machine Faults Diagnosis Technique

1.2.1 Definition of machine faults diagnosis technique

Machine faults diagnosis technique is on-line faults diagnosis technique, it means this technique obtains the condition of machine on-line. And it finds out the

causation and location of the fault. Then it forecasts the most possible condition machine could be in the future. Fault diagnosis technique consists of threes parts: First is to realize the actuality of purpose equipment. Second is to realize the abnormity or feature of the fault. Third is to predict and forecast the trend of equipment fault and state. It should be known that prediction is based on signal or symptom of a certain machinery to do diagnosis; but forecast employs the probability and statistics method to speculate the result.

1.2.2 The approach of machine faults diagnosis technique

The content of faults diagnosis is composed of condition monitoring, analysis and diagnosis and fault prediction, the detailed procedures are listed as fellow:

1. Data acquisition: during the process of machine running change of force, heat, vibration and energy, diversified signals exists synchronously. And then according to the need of diagnosis, different signals are selected by which can stand for running condition of equipment such as vibration, pressure, temperature and so on. The signals mentioned are obtained by various sensors.
2. Data processing: In this procedure, the acquired data is processed by mathematic and statistics methods to calculate features which can represent the machine state well. For example, transforming the signal from time-domain to the frequency-domain to do the analysis is one method of signal processing.
3. Status identification: comparing the features which calculated features, the difference found between the two data can be used to detect the character and category of the fault. According to output of diagnosis system, the diagnosis policy will be made.
4. Diagnosis decision-making: after making the policy of diagnosis, system decides the certain countermeasure and plan, and according to the condition of equipment and change of feature, trend analysis will be done. The figure

1.1 shows the whole diagnosis system.

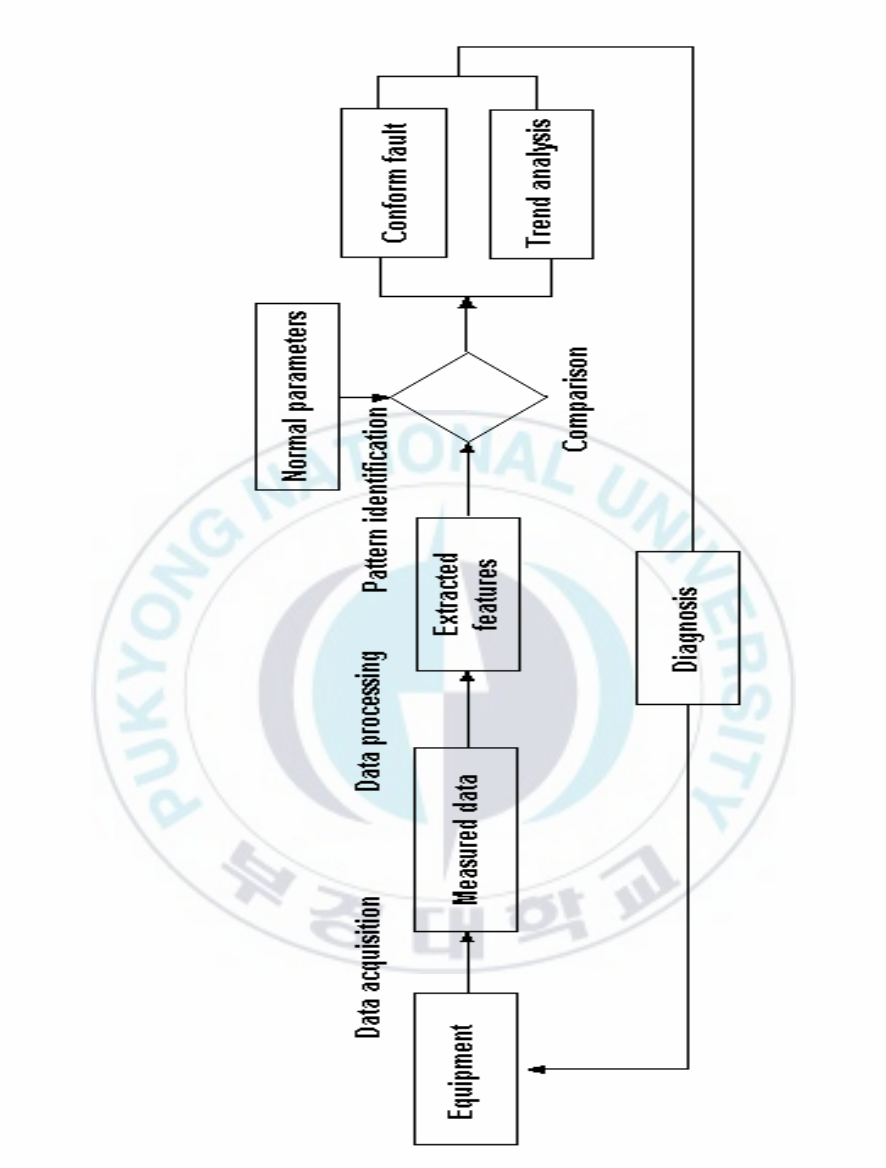


Figure 1.1 Flowchart of diagnosis system

1.3 Methodologies of Machine Faults Diagnosis Technique

The complexity of the machine faults and relationship between faults and symptoms tell us that the machine faults diagnosis is always considered as an

exploring procedure. In machine faults domain, emphases of diagnosis technique do not focus on the fault itself but on the diagnosis methods. Due to the complexity of diagnosis, it is impossible to get correct output just via single method. The succeed diagnosis must combine various method. So the fact is that researchers should integrate different techniques, knowledge and methodology from diversified field. It is also important characteristic that diagnosis technique is an intersectional science.

1.3.1 Conventional faults diagnosis method

One of the conventional methods is to utilize the physics and chemistry theory and techniques to detect the multifarious physical and chemic phenomena of equipment to find the fault out directly. For example: By monitoring of chemic composition, vibration, acoustics, lights, electromagnetic and thermal radiation signal to detect and diagnose the fault immediately. The advantages of this method are visible, fast, effective, but disadvantage is that it is just suit for partition of faults.

Another method is most popular and well-developed. It diagnoses base on the relation between the faults and symptoms. Taking the rotational machinery as an example, the symptom of rotational machine fault is the characteristics of vibration signal in time and frequency domain. Hence, engineers put the attention on the research of fault mechanism and corresponding symptoms. During the diagnosis procedures, experts analyze the measured signal, extract the features, and then find the corresponding symptoms from the features. The symptoms are used to do the fault diagnosis. But it should be emphasized again that That is the reason why fault diagnosis is very complicated, therefore usually fault diagnosis is a procedure of reduplicate experiment.

1.3.2 Intelligent fault diagnosis method

Intelligent fault diagnosis method is established on the conventional methods and integrates the principle and technique of Artificial Intelligence, which is a

new approach of fault diagnosis. This technique is widely employed in many diagnosis fields and leads the development of fault diagnosis industry.

Artificial Intelligence makes the computer to finish the tasks which need the human intelligent before, i.e. consequence, comprehension, programming, decision-making, abstracting and learning…etc. Expert System is one form of AI which is introduced to diagnosis field sophisticatedly.

Expert System consists of repository, logistic system and storage space (including database). Furthermore, a realized expert System should have knowledge acquisition module, repository management module, explanation module, display module and man-machine conversation module…etc.

The problems of Expert System are knowledge acquisition and knowledge representation. Knowledge acquisition is the bottleneck of Expert System, the reasonable representation method can organize the knowledge effectively, enhance the capability of Expert System. For the sake of extending the Expert System, so much work has to be done. Such as: To Analyze the mechanism of machine fault, set the mathematic model for analyzing in theory; To do test and experiment on the equipment; To summarize the diagnostic experience of specialist and transfer this knowledge to the form which computer understands; To research the theory and method of machine learning. All of work introduced makes the Expert System more and more excellent.

1.4 Motivation of the Study

As mentioned above, the Expert System are well developed and widely applied. The strength of if is significant, but the weakness is also distinct. Performance of Expert System has strong connection with its repository which is fully constructed or not. The problem is knowledge (experience) acquisition often limits the capability of Expert System, because sometimes it is too difficult to establish an integrated repository.

On this occasion, another diagnosis method which is named mathematical

diagnosis method catches the researchers' eyes. This method employs the latest research output of other kinds of science and especially some effective mathematical tools, such as machine learning methods, like decision tree (DT), artificial neural network (ANN), support vector machines (SVM) etc. The new techniques and their extended research increase the intelligent, preciseness and applicability of diagnosis domain. The potential of machine learning based fault diagnosis inspirits researchers to find the opportunities to improve the performance of existed algorithm.

While my passion of developing machine learning based machinery faults diagnosis methods are increasing, the Ensemble Theory offers the chance to carry this object out. The simple definition of Ensemble Theory is that an ensemble consists of a set of individually trained classifiers (such as ANN and Decision Tree) whose predictions are combined when classifying novel instance.

The geol of the thesis is to introduce and investigate a novel machinery faults diagnosis methodology based on random forests algorithm [1, 2]. I believe that the research on this algorithm is worthy as developing a new accurate diagnosis mechanism and also helpful for the continuous work on Ensemble Theory.

II. The Theoretical Background of Thesis

2.1 Artificial Intelligence

Artificial intelligence (AI) is defined as intelligence exhibited by an artificial entity. Such a system is generally assumed to be a computer. Although AI has a strong science fiction connotation, it forms a vital branch of computer science, dealing with intelligent behavior, learning and adaptation in machines. Research in AI is concerned with producing machines to automate tasks requiring intelligent behavior. Examples include control, planning and scheduling, the ability to answer diagnostic and consumer questions, handwriting, speech, and facial recognition. As such, it has become a scientific discipline, focused on providing solutions to real life problems. AI systems are now in routine use in economics, medicine, engineering and the military … etc.

AI divides roughly into two schools of thought: Conventional AI and Computational Intelligence (CI). Conventional AI mostly involves methods now classified as machine learning, characterized by formalism and statistical analysis. This is also known as symbolic AI, logical AI, neat AI and Good Old Fashioned Artificial Intelligence (GOFAI). Methods include:

- Expert systems: apply reasoning capabilities to reach a conclusion. An expert system can process large amounts of known information and provide conclusions based on them.

- Case based reasoning

- Bayesian networks

- Behavior based AI: a modular method of building AI systems by hand

Computational Intelligence involves iterative development or learning (e.g. parameter tuning e.g. in connectionist systems). Learning is based on empirical data and is associated with non-symbolic AI, scruffy AI and soft computing. Methods mainly include:

- Neural networks: systems with very strong pattern recognition capabilities.

- Fuzzy systems: techniques for reasoning under uncertainty, has been widely used in modern industrial and consumer product control systems.

- Evolutionary computation: applies biologically inspired concepts such as populations, mutation and survival of the fittest to generate increasingly better solutions to the problem. These methods most notably divide into evolutionary algorithms (e.g. genetic algorithms) and swarm intelligence.

With hybrid intelligent systems attempts are made to combine these two groups. Expert inference rules can be generated through neural network or production rules from statistical learning such as in ACT-R.

A promising new approach called intelligence amplification tries to achieve

artificial intelligence in an evolutionary development process as a side-effect of amplifying human intelligence through technology.

2.2 Machine Learning

As a broad subfield of artificial intelligence, Machine learning is concerned with the development of algorithms and techniques, which allow computers to "learn". At a general level, there are two types of learning: inductive, and deductive. Inductive machine learning methods create computer programs by extracting rules and patterns out of massive data sets. Machine learning overlaps heavily with statistics, since both fields study the analysis of data, but unlike statistics, machine learning is concerned with the algorithmic complexity of computational implementations. Many inference problems turn out to be NP-hard or harder, so part of machine learning research is the development of tractable approximate inference algorithms.

Machine learning has a wide spectrum of applications including search engines, medical diagnosis, bioinformatics and chemoinformatics, detecting credit card fraud, stock market analysis, classifying DNA sequences, speech and handwriting recognition, object recognition in computer vision, game playing and robot locomotion.

2.2.1 Machine learning algorithm types

Machine learning algorithms are organized into taxonomy, based on the desired outcome of the algorithm. Common algorithm types include:

l Supervised learning where the algorithm generates a function that maps inputs to desired outputs. One standard formulation of the supervised learning task is the classification problem: the learner is required to learn (to approximate the behavior of) a function which maps a vector $[X_1, X_2 ... X_n]$, into one of several classes by looking at several input-output examples of the function.

l Unsupervised learning: which models a set of inputs: labeled examples are not available.

l Semi-supervised learning which combines both labeled and unlabeled examples to generate an appropriate function or classifier.

l Reinforcement learning where the algorithm learns a policy of how to act given an observation of the world. Every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm.

l Ttransduction is similar to supervised learning, but does not explicitly construct a function: instead, tries to predict new outputs based on training inputs, training outputs, and new inputs.

l Learning to learn where the algorithm learns its own inductive bias based on previous experience.

The performance and computational analysis of machine learning algorithms

is a branch of statistics known as computational learning theory.

Some of machine learning topics are well known as the powerful tools in many different fields, for example artificial neural networks, decision trees, k-Nearest Neighbor, Support vector machines and so on. The machinery faults diagnosis industry is also one of its application fields. Usually anyone of these algorithm has a good performance when diagnose the fault based on it, especially artificial neural networks, decision trees and Support vector machines. But sometimes we meet the problem that it is so hard to promote the capability of this algorithm itself after it toughs its limitation. Without considering its probability, it needs the long term research and great effort to be spent on. So one way is to move the points form improving the algorithm endlessly to thinking about how to use the algorithm. Hence, the ensemble theory exists which makes the work above possible.

2.3 Ensemble Theory

Many researchers have investigated the technique of combining the predictions of multiple classifiers to produce a single classifier [1]. The resulting classifier is generally more accurate than any of the individual classifiers making up the ensemble. Both theoretical and empirical research [2, 3] has demonstrated that a good ensemble is one where the individual classifiers in the ensemble are both accurate and make their errors on different parts of the input space. Two popular methods for creating accurate ensembles are Bagging [1] and Boosting [4]. These methods rely on re-sampling techniques to obtain different training sets

for each of the classifiers. Previous work has demonstrated that Bagging and Boosting are very effective for decision trees [5]. But without the concerning selecting training parameter problems, neural networks and SVM are also fit for ensemble theory. The rest of this section will discuss conventional ensemble methodology especially these two popular methods.

The basic framework for a classifier ensemble is shown in Fig. 2.1. In this example, neural networks are the basic classification method, though conceptually any classification method, such as decision trees, can be substituted in place of the networks. Each network in Fig 2-1ls ensemble, network 1 through network N in this case, is trained using the training instances for that network. Then, for each example, the predicted output of each of these networks, o_i in Fig 1, is combined to produce the output of the ensemble, $\hat{o}$ in Fig. 2-1. Many researchers [1, 2, 6] have demonstrated that an effective combining scheme is to simply average the predictions of the network.

2.3.1 Classifier ensembles

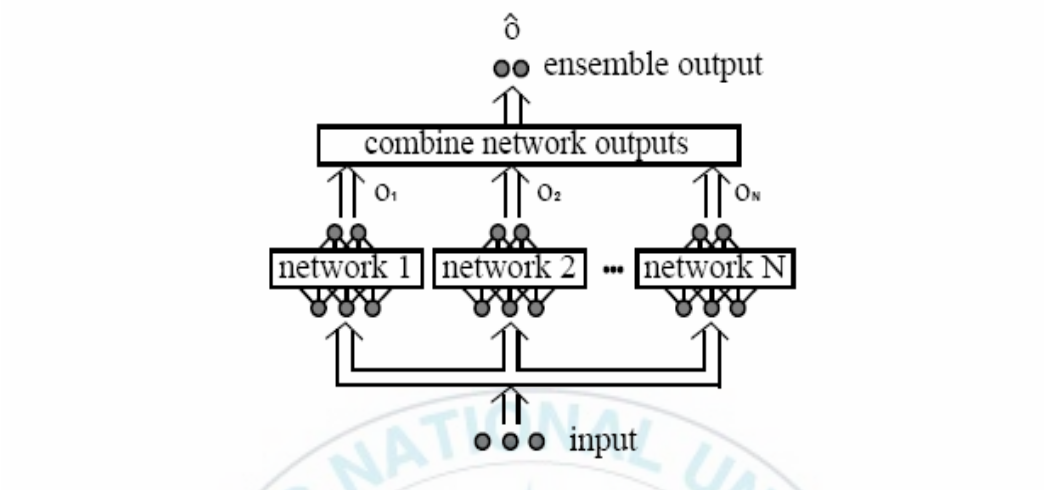


Figure 2.1 A classifier ensemble of neural networks

Combining the output of several classifiers is useful only if there is disagreement among them. Obviously, combining several identical classifiers produces no gain. Hansen and Salamon [7] proved that if the average error rate for an example is less than 50% and the component classifiers in the ensemble are independent in the production of their errors, the expected error for that example can be reduced to zero as the number of classifiers combined goes to infinity; however, such assumptions rarely hold in practice. Krogh and vedelsby's paper [2] proved that the ensemble error can be divided into a term measuring the average generalization error of each individual classifier and a term measuring the disagreement among the classifiers. What they formally showed was that an ideal ensemble consists of highly correct classifiers that disagree as much as possible.

As a result, methods for creating ensembles center around producing classifiers that disagree on their predictions. Generally, these methods focus on altering the training process in the hope that the resulting classifiers will produce different predictions. For example, neural network techniques that have been employed include methods for training with different topologies, different initial weights, different parameters, and training only on a portion of the training set. At the fellow parts, two popular ensemble methods Bagging and Boosting.

A sample of a single classifier on an imaginary set of data.	
	(Original) Training Set
Training-set-1:	1, 2, 3, 4, 5, 6, 7, 8

A sample of Bagging on the same data.	
	(Resampled) Training Set
Training-set-1:	2, 7, 8, 3, 7, 6, 3, 1
Training-set-2:	7, 8, 5, 6, 4, 2, 7, 1
Training-set-3:	3, 6, 2, 7, 5, 6, 2, 2
Training-set-4:	4, 5, 1, 4, 6, 4, 3, 8

A sample of Boosting on the same data.	
	(Resampled) Training Set
Training-set-1:	2, 7, 8, 3, 7, 6, 3, 1
Training-set-2:	1, 4, 5, 4, 1, 5, 6, 4
Training-set-3:	7, 1, 5, 8, 1, 8, 1, 4
Training-set-4:	1, 1, 6, 1, 1, 3, 1, 5

Figure 2.2: Hypothetical runs of bagging and boosting

2.3.2 Bagging classifiers

Bagging [1] is a “bootstrap” [8] ensemble method that creates individuals for its ensemble by training each classifier on a random redistribution of the training set. Each classifier’s training set is generated by randomly drawing, with replacement, N examples – where N is the size of the original training set; many of the original examples may be repeated in the resulting training set while others may be left out. Each individual classifier in the ensemble is generated with a different random sampling of the training set.

Fig. 2.2 shows the process of Bagging and Boosting. Assume there are eight training examples. Assume example 1 is an “outlier” and is hard for the component learning algorithm to classify correctly. With Bagging, each training set is an independent sample of the data thus, some examples are missing and others occur multiple times. The Boosting training sets are also samples of the original data set, but the “hard” example occurs more in later training sets since Boosting concentrates on correctly predicting it.

It gives a sample of how bagging might work on a imaginary set of data. Since Bagging re-samples the training set with replacement, some instances are represented multiple times while others are left out. So Bagging’s Training-set-q might conain examples 3 and 7 twice, but does not contain either example 4 or 5. As a result, the classifier trained on training-set-1 might obtain a higher test-set error than the classifier using all of the data. In fact, all four of Bagging’s component classifiers could result in higher test-set error; however, when

combined, these four classifiers can produce test-set error lower than that of the single classifier. the diversity among these classifiers generally compensates for the increase in error rate of any individual classifier.

Breiman [1] showed that Baagging is effective on “unstable” learning algorithms where small changes in the training set result in large changes in predictions. Breiman claimed that neural networks and decision trees are examples of unstable learning algorithms.

2.3.3 Boosting classifiers

Boosting [9] encompassed a family of methods. The focus of these methods is to produce a series of classifiers. The training set used for each member of the series is chosen based on the performance of the earlier classifiers in the series. In Boosting, examples that are incorrectly predicted by previous classifiers in the series are chosen more often than examples that were correctly predicted. Thus Boosting attempts to produce new classifiers that are better able to predict examples for which the current ensemble’s performance is poor. But in bagging, the re-sampling of the training set is not dependent on the performance of the earlier classifiers.

Fig. 2.2 shows a hypothetical run of Boosting. Note that the first training set would be the same as Bagging; later training sets accentuate examples that were misclassified by the earlier member of the ensembles. In this figure, example 1 is a “hard” example that previous classifiers tend to misclassify. With the second

training set, example 1 occurs multiple times, as do examples 4 and 5 since they were left out of the first training set and, in this case, misclassified by the first learner. For the final training set, example 1 becomes the predominant example chosen whereas no single example is accentuated with Bagging; thus, the overall test-set error for this classifier might become very high. Despite this, however, Boosting will probably obtain a lower error rate when it combines the output of these four classifiers since it focuses on correctly predicting previously misclassified examples and weights the predictions of the different classifiers based on their accuracy for the training set.

Previous work has demonstrated that Bagging and Boosting are very effective for decision trees. Discussions with previous researchers reveal that many authors concentrated on decision trees due to their fast training speed and well-established default parameter settings. Other AI methods, neural networks and SVM, present difficulties for testing both in terms of the significant processing time required and in selecting training parameters. So as the primary research on ensemble theory, a novel and powerful ensemble method, Random Forest Algorithm, is investigated in my thesis. However, there are distinct advantages to including neural networks and SVM in my future study. First, previous empirical studies have demonstrated that individual neural net works and SVM produce highly accurate classifiers that are sometimes more accurate than corresponding decision trees. Second, neural networks have been extensively applied across numerous domains. Finally, by studying neural networks in addition to decision trees we can examine how Bagging and Boosting are

influenced by the learning algorithm, giving further insight into the general characteristics of these approaches.

There are also a number of interesting conclusions of Bagging and Boosting. The first is that a Bagging ensemble generally produces a classifier that is more accurate than a standard classifier. For Boosting, however, we note more widely varying results. For a few data sets Boosting produced dramatic reductions in error, but for other data sets it actually increases in error over a single classifier.

2.4 Random Forest

RF which derive from decision tree classifier is an assembled method, it grows tree using CART (acronym of *classification* and *regression trees*) methodology to maximum size and without pruning. Therefore, basic principles of CART methodology will be provided here.

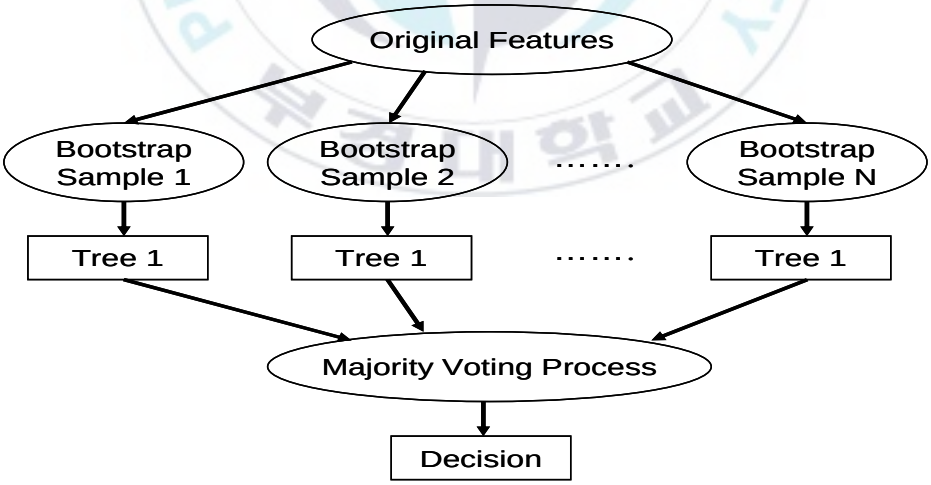


Fig. 2.3. Construction of random forest.

2.4.1 Classification and regression tree

CART grows classification and regression trees to predict continuous dependent variables (regression) and categorical predictor variables (classification) [14]. An example of a classification tree is shown below. The target variable is “Species”, the species of Iris. We can see from the tree that if the value of the predictor variable “Petal length” is less than or equal to 2.45 the species is Setosa. If the petal length is greater than 2.45, then additional splits are required to classify the species.

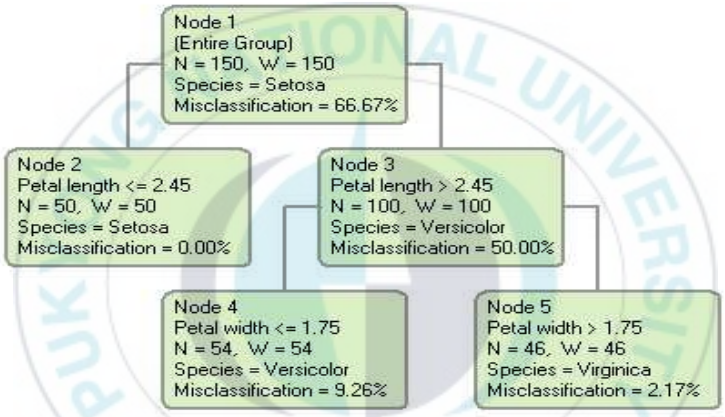


Figure 2.4: An example of classification tree

2.4.2 The predictive accuracy of CART

Accuracy is the most important feature of a classification tree. All classification procedures, however, including CART, can produce errors. The CART procedure does not make any distributional assumptions on covariates; hence, hypothesis testing is not possible. Confidence in CART’s performance, therefore, has to be based primarily on an assessment of the extent of misclassification it generates from data sets with known class distributions and on knowledge of and experience with the subject matter under study. And this

method is also suitable for test the accurate rate of Random forest algorithm.

The best way to test the predictive accuracy of a tree is to take an independent test data set with known class distributions and run it down the tree and determine the proportion of cases misclassified. In empirical studies, the possibility of getting such a data set is remote. To overcome this difficulty, Breiman [15] provide three procedures for estimating the accuracy of tree-structured classifiers. In this thesis, one of them is applied and explained here. Let:

$c(X)$ or c = a tree-structured classifier, where X is a vector of characteristics variables that describe an observation;

$R^*[c(X)]$ = the classifier's "true" misclassification rate; and

L = the learning sample (the sample data from which to construct a classification tree)

The three estimation procedures below have two objectives: constructing a classification tree, $c(X)$, and then finding an estimate of $R^*[c(X)]$.

Re-substitution Estimates $R[c(X)]$. This estimates the accuracy of the true misclassification rate, $R^*[c(X)]$, as follows:

- Build a classification tree, $c(X)$, from the learning sample L, and save it.

- Apply the tree, $c(X)$, to the data set from which it is built. That is let the observations in the sample run down the tree one at a time.

- Compute the proportion of cases that are misclassified. This proportion is the re-substitution estimate, $R\left[c(X)\right]$, of the true misclassification rate, $R^{*}\left[c(X)\right]$.

The re-substitution estimate tests the accuracy of a classifier by applying it to observations for which the classes are known. The major weakness of this estimator of the error rate is that it is derived from the same data set from which the tree is built; hence, it underestimates the true misclassification rate. The error rate is always low in such cases.

2.4.3 Methodology for building a classification tree

In constructing a classification tree, CART makes use of prior probabilities (priors). A brief review of priors and their variations as used in CART is provided.

Prior probabilities play a crucial role in the tree-building process. Three types of priors are available in CART: prior data, priors equal, and priors mixed. They are either estimated from data or supplied by the analyst.

In the following discussion, let

N = number of cases in the sample,

N_j = number of class j cases in the sample, and

π_j = prior probabilities of class j cases.

- Prior data assumes that distribution of the classes of the dependent variable in the population is the same as the proportion of the classes in the sample. It is estimated as $\pi_j = \frac{N_j}{N}$.

- Priors equal assumes that each class of the dependent variable is equally likely to occur in the population. For example, if the dependent variable in the sample has two classes, then pro(class 1) = pro(class 2) = 0.5

- Prior mixed is an average of prior equal and prior data for any class at a node.

2.4.3 Components for building a classification tree

Three components are required in the construction of a classification tree: (1) a set of questions upon which to base a split; (2) splitting rules and goodness-of-split criteria for judging how good a split is; and (3) rules for assigning a class to each terminal node. Each of these components is discussed below.

2.4.3.1 Type and format of questions

Two question formats are defined in CART: (1) Is $X \le d$?, if X is a continuous variable and d is a constant within the range of X values. For example, is $income \le 2000$? or is $Z = b$?, if Z is a categorical variable and b is one of the integer values assumed by Z. For example, is $sex = 1$?.

The number of possible split points on each variable is limited to the number of distinct values each variable assumes in the sample. For assumes N distinct points in the sample, then the maximum number of split points on X is equal to N. if Z is a categorical variable with m distinct points in a sample, then the number of possible split points on Z equals $2^{m-1}-1$ [15]. Unless otherwise specified, CART software assumes that each split will be based on only a single variable.

2.4.3.2 Splitting rules and goodness-of-split criteria

This component requires definition of the impurity function and impurity measure.

Let j = 1,2,…,k be the number of classes of categorical dependent variables; then define $p(j|t)$ as class probability distribution of the dependent variable at node t, such that $p(1|t)+p(2|t)+p(3|t)+...+p(k|t)=1$, j = 1,2,…, k. Let $i(t)$ be the impurity measure at node t. then define $i(t)$ as a function of class

probabilities $p(1|t), p(2|t), p(3|t), \ldots$. Mathematically, $i(t) = \phi[p(1|t), p(2|t), ..., p(j|t)]$

. The definition of impurity measure is generic and allows for flexibility of functional forms.

There are three major splitting rules in CART: the Gini criterion, the towing rule, and the linear combination splits. In addition to these main splitting rules, CART users can define a number of other rules for their own analytical needs. CART uses the Gini criterion as its default splitting rule. The towing rule is discussed in detail in Breiman' paper [15] and will not be covered here.

The Gini impurity measure at node t is defined as $i(t) = 1 - S$, where S (the impurity function) $= \sum p(j|t)$, for j = 1,2,…,k [14]

The impurity function attains its maximum if each class in the population occurs with equal probability. That is $p(1|t) = p(2|t) = ... = p(j|t)$. On the other hand, the impurity function attains its minimum (= 0) if all cases at a node belong to only one class. That is , if node t is a pure node with a zero misclassification rate, then $i(t) = 0$.

Let s be a split at node t. then, the goodness of split "s" is defined as the decrease in impurity measured by:

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

Where:

s = a particular split,

P_L = the proportion of the cases at node t that go into the left child node, t_L,

P_R = the proportion of cases at node t that go into the right child node, t_R,

$i(t_L)$ = impurity of the left child node, and

$i(t_R)$ = impurity of the right child node.

2.4.3.3 Class Assignment Rule

There are two rules for assigning classes to nodes. Each rule is based on one of two types of misclassification costs.

1. The Plurality Rule: Assign terminal node t to a class for which $p(j|t)$ is the highest. If the majority of the cases in a terminal node belong to a specific class, then that node is assigned to that class. The rule assumes equal misclassification costs for each class. It does not take into account

the severity of the cost of making a mistake. This rule is a special case of rule 2.

2. Assign terminal node t to a class for which the expected misclassification cost is at a minimum. The application of this rule takes into account the severity of the costs of misclassifying cases or observations in a certain class, and incorporates cost variability into a Gini splitting rule.

When dealing with famine vulnerability, for example, misclassifying a vulnerable household as invulnerable has more severe consequences than misclassifying a invulnerable household as vulnerable. Variable costs can be accounted for by defining a matrix of variable misclassification costs that can be incorporated into the splitting rules.

Let $c(i|j)$ = the cost of classifying a class j case as a class I case:

$$c(i|j) \geq 0 \quad \text{if} \quad i \neq j, c(i|j) = 0 \quad \text{if} \quad i = j \qquad (2)$$

Now, assume that there are two classes in a problem. Let

$\pi_t(1)$ = prior probability of class 1 at node t,

$\pi_t(2)$ = prior probability of class 2 at node t,

$r_1(t)$ = the cost of assigning node t to class 1, and

$r_2(t)$ = the cost of assigning node t to class2.

Given priors and variable misclassification costs, $r_1(t)$ and $r_2(t)$ are estimated as follows:

$$r_1(t) = \pi(1)\,c(2|1)$$
$$r_2(t) = \pi(2)\,c(1|2) \qquad (3)$$

According to rule 2, if at node t, $r_1(t) < r_2(t)$, node t is assigned to class1. If $c(2|1) = c(1|2)$, then rule 1 applies and a node is assigned to a class for which the prior probability is the highest.

2.4.3.4 Steps in building a CART like tree

The tree building process begins with departing the root node into binary nodes by a very simple question of the form is $\mathbf{X} \leq d$? Initially, all observations are located in the root node. CART implement 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 using the formula (1).

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 reduced 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.

Step 6: Stopping tree building, CART stops the splitting process when:

1) There is only one observation in each of the child nodes;

2) All observations within each child node have the identical distribution of predictor variables, making splitting impossible.

3) The user set an external limit on the number of levels in the maximal tree in previously.

Stand by there steps a CART algorithm base decision tree without pruning and optimizing will be built.

2.4.4 Random forest algorithm

RF has greatly improved 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 random vectors generation via random selection procedure from integrated training set. The constituent in this method is that we prepare k random vectors, Θk, which is independent of the past random vectors $\Theta 1$, $\Theta 2$, $\Theta 3$, …, Θk-1 but with the same distribution to build the trees among the RF. Corresponding individual classifier is noted by $C(\mathbf{X}, \Theta k)$. And then they vote for the most popular class. Breiman names these procedures as random forests. A definition drawn from original paper is available here [10, 11].

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

2.4.4.1 Two Randomized procedures in RF Tree building

As mentioned below, RF enhances 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, assuming 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*. 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.

The two distinctive randomized procedures exist among four steps below. That is, RF extracts a fixed quantity from training set randomly, or names it *bagging* process [16]. 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. Thus, each classifier in the ensemble is 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. 2.5) 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 data is used to build the single tree classifier.

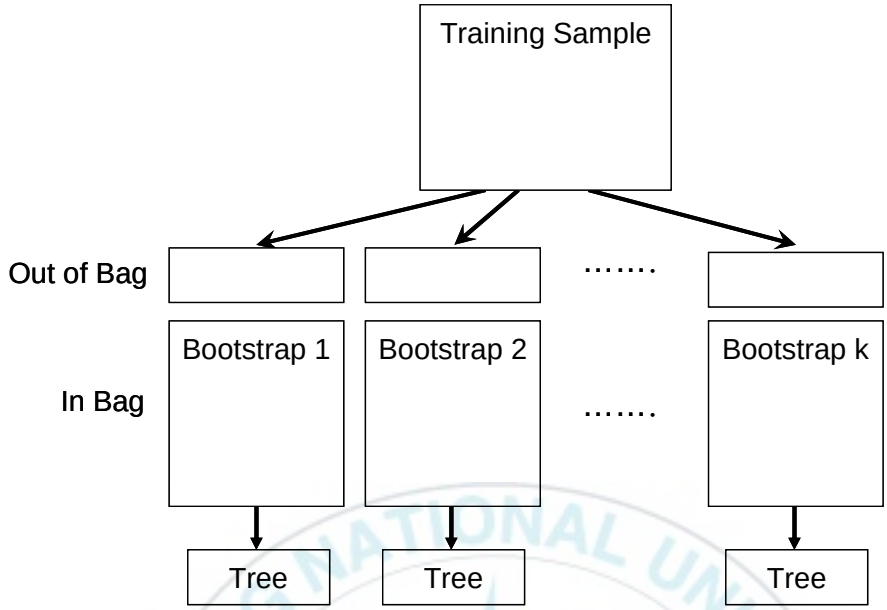


Fig. 2.5. Schematic of bagging using the decision tree as the base classifier

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 randomly 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.

2.4.4.2 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 too long or where training examples are rare, the learner

may be limited in very specific random features of the training data that have no causal relation to the target function. But RF can avoid the over-fitting completely, which was proved in Refs [10]. To affirm this point, we define a margin function at first.

Given an ensemble of a series of classifiers C1(**X**), C2(**X**), …, C*k*(**X**), and with the training set drawn at random from the distribution of the random vector *Y*, **X**, define the margin function as

$$mg(\mathbf{X},Y) = av_k \, \mathrm{I}(C_k(\mathbf{X}) = Y) - \max_{j \neq Y} av_k \, \mathrm{I}(C_k(\mathbf{X}) = j) \qquad (4)$$

where I(·) is the indicator function. The margin measures the extent to which the average number of votes at **X**, *Y* for the right class exceeds the average vote for any other class. The larger in the margin, the more confidence in the classification.

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

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

Theorem 1 As the number of trees increases, for almost surely all sequences Θ1, PE^* converges to:

$$P_{\mathbf{X},Y}\left(P_\Theta\left(C(\mathbf{X},\Theta) = Y\right)\right) - \max_{j \neq Y} P_\Theta\left(\left(C(\mathbf{X},\Theta) = j\right) < 0\right) \qquad (5)$$

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 do not

overfit when more trees are added, meanwhile it result in a limiting value of the generalization error. This is another important feature of RF beside the two randomized procedure mentioned above.

2.4.4.3 Accuracy of RF Depending on Strength and Correlation

In last section, the anti-overfitting characteristic of RF is proved. But we concern more about its accuracy. According to the analysis built in references, an upper bound of RB 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 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 random forest is:

$$mr(X,Y) = P_{\Theta}\left(C(\mathbf{X},\Theta) = Y\right) - \max_{j \neq Y} P_{\Theta}\left(C(\mathbf{X},\Theta) = j\right) \qquad (6)$$

The raw margin function is:

$$rmg(\Theta,\mathbf{X},Y) = I\left(C(\mathbf{X},\Theta) = Y\right) - I\left(C(\mathbf{X},\Theta) = \hat{j}(\mathbf{X},Y)\right) \qquad (7)$$

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

$$S = E_{X,Y} mr(\mathbf{X},Y) \qquad (8)$$

Then we compute the variance of margin function:

$$\operatorname{var}(mr) = \bar{\rho}\left(E_{\Theta} sd(\Theta)\right)^2 \le \bar{\rho} E_{\Theta} \operatorname{var}(\Theta) \tag{9}$$

Write

$$\mathrm{E}_{\Theta} \operatorname{var}(\Theta) \le \mathrm{E}_{\Theta}\left(\mathrm{E}_{\mathbf{X},Y} rmg(\Theta, \mathbf{X}, Y)\right)^2 - S^2 \le 1 - S^2 \tag{10}$$

Considering function (9), (10) and Chebychev inequality, theorem 2 can be concluded.

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

$$PE^* \le \bar{\rho}\left(1 - S^2\right) / S^2 \tag{11}$$

Although the bound is likely to be loose, it fulfills the same suggestive function for random forest as VC-type bounds do for other types of classifiers. It shows that the two ingredients involved in the generalization error for random forests 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, that is the smaller this ration is, the better performance RF provided.

2.4 Genetic Algorithm

RF is strengthened by a standard genetic algorithm (GAs) [19] 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 searches procedure that simulates the evolutionary process by applying genetic operators. Comparing 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. 2.2) 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 step 3 to 5 until the evolved result satisfy 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.

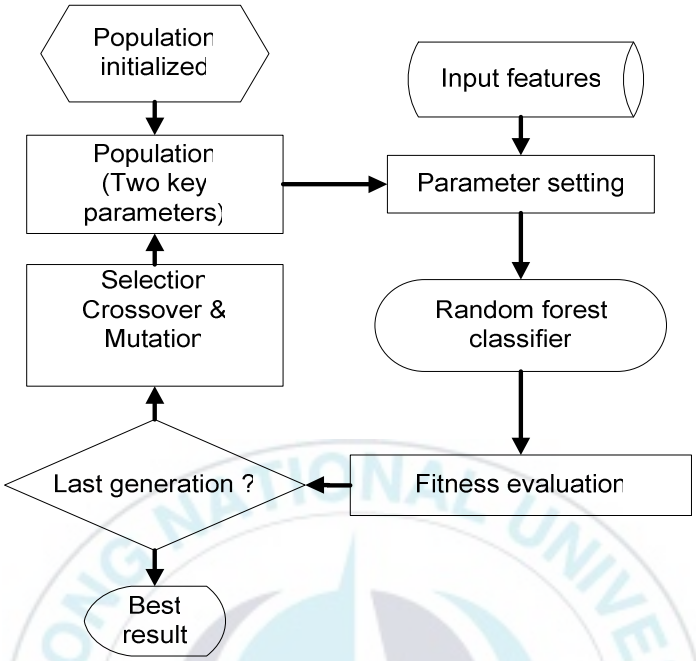


Fig. 2.6. Flowchart of genetic algorithm.

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, *Np* is described as following equations:

$$N_p(i) = \frac{B_s(i)}{1-\left(1-B_s(i)\right)^{N_g(i)}} \quad (11)$$

where *i* is an individual, *Ng* is the number of generation. *Bs* is probability of selecting best individual from the current population.

The selection probability of each individual is:

$$P_s(i) = N(i)\left(1 - B_s(i)\right)^{I(i)} \quad (12)$$

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.

III. Application and Optimization of Random Forest Algorithm on Induction Motor Fault Diagnosis

3.1 The Significance of Intelligent Diagnosis of Rotating Machine

The rotating machine plays an important role of modern industry, and is equipped in many crucial departments. This situation causes the problem that its breakdown will result in a huge losing. Therefore, fault diagnosis of machines is gaining importance in industry because of its capability to increase reliability and to decrease possible loss of production due to machine breakdown. Efficient and accurate faults categorized have been critical to machinery operated in normal condition. There are several well-known machine learning methods which also named artificial intelligence, such as artificial neural network (ANN), support vector machines (SVM) etc. The new techniques and their extended research increase the intelligent, preciseness and applicability of diagnosis domain. It exhibits the great potential of combining machine learning methodology and machinery faults diagnosis theoretic. While the passion of developing machine learning based machinery faults diagnosis methods are increasing, there are a number of obstacles in the presence of researchers. That is, the correct diagnosis of a fault is rather complicated. The reasons are listed as follows:

· Different kinds of faults may result in a certain symptom, or feature extracted from raw data.

· Because of the background noise, some faults are difficult to be

distinguished 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 bar and rotor etc.

Hence, the machine learning based fault diagnosis method which is employed to make hypotheses should powerful enough to category the malfunctions in a correct way. So that, improving the capability of diagnosis is the main motivate to inspirit researchers syncretizing existent technologies and exploring new theories.

In this thesis, we introduce and investigate a novel rotating machinery faults diagnosis methodology based on random forests algorithm (RF) [10, 11]. It built a large amount of decision trees out of sub-dataset from a unique original training set by using *bagging*, acronym of *bootstrap aggregating* which is a meta-algorithm to improve classification and regression models according to stability and classification accuracy. Bagging also reduces variance and helps to avoid over-fitting. This procedure extracts cases randomly from original training data set and the bootstrap sets are used for construct each of decision trees in the RF. Each tree classifier is named component predictor. The RF makes decision by counting the votes of component predictors on each class and then selecting the winner class in terms of number of votes to it.

Since first introduced by Breiman, RF has been employed in various fields such as astronomy, micro-array analysis and drug discovery and otherwise [12, 13]. RF provides good performance in applications in these fields. RF can be a competitor for rotating machinery faults diagnosis, because of these distinctive features as below [10]:

· It is unexcelled in accuracy among current algorithms

· It runs efficiently on large data bases.

· It can estimates of what variables are important 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 he generalization error as the forest building progresses.

Due to these features and its board application, we investigate the performance of RF based faults diagnosis methodology.

As the backbone of modern industry, induction motors play an important role in manufacture, transportation and so on. The squirrel cage induction motor's versatility and ruggedness continue to make it the workhorse of the industry, but that doesn't mean it's invincible. Pushing it too hard for too long can cause the stator, rotor, bearings, and shaft to fail. Numerous industry surveys document which parts fail and how, but very little data is available to explain the reason.

As the industry's approach to maintenance and repair gradually evolves from

reactive and preventive to diagnostic and predictive, it's important to pay more attention to root cause failure analysis. Neglecting to do so often will cause your motors to repeatedly fail and cost you valuable resources and time. So a general study on induction motor faults diagnosis was done. And the result is shown in next section.

In this chapter, we also confirm the possibilities of using random forests algorithm (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 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 [15]. Then random forest diagnosis system detects the certain faulty type bases on these features. So the experiments have designed to indicate the validity and reliability of RF based fault diagnosis method. Experimental result shows the optimized RF based method achieves a very high accuracy by combining RF with GA.

3.2 Induction Motor Faults Diagnosis

3.2.1 Failure surveys on induction motor

It's common to use the results of failure surveys to diagnose the cause of a specific motor failure, but it can be a costly mistake. Most failure survey data for electric motors is influenced by the particular industry, the geographic location, and the combination of the motors in use. Therefore, specific numbers may not

always be relevant.

Not only that, most failure surveys focus on the component that actually failed while neglecting to address the root cause of that failure. For example, such a survey may tell you that a bearing failed, but that isn't the root cause; it's simply the component that failed. The root could be one of several things, but it's not specified.

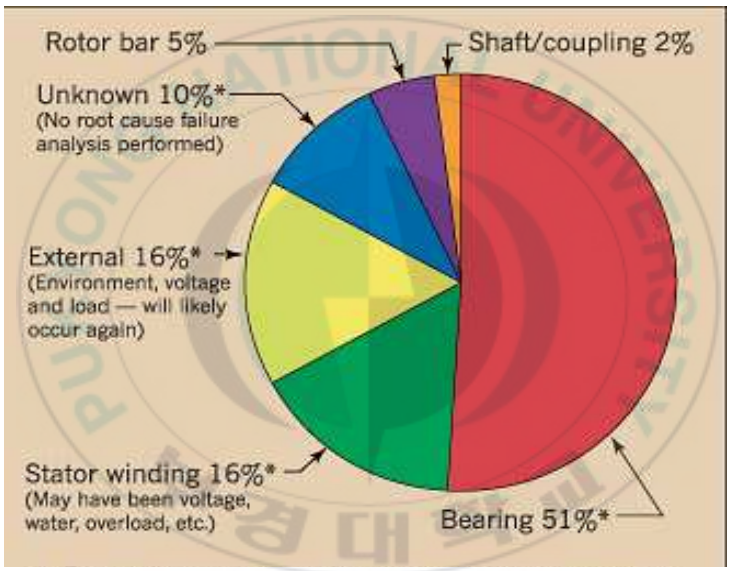


Figure 3.1 IEEE study on induction motors failures

The data provided by the Institute of Electrical and Electronics Engineers (IEEE) study shown in Fig. 1 above is helpful because in addition to identifying failed components, it suggests the most likely causes of failure based on which component failed. However, that's still not enough. It's your responsibility to conduct a thorough analysis to find the definitive root cause of that particular component's failure. These percentages in Figure 4.3 may vary based on industry

or location.

The real challenge lies in reducing the large category of “unknown” failures. It's these “unknown” failures that make analyzing the entire motor system so critical.

3.2.2 Summary of motor stresses

Most motor failures are caused by a combination of various stresses that act upon the bearings, stator, rotor, and shaft. If these stresses are kept within the design capabilities of the system, premature failure shouldn't occur. However, if any combination of the stresses exceeds the design capacity, the life of the system may be drastically reduced and catastrophic failure could occur.

These stresses are classified as follows:

l Bearing stresses: Thermal, dynamic and static loading, vibration and shock, environmental, mechanical, electrical

l Stator stresses: Thermal, electrical, mechanical, and environmental

l Rotor stresses: Thermal, dynamic, mechanical, environmental, magnetic, residual, and miscellaneous

l Shaft stresses: Dynamic, mechanical, environmental, thermal, residual,

electromagnetic

Detailed Summary of Motor Stresses		
Motor component	**Stress type**	**Actual stress or damage**
Bearings	Thermal	Friction, lubricant, ambient
	Dynamic and static loading	Radial, axial, preload, misapplication
	Vibration and shock	Rotor, driven equipment, system
	Environmental	Condensation, foreign materials, excessive ambient, poor ventilation
	Mechanical	Loss of clearances, misalignment, shaft and housing fits
	Electrical	Rotor dissymmetry, electrostatic coupling, static charges, variable-frequency drives
Stator	Thermal	Thermal aging, thermal overload, voltage variation, voltage unbalance, ambient, load cycling, starting and stalling, poor ventilation
	Electrical	Dielectric aging, transient voltages, partial discharge (corona), tracking
	Mechanical	Winding movement, damaged motor leads, improper rotor-to-stator geometry, defective rotor, flying objects
	Environmental	Moisture, chemical, abrasion, poor ventilation, excessive ambient
Rotor	Thermal	Thermal overload, thermal unbalance, excessive rotor losses, hot spots/sparking, incorrect direction of rotation, locked rotor
	Dynamic	Vibration, loose rotor bars, rotor rub, transient torque, centrifugal force/overspeed, cyclical stress
	Mechanical	Casting variations/voids, loose laminations and/or bars, incorrect shaft-to-core fit, fatigue or part breakage, improper rotor-to-stator geometry, material deviations, improper mounting, improper design or manufacturing practices
	Environmental	Corrosion, abrasion, foreign materials, poor ventilation, excessive ambient temperature, unusual external forces
	Magnetic	Rotor pullover, uneven magnetic pull, lamination saturation, noise, circulationg currents, vibration, noise, electromagnetic effect
	Residual	Stress concentrations, uneven cage stress
	Miscellaneous	Misapplication, effects of poor design, manufacturing variations, inadequate maintenance, improper operation, improper mounting
Shaft	Dynamic	Cyclic loads, overload, shock
	Mechanical	Overhung load and bending, torsional load, axial load
	Environmental	Corrosion, moisture, erosion, wear
	Thermal	Temperature gradients, rotor bowing
	Residual	Manufacturing processes, repair processes
	Electromagnetic	Excessive radial load, out-of-phase reclosing

Table 3.1 Detailed summary of motor stresses

3.2.3 Arriving at correct conclusion

When analyzing a motor failure, it's important to not make assumptions. The service center rarely knows much about the motor application, much less the power supply and/or maintenance history. The individual dealing with the service center may not be the person who removed the motor from service or the operator who is familiar with the motor or its application, meaning that it's imperative those individuals compile all of the facts before concluding anything.

Incorrect, incomplete, or even misleading information is the norm. But it doesn't have to be that way. Never assume a piece of evidence exists just to force the conclusion to fit the facts. When a conclusion is built around erroneous information mingled with facts, the root cause of failure is seldom correct. The result will be additional failures or the assignment of blame to the wrong parties.

3.3 Experiment Platform and Motor Faults Data Description

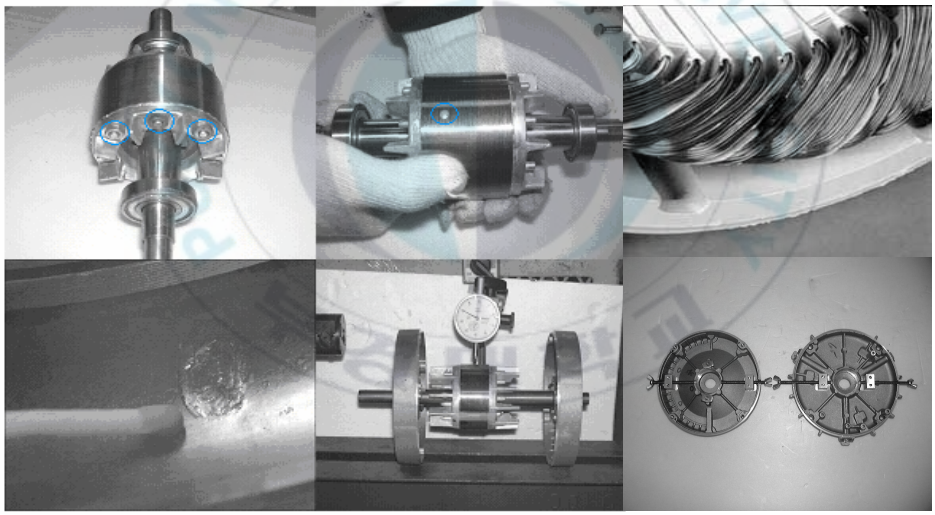
The experiments are designed to simulate six most universal categories of induction motors faults which are broken rotor bar, bowed rotor, bearing outer race fault, rotor unbalance, adjustable eccentricity motor (misalignment) and phase unbalance, first four motor faults are shown in Fig. 3.2 as an example. 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 0.5kW, 60Hz, 4-pole induction motors, pulleys, belt, shaft and fan with changeable blade

pitch angle.

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. Fig. 4.3 shows the platform of the experiment.

After measuring the raw data, a preprocessing and feature extraction are implemented on the data to obtain the most important features for the RF based diagnosis methodology [17]. Finally, there are 63 features left which are prepared for the next procedure, induction motor faults diagnosis by RF.

Rotor Unbalance **Rotor bar broken** **Stator fault**



Faulty bearing **Bowed rotor** **Eccentricity**

Figure 3.2 Faults on induction motors

Figure 3.3 Experiment platform

3.4 Discussion and Analyze

In this section, RF was run on the induction motor faults data. The experimental results for random forest based method are given in Table 3.2. And confusion matrixes for the training data in RF are given by Tables 3.3 shows the accuracies of each fault class for testing data with 1200 trees and selecting 1 variable every split.

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

Table 3.2 Faults diagnosis accuracies based on RF

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	7	3	0	0	0	70
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	1	0	0	0	9	0	90
9	0	0	0	0	0	0	0	0	10	100

Table 3.3 Accuracy of each fault class for testing data with 1200 trees and selecting 1 variable each split

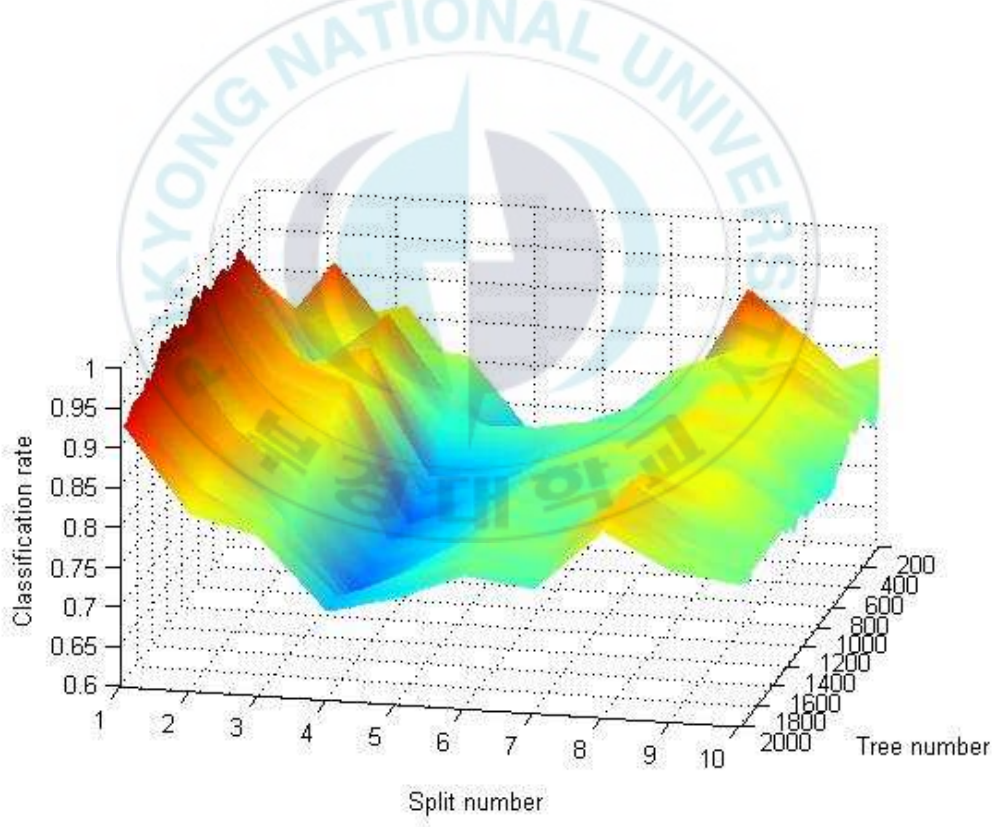


Fig. 3.4. Classification rate against random split number and tree number

Fig. 3.4 shows the classification rate according to the experiment which represe nts three characteristics of RF very clearly. First, compared with the number of co

mponent classification trees, the parameter, random split number at each node, is more sensitive to the classification accuracy. Hence a prudential searching proced ure is necessary to find the best split variables number by an experimental way. Se cond, if the split variables number is decided, the sum of individual tree classifier should achieve an appropriate quantity to get a better performance. Last one, whe n we increase trees into a high number, for example 5000 or 10000, there is no ov er-fitting occurred but a little undulating exists.

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	70
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	1	0	9	0	90
9	0	0	0	0	0	0	0	0	10	100

Table 3.4 Accuracy of each fault class for test data with 907 trees and selecting 1 variable every split

Table 3.4 indicates that incorrect diagnosis of RF based methodology often occurs at certain fault category. So we can apply some assistant diagnosis method which are function in that specific kind of fault to improve the diagnosis precision.

In general, the normal RF has achieved the satisfied fault diagnosis accuracy. B ut it should be noticed that two parameters, the number of trees and random split n umber, which greatly affect classification result are set manually. It means accurac y of normal RF depends on researcher’s experience. This situation exists at almost all the applications of RF. So that applying the genetic algorithm to do the param eter optimization is exigent. The effect of this cooperation is proved by using the s ame data. According to the pervious research, in order to reduce executed time of 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 1to 10 re spectively.

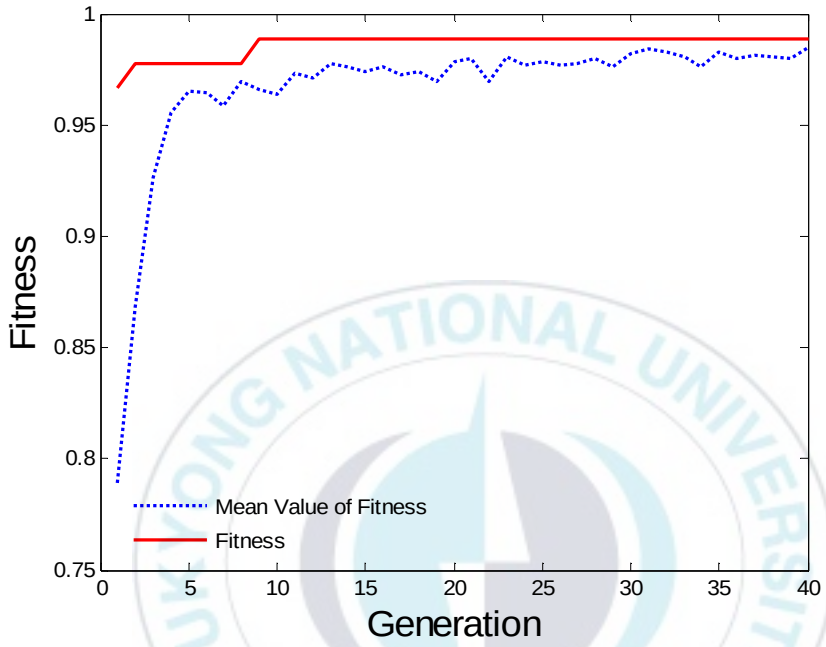


Fig. 3.5. Optimization trace within 40th generation.

Fig. 3.5 shows the trace information of every generation. Fitness adopts the classification accuracy of the test data set. The upper curve 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 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 the 98.89%, 3.33% higher than the best value of normal RF. It means GA can enhance the capability of RF algorithm distinctly.

We also do the comparisons of some familier diagnosis methods. Such as, ANN(BP nural network), support vector machines and C4.5 classification tree. Obviously, RF algorithm has great increased the capability of tree classification

method, from 77.78% to 95.56. Others can also be seen from table 3 that normal RF just 0.69% weaker than BP neural network, but its speed is higher. And RF optimized by GA almost touches the same accuracy of SVM. The RF, a novel algorithm, goes near to a well developed method. This means all we have done is significatively and the farther research is important and necessary.

C4.5	BP-NN	SVM	RF	RFOGA
77.78%	96.25%	99.15%	95.56%	98.89%

Table 3.5 Result and comparisons of ANN,SVM, C4.5, RF and Optimized RF by GA

3.5 Conclusion

The purpose of this chapter is to confirm the possibilities of using 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 amount of decision trees to improve the single tree classifier. Although there are

several existed 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 faults diagnosis. Evaluation of the RF based method has been demonstrated by a case study on induction motor faults diagnosis. Experimental results indicate the validity and reliability of RF based fault diagnosis method. In this paper, the RF and optimized RF based faults diagnosis method of rotating machinery were investigated. The performance of two methods was proved by the faults 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.

IV. Application of RFOGA to Elevator Induction Motor Fault Diagnosis

4.1 Introduction

As elevators are more widely used as transportation means in buildings, importance of guaranteeing elevators working under normal state is also becoming more significant. The sudden breakdown of elevator mechanical system will be result in very inconvenient consequences which may disturb normal step of human life and manufacturing processes and cause a huge loss of time and productivity.

Induction motor is core component of elevator mechanical system. Under long time and under-the-clock running, the degradation and malfunction of elevator induction motor are possibly occurred. Further, the faults of motor may be inherent to the machine itself or caused by severe operating conditions [20]. And it is difficult to trace the root cause too. Therefore, to apply an intelligent fault diagnosis system to elevator door is crucial demand [21].

With this purpose, after testing RFOGA onto induction motor fault stimulation platform, in this chapter, RFOGA based intelligent system was applied to diagnosis of elevator motor faults by using vibration and current signal separately. Because of the advantage of vibration signal, like relative low interference of background noise, good discriminability to different fault types,

and high performance of RFOGA, a 100% diagnostic accuracy is achieved with vibration signal only. However, due to the exigent requirement of vibration data acquisition in real-world, it should be considered when vibration signal is unavailable or incomplete, frequently fault diagnosis accomplished by employing current signal is more convenient and economical. Therefore, RFOGA elevator motor fault diagnosis using stator current signals is evaluated in this paper as well.

The RFOGA intelligent system for elevator motor fault diagnosis works as follow: first, raw data is collected from multiple sensors and values of features of the raw data are calculated that extract most of important information. The generated feature sets are then grouped as the original input of the system to be sent into RFOGA for diagnosing motor faults. The rest of this chapter is arranged as: explanation of experiment apparatus and data, experimental result discussion and conclusion of this chapter.

4.2 Experiment Apparatus and Data Description

In order to demonstrate the effectiveness of the proposed system in real-world operating conditions, an experiment was carried out using an induction motor system of elevator as shown in Fig. 4.1.

The test objects are ten 15 kW, 50 Hz and 4-pole induction motors for elevators. The basic specifications of them are shown in Table 4.1. This motor was set to operate at full-load conditions. One of the motors is normal (healthy), which is used as a benchmark for comparing with faulty motors. The others are faulty motors with rotor unbalance, stator eccentricity, rotor eccentricity, broken rotor bar, bearing housing looseness, bearing inner race looseness, ball fault, bearing

outer race fault and inner race fault, as shown in Fig. 4.2. The conditions of faulty induction motors are described in Table 4.2.

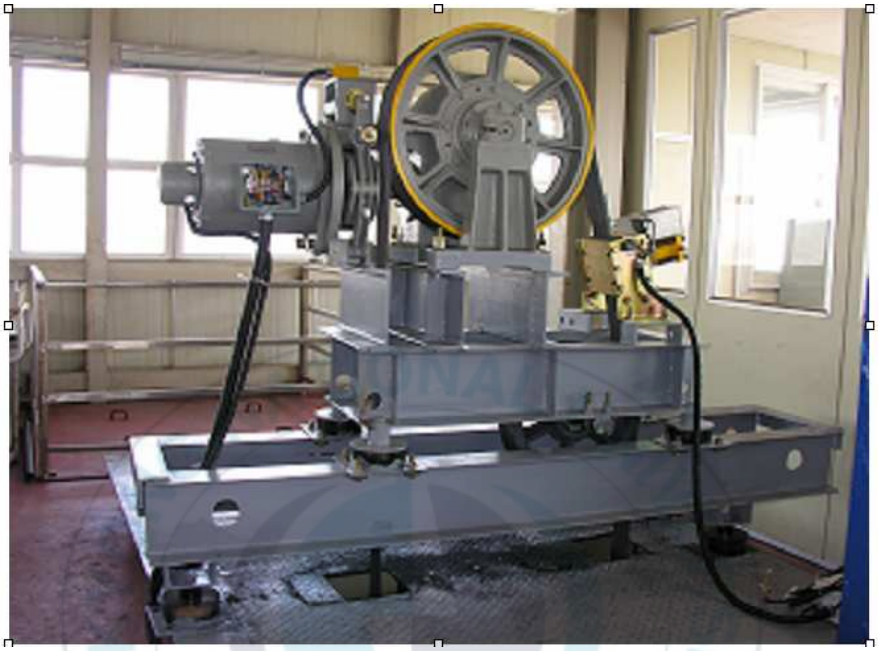


Figure 4.1 Experiment apparatus

Table 4.1 Basic specification of the elevator induction motor

Type	Induction motor
Voltage	340 V
Current	34.2 A
Rotating speed	1450 rpm
Line frequency	50 Hz
Bearing (DE)	#6310
Bearing (NDE)	#6308
Weight	1402 N
Power	15 kW

Number of stator slot	36

Three accelerometers and one AC current probes were used to measure the vibration signals of horizontal, vertical, axial directions and stator current signal to evaluate the fault diagnosis system. The maximum frequency of sampling signals was 3 kHz and the number of sampled data was 16384. Sampling time is 2.133 seconds and Hanning window was chosen for filtering. Each condition was measured for two times.

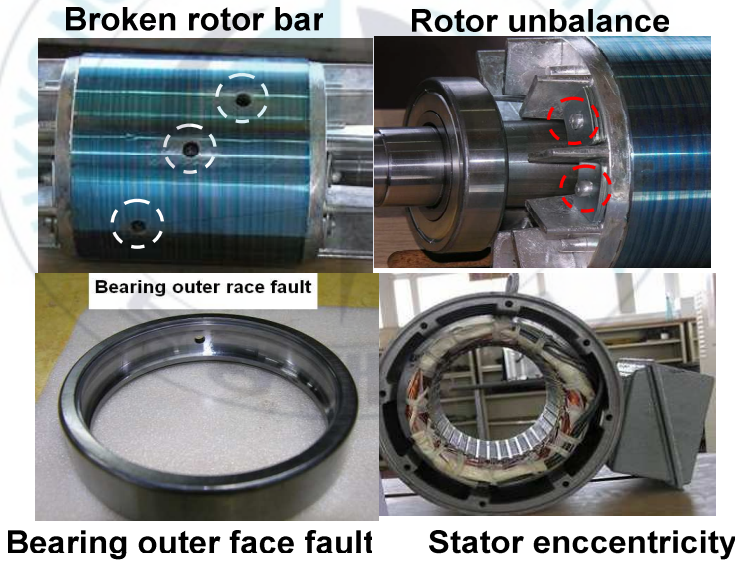


Figure 4.2 fault examples of induction motors

Table 4.2 Description of fault types of the motor tested

Fault types	Fault parameter
Rotor unbalance	In-phase, 60 gmm/kg
Stator eccentricity	30% (+0.23 mm)
Rotor eccentricity	Out-of-phase, 80 gmm/kg
Broken rotor bar	1 spot
Bearing housing looseness	Between outer race and housing
Inner race looseness	Between shaft and inner race
Ball fault	Diameter 2 mm, depth 1.5 mm
Outer race fault	Diameter 2mm depth 2mm
Inner race fault	Diameter 2 mm, depth 2 mm

The permitted measuring time for each fault is 15 seconds containing three running conditions: speed-up, steady and slow-down. Another real limitation is that many times of measurement per fault is nearly impossible, or else the elevator will break down severely. In this experiment, each fault was measured for two times, then steady signals were picked out for analyze. Considering the limit raw data that is not enough for RFOGA diagnosis system, an overlap method was employed to solve the problem. This method picks out each sample using an overlap rate predetermined from collected steady signal in sequence. The overlap rate was set as 0.75 in this experiment.

Using the overlap method, we extended the steady signal of one time measurement into 10 times. So finally we acquire 20 samples per fault and total samples are 200. Among them, 100 samples were divided for training classifiers, 100 samples for testing performance of RFOGA.

After measuring the raw data, a preprocessing and feature extraction are

implemented on the data to obtain the most important features for the RF based diagnosis methodology which is same with the former experiment [17]. Finally, 21 values of features are acquired from each sensor consisting of the time domain (10 features), frequency domain (3 features) and regression estimation (8 features).

4.3 Experiment Result and Discussion

In this section, firstly RF was run on the induction motor faults data without cooperation with GA for saving experiment time. Because base on former research on RF, it is shown that RF often provide the satisfied result with vibration signal and random split number (RFN) 1, which is mentioned by Breiman as well. Fig. 5.3 provides the test result which RSN is set to 1 and tree number (TN) varies from 50 to 2050 with interval 100.

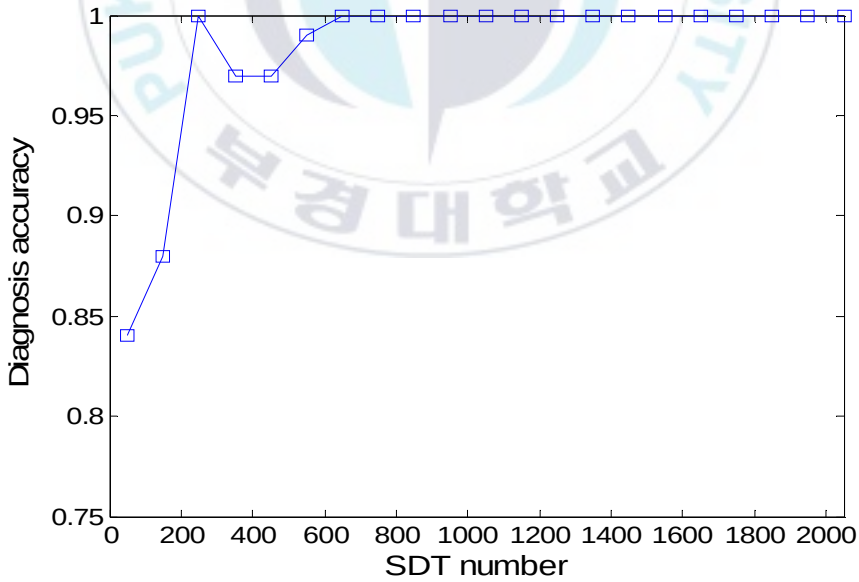


Figure 4.3 Test result with RSN equal to 1 and TN varying from 50 to 2000

Fig. 4.3 expressly shows that RF can arrive at 100% even without support of

GA. Thus it makes using vibration and current signal synchronously meaningless, for current signal just will decrease the accuracy when it works together with vibration signal. But in the former section, it is mentioned that vibration signal could be invariable or incomplete sometimes, thereby an investigation of the feasibility for applying current signal only is necessary. Fig. 4.4 provides the inter-relationship between output of RF and parameters when current signal is employed alone.

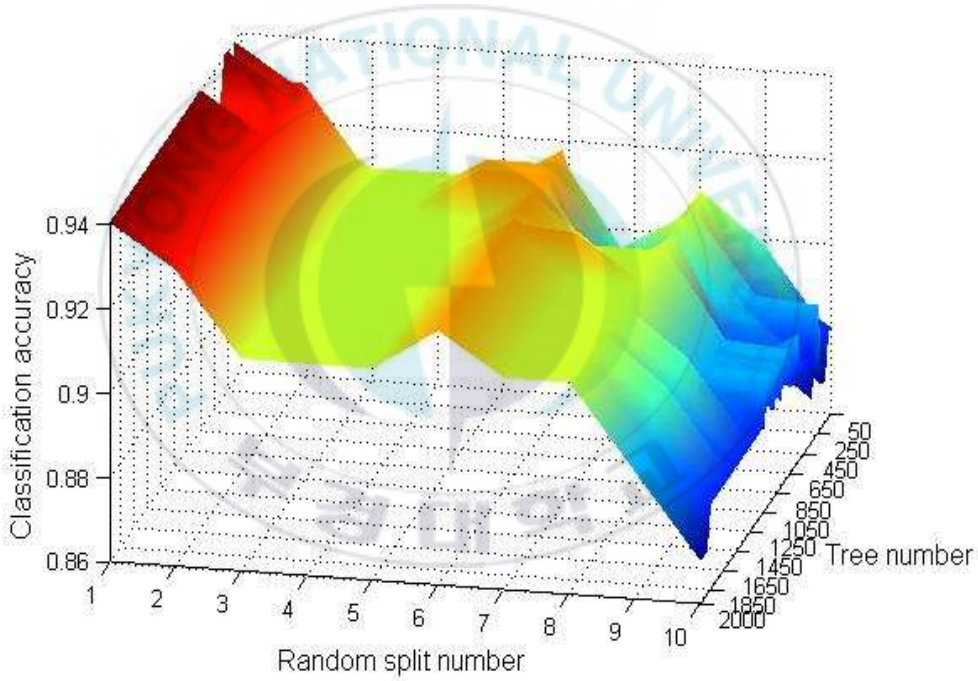


Figure 4.4 Classification rate against RSN and TN

Fig 4.4 improves the three characteristics of RF again. First, compared with the number of component classification trees, the parameter, random split number at each node, is more sensitive to the classification accuracy. Second, if the split variables number is decided, the sum of individual tree classifier should achieve

an appropriate quantity to get a better performance. Last one, when we increase trees into a high number.

Hence GA is adopted here to find the best combination of RSN and TN, and optimization result is given by figure 4.5. The definition of GA' parameters are not changed contrasting to former experiment. And then table 4.3 indicates the accuracies of each fault class for training and testing data while RN equals to 907 trees and RSN equals to 1.

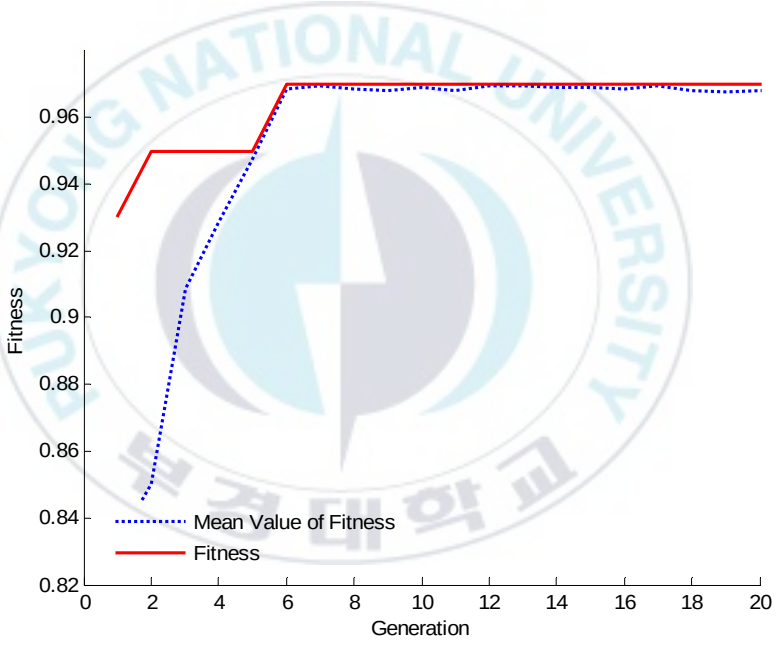


Figure 4.5 Optimization trace within 20th generation.

Figure 4.5 indicates RFOGA based system provides 98% precision. It shows RFOGA does not only enhance the diagnosis accuracy comparing with RF only, almost 4%, but also convinces the efficiency of applying RFOGA with current signal.

The optimized diagnosis output is given by table 4.3. The evolved combination of RSN and TN which are 976 and 1 respectively is provided as well

Comparing diagnosis output from table 4.4 with former one which is provided by table 3.4, it is found that RFOGA makes wrong diagnosis on different type of fault, thus a new characteristic is exposed that RFOGA is not defective on certain fault type, in other words RFOGA is competent for faults detection task without considering the fault type is fit for the system or not.

Table 4.3 Output of RFOGA on elevator induction motor

Class No.	1	2	3	4	5	6	7	8	9	10	Accuracy (%)
1	10	0	0	0	0	0	0	0	0	0	100
2	0	10	0	0	0	0	0	0	0	0	100
3	0	0	10	0	0	0	0	0	0	0	100
4	0	0	0	7	0	0	0	3	0	0	70
5	0	0	0	0	10	0	0	0	0	0	100
6	0	0	0	0	0	10	0	0	0	0	100
7	0	0	0	0	0	0	10	0	0	0	100
8	0	0	0	0	0	0	0	10	0	0	100
9	0	0	0	0	0	0	0	0	10	0	100
10	0	0	0	0	0	0	0	0	0	10	100

4.4 Conclusion

In this chapter, the performance of RFOGA on a real-world application was investigated. Excellent results were achieved in the fault diagnosis of elevator motor using vibration or current signal. By considering difficulties to measure vibration signal in real condition, an effective and cost saving approach has been proposed based on RF that only require analyzing current signals. And it has competitive diagnosis accuracy 94%. Genetic algorithm can improve the accuracy

rates remarkably. It increases the diagnosis accuracy from 94% to 97% when apply the current signal.

The comparison with experiment on fault diagnosis of normal induction mot or denotes the universal adaptability of RFOGA system, because it do not have dis tinct soft spot on some induction motor fault types. So this algorithm has widely p erspective on induction motor fault diagnosis of various fields.

V. Conclusion and Future Work

In this thesis, the importance of machinery fault diagnosis is stated. Both of the history and popular methods used nowadays are included here. The AI Machine learning and ensemble theory are also discussed to lead the readers to know how Random Forest Algorithm comes from.

Most import is that the RF and optimized RF based faults diagnosis methodology of rotating machinery were investigated in the thesis. The performance of two methodologies was proved by the faults diagnosis test of an induction motor. The optimized approach attains a high correct rate of diagnosis, 98.89%. And the comparison result also shows that optimized RF based method is competitive with other classification method. In addition, the assemble classification trees method and even faster then some of them [10, 18], it is proved by other multi-classes classification applications.

But the weakness of RFOGA is also distinct. The reason is that RFOGA is based on decision tree (DT) classifier, and the capability of DT is not outstanding among a lot of existed AI methods. Affirmatively DT has been losing the interest of researchers. But ensemble theory is an excellent and high-speed developing methodology. It can be cooperated not only with DT but also many other kind of advanced and precise AI methods, i.e. ANN, SVM .etc.

All in all, extended research will focus on two parts. In the near future, my task is to improve on this hybrid method RFOGA: GA is not only for the parameter optimization, it can be used to select best combination of sub-classification trees from the forest to get the more accurate result. The second part, we will decrease the redundancy of the RF and try other optimization algorithm or more effective voting principle. For long views, a combination of ensemble theory, various precise AI algorithms and GA, or other optimization method, will be explored.

Reference

[1] Breiman, L (1996c). Stacked regressions. Machine Learning, 24(1) 49-64

[2] kohavi,R & Vedelsby, J. (1995). Neural network ensembles, cross validation, nd active learning. In Tesauro, G., Touretzky, D., & Leen, t. (Eds.), Advances in Neural Information Processing Systems. Vol.7, OO, 231-238 Cambridge, MA. MIT Press

[3] Hashem, S. (1997). Optimal linear combinations of neural networks. Neural Networks 10 (4), 599-614

[4] Freund, Y., & Schapre, R. (1996). Experiments with a new boosting algorithm. In Proceedings of the Thirteenth International Conference on Machiine Learning, PP. 148-156

[5] Bauer, E,. & Kohavi, R. (1999). An empirical comparision of voting classification algorithms: Bagging, Boosting, and variants. Machine Learning, 36, 105-139

[6] Alpaydin, E. (1993). Multiple networks for function learning. In Proceeding of the 1993 IEEE International Conference on Neural Networks, Vol. I, PP. 27-32 San Francisco

[7] Hansen, L., & Salamon, P. (1990). Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, 993-1001

[8] Efron, B., & Tibshirani, R. (1993). An Introduction to the Bootstrap. Chapman and Hall, New York

[9] Freund, Y., & Schapire, T. (1996). Experiments with a new boosting algorithm. In Proceedings of the Thirteenth International Conference on Machine Learning, pp. 148-156 Bari, Italy

[10] Breiman L. (2001) Random forests. Machine Learning, 40(1), 5-32

[11] Breiman L., Random Forest User Notes. ftp://ftp.stat.berkeley.edu/pub/users/breiman/notes_on_random_forests_v2.p

df

[12] Gislason PO, Benediktsson JA & Sveinsson JR (2004) Random forest classification of multisource remote sensing and geographic data. Proceedings of IGARSS ’04 IEEE International, 2, 1049-1052

[13] Remlinger K (2003) Introduction and application of random forest on high throughput screening data from drug discovery. Proceedings of Workshop for the SAMSI Program on Data Mining and Machine Learning

[14] L. Breiman L (1996) Bagging predictors. Machine Learning, 24, 123-140

[15] Yang BS, Han T & An JL (2004) ART-KOHONEN neural network for fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing, 18, 645-657

[16] Pal M (2003) Random forests for land cover classification. Geoscience and Remote Sensing Symposium, IGARSS

[17] Dargupta H & Dutta H (2004) Orthogonal decision trees. Proceedings of 4th IEEE International Conference on Data Mining, 427-430

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

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

[20] A.H. Bonnett, Root causes AC motor failure analysis with a focus on shaft failures, IEEE Transactionson Industry Applications, 36 (5) (2003) 1435-1448.

[21] Shui Yuan et al., Intelligent diagnosis in electromechanical operation systems, IEEE, in Proceedings of International Conference on Robotics & Automation, New Orleans, LA, 2004.

Acknowledgements

This thesis could not have been completed without the help and continuous support from professors, colleagues and friends to whom I am most grateful.

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Bo-Suk Yang, Pukyong National University. He has directed my two years research and study, provided me many valuable direction and ideas. The most important thing is that Professor does not only give the supervision, direction on academe but also on the Weltanschauung.

I would like to appreciate to , Prof. Dong-Jo Kim, Prof. Soo-Jong Lee for their detailed comments and suggestion during the final phases of the preparation of this thesis.

I am also grateful to all members of Intelligent Mechanics Lab. for giving me a comfortable and active environment to achieve my study: Dr. Tian-Han, Mr. Nui-Gang, Dr Jin-Dae Song. Dr Yong-Mo Kong, Mr. Jong-Yong Ha, Mr. Achmad Widodo, Mr. Jong-Duk Son, Jae-Gab Lee, Seon-Hwa Kim, Ae-Hee Song Baek-Seok Kim, Min-Chan Shim, who have helped me a lot when I put the first step here to get aware of Korea life and the activities in Lab

Last but not least, it is my pleasure to thank my parents and my parents-in-law, my brothers, my sisters, my brother-in-law, my sister-in-law, my nephews and my all relatives for their endless encouragement.