JPE 8-1-10

# Decision Tree with Optimal Feature Selection for Bearing Fault Detection

Ngoc-Tu Nguyen* and Hong-Hee Lee†

†*School of Electrical Engineering, University of Ulsan, Ulsan, Korea

**ABSTRACT**

In this paper, the features extracted from vibration time signals are used to detect the bearing fault condition. The decision tree is applied to diagnose the bearing status, which has the benefits of being an expert system that is based on knowledge history and is simple to understand. This paper also suggests a genetic algorithm (GA) as a method to reduce the number of features. In order to show the potentials of this method in both aspects of accuracy and simplicity, the reduced-feature decision tree is compared with the non reduced-feature decision tree and the PCA-based decision tree.

**Keywords:** bearing fault, diagnostics, decision tree, genetic algorithm, principal component analysis

## 1. Introduction

Induction motors are becoming an increasingly important part in recent industrial processes. Predictive condition monitoring and fault diagnosis can prevent the motor from breaking by alarming against future damages. Some research shows that more than 40% of motor damage cases are related to the bearing. Many methods have been developed to detect motor failures, such as the decision tree [1]-[4], support vector machine [1], [5], artificial neural network [7], [9], etc.

The use of decision tree for bearing diagnosis [1] and multi-fault [2]-[4] has proven good performance in the classification. In order to guarantee a good classification result, the input data need special preprocessing. Recently, many methods have been suggested for data preparation, which use feature selection and feature extraction techniques. Independent component analysis (ICA) and principal component analysis (PCA) are popular feature extraction techniques. They are used to decrease the data dimension by extracting as much useful information as possible from the given data set. A decision tree using PCA technique for fault diagnosis has shown encouraging results [2].

In this paper, the bearing condition is detected by using the decision tree with a feature selection technique. Three axes vibration signals are measured and processed to give 18 time domain features. In order to reduce the number of the features, a GA is suggested to select optimal ones. The main difference between the proposed decision tree and the PCA-based decision tree is the way in which data is processed. The PCA can extract almost useful information from the given data to form a new smaller dimension data set, but it also simply removes a little useful information. There is no guarantee that this removed information is not necessary for bearing fault detection. Meanwhile, the GA

Manuscript received May 14, 2007; revised Dec. 12, 2007
† Corresponding Author: hhlee@mail.ulsan.ac.kr
Tel: +82-52-259-2187, Fax: +82-52-259-1686, Univ. of Ulsan
*School of Electrical Engineering, Univ. of Ulsan

chooses a small part of the given data based on the distance criteria technique. Selecting a small part in the feature set can remove the redundant and irrelevant information that spoils the classification performance. To classify the bearing conditions, the selection method is chosen by only the most appropriate features which have a small average distance inside certain classes and a big average distance between different classes. Efficiency of the proposed method for bearing diagnostics is shown in the experimental section with comparing to PCA-based technique.

## 2. Decision Tree Algorithm

The decision tree is a diagnostic tool that builds the knowledge-based system by the inductive inference of case histories. A typical decision tree contains:

- Leaf nodes (or answer nodes) which contain names.
- Decision nodes (or non-leaf nodes) that specify some test to be carried out on a single attribute value, with one branch and sub-tree for each possible outcome of the test.

The structure of the decision tree highly depends on how to select the root of the tree. The criterion for selecting the root of the tree is Quinlan's information theory (information gain) [6]. According to this criterion, the information conveyed by a message depends on its probability. The construction of the decision tree is based on a training set T, which is a set of cases. Each case specifies the values for a collection of attributes and for a class. Let the classes be denoted by $\{C_1, C_2, \ldots, C_k\}$. Suppose there is a possible test with n outcomes that partition the training set T into the subsets $T_1, T_2, \ldots, T_n$. Assume S is any set of cases, $freq(C_i, S)$ is the number of cases in S that belong to class $C_i$, and $|S|$ is the number of cases in set S. If one case is selected randomly from set S and it belongs to class $C_j$, the message has probability

$$freq(C_j, S)/|S| \tag{1}$$

and the information it conveys is

$$-\log_2\left(freq(C_j, S)/|S|\right) \quad \text{bits} \tag{2}$$

The expected information needed to identify the class of case in S is

$$\text{info}(S) = -\sum_{j=1}^{k} freq(C_j, S)/|S| \times \log_2\left(freq(C_j, S)/|S|\right) \tag{3}$$

When it is applied to the set of training cases, info(T) measures the average amount of information needed to identify the class of a case in T.

A similar measurement after T has been partitioned in accordance with n outcomes of a test X

$$\text{info}_X(T) = \sum_{i=1}^{n} \left(|T_i|/|T|\right) \times \text{info}(T_i) \quad \text{bits} \tag{4}$$

The quantity

$$\text{Gain}(X) = \text{info}(T) - \text{info}_X(T) \tag{5}$$

measures the information that is gained by partitioning T in accordance with the test X. The gain criterion selects a test to maximize this information gain. In this paper, each feature represents a continuous attribute of the decision tree. A training set and a test set are collected to build and evaluate the decision tree.

## 3. Time Domain Features

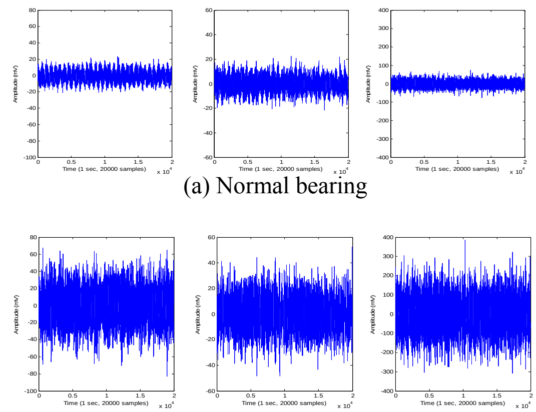
Tri-axial accelerometer is mounted to measure the vibration signal in x, y, and z directions. Six features are extracted from each signal direction. The feature set has a total of 18 features which are used to train the decision tree.

Table 1 Time-domain features

| Feature | Equation |
|---|---|
| Root mean square | $rms = \sqrt{\dfrac{\sum_{n=1}^{N} (x(n))^2}{N}}$ |
| Variance | $\text{var} = \sigma^2 = \dfrac{\sum_{n=1}^{N} (x(n) - mean(x))^2}{(N-1)}$ |
| Skewness | $skewness = \dfrac{\sum_{n=1}^{N} (x(n) - mean(x))^3}{(N-1)\sigma^3}$ |
| Kurtosis | $kurtosis = \dfrac{\sum_{n=1}^{N} (x(n) - mean(x))^4}{(N-1)\sigma^4}$ |
| Crest factor | $crest = \dfrac{\max \lvert x(n) \rvert}{rms}$ |
| Maximum value | $\max = \max \lvert x(n) \rvert$ |

The dataset are formed as follows: root mean square(a), variance(a), skewness(a), kurtosis(a), crest factor(a), maximum(a), root mean square (h), variance(h), skewness(h), kurtosis(h), crest factor(h), maximum(h), root mean square (v), variance(v), skewness(v), kurtosis(v), crest factor(v), maximum(v). The terms (a), (h), and (v) are the abbreviation of axial, horizontal, and vertical directions, respectively.

Vibration data of defective and normal bearings are collected to form the feature set that is used to train the decision tree. Fig. 1 shows time signals of a normal and defective bearing in three directions. The 18 features extracted from the signals in Fig. 1 are shown in Fig. 2.



(a) Normal bearing

(b) Defective bearing

Fig. 1 Vibration time signals in 3 directions (from left to right: horizontal, axial, and vertical) of a normal (a), defective bearing (b)

## 4. Feature Reduction

There are two techniques for feature reduction that are used in this work, PCA and GA techniques. Data are preprocessed by them and used to train the decision tree.

### 4.1 Principal component analysis

Principal component analysis (PCA) is a technique for simplifying the data, by extracting the most relevant information from the original dataset and forming a new lower dimension data for analysis. An N-dimensional (zero mean) dataset $x_i$ (i=1, 2, …, m, N<m) is projected on the eigenvectors of its covariance matrix

$$v = U^T x_i \tag{6}$$

where U is an orthogonal matrix containing the eigenvectors of the data covariance matrix C

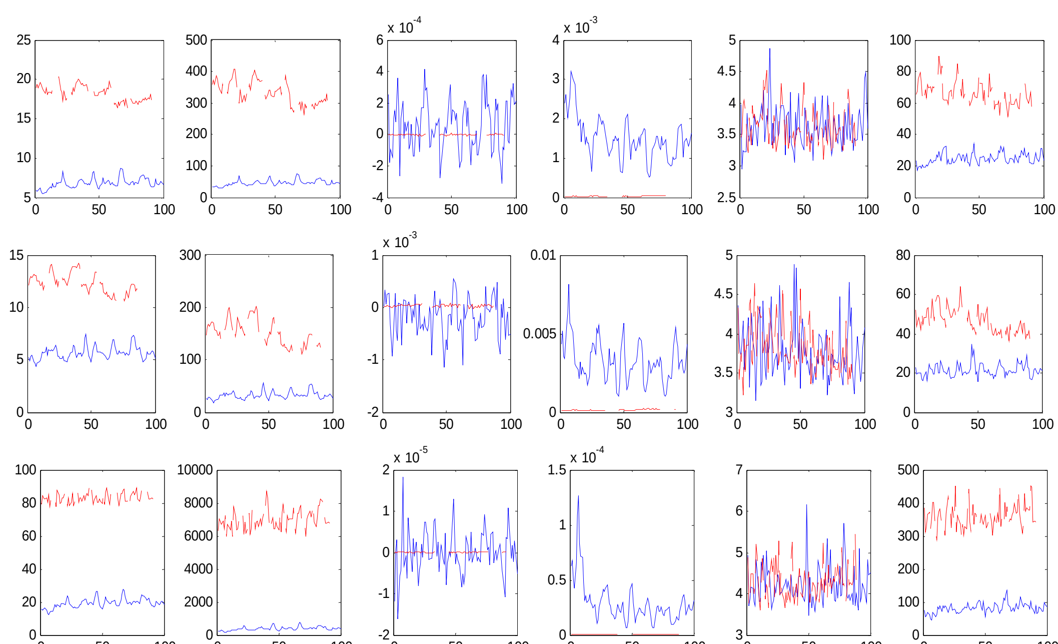


Fig. 2 Extracted features, (- blue) normal and (-- red) defective bearing

$$C = (1/m)\sum_{i=1}^{m} x_i x_i^T \tag{7}$$

The eigenvalues of C is computed and sorted in decreasing order to form matrix U

$$\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_k \geq ... \geq \lambda_N \tag{8}$$

For dimension reduction and preserving the most information in the data, only the most significant eigenvectors are kept. In order to choose the number of eigenvector k, the following criterion is used:

$$\left(\sum_{i=1}^{k} \lambda_i\right) / \left(\sum_{i=1}^{N} \lambda_i\right) > \text{threshold value} \tag{9}$$

The dataset used to train bearing diagnostic decision tree has 18 dimensions, and is processed by PCA algorithm. The threshold value should be chosen as large as possible. In this case, with k = 9 (threshold ≈ 1.0), the new dataset has 9 equivalent features.

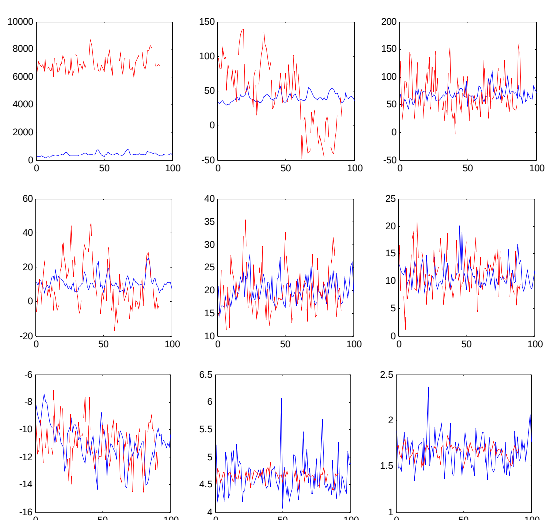


Fig. 3 Reduced feature set based on PCA algorithm: (- blue) normal, and (-- red) defective bearing.

### 4.2 Feature selection using genetic algorithm (GA)

GA is used to give a solution by simulating the evolutionary processes of survival of the fittest, which ensures that the best members of the population are retained. The algorithm begins with a set of solutions (chromosomes) that are called the population. A fitness function of each chromosome is calculated in each generation, and then some chromosomes are selected based on their fitness. These selected chromosomes are modified by crossover and mutation techniques to reproduce a new generation. The reproduction is repeated until the best solution is created.

In this paper, a GA is used for feature selection. This method can select the best features in the feature set which will improve the classification performance. The GA chooses a subset in 18 features that are extracted from three axes vibration time signals.

The chromosome of the GA represents the feature sequences. Binary code of the chromosome includes 18 bits which represent the selection of features. Bit “1” represents the selected feature and bit “0” represents the abandoned feature.

The objectives of the GA are:

The average within-class distance

$$J_c = \sum_{i=1}^{c} p_i J_i \tag{10}$$

$$J_i = \left(1/n_i\right)\sum_{k=1}^{n}\left(x_k^i - m_i\right)^T \left(x_k^i - m_i\right),\ i = 1, \ldots, c \tag{11}$$

where, c is the number of classes; $m_i$ is the mean vector of class i; $n_i$ is the number of samples in class i; $p_i$ is the ratio factor between number of samples in class i and total samples.

The average between-class distance

$$J_b = \sum_{i=1}^{c} p_i \left(m_i - m\right)^T \left(m_i - m\right) \tag{12}$$

where m is the mean vector of all classes.

The GA is processed with two objectives: the first objective is to minimize the average within-class distance $J_c$ and the second is to maximize the average between-class distance $J_b$. The fitness function of the GA can be defined as:

$$J = J_c + \left(1/J_b\right) \tag{13}$$

The chromosome which minimizes the fitness function is chosen. Therefore, according to the smaller value of J, the optimal features can be selected. After choosing the features, C4.5 algorithm can be applied to build the decision tree for classifying bearing condition. C4.5 is an algorithm used to build the decision tree, introduced by J. R. Quinlan. More detail about C4.5 can be found in [6].

## 5. Experimental Results

The local optimum selection of features obtained by the GA is used to train the decision tree. Test results are compared with the normal 18 features (non reduced-feature) decision tree and the decision tree based on PCA algorithm (PCA based decision tree). Parameters of the GA are set as follows: population is 48, length of chromosome is 18, mutation probability is 0.1, crossing probability is 0.7, 1 crossing point, and number of generation is 50. The experiment result is 000100100000000100 which means the features F4, F7, and F16 are selected.

Table 2 Performance comparison of non reduced-feature, PCA based, and reduced-feature decision trees

| Type | Size | Evaluation on training data | Evaluation on test data |
|---|---|---|---|
| Non reduced-feature decision tree | 25 | 99.6% | 95.8% |
| PCA-based feature decision tree | 23 | 99.9% | 100% |
| Reduced-feature decision tree | 11 | 99.8% | 95.8% |

The same training set with 1372 samples and 192 samples test set is used to evaluate the trees. Evaluation results are shown on Table 2.

From Table 2, the reduced-feature decision tree using the GA has the smallest size compared to the others, but it has the same accuracy as the non reduced-feature one when they are evaluated on test set. The PCA based decision tree has highest accuracy but its size is still large.

Fig. 4 shows the reduced-feature decision tree, which only uses 2 features, F4 and F7. From Table 2, this tree has a significant improvement compared to the non reduced-feature tree. Fig. 5 shows the PCA-based decision tree which is a little bit smaller than the non reduced-feature one in Fig. 6.

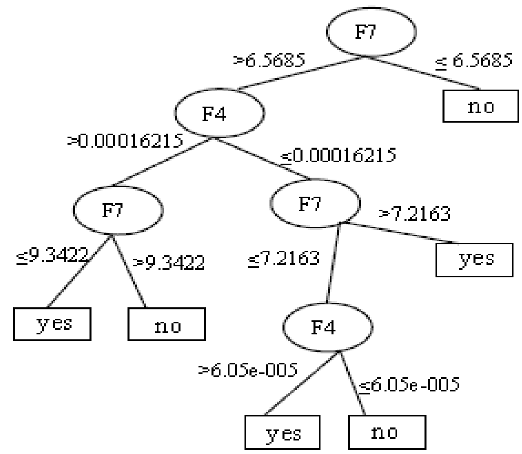



Fig. 4 Reduced-feature decision tree

Compared to the PCA based and non reduced-feature trees, the decision tree for bearing diagnosis using GA based feature selection has the most compact structure. This tree has a depth level 4 compared to 6 of the PCA based tree and 8 of the non reduced-feature tree. As a result, the feature selection reduces the complexity of the decision tree, and therefore increases the efficiency of the classification.

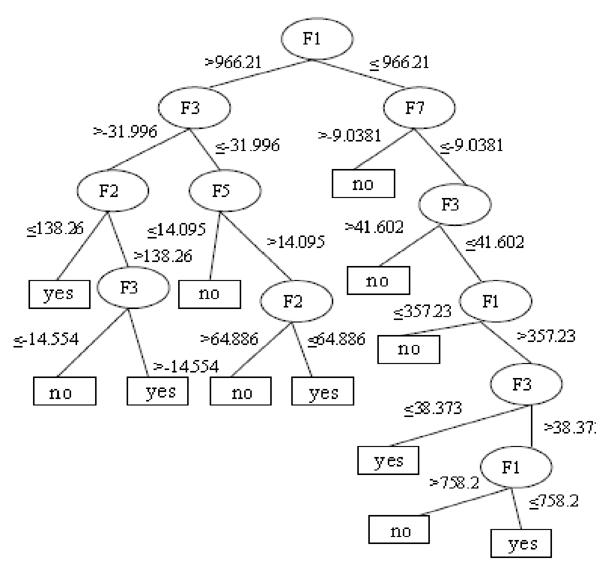



Fig. 5 PCA-based decision tree

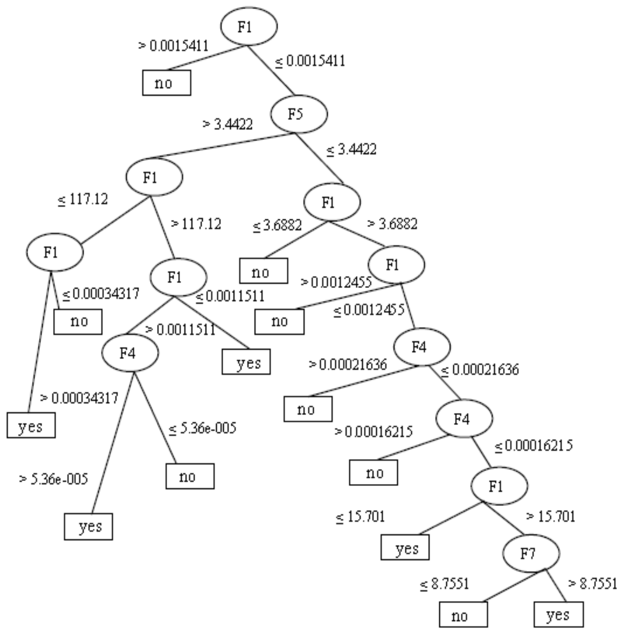



Fig. 6 Non reduced feature decision tree

## 6. Conclusions

In this paper, the decision tree is applied for bearing fault diagnosis with the optimal feature selection which can reduce the tree's size and depth level while the same accuracy can be kept. For simplicity, only two states (fault and normal) of bearing condition are considered, but it is possible to apply this to the multi-fault type system.

The drawback of the decision tree method is the discrete output. Therefore the decision tree cannot give the severity level of fault bearing because it requires the system to have continuous output. Another problem of the decision tree is the sensitivity to noise. If there is a small amount of noise, which is added to attribute values, then the tree can give wrong results. Besides these weak points, the decision tree has a simple construction that can be understood easily. In this paper, the proposed decision tree has very high accuracy when it is evaluated with the test set.

## Acknowledgments

The authors would like to thank the University of Ulsan, the Ministry of Commerce, Industry and Energy (MOCIE) and Ulsan Metropolitan City that partly supported this research through the Network-based Automation Research Center (NARC).

## References

[1] V. Sugumaran, V. Muralidharan, K. I. Ramachandran, "Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing", Mechanical Systems and Signal Processing 21, pp. 930~942, 2007.

[2] W. Sun, J. Chen, J. Li, "Decision tree and PCA-based fault diagnosis of rotating machinery", Mechanical Systems and Signal Processing, 2006.

[3] B. S. Yang, C. H. Park, H. J. Kim, "An Efficient Method of Vibration Diagnostics for Rotating Machinery using a Decision Tree", International Journal of Rotating Machinery, Vol.6, No.1, pp. 19~27, 2000.

[4] D. S. Lim, B. S. Yang, D. J. Kim, "An Expert System for Vibration Diagnosis of Rotating Machinery using Decision Tree", International Journal of COMADEM, pp.31~36, 2000.

[5] A. Widodo, B. S. Yang, T. Han, "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors", Expert Systems with Applications 32, pp. 299~312, 2007.

[6] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publisher, Inc., 1993.

[7] B. Samanta, K. R. Al-Balushi, "Artificial Neural Network based fault diagnostics of rolling element bearings using time-domain features", Mechanical Systems and Signal Processing 17, pp. 317~328, 2003.

[8] H. Gu, Z. Gao, F. Wu, "Selection of Optimal Features for Iris Recognition", in International Symposium on Neural Networks, China, pp. 81~86, 2005.

[9] B. Samanta, K. R. Al-Balushi, S. A. Al-Araimi, "Artificial neural networks and genetic algorithm for bearing fault detection", Soft Coput., pp. 264~271, 2006.

[10] T. Lindh, J. Ahola, P. Spatenka, A-L Rautiainen, "Automatic bearing fault classification combining statistical classification and fuzzy logic", in NORPIE, 2004.

[11] J. S. Rao, Vibratory Condition Monitoring of Machines, Alpha Science International Ltd., pp. 361~382, 2000.

**Ngoc-Tu Nguyen** He received his B.S. and M.S. degrees in the Faculty of Electronics and Electrical Engineering, Ho Chi Minh City Uni. of Technology, Vietnam. He is currently a Ph.D. student at the School of E.E., University of Ulsan, Korea. His research interests are electrical machines, fault diagnostics and power electronics.

**Hong-Hee Lee** He received his B.S., M.S., and Ph.D. degrees from Seoul National University, Seoul, Korea. Currently, he is a Professor at the School of Electric-Electronic Information System Engineering, University of Ulsan, Ulsan, Korea. He is also Director of the Network-based Automation Research Center (NARC), which is sponsored by the Ministry of Commerce, Industry and Energy (MOCIE). His research interests are power electronics, network-based motor control, and control network. He is a member of IEEE, KIEE, KIPE, and ICASE.