

## Optimal feature selection using genetic algorithm for mechanical fault detection of induction motor

Ngoc-Tu Nguyen, Hong-Hee Lee\* and Jeong-Min Kwon

*School of Electrical Engineering, 7-409, Namgu, Mugeo 2-dong, Ulsan Univ., Ulsan, South Korea*

(Manuscript Received May 7, 2007; Revised September 18, 2007; Accepted September 19, 2007)

---

### Abstract

Time-domain vibration signals are measured in all horizontal, axial, and vertical directions for induction motor mechanical fault diagnostics. Many features are extracted from these signals. The problem is how to find the good features among the feature set in order to receive reliable classifications. Based on specific distance criteria, a genetic algorithm (GA) is introduced to reduce the number of features by selecting optimized ones for fault classification purpose. A decision tree and multi-class support vector machine are used to illustrate the potentiality and efficiency of this selection method. Comparisons show that the diagnostic systems after selecting specific features perform better than the original system.

*Keywords:* Mechanical fault detection; Induction motor; Genetic algorithm; Support vector machine; Decision tree

---

### 1. Introduction

Induction motors more and more play an important part in modern industrial processes. What is needed is predictive condition monitoring and fault diagnosis to prevent motor breakage by warning against future damages. Mechanical faults are some of the most common faults of induction motors. In this paper, three types of mechanical faults are considered as bearing looseness, bearing damage, and rotor unbalance. In order to guarantee good classification results, input data require special preprocessing. Of the many methods suggested for data preparation, two common ones are feature selection and feature extraction. Principal component analysis (PCA) and independent component analysis (ICA) are two popular feature extraction techniques that are used to decrease the data dimension by extracting as much information as possible from the given data set [1, 2]. These methods change the original data in order to form a new one.

Meanwhile the feature selection simply selects a subset of existing features without any changes. There are some feature selection methods such as decision tree based method [3], GA in [6, 7], etc.

Recently, many methods have been developed to detect motor failures, such as using decision tree [2-5], support vector machine [1, 3, 8, 9], an artificial neural network [9-11], adaptive network-based fuzzy inference system [12], or fuzzy logic in [13]. The use of decision tree and support vector machine (SVM) in these works has provided good performance in motor fault classification.

In this paper, a feature selection technique is proposed to process data which is applied to decision tree and support vector machine classifiers. The time features of vibration signals are extracted in order to collect information about the motor status. There are 18 features obtained from three-dimensional vibration data. For preprocessing, the most important features are kept by a feature selection method. In order to remove noise, irrelevant or redundant information and also to reduce the number of features, a GA is proposed to select optimal features. The GA chooses a

---

\*Corresponding author. Tel.: +82 52 259 2187, Fax.: +82 52 259 1686  
E-mail address: hhlee@mail.ulsan.ac.kr  
DOI 10.1007/s12206-007-1036-3

part of the given data based on the distance criteria technique. It chooses only the most appropriate features for classifying which have the smallest within-class distance and the biggest between-class distance. The C4.5 algorithm [14] and the radial basis function kernel SVM are applied to new data and the results are shown in the experiment section.

## 2. Decision tree and support vector machine

### 2.1 Decision tree

A decision tree is a diagnostic tool that builds a knowledge-based system by inductive inference from case histories. A decision tree contains leaves, which represent classifications and branches, which represent conjunctions of features that lead to the classifications. The structure of a decision tree highly depends on how a test is selected as the root of the tree. The criterion for selecting the root of the tree is Quinlan’s information theory (information gain) [14]. According to this criterion, the information conveyed by a message depends on its probability. The construction of 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  $\{C_1, C_2, \dots, C_k\}$ . Suppose there is a possible test with  $n$  outcomes that partitions the training set  $T$  into subsets denoted by  $\{T_1, T_2, \dots, T_n\}$ . Let  $S$  be any set of cases,  $\text{freq}(C_i, S)$  be the number of cases in  $S$  that belong to class  $C_i$ , and  $|S|$  be the number of cases in set  $S$ . If one case is selected at random from set  $S$  and it belongs to class  $C_j$ , the message has the probability

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

and the information it conveys is

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

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

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

When it is applied to the set of training cases,  $\text{info}(T)$  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  $X$ :

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

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 in order to maximize this information gain.

### 2.2 Support vector machine

Support vector machine (SVM) is a supervised learning method for classification and regression. For illustration, two-class classification is used as an example. SVMs operate by finding a hyper-plane in the space of all possible inputs. This hyper-plane will try to split the positive samples from the negative samples. The separation will be chosen to have the largest distance from the hyper-plane to the nearest of the positive and negative samples. With a given data set  $\{(x_i, y_i), i = 1, 2, \dots, n\}$  where  $x_i \in R^n$  and  $y_i$  is either 1 or -1, the dividing hyper-plane can be expressed as

$$\omega \cdot x + b = 0 \tag{6}$$

where  $\omega$  is the weight vector, which is perpendicular to the separating hyper-plane. The bias  $b$  allows the separation to increase. In the case of linearly separable vectors, the data points will be classified by

$$\omega \cdot x + b \geq 1 \text{ for all } y_i = 1 \tag{7}$$

$$\omega \cdot x + b \leq -1 \text{ for } y_i = -1 \tag{8}$$

In order to maximize the separation  $2/|\omega|$  between two classes, a quadratic programming optimization must be solved:

$$\begin{aligned} &\text{Minimize } (1/2) \|\omega\|^2 \\ &\text{Subject to } y_i(\omega \cdot x + b) \geq 1, \quad 1 \leq i \leq n. \end{aligned} \tag{9}$$

This optimization (9) is solved to get the solutions  $\omega^*$  and  $b^*$  and the discrimination function is obtained:

$$f(x) = \text{sgn}(\omega^* \cdot x + b^*) \tag{10}$$

In the case that the linear boundary in the input spaces is not enough to separate into two classes

properly, nonlinear classification is suggested. SVM uses kernel function to map the data from input space to feature space, while it applies the kernel function to maximize the margin of the hyper-planes. The basis form of the discrimination function becomes

$$f(x) = \text{sgn} \left( \sum_{i=1}^n y_i \alpha_i k(x_i, x_j) + b \right) \quad (11)$$

There are some common kernel functions such as

Polynomial kernel:  $k(x_i, x_j) = (1 + x_i \cdot x_j)^d$  (12)

Radial basis function kernel:  
 $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$  (13)

Gaussian radial basis function kernel:  
 $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (2\sigma^2))$  (14)

Sigmoid kernel:  $k(x_i, x_j) = \tanh(kx_i \cdot x_j - c)$  (15)

More detail about the SVM can be found in [16, 17]

### 3. Feature extraction and selection

#### 3.1 Feature extraction

In this paper, in order to diagnose the motor conditions, specific time domain features are extracted including the root mean square (rms), variance, skewness, kurtosis, crest factor and maximum value. These features represent the energy, vibration amplitude, and the time series distribution of the signal in the time domain.

Skewness is a measure of symmetry, or the lack of symmetry in a signal.

$$\text{skewness} = \frac{\sum_{n=1}^N (x(n) - \text{mean}(x))^3}{(N-1)\sigma^3} \quad (16)$$

$$\sigma = \sqrt{\frac{\sum_{n=1}^N (x(n) - \text{mean}(x))^2}{(N-1)}} \quad (17)$$

Kurtosis is a measure of whether the data has peaked or is flat relative to a normal distribution.

$$\text{kurtosis} = \frac{\sum_{n=1}^N (x(n) - \text{mean}(x))^4}{(N-1)\sigma^4} \quad (18)$$

The crest factor is a measure of how much impact-

ing is occurring in the time waveform. Impact in the time waveform may indicate rolling element wear or cavitations.

$$\text{crest} = \frac{\max |x(n)|}{\text{rms}} \quad (19)$$

$$\text{rms} = \sqrt{\frac{\sum_{n=1}^N (x(n))^2}{N}} \quad (20)$$

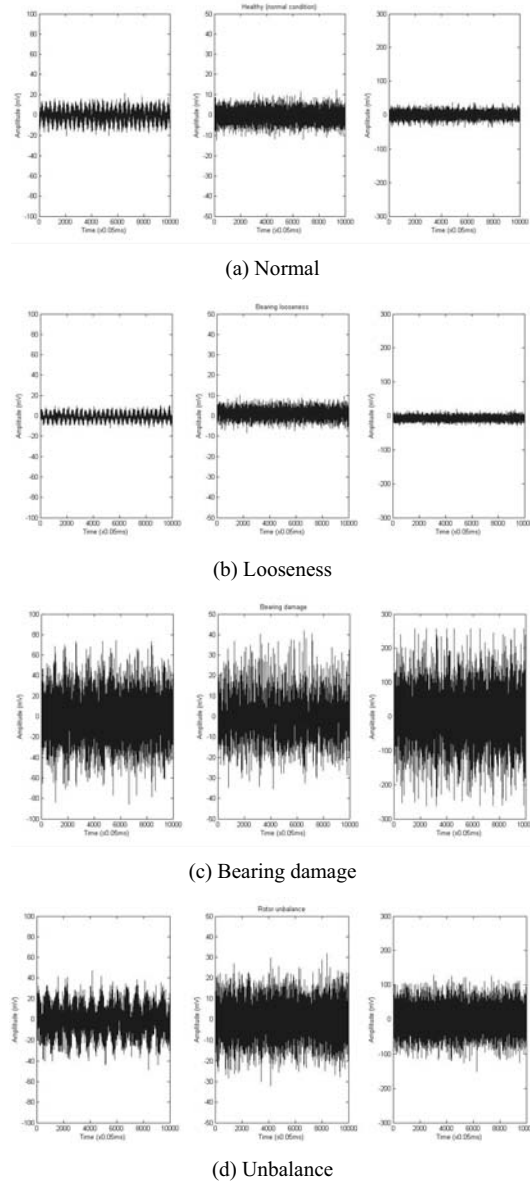


Fig. 1. Time signal waveforms in 3 dimensions (horizontal-axial-vertical): (a) Normal condition, (b) Bearing looseness, (c) Bearing damage, (d) Rotor unbalance.

Variance ( $\sigma^2$ ) is a measure of the dispersion of a waveform about its mean and is called the second moment of the signal. Finally, maximum amplitude and rms of a signal indicate the severity of the defects.

Three accelerometers were installed in order to measure the tri-axial vibration signal, where six features are extracted from each signal. The feature set has a total of 18 features and is used to train the decision tree. For normalization, each feature was divided by the corresponding absolute maximum value. All features will be normalized with a range of 0.0 to 1.0, except for the skewness with a range of -1.0 to 1.0.

The features are formed as follows: {rms (a), variance (a), skewness (a), kurtosis (a), crest factor (a), maximum (a), rms (h), variance (h), skewness (h), kurtosis (h), crest factor (h), maximum (h), rms (v), variance (v), skewness (v), kurtosis (v), crest factor (v), maximum (v)}

Fig. 1 shows a few time signal waveforms of four motor conditions in all three dimensions collected by the accelerometers. There are four motor states in which data are collected including the bearing damage, bearing looseness, rotor unbalance, and normal condition.

### 3.2 Genetic algorithm

The GA provides 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. Solutions from one population are reproduced to create a new generation in the population. Mutations occur randomly in each population. First, the GA generates an initial population of random chromosomes. After that, the fitness of each chromosome is calculated. The next generation is produced by selecting pairs of parents. During reproduction, crossover of the genes and random mutation occur in some of the children. Reproduction continues until the best solution is created. In the GA process, randomness plays a central role in which many random procedures are called during execution of the GA.

In this paper, the multi-objectives GA is used for feature selection. Eighteen features are extracted from three axes of vibration time signals. Each signal provides six features including the root mean square (rms), variance, skewness, kurtosis, crest factor, and maximum value.

The GA chromosome represents a sequence of features, in which the binary code includes 18 bits that represent the selection of features, where “1” represents a selected feature and “0” represents an abandoned feature.

The objectives of the GA are a smaller within-class distance and a larger between-class distance. The within-class distance is given by

$$J_c = \sum_{i=1}^c p_i J_i. \tag{21}$$

$$J_i = (1/n_i) \sum_{k=1}^n (x_k^i - m_i)^T (x_k^i - m_i). \tag{22}$$

Class  $i = 1, \dots, c$ ;  $m_i$  is the mean vector of class  $i$ ;  $n_i$  is the number of samples in class  $i$ ;  $p_i$  is the number of samples in class  $i$ .

The between-class distance is as follows

$$J_b = \sum_{i=1}^c p_i (m_i - m)^T (m_i - m). \tag{23}$$

$m$  is the mean vector of all of the classes.

The GA is processed with two objectives. The first objective is to determine a minimum in the within-class distance, and the second is to determine a maximum in the average between-class distance. For this purpose, the fitness function is defined as

$$J = J_i + (1/J_b). \tag{24}$$

The chromosome that minimizes the fitness function is chosen, and therefore optimal features can be selected. After the features are chosen, the C4.5 algorithm and the SVM are applied to classify the condition of the motor.

### 4. Experimental results

Parameters of the feature selection GA are set as follows: the population is 96, the length of chromosome code is 18, 1 crossing point, and number of generation is 500. Selection result shows the best combination of features satisfying the given objective, e.g., 100001000110100100 means that features F1, F6, F10, F11, F13 and F16 were selected, remaining were abandoned. The selected feature subsets are used to train the decision tree and the support vector machine. The C.45 algorithm is used for the decision tree and the radial basis function is chosen as the ker-

nel function of the support vector machine system. There are only two SVM parameters needed:  $C$  and  $\gamma$  [17],  $C = 1.2$  and  $\gamma = 1.7917$  are chosen in this work.

The same training set with 559 samples and 307 samples for a test set is used to evaluate the trees and the SVM classifications. Though the sizes of decision trees are almost similar, after selection the overall accuracy is a bit better.

Tables 1 and 2 show the diagnosis results that were evaluated on the training and test data of the decision tree and the SVM classifiers. A decision tree without

feature selection has 95.1% accuracy when evaluated on the test data, while the trees with feature selection have accuracy up to 98.7%–99.7%. Meanwhile, the SVM classifier provides 94.14% accuracy when built without feature selection. The accuracy increases to 94.79%–99.67% with feature selections. Generally, after features are selected, the diagnostic performance is better for both the decision trees and the SVM classifiers.

For SVM in Table 1, the accuracy of an unbalanced state in case of 4-feature data is rather low compared to others. This may be caused by the lack of appropriate features in this data set that can help to discriminate this fault from the others. Feature selection can help to obtain smaller data without changing the original data. But discarding some features will not guarantee that all useful data can be kept. This is a drawback of feature selection technique.

From 18 features at the beginning, the selection algorithm forms new feature subsets which have 2 to 6 features in different trials. In order to choose the best one of these feature subsets, a small number of features and high classification accuracy criteria are considered. As seen in Tables 1 and 2, the 2-feature subset is the one that somehow satisfies above two criteria. The performance of these selected features showed that they improved the classification accuracy in both decision tree and support vector machine classifiers.

In order to confirm the efficiency of the proposed selection, Table 3 compares the classification accuracies of SVM for the selected 2-feature subset and a random 2-feature subset. In case of the feature selected randomly, the experiments under the same condition are repeated ten times to get the average results. Table 3 shows that the proposed selection provides a better performance compared to the random selection.

Table 1. SVM classifier performances before and after feature selections.

SVM and selected features	On test data (%)				Average evaluation on training data	Average evaluation on test data
	Normal	Looseness	Bearing damage	Unbalance		
All 18 features (F1–F18)	92.86	98.48	95.6	90.43	99.82%	94.14%
F6, F13	92.86	100	100	98.93	94.63%	98.37%
F6, F13, F18	85.71	100	100	98.93	94.99%	97.07%
F6, F10, F12, F13	96.43	100	96.7	88.3	96.96%	94.79%
F1, F2, F6, F13, F14	100	100	100	98.93	94.99%	99.67%
F1, F6, F10, F11, F13, F16	100	100	100	98.93	96.06%	99.67%

Table 2. Decision tree (DT) performances before and after feature selections.

Decision tree and selected features	Size	On test data (%)				Average evaluation on training data	Average evaluation on test data
		Normal	Looseness	Bearing damage	Unbalance		
All 18 features (F1–F18)	13	100	100	90.11	93.62	99.8%	95.1%
F6, F13	15	100	100	100	98.93	98.4%	99.7%
F6, F13, F18	15	100	100	100	98.93	98.4%	99.7%
F6, F10, F12, F13	21	100	100	100	95.74	99.3%	98.7%
F1, F2, F6, F13, F14	15	100	100	100	95.74	98.9%	98.7%
F1, F6, F10, F11, F13, F16	15	100	100	100	95.74	98.6%	98.7%

Table 3. Performance comparison for SVM classifier (under the same condition).

Feature	On test data (%)				Average evaluation on test data
	Normal	Looseness	Bearing damage	Unbalance	
SVM (F6, F13)	92.86	100	100	98.93	98.37%
SVM (2 features selected randomly)	65.89	87.27	90	82.34	82.67%

## 5. Conclusions

Optimization of the feature selection by using the distance-based feature selection method can improve the classification performance of induction motor mechanical fault diagnosis. The selection results which were applied for the decision tree and the SVM performed better compared to classification systems without feature selection. The experiment also shows that the decision tree and the SVM with optimal feature selection are promising classifiers for induction motor fault diagnosis.

Along with the SVM, the decision tree with optimal selection features has higher accuracy and performs better by removing the redundant and irrelevant information. This proves that the distance-based selection method using GA is an efficient way to improve diagnostic performance.

## Acknowledgment

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).

## Nomenclature

(a)	: Axial
(h)	: Horizontal
(v)	: Vertical
DT	: Decision tree
Fi	: $i^{\text{th}}$ feature
GA	: Genetic algorithm
ICA	: Independent component analysis
PCA	: Principal component analysis
Rms	: Root mean square
SVM	: Support vector machine

## References

- [1] A. Widodo, B. S. Yang and T. Han, Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors, *Expert Systems with Applications* 32 (2007) 299-312.
- [2] W. Sun, J. Chen and J. Li, Decision tree and PCA-based fault diagnosis of rotating machinery, *Mechanical Systems and Signal Processing*, (2006).
- [3] V. Sugumaran, V. Muralidharan and 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 (2007) 930-942.
- [4] B. S. Yang, C. H. Park and H. J. Kim, An efficient method of vibration diagnostics for rotating machinery using a decision tree, *International Journal of Rotating Machinery* 6 (1) (2000) 19-27.
- [5] D. S. Lim, B. S. Yang and D. J. Kim, An expert system for vibration diagnosis of rotating machinery using decision tree, *International Journal of COMADEM* (2000) 31-36.
- [6] C. L. Huang and C. J. Wang, A GA-based feature selection and parameters optimization for support vector machines, *Expert Systems with Applications* 31 (2006) 231-240.
- [7] H. Gu, Z. Gao and F. Wu, Selection of optimal features for iris recognition, *International Symposium on Neural Networks, China* . (2005) 81-86.
- [8] Y. Yang, D. Yu and J. Cheng, A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM, *Measurement* (2006) doi: 10.1016/j.measurement.2006.10.010.
- [9] L. B. Jack and A. K. Nandi, Fault detection using support vector machines and artificial neural networks, Augmented by Genetic Algorithms, *Mechanical Systems and Signal Processing* 16 (2002) 373-390.
- [10] B. Samanta and K. R. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, *Mechanical Systems and Signal Processing* 17 (2003) 317-328.
- [11] B. Samanta, K. R. Al-Balushi and S. A. Al-Araimi, Artificial neural networks and genetic algorithm for bearing fault detection, *Soft Comput.* (2006) 264-271.
- [12] Y. Lei, Z. He, Y. Zi and Q. Hu, Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs, *Mechanical Systems and Signal Processing* (2007) doi:10.1016/j.ymsp.2006.11.003.
- [13] T. Lindh, J. Ahola, P. Spatenka and A. L. Rautiainen, Automatic bearing fault classification combining statistical classification and fuzzy logic, *NORPIE*, (2004)
- [14] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publisher, California, USA, (1993).
- [15] J. S. Rao, *Vibratory Condition Monitoring of Ma-*

- chines, Alpha Science International Ltd., (2000) 361-382.
- [16] C. C. Chang and C. J. Lin, LIBSVM: a library for support vector machines, Software available at <http://www.csie.ntu.edu.tw/~cjlin.libsvm>, (2001).
- [17] C. W. Hsu, C. C. Chang and C. J. Lin, A practical guide to support vector classification. Available at [www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf](http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf).
- [18] The C4.5 code comes from the Internet (<http://rulequest.com/personal/c4.5r8.tar.gz>).