Reliable Condition Monitoring of an Induction Motor using a Genetic Algorithm based Method

Won-Chul Jang, Myeongsu Kang, Jaeyoung Kim, Jong-Myon Kim*
School of Electrical Engineering
University of Ulsan
Ulsan, South Korea
{nasha0339, ilmareboy, kjy7079, jongmyon.kim}@gmail.com

Hung Ngoc Nguyen
Control Engineering Dept.
Le Quy Don University
Hanoi, Vietnam
ngochung212@gmail.com

***Abstract*—Condition monitoring is a vital task in the maintenance of industry machines. This paper proposes a reliable condition monitoring method using a genetic algorithm (GA) which selects the most discriminate features by taking a transformation matrix. Experimental results show that the features selected by the GA outperforms original and randomly selected features using the same *k*-nearest neighbor (k-NN) classifier in terms of convergence rate, the number of features, and classification accuracy. The GA-based feature selection method improves the classification accuracy from 3% to 100% and from 30% to 100% over the original and randomly selected features, respectively.**

Keywords—feature selection; genetic algorithm; transform matrix, k-nearest neighbor classifier

I. INTRODUCTION

Pattern recognition is the scientific discipline whose goal is to classify objects into a number of categories of classes based on the set of feature data. Computer-aided diagnosis is an important application of pattern recognition [1], aiming at assisting engineers and operators in making a decision of machine condition.

For designing a classification system to determine machine's condition, an optimal feature selection for each machine condition is an important task, which highly affects the classification performance of the system [1]. In fact, it is difficult to determine an optimal set of features because sensor noises are usually included into the original signal. Many existing algorithms including using Fisher discriminate analysis (FDA), principal component analysis (PCA), locally linear embedding (LLE) and its variant as supervised locally linear embedding (SLLE) tried to extract the useful features by analyzing each feature component and the relationship between them [2-6]. However, the efficiency of these methods on the classification accuracy is not guaranteed due to the lack of the classifier information [7], [8].

To address this issue, this paper utilizes a genetic algorithm (GA) [9-11] incorporating feature selection and optimization as a part of the learning process to select a set of useful features. In experiments, original features are generated from vibration signals of an induction motor. Then, GA incorporated with a classifier is applied to generate a transform matrix. This matrix is used to generate optimal features from the original ones. In order to validate the effectiveness of the optimal feature space, a *k*-nearest neighbors (k-NN) classifier is utilized as a learning process to compute the classification accuracy of the proposed method.

This paper is organized as follows. Section 2 discusses the proposed GA-based feature selection method, and Section 3 analyzes the experimental results. Finally, Section 4 concludes this study.

II. PROPOSED FEATURE SELECTION METHOD

Classification is an important task of condition monitoring for machinery in non-stationary operations. The classification accuracy of the condition monitoring system highly depends on the section of significant feature vectors. For example, if feature vectors overlap in many dimensions, a classifier cannot classify each case well, degrading classification performance of the system. To select the most important features for each case, redundant features (or dimensions) should be discarded to prevent misclassification.

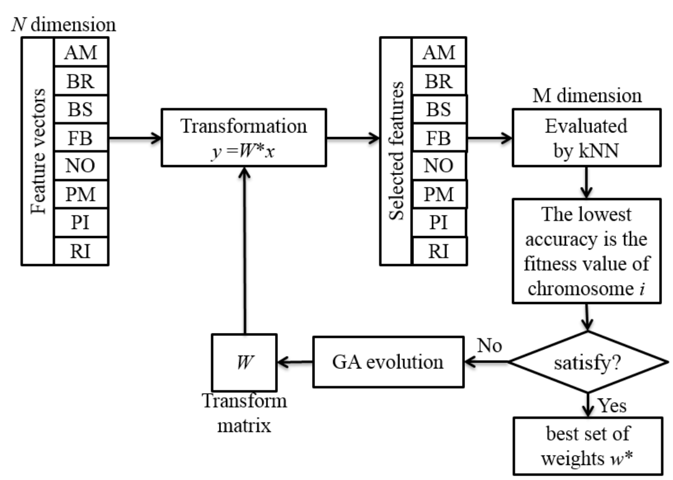


Fig. 1. The proposed GA-based feature selection scheme.

To select an optimal feature vector for each fault of an induction motor, this paper presents a GA-based feature

*Corresponding author.

978-1-4799-4509-2/14/$31.00 ©2014 IEEE

selection method, as shown in Fig. 1. In the training stage, the GA is applied on generations to find the best solution such as an optimal transform matrix that helps to select the best features for each case of an induction motor. The followings detail each step in the proposed feature selection method.

A. Input Data

In this study, we acquired seven types of faulty data along with normal data: angular misalignment (AM) faults, broken rotor bar (BR) faults, bowed rotor shaft (BS) faults, fault bearing (FB), parallel misalignment (PM) faults, phase imbalance (PI) faults, and rotor imbalance (RI) faults. The acquired vibration signals were sampled at 8 kHz, and we used 105 one-second-long vibration signals to represent each fault symptom.

B. Transform Matrix

A transform matrix is used to select significant features from the original feature space (*N* dimensions) for improving classification accuracy. For example, if the new feature space has *M* dimensions (*M*<<*N*), the transform matrix consists of *M* rows and *N* columns, as shown in Fig. 2. In each row, only one element is '1', and the others are '0's. The column index of '1'element is represented as an essential feature in the new feature space. By multiplying the transform matrix with the original feature vector, essential dimensions are generated to form a new feature vector.

$$\overline{y} = W * \overline{x} \tag{1}$$

where $W = [w(i,j)]$, $w(i,j)$ is '0' or '1', i is from 1 to M, and j from 1 to N.

Example : three essential dimensions at i, j, k

Location	1	2	…	*i*	…	*j*	…	*k*	…	*N*
Value	…	…	…	…	…	…	…	…	…	…

Corresponding mask for dimensions i :

Location	1	2	…	*i*	…	*j*	…	*k*	…	*N*
Value	…	…	…	1	…	…	…	…	…	…

Corresponding mask for dimensions j :

Location	1	2	…	*i*	…	*j*	…	*k*	…	*N*
Value	…	…	…	0	…	1	…	…	…	…

Corresponding mask for dimensions k :

Location	1	2	…	*i*	…	*j*	…	*k*	…	*N*
Value	…	…	…	0	…	0	…	1	…	…

Transform matrix has size 3xN :

Location	1	2	…	*i*	…	*j*	…	*k*	…	*N*
1	0	0	…	1	…	0	…	0	…	0
2	0	0	…	0	…	1	…	0	…	0
3	0	0	…	0	…	0	…	1	…	0

Fig. 2. A transform matrix with three essential dimensions.

This approach takes a long time to find suitable dimensions. For example, the approach needs to search 2^31 transform matrices to find the best one with 31 dimensions. In addition, the approach requires to search 2^102 transform matrices to find the best one with 102 dimensions. To overcome this problem, a genetic algorithm (GA) is utilized in this study.

C. Genetic Algorithm

A genetic algorithm consists of four main steps: encoding, parent selection, crossover/mutation, and replacement, where a fitness function is defined in advance. The fitness function is a vital component that calculates the performance of each chromosome [9,10,11]. In this study, the initial generation consists of 500 chromosomes (500 transform matrices). For each solution, its performance is evaluated by a *k*-NN classifier. If a solution gives the highest classification accuracy, it is the best solution in this generation.

1) Fitness function

A fitness function reacts between selected features with a classifier in this study. The accuracy of the *k*-NN classifier per each transform matrix represents as a fitness value of the solution such as

$$Fitness = J(w) \mid T = \gamma \frac{Totpats - CorrectPaths}{Totpats} + \delta \frac{n\min/K}{Totpats}, \tag{2}$$

where *Totpats* is the number of samples to be examined and *CorrectPats* is the number of samples which is correctly classified. *nmin* is the number of the nearest neighbors which are not used in the subsequent determination of class, and *K* is the number of the nearest neighbors. The constants γ and δ are used to tune the processing of the algorithm, and they are 2 and 0.2, respectively, in this study.

Three stop-criteria are defined in this study: (1) if the minimum fitness value of each generation is higher than the expected value (threshold); (2) if the average fitness value of generation is higher than the expected value, and (3) if the GA loop ends after the given number of iterations. The first case is harder to converge than the others.

2) Encoding

Feature vector		**Decision**	**Encoding values**
1	…	Redundancy	0
…	…	…	…
I	…	Essence	1
…	…	…	…
J	…	Essence	1
…	…	…	…
N	…	Redundancy	0

Location | Feature values | Chromosome

Fig. 3. Encoding chromosome bit-string.

A single bit is encoded for each component of the feature vector, where binary encoding is addressed to this problem. Each chromosome represents each transform matrix. As aforementioned, we encode the chromosome bit-string as follows: (1) the length of chromosome is equal to the number of dimensions of the original feature vector (*N*); (2) if *i*th dimension of the original vector is kept, *i*th element of the chromosome is set to 1. Otherwise, set to 0. The number of 1's elements of chromosome is M, and the position of 1's elements in chromosome specifies which dimension is preserved, as shown in Fig. 3.

3) Parent Selection

Many parent selection methods have been proposed and they are closed to the natural selection. In this study, a Roulette-wheel selection method is applied because it gives a reasonable convergence and divergence ratio.

4) Crossover and Mutation

Crossover and mutation with a probability value are applied to expand the search range of the genetic algorithm over generations. The probability value is a controlling parameter to keep the diversity of solutions. If the value is high, the searching process is likely to diverge. When the probability is low, a local optimal solution is also easily found.

5) Replacement policy

In this study, the worst-case replacement is used as a replacement policy. The fitness value of new offsprings is calculated and compared to those of the worst solution in the current population. If offspring's fitness is higher than the worst solution, it is replaced by the offspring. Otherwise, the offspring is omitted. In this study, a steady-state GA is utilized.

Fig. 4 illustrates a flowchart of the proposed GA-based approach.

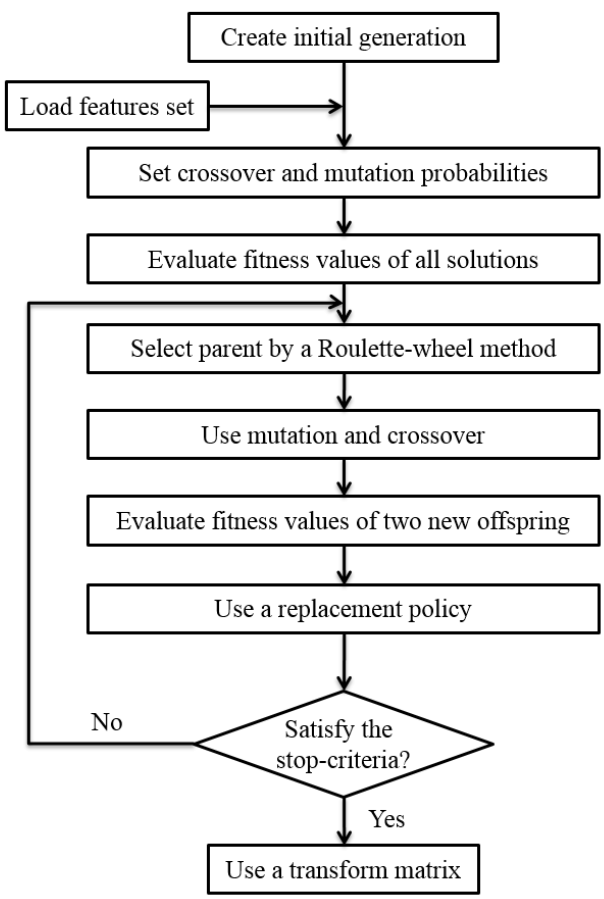


Fig. 4. Flowchart of the proposed GA-based approach.

III. EXPERIMENTAL RESULTS

In this study, we utilizes a GA to select optimal features (or to remove the redundant features) for improving classification accuracy of a condition monitoring system. To evaluate the performance of the proposed GA-based approach, we use two feature sets: one includes 31 dimensions and the other includes 102 dimensions using two different feature generation methods [7], [8], as shown in Table 1, where each class has 105 samples. In each set, 15 features of each class were randomly selected to generate labeled data for the *k*-Nearest Neighborhood (*k*-NN) classifier and the others are unlabeled data.

TABLE 1. Selected two different feature sets

	Set 1	*Set 2*
Number of subsets	8	8
Number of samples	105	105
Number of dimensions	31	102

In addition, parameters of the GA are different for these datasets because of their properties, as shown in Table 2. The quality of feature set 1 is much better than that of the set 2 even if the crossover and mutation probabilities are low, as shown in Fig. 5. This result demonstrates that the set 1 already includes distinctive features. On the other hand, the set 2 takes a long time to find an optimal solution because there exists a lot of redundant features.

TABLE 2. Parameters of GA for two different feature sets

Parameters of GA	***Set 1***	***Set 2***
Number of chromosomes	50	50
Maximum iterations	500	500
Parent selection	Roulette wheel	Roulette wheel
Crossover method	Multi-points	Multi-points
Mutation method	One point	One point
Probability of crossover and mutation	100%	8%
Replacement policy	Worst case	Worst case
Expected fitness value	100%	100%

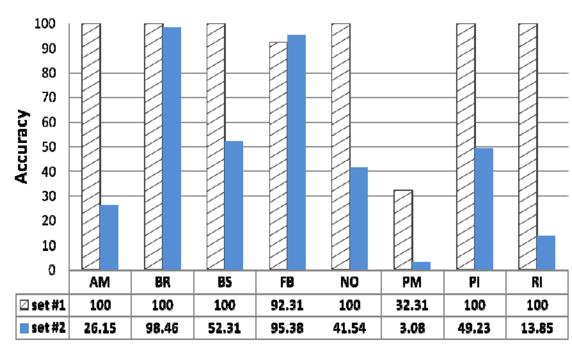
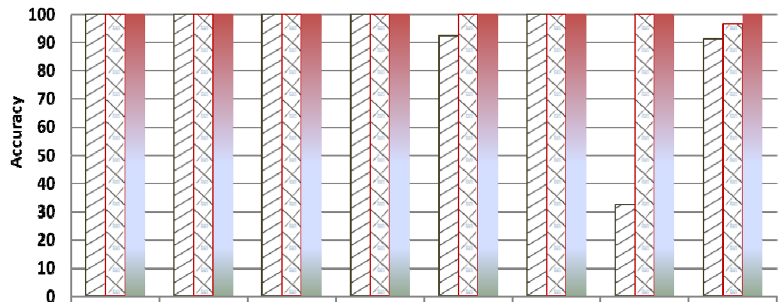
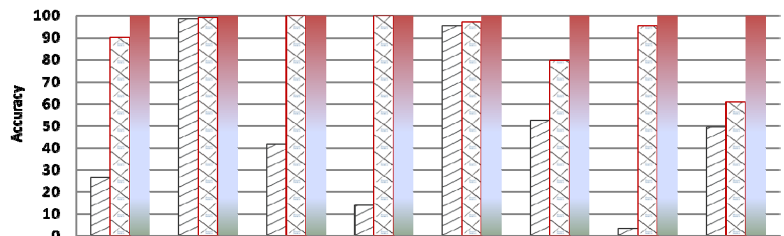


Fig. 5. Classification accuracy of the *k*-NN classifier using two different feature sets.



	AM	BR	NO	RI	FB	BS	PM	PI
Original(31 dims)	100	100	100	100	92.31	100	32.31	91.2
Random(16 dims)	100	100	100	100	100	100	100	96.4778
Proposed(14 dims)	100	100	100	100	100	100	100	100

Fig. 6. Performance of original, randomly selected, and the GA-based feature selection using the same the *k*-NN classifier for *set* 1.



	AM	BR	NO	RI	FB	BS	PM	PI
Original (102 dims)	26.15	98.46	41.54	13.85	95.38	52.31	3.08	49.23
Random(51 dims)	89.8889	99	99.778	99.7133	97	79.5778	95.3911	60.5689
Proposed(38 dims)	100	100	100	100	100	100	100	100

Fig. 7. Performance of original, randomly selected, and the GA-based feature selection using the same the *k*-NN classifier for *set* 2.

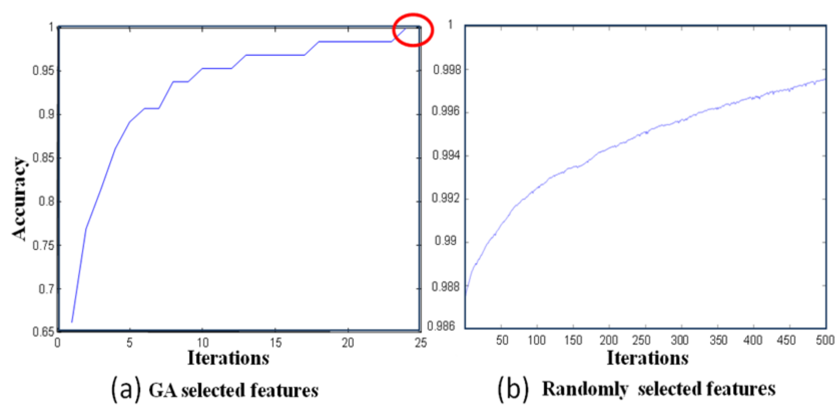


Fig. 8. Classification accuracy of *set 1*.

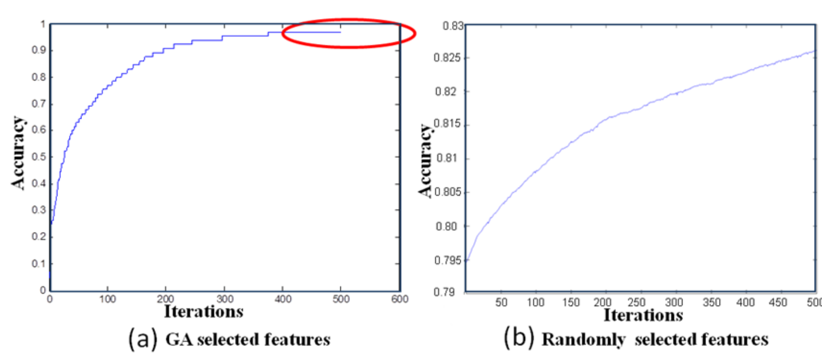


Fig. 9. Classification accuracy of *set 2*.

The GA-based feature selection method improves the classification accuracy. After applying GA, the number of dimensions in the set 1 is reduced from 31 to 14 (56%) after 23 iterations, and the number of dimensions in the set 2 is reduced from 102 to 38 (62%) after 500 iterations. In addition, features selected by the GA are more effective than original and randomly selected features in the classification accuracy, as shown in Fig. 6 and Fig. 7, respectively.

Figs. 8 and 9 present additional data showing the classification accuracy (or fitness value) using the features selected by the GA and randomly search methods over iterations for set 1 and set 2, respectively. Due to the hierarchical information, an overall trend of fitness values using the GA is smoother and higher than those using the randomly search method, where the blind search is used in all the generations. In addition, the convergence speed using the GA method highly depends on the characteristics of the feature set. In general, more dimensions of the feature vector are, lower convergence speed is. For example, GA requires 500 loop iterations for the 102 dimensions (set 2), while it requires only 24 iterations for set 1 which has 31 dimensions of the feature vector.

IV. Conclusion

This paper proposed a GA-based feature selection method for reliable condition monitoring of an induction motor. The proposed GA-based method can help to select an optimal feature vector from all the features of each case. Experimental results using eight different vibration signals indicated that the GA-based feature selection method reduces the number of dimensions in the set 1 data from 31 to 14 (56%) and the number of dimensions in the set 2 data from 102 to 38 (62%) while improving the classification accuracy from 3% to 100% in the set 1 and from 30% to 100% in the set 2 over the randomly search method using the same *k*-NN as a classifier.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1A2043644), and this research was also supported by the Leading Industry Development for Economic Region (LeadER) grant funded the MOTIE (Ministry of Trade, Industry and Energy), Korea in 2014 (No. R0001220).

References

[1] Sergios Theodoridisi, Konstantinos Koutroumbas "Pattern Recognition", 2nd edition, Elsevier Academic Press, 2003.

[2] B.–S. Yang, T. Han, and W.W. Hwang, "Fault Diagnosis of Rotating Machinery based on Multi-Class Support Vector Machines," *Journal of Mechanical Science and Technology*, Vol. 19, No. 3, pp. 845–858, 2005.

[3] A. Widodo, B.–S. Yang, and T. Han, "Combination of Independent Component Analysis and Support Vector Machine for Intelligent Faults Diagnosis of Induction Motors," Expert System with Application, Vol. 32, pp. 299-312, 2007.

[4] S. Abbasion, A. Rafsanjani, A. Farshidianfar, N. Irani, "Rolling Element Bearing Multi-Fault Classification Based on Wavelet Denosing and Support Vector Machine," Mechanical Systems and Signal Processing, Vol. 21, pp. 2933-2945, 2007.

[5] A. Widodo, B. –S. Yang, T. Han, and D. J. Kim, "Fault Diagnosis of Induction Motor using Independent Component Analysis and Multi-Class Support Vector Machine," Proceedings of the 11th Asia-Pacific Vibration Conference, pp. 144–149, 2005.

[6] T. Han, B.–S. Yang, and Z.–J. Yin, "Feature-based Fault Diagnosis System of Induction Motors using Vibration Signal," Journal of Quality in Maintenance Engineering, Vol. 13, No. 2, pp. 163-175, 2007.

[7] Hung Nguyen, Myeongsu Kang, Jongmyon Kim, "An Effective Feature Extraction method for Fault Diagnosis of Induction Motor," Korean Society of Computer and Information, (accepted)

[8] Yaguo Lei, Zhengjia he, Yanyang Zi, "Application of an intelligent classification method to mechanical fault diagnosis", Journal of Expert System with Applications, 2009.

[9] David Beasley , David R. Bull , Ralph R. Martin, "An Overview of Genetic Algorithms: Part 1, Fundamentals," 15(2), University Computing, 1993, pp. 58-69.

[10] David Beasley , David R. Bull , Ralph R. Martin, "An Overview of Genetic Algorithms: Part 2, Research Topics", 15(4), University Computing, 1993, pp. 170-181.

[11] Darrell Whitley, "A Genetic Algorithm Tutorial", Technical Report, Colorado Advanced Software Institute, 1995.