A GA-based Feature Optimization Technique for Bearing Fault Diagnostics

Jie Liu[†], **Wilson Wang**[††], **Farid Golnaraghi**[†]

[†]Dept. of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1
[††]Dept. of Mechanical Engineering, Lakehead University, Thunder Bay, ON, Canada P7B 5E1

Summary
Rolling-element bearings are widely used in various mechanical and electrical systems. A reliable online bearing fault diagnostic technique is critically needed in industries to detect the occurrence of a fault so as to prevent system's performance degradation and malfunction. To improve the fault diagnostic reliability and efficiency, a genetic algorithm based feature optimization technique is proposed in this work. In this scheme, the discrete wavelet packet analysis is utilized to decompose the raw vibration signal into several constituent signatures, from which the bearing health condition related features are formulated. Taking these features as a fundamental search space, the genetic algorithm based technique is adopted to choose the representative features that carry more discriminatory information for bearing health condition assessment. This optimization process is guided by a suggested fitness function. A neural fuzzy system is utilized for diagnostic classification operations. The performance of the proposed technique is evaluated by experimental tests.
Key words:
Genetic algorithm, Feature optimization, Fault classification, Rolling element bearing.

1. Introduction

Rolling element bearings are widely used in various types of mechanical and electronic systems. Accordingly, a reliable bearing fault detection technique is critically needed to prevent system's performance degradation, malfunction or even catastrophic failures. Several methods have been proposed in the literature for bearing fault detection and fault type classification. Based on the properties of information carriers, these techniques can be classified into vibration monitoring [1]-[4], acoustic signal processing [5], lubricant analysis [6], temperature measurement [7], and electric current analysis [8]. Among them, vibration monitoring is the most commonly used approach in practice due to the ease of measurement and analysis. A detailed review of the bearing condition monitoring techniques based on vibration measurement can be found in [9]. In applications, most proposed techniques still rely on human interpretation to some extent. When a large number of features are present, however, several concerns have to be properly considered. Firstly, some features may provide confusing information to the diagnosis operations. Secondly, if a data-driven automatic classifier is employed, the number of data sets required for system training, as well as the computation efforts, may increase dramatically as the number of input features increases. Consequently, it is necessary to examine the whole set of candidate features and manually select the most representative ones based on expertise and/or certain criteria [10],[11]. This procedure, however, is usually time consuming, and can only be applied to some relatively simple classification applications. Therefore, a highly automatic feature optimization technique is highly demanded.

A rolling element bearing is not a simple mechanical component, but a relatively complex system, which consists of an inner ring, an outer ring, a cage, and a series of rolling elements. Whenever a fault happens on a bearing component, stationary and/or non-stationary impacts are generated, which excite the bearing and its support structures. In this study, a genetic algorithm (GA) based feature optimization technique is proposed to extract the representative features for bearing fault diagnostics. Based on a designated fitness function, the GA can eliminate human intervention and automatically formulate the optimal features. Instead of using a thorough and complex search space, the fundamental search base in this work is based on energy distribution ratios from several constituent signals that are processed by the use of discrete wavelet packet (DWP) analysis.

The paper is organized as follows: Section 2 describes the experimental setup used in this work and the related signal processing techniques to extract the representative features. The GA-based feature optimization technique is presented in Section 3, whereas its diagnostic reliability is validated based on experimental tests.

2. Feature Preparation

Manuscript received September 5, 2007
Manuscript revised September 20, 2007

2.1 Experimental Setup

The experimental setup used in this work is schematically shown in Fig. 1. The rotor system is supported by two cylindrical roller bearings (SKF NJ204 ECP) fitted in the solid housings. A 3-hp, three-phase induction motor is employed to drive a shaft via a time belt and a self-aligning coupling. The speed controller allowed the system to operate in the range from 600 to 1800 rpm. The radial load is provided by a static disk installed on the shaft and between the bearings. Vertical vibrations are measured by two piezoelectric accelerometers (Dytran 3035AG) that have a bandwidth up to 10 kHz, a sensitivity of 100 mV/g, and a full scale range of ± 50g. The collected vibration signals are properly amplified by charge amplifiers (Dytran 4105C), and then are stored to a data recorder through an anti-aliasing filter. The sampling frequency in this test is set at 6 KHz.

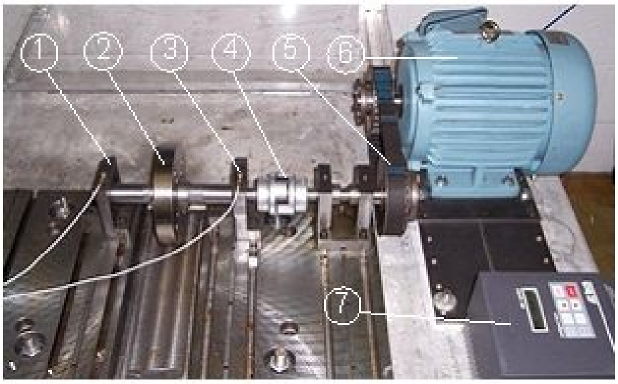

Fig. 1 Experimental setup: 1-test bearing housing; 2-balanced disk; 3-piezoelectric accelerometer; 4-coupling; 5-power transmission; 6-motor; 7-speed control.

Four types of bearing health condition cases are considered in this work: healthy bearings (HY), bearings with rolling element damages (ED), bearings with inner-race defects (IR), and bearings with outer-race defects (OR). These simulated defects are artificially introduced onto the corresponding bearing components. Fig. 2 shows three types of bearing faults in tests, whose dimensions range from 0.5 mm to 2.0 mm.

Fig. 2 Three types of bearing faults.

Each bearing is well lubricated before tests. Four bearing health condition cases have been tested. Under each condition, the bearing is driven over five different speeds, from 600 rpm to 1800 rpm at a step size of 300 rpm. At each speed, 50 segments of vibration signal are collected with the time interval of 5 minutes. Consequently, 1000 segments of raw signals are recorded for analysis. Fig. 3 shows some samples of vibration signals at shaft speed 1200 rpm.

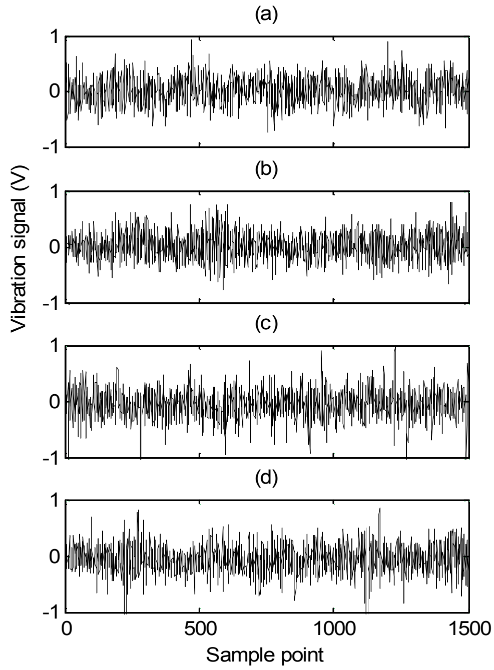

Fig. 3 Vibration signals: (a) from a healthy bearing; (b) from a rolling-element damaged bearing; (c) from an inner-race defective bearing; (d) from an outer-race defective bearing.

2.2 Feature Extraction

For a rolling element bearing, usually a defect occurs on the fixed ring race first because the fixed ring material is subject to more dynamic load cycles. Consider a general bearing with a fixed outer ring, and suppose that a defect (e.g., a fatigue pit) has happened on the outer ring race. Each time a rolling element rolls over the pit, an impulse is generated due to the impact. This impulse excites vibration resonance of the bearing and the surrounding structures. In theory, the excited transient modes due to an outer race defect do not vary because the defect angular position remains the same as each impact occurs, as long as no slippage occurs among the bearing components. On the other hand, for a rolling element fault or an inner race defect, the generated impulse transient modes may change in properties because the impact occurs at a different angular position as the bearing rotates. As a result, the magnitudes of the impulse transients and the excited resonance modes vary over time. Different type of defects induces transients within different frequency bands, which generates different energy distributions over distinct bandwidths. In this work, the DWP is employed to decompose the raw signal into several constituent signatures, each of which drops within a specific bandwidth. After the energy ratio of each constituent signature with respect to the raw signal is calculated, advanced investigation is taken to evaluate the potential for the resulting indices to carry valuable features related to bearing health conditions.

The DWP analysis is a generalized wavelet transformation. It applies multiple band filters to decompose a signal into a series of packets that contain the shifted and scaled versions of the mother wavelet, and thus it has the possibility to provide more information for signal analysis [12]. In this work, a three-level DWP decomposition is adopted. Consequently, eight wavelet packets are yielded. In each packet, the detail and approximation coefficients are utilized to reconstruct the original signal by reconstruction filters and up-sampling. The reconstructed signatures represent the original signal over different frequency bands. For example, using daubechies-12 wavelet as mother wavelet and Shannon entropy least cost function as the basis criterion, the original unbiased signal $S(i,k)$, generated by a bearing with an inner race defect as partly shown in Fig. 3, can be decomposed into eight constituent signatures $S_j(i,k)$.

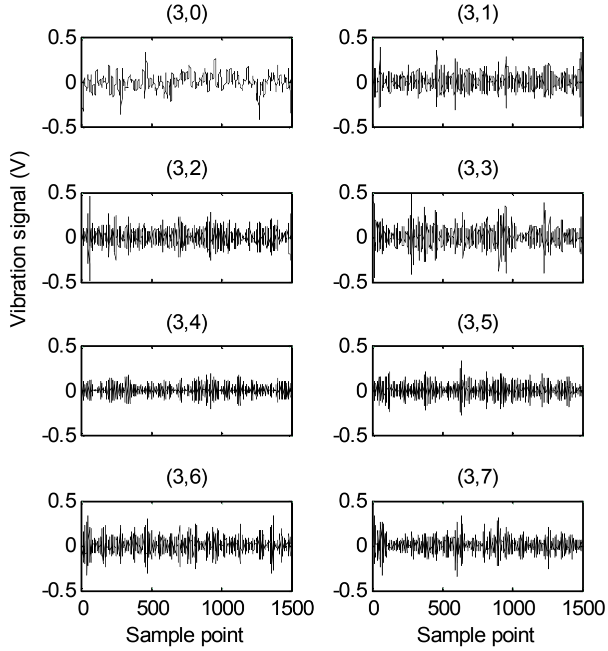

Fig. 4 Constituent signatures in each packet at the third level when three-level DWP is applied.

Fig. 4 illustrates the decomposed signatures over eight distinct bandwidths, where i represents the ith raw signal or decomposed signature, $j = 0, 1, ..., 7$ denotes the jth packet, and k is the kth sampling data. Each reconstructed signature contains some specific vibration energy distribution information, which can be mathematically quantified by a root mean square (RMS) quantity.

The obtained features reflect the energy distribution over different frequency bands, and can be formulated by

$$\gamma(i,j) = \left(\frac{\sum_{k=1}^{NR} S^2{}_j(i,k)}{\sum_{k=1}^{NR} S^2(i,k)} \right)^{1/2}, \tag{1}$$

where R is the number of shaft revolutions to be considered, and N is the number of samples per revolution. N is related to sampling frequency f_s (6000 samples/sec in this case) and the shaft rotation speed f, that is, $N = f_s / f$. Test results have shown that these RMS ratios $\gamma(i,j)$ vary as the bearing health condition changes; correspondingly, the information involved in such a RMS ratio variation may have a potential to be an indicator for bearing condition monitoring.

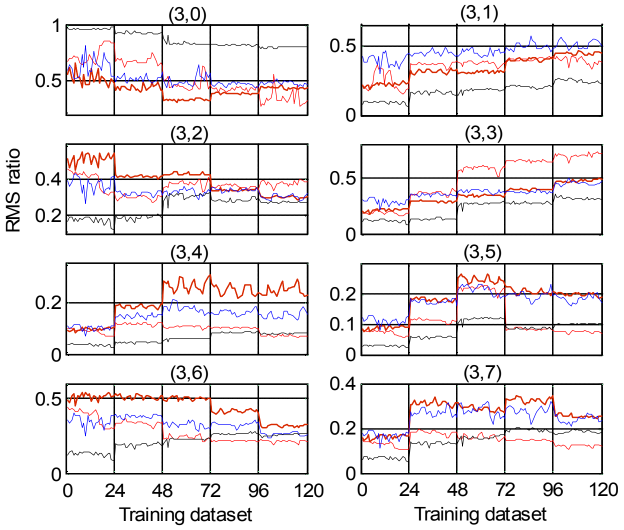

Fig. 5 RMS ratios under different bearing operating conditions: HY bearing (black curves); ED bearing (red curves); IR bearing (blue curves, OR bearing (doted red); 0-24, 600 rpm; 25-48, 900 rpm; 49-72, 1200 rpm; 73-96, 1500 rpm; 97-120, 1800 rpm.

In total, 1000 data sets of $\gamma(i, j)$ are derived, and Fig. 5 shows some of these features. Each data set consists of eight features corresponding to different bearing operating conditions. In order to differentiate these bearing conditions for system training and validation, in this work, the healthy bearing condition is encoded as 0, whereas 1, 2 and 3 are for faulty bearings with the rolling element defects, inner-race defects, and outer-race defects, respectively. The classification thresholds are defined as half state between the nearest encoded numbers. These 1,000 data sets are divided into four groups: 600 pairs for system training, 200 for the training process cross-checking, while the remaining 200 data to validate the proposed feature optimization technique.

3. A GA-based Feature Optimization Technique

The key requirement for an online bearing condition monitoring system is its diagnostic reliability and efficiency. Thus, when a number of potential features are present, an effective feature optimization technique is highly demanded. The objective of this work is to develop an advanced feature optimization technique for a more positive assessment of bearing health conditions. As illustrated in Fig. 6, the proposed feature reformulation scheme consists of the following units: 1) the DWP module is to investigate the energy distributions over several frequency bands for the feature formulation. x_1 to x_8 correspond to the features extracted from packets (3,0) to (3,7), respectively. 2) The GA unit is applied to choose the optimal feature formulation through genetic evolution processes, guided by the proposed fitness function as discussed in the following context. 3) The adaptive neural fuzzy (NF) system [13] is adopted for bearing condition diagnostics.

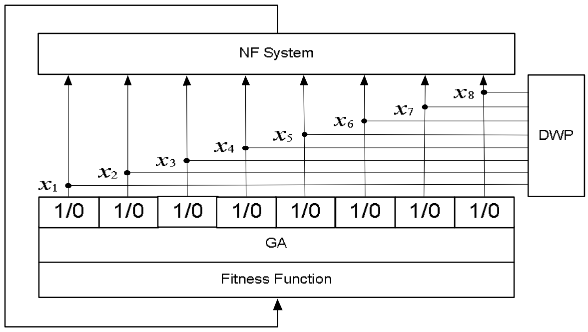

Fig. 6 Architecture of the GA-based feature optimization scheme.

3.1 GA Implementation

The GA is an optimization paradigm based on the principles of natural selections. Each individual within populations is encoded by means of a binary string that can be regarded as a set of genes constituting the chromosomes. Three GA operators are applied to generate the new generation: selection, mutation and crossover. These operations drive the evolution of populations towards an optimal solution based on a prescribed fitness function. A general introduction to the GA can be found in [14].

In this work, each chromosome consists of eight bits, and each bit corresponds to one specific feature with a value of either "0" or "1", representing the absence or presence of the corresponding feature, respectively. The GA has a population size of N_d individuals (N_d =20 in this work). The initial genomes are randomly generated. A modified fitness roulette method is adopted as a selection strategy to choose the related parents. First, the individuals or chromosomes are listed from the first to last as d_i (i = 1, 2, . . ., N_d) according to their fitness values. The first two with the largest fitness values (i.e., d_1 and d_2) are called elitists, and are reserved into the next generation without change, which can keep the excellent performance of the population and guarantee the convergence of the evolution process. The remaining individuals are then processed by the roulette wheel strategy, which is adopted to choose the better individuals of a group to improve the quality of the population and to overcome the premature due to the over-production of some good individuals. The selection probability of the ith individual d_i based on the roulette wheel strategy is given by

$$P(C_i) = \frac{f(d_i)}{\sum_{i=3}^{N_d} f(d_i)}, \qquad i \geq 3, \tag{2}$$

where $f(d_i)$ is the fitness value of individual d_i . It is seen that the probability for each individual to be selected is directly proportional to its fitness. In order to promote the population diversity and to prevent the algorithm from premature convergence, the number of individuals selected from both elitists and the roulette wheel strategy is limited to 5 for the generation of a new population. The remaining ($N_d - 5$) individuals are processed by the crossover and mutation operations. In this work, the crossover probability is set at 0.8, whereas the mutation probability is 0.2.

The fitness function is tailored to satisfy the following three requirements: 1) the high classification accuracy must be guaranteed, 2) the dimension of the input vector is expected to be trimmed down, and 3) the difference between the classification results and the real target should be as small as possible. Hence, the fitness function is proposed to comprise of three weighted elements,

$$f = L \times P_L + N_{ic} \times P_N + E_{rms} \times P_E \times \sqrt{M}, \tag{3}$$

where $L = \sum_{i=1}^{8} b_i$ is the number of the features chosen by an individual; b_i denotes the ith bit value of a chromosome; N_{ic} is the number of incorrect classifications counted from the processing results given by the NF system; E_{rms} is the root mean squared error between the output of the NF system and the real target over all the training data sets; M is the number of the training data sets, whereas $\sqrt{M}$ is for normalization; P_L, P_E and P_E are the percentage weights specified by the user, which represent the priority levels of three elements of L , N_{ic} and E_{rms} , respectively, $P_L + P_N + P_E = 1$. In implementation, more emphasis should be placed on the reliability of classification; that is, for an individual leading to incorrect bearing fault diagnostic, higher percentage of penalty should be applied, forcing such a candidate to have less chance to be selected into the next generation. In this work, P_E is set at 60%, whereas P_L and P_E are set at 20% and 20%, respectively.

It is seen that the fitness function takes the performance of the diagnostic system with respect to different feature combinations; therefore, the GA is needed to cooperate with the NF scheme to accomplish the task of population evolution. Artificial neural networks and the NF systems have been used in machinery condition monitoring for decades, and their advantages over classical model based schemes have been demonstrated in the previous studies [10],[15],[16]. Each feature may carry part information related to bearing health conditions. To integrate the advantages of several features for a more positive assessment of bearing health conditions, a zero order Takagi-Sugeno NF scheme is employed in this work. Fig. 7 shows the structure of this NF scheme. x_1 to x_L are the features chosen by the GA or the user; q_i (i = 1, …, m) are constants that will be fine-tuned in the training process; and m represents the number of the fuzzy rules involved. These fuzzy rules can be fully or partially populated, depending on specific applications. In this work, fully populated fuzzy association is applied. Also, sigmoid membership functions are utilized considering the high-level nonlinearity of the model of interest. A hybrid training algorithm based on the least-squares estimate and steepest gradient method is applied to train the NF classification scheme [17].

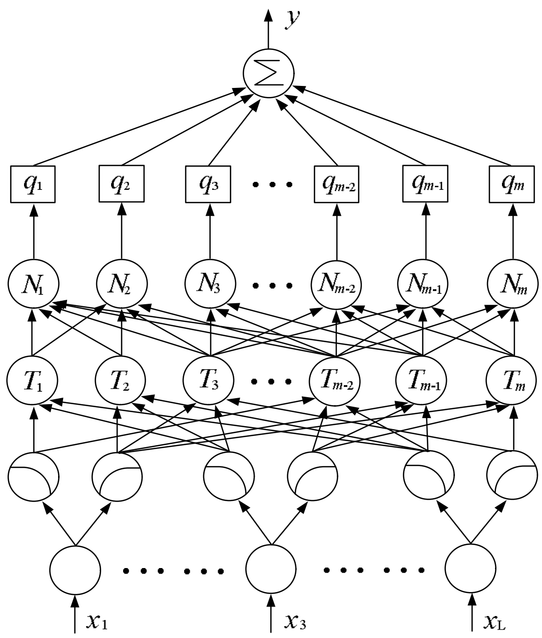

Fig. 7 Network architecture of the zero-order Takagi-Sugeno NF classifier.

3.2 Performance Evaluation

For the purpose of comparison, all eight features are first processed by using the NF classifier. The program takes 18 minutes to run on a Pentium III 1 GHz computer with 512 Mb RAM. The processing results are plotted in Fig. 8. It can be seen that three test cases are mistakenly diagnosed (or three false alarms). A healthy condition is misclassified as a state with a rolling element defect. A rolling element damage condition is misclassified as a healthy condition. An inner-race defect condition is misclassified as a faulty state with an outer-race bearing fault. Furthermore, the high requirement on computing time and training resources (e.g., representative training data) impose a challenge to this algorithm for its real-world applications.

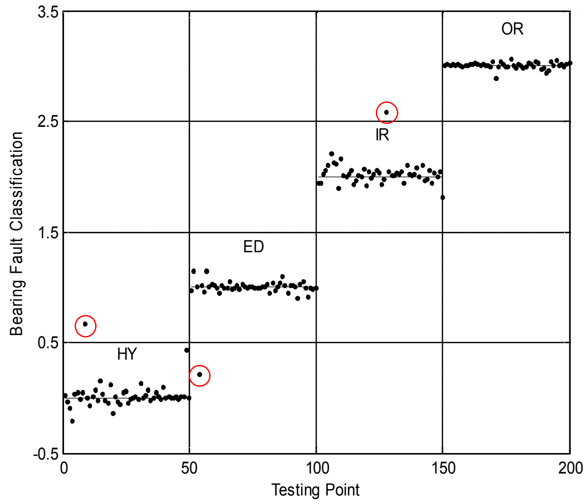

Fig. 8 Classification results using all eight features (mistakenly diagnosed cases are circled).

Considering the size of search space, the GA is allowed to run for 5 consecutive generations. Consequently, the fitness function is executed 100 times in total, instead of the complete space search (i.e., $\sum_{i=1}^{8} C_8^i = 255$ cases). The surviving individuals and their corresponding scores are listed in Table 1. It can be seen that the optimal individual is "1001 1111," which represents the feature formulation from packets (3, 0), (3, 3), (3, 4), (3, 5), (3, 6) and (3, 7). The optimal feature formation can reduce the fitness function score to a level as low as 0.7923, which represents the best fitness condition.

Table 1: Best individuals and their scores in the final population

Individual	0111 1011	1001 1111	1100 1111	1101 0111	1011 1111	1101 1010	1101 1011
Score	1.78 2	0.79 2	2.91 2	1.99 0	1.83 8	1.82 2	1.92 7

To validate the efficiency of the proposed technique for feature reformulation, the optimized features are further processed by the NF classifier. By a series of tests and comparisons, higher classification accuracy has been achieved for this bearing fault diagnostic operation. As demonstrated in Fig. 9, for a total of 200 validating data sets, except one false alarm (i.e., a healthy state is beyond the threshold [-0.5, 0.5]), all other bearing conditions are correctly classified. This diagnostic reliability is much higher than that achieved prior to feature optimization. Furthermore, the program computation time is dramatically reduced (around 30 seconds) with the same computer, Pentium III 1 GHz with 512 Mb RAM.

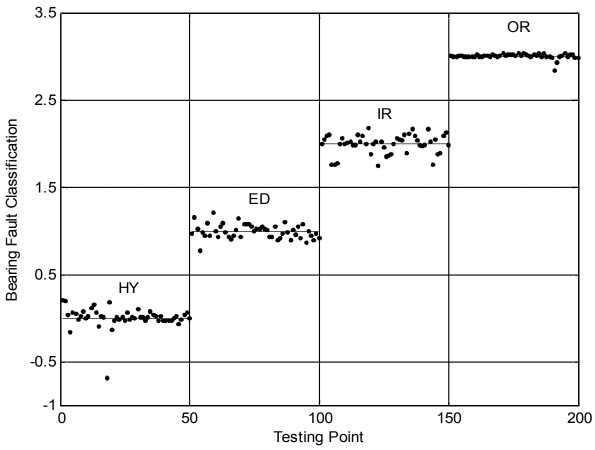

Fig. 9 Classification results using six GA-optimized features.

4. Conclusion

A GA-based feature optimization technique is proposed in this paper for bearing fault diagnosis. Representative features corresponding to different bearing health conditions are extracted from the raw vibration signals by using the DWP analysis. Two diagnostic methods are examined for the bearing fault detection and fault type classification. The diagnostic system based on all the originally extracted features and an NF classifier is sluggish in training, and its reliability is low in bearing fault diagnostics. Based on the developed GA-based feature optimization technique, potential features are effectively optimized. Its training efficiency and excellent classification reliability have been demonstrated by experimental tests.

Acknowledgments

This project was financially supported by the Natural Sciences and Engineering Council of Canada (NSERC) and Mechworks Systems Inc.

References

[1] Y. T. Su, S. J. Lin, "On initial fault detection of a tapered roller bearing: Frequency domain analysis," *J. Sound Vib* , vol. 155, pp. 75-84, 1992.

[2] B. Samanta, K. R. Al-Balushi, "Artificial neural network based fault diagnostics of rolling element bearings using time-domain features," *Mech. Syst. Signal Processing*, vol. 17, pp. 317-328, 2003.

[3] W. Wang, F. Golnaraghi, F. Ismail, "Condition monitoring of multistage printing presses," *J. Sound Vib*, vol. 270, pp. 755-766, 2004.

[4] R. Q. Yan, R. X. Gao, "An efficient approach to machine health diagnosis based on harmonic wavelet packet transform", *Robotics and Computer-Integrated Manufacturing*, vol. 21, pp. 291-301, 2005.

[5] R. B. W. Heng, M. J. M. Nor, "Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition," *Applied Acoustics*, vol. 53, pp. 211-226, 1998.

[6] L. Rende, T. Dehua, Using oil analysis to study the wear condition of bearing in trunnion of converter during/after run-in period, *Proc. of the 5th International Conference on Quality, Reliability and Maintenance QRM 2004*, pp. 101-104, 2004.

[7] J. Yi, J. Henao-Sepulveda, M. Toledo-Quinones, Wireless temperature sensor for bearing health monitoring, *Proc. of SPIE 5391*, pp. 368-376, 2004.

[8] R. R. Schoen, T. G. Habetler, F. Kamran, R. G. Bartheld, "Motor bearing damage detection using stator current monitoring," *IEEE Trans. Industry Applications*, vol. 13, pp. 1274-1279, 1995.

[9] N. Tandon, A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology International*, vol. 32, pp. 469-480, 1999.

[10] X. Lou, K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mech. Syst. Signal Processing*, vol. 18, pp. 1077-95, 2004.

[11] B. A. Paya, I. I. Esat, M. N. M. Badi, "Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor," *Mech. Syst. Signal Processing*, vol. 11, pp. 751-65, 1997.

[12] R. M. Rao, A. S. Bopardikar, *Wavelet transforms: introduction to theory and applications*. Addison-Wesley, England, 1998.

[13] J. S. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Trans Systems, Man and Cybernetics*, vol. 23, pp. 665-685, 1993.

[14] D. A. Coley, *An introduction to genetic algorithms for scientists and engineers*. River Edge, NJ: World Scientific, Singapore, 1999.

[15] L. B. Jack, A. K. Nandi, A. C. McCormick, "Diagnosis of rolling element bearing faults using radial basis function networks," *Applied signal processing*, vol. 6, pp. 25-32, 1999.

[16] M. Subrahmanyam, C. Sujatha, "Using neural networks for the diagnosis of localized defects in ball bearings," *Tribology International*, vol. 30, pp. 739-752, 1997.

[17] W. Wang, F. Ismail, F. Golnaraghi, "A neuro-fuzzy approach for gear system monitoring," *IEEE Trans. Fuzzy Systems*, vol. 12, no. 5, pp. 710-723, 2004.