ORIGINAL ARTICLE

Rolling Element Bearing Condition Monitoring Based on Vibration Analysis Using Statistical Parameters of Discrete Wavelet Coefficients and Neural Networks

Vahid Kazemi Golbaghi[1,*], Mehdi Shahbazian[1], Bahman Moslemi[1], and Gholamreza Rashed[1]

[1]*Petroleum University of Technology, Iran, Islamic Republic of*
(**Received** 12 August 2016; **Accepted** 13 January 2017; **Published on line** 1 June 2017)
*Corresponding author: vahid.kazemi.g@gmail.com
DOI: 10.5875/ausmt.v7i2.1201

Abstract: There are several techniques that can be used to determine the condition of a rolling element bearing. In this paper, vibration analysis is used to conduct fault diagnosis of a bearing. Vibration signal noise was eliminated using hard thresholding wavelet analysis. The best mother wavelet for the denoising process was selected using the minimum Shannon entropy criterion. Statistical parameters and other signal properties such as energy and entropy are powerful tools for analyzing vibration signals. These features were calculated in the time and wavelet domains and applied to Artificial Neural Networks (ANNs) as the feature vector to classify the condition of a bearing into one healthy and three faulty conditions. The ANN parameters were separately optimized using three optimization algorithms. The comparison of the results shows that if the ANN parameters are properly optimized, the statistical parameters in the time-frequency domain can optimize accuracy.

Keywords: rolling element bearing, condition monitoring, vibration analysis, discrete wavelet transform, statistical parameters, optimized artificial neural networks

Introduction

Safety and reliability are crucial issues in mechanical system maintenance. Rolling element bearings are key components in rotary machines and play a vital role in machine condition. Various techniques have been proposed to monitor bearing conditions based on performance evaluation, vibration signals, motor stator current, shock pulse, acoustic emission, thermography, and wear debris monitoring [1]. Vibration analysis has many advantages and is used widely in monitoring the conditions of bearings and other rotary machine components [2].

Most normal bearing vibration is generated by low frequency components as a result of shaft-rotation, load fluctuation, etc. As a rolling element passes over a defect in a faulty bearing, it will generate periodic impulses which may be masked by noise [3].

It is preferable to remove these noises prior to signal analysis. Thresholding methods are among the more commonly used approaches to reduce noise in vibration signals. Generally, soft thresholding reconstructs a 'noise-free' signal, whereas hard thresholding better preserves some features such as peak height, but this approach provides less smooth substitution [4]. Abbasion et al. used the Meyer wavelet and soft thresholding to denoise the vibration signal of a bearing [5]. Nikolaou et al. defined a specific thresholding value and set the coefficients below that to zero [6].

Signal-processing methods should be used to prepare denoised signals for future application of intelligent fault classification algorithms. it's the adaptability and multi-resolution capability of wavelet analysis make it a powerful tool for rotary machine fault diagnosis [7]. Use of continuous wavelets may produce redundant information [7] which incurs an undesired information cost [8]. However discrete wavelet transform (DWT) provides an efficient way to create feature vectors

Copyright © 2017 International Journal of Automation and Smart Technology

[9].

Different methods have been used to select the best mother wavelet. Rafiee et al. [10] and Paya et al. [11] used db4 mother wavelet by several trials. Nikolaou and Antoniadis [6] used computational cost function and accuracy criteria, and asserted that proper results can be obtained using db12. In 2009, Rafiee and his coworkers proposed a novel method using genetic algorithm to select db11 as the best mother wavelet [12]. Rafiee et al later [13] studied 324 mother wavelets and concluded that Daubechies 44 (db44) has the most similar shape for both gear and bearing vibration signals.

A good feature extraction from wavelet coefficients can help to more efficiently identify machine defects [7]. Statistical parameters are one of the most commonly used feature sets used in machine condition monitoring. Samanta used statistical parameters as the feature vector to determine the condition of a gearbox [14]. This set of features can also be generated by DWT coefficients at any transform resolution level. Lou decomposed the vibration signal of a specific bearing in normal and faulty conditions and used the standard deviation of the discrete wavelet transform coefficient as the feature vector for later analysis [9]. Rafiee et al. proposed using the standard deviation of the wavelet packet coefficient as the feature vector for artificial neural network training [12]. In 2009, Lei et al. extracted six statistical parameters from all frequency bands of the wavelet packet transform and Intrinsic Mode Functions (IMFs) of Empirical Mode Decomposition (EMD) [15].

An artificially intelligent system should then be applied to the extracted features to automatically identify system faults. Artificial neural networks (ANNs) are a type of information processing method that models information processing in the human brain [15], and are one of the most useful methods for condition monitoring fault classification [12]. Jack found that ANN outperformed SVM in terms of system fault diagnosis accuracy, and did so with less training time [16]. In 2007, Rafiee et al. used ANN to diagnosis faults in a gearbox [10].

Vahid Kazemi Golbaghi received his B.Sc. in technical inspection engineering from Petroleum University of technology (PUT), Iran, Ahwaz. Currently, he is a M.Sc. student of instrumentation and automation in PUT. His field of interest is intelligent fault diagnosis and vibration signal analysis.

Dr. Mehdi Shahbazian Received his MSc degree in Electronic Engineering from University of Tehran, Iran and his PhD degree in Signal Processing from the University of Surrey in UK. He is currently an associate professor in Department of Instrumentation and Automation Engineering at PUT. His research areas are Intelligent Systems and wavelet analysis and their applications to Modeling, Identification, Control, Fault diagnosis and Soft sensors.

Dr. Bahman Moslemi received his M.Sc. and PHD in mathematics from Shahid Chamran University, Iran, Ahwaz. He is currently assistant professor in department of science at PUT and his field of interest is applied mathematics and its applications in engineering.

Dr. Gholamreza Rashed received His M.Sc. and PhD in mechanical engineering in Khajeh Nasir Toosi University of Technology, Iran, Tehran. He currently is associated professor at PUT and his field of interest is vibration and its applications in fault diagnosis and fault detection of materials and turbo-motors.

General Procedure

The general procedure of this paper can be summarized in the following steps:

1. Denoise the original signals generated by a rolling element bearing using the hard thresholding wavelet denoising method.
2. Perform one level discrete wavelet transform decomposition of the denoised data, using six different mother wavelets.
3. Select the best mother wavelet based on the minimum Shannon entropy criterion.
4. Calculate nine features of all signals in the time domain and eighteen features in the wavelet domain.
5. Train two neural network classifiers based on time domain and wavelet domain features.
6. Compute neural network weights using three different evolutionary algorithms: GOA, PSO and COA.

For all data sets, the statistical features of the signals, with energy and entropy were used for classification of bearing conditions

Experimental Setup

The experiment was performed using a 2 hp Reliance Electric motor. By means of a self-aligning coupling, the motor was connected to a dynamometer and a torque sensor. The torque load levels were regulated by controlling the dynamometer. A 6205-2RS JEM SKF deep groove ball bearing supported the motor shaft. An Electro-discharge machine (EDM) was used to produce a single point 0.014” fault in the test bearing. Vibration data was collected by attaching an accelerometer with a maximum bandwidth of 5000 Hz and a 1 V/g output to the housing with magnetic bases. The load applied on the motor was 1hp and the shaft rotational speed was 1772 rpm. The data acquisition system consisted of a specially designed amplifier to produce vibration signals with high bandwidth and a data recorder with a sampling frequency of 48000 Hz. A low-pass filter was used for anti-aliasing.

The experimental data and the system are taken from Case Western Reverse Lab and used with the permission of Dr. Kenneth A. Loparo. The data set was classified in four different conditions: normal bearing, bearing with inner race fault, bearing with outer race fault and bearing with rolling element fault. All of the vibration

Copyright © 2017 International Journal of Automation and Smart Technology

signals in each condition were cut into 100 samples each with a length of 3700 points. Figure 1(a) to (d) show one sample of each bearing condition.

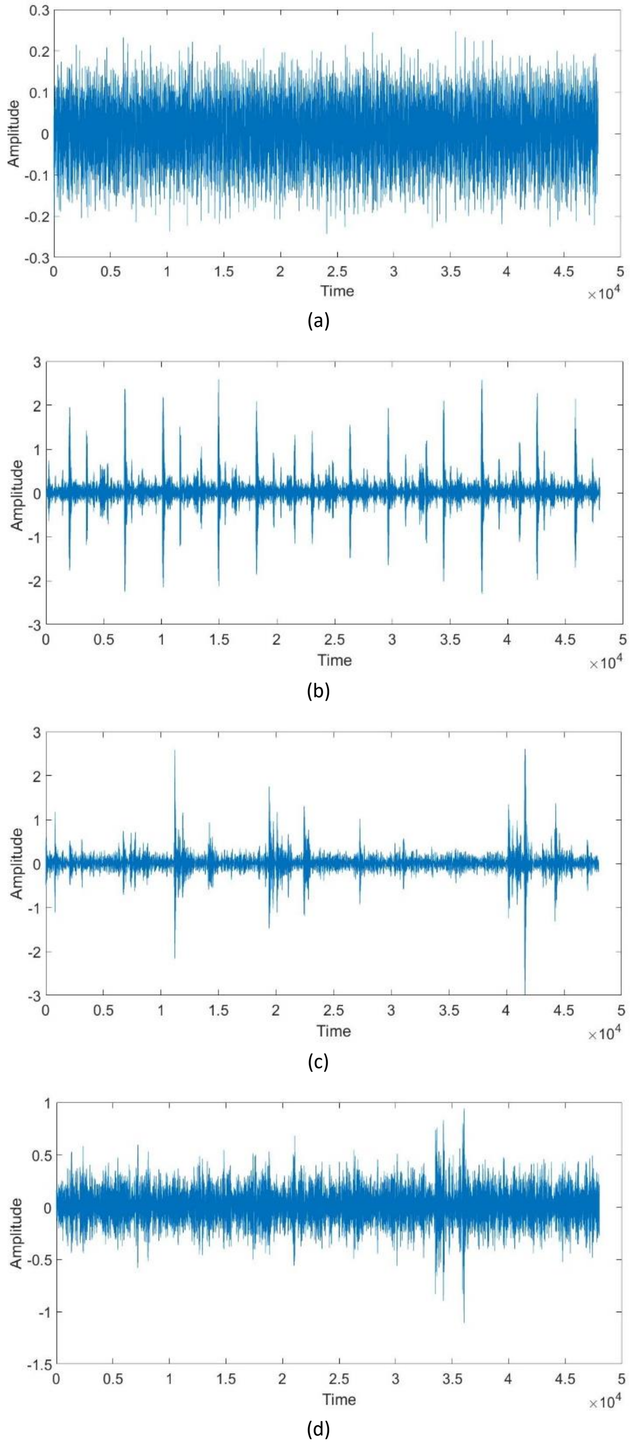

Figure 1. Vibration signals in different conditions: (a) healthy (b) inner race fault (c) rolling element fault (d) outer race fault.

Signal Denoising Using Wavelet Transform

Wavelet Fundamentals

In 1882, Morelet et al. introduced wavelets as a family of functions constructed by a single function called the "mother wavelet" [17]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), a, b \in \mathbb{R}, a \neq 0, \quad (1)$$

where a and b are respectively called scaling and translation parameters. The continuous wavelet transform of a finite energy signal $\left(f(t) \in L^2(R)\right)$ is defined by Grossmann in the following form:

$$W_f(a,b) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} f(t)\psi^*\left(\frac{t-a}{b}\right) \mathrm{d}t, \quad (2)$$

where $W_f(a,b)$ is the continuous wavelet transform of f and $\psi^*$is the conjugation of the mother wavelet. The function f can be reconstructed by inverse wavelet transform which is formulated as follows:

$$f(t) = C_\psi^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_\psi[f](a,b)\psi_{a,b}(t)(a^{-2}da)db, \quad (3)$$

where C_ψ satisfies the admissibility condition

$$C_\psi = 2\pi\int_{-\infty}^{\infty}\frac{|\hat{\psi}(\omega)|}{|\omega|}d\omega < \infty, \quad (4)$$

and $\hat{\psi}(\omega)$ is the Fourier transform of the mother wavelet $\psi(t)$.

Discrete Wavelet Analysis

The Discrete Wavelet Transform (DWT) consists of introducing a signal $s(x)$ to low-pass (L) and high-pass (H) filters. As the results, two vectors cA_1 and cD_1 will be obtained for which the elements of vector cA_1 correspond to the low frequencies of the signal and are called approximation coefficients while the elements of the vector cD_1 correspond to the highest frequencies and are called detail coefficients [18]. The procedure can be repeated with the elements of the vector cA_1 and successively with each new vector cA_j obtained. The signal can be decomposed into N scale, and one approximation coefficient and N detail coefficients can be obtained.

In this scheme, a and b are given by [9]:

$$a = 2^j, b = k2^j, (\mathrm{j}, \mathrm{k}) \in Z^2 \quad (5)$$

if

$$\psi_{j,k}(x) = 2^{-\frac{1}{2}}\psi(2^{-j}x - k) \quad (6)$$

$$\emptyset_{j,k}(x) = 2^{-\frac{1}{2}}\emptyset(2^{-j}x - k) \quad (7)$$

the wavelet ψ is substituted by a wavelet filter with impulse g; and a the scaling function $\emptyset$ is replaced by scaling filter with impulse response h. g and h are defined on a regular grid ΔZ; where Δ is the sampling period. The mathematical description of the discrete wavelet analysis can thus be represented as [9]:

$$\mathrm{C(a,b)} = \mathrm{c(j,k)} = \sum_n s(n)g_{j,k}(n), \quad j,k \in N \quad (8)$$

Copyright © 2017 International Journal of Automation and Smart Technology

Moreover, discrete synthesis produces [9]:

$$s(x) = \sum_{j\in z}\sum_{k\in z} c(j,k)\psi_{j,k}(x) \quad (9)$$

The detail at level *j* and the approximation at level *J* are respectively defined as [9]:

$$D_j(x) = \sum_{k\in z} c(j,k)\psi_{j,k}(x) \quad (10)$$

$$A_{j-1} = \sum_{j>J} D_j \quad (11)$$

During the decomposition, the signal s(x) and vectors cA_j undergo a downsampling. Vector A_j and D_j which are respectively called approximation and detail and satisfy the relations [18]:

$$A_{j-1} = A_j + D_j \quad (12)$$

$$s = A_j + \sum_{i<j}^{N} D_j \quad (13)$$

where i and j are integers. Figure 2 represents an example of three level wavelet decomposition.

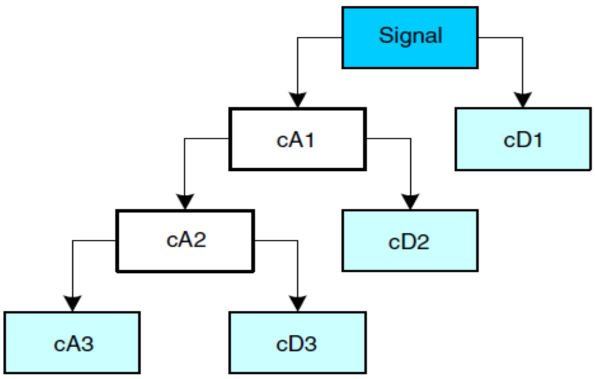

Figure 2. Multiple level wavelet decomposition scheme [19].

Wavelet Denoising by Thresholding

Suppose that the desired signal f is blended by σ-variance zero-mean Gaussian white noise as follows:

$$g[n] = f[n] + \sigma z[n] \quad (14)$$

To separate the signal from noise Donoho and Johnstone proposed a method in the wavelet domain called wavelet thresholding as follows:

(1) Transfer the signals to the time-frequency domain by wavelet transform and obtain the scaling coefficient and wavelet coefficient.

(2) Determine a threshold value and apply it to the wavelet coefficients.

(3) Reconstruct the denoised signal by means of the remaining wavelet coefficient [3].

Determining the mother wavelet and thresholding value plays a vital role in this method. The mother wavelet should be selected such that the information cost function is minimized [8]. Different criteria exist to describe the cost function, but the most common criterion is Shannon entropy. For a discrete signal X of unit energy, the Shannon entropy of the decomposition coefficient sequence $X = \{x^k\}$ is computed by [19]:

$$M(X) = -\sum_k (x^k)^2 \log(x^k)^2 \quad (15)$$

To achieve this goal, wavelet denoising is performed using 10 mother wavelets. As shown in Table 1 the entropy of all gathered data is calculated and compared. The mother wavelet that regenerates the signal with minimum entropy is selected as the best mother wavelet.

Thresholding value is a combination of adaptive threshold selection using the principle of Stein's Unbiased Risk Estimate and fixed-form threshold using the MATLAB function 'heursure'.

Figure 3 shows one sample of vibration signal before and after denoising.

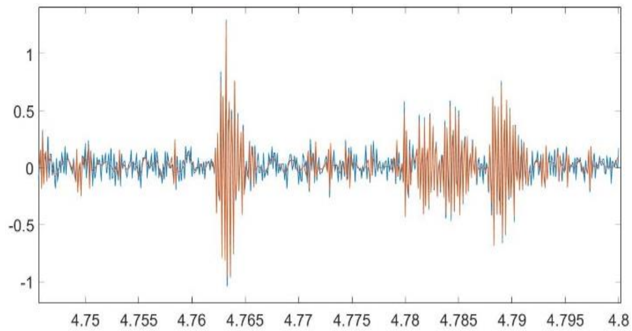

Figure 3. Inner race fault signal. The orange signal shows the denoised signal and the blue signal is the original signal

Mother Wavelet Selection for Discrete Wavelet Transform

Discrete wavelet transforms the signals to the time-frequency domain. Equation (18) was used to select the best mother wavelet minimum entropy criteria. For this goal, the mother wavelets used most commonly widely in condition monitoring were studied, including db4, db8, db10, db11, db12 and db44.

Summing calculated entropy of all signals in each condition obtained the following results in table 2.

As shown in table 1, db8 had the minimum entropy, so it was selected as the mother wavelet to transform the signals to the time-frequency domain.

Denoised signals were decomposed by DWT in only one decomposition level and the statistical properties of each sub signal are evaluated in both high and low band filters.

Feature Extraction

The dimensionless statistical characteristics can effectively identify system faults [15]. Entergy and entropy are two other parameters which are widely used in monitoring the condition of rotary machines. The

Copyright © 2017 International Journal of Automation and Smart Technology

following features for the discrete vibration signals are extracted: root Mean Square, standard deviation, crest factor, kurtosis, skewness, impulse factor:

Table 1. Shannon entropy of the signals after denoising. Highlighted cells shows the minimum entropy for each state.

	Healthy	Outer race fault	Inner race fault	Ball fault
Meyer	3.885e+03	4.283e+03	4.576e+03	4.240e+03
Db2	3.885e+03	4.272e+03	4.575e+03	4.227e+03
Db3	3.888e+03	4.279e+03	4.577e+03	4.232e+03
Db4	3.884e+03	4.281e+03	4.578e+03	4.236e+03
Db5	3.883e+03	4.282e+03	4.578e+03	4.240e+03
Db6	3.885e+03	4.283e+03	4.577e+03	4.240e+03
Db7	3.882e+03	4.282e+03	4.578e+03	4.239e+03
Db8	3.884e+03	4.284e+03	4.577e+03	4.241e+03
Db9	3.884e+03	4.283e+03	4.577e+03	4.240e+03
Db10	3.884e+03	4.283e+03	4.577e+03	4.240e+03

Table 2. Shannon entropy for mostly used mother wavelets.

	Db4	Db8	Db10	Db11	Db12	Db44
Ent	3.66E+05	3.37E+05	3.39E+05	3.47E+05	3.45E+05	3.41E+05

$$IF(X(n)) = \frac{Max(X(n))}{\frac{1}{n}\sum_{n=1}^{N}|x(n)|} \quad (16)$$

FM4 parameter:

$$FM4(X(n)) = \sum_{n=1}^{N} \frac{4th\ moment(X(n))}{Var(X(n))} \quad (17)$$

where *Var* is the variance of the signal,

Energy

$$Enr(X(n)) = \sum_{n=1}^{N} (x(n))^2 \quad (18)$$

and Entropy. First, all of these parameters are extracted from the de-noised vibration signals in the time domain to determine bearing operation conditions. Moreover, the same parameters are calculated for both approximation and detail coefficients of the discrete wavelet transform. The extracted features are normalized in the range of [-1 1] and then applied to the Artificial Neural Networks (ANNs) to automatically determine the bearing's condition.

Artificial Neural Network Architecture

One of the most important parts of the fault diagnosis method is pattern classification. Nowadays, ANNs are one of the most common and powerful ways to classify patterns. An artificial neural network is built of connections which receive and send information [20]. Feed-forward is one of the most commonly used ANN topologies due to its simple construction, variety of the existing training algorithms, and excellent performance [21]. A feed-forward artificial neural network is constructed of three main layers: Input nodes, where the number of nodes equals the dimensions of the feature vector's dimensions; the hidden layer; and the output layer where the number of nodes equals the number of classes. These layers are connected to each other by weighted lines, with the information moving from the input layer to the hidden layers and finally to the output layer. The calculated values in each node are added to a constant value called the biasing value. The output of each node (neuron) is presented as a function called the transfer or activation function. The output value of each node is calculated by following equation:

$$z_j(t) = f\left(\sum_{i=1}^{n}\sum_{j=1}^{m} W_{ij} x_i(t) + b_j\right) \quad (19)$$

where z_j is the output of the j^{th} neuron, W_{ij} is the weight of the connection and b_j is the bias of the j^{th} neuron [21].

Almost any transfer function can be used in a feedforward neural network, but for classification problems, nonlinear transfer functions work best [22]. The activation function for the hidden and output layers is considered to be a tangent sigmoid.

The number of neurons in the hidden layer is set at 10 by trial and error. 70% of the data is used for network training, while the remaining 30% is used for testing. Least mean squares is chosen as the error function. Many optimization algorithms can be used to optimize network parameters. In the following sections, the evolutionary algorithms used in this paper are described briefly.

Genetic Algorithm (GA)

Genetic algorithm imitates natural genetic mechanisms to produce optimal solutions [23]. The main idea is to preserve a primary chromosome, eliminating weaker solutions in a process of elimination [24]. In the first case, GA is used to modify ANN parameters [25]. Generally, a simple GA consists of three main procedures: (1) selection of parents (2) crossover and (3) mutation

Copyright © 2017 International Journal of Automation and Smart Technology

[12].

The initial population size was considered to be equal to 600 genomes which are randomly distributed in the solution space. A relatively high population size is selected to ensure that the interchanges among the genomes are relatively high and slow algorithm convergence.

Tournament selection is used to select the parents in each iteration, and the maximum number of iterations (401) is used as the stopping criteria for the algorithm.

The training MSE curves are shown in Figure 4.

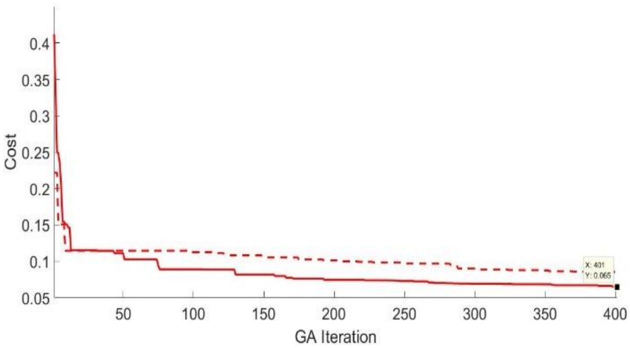

Figure 4. MSE curves for ANNs in both time and time-frequency domain using GOA.

Particle Swarm Optimization

Proposed by Kennedy and Eberhart in 1995, the PSO is inspired by the unpredictable behavior of flocking birds [26]. This method is an iterative optimization algorithm in which a group of solutions in the search space is iteratively updated to localize the optimum solution [27].

PSO uses a direct search method to find the best solution. All particles represent probable answers to the problem and move in a D-dimensional space according to two concepts: first, all particles remember the position which minimizes its cost (its 'personal best'); second, all particles communicate with each other to recognize the best particle ('global best'). Each particle randomly moves toward its best position and the global best position according to the following equation:

$$v_i^d(k+1) = v_i^d(k) c_1 rand_1^d(\mathrm{k}) \left(pBest_i^d(k) - x_i^d(k)\right) + c_2 rand_2^d(k) \left(gBest_i^d(k) - x_i^d(k)\right) \quad (20)$$

Then, the updated particle position can be determined by:

$$x_i^d(k+1) = x_i^d(k) + v_i^d(k+1)(k+1)(k+1) \quad (21)$$

Where x_i^d Is the corresponding position of the particle, k is the current iteration step, c_1 and c_2 are acceleration factors and $rand_1^d$ and $rand_2^d$ are random weights uniformly distributed in the range (0, 1).

The above process is repeated in each iteration until the ending criteria is met. This criteria could be a certain number of iterations or any other threshold determined by the user.

Number of particles and iterations are considered equal to genomes and iterations in previous algorithms for easier comparison. As shown in Figure 5, PSO greatly reduces network MSE.

Like other iterative optimization algorithms, PSO has limitations: premature convergence into a suboptimal solution and slow algorithm convergence.

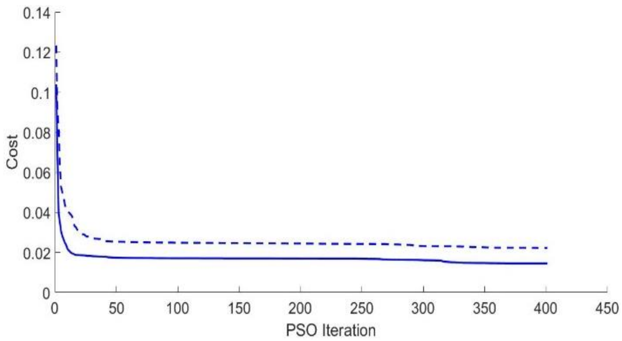

Figure 5. MSE curve for ANNs in both time and time-frequency domain using PSO.

Cuckoo Optimization Algorithm (COA)

This algorithm is inspired the egg laying behavior of the cuckoo bird. Similar to other evolutionary algorithms, COA starts with some initial population of cuckoos living in discrete groups. The cuckoos lay their eggs in the nests of other host birds according to their descent and are classified into several groups. Master race eggs will survive in the host birds' nest, while retardation eggs will be killed by the host bird. The remaining eggs will hatch and become adult cuckoos. To maximize egg survival, cuckoos search the nearby area for suitable host nests.

These cuckoos gradually create societies in the feature space and the procedure continues. All of these societies are in communication with each other and cuckoos will be inclined to migrate to an optimal habitat, thus iteratively converging on the best situation.

All the cuckoos situate in a specific 'habitat' -which is an N_{var}-dimensional matrix representing the potential solution of unknowns of the problem. 'Profit', which presents the cost value of each cuckoo habitat, is defined as:

$$profit = -Cost = -f_c(x_1, x_2, \cdots x_{N_{var}}) \quad (22)$$

where $x_1, x_2, \ldots$ determine the cuckoo habitats. The negative sign in the formula is due to the method seeking to maximize profit value. Real world cuckoos can lay anywhere from 5 to 20 eggs, but only within a deterministic distance. All of these concepts are considered in the algorithm design. The number eggs laid is determined by the user and 'Egg Laying Radius (ELR)' is

Copyright © 2017 International Journal of Automation and Smart Technology

determined by the following formula:

$$ELR = \alpha \times \frac{Numberof\ current\ cuckoo's eggs}{Total\ number\ of\ eggs} \times (var_{hi} - var_{low}) \quad (23)$$

where alpha is an integer which determines maximum laying radius.

As previously mentioned, unsuitable eggs are detected and thrown out by the host bird. In this algorithm, high cost generated points represent the unsuitable eggs and will be vanished by the algorithm.

Survivor eggs will hatch and multiply within their host settlement the following gestation period at which point cuckoos search for better habitats offering more suitable nesting opportunities. Immigration in the algorithm is performed by means of a parameter called λ. Cuckoos traverse λ% of the path toward the goal and with a deviation equal to λ from the direct path. λ ranges from 0 and 1 and differs for each cuckoo.

The gradual increase in the number of cuckoos is restricted by hunters and limited sources of sustenance, and is applied in algorithm through a parameter called 'maximum number of live cuckoos in the environment' and determined byN_{max} .

After a while, the total cuckoo population will converge to a habitat with maximum profit. When more than 95% of the cuckoo population converges, the process ends.

Similar to other evolutionary algorithms, COA starts with an initial population, numbering 15 in this case. Each cuckoo can lay two to five eggs in each iteration, with cuckoos lifespans lasting a maximum of 50 iterations. Cuckoos are classified in three individual groups using the k-means clustering method. Like the other two algorithms, the stopping criteria is maximum number of iterations (401).

Figure 6 shows the results of the cuckoo optimization algorithm. In comparison with standard GA and PSO, COA can find the global best position more accurately.

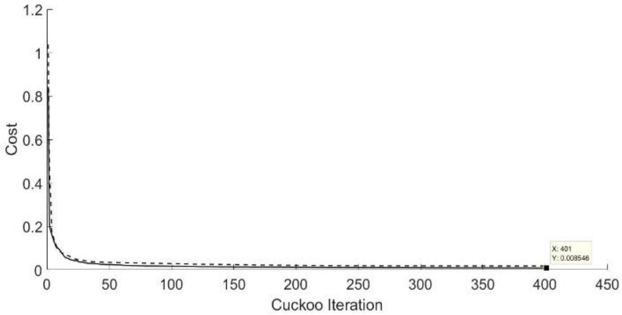

Figure 6. MSE curve for ANNs in both time and time-frequency domain using COA

Results and discussion

As it was previously explained, the statistical properties for the denoised vibration signals are evaluated in both the time and wavelet domains. This produces nine features in the time domain and eighteen features in the time-frequency domain. These features are then individually applied the feature vector in an artificial neural network classifier. The prosperity percentage, which is the ratio between correctly classified data to all data applied to neural networks for training and testing for all optimization algorithms (GA, PSO and COA), and the mean squared error for all cases is presented in table 3 and 4. Table 3 shows the results in the time domain and Table 4shows the results of the wavelet domain features.

These tables show that a single level of signal decomposition by DWT significantly improved the fault diagnosis results. Another way to improve classification results is to select the most proper optimization algorithm. As shown in tables 3 and 4, the cuckoo optimization algorithm provides the best classification results, along with minimum time cost. Finally as shown in Table 4, DWT decomposition and cuckoo ANN training

Table 3. Bearing condition monitoring results obtained by GOA, PSO and COA using time domain parameters.

	Prosperity percentage for train data	Prosperity percentage for test data	MSE train	MSE test	Averaged elapsed time
GA	94.09%	95.5%	0.0943	0.0958	2621
PSO	95%	95.5%	0.0215	0.0214	2547
COA	95.81%	95.75%	0.0182	0.0174	924

Table 4. Bearing condition monitoring results obtained by GOA, PSO and COA using time-frequency domain parameters.

	Prosperity percentage for train data	Prosperity percentage for test data	MSE train	MSE test	Averaged elapsed time
GA	95.06%	96%	0.0684	0.0664	2476
PSO	99.97%	99.75%	0.0108	0.0102	3526
COA	100%	100%	0.0072	0.0061	896

Copyright © 2017 International Journal of Automation and Smart Technology

produces the best bearing fault diagnosis method with perfect classification (100%) and lower CPU time consumption.

Conclusion

This research seeks to determine bearing condition based on discrete wavelet analysis. Raw vibration signals are denoised using the wavelet hard thresholding method. A feature vector was constructed using statistical features, energy and entropy from the denoised signals. These features were calculated in both the time and time-frequency domains (using DWT). ANN was used as the classifier to categorize the data to four different conditions. ANN parameters were optimized using three evolutionary optimization algorithms: GOA, PSO and COA. Simulation results show that the optimization algorithm produces a significant improvement in classification accuracy, and COA significantly outperforms both PSO and GA. Comparing the results in the time and wavelet domains finds that DWT signal decomposition, even in a single level of decomposition, can greatly improve fault classification performance.

Reference

[1] L. Wang and R. X. Gao, *Condition monitoring and control for intelligent manufacturing*, Berlin, Heidelberg, Germany: Springer Science & Business Media, 2006.

[2] R. B. Randall, *Vibration-based condition monitoring: industrial, aerospace and automotive applications*, Hoboken, New Jersey: John Wiley & Sons, 2011.

[3] X. Wang, Y. Zi, and Z. He, "Multiwavelet denoising with improved neighboring coefficients for application on rolling bearing fault diagnosis" *Mechanical Systems and Signal Processing*, vol. 25, no. 1, pp. 285-304, 2011.
doi: 10.1016/j.ymssp.2010.03.010

[4] D. L. Donoho and I. M. Johnstone "Threshold selection for wavelet shrinkage of noisy data," in Proceedings of *the 16th Annual International Conference of the IEEE*, Baltimore, MD, Nov. 3-6, 1994, pp. A24-A25.
doi: 10.1109/IEMBS.1994.412133

[5] S. Abbasion, A.Rafsanjani, A. Farshidianfar, and N. Irani, "Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine" *Mechanical Systems and Signal Processing*, vol. 21, no.7, pp. 2933-2945, 2007.
doi: 10.1016/j.ymssp.2007.02.003

[6] N. Nikolaou and I. A. Antoniadis, "Rolling element bearing fault diagnosis using wavelet packets," *NDT & E International*, vol.35, no.3, p. 197-205, 2002.

[7] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol.96, part A, pp. 1-15, 2014.
doi: 10.1016/j.sigpro.2013.04.015

[8] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on information theory*, vol. 38, no. 2, pp. 713-718, 1992.
doi: 10.1109/18.119732

[9] X. Lou and K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mechanical systems and signal processing*, vol.18, no. 5, pp. 1077-1095, 2004.
doi: 10.1016/S0888-3270(03)00077-3

[10] J. Rafiee, F. Arvani, A. Harifi, and M. H. Sadeghi, "Intelligent condition monitoring of a gearbox using artificial neural network," *Mechanical systems and signal processing*, vol. 21, no. 4, pp. 1746-1754, 2007.
doi: 10.1016/j.ymssp.2006.08.005

[11] B. A. Paya, I. I. Esat, and M. Badi, "Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor," *Mechanical systems and signal processing*, vol. 11, no.5, pp. 751-765, 1997.
doi: 10.1006/mssp.1997.0090

[12] J. Rafiee, P. W. Tse, A. Harifi, and M. H. Sadeghi, "A novel technique for selecting mother wavelet function using an intelli gent fault diagnosis system," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4862-4875, 2009.
doi: 10.1016/j.eswa.2008.05.052

[13] J. Rafiee, M. A. Rafiee, and P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4568-4579, 2010.
doi: 10.1016/j.eswa.2009.12.051

[14] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithm," *Mechanical Systems and Signal Processing*, vol. 18, no. 3, pp. 625-644, 2004.
doi: 10.1016/S0888-3270(03)00020-7

[15] Y. Lei, Z. He, and Y. Zi, "Application of an intelligent classification method to mechanical fault diagnosis," *Expert Systems with Applications*, vol. 36, no. 6, pp. 9941-9948, 2009.
doi: 10.1016/j.eswa.2009.01.065

[16] L. Jack and A. Nandi, "Fault detection using support vector machines and artificial neural networks,

Copyright © 2017 International Journal of Automation and Smart Technology

augmented by genetic algorithms," *Mechanical systems and signal processing*, vol. 16, no. 2, pp. 373-390, 2002.
doi: 10.1006/mssp.2001.1454

[17] L. Debnath and F. A. Shah, *Wavelet transforms and their applications,* Boston: Birkhäuser Basel, 2002.

[18] A. Djebala, N. Ouelaa, and N. Hamzaoui, "Detection of rolling bearing defects using discrete wavelet analysis," *Meccanica*, vol. 43, no. 3, pp. 339-348, 2008.
doi: 10.1007/s11012-007-9098-y

[19] B. Liu, "Selection of wavelet packet basis for rotating machinery fault diagnosis," *Journal of Sound and Vibration*, vol. 284, no. 3-5, pp. 567-582, 2005.
doi: 10.1016/j.jsv.2004.06.047

[20] J. C. Garcia-Prada, C. Castejon, and O. Lara, "Incipient bearing fault diagnosis using DWT for feature extraction," in proceeding of *12th IFToMM*, Besançon, France, June 18-21, 2007.
doi: 10.1016/j.eswa.2009.11.006

[21] J. A. Flores, *Focus on artificial neural networks,* New York: Nova Science Publishers, 2011.

[22] K. L. Priddy and P. E. Keller, *Artificial neural networks: an introduction,* Bellingham: Washington SPIE Publications, 2005.

[23] D. E. Goldberg, *Genetic algorithms in search optimization and machine learning*, Boston: Addison-Wesley Longman Publishing Co, 1989.

[24] F. Herrera, "Genetic fuzzy systems: Status, critical considerations and future directions," *International Journal of Computational Intelligence Research*, vol. 1, no. 1, pp. 59-67.
doi: 10.5019/j.ijcir.2005.23

[25] I. G. Damousis, K. J. Satsios, D. P. Labridis, and P. S. Dokopoulos, "Combined fuzzy logic and genetic algorithm techniques—application to an electromagnetic field proble," *Fuzzy sets and systems*, vol. 129, no. 3, pp. 371-386, 2002.
doi: 10.1016/S0165-0114(01)00137-3

[26] J, Kennedy and R. Eberhart, "Particle swarm optimization," in proceeding of *IEEE International Conference on Neural Networks*, Perth, WA, Australia, Nov. 27-Dec. 1, 1995, pp. 1942-1948.
doi: 10.1109/ICNN.1995.488968

[27] Y, Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in proceeding of *Evolutionary Computation*, Washington, DC, USA, July 6-9, 1999, pp. 1945-1950.
doi: 10.109/CEC.1999.785511

[28] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in proceeding of *Nature & Biologically Inspired Computing,* Coimbatore, India, Dec. 9-11, 2009, pp. 210-214.
doi: 10.1109/NABIC.2009.5393690

Copyright © 2017 International Journal of Automation and Smart Technology