A Bearing Fault Diagnosis Method Based on Autoencoder and Particle Swarm Optimization – Support Vector Machine

Duy-Tang Hoang[1] and Hee-Jun Kang[2(✉)]

[1] Graduate School of Electrical Engineering,
University of Ulsan, Ulsan, South Korea
hoang.duy.tang@gmail.com
[2] School of Electrical Engineering, University of Ulsan, Ulsan, South Korea
hjkang@ulsan.ac.kr

Abstract. Rolling element bearing is an important part of rotary machines. Bearing fault is a big issue because it can cause huge cost of time and money for fixing broken machines. Thus, early detecting fault of bearing is a critical task in machine health monitoring. This paper presents an automatic fault diagnosis of bearing based on the feature extraction using Wavelet Packet Analysis, feature selection using Autoencoder, and feature classification using Particle Swarm Optimization - Support Vector Machine. First, bearing vibration signals are decomposed at different depth levels by Wavelet Packet Analysis. Then the wavelet packet coefficients are used to compute the energy value of the corresponding wavelet packet node. After that, an Autoencoder is exploited to select the most sensitive features from the feature set. Finally, classification is done by using a Support Vector Machine classifier whose parameters are optimized by Particle Swarm Optimization. The effectiveness of the proposed intelligent fault diagnosis scheme is validated by experiments with bearing data of Case Western Reserve University bearing data center.

Keywords: Artificial neural network · Autoencoder · Bearing fault diagnosis
Particle Swarm Optimization · Support Vector Machine
Wavelet Packet Analysis

1 Introduction

Rolling Element Bearing (REB) is a critical component of rotary machine. Healthy operation of bearings is the necessary factor for rotary machines to operate efficiently. REBs are account for almost 45−55% of machine failures [1], which can cost a huge amount of time and money for maintaining. As a result, it is very important to early detect the existing faults in bearings. It will help to cut down costs necessary for emergency maintenance, replacement and delay production.

Due to the ease of measurement and the ability to provide a lot of dynamic information reflecting the condition of the mechanical systems, vibration signal-based fault diagnosis has been widely applied in machine health monitoring [1–4]. The abnormal vibration is the first clue of rotary component failure. Vibration analyzing can

detect all types of faults, either localized or distributed. Furthermore, low-cost sensors, accurate results, simple setups, specific information on the damage location, and comparable rates of damage are other benefits of the vibration analysis method [5]. The condition monitoring of a rolling element bearing based on vibration signals can be considered as a pattern recognition problem which has been successfully applying intelligent diagnosis methods. Generally, a general intelligent diagnosis methodology includes three steps as follows: feature extraction, feature reduction, and feature classification.

Feature extraction is the first step mapping the original fault signals onto statistical parameters which reflect the working status of machines. Feature extraction requires expert knowledge, human labor, and signal processing techniques. To extract representative features indicating health conditions of bearings, vibration signals can be analyzed in time domain [2], frequency domain [3], or time-frequency domain [6, 7].

The output of feature extraction is a feature set which consists of features reflecting the characteristic of signals. Normally, the feature set has high dimensionality with a lot of features. The high dimensional feature set often reduces the classification accuracy of fault diagnosis system. Moreover, there is no guarantee that all features are equally usefully in reflecting the health status of bearings [8]. As a result, feature selection is often exploited to reduce the dimensionality of the feature set.

Two approaches are available to perform dimensionality reduction are feature selection and feature extraction. In feature selection approach, a subset of all the features is selected without the transformation. On the other hand, in feature extraction approach, a new feature set is created by the transformation of the existing features. The most well-known algorithm in this approach is principal component analysis (PCA) and its derivatives.

Recently, unsupervised learning autoencoder (AE) has been applied in signal-based fault diagnosis [9, 10]. Hongmei Liu et al. [9] used a stacked autoencoder which formed by stacking several AEs, to extract fault features. Three AEs are stacked together to extract fault features in frequency domain from a spectrogram of vibration signals. This approach obtained high performance but too complex and still missing features in time domain. Tao et al. [10] also used AE with many layers, but directly on raw vibration signal and can only extracted time domain feature.

In feature classification step, after the high sensitive feature set is determined, machine learning based classifiers are employed to detect the health condition of bearings. Among current classification algorithms, artificial neural network (ANN) has proved to be a powerful tool with high accuracy fault detection. However, ANN is not suitable for handling algorithm with few samples for training [11]. Moreover, ANN has some drawbacks, including generalization ability, and slow convergence [8]. Support Vector Machine (SVM) is a supervised machine learning algorithm which based on statistical learning theory. Compared to ANN, the advantages of SVM are (1) better generalization, and (2) does not require many samples for training.

Particle Swarm Optimization (PSO), inspired by the bird swarm behavior of preying on food, emerged as a powerful optimization technique. In PSO, the set of candidate solutions to the optimization problem is defined as a swarm of particles which may flow through the parameter space defining trajectories which are driven by their own and neighbors' best performances [12]. The concept and implementation of

PSO are simple so it is widely applied in many optimization problems, including optimization of SVM parameters. The parameter selection method based on PSO can not only ensure the learning ability of SVM but also improve the generalization ability and accordingly improve the integrated performance of SVM classification [13].

In this paper, a novel bearing fault diagnosis method is proposed. First, bearing vibration signals are analyzed in time-frequency domain by Wavelet Packet Analysis, the energy features from nodes of wavelet packets at different decomposition levels are calculated to characterize the health conditions of the bearing. After that, an AE is applied to reduce the dimensionality of the feature set, only select the most sensitive features. Finally, for feature classification, a SVM classifier whose parameters are optimized by PSO is exploited. To evaluate the effectiveness of the proposed scheme, experiments are carried out with bearing data from Case Western Reserve University (CWRU) bearing data center [14]. Moreover, the proposed method is also investigated with the signals under various low signal-to-noise ratio (SNR) conditions to show the effectiveness under noisy environment.

The remainder of this paper is organized as follows. Section 2 presents brief overviews of AE and PSO. Section 3 explains the proposed algorithm in detail. Section 4 described the experiments. Finally, in Sect. 5 we conclude the paper.

2 Background

2.1 Autoencoder

AE was first introduced by Rumelhart et al. [15] as an unsupervised machine learning algorithm. As shown in Fig. 1, an AE is a feed-forward neural network with three layers: input layer, hidden layer, and output layer where the input and output layer have the same size. The structure of an AE can be considered as an encoder followed by a decoder. The encoder includes the input layer and hidden layer, try to map the input vector to the hidden layer. In opposite, the decoder tries to reconstruct the input vector from the hidden layer vector. Taking an input vector x^i, the computation of AE includes two steps: encoding and decoding as follow:

$$a^i = f\left(W_e x^i + b_e\right) \tag{1}$$

$$\hat{x} = f\left(W_d a^i + b_d\right) \tag{2}$$

where W_e and b_e are respectively the weight matrix and the bias vector of the encoder; W_d and b_d are respectively the weight matrix and the bias vector of the decoder; $f(.)$ denotes the activate function.

With an input set includes m samples $\{x^i, i = 1 : m\}$, the AE will produce m output samples $\{\hat{x}^i, i = 1 : m\}$. Since the goal of the AE is to make the reconstructed vector $\hat{x}$ as close as possible to the input x, the cost function can be defined as:

$$J(W_e, W_d, b_e, b_d) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{x}^i - x^i \right)^2 \tag{3}$$
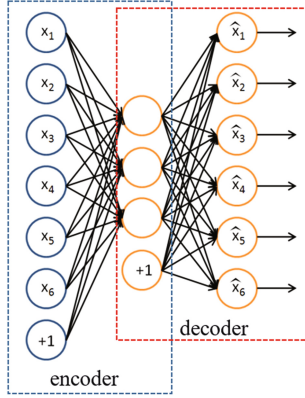

Fig. 1. Autoencoder

2.2   Particle Swarm Optimization

PSO introduced by Kennedy et al. [16] was inspired by the movement of organisms in a bird flock or fish school. The goal of PSO is finding out the solution for the optimization problem. Each candidate solution is called a particle and represents a point in a d-dimensional space, d is the variable number of the function need to be optimized. PSO performs searches using a population (swarm) of particles that are updated their characteristic after each step of movement. The principle algorithm of PSO can be explained as follows [12].

Assume that a swarm includes n particles $X : \{x_1, x_2, \dots x_n\}$. Each particle is characterized by three properties:

- Position $x_i = \begin{bmatrix} x_i^1 & x_i^2 & \dots & x_i^d \end{bmatrix}$
- Velocity $v_i = \begin{bmatrix} v_i^1 & v_i^2 & \dots & v_i^d \end{bmatrix}$
- Personal best p_i the position of the best solution obtained so far by that particle

In each iteration, each particle moves to a new position, the movement in the searching space of a specific particle is computed by:

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{4}$$

where t and $t+1$ indicate two successive iterations. The velocity vectors govern the way particles move across the search space and are made of the contribution of three terms: the first one, defined the inertia or momentum prevents the particle from drastically changing direction, by keeping track of the previous flow direction; the second

term, called the cognitive component, accounts for the tendency of particles to return to their own previously found best positions; the last one, named the social component, identifies the propensity of a particle to move towards the best position of the whole swarm (or of a local neighborhood of the particle, depending on whether a global or partial PSO is implemented). Based on these considerations, the velocity of particles is defined as:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 (p_i - x_i(t)) + c_2 r_2 (g - x_i(t)) \tag{5}$$

In the update velocity equation, each of term has its own distinct roles. The first term $\omega v_i(t)$ is the inertia component which keeps the particle move in the old direction. The value of the inertial coefficient ω is often chosen in the range $[0.8, 1.2]$, which can either dampen or accelerate the particle in its original direction.

The second term $c_1 r_1 (p_i - x_i(t))$, called the cognitive component, acts as the particle's memory, make it has the tendency to return to the position of its best fitness. The third term $c_2 r_2 (g - x_i(t))$, called the social component, causes the particle to move to the best position (g) the swarm has found until current step. The cognitive coefficient c_1 and social coefficient c_2 affect the size of the movement step of particles in each iteration. They are usually chosen close to 2. On the other hand, r_1 and r_2 are diagonal matrices of random numbers generated from a uniform distribution in $[0.1]$ cause the corresponding components to have a stochastic influence on the velocity update. Accordingly, the trajectories drawn by the particles are semi-random in nature, as they derive from the contribution of systematic attraction towards the personal and global best solutions and stochastic weighting of these two acceleration terms.

Consider the optimization problem:

$$\begin{aligned} &Given\ f : R^D \rightarrow R \\ &Find\ x_{opt}\ \vee f\left(x_{opt}\right) \geq f(x)\,\forall x \in R^D \end{aligned} \tag{6}$$

The problem is solved by PSO algorithm described step by step as follows.

Step 1. Initialization: For each of the n particles:

1. Randomly initialize the position $x_i(0)$
2. Initialize the particle's best position by its initial position $p_i = x_i(0)$
3. Calculate the fitness of each particle $f(x_i(0))$
4. Initialize the global best by g by the position of particle which has the best fitness

Step 2. Until stop condition is met, repeat:

5. Update velocity by Eq. (2)
6. Update position by Eq. (1)
7. If $f(x_i(t+1)) \geq f(p_i)$, then $p_i = x_i(t+1)$
8. If $f(x_i(t+1)) \geq f(g)$, then $g = x_i(t+1)$

When the stop condition is satisfied, the iterative process stops, the best solution is presented by global best value g.

3   Proposed Bearing Fault Diagnosis Method

The proposed new bearing fault diagnosis scheme has 3 main steps: feature extraction, feature selection, and feature classification. Each step is described in detail as follows.

Step 1: Feature Extraction

At this step, vibration signals are decomposed into various levels by Wavelet Packet Analysis (WPA). All nodes at all levels are considered because it is difficult to declare definitively that those at a certain depth are better than those at another. Subsequently, we calculate wavelet packet energy for every node and use those as time-frequency features of the signal.

Consider a time-domain vibration signal $x(t)$ consists of S sample. By decomposing $x(t)$ into N levels, we obtain the result of the decomposition as follows. Energy of each node in the decomposition is computed by [17]:

$$E_j^n = \sqrt{\frac{1}{M} \sum_{t=1}^{M} \left(x_j^n(t) \right)^2} \qquad (7)$$

where $M = \frac{S}{2^j}$ is the number of samples at the node x_j^n. Since the decomposition has $\sum_i^N 2^i$ nodes, we obtain the corresponding number of node energy features.

Step 2: Feature Selection

In this second step, an AE is exploited to reduce the dimensionality by mapping the energy features into high sensitive features. The energy feature set from step 1 is considered as the input of the AE. The training process using backpropagation with the cost function defined in (6). After training, the output of hidden layer of the SA becomes the input for classification step.

Step 3: Feature Classification

In this step, a SVM is employed to classify the selected features. The classification accuracy of the SVM is affected by three factors: kernel function, kernel parameters, and penalty parameter. PSO algorithm is used to find those parameters of the SVM. After finding out the SVM model with best parameters, the feature set is given into SVM to recognize the type of faults.

4   Experiment

4.1   Test-Bed

To evaluate the effectiveness of the proposed method, experiments are carried out with data of faulty bearing from the Case Western Reserve University (CWRU) bearing data center. This dataset is public and widely used to validate the effectiveness of bearing fault diagnosis algorithms. The test-bed shown in Fig. 2 includes a dynamometer (right), a 2 HP motor (left), and a torque transducer/encoder (center). The test-bed also consists of a control electronics but not shown in the figure. The motor shaft is supported by the test bearings. Single point faults were introduced to these bearings using electro-discharge machining with fault diameters of 7 mils, (1 mil = 0.001 inches). Vibration data are collected by using accelerometers, which are attached to the housing with magnetic bases. Accelerometers are placed at the 12 o'clock position at both the drive housing. Vibration signals are collected using a 16 channel DAT recorder, including three operating conditions: fault at ball, fault at inner race, and fault at outer race. These operating conditions are operated with bearings 6205-2RS JEM SKM, which are deep groove ball bearing type. All experiments are conducted for one load condition (2 HP load), where the rotation speed was 1797 revolutions per minute (rmp). Data were collected at 12 kHz sampling frequency from both drive end (DE) and fan end (FE). Figure 3 shows the vibration signals of three operating conditions.
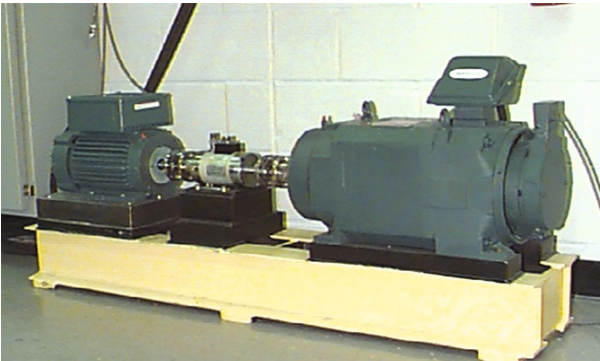

Fig. 2.   Bearing fault diagnosis test-bed

4.2   Vibration Signal Pre-processing

Vibration signals are from both fan end and drive end of the motor shaft at sample rate 12000 Hz. To have enough samples for the training process of the machine learning based classifier, at first, each vibration signal of each health condition is split into non-overlapping segments with the same length. For every condition, 100 samples are acquired, so totally we obtain 300 sample for three bearing health conditions.
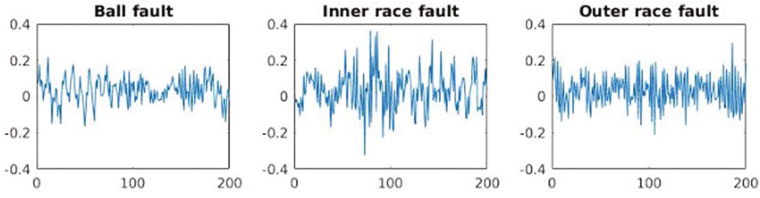
Fig. 3. Vibration signals

4.3 Feature Extraction and Dimensionality Reduction

In the next step, we compute the energy features using WPA. The mother wavelet is Daubechies 4, each segment of the signal is decomposed to 4 level. For each sample, we have $\sum_i^4 2^i = 30$ nodes corresponding to 30 energy features. The energy features are calculated from all 30 nodes of the decomposition tree.

AEs are exploited to extract most sensitive features from the energy feature set in other to reduce the dimension. The input size of every AE is equal to the number of energy features, while the size of the hidden layer can be varied. Our goal when using AEs is to compress the data feature as much as possible. So, we investigate AEs with various hidden layer size from 1 to 5 (neuron). Backpropagation is used to train the AEs. Finishing the training, the hidden vectors are used as compressed feature set.

4.4 PSO-SVM Classification

In our experiments, to classify features, SVM classifiers with RBF kernel are used. The classification performance highly depends on the parameters C, γ. To search the optimum values of those parameters, we exploited the PSO algorithm with parameters as follows.

- Number of particles in population: 6
- Searching dimension: 2 (including two values need to be optimized: C and γ)
- Searching space: $C \in [0, 100]$, $\gamma \in [0, 1]$
- Inertial coefficient, cognitive coefficient, and social coefficient are selected by the method of Clerc et al. [18] proposed as follows.

$$\chi = \frac{2\kappa}{\left| 2 - \phi - \sqrt{\phi^2 - 4\phi} \right|}, \tag{8}$$

$$\omega = \chi, \ c_1 = \chi\phi_1, \ c_2 = \chi\phi_2, \ 0 < \kappa \leq 1, \ \phi_1 + \phi_2 = \phi \geq 4$$

By choosing $\kappa = 1$, $\phi_1 = \phi_2 = 2.05$, the cognitive coefficient and social coefficient are $c_1 = c_2 = 1.496$, the inertial coefficient $\omega = 0.73$.

In the last step – feature classification, we use k-fold ($k = 5$) cross validation for SVM classification with parameters found by the above PSO method. The feature sets are divided into k subsets; each subset is used for once for testing while the remain $k - 1$ subsets are used for training. The final classification accuracy is the average of k

classification times. The results with different feature sets supplied by varying the size of AE hidden layers are shown in Table 1. We can see that even with the 2-feature set (AE with 2 neurons in hidden layer), the classification results still achieve satisfactory performance (99:05%). And from 3 neurons above, the classification accuracy is 100%.

Table 1. Classification results of different AE models

Size of hidden layer in AE	1	2	3	4	5
Classification accuracy (%)	93.39	99.05	100	100	100

4.5 Robustness Investigation

The proposed scheme can classify bearing fault efficiently with original vibration signal. However, in real industrial environments, the sensory signals are contaminated by noise [7]. Thus, now we analyze the robustness of the proposed scheme under low signal-to-noise ratio (SNR) condition. The additive Gaussian white noise (AWGN) with different standard variances are added to the original vibration signals to mimic the low SNR. The SNR is defined as follows:

where P_{signal} and P_{noise} are the power signal and noise respectively. Figure 4 shows the noised signal which made by adding the original signal with the AGWN.
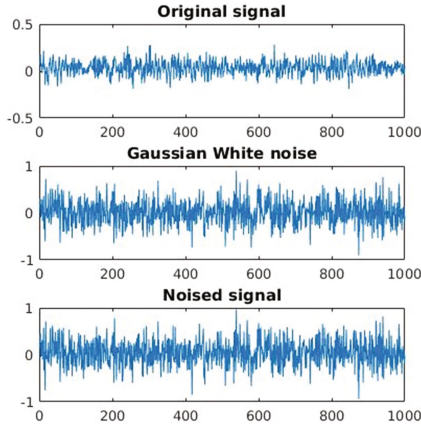
Fig. 4. A noisy signal with SNR $= -10\,\mathrm{dB}$

Table 2 shows the fault classification accuracy of the proposed scheme with SNR value varies in $[-2, -4, -6, -8, -10]$. Results show that even under very low SNR $(-6\,\mathrm{dB})$, the classifier is still capable to achieve absolute accuracy. Under the worst case (SNR $= -10\,\mathrm{dB}$), classification accuracy is 93.34% .

Table 3 shows the result of comparison between the proposed method with some other techniques mentioned in the publication [7] under the worst scenario

$SNR = -10$ dB. The comparison shows that the proposed scheme yields much superior classification accuracy robustness against noise compared to other methods.

Table 2. Classification results with different SNR values

SNR (dB)	−2	−4	−6	−8	−10
Classification accuracy (%)	100	100	100	98	94.34

Table 3. Classification results of different methods ($SNR = -10$ dB)

Scheme	Accuracy (%)
Malhi [3]	38.09
Lou [6]	33.33
Seker [19]	78.25
Li [20]	86
Yaqub [7]	91.23
Proposed method	94.34

5   Conclusion

In this paper, a feature reduction method by using AE was proposed and successfully applied in bearing fault diagnosis. At first, vibration signals are decomposed by WPA. Dual time-frequency domain features are extracted by computing the energy of every node in the decomposing tree. Unsupervised learning is applied to train simple AE with only 1 hidden layer to extract most sensitive features. Finally, features are classified by RBF kernel SVM with optimal parameters are optimized by PSO algorithm. Our proposed method can achieve very high accuracy and robustness even under very poor SNR condition. The effectiveness is also presented through comparisons with other existing fault diagnosis methods.

References

1. Rai, A., Upadhyay, S.: A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. Tribol. Int. **96**, 289–306 (2016)
2. Samanta, B., Al-Balushi, K.: Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. Mech. Syst. Sig. Process. **17**(2), 317–328 (2003)
3. Malhi, A., Gao, R.X.: Pca-based feature selection scheme for machine defect classification. IEEE Trans. Instrum. Measur. **53**(6), 1517–1525 (2004)

4. Yen, G.G., Lin, K.-C.: Wavelet packet feature extraction for vibration monitoring. IEEE Trans. Ind. Electron. **47**(3), 650–667 (2000)

5. Kharche, P.P., Kshirsagar, S.V.: Review of fault detection in rolling element bearing. Int. J. Innovative Res. Adv. Eng. **1**(5), 169–174 (2014)

6. Lou, X., Loparo, K.A.: Bearing fault diagnosis based on wavelet transform and fuzzy inference. Mech. Syst. Sig. Process. **18**(5), 1077–1095 (2004)

7. Yaqub, M.F., Gondal, I., Kamruzzaman, J.: Inchoate fault detection framework: Adaptive selection of wavelet nodes and cumulant orders. IEEE Trans. Instrum. Measur. **61**(3), 685–695 (2012)

8. Shen, C., Wang, D., Kong, F., Peter, W.T.: Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. Measurement **46**(4), 1551–1564 (2013)

9. Liu, Z., Cao, H., Chen, X., He, Z., Shen, Z.: Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. Neurocomputing **99**, 399–410 (2013)

10. Tao, S., Zhang, T., Yang, J., Wang, X., Lu, W.: Bearing fault diagnosis method based on stacked autoencoder and softmax regression. In: 2015 34th Chinese Control Conference (CCC), pp. 6331–6335. IEEE (2015)

11. Widodo, A., Yang, B.-S.: Support vector machine in machine condition monitoring and fault diagnosis. Mech. Syst. Sig. Process. **21**(6), 2560–2574 (2007)

12. Marini, F., Walczak, B.: Particle swarm optimization (PSO). a tutorial. Chemometr. Intell. Lab. Syst. **149**, 153–165 (2015)

13. Zhang, X., Guo, Y.: Optimization of SVM parameters based on PSO algorithm. In: Proceedings of Fifth International Conference on Natural Computation. ICNC 2009, vol. 1, pp. 536–539. IEEE (2009)

14. Loparo, K.A.: Bearing data center, Case Western Reserve University (2013)

15. Rumelhart, D.E., Hinton, G.E., Williams, R.J., et al.: Learning representations by back-propagating errors. Cogn. Model. **5**(3), 1 (1988)

16. Kennedy, J.: Particle swarm optimization. In: Encyclopedia of Machine Learning. Springer, Heidelberg, pp. 760–766 (2011)

17. Zarei, J., Poshtan, J.: Bearing fault detection using wavelet packet transform of induction motor stator current. Tribol. Int. **40**(5), 763–769 (2007)

18. Clerc, M., Kennedy, J.: The particle swarm explosion, stability, and convergence in a multidimensional complex space. IEEE Trans. Evol. Comput. **6**(1), 58–73 (2002)

19. Seker, S., Ayaz, E.: Feature extraction related to bearing damage in electric motors by wavelet analysis. J. Franklin Inst. **340**(2), 125–134 (2003)

20. Li, F., Meng, G., Ye, L., Chen, P.: Wavelet transform-based higher-order statistics for fault diagnosis in rolling element bearings. J. Vibr. Control **14**(11), 1691–1709 (2008)