An integrated condition monitoring method for rolling element bearings based on perceptual vibration hashing and SOM

Shiwei Yan¹, Haining Liu*, Fajia Li¹, Huanyong Cui¹,
¹School of Mechanical Engineering, University of Jinan, Jinan, China

*Corresponding author: me_liuhn@ujn.edu.cn

Abstract. Rolling element bearings are widely used in rotating machinery. Bearing faults will result in damage to property. So, the condition monitoring of bearings is of great significance, but few methods can achieve both degradation assessment and fault diagnosis. In this paper, an integrated condition monitoring method for rolling element bearings based on perceptual vibration hashing (PVH) and self-organizing maps (SOM) is proposed. Distance matrix based on PVH is used as a health indicator for degradation assessment, in which the baseline of healthy state is selected based on the clustering centre of SOM instead of experience. When the value of health indicator exceeds the pre-set threshold, visualized fault diagnosis can also be achieved by training the SOM network. The effectiveness of the developed method is verified with the vibration data from accelerated degradation tests of rolling element bearings.

Keywords: Bearing fault diagnosis; bearing degradation assessment; Perceptual Vibration Hashing; Self-organizing maps.

1. Introduction

The rolling element bearings are one of the most widely used parts in rotating machinery. According to statistics, most failures of the rotating machinery are caused by bearing faults, which leads to loss of property. So, the degradation assessment and accurate identification of the bearing fault types are of great significance to ensure the safe operation of mechanical equipment and avoid catastrophic accidents.

A feature is a compact representation containing the information of the machine condition. Statistical parameters in the time and frequency domain such as root mean square (RMS), kurtosis, and crest factor are widely used for condition monitoring. Time-frequency joint analysis methods such as empirical mode decomposition (EMD) [1] and wavelet-based methods [2] are widely studied and used. However, one feature can only represent one aspect of the bearing health state. It might result in the loss of some important information so that the effectiveness of the feature extraction is not satisfactory. Perceptual vibration hashing by sub-band coding is proposed [3]. The extracted sub-band features can effectively reveal the frequency distribution pattern in the vibration signals, which are further compacted and symbolically represented by condition hashes. It can represent the machine condition comprehensively and improve the computational speed. The distance matric against healthy machine condition hash (MCH) is obtained which can be used for degradation assessment.

In recent years, with the gradual maturity of the artificial intelligence (AI) technology, using the AI to fault diagnosis of bearings has become a research hotspot. BP neural network with its improved
algorithm is an extensively researched method [4,5], but it is a supervised learning method which need advance knowledge. Its learning speed slows with the increase of the data dimension. Self-organizing maps (SOM) is a kind of neural network with simple structure and good classification effect, which is an unsupervised learning method with high learning speed. SOM does not need too much experience knowledge to achieve a good classification effect. It also has the advantages of efficient data visualization.

Based on above analysis, it can be seen that the degradation assessment and fault diagnosis of rolling element bearings is of great significance, but few methods can achieve two processes simultaneously. In this paper, an integrated condition monitoring method for rolling element bearings based on PVH and SOM is proposed. Distance matrix based on PVH is obtained for degradation assessment, in which the healthy state is selected by the clustering centre of SOM network instead of experience. When the health indicator exceeds the pre-set threshold, a visualized bearing fault diagnosis method based on PVH and SOM is introduced. The effectiveness of proposed methods is validated through the experiment.

2. Theoretical Foundation

2.1. Perceptual Vibration Hashing by Sub-band Coding

The purpose of the perceptual vibration hashing is to transmit the vibration signal into MCHs, which can represent the equipment condition in a compact form. The process is shown below:

1) Sub-band division by wavelet packet transform (WPT) [6].
2) Blocking: The sub-band signal is truncated into several blocks so that a sub-band matrix can be made for two-dimensional discrete cosine transform (2D-DCT) [7].
3) 2D-DCT: 2D-DCT is implemented on each frequency band for decorrelation and then a coefficient matrix is obtained.
4) Feature extraction: A sub-band feature vector for raw signal is obtained by cascading feature vectors extracted from each sub-band coefficient together.
5) Hashing with Symbolic Aggregate Approximation (SAX) [8]: The MCH which is the symbolic representation of feature vectors is obtained through SAX.

Finally, a piece of a 1024-point vibration signal can be transformed into a compact MCH with a few numbers. A distance function can be used for results of SAX to quantize the similarity of MCHs. A more detailed description of it can be found in [3].

2.2. Self-organizing Map Network

SOM proposed by Kohonen [9] is a kind of artificial neural network (ANN) which is suitable for data clustering. It is trained using unsupervised competitive learning, which can produce a low-dimensional space to represent the high-dimensional input data. The training process of it is show below.

1) Initialize all weights $w_{ij}$ with small random values. Set the initial topological neighborhood, initial learning rate, total number of iterations, and the number of the nodes in competitive layer.
2) Choose an input vector $x$ randomly from the training data.
3) Calculate the distance between each neuron and the input vector so that the neuron $j$ called the winner neuron or best matching unit (BMU) is determined, of which weight vector $w_j$ is closest to the input vector. The closest distance $D_{\text{min}}$ is given:

$$D_{\text{min}} = \{D_j\} = \min\{(x - w_j)^2\}$$

(1)

4) Update the weight vectors of the BMU and its neighbors to make the nodes more like the input vector. The closer the node and BMU are, the more the weight is adjusted.
5) Repeat from the step 2 until the specified number of iterations.
3. An Integrated Condition Monitoring Method Based on PVH and SOM

3.1. Selection of Optimum Healthy State Based on SOM

For a mechanical system, the beginning of the operation indicates that the machinery is new but the machine condition at the beginning cannot represent the optimum healthy state. Due to the advantages of visualization and unsupervised clustering, SOM can present the data distribution clearly in the U-Matrix and weight space. In order to select an optimum healthy state as the baseline in distance metric, a new method is proposed below:

1) Raw vibration signals are processed by PVH to get MCHs.
2) MCHs are used to train SOM neural network and the weight coefficient of neurons can be obtained from the trained SOM neural network. The mean of distance sum between one neuron and others can be calculated and then the neuron located in the cluster centre with minimum mean distance can be obtained.
3) Calculate the distance between samples and the central neuron. Then, the sample with minimum distance can be obtained, which represent the optimum healthy state.
4) When new vibration signals are obtained, repeat from step 1 to achieve the update of optimum healthy state selection.

![Figure 1. Flowchart of the proposed integrated condition monitoring method.](image)

3.2. Degradation Assessment

For the degradation assessment, as is shown in figure 1, the distance matric based on PVH is set as the health indicator. Firstly, raw vibration signals are transmitted into compact MCHs through PVH. Secondly, select an optimum healthy state using the method in section 3.1 as the baseline in distance metric. Finally, a distance matric against a condition hash in optimum healthy state is obtained which can provide a quantitative assessment of degradation. When the value of this health indicator exceeds the pre-set threshold, a bearing fault diagnosis method based on PVH and SOM is used for diagnosis.

3.3. Fault Diagnosis

The proposed diagnosis method is shown in figure 1, firstly, the raw vibration signals containing various bearing running states are initially selected. Secondly, the selected raw signals are processed by PVH to get MCHs. Thirdly, the MCHs are divided into training data and testing data randomly to be input feature vector of the SOM. After the input feature vectors are normalized, the SOM is trained iteratively. The fault type of the bearings can be described by the matched regions in the SOM. Finally, when testing data is input the trained SOM, the fault types of bearings can be recognized automatically, which can be visualized clearly though the hit units on the map. When early degradation occurs in distance curve, corresponding MCH is put into the trained SOM and then the fault diagnosis can be achieved.
4. Experimental Results

4.1. Run-To-Failure Bearing Vibration Data
In order to validate the effectiveness of the proposed approach, the vibration signal data is obtained from the XJTU-SY bearing datasets [10] which provides the run-to-failure data of 15 rolling element bearings. Table 1 shows the information of the dataset used in this paper. The type of the bearings is LDK UER204. In the process of the experiment, the sampling frequency is 25600 Hz and the rotating speed is 2250 r/min under the load of 11 kN. The dataset contains four types of the bearing health condition including the normal, inner race fault, outer race fault and cage fault.

Table 1. Description of the bearing dataset.

| Bearing type | Fault type         | Dataset   | Rotating speed (r/min) | Working condition (kN) |
|--------------|--------------------|-----------|------------------------|------------------------|
| LDK UER204   | Normal             | Bearing 2_1 | 2250                  | 11                     |
| LDK UER204   | Inner race fault   | Bearing 2_1 | 2250                  | 11                     |
| LDK UER204   | Outer race fault   | Bearing 2_5 | 2250                  | 11                     |
| LDK UER204   | Cage fault         | Bearing 2_3 | 2250                  | 11                     |

4.2. Degradation Assessment with Distance Curves
To show the effectiveness of the proposed method in degradation assessment, 15712 pieces of MCHs were obtained from Bearing 2_1. Then the distance curves calculated based on different MCHs as baselines were obtained. The comparison of different distance curves is shown in figure 2. Figure 2(a) shows the distance curve of the whole life calculated based on the first MCH as the baseline of the healthy bearing and figure 2(b) presents the distance curve calculated based on the MCH selected by the proposed method in section 3.1.

![Figure 2](image-url) Comparison of distance curves

Through the comparison of the distance curves, it can be seen that the decreasing tendency at the beginning in figure 2(b) can indicate the running-in period of the new machinery. Then, the distance curve remains at a relatively stable value, which indicates that the bearing is in a healthy state before the degradation. After the early degradation, the “healing” phenomenon is presented clearly and the increasing tendency is stronger compared with figure 2(a). When the value of distance exceeds a pre-set threshold, corresponding MCHs can be used as input of the trained SOM to achieve fault diagnosis.

4.3. Bearing Fault Diagnosis

4.3.1. Data Processing based on PVH. 300 rows of 1024-point vibration signal of each health condition were obtained and 1200 rows were got in total. In the process of PVH, the bearing dataset of each health condition is transformed respectively into 300 samples with 128-di­mensional sub-band feature vectors and the sub-band features were hashed with SAX. The hashing procedure is shown in figure 3, in which 16 segments are used for PAA and a number size of 8 is used for symbolic representation. For example,
a piece of vibration signal under normal condition in figure 3 is transformed into the MCH \{787433447454222\}. Finally, 300 pieces of MCH under each health condition were obtained.

![Figure 3. Illustration of SAX hashing](image)

4.3.2. *Fault diagnosis using SOM.* 5-fold cross-validation was used to generate training data and testing data. The processed dataset which was arranged at random was divided into 5 folds. For each training process, 4 folds which included 960 samples containing 4 types of health states were selected as training data and 1 fold including 240 samples was left as testing data.

![Figure 4. Training and testing results.](image)

The size of the competitive layer was 14×11. After repeated training, as is shown in figure 4, the neuron labels of clustering results on training data are presented by various letters in the hexagons where the “N, I, O, C” respectively represent the normal, inner race fault, outer race fault and cage fault.

| Fault type       | Accuracy of model 1 (%) | Accuracy of model 2 (%) | Accuracy of model 3 (%) | Accuracy of model 4 (%) | Accuracy of model 5 (%) |
|------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| Normal           | 96.67                   | 93.33                   | 93.33                   | 95                      | 93.33                   |
| Inner race fault | 98.33                   | 91.67                   | 96.67                   | 95                      | 93.33                   |
| Outer race fault | 100                     | 100                     | 100                     | 96.67                   | 100                     |
| Cage fault       | 96.67                   | 100                     | 100                     | 100                     | 98.33                   |
| Average          | 97.92                   | 96.25                   | 97.50                   | 96.67                   | 96.25                   |

After the training of the network was finished, to diagnose the bearing faults and validate the accuracy of the model, testing data was put into the network. Bearing health states could be classified automatically. Diagnosis results are shown in figure 4 where red, green, blue and yellow respectively
represent the normal, inner race fault, outer race fault and cage fault. Most of the faults were clustering in the true area. The accuracy of the SOM models under 5-fold cross-validation is calculated which is shown in table 2.

The SOM accuracy of different bearing faults is listed and an overall diagnosis accuracy can be calculated which is 96.92%. The results show that the SOM neural network can achieve a good performance on classification and visualization.

5. Conclusions
This paper presents an integrated condition monitoring method for rolling element bearings based on PVH and SOM, which can achieve both degradation assessment and fault diagnosis. Dynamic degradation assessment can be tracked with the distance curve of MCHs, which is calculated based on an optimum healthy state as a baseline. After early degradation occurs, fault diagnosis starts directly by putting corresponding MCHs into the trained SOM, which can achieve an excellent effectiveness on fault classification and visualization. Convenience of the whole process of degradation assessment and fault diagnosis is improved by the proposed integrated condition monitoring method.

Acknowledgments
This work was supported in part by the National Natural Science Foundation of China under Grant 51605191, in part by the Shandong Provincial Key Research and Development Program under Grant 2019GGX105009.

References
[1] Lei Y, Lin J, He Z, et al. A review on empirical mode decomposition in fault diagnosis of rotating machinery[J]. Mechanical systems and signal processing, 2013, 35(1-2): 108-126.
[2] Yan R, Gao R X, Chen X. Wavelets for fault diagnosis of rotary machines: A review with applications[J]. Signal processing, 2014, 96: 1-15.
[3] Liu H, Wang Y, Li F, et al. Perceptual vibration hashing by sub-band coding: an edge computing method for condition monitoring[J]. IEEE Access, 2019, 7: 129644-129658.
[4] Wu J, Yang X L. Fuzzy BP neural network and its applications in fault diagnosis[J]. Systems Engineering and Electronics, 2001, 10.
[5] Tang H, Wu Y, Hua G, et al. Internal leakage fault diagnosis of hydraulic cylinder using PCA and BP network[J]. Journal of Central South University of Science and Technology, 2011, 42(12): 3709-3714.
[6] Mallat S G . A Wavelet Tour of Signal Processing: The Sparse Way. China Machine Press, 2009.
[7] Ahmed N, Natarajan T, Rao K R. Discrete cosine transform[J]. IEEE transactions on Computers, 1974, 100(1): 90-93.
[8] Lin J, Keogh E, Lonardi S, et al. A symbolic representation of time series, with implications for streaming algorithms[C]//Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery. 2003: 2-11.
[9] Kohonen T . Self-Organizing Maps[J]. series in information science, 2001.
[10] Wang B , Lei Y , Li N , et al. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings[J]. IEEE Transactions on Reliability, 2018:1-12.