Postprint: Bearing Fault Diagnosis Based on Wavelet Packet-AR Spectrum and Deep Learning

Authors: He Siyan, Liu Ya, Tian Xincheng

Date: 2018-07-09T00:00:00+00:00

Abstract

To address the non-stationary and nonlinear characteristics of bearing fault signals, a method combining wavelet packet decomposition and auto-regressive (AR) spectral estimation is employed to extract vibration signal features. To improve diagnostic accuracy, a Deep Belief Network (DBN) is proposed for diagnostic model training. First, wavelet packet decomposition and auto-regressive spectral estimation are performed on bearing vibration signals, and the energy in different frequency bands is calculated to achieve bearing fault feature extraction. Second, the extracted features are used as input vectors for the Deep Belief Network for model training. Finally, the trained model is utilized for fault diagnosis. To verify the effectiveness of the proposed method, the rotating bearing dataset provided by Case Western Reserve University is employed, and comparative experiments are conducted between the proposed algorithm and three fault diagnosis methods. The experimental results demonstrate that the proposed method exhibits superior diagnostic performance.

Full Text

Preamble

Title: Research on Bearing Fault Diagnosis Based on Wavelet Packet-AR Spectrum and Deep Learning

Authors: He Siyan[1], Liu Ya[2]†, Tian Xincheng[2]
[1] Department of Intelligent Manufacturing Engineering, Shandong College of Electronic Technology, Jinan 250200, China
[2] School of Control Science & Engineering, Shandong University, Jinan 250061, China

Abstract: Bearing fault signals exhibit non-stationary and non-linear characteristics. This paper employs a combination of wavelet packet decomposition (WPD) and auto-regressive (AR) spectrum estimation for feature extraction from vibration signals. To improve diagnostic accuracy, a deep belief network (DBN) is proposed for diagnostic model training. First, WPD and AR spectrum estimation are applied to bearing vibration signals to calculate the energy in different frequency bands for feature extraction. Second, the extracted features serve as input vectors to the DBN for model training. Finally, the trained model is used for fault diagnosis. To validate the proposed method, experiments are conducted using the rotating bearing dataset from Case Western Reserve University, comparing the algorithm with three other fault diagnosis methods. Experimental results demonstrate that the proposed method achieves superior diagnostic performance.

Keywords: wavelet packet decomposition; feature extraction; deep belief network; fault diagnosis

0 Introduction

Rolling bearings are among the most widely used mechanical components in electromechanical equipment, and their operational status directly affects overall equipment performance. As one of the most vulnerable components, condition monitoring and fault diagnosis of bearings are particularly critical [1, 2]. Equipment in abnormal states exhibits changes in vibration characteristics, making vibration signal analysis widely applicable in rotating machinery fault diagnosis. Feature extraction based on signal decomposition has been extensively studied [3, 4]. When bearing faults occur, vibration signals often display non-stationary characteristics due to factors such as rigidity, non-linearity, and friction, making Fourier transform inadequate for accurate feature extraction.

Empirical Mode Decomposition (EMD) is theoretically suitable for non-linear non-stationary processes, but its significant "mode mixing" problem creates limitations in obtaining effective information [5]. In recent years, scholars have proposed feature extraction methods based on Ensemble Empirical Mode Decomposition (EEMD), which partially solves the mode mixing problem but still suffers from high computational complexity [6]. Wavelet packet decomposition can perform multi-channel filtering on detection signals, dividing them into different frequency bands through interaction with wavelets at various frequencies, thereby reducing signal interference. Additionally, AR spectrum estimation has extrapolation capabilities that enable effective analysis of short-sample signals [7].

Currently, numerous domestic and international researchers have investigated bearing fault diagnosis. Common fault recognition algorithms include Grey Relational Analysis (GRA), Support Vector Machine (SVM), and Back Propagation Neural Network (BPNN) [8, 9]. While GRA is easy to implement and understand, it suffers from insufficient classification accuracy. SVM is suitable for small-sample training but faces challenges in determining penalty coefficients and requires kernel functions to satisfy Mercer' s theorem. BPNN is prone to falling into local optima.

Deep learning represents a new wave in neural network development. Following AlphaGo' s victory over human Go champions, deep learning has become a focal point in both academia and industry, playing important roles in image processing, audio analysis, and artificial intelligence. In speech recognition, for instance, Deng et al. discovered that pre-training a deep feedforward neural network using a Deep Belief Network (DBN) followed by fine-tuning significantly improved recognition accuracy [10–12]. Considering the similarities between bearing fault diagnosis and speech recognition in pattern classification, this paper applies DBN to bearing fault diagnosis.

A DBN consists of a visible layer and multiple hidden layers, capable of extracting multi-level representations from training data and reconstructing input data. The DBN training process is layer-wise, making it an effective and efficient network in deep learning. However, as data dimensionality increases, the number of samples required for learning grows exponentially, and learning from large datasets demands more memory and processing power [13, 14]. Therefore, this paper performs feature extraction before DBN training. The proposed method combines wavelet packet decomposition-AR spectrum estimation for feature extraction to obtain bearing state information, followed by DBN training for classifier implementation.

---

1 Basic Theory

The proposed fault diagnosis algorithm comprises two main components: feature extraction based on wavelet packet-AR spectrum estimation and fault classification based on a Deep Belief Network. The algorithm flowchart is shown in [Figure 1: see original paper].

Feature extraction constructs new features from raw vibration data that effectively describe the original data while optimizing model performance on unknown data.

1.1 Feature Extraction Based on Wavelet Packet Decomposition and AR Spectrum Estimation

Wavelet packet decomposition separates feature information from interference, providing noise reduction. AR spectrum estimation is applied to the decomposed signals in each frequency band to extract energy features containing bearing state information. The process consists of four steps:

1) **Wavelet Packet Decomposition:** Perform n-layer decomposition on vibration signals to obtain 2^n wavelet packet coefficients.

2) **Band-wise Signal Reconstruction:** Let $X_i(n)$ denote the reconstructed signal for the i-th frequency band. The total signal can be expressed by formula (1):

$$X(n) = \sum_{i=0}^{2^n-1} X_i(n)$$

where n represents the decomposition layer number and 2^n is the number of frequency bands.

3) **AR Spectrum Analysis:** Perform AR spectrum estimation on each reconstructed band signal to obtain AR spectra containing only specific frequency information.

4) **Wavelet Packet-AR Spectrum Band Energy Calculation:** The energy values of 2^n frequency bands form an m-dimensional feature vector characterizing bearing condition.

1.2 Deep Belief Network

The DBN, proposed by Hinton et al. in 2006, is built upon Restricted Boltzmann Machines (RBM) [15, 16]. An RBM consists of a visible layer and a hidden layer with no intra-layer connections but full inter-layer connections, as shown in [Figure 2: see original paper].

As illustrated in [Figure 2: see original paper], the visible layer (v) receives input signals, while the hidden layer (h) acts as a feature extractor. With m neurons in the visible layer (indexed by i) and n neurons in the hidden layer (indexed by j), the probability distribution over hidden and visible units is defined by an energy function:

$$E(v,h) = -\sum_{i=1}^{m}\sum_{j=1}^{n} W_{ij} v_i h_j - \sum_{i=1}^{m} b_i v_i - \sum_{j=1}^{n} c_j h_j$$

where W_{ij} is the connection weight between visible neuron i and hidden neuron j, b_i is the bias of visible neuron i, and c_j is the bias of hidden neuron j.

The RBM is a probabilistic graphical model with joint probability distribution:

$$p(v,h) = \frac{1}{Z}\exp(-E(v,h))$$

where Z is the normalization factor:

$$Z = \sum_{v,h} \exp(-E(v,h))$$

Hidden layer nodes are conditionally independent, enabling maximum likelihood learning through simple alternating updates of all hidden units and visible units in the same layer.

To enhance RBM learning, the Contrastive Divergence (CD) algorithm is employed. Starting from visible units, it updates all hidden units in parallel, reconstructs visible units from hidden units, and finally updates hidden units again. The first step in DBN training uses CD to learn features from visible units. These learned feature activations then become the visible units for learning features in the second hidden layer. This process continues until the final hidden layer is trained, completing the DBN training process as shown in [Figure 3: see original paper].

The DBN learning process, shown in [Figure 3: see original paper], consists of two main phases: pre-training and fine-tuning. Pre-training generates excellent initial parameter values and addresses the slow training speed of large networks through layer-wise RBM training stacked into a DBN [17]. Fine-tuning uses sample feedback to continuously adjust network output toward desired results [18].

---

2 Experimental Validation

2.1 Experimental Data Description

To validate the proposed fault diagnosis method, vibration signals from the Bearing Data Center of Case Western Reserve University' s Electrical Engineering Laboratory are used. The bearing data acquisition system test bench is shown in [Figure 4: see original paper].

The dataset was collected at a 12 kHz sampling frequency under four different load conditions (Load 0 = 0 HP/1797 RPM, Load 1 = 1 HP/1772 RPM, Load 2 = 2 HP/1750 RPM, Load 3 = 3 HP/1730 RPM). The test bench simulates three fault types: inner race fault, outer race fault, and rolling element fault, each with three damage levels of 0.18 mm, 0.36 mm, and 0.54 mm. This study uses signals with 0.18 mm fault size and normal signals to verify early fault diagnosis performance. [Figure 5: see original paper] shows the time-domain waveforms of vibration signals for four bearing states under Load 0, each containing 120,617 data points. Using a window of 1024 points, the raw vibration signals are segmented into 100 samples per bearing state.

As shown in [Figure 5: see original paper], vibration amplitudes are smaller during normal operation. Fault signals exhibit increased amplitude, but some states have similar time-domain waveforms combined with environmental interference, making bearing state differentiation difficult. Therefore, feature extraction is necessary.

2.2 Vibration Signal Feature Extraction

Each load condition contains 400 samples, with each sample comprising 1024 data points. The wavelet packet decomposition-AR spectrum estimation method is applied to each sample, yielding feature vectors composed of 8 frequency band energy values. Partial data are shown in .

shows feature extraction data where fault types 1, 2, 3, and 4 represent inner race fault, outer race fault, rolling element fault, and normal condition, respectively.

To visualize the effectiveness of feature extraction, [Figure 6: see original paper] (left) presents a scatter plot from Principal Component Analysis [19] of features obtained by the proposed method, while [Figure 6: see original paper] (right) shows a scatter plot of three time-domain features (variance, root mean square, kurtosis).

The left plot in [Figure 6: see original paper] demonstrates more concentrated feature information, with same operating conditions clustering well and different conditions effectively separated. In contrast, the right plot shows significant overlap between normal and rolling element fault states, making them difficult to distinguish. This validates the effectiveness of the proposed feature extraction method.

2.3 DBN-Based Fault Diagnosis Model Training

Each training sample consists of a feature sequence and a class label. The 200 training samples from Load 0 are input into the DBN for model training. To avoid sample imbalance, an equal number of samples is selected for each state.

The DBN model structure parameters are shown in . The input layer node count is determined by sample feature dimension, and the output layer node count by state categories. After testing various DBN depths with the same input data, a stack of 2 RBMs is adopted. Following Geoffrey Hinton' s recommendations [20], DBN learning parameters are set as shown in .

The mean squared error curve for the training set during model fine-tuning is shown in [Figure 7: see original paper]. The model achieves good training performance and stabilizes at 100 iterations. While increased iterations may improve fault recognition, computational time rises significantly. Considering both recognition effect and computational cost, 100 iterations are selected.

2.4 Experimental Results Analysis

To evaluate model training effectiveness, the remaining 200 samples from Load 0 form the test set, as shown in .

Inputting the test set from into the trained DBN model, output data are analyzed to calculate diagnostic accuracy, defined as the ratio of correct results to total results. Results are shown in [Figure 8: see original paper].

In [Figure 8: see original paper], red markers indicate predicted labels, black dots represent true labels, and blue line values of 0 denote correct predictions while -1 indicates misclassification. Among 200 test samples, only the 50th sample was misclassified as outer race fault type 2, achieving 99.5% accuracy.

To reduce randomness effects, the experiment is repeated 10 times. Accuracy rates are 99.5%, 99%, 99.5%, 100%, 99.5%, 99.5%, 99.5%, 99%, 99.5%, and 100%, with an average accuracy of 99.5%, confirming high diagnostic precision and robustness.

---

3 Comparative Analysis

3.1 Comparison with Other Algorithms

DBN is compared with GRA, SVM, and BPNN by calculating average diagnostic accuracy and standard deviation across 10 trials for each method, as shown in .

The proposed method demonstrates higher accuracy and stability than GRA, SVM, and BPNN, attributed to the strong self-learning capability of the deep learning approach.

3.2 Effect of Variable Load on Algorithm Accuracy

To further verify model generalization, the same data processing method is applied using experimental data from Load 0, Load 1, Load 2, and Load 3 as training sets, with different load datasets forming test samples. Recognition results are shown in .

The proposed method achieves over 95% classification accuracy even under variable load conditions, demonstrating robustness and generalization capability while maintaining classification precision.

3.3 Effect of Different Sample Sizes on Accuracy and Efficiency

Using 400 sample groups from Load 0, the ratio of training to test samples is varied while keeping total sample count constant. Datasets and results are shown in .

shows that as training sample size increases, the proposed method' s accuracy continues to grow, consistently exceeding 99.5%, effectively avoiding overfitting. Generally, larger training sample sizes yield higher test accuracy but require more computational time. Since feature extraction reduces input data dimensionality for deep learning, time consumption does not increase dramatically. Direct input of raw data into DBN would increase time consumption by tens of times, especially with larger training sets.

---

4 Conclusion

This paper proposes a bearing fault diagnosis algorithm based on wavelet packet-AR spectrum estimation and deep learning, applying it to bearing signal analysis.

Results show the method achieves 99.5% diagnostic accuracy under identical load conditions and over 95% under variable loads, verifying higher reliability.

The method offers two advantages: (1) Feature extraction filters noise and significantly reduces model training time; (2) DBN' s strong self-learning capability substantially improves diagnostic accuracy.

Future research directions include applying this method to more machinery types and investigating fault diagnosis techniques under complex operating conditions.

References

[1] Wei Zexian, Wang Yanxue, He Shuilong, et al. A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection [J]. Knowledge-Based Systems, 2017, 116: 1-12.

[2] Wu Zhe. Research on Fault Diagnosis and Prognosis Method and Its Application for Rotating Machinery [D]. Beijing: Beijing Jiaotong University, 2016.

[3] Deng Feiyue. Research on fault feature extraction and diagnosis method of rolling element bearing [D]. Beijing: North China Electric Power University, 2016.

[4] Xing Xin, Cui Yahui, Liu Xiaolin, et al. A fault diagnosis method for rolling bearings based on adaptive extraction of effective signals [J]. Noise and Vibration Control, 2018, 38 (2): 150-153.

[5] Chen Jinglong, Li Zipeng, Pan Jun, et al. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review [J]. Mechanical systems and signal processing, 2016, 70: 1-35.

[6] Zhao Hongshan, Xu Fanhao, Xu Wenqi, et al. Feature extraction method of transformer vibration based on ensemble empirical mode decomposition subband [C]// Proc of IEEE International Conference on Power System Technology. 2016: 1-6.

[7] Zhang Lingling, Xiao Jing. Case study of mechanical fault diagnosis technology based on MATLAB [M]. Beijing: Higher Education Press, 2016: 140-150.

[8] Chang Chun, Wang Guowei, Mei Jianmin, et al. Connecting rod bearing fault diagnosis based on the wavelet packet-auto regressive model spectrum and support vector machine [J]. Journal of Military Transportation University, 2015, 17 (4): 40-44.

[9] Liu Ruonan, Yang Boyuan, Enrico Zio, et al. Artificial intelligence for fault diagnosis of rotating machinery: A review [J]. Mechanical Systems and Signal Processing, 2018, 108: 33-47.

[10] Wu Kui, Wang Xianyong, Sun Jie, et al. Fault diagnosis method based on the deep belief network [J]. Computer Measurement & Control. 2017, 25 (10): 43-47.

[11] Hinton G, Deng Li, Yu Dong, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups [J]. IEEE Signal Processing Magazine, 2012, 29 (6): 82-97.

[12] Jiang Mingyang, Liang Yanchun, Feng Xiaoyue, et al. Text classification based on deep belief network and softmax regression [J]. Neural Computing and Applications, 2018, 29 (1): 61-70.

[13] Lei Yaguo, Jia Feng, Zhou Xin, et al. A deep learning-based method for machinery health monitoring with big data [J]. Journal of Mechanical Engineering, 2015, 51 (21): 49-56.

[14] Li Yanfeng, Wang Xinqing, Zhang Meijun, et al. An approach to fault diagnosis of rolling bearing using svd and multiple DBN classifiers [J]. Journal of Shanghai JiaoTong University, 2015, 49 (05): 681-686.

[15] Hinton G E, Osindero S, The Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18: 1527-1554.

[16] Fischer A, Igel C. An introduction to restricted Boltzmann machines [C]// Proc of Iberoamerican Congress on Pattern Recognition. Berlin: Springer, 2012: 14-36.

[17] Larochelle H, Erhan D, Courville A, et al. An empirical evaluation of deep architectures on problems with many factors of variation [C]// Proc of the 24th International Conference on Machine Learning. New York: ACM Press, 2007: 473-480.

[18] Ackley D H, Hinton G E, Sejnowski T J. A learning algorithm for Boltzmann machines [M]// Readings in Computer Vision. 1987: 522-533.

[19] Karl P L. On lines and planes of closest fit to systems of points in space [J]. Philosophical Magazine Series 6, 1901, 2 (11).

[20] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18 (7): 1527-1554.

Note: Figure translations are in progress. See original paper for figures.