Bearing fault diagnosis based on improved ensemble learning and deep belief network

Tianchen Liang1,3, Shuaipeng Wu1, Wenjing Duan1 and Ruishan Zhang2
1 Northwestern polytechnical university
2 Xi’an aircraft industrial company
3 E-mail: liangtc@mail.nwpu.edu.cn

Abstract. The collected bearing signals are easily interfered by strong ambient noise due to complex operating conditions. It’s a challenge to identify faults accurately and to reduce dependence on model hyper-parameters for intelligent diagnostic methods. This paper proposes an improved ensemble method based on deep belief network (DBN) for fault diagnosis of rolling bearings. Firstly, a series of DBNs with different hyper-parameters are constructed and trained. Secondly, the improved ensemble method is used to acquire the weights matrix for each DBN. Finally, each DBN votes together in accordance with its respective weight matrix to get the final diagnosis result. This method is applied to rolling bearing data from Case Western Reserve University. Experiments show that the effect of fault diagnosis is significantly improved because the feature learning ability of different DBNs is fully taken advantage of.

1. Introduction
Rolling bearings are the most widely used components in rotating machinery. Their operating conditions directly affect the overall performance and safety of mechanical equipment [1]. On the other hand, different types of failures inevitably occur due to overload, impact and fatigue during the operation of rolling bearings [2]. Therefore, the intelligent diagnosis of bearing faults is of great practical significance.

Under the circumstance of mechanical big data age, data-driven fault diagnosis methods play an increasingly important role, which mainly include signal processing methods and intelligent diagnosis methods [3-4]. Intelligent diagnosis first extracts fault features from the original vibration signal, and then intelligently identifies fault types based on features. Artificial neural network (ANN) and support vector machine (SVM) are two popular intelligent methods in recent years [5-6]. However, SVM and ANN often need to manually extract time domain or frequency domain information to design high quality features which is time-consuming and relies on engineering experience and prior knowledge. Therefore, it’s urgent to get rid of manual feature extraction for real intelligent diagnosis.

Recently, breakthroughs in deep learning have provided new solutions for the intelligent fault diagnosis of rotating machinery. Deep learning methods can automatically capture potential high-value features from raw data. This gradually makes intelligent diagnostic methods free from the dependence on signal processing techniques and manual feature extraction [7]. Scholars studied the fault diagnosis of rotating machinery for bearings, gearboxes and aircraft engines using DBNs, stacked auto-encoders, and Improved convolutional DBN [8-11]. However, these diagnostic methods focus on single learning model, which limits the generalization ability. To achieve the best diagnostic accuracy,
the hyper-parameters for the single model need to be carefully designed. This means that although these methods avoid the manual extraction of features, hyper-parameters still need to be well adjusted.

Ensemble learning is a classical machine learning method that combines a series of individual learners to obtain better effect than each single learner does. The general ensemble learning algorithm only considers the learner's overall classification ability but ignores its ability for different categories. This paper presents a diagnosis method for rolling bearings based on DBNs and improved ensemble learning algorithm.

2. Deep belief network theory

2.1. Restricted Boltzmann Machine (RBM)

RBM is composed of visible layer and hidden layer. The visible layer neurons only link with each neuron in the hidden layer and there is no connection among neurons in the same layer. RBM structure is shown in figure 1.

![Figure 1. Classic structure of RBM.](image)

Where $n_v$ is the number of the visible layer neurons. $n_h$ represents the number of the hidden layer neurons. $v = (v_1, v_2, ..., v_{n_v})^T$ is the state vector in the visible layer. $h = (h_1, h_2, ..., h_{n_h})^T$ is the state vector in the hidden layer. $a = (a_1, a_2, ..., a_{n_v})^T \epsilon \mathbb{R}^{n_v}$ is offset vector in the visible layer, $b = (b_1, b_2, ..., b_{n_h})^T \epsilon \mathbb{R}^{n_h}$ is offset vector in the hidden layer. $W \epsilon \mathbb{R}^{n_h \times n_v}$ is the weight matrix.

RBM neurons include Binary unit, Gaussian unit and Rectified linear unit. Both Binary unit and Gaussian unit are used in this paper. Both Binary unit and Gaussian unit are used in this paper. The following comes the basic theory for RBM with Binary visible units and Binary hidden units. For a set of states ($v$, $h$) and $\Theta = \{a, b, W\}$, energy function is defined as:

$$E_{\Theta} = -a^T v - b^T h - h^T W v$$

According to this energy function, Joint probability distribution of states ($v$, $h$) is calculated as:

$$P_{\Theta}(v, h) = \frac{1}{Z_{\Theta}} e^{-E_{\Theta}(v, h)}$$

$$Z_{\Theta} = \sum_{(v, h)} e^{-E_{\Theta}(v, h)}$$

According to $P_{\Theta}(v, h)$, probability distributions and conditional probability distribution as:

$$P_{\Theta}(v) = \sum_{h} P_{\Theta}(v, h) = \frac{1}{Z_{\Theta}} \sum_{h} e^{-E_{\Theta}(v, h)}$$

$$P_{\Theta}(h) = \sum_{v} P_{\Theta}(v, h) = \frac{1}{Z_{\Theta}} \sum_{v} e^{-E_{\Theta}(v, h)}$$

$$P(h_k = 1 | v) = \text{sigmoid}(b_k + \sum_{i=1}^{n_v} W_{k,i} v_i)$$

$$P(v_k = 1 | h) = \text{sigmoid}(a_k + \sum_{j=1}^{n_h} W_{j,k} h_j)$$

Sigmoid is defined as $\text{sigmoid}(t) = 1/(1 + e^{-t})$. ReLU is defined as $\text{ReLU}(t) = \max(0, x)$.

RBM training is the maximum likelihood estimation of the parameter $\Theta = \{a, b, W\}$, maximum likelihood estimation can be expressed as:

$$L_{\Theta} = \prod_{i=1}^{N} P(v^i)$$

Log likelihood function can be expressed as:

$$\log L_{\Theta} = \sum_{i=1}^{N} \log P(v^i)$$
\[
\ln(L_{\theta}) = \ln\left(\prod_{i=1}^{N} P(v^{i})\right) = \sum_{i=1}^{N} \ln(P(v^{i}))
\]

Contrastive Divergence [12], also known as CD-k algorithm, is a classical way to train RBM. Its basic idea is to use training samples as the starting point of the Markov Chain Monte Carlo which only needs a few state transitions to achieve RBM distribution. Then \(v^{(k)}\) is obtained by the k-step Gibbs sampling. The gradient formula is as follows:

\[
\frac{\partial \ln P(v)}{\partial W_{ij}} \approx P(h_{i} = 1 \mid v^{(0)})v_{j}^{(0)} - P(h_{i} = 1 \mid v^{(k)})v_{j}^{(k)} \quad (10)
\]

\[
\frac{\partial \ln P(v)}{\partial a_{i}} \approx v_{i}^{(0)} - v_{i}^{(k)} \quad (11)
\]

\[
\frac{\partial \ln P(v)}{\partial b_{i}} \approx P(h_{i} = 1 \mid v^{(0)}) - P(h_{i} = 1 \mid v^{(k)}) \quad (12)
\]

2.2. Construction of DBN based on RBM

Deep belief network is formed by stacking Restricted Boltzmann Machines one by one. The previous RBM output is used as the input of the next RBM. It can be defined as:

\[X_{i+1} = H_{i}, \ i \in \{1, 2, ..., N - 1\} \quad (13)\]

where \(i\) is the \(i\)th RBM. \(N\) is RBM number. \(H_{i}\) is the output of the \(i\)th RBM. \(X_{i+1}\) is the input of \(i + 1\)th RBM. Through layer-by-layer unsupervised feature learning, fault features \(H_{\text{final}}\) is obtained. Softmax is used to identify the fault according to features.

3. Ensemble learning

3.1. Improved ensemble learning algorithms

Ensemble learning is accomplished by training multiple learners and then combining them to get better performance. Combining strategy is a key step in ensemble learning. Simple voting is the most commonly used combining strategy at present [13]. But it only considers overall classification ability for all labels and ignores the ability to distinguish different categories. In order to solve this problem, we bring up a brand new combining strategy called improved weight voting (IWV) to integrate learners, which is expressed as:

\[H_{\text{IWV}}(X) = c_{\arg \max \sum_{i=1}^{T} w_{ij}h_{i}^{j}(x)} \quad (w_{ij} \geq 0, \ \sum_{i=1}^{T} w_{ij} = 1) \quad (14)\]

Where \(H_{\text{IWV}}(X)\) is the final output of IWV. \(w_{ij}\) is the weight of the \(i\)th learner for the \(j\)th label. Bayesian optimal discriminant function is defined as:

\[H^{j}(X) = \log(P(c_{j})P(\ell \mid c_{j})) \quad (15)\]

\[P(\ell \mid c_{j}) = \prod_{i=1}^{T} P(\ell_{i} \mid c_{j}) \quad (16)\]

\[H^{j}(X) = \log P(c_{j}) + \sum_{\ell=1}^{T} \log \frac{p_{ij}}{1-p_{ij}} + \sum_{i=1}^{T} \log(1 - p_{ij}) \quad (17)\]

Where \(\ell = (\ell_{1}, \ell_{2}, ..., \ell_{T})\) represents individual classifier output label. \(p_{i,j}\) is the accuracy of individual classifier \(h_{i}\) for label \(c_{j}\). \(\ell_{i} = c_{j}\) can be replaced by \(h_{i}^{j}(x)\), which is:

\[\sum_{\ell=1, \ell=c_{j}}^{T} \log \frac{p_{ij}}{1-p_{ij}} = \sum_{i=1}^{T} h_{i}^{j}(x) \log \frac{p_{ij}}{1-p_{ij}} \quad (18)\]

The voting weight can be calculated as:

\[w_{ij} \propto \log \left(\frac{p_{ij}}{1-p_{ij}}\right) \quad (19)\]

\[w_{ij} = \frac{\log \left(\frac{p_{ij}}{1-p_{ij}}\right)}{\sum_{i=1}^{T} \log \left(\frac{p_{ij}}{1-p_{ij}}\right)}, \ \ j \in \{1, 2, ..., N\} \quad (20)\]

3.2. General procedure of the proposed method

This paper ensembles multiple DBNs for intelligent fault diagnosis of rolling bearings. The method consists of three parts: designing and training different hyper-parametric DBN models, constructing
the ensemble-DBNs and getting the final diagnosis result. The framework of the proposed method is shown in figure 2, and the detailed steps of the proposed methods are as follows:

- **Step 1.** Bearing data is collected by sensors;
- **Step 2.** Without any signal processing and feature extraction, the original vibration signal is normalized and then divided into training and testing samples;
- **Step 3.** On the training set, DBNs with different network hyper-parameters is trained by CD-k algorithm, and the performance of each model is obtained by cross-validation;
- **Step 4.** Determine the final combining weight based on the performance of each model in cross-validation;
- **Step 5.** Put the testing samples into the diagnostic model and obtain the diagnostic results.

![Figure 2. The framework of the proposed method.](image)

4. **Experimental verification**

4.1. **Experimental data description**

The bearing fault data from Case Western Reserve University is used for experimental analysis. Experimental device is shown in figure 3. The sample settings are shown in table 1. The vibration signal is collected at 1797 rpm motor speed. More detailed data information is available in [14]. Each sample contains 400 data points and the data description is shown in table 1. Figure 4 shows raw vibration signals, in which each kind of signal is ordered according to its label in table 1.

![Figure 3. Experimental device.](image)  
![Figure 4. Experimental vibration signals.](image)
| Bearing fault location | Damage degree (mile) | Training/Testing sample number | label |
|------------------------|----------------------|-------------------------------|-------|
| normal                 | 0                    | 240/60                        | 1     |
| ball fault             | 0.007                | 240/60                        | 2     |
| inner race fault       | 0.007                | 240/60                        | 3     |
| outer race fault       | 0.007                | 240/60                        | 4     |
| ball fault             | 0.014                | 240/60                        | 5     |
| inner race fault       | 0.014                | 240/60                        | 6     |
| outer race fault       | 0.014                | 240/60                        | 7     |
| ball fault             | 0.021                | 240/60                        | 8     |
| inner race fault       | 0.021                | 240/60                        | 9     |
| outer race fault       | 0.021                | 240/60                        | 10    |

4.2. Base model comparison

In this experiment, three intelligent models are compared with DBN: Back propagation neural network (BP), Radial basis function neural network (RBF) and SVM. The input is the original signal that has not been processed. The specific parameters are set as follow:

1. DBN adopts Sigmoid activation function. The network structure is [400 200 100 10]. The learning rate and momentum rate are 0.05 and 0.6 respectively, and the iteration is 800;
2. BP adopts Sigmoid activation function. The network structure is [400 100 100 10]. The learning rate is 0.1, and the iteration is 800;
3. RBF, radial basis function adapts Gaussian function. The network structure is [400 240 10], and expansion speed is 0.3;
4. SVM, kernel function adapts Gaussian function. Penalty factor and Gaussian nuclear radius are 20 and 0.3 respectively.

The four models are applied in ten repeated tests. Figure 5 shows the test accuracy of the ten repeated tests. The average accuracy is shown in table 2. The average accuracy of DBN is 84.2% which is far higher than BP neural network with 58.05% accuracy, RBF neural network with 52.5% accuracy and SVM with 47.1% accuracy. The excellent performance of DBN is due to its powerful feature representation ability.

![Figure 5. The correct rate of recognition of the ten repeated tests.](image)

Table 2. Average recognition accuracy.

|       | BP     | RBF    | SVM    | DBN    |
|-------|--------|--------|--------|--------|
| Average accuracy (%) | 58.05  | 52.5   | 47.1   | 84.2   |
4.3. Ensemble learning diagnosis
This section verifies the actual effect of the IWV proposed in this paper. The improved voting ensemble of 10 DBN is compared with each of the 10 DBN with different hyper-parameters and simple voting ensemble of 10 DBN. The model parameters and the diagnostic results are shown in table. 3. The learning rate and momentum rate are 0.05 and 0.6 respectively, B and G represent Binary unit and Gaussian unit respectively.

Figure 6 shows the accuracy of the ten repeated experiments. The testing accuracy of proposed IWV is 96.40%, 96.90%, 97.33%, 96.76%, 96.51%, 96.99%, 97.68%, 97.09%, 96.89% and 96.97% respectively. Figure 6 shows the average test accuracy rate is 96.95%. Compared with the best DBN with 85.60% accuracy and simple voting ensemble with 91.21% accuracy, the proposed method is more effective without any parameters adjustment.

| Table 3. Parameter setting of DBNs. |
|-----------------------------------|
| Activation function | Network structure | Unit type | Average accuracy (%) |
| DBN 1 | Sigmoid | [400 200 100 10] | BB | 83.05 |
| DBN 2 | Sigmoid | [400 150 150 10] | BB | 85.60 |
| DBN 3 | Sigmoid | [400 100 100 10] | BB | 83.06 |
| DBN 4 | Sigmoid | [400 200 100 10] | GB | 84.71 |
| DBN 5 | Sigmoid | [400 150 150 10] | GB | 83.72 |
| DBN 6 | ReLU | [400 200 100 10] | BB | 81.06 |
| DBN 7 | ReLU | [400 150 150 10] | BB | 84.59 |
| DBN 8 | ReLU | [400 100 100 10] | BB | 84.61 |
| DBN 9 | ReLU | [400 200 100 10] | GB | 84.90 |
| DBN 10 | ReLU | [400 150 150 10] | GB | 83.06 |

5. Conclusions
This paper proposes an improved ensemble method based on deep belief network (DBN) for fault diagnosis of rolling bearings. This method can be divided into the following three steps: Firstly, DBNs are constructed according to different hyper-parameters. Secondly, the proposed IWV is used to determine the weight matrix for each DBN. Finally, Each DBN votes together in accordance with its respective weight matrix to get the final diagnosis result.

Experiments prove: (1) Compared with shallow models, DBN can more effectively capture fault features from raw data and avoid manual features extraction. (2) The proposed IWV can make full use
of the learning ability of single model, which avoid the model's dependence on hyper-parameters. In the future we will continue to focus on the application of integration strategies and stacked auto-encoders.

References

[1] Jiang H, Li C and Li H 2013 An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis Mechanical Systems & Signal Processing 36(2) 225-239
[2] Lei Y, Lin J, He Z, et al. 2013 A review on empirical mode decomposition in fault diagnosis of rotating machinery Mechanical Systems & Signal Processing 35(1–2):108-126
[3] Mao W, He L, Yan Y, et al. 2016 Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine Mechanical Systems & Signal Processing 83:450-473
[4] Lv Y, Yuan R, Song G 2016 Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing Mechanical Systems & Signal Processing 81:219-234
[5] Shao H, Jiang H, Zhang H, et al. 2018 Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing Mechanical Systems & Signal Processing 100:743-765
[6] Shao H, Jiang H, Zhao H, et al. 2017 A novel deep autoencoder feature learning method for rotating machinery fault diagnosis Mechanical Systems & Signal Processing 95:187-204
[7] Shao H, Jiang H, Zhao H, et al. 2016 An enhancement deep feature fusion method for rotating machinery fault diagnosis Knowledge-Based Systems 119
[8] Tamilselvan P, Wang P 2013 Failure diagnosis using deep belief learning based health state classification Reliability Engineering & System Safety 115(7):124-135
[9] Tran V T, Althobiani F, Ball A 2014 An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks Expert Systems with Applications 41(9):4113-4122
[10] Sun W, Shao S, Zhao R, et al. 2016 A sparse auto-encoder-based deep neural network approach for induction motor faults classification Measurement 89:171-178
[11] Xing S, Lei Y, Jia F, et al. 2017 Intelligent fault diagnosis of rotating machinery using locally connected restricted boltzmann machine in big data era [C]// IEEE International Conference on Industrial Engineering and Engineering Management. IEEE 1930-1934
[12] Hinton G E 2014 Training products of experts by minimizing contrastive divergence Neural Computation 14(8):1771-1800
[13] Polikar R 2012 Ensemble Learning [M]// Ensemble Machine Learning. Springer US 1-34
[14] Loparo K 2012 Case Western Reserve University Bearing Data Centre Website [Online] Available: http://cseroups.case.edu/bearingdatacenter/pages/download-data-file