Microelectronics Reliability xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Microelectronics Reliability

journal homepage: www.elsevier.com/locate/microrel

Introductory invited paper

Deep neural networks-based rolling bearing fault diagnosis

Zhiqiang Chen [a,b,c,*], Shengcai Deng [c], Xudong Chen [a,b], Chuan Li [a], René-Vinicio Sanchez [d], Huafeng Qin [a,b]

[a] National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China
[b] Chongqing Engineering Laboratory for Detection Control and Integrated System, Chongqing Technology and Business University, Chongqing 400067, China
[c] Chongqing Key Laboratory of Electronic Commerce and Supply Chain, Chongqing Technology and Business University, Chongqing 400067, China
[d] Department of Mechanical Engineering, Universidad Politécnica Salesiana, Cuenca, Ecuador

ARTICLE INFO

Article history:
Received 6 December 2016
Received in revised form 7 February 2017
Accepted 8 March 2017
Available online xxxx

Keywords:
Rolling bearing
Fault diagnosis
Deep Boltzmann Machines
Deep Belief Networks
Stacked Auto-Encoders

ABSTRACT

Rolling bearing is one of the most commonly used components in rotating machinery. It's so easy to be damaged that it can cause mechanical fault. Thus, it is significant to study fault diagnosis technology on rolling bearing. In this paper, three deep neural network models (Deep Boltzmann Machines, Deep Belief Networks and Stacked Auto-Encoders) are employed to identify the fault condition of rolling bearing. Four preprocessing schemes including feature of time domain, frequency domain and time-frequency domain are discussed. One data set with seven fault patterns is collected to evaluate the performance of deep learning models for rolling bearing fault diagnosis, which is based on the health condition of a rotating mechanical system. The results proved that the accuracy achieved by Deep Boltzmann Machines, Deep Belief Networks and Stacked Auto-Encoders are highly reliable and applicable in fault diagnosis of rolling bearing.

1. Introduction

Rolling bearing is the key component of rotating machinery, whose running state has an important effect on the healthy condition of rotating machinery equipment. It's very necessary to monitor the condition of rolling bearing and identify its faults to avoid fatal breakdowns of machines and prevent loss of production and human casualties [1]. Accordingly, a reliable bearing health-condition-monitoring system is very useful to a wide array of industries.

Traditional maintenance operations by observing the variations and/or trends of some condition-monitoring indices is usually time consuming and not always reliable when multiple features (techniques) are applied for fault diagnostics, particularly as the data are noise affected [2]. The availability of an important number of condition parameters that are extracted from rolling bearing signals, such as vibration signals, has motivated the use of data-driven fault diagnosis. Q. Miao and Zhou [3] special investigated the relation between vibration signal characteristics and fault diagnosis for planetary gearbox. Some automatic decision-making techniques have been proposed in the literature [4–9], which can broadly be classified into mathematical model-based methods and data-driven techniques. The latter technique is employed in this paper, because an accurate numerical model is usually difficult to derive from complex mechanical systems, particularly when the machine is operating in an uncertain noisy environment, and data-driven techniques were widely applied to the fault diagnosis of rolling bearing [2].

Classical data-driven fault diagnosis techniques include statistical classifiers, geometric methods, and polynomial classifiers [10]. These techniques cannot be used for time varying systems, because these models depend on statistic measurements (e.g., density and probability) of the vibration data. Under data-driven, inference-based intelligent tools such as neural networks [11,12], fuzzy logic [13,14], synergetic schemes [15], neuron fuzzy (NF) paradigms [16,17] and Bayesian inference [18] are widely used for rolling bear fault diagnosis. In addition to, some machine learning approaches such as support vector machine and their related models also are commonly used, because of the simplicity for developing industrial applications. Signal analysis-based techniques such as wavelet decomposition and its variants also are commonly used as classical tools for the robust health evaluation of rotating machinery [19,20].

Although some intelligent tool-based techniques have been proposed for bearing fault diagnostics [21–24],it still remains as a challenging task in research and development, because a bearing is not a mechanical component (e.g., a gear or shaft), but a complex system with inner and outer rings, as well as many rolling elements. The bearing signal is non-stationary in general, particularly when slip occurs among the bearing elements (e.g., the rolling elements and ring raceways) [2].

Nowadays, deep learning techniques are becoming a promising tool for fault characteristic mining and intelligent diagnosis of rotating

* Corresponding author.
E-mail addresses: czq@ctbu.edu.cn (Z. Chen), rsanchezl@ups.edu.ec (R.-V. Sanchez).

http://dx.doi.org/10.1016/j.microrel.2017.03.006
0026-2714/

machinery with massive data [25]. Chen et al. [26] introduced convolution neural networks (CNN) to identify and classify the faults of gearbox. Tran et al. [27] suggested a deep belief network based application to diagnose reciprocating compressor valves. C. Li et al. [28] proposed multimodal deep support vector classification for gearbox fault diagnosis, where Gaussian-Bernoulli Deep Boltzmann Machines (GDBMs) were used to extract the feature of the vibratory and acoustic signal in time, frequency and wavelet modalities, respectively; and then the extracted features are integrated for fault diagnosis by using GDBMs. Li's research [28] indicated Gaussian-Bernoulli Deep Boltzmann Machines is effective for the gearbox fault diagnosis. We have presented a multi-layer neural network (MLNN) for gearbox fault diagnosis [29], where the weights of deep belief network are used to initialize the weights of the constructed MLNN. More extensive engineering verification is necessary to cognize and popularizes deep learning techniques.

In this paper, three deep neural networks (Deep Boltzmann Machines [30], Deep Belief Networks [31] and Stacked Auto-Encoders [32]) are applied to the rolling bearing fault diagnosis. The following aspects are focused on: the performance evaluation of the feature extraction and fault identification based on the raw vibration signal and the feature representations from time, frequency and time-frequency domain for the rolling bearing fault. The rest of this paper is organized as follows: Three deep learning models are introduced briefly in Section 2. Preprocessing schemes of vibration signal are presented in Section 3. In Section 4, the experiment setups are described in detail, such as how to acquire data and extract features. Results and discussions are presented in Section 5. The Conclusions of this work are given at the end.

2. Methodology

Fig. 1 outlines the schematic of the developed rolling bearing fault diagnosis system. In this section, the used deep neural networks (DNN) will be briefly introduced: Deep Boltzmann Machines (DBM), Deep Belief Networks (DBN) and Stacked Auto-Encoders (SAE).

2.1. Deep Boltzmann Machine

Deep Boltzmann Machine is a network of symmetrically coupled stochastic units, which is an undirected graphical model with bipartite connections between adjacent layers of hidden units. DBM model can be regarded as a deep learning model which is a stack of restricted Boltzmann machines (RBM) [33]. Fig. 2 indicates a Deep Boltzmann

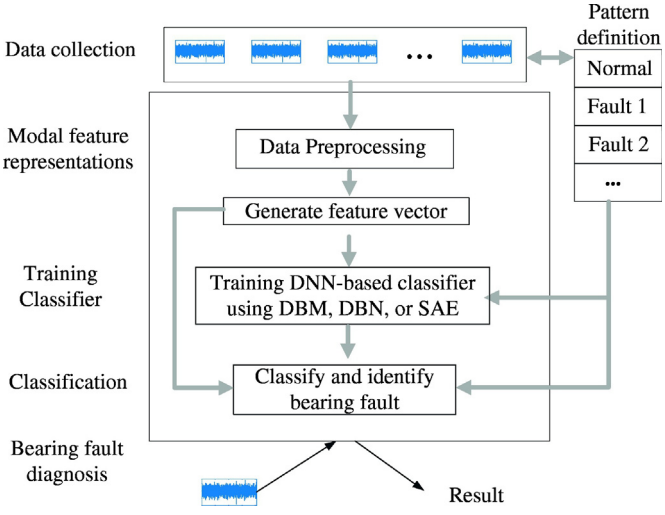

Fig. 1. Schematic of the developed rolling bearing fault diagnosis system.

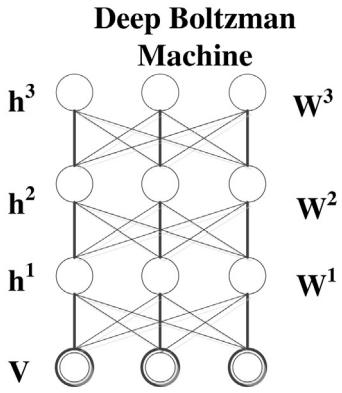

Fig. 2. Deep Boltzmann Machine with three hide layers.

Machine with three hide layers. The energy of the state $\{\mathrm{v}, \mathrm{h}^1, \mathrm{h}^2, \mathrm{h}^3\}$ is defined as [34]:

$$E\left(\boldsymbol{v}, \mathrm{h}^1, \mathrm{h}^2, \mathrm{h}^3, \theta\right) = -\mathrm{v}^T W^1 \mathrm{h}^1 - \mathrm{h}^{1T} W^2 \mathrm{h}^2 - \mathrm{h}^{2T} W^3 \mathrm{h}^3 \tag{1}$$

where the model parameters $\theta = \{W^1, W^2, W^3\}$ represents visible-to-hidden and hidden-to-hidden symmetric interaction terms. The probability that the model assigns to a visible vector $\mathbf{v}$ is

$$P(\mathrm{v}, \theta) = \frac{1}{Z(\theta)} \sum_{\mathrm{h}^1, \mathrm{h}^2, \mathrm{h}^3} \exp\left(-E\left(\mathrm{v}, \mathrm{h}^1, \mathrm{h}^2, \mathrm{h}^3, \theta\right)\right) \tag{2}$$

The conditional distributions over the hidden (L expresses the number of hidden layers) and visible states are defined as in (3) and (4):

$$P\left(\mathrm{v}_i | \mathrm{h}^1, \theta\right) = N\left(\mathrm{v}_i | \sum_j \mathrm{h}_j^1 W_{ij} + b_i, \sigma_i^2\right) \tag{3}$$

$$P\left(\mathrm{h}_j^l | \mathrm{h}^{l-1}, \mathrm{h}^{l+1}, \theta\right) = s\left(\sum_j \mathrm{h}_j^{l-1} W_{ij}^{l-1} + \sum_k \mathrm{h}_k^{l+1} W_{jk}^l + b_j^l, \sigma_i^2\right) \tag{4}$$

where l expresses l^{th} hidden layer and b is the bias value of neuron.

The training steps of the DBM model are as follows [34]:

Step 1: Train the first-layer RBM by using one-step contrastive divergence learning with mean-field reconstructions of the visible vectors. The generative probabilistic model can be written as:

$$\begin{aligned} P(\mathrm{v}, \theta) &= \textstyle\sum_{\mathrm{h}^1} P\left(\mathrm{h}^1; W^1\right) P\left(\mathrm{v} | \mathrm{h}^1; W^1\right), P\left(\mathrm{h}^1; W^1\right) \\ &= \textstyle\sum_{\mathrm{v}} P\left(\mathrm{h}^1, \mathrm{v}, W^1\right) \end{aligned} \tag{5}$$

Table 1
Time domain features.

Description of features	Formulate
Mean value	$\overline{X} = \frac{1}{N}\sum_{n=1}^{N} X(n)$
Standard deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N} (X(n)-\overline{X})^2}$
Mean square root	$X_{\mathrm{RMS}} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} x(t)^2}$
Kurtosis	$K = \frac{1}{N}\sum_{n=1}^{N} \frac{(X(n)-\overline{X})^4}{\sigma^4}$
Shape factor	$S_f = \frac{X_{\mathrm{RMS}}}{\lvert\overline{X}\rvert}$
Crest factor	$C_f = \frac{\max(x(t))}{X_{\mathrm{RMS}}}$
Impulse factor	$I_f = \frac{\max(x(t))}{\lvert\overline{X}\rvert}$

Please cite this article as: Z. Chen, et al., Deep neural networks-based rolling bearing fault diagnosis, Microelectronics Reliability (2017), http://dx.doi.org/10.1016/j.microrel.2017.03.006

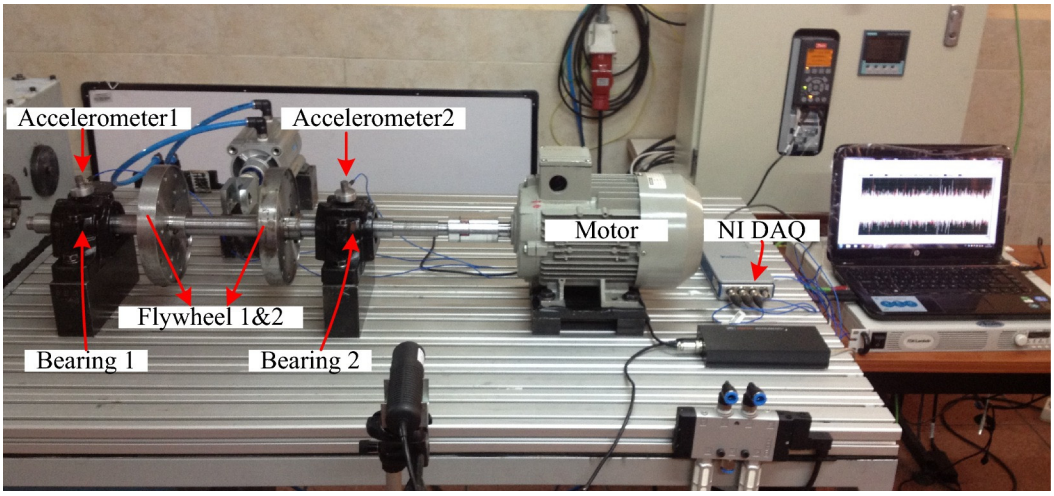

Fig. 3. Experimental platform1 fault simulator schematic.

Step 2: Use the reconstruction vectors to train the second RBM, replace $P(h^1; W^1)$ with$P(h^1; W^2) = \sum_{h^2} P(h^1, h^2; W^2)$.

Step 3: Repeat the above procedures until all the RBM are trained well.

Step 4: Update weights W and initialize the weight of neural network with new W.

Step 5: Fine-tune parameters with standard back propagation algorithm.

2.2. Deep Belief Networks

Deep Belief Networks (DBNs) [31] can be viewed as another greedy, layer-by-layer unsupervised learning algorithm that consists of learning one layer from a stack of RBMs at one time. The top two layers form a restricted Boltzmann machine, but the lower layers form a directed sigmoid belief network [34]. After learning the first RBM in the stack, the generative model can be written as [34]:

$$P(\mathbf{v}, \theta) = \sum_{h^1} P\left(\mathbf{h}^1; W^1\right) P\left(\mathbf{v}|\mathbf{h}^1; W^1\right) \tag{6}$$

The training algorithm for DBNs proceeds as follows. Let X be a matrix of inputs which are regarded as a set of feature vectors.

Step 1: Train a restricted Boltzmann machine on X to obtain its weight matrix, W, and use this as the weight matrix between the lower two layers of the network.

Step 2: Transform X with the RBM to produce a new data X'.

Step 3: Repeat this procedure with $X \leftarrow X'$ for the next pair of layers, until the top two layers of the network are reached.

Step 4: Fine-tune all the parameters of this deep architecture under the supervised criterion.

Table 2
The definition of fault pattern.

Designator	Description
H1H2P1	Healthy bearing 1, healthy bearing 2
I1H2P2	Inner race fault bearing 1, healthy bearing 2
O1H2P3	Outer race fault bearing 1, healthy bearing 2
B1H2P4	Ball fault bearing 1, healthy bearing 2
I1O2P5	Inner race fault bearing 1, outer race fault bearing 2
I1B2P6	Inner race fault bearing 1, ball fault bearing 2
O1B2P7	Outer race fault bearing 1, ball fault bearing 2

2.3. Stacked Auto-Encoders

The Auto-Encoder is trained to encode the input X into some representation $C(X)$ so that the input can be reconstructed from that representation [32]. Hence the target output of the auto-encoder is the auto-encoder input itself. If there is one linear hidden layer and the mean squared error criterion is used to train the network, then the k hidden units will learn to project the input in the span of the first k principal components of the data. Auto-encoders have been used as building blocks to build and initialize a deep multi-layer neural network [35]. The training procedure is similar to the one for Deep Belief Networks. The principle is exactly the same as the one previously proposed for training DBNs, but auto-encoders instead of RBMs are used as follows [36]:

Step 1: Train the first layer as an auto-encoder to minimize certain forms of reconstruction errors of the raw input.

Step 2: The hidden units' outputs of the auto encoder are used as input for another layer, also trained to be an auto-encoder.

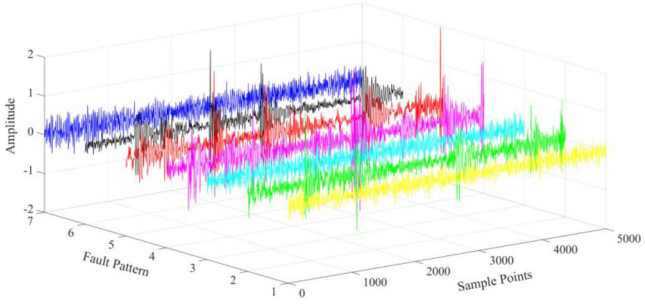

Fig. 4. Vibration signals of 7 fault patterns rolling bearing under F3 and L3.

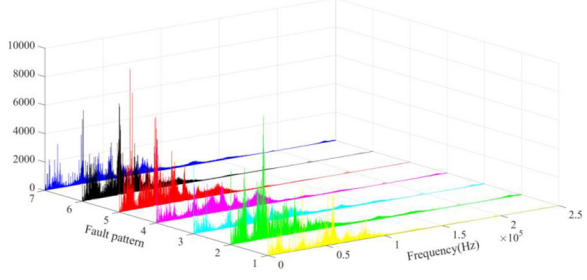

Fig. 5. FFT corresponding to vibration signal shown in Fig.4.

Please cite this article as: Z. Chen, et al., Deep neural networks-based rolling bearing fault diagnosis, Microelectronics Reliability (2017), http://dx.doi.org/10.1016/j.microrel.2017.03.006

Table 3
Parameter setting for Scheme 1.

Description	Parameters
Network architecture (n_h and n_{neuron})	{1250} {1250, 625} {1250, 625, 300} {1250, 625, 300, 150} {1250, 625, 300, 150, 100, 50, 50, 50, 50}
lRate	1.0, 0.5, 0.1, 0.05, 0.01
$epoch_1$	1, 2, 4, …, 20
$epoch_2$	10, 30, 50, 100, …, 200
Activation function	sigmoid

Step 3: Iterate as in step (2) to initialize the desired number of additional layers.

Step 4: Take the last hidden layer output as input to a supervised layer and initialize its parameters (either randomly or by supervised training, keeping the rest of the network fixed).

Step 5: Fine-tune all the parameters of this deep architecture with respect to the supervised criterion. Alternately, unfold all the auto-encoders into a very deep auto-encoder and fine-tune the global reconstruction error.

3. Preprocessing scheme of vibration signal

As shown in Fig. 1, the first step of the vibration-based analysis method is to extract representative features by using appropriate signal-processing techniques. In order to evaluate the performance of the feature extraction and fault identification of the developed DNN-based fault diagnosis system, the following four pretreatment schemes will be focused on.

Scheme 1: The raw vibration signals in time domain from the measurements are directly used as the input parameters of the deep neural network model. The corresponding input parameter is expressed as vector V_I;

Scheme 2: Extract the energy spectrum features of signal in time-frequency domain. The corresponding energy spectrum feature is expressed as vector V_{II};

Scheme 3: Extract the features of signal in time and frequency domain. The corresponding feature is expressed as V_{III};

Scheme 4: Combine VII with V_{III}, that is to say, the feature is $V_{IV} = V_{II} + V_{III}$.

3.1. Definition of the feature vector V_{II}

The feature vector V_{II} is from energy spectrum features of time-frequency domain. The vibration signal can be decomposed into different frequency bands by using wavelet packet decomposition with six layers. The varying of the energy spectrum of vibration signals can reflect the health condition of rolling bearing. Thus, energy spectrum features are extracted from each frequency band as the input parameters of DNN model. The steps of extracting energy spectrum features are as follows:

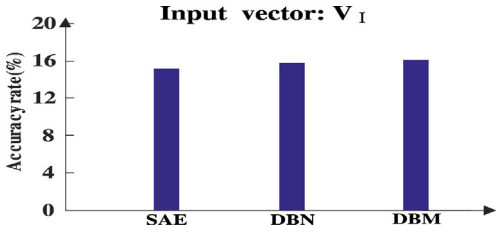

Fig. 6. The classification accuracy by using Scheme 1.

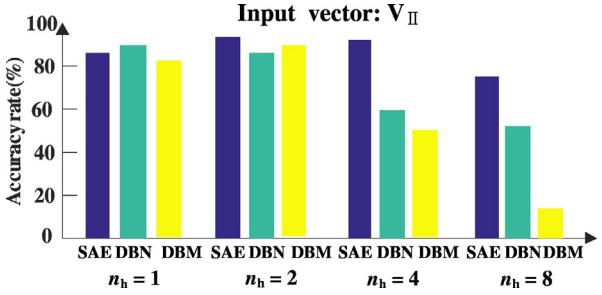

Fig. 7. Investigation of n_h for Scheme 2.

Step 1: Decompose vibration signal by using wavelet packet decomposition method, and obtain coefficients of wavelet decomposition. Here the depth of decomposition is set to 6. The coefficient vector of signal is defined as

$$X_6^0, X_6^1, X_6^2, X_6^3, X_6^4, \ldots, X_6^{61}, X_6^{62}, X_6^{63} \tag{7}$$

where $X_6^j (j = 0, 1, \ldots, 63)$ represents the j-th coefficient vector of signal.

Step 2: Reconstruct the signal basing on the coefficients of wavelet decompositionX_6^j.

$$s = \sum_{j=0}^{63} s_6^j \tag{8}$$

where $s_6^j (j = 0, 1, \ldots, 63)$ represents the reconstructed signal of j-th sub-band signal, and corresponding to$X_{jk} (j = 0.1,2\ldots7; k = 0.1,2\ldots N)$, it represents the amplitude of reconstruction signal of $s_3^j(t)$, and N is the number of sample point of the vibration signal.

Step 3: Calculate the energy values of each sub-band signal.

$$E_6^j = \int \left| s_6^j(t) \right|^2 dt = \sum_{k=1}^{N} |X_{jk}|^2 \tag{9}$$

Step 4: Construct the representations of energy spectrum feature as

$$E = \sum_{j=0}^{63} E_6^j E_6^j \tag{10}$$

$$V_{II} = \left[\frac{E_6^0}{\sqrt{E}}, \frac{E_6^1}{\sqrt{E}}, \frac{E_6^2}{\sqrt{E}}, \ldots, \frac{E_6^{63}}{\sqrt{E}} \right] \tag{11}$$

3.2. Definition of the feature vector V_{III} and V_{IV}

The feature vector V_{III} represents the features of the time and frequency domain, which is composed of the statistical features of time domain and the root mean square value (RMS) of the sub frequency band based on fast Fourier transform.

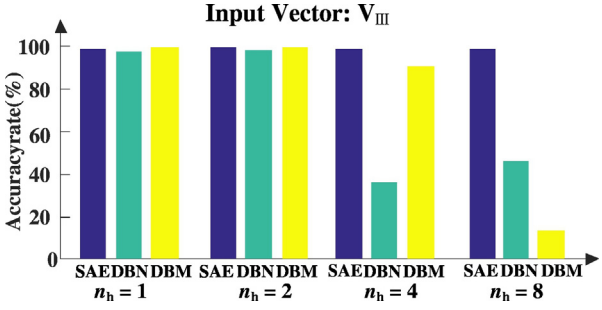

Fig. 8. Investigation of n_h for Scheme 3.

Please cite this article as: Z. Chen, et al., Deep neural networks-based rolling bearing fault diagnosis, Microelectronics Reliability (2017), http://dx.doi.org/10.1016/j.microrel.2017.03.006

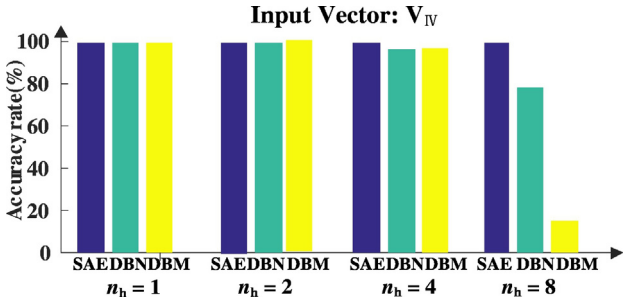

Fig. 9. Investigation of n_h for Scheme 4.

Statistical features of the time domain vibration signal can indicate the health condition of rotating machinery. In this paper, mean value, standard deviation, mean square root, kurtosis, shape factor, crest factor and impulse factor are chosen as time domain features. They are calculated by the formulation of Table 1, where $x(t)$ represents sample signal, N represents the sample length of vibration signal, and $\max(x(t))$ represents maximal value of signal.

Corresponding to the fast Fourier transform spectrum of a rolling bearing vibration signal, the RMS value of each frequency band can keep track of the energy in the spectrum peaks, thus, it can also reflect the health state of machinery. The RMS value of each sub frequency band is used as the representation of feature in frequency domain, which can be calculated as in (12) and (13).

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-i2\pi ft} dt \tag{12}$$

$$\mathrm{FFT}(i)_{\mathrm{RMS}} = \frac{\sqrt{\sum_{m=1}^{N} \mathrm{FFT}(m)^2}}{\sqrt{N}} \tag{13}$$

where $X(f)$ is the fast Fourier transform spectrum of a rolling bearing vibration signal, $\mathrm{FFT}(i)_{\mathrm{RMS}}$expresses the RMS value of the i-th sub frequency band, $FFT(m)$ represents the m-th spectrum line of the i-th sub frequency band, and N is the number of its spectrum lines.

V_{III} is formed as follows: N_{RMS} RMS values, mean value, standard deviation, mean square root, kurtosis, shape factor, crest factor and impulse factor. The feature vector V_{IV} is composed of the first and the second one.

4. Experiment setup

To validate the effectiveness of the deep learning models for rolling bearing fault diagnosis, a rolling bearing fault diagnosis experimental platform (fabricated by the lab of the Universidad Politécnica Salesiana, Ecuador) is carried out the experiments. Fig. 3 shows their fault simulator schematic, respectively. The signal collection parameters are described as follows:

(1) Sampling frequency: 50 kHz (for accelerometer).
(2) Length of each test: 20 Sec.

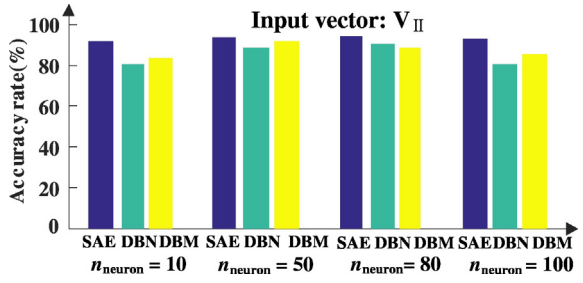

Fig. 10. Investigation of n_{nueron} for Scheme 2.

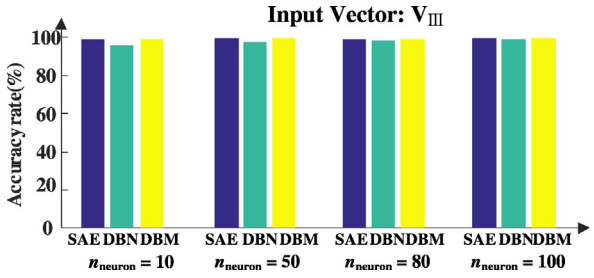

Fig. 11. Investigation of n_{nueron} for Scheme 3.

(3) Repetition of the tests: 5 times for each test (R1, R2, …, R5).
(4) Rotation Frequency: 8 Hz (F1), 10 Hz (F2), 15 Hz (F3).
(5) Loads: No flywheel (L1), 1 flywheel (L2), 2 flywheels (L3).

In this way, 7 kinds of the fault patterns of rolling bearing are defined in Table 2, and 315 (5*3*3*7) tests are performed. For each test, two accelerometers are used. Hence the total number of the dataset is 630. Fig. 4 shows the vibration signal of rolling bearing of 7 fault patterns defined in Table 2 under the condition of F3 and L3. Fig. 5 indicates their corresponding frequency spectrum by using fast Fourier transform.

5. Experimental results and analysis

In this section, the performance of three deep learning models for rolling bearing fault diagnosis based on four schemes defined in Section 3 is evaluated. The accuracy rate of the fault patterns classification is employed as the criteria for evaluation, which is defined as in (14).

$$\mathrm{AccuracyRate} = \frac{\mathrm{Thecorrectclassficationnumberofsamples}}{\mathrm{Thetotalnumberofsamples}} \tag{14}$$

The training mechanism of three deep neural network models includes two stages: the first stage is to use bottom-up unsupervised learning. This process can be regarded as a process of feature learning. The second stage is to use top-down supervised learning, which usually applies the gradient descent method to fine-tuning the whole network parameters. The number of hidden layer (n_h), the number of neurons of each hidden layer (n_{neuron}), the learning rate (lRate), the iterations of two training stages ($epoch_1$ and $epoch_2$), and the activation function of hidden layer (the default setting is sigmoid function) are the major parameters of DBM, DBN and SAE. To fairly evaluate them for rolling bearing diagnosis, these parameters are fine-tuned for each scheme.

5.1. Raw data-based evaluation

Scheme 1 is to use the raw vibration signals in time domain from the measurements as the input parameters of the deep neural network model. Based on the dataset described in Section 4, 12,600 samples are generated for Scheme 1, where each test generates 40 samples and the length of sample is 0.05 s (2500 sample points), which means

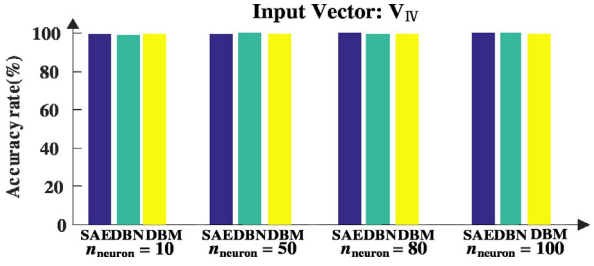

Fig. 12. Investigation of n_{neuron} for Scheme 4.

Please cite this article as: Z. Chen, et al., Deep neural networks-based rolling bearing fault diagnosis, Microelectronics Reliability (2017), http://dx.doi.org/10.1016/j.microrel.2017.03.006

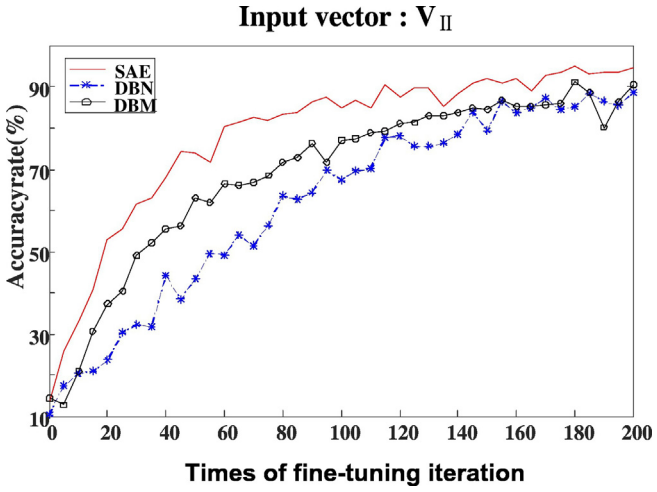

Fig. 13. Investigation of epoch$_2$ for Scheme 2.

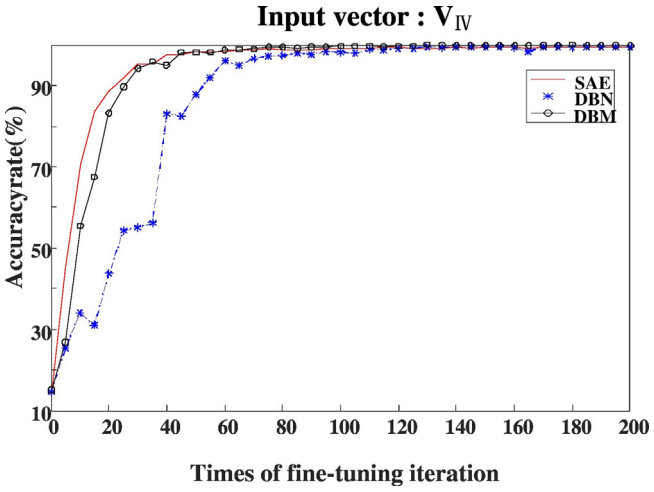

Fig. 15. Investigation of epoch$_2$ for Scheme 4.

the number of the neurons of the input layers is 2500. 80% of these samples which are used to train the neural networks, and the remaining samples are used as test ones. The parameter settings indicated in Table 3 are considered, respectively.

Base on the setting in Table 3, a large number of experiments were performed to fine-tune parameter setting for the raw data-based classification. Fig. 6 presents the best classification accuracy using DBM, DBN and SAE. It is very obvious that the raw data-based classification accuracy is not good, which means that the feature exaction is still necessary for deep neural networks to better identify the fault of rolling bearing. There may be exist optimized parameter setting for raw data-based classification, however, finding the most appropriate combination of parameters occurring in a DNN is a difficult task.

5.2. Preprocessing data-based evaluation

Schemes 2–4 are to extract the feature in time domain, frequency domain and time-frequency domain, and then these features are used as the input data of DBM, DBN and SAE, respectively. Based on the dataset described in Section 4, 6300 samples are generated for Schemes 2–4, where the length of sample is 0.5 s and each test generates 10 samples. 80% of these samples are used to train the neural networks, and the remaining samples are used as test ones. Parameter fine-tuning is still necessary for each scheme.

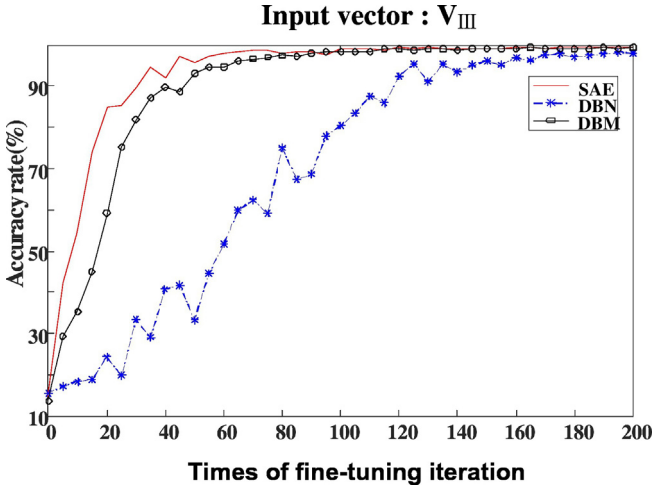

Fig. 14. Investigation of epoch$_2$ for Scheme 3.

The number of hidden layer (n_h), the neuron number of each hidden layer (n_{neuron}), and the iterations of two training stages (epoch$_1$ and epoch$_2$) have a great effect on the performance of deep neuron networks. So they are investigated in detail in the following. The default setting of these parameters is as follows: $n_h = 2$, $n_{neuron} = 50$, epoch$_1 = 1$, epoch$_2 = 200$, lRate $= 1.0$, and the activation function of hidden layers is sigmoid function. When discussing a certain parameter, the other ones use default setting.

There exists a kind of opinion that more hidden layers will bring better classification accuracy. In order to verify this issue, different setting of n_h from 1 to 8 were considered. Figs. 7–9 present some experiment results by using Schemes 2–4. As shown in Figs. 7–9, if the architecture gets deeper for each scheme, it will become more difficult to obtain good results. When $n_h = 1$ or 2, it is clear that the classification accuracy of each deep model will be the best for each scheme. When $n_h = 4$ or 8, DBN and DBM will obviously deteriorate.

Experimental evidence suggests that training deep architectures is more difficult than training shallow architectures. The deeper architecture of deep neural network does not necessarily lead to better results.

Figs. 10–12 reveal the effect of the neuron number of each hidden layer (n_{neuron}). It is not sensitive to vary the size of n_{neuron} for Schemes 3 and 4 shown in Figs. 11 and 12. As for Scheme 1, varying n_{neuron} of DBN and DBM has minor effect on the classification accuracy. So the following conclusion can be drew that it is not sensitive to vary the size of n_{neuron} for rolling bearing diagnosis by using DBN, SAE and DBM. n_{neuron} is set to 50 for all the following experiments.

Epoch$_1$ and epoch$_2$ represent the training time in the pre-training and fine-tuning stage of three DNN models, respectively, which also affects the performance of DNN-based classifier. If the training time is too long, it is possible to lead to "overfitting"; it also possibly results in a lack of training on the contrary. When epoch$_1$ is equal to 1, the good classification accuracy can be obtained. So the parameter epoch$_2$ is fine-tuned. Figs. 13–15 reveal the convergence process of the classification accuracy for each scheme by using DBM, DBN and SAE. Fig. 13 shows that when epoch$_2 = 200$, the classification accuracy of each DNN model is still less than 90% for Scheme 2. Fig. 14 shows that DBN is slower than DBM and SAE to reach 99% classification accuracy for

Table 4
Diagnostic testing results using different classification schemes.

Scheme	DBM	SAE	DBN
Scheme 1	16.12%	15.12%	15.77%
Scheme 2	88.62%	93.46%	86.15%
Scheme 3	99.23%	99.38%	98.31%
Scheme 4	99.15%	99.62%	99.38%

Please cite this article as: Z. Chen, et al., Deep neural networks-based rolling bearing fault diagnosis, Microelectronics Reliability (2017), http://dx.doi.org/10.1016/j.microrel.2017.03.006

Scheme 3; Fig. 15 shows that when $epoch_2 = 100$, DBM, DBN and SAE can be successful to reach 99% classification accuracy for Scheme 4.

5.3. Comparison and analysis

As is stated above, the following parameter setting can bring rather good classification accuracy for each scheme: $n_h = 2$, $n_{neuron} = 50$, $epoch_1 = 1$, $epoch_2 = 200$, lRate = 1.0, and the activation function of hidden layers is sigmoid function. Table 4 presents the diagnostic testing results using different classification schemes. As shown in Table 4, DBM, SAE and DBN are highly reliable and effective in fault diagnosis of rolling bearing. However, the raw data-based (Scheme 1) classification accuracy is not good. As a result, they still depend on the good feature extraction of vibration signal. Scheme 4 is the best scheme, which slightly better than Scheme 3. Scheme 2 is far more favored than Schemes 3 and 4. In other words, the feature extraction scheme including time domain, frequency and time-frequency domain feature is the best one. Among the three DNN models (DBM, SAE and DBN), SAE is slightly better than DBM and DBN. As shown in Fig. 15, DBN needs more training time.

6. Conclusions

In this paper, Deep Boltzmann Machines, Deep Belief Networks and Stacked Auto-Encoders were applied to diagnose the fault of rolling bearing. Four preprocessing schemes are discussed. These classification schemes are extensively evaluated, which are based on a large number of experiments. Some interesting findings from this study are given below:

(1) DBM, DBN and SAE are efficient in rolling bearing fault diagnosis, whose classification accuracy achieves more than 99%. These classifiers have a good potential to provide helpful maintenance guidelines for industrial systems.
(2) The raw data-based (Scheme 1) input is not good. Finding the most appropriate combination of parameters occurring in a DNN is a difficult task for the raw data-based classification. So feature extraction of vibration signal is still a necessary step for the DNN-based classifiers.
(3) The feature extraction scheme including time domain, frequency and time-frequency domain feature is superior to those single mode features such as Schemes 2 and 3.
(4) The deeper architecture of deep neural network does not necessarily lead to better results.

Acknowledgements

This work is supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission [No·KJ1500607, No·KJ1400629], Science Research Fund of National Research Base of Intelligent Manufacturing Service (950216061), the National Natural Science Foundation of China [61402063, 61502063].

References

[1] D. Wang, P.W. Tse, K.L. Tsui, An enhanced Kurto-gram method for fault diagnosis of rolling element bearings, Mech. Syst. Signal Process. 35 (1–2) (2013) 176–199.
[2] I.E.E.E. Jie Liu, Wilson Wang, Farid Golnaraghi, An enhanced diagnostic scheme for bearing condition monitoring, IEEE Trans. Instrum. Meas. 59 (2) (2010) 309–321.
[3] Q. Miao, Q.H. Zhou, Planetary gearbox vibration signal characteristics analysis and fault diagnosis, Shock. Vib. 2015 (2015), 126489. .
[4] Y. Wang, Z. He, J. Xiang, Y. Zi, Application of local mean decomposition to the surveillance and diagnostics of low-speed helical gearbox, Mech. Mach. Theory 47 (1) (2012) 62–73.
[5] Y. Lei, D. Kong, J. Lin, M.J. Zuo, Fault detection of planetary gearboxes using new diagnostic parameters, Meas. Sci. Technol. 23 (5) (2012), 055605. .
[6] S. Hou, M. Liang, Y. Zhang, C. Li, Vibration signal demodulation and bearing fault detection: a clustering-based segmentation method, Proc. IME C J. Mech. Eng. Sci. 228 (11) (2014) 1888–1899.
[7] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Signal Process. 20 (7) (2006) 1483–1510.
[8] L. Guo, J. Chen, X. Li, Rolling bearing fault classification based on envelope spectrum and support vector machine, J. Vib. Control. 15 (9) (2009) 1349–1363.
[9] Q. Miao, C. Tang, W. Liang, M. Pecht, Health assessment of cooling fan bearings using wavelet-based filtering, Sensors 13 (1) (2013) 274–291.
[10] R. Duda, P. Hart, D. Stork, Pattern Classification, Wiley, Hoboken, NJ, 2001.
[11] L.B. Jack, A.K. Nandi, A.C. McCormick, Diagnosis of rolling element bearing faults using radial basis function networks, Appl. Signal Process. 6 (1999) 25–32.
[12] B. Samanta, K.R. Al-Balushi, Artificial neural-network-based fault diagnostics of rolling-element bearings using time-domain features, Mech. Syst. Signal Process. 17 (2) (Mar. 2003) 317–328.
[13] R. Isermann, On fuzzy logic applications for automatic control, supervision, and fault diagnosis, IEEE Trans. Syst. Man Cybern. Syst. Humans 28 (2) (Mar. 1998) 221–235.
[14] G. Castellano, A. Fanelli, C. Mencar, An empirical risk functional to improve learning in a neuro fuzzy classifier, IEEE Trans. Syst. Man Cybern. B Cybern. 34 (1) (Feb. 2004) 725–731.
[15] O. Uluyol, K. Kim, E. Nwadiogbu, Synergistic use of soft computing technologies for fault detection in gas turbine engines, IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 36 (6) (Jul. 2006) 476–484.
[16] W. Wang, An intelligent system for machinery condition monitoring, IEEE Trans. Fuzzy Syst. 16 (1) (Feb. 2008) 110–122.
[17] W. Wang, An enhanced diagnostic system for gear system monitoring, IEEE Trans. Syst. Man Cybern. B Cybern. 38 (1) (Feb. 2008) 102–112.
[18] D. Wang, Q. Miao, Smoothness index-guided Bayesian inference for determining joint posterior probability distributions of anti-symmetric real Laplace wavelet parameters for identification of different bearing faults, J. Sound Vib. 345 (9) (2015) 250–266.
[19] D. Wang, Q. Miao, R. Kang, Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition, J. Sound Vib. 324 (3–5) (2009) 1141–1157.
[20] Q. Miao, V. Makis, Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models, Mech. Syst. Signal Process. 21 (2) (2007) 840–855.
[21] J.B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, F. Fnaiech, Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals, Appl. Acoust. 89 (2015) 16–27.
[22] Y. Li, M. Xu, Y. Wei, W. Huang, Bearing fault diagnosis based on adaptive multiscale fuzzy entropy and support vector machine, J. Vibroeng. 17 (3) (2015) 1188–1202.
[23] D.L. Souzal, M.H. Granzotto, G.M. Almeida, L.C. Oliveira-Lopes, Fault detection and diagnosis using support vector machines - a SVC and SVR comparison, J. Saf. Eng. 3 (1) (2014) 18–29.
[24] Q. Miao, L. Xie, H. Cui, W. Liang, M. Pecht, Remaining useful life prediction of Lithium-ion battery with unscented particle filter technique, Microelectron. Reliab. 53 (6) (2013) 805–810.
[25] Feng Jia, Yaguo Lei, Jing Lin, Xin Zhou, Na Lu, Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mech. Syst. Signal Process. 72-73 (2016) 303–315.
[26] Z.Q. Chen, C. Li, R.V. Sanchez, Gearbox fault identification and classification with convolutional neural networks, Shock. Vib. 2015 (2) (2015) 1–10.
[27] V.T. Tran, F.A. Thobiani, A. Ball, An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks, Expert Syst. Appl. 41 (9) (2014) 4113–4122.
[28] C. Li, R. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R. Vásquez, Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis, Neurocomputing 168 (2015) 119–127.
[29] Z. Chen, C. Li, R.V. Sánchez, Multi-layer neural network with deep belief network for gearbox fault diagnosis, J. Vibroeng. 17 (5) (2015) 2379–2392.
[30] R. Salakhutdinov, G. Hinton, Deep Boltzmann machines, J. Mach. Learn. Res. 5 (2) (2009) 448–455.
[31] G.E. Hinton, S. Osindero, Y. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (2006) 1527–1554.
[32] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.A. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, J. Mach. Learn. Res. 11 (2010) 3371–3408.
[33] Y. Freund, D. Haussler, Unsupervised learning of distributions on binary vectors using two layer networks, Technical Report UCSC-CRL-94-25, University of California, Santa Cruz, 1994.
[34] R. Salakhutdinov, G. Hinton, An efficient learning procedure for deep machines, Neural Comput. 24 (8) (2012) 1967–2006.
[35] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, in: B. Schölkopf, J. Platt, T. Hoffman (Eds.), Advances in Neural Information Processing Systems 19 (NIPS'06), MIT Press 2007, pp. 153–160.
[36] Y. Bengio, Learning deep architectures for AI, Found. Trends Mach. Learn. 2 (1) (2009) 1–127.

Please cite this article as: Z. Chen, et al., Deep neural networks-based rolling bearing fault diagnosis, Microelectronics Reliability (2017), http://dx.doi.org/10.1016/j.microrel.2017.03.006