Neurocomputing 335 (2019) 327–335

Contents lists available at ScienceDirect

Neurocomputing

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

A survey on Deep Learning based bearing fault diagnosis

Duy-Tang Hoang[a], Hee-Jun Kang[b,*]

[a] Graduate School of Electrical Engineering, University of Ulsan, Ulsan, South Korea
[b] School of Electrical Engineering, University of Ulsan, Ulsan, South Korea

ARTICLE INFO

Article history:
Received 7 March 2018
Revised 30 May 2018
Accepted 10 June 2018
Available online 2 November 2018

Keywords:
Bearing fault diagnosis
Deep Learning

ABSTRACT

Nowadays, Deep Learning is the most attractive research trend in the area of Machine Learning. With the ability of learning features from raw data by deep architectures with many layers of non-linear data processing units, Deep Learning has become a promising tool for intelligent bearing fault diagnosis. This survey paper intends to provide a systematic review of Deep Learning based bearing fault diagnosis. The three popular Deep Learning algorithms for bearing fault diagnosis including Autoencoder, Restricted Boltzmann Machine, and Convolutional Neural Network are briefly introduced. And their applications are reviewed through publications and research works on the area of bearing fault diagnosis. Further applications and challenges in this research area are also discussed.

1. Introduction

Deep Learning (DL) is a branch of Machine Learning (ML), where multiple layers of data processing units are assembled to form deep architectures to extract multiple levels of data abstraction. The concept of Deep Learning appeared in the 1980s, but it has become more popular recently because of two main reasons [1]:

- The increase of the computational ability of processing units, especially Graphics Processing Units (GPUs), while the cost is reducing: DL algorithms often require strong computation efforts. The low cost and high computational ability help to implement, train and perform DL algorithms more easily and quickly.
- The recent advances in ML research: Before the year of 2006, training deep architectures was very difficult and had unsuccessful result. A rational reason explaining for this problem is that the gradient-based optimization starting from random initialization of weight matrices often has poor results [2]. In 2006, Hinton et al. proposed an efficient method for training networks with deep structures, called greedy layer-wise training [3]. This work can be considered as the breakthrough which opened the fascinating era of ML research with deep architectures.

DL architectures such as Convolutional Neural Network (CNN), Stacked Autoencoder (SAE), Deep Belief Network (DBN), Deep Boltzmann Machine (DBM), and Recurrent Neural Network (RNN) have been applied successfully in many areas, includes computer vision [4,5], natural language processing [6,7], medical image analysis [8,9], driverless car [10,11], and machine health monitoring [12,13]. In the machine health monitoring area, bearing fault diagnosis is an important part because Rolling Element Bearings (REBs) are indispensable elements in rotary machines. REBs are not only the most critical components but also the main contributor to the system failures, 45–55% of equipment failure cases caused by broken of REBs [14,15]. Any unexpected failure of bearings may cause sudden breakdown of the machine, even of the entire system, leading to huge financial losses and time-consuming. Therefore, fault diagnosis of REBs is a major concern and has drawn substantial consideration of researchers.

The condition monitoring of a REB can be considered as a pattern recognition task which has been successfully solved by intelligent diagnosis methods. According to the current literature, a general intelligent diagnosis methodology includes four steps as follows: data acquisition, feature extraction, feature selection, and feature classification [16].

Data acquisition step collects signals which reflect the health status of bearings from sensor systems. These signals are vibration signals [17], acoustic emission signals [18], or electrical motor currents [19].

Feature extraction maps the original signals onto the statistical parameters which convey the information about the machine status. In the problem of pattern recognition, to obtain a high accuracy recognition result, the design of feature extractor takes a very important role [20]. Features of signals can be extracted from time domain [21], frequency domain [22], and time-frequency domain

* Corresponding author.
E-mail address: hjkang@ulsan.ac.kr (H.-J. Kang).

https://doi.org/10.1016/j.neucom.2018.06.078
0925-2312/

[23,24]. The original bearing fault signals collected from rotary machines are in time domain. Bearing fault signals are also can be investigated in frequency domain and time-frequency domain by using appropriate tool to transform them into the corresponding domains. Fourier Transform (FT) is the most popular signal processing tool using for transforming signal into frequency domain. Time-frequency domain features can be extracted by Wavelet Packet Transform (WPT) [25], Dual-tree Complex Wavelet Transform (DT-CWT) [26], and Short-time Fourier Transform (STFT) [27].

The extracted feature set often has high dimension so that it may contain some redundant and irrelevant features. High dimension feature set with many features may make it difficult to identify bearing faults. A large feature set often requires high computation effort, increases the learning time, and reduces the performance of classifiers. Thus, selecting the most discriminant features is an important step. Feature selection not only reduces the computation time but also increases the classification accuracy. Generally, there are two popular approaches for selecting features. The first approach is to generate a new feature set with lower dimension from the extracted feature set. Principal Component Analysis (PCA) [22] and Independent Component Analysis (ICA) [28] are two well-known methods of this approach. The second approach is to eliminate non-sensitive or useless features based on certain benchmarks. Sequential Selection [29] is a popular method of this approach.

Feature classification: once the salient features are selected, they are fed into an ML-based classifier such as k-Nearest Neighbor (kNN) [30], Artificial Neural Network (ANN) [31,32], and Support Vector Machine (SVM) [33] to identify the bearing fault. Among current ML based classification algorithms, ANN was proved as a powerful tool with high accuracy. There are many types of ANN were proposed such as Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Radial Basic Function Neural Network (RBFNN), Probabilistic Neural Networks (PNN), and Radial Basis Probabilistic Neural Network (RBPNN), which was developed from RBFNN and PNN [34].

The ML-based bearing fault diagnosis has been applied widely for decades, but this approach has some drawbacks. Designing a feature extractor requires human expertise and signal processing techniques. However, the feature extraction does not have a common procedure for every task [35]. Moreover, the shallow structures of conventional ML algorithms have very limited ability for learning the non-linear relation of extracted features [36].

DL algorithms can learn multiple layers of representations from input data by deep architectures with many layers of data processing units [37]. The output from a layer will be the input for its successive layer. Each layer can learn a higher level of data presentations from its preceding layer output. Therefore, DL architectures can automatically extract multiple complex features from the input data without human engineers: layers of features are extracted from raw data by a general purpose learning procedure [38]. With that ability, DL architectures possess the ability to deal with the difficulty of conventional ML. DL models have been applied successfully in a vast number of areas such as computer vision, voice recognition, signal processing, and medical image analysis. Not only in single modalities (e.g., text, images or audio), applications of DL for multiple modalities are also attracting attention of researchers [39]. Moreover, recently DL has been approaching the area of machine health monitoring, especially the bearing fault diagnosis topic since REBs play important roles in rotary machines. The intention of this paper is to make a systematic review of the achievements of DL based bearing fault diagnosis research, to highlight their contributions, and to discuss the challenges. This paper is constructed as follows. Section 2 briefly presents the background concepts of three most popular DL algorithms applied in bearing fault diagnosis, includes Autoencoder (AE), Restricted Boltzmann Machine (RBM), and Convolutional Neural Network (CNN). Section 3 reviews DL based bearing fault diagnosis publications. Section 4 concludes the paper and discusses further applications and challenges.

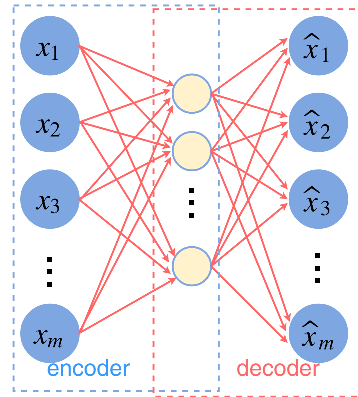

Fig. 1. Structure of AE.

2. Overview of Deep Learning

Up to date, there are a lot of DL structures applied in various areas. In this section, we supply a brief introduction about the three most popular and fundamental structures in DL.

2.1. Autoencoder and its deep models

2.1.1. Autoencoder

Autoencoder (AE) is an unsupervised DL algorithm. It was first proposed by Rumelhart et al. [40]. As shown in Fig. 1, an AE is a special neural network consists of three layers: input layer, hidden layer, and output layer. The difference is that in the structure of AE, the input and output layer have the same number of neurons.

The structure of an AE can be considered as an encoder which integrated with a decoder. The encoder includes the input layer and the hidden layer, mapping the input vector to the hidden layer. The decoder takes the output of hidden layer to recreate the input values. Consider an input vector x, the forward direction computation of AE includes two steps: encoding and decoding. The encoding step maps the input vector into the hidden layer:

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

and the decoding step tries to reconstruct input values from hidden values:

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

where W_e, b_e and W_d, b_d are respectively the weight matrix - bias vector of the encoder and decoder. $f(.)$ denotes the activation function. With an input set including m samples $x^i, i = 1 : m$, the AE will produce m output samples $\hat{x}^i, i = 1 : m$. The loss function is defined the following equation with squared error as:

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

or with the cross-entropy as:

$$J(W_e, W_d, b_e, b_d) = \frac{1}{m} \sum_{i=1}^{m} x^i \log(\hat{x}^i) + (1 - x_i) \log(1 - \hat{x}^i) \tag{4}$$

The training process minimizes the loss function and optimizes the AE parameters to reconstruct the output vector $\hat{x}$ so that the reconstruct error $(x - \hat{x})$ is as small as possible. After being trained, the AE can reconstruct the output from the original input with

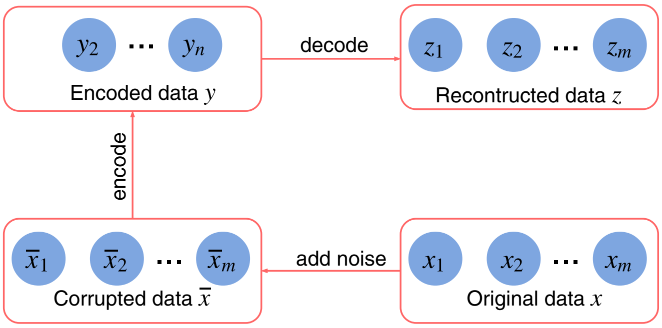

Fig. 2. Structure of DAE.

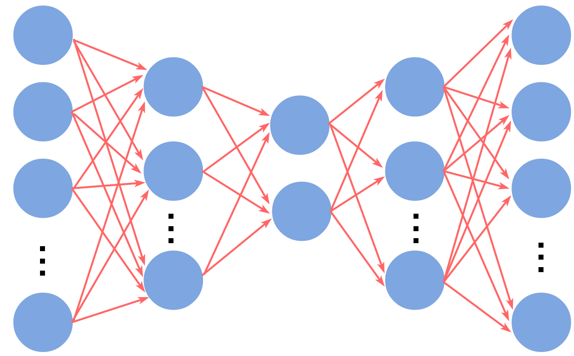

Fig. 3. Stacked Autoencoder.

an arbitrary accuracy [40]. Since the output is reconstructed from the hidden vector, we can say that the hidden vector is a representation of the input data, i.e., the AE has learned representative features from the original input data and mapped into the hidden vector.

2.1.2. Denoising Autoencoder

Denoising Autoencoder (DAE) first proposed by Vincent et al. [41] is another type of AE. A DAE has the same structure with the original AE, the difference is the way of feeding the input data into the network. First, the original input data is corrupted before being passed to the network. The procedure of encoding and decoding is similar to that of a typical AE. However, the DAE is trained to reconstruct the original input from the corrupted data version. The corrupted data can be produced by adding Gaussian noise, Masking noise, and Salt-and-pepper noise. The schematic diagram of the procedure of DAE is shown in Fig. 2.

The original input data x^i is added noise to obtain the corrupted data $\bar{x}^i$. The corrupted input data are mapped into the hidden layer by:

$$y^i = f(W_e \bar{x}^i + b_e) \tag{5}$$

The the decoder phase of the AE reconstructed the output by:

$$z^i = f(W_d y^i + b_d) \tag{6}$$

The loss function is calculated by:

$$J(W_e, W_d, b_e, b_d) = \frac{1}{2m} \sum_{i=1}^{m} (z^i - x^i)^2 \tag{7}$$

2.1.3. Stacked Autoencoder

A SAE is a deep model of AE. The architecture of a SAE is constructed by stacking multiple AEs together to form a deep model with many layers. The training process of an SAE called greedy layer-wise training proposed by Hinton et al. [3], a detail explanation and pseudocode of this algorithm was also supplied by Bengio et al. [2]. Consider a SAE constructed by n AE as shown in Fig. 3.

The encoding operation is described by the following equation:

$$a^k = f(W_e^k a^{k-1} + b_e^k), \ \ k = 1:n \tag{8}$$

where k denotes the kth AE. a^k denotes the encoding result of kth AE. When $k = 1, \ a^0 = \mathbf{x}$ is the input data. The decoding operation is described by the following equation:

$$c^k = f(W_d^{n-(k-1)} c^{k-1} + b_d^{n-(k-1)}), \ \ k = 1:n \tag{9}$$

when $k = 1, \ c^0 = a^n$, when $k = n, \ c^n = \hat{\mathbf{x}}$ is the reconstructed data of the input data $\mathbf{x}$.

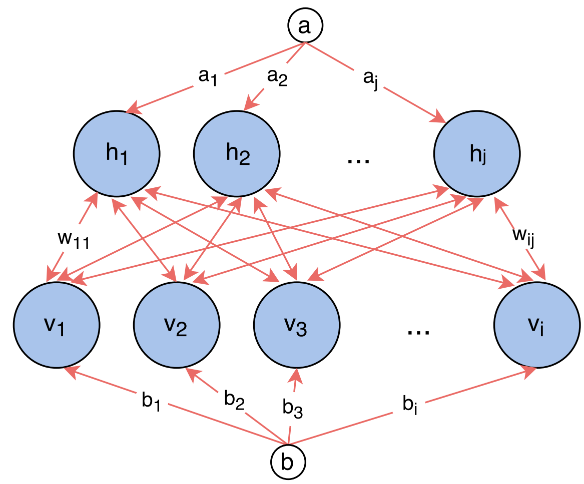

Fig. 4. Structure of RBM.

2.2. Restricted Boltzmann Machine and its deep models

2.2.1. Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM) is an unsupervised learning algorithm introduced by Smolensky [42]. The structure of a RBM is shown in Fig. 4.

In the visible layer, visible nodes are denoted by v_i. In the hidden layer, hidden nodes are denoted by h_j . Nodes in the same layers are not connected together. The weight which connects node v_i and h_j is denoted by w_{ij}. Each node has its own bias value. Visible node v_i has the corresponding bias value b_i ; hidden node h_j has the corresponding bias value c_j. The relationship of the visible layer and hidden layer is defined by energy function given by:

$$E(v, h) = -\sum_i v_i b_i - \sum_j h_j c_j - \sum_{i,j} v_i h_j w_{ij} \tag{10}$$

The eventual goal of RBM training is to optimize the parameter set $\theta = w_{ij}, b_i, c_j$ that minimizes the model energy and balances the model at a finite state. Every possible case of a pair visible–hidden node is assigned a probability by the energy function:

$$p(v, h) = \frac{1}{Z} e^{-E(v,h)} \tag{11}$$

where Z is the partition function, calculated by summing all possible visible–hidden node pairs:

$$Z = \sum_{v,h} e^{-E(v,h)} \tag{12}$$

Since nodes in the same layer are not connected, the conditional probability distributions of each unit are described by the

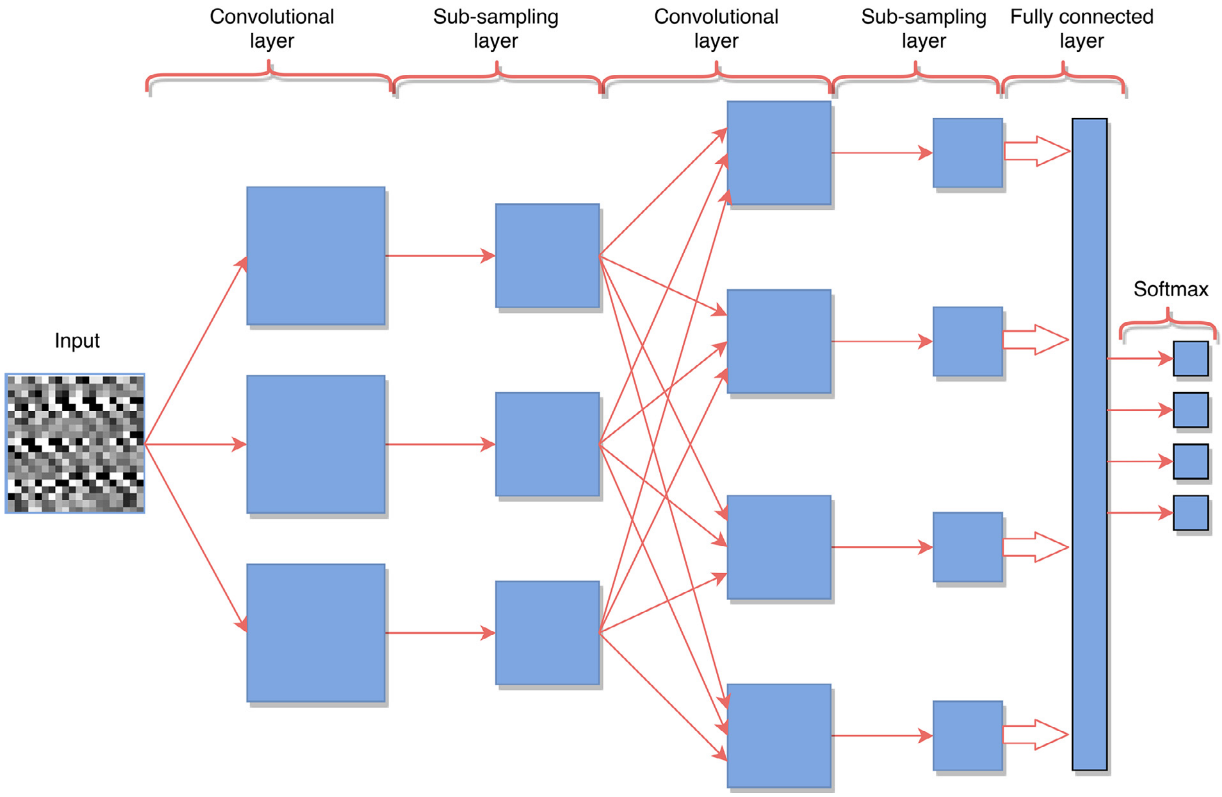

Fig. 5. Convolutional Neural Network.

following equations:

$$p(v_i = 1|h) = \frac{1}{1 + exp(-c_j - \sum_j w_{ij}h_j)} \tag{13}$$

$$p(h_j = 1|v) = \frac{1}{1 + exp(-b_i - \sum_i w_{ij}v_i)} \tag{14}$$

The training process for RBM is to maximize the joint probability. The parameters can be trained by Contrastive Divergence (CD) algorithm proposed by Hinton [43].

2.2.2. Deep models of Restricted Boltzmann Machine

DBN proposed by Hinton [44] is a deep model constructed by stacking multiple RBMs, where the input of a layer is the output of the preceded layer. DBNs can also be trained with the greedy layer-wise training [2].

DBM introduced by Salakhutdinov and Hinton [45] is a deep model with many hidden layers stacked into a hierarchy structure. The difference between a DBM and a DBN is that a DBM is an undirected model while a DBN is a directed model. DBMs can also be trained efficiently by the greed layer-wise algorithm [46].

2.3. Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of feed-forward neural network which constructed by three types of layer: Convolutional Layer (CL), Pooling Layer (PL), and Fully-connected Layer. The Fully-connected Layer has the same structure and the way of operation with the conventional feed-forward neural network. Advantages of CNN come from the differences in structures and operations of CLs and PLs. An illustration of CNN structure for processing 2-D data is shown in Fig. 5.

A CL consists of multiple learnable kernels. Each kernel has a trainable weight and bias. The CL convolves the input data with kernels in that layer. The result of the convolution operation then will be fed into an activate function to produce the final output of that CL. The math operation in the lth layer between the specific jth and the input data x^{l-1} can be described by the following equation:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_j^l + b_j^l\right) \tag{15}$$

where $(*)$ represents the convolution operation. The explanation for the above equation is as follows. Assume that the input data x^{l-1} includes m 2-D matrices. Every input matrix $x_i^{l-1}(\ i \in m)$ is convolved with the kernel k_j. Then the sum of all convolution operation results will be added with the bias. Finally, the result will be fed into the activate function f to produce the final output of kernel j. The purpose of CL is to extract local features from input data. Weight sharing is a term often used for the way of using kernels for all input maps of data.

After each CL, there is a PL. The purpose of a PL is to reduce the spatial size of the feature maps produced by the preceded layer. The operation of a PL is a down-sampling operation which exploits max-sampling or average-sampling.

3. Deep Learning applications in bearing fault diagnosis

3.1. Autoencoder

AE and its deep models are unsupervised learning algorithms. They are promising methods to overcome the difficulties of feature extraction by learning the features in high-level representations automatically. Moreover, with quite simple structure and easy to implement, AE is the most popular algorithm of DL applied in bearing health monitoring.

Jia et al. [36] stacked multiple AEs to extract features from raw bearing vibration signals. The extracted features then were passed through a softmax regression to recognize fault types of bearing. This method used raw vibration data directly, but the performance was not good because with the complex signal the original AE with MSE loss function is lack of robustness [47]. To overcome that drawback of AE, Shao et al. [48] proposed a novel loss function for AE by adopting maximum correntropy, which is a nonlinear and local similarity measure, insensitive for complex and non-stationary background noises [49]. The hyper-parameters of the SAE were selected by exploiting the Artificial Fish Swarm Optimization.

The original AE and its deep stacked models have another drawback. They cannot guarantee the usefulness of extracted features [50]. Xia et al. [51] proposed the method of automatically extracting features by using stacked DAE. Masking noise [41] was used to corrupt the input data by forcing a portion of data to be zero. Also used DAE, but instead of corrupting only the input data, Lu et al. [52] applied the data destruction process in all layer of SAE.

Shao et al., in their another publication [53], proposed an enhanced deep AE model from the combination of DAE and Contrastive AE (CAE) [54]. First, low-level features are extracted from the raw vibration signals by using a DAE. Then, high-level features are extracted by CAE from the low-level features. Since the DAE is trained by the corrupted version of the original vibration signals, it is more robust than the original AE. The corrupting data process was done by adding the Gaussian noise to the original signals.

In above publications, raw vibration signals in time domain were directly used as input data of AE deep models. Some other authors used AE models to extract features from processed signals. Liu et al. [55] exploited stacked sparse AE to learn features from spectrogram data of recorded sound signals from fault bearing. The spectrogram data was produced by STFT. Also using data in frequency domain with a similar deep AE model, Junbo et al. [56] employed digital wavelet frame to denoise vibration signals before the feature learning process. Thirukovalluru et al. [57] used the low-level features in frequency domain and time-frequency domain. Power spectrum of vibration signals was calculated by FFT. The power spectrum was split into 256 equal bins. From every single bin, the energy ratio to total signal energy was calculated to form the corresponding feature. Features in time-frequency domain were extracted by WPT. Features were extracted from WPT tree by calculating the energy of each node. Finally, low-level features from both domain were used to trained SDA to learn high-level features. Guo et al. [58] exploited features from three domains to build a low-level feature fusion. Then a deep model constructed by stacking sparse AEs was employed to extract high-level features.

Mao et al. [59] proposed a deep model called Autoencoder-Extreme Learning Machine (AE-ELM) to address the high computation issue of deep AE models. ELM is an efficient training algorithm for training single-hidden layer feed-forward neural network [60]. Not only fast learning and low computation, AE-ELM also has high classification accuracy [61,62].

Winner-take-all Autoencoder (WTA-AE) was introduced by Makhzani and Frey [63] to address the issues of sparse AE. Those issues are as follows. To form a sparse AE, the loss function of a normal AE is modified by adding a penalty function proportional to the KL divergence. This method is only suitable for certain target sparsity and hard to select proper parameters. Moreover, the KL divergence only works with sigmoidal activation function, and it is unclear how to apply with ReLU AE. In the proposed WTA-AE, only the target sparsity rate needs to be turned. Moreover, WTA-AE can work with any target sparsity rate and can be trained very fast. Li et al. [64] proposed a deep model using a modified version of WTA-AE. Their model was modified by adding a penalty term on the hidden layer and a penalty term on weights.

Jia et al. [65] addressed two shortcomings of traditional AE including (1) AEs may extract similar features and (2) extracted features suffer from shift variant property. To overcome these drawbacks, they proposed a Local Connection Network (LCN) constructed by Normalized Sparse Autoencoder (NSAE), called NSAE-LCN. In their approach, NSAE is a modification of sparse AE. The differences include: (1) use ReLU as the activation function instead of sigmoidal function since ReLU function is more efficient in training process [66], (2) get rid of bias because bias is ineffective in feature extraction [67], (3) replace KL divergence function by norm, and (4) add to the loss function a soft orthonormal constraint.

In industrial production where load and environment vary quite often, machines and bearings must work in various states or in multimode. For the fault diagnosis in multimode, the characteristics of data in each mode are different. Zhou et al. [68] showed that in bearing fault classification, considering a multimode as a single mode will lead to low accuracy since the feature extraction is inaccurate. Zhou et al. introduced a deep AE model to solve the mentioned problem in multimode bearing fault classification. At first, an SAE model was trained to mode partition. Then, a set of another SAEs was constructed for observing data in each mode, the goal of this set is to determine the corresponding mode of given data. Finally, with the determined mode, another SAE was built to classify the type of bearing fault.

3.2. Restricted Boltzmann Machine

In this section, we review publications in which deep models of RBM were employed for bearing fault diagnosis. In an introductory paper, Chen et al. [69] introduced two feature extraction methods for bearing fault diagnosis, one used DBM and another one used DBN. They carried out four different experiments with four separate sets of low-level features:

(1) Raw time domain vibration signals directly.
(2) Energy spectrum features in time-frequency domain.
(3) Features from time domain and frequency domain.
(4) Features from all domains.

With the first feature set with raw time domain signals, the classification accuracy was poor. The 4th feature set which was a combination of low-level features of three domain generated the best performance. Deng et al. [70] implemented similar experiments with different test-bed. Three sets of low-level features were extracted before learning high-level features by a DBM. With the combining set of all features from the three domains, DBM led to the best performance. Li et al. [71] conducted experiments for gearbox and bearing fault data, with a DBM model to learn high-level features from low-level feature set extracted from all of three domains. FFT and WPT respectively were employed to represent raw signals in frequency domain and time-frequency domain. In each domain, nine statistical features were extracted. They verified that the selection of low-level feature domain has a profound effect on deep statistical feature learning with DBM.

Gan et al. [72] combined several DBNs layer by layer to form a deep model for feature learning of mechanical system, called Hierarchical Deep Network (HDN). Energy features were extracted from time-frequency domain by using WPT. The HDN was constructed by two layers of DBNs. The first layer used low-level features to identify location of the faults. The second layer used the result of the first layer to identify the size of bearing fault.

Shao et al. [73] used PSO to optimize DBN structure for bearing fault diagnosis. Their DBN model with three hidden layers was constructed by stacked RBMs. First, DBN was pre-trained and fine-tuned by greedy layer-wise training with low-level features extracted in time domain. Then PSO algorithm was exploited to select hyper-parameters including the size of hidden layers, the learning rate, and the momentum coefficient. In their another publication [74], they presented an adaptive DBN. First, the raw vibration signals were decomposed into 8 components with different frequency bands by dual-tree complex wavelet transform. Then from each component, nine statistical features were extracted. These low-level features would be used to learn high-level features by the adaptive DBN. In the training process of their model, the learning rule and momentum coefficient can be altered at each training loop. Start with a small value of learning rate, if the reconstruction error of the next training loop increases, then the learning rate was reduced and vice versa.

Tao et al. [75] considered the fault diagnosis system with multisignal fusion. Compared to a single sensor system, multisignal fusion technique can provide more accurate and high sensitive fault features [76]. Raw vibration signals from all sensors are processed to extracted statistical features in the time domain. Learning high-level features were conducted by DBN model. The experiment results verified that DBN can handle multisignal fusion with higher accuracy compared to traditional machine learning algorithms.

3.3. Convolutional Neural Network

CNN can directly extract sensitive features from data in high dimensionality, such as images and videos. Vibration signals (1-D data) in time domain can be presented in a 2-D format such as a matrix form or frequency spectrum image. Motivated by those facts, some researched tried to converted vibration signal to 2-D form to exploit the ability of CNN. Janssens et al. [77] proposed a feature learning by a CNN model. Vibration signals from two sensors were used. First, the original vibration signals were represented in the frequency domain by DFT. Then a pair of frequency spectrum of signals from two single sensors were stacked together to construct a single sample in a matrix form.

Guo et al. [78] proposed hierarchical CNN with adaptive learning rate to classify bearing faults and further identify their severity. Their hierarchical model included two layers. The first layer was a CNN, the task of this layer was to identify the type of bearing faults. The second layer consisted of three CNN determined the size of fault occurred in bearing. Normally, the learning rate does not change during the training process. However, too small learning rate slows down the training process, too big learning rate may increase loss error and make the learning process not converge. To address this problem, an adaptive learning rate was proposed. Raw vibration signals in time domain were used directly but each sample was rearranged in a square matrix form. The experiment results validated that their hierarchical deep model outperformed typical CNN model.

Ding and He [79] transformed bearing fault diagnosis problem into an image classification problem. First, vibration signals in time domain were converted into image form by using WPT combined with phase space reconstruction. This type of image is called WPE image which contains the local relationship of WP nodes and energy fluctuation of the vibration signal. Their model included a multiscale layer which is a combination of the last CL with its preceded PL. In a normal CNN model, features learned by the last CL are robust but lack of precise details [80]. By the way of using multiscale layer, the last feature layer can preserve the global and local information simultaneously.

In the above publications, the CNN model required data in 2-D form. In [78], 32 segments of the signal in time domain were stacked together to form a matrix. In [77], frequency spectrums of signals from two single sensors were stacked together to form a matrix. Ding and He constructed WPE from vibration signals. Some other authors proposed approaches with CNN models which used 1-D data in time domain directly. In [81], Eren applied one-dimensional (1-D) CNN for vibration signal of bearing, while in [82], Ince et al. used motor current signals. Requiring only 1-D convolutions, one-dimensional CNN is very effective in computation and can be easy implemented feasibly and cheaply on hardware system [83]. In [84], Wang et al. also used 1-D data with a CNN model, but their model had parameters optimized by PSO algorithm. Two experiments were conducted, the first one used raw vibration signals in time domain directly, while the seconds one adapted WPT to extract low-level features in the time-frequency domain. Their experiments showed that by learning high-level features from low-level ones, the accuracy of fault diagnosis is higher. Zhang et al. [85] proposed a deep model using Wide First-layer Kernels for CNN (WDCNN) to deal with raw vibration data directly. The first CL consisting of wide kernels were used to extract features. With wide kernels, high-frequency noise can be suppressed easier than small kernels. In their model, Batch normalization layers were added after the CLs to reduce the shift of internal covariance and reduce the training time. In their another publication [86], Zhang et al. introduced a training method for their CNN model to diagnose bearing faults of machines which work under noisy environments. The dropout technique, a simple method to avoid overfitting [87], was integrated in the first CL. At each batch of training, the dropout rate was selected randomly. In addition, ensemble learning based on majority voting [88] was employed to improve the accuracy and stability of their model.

3.4. Other types of deep model

There are some publications in which deep models do not just base on a single DL algorithm. This section will review those models which are combinations of several DL algorithms such as AE, RBM, and CNN.

To deal with a multisensory system for bearing fault diagnosis, Chen and Li proposed a deep model called SAE-DBN which is a combination of sparse AE and DBN [89]. First, statistical features were extracted from time domain and frequency domain. Then these low-level features were fed into a SAE to learn high-level features. Finally, a DBN was employed to identify fault of bearing by using features learned by SAE.

Zhang et al. [90] introduced a deep model which is similar to a traditional NN with many hidden layers, using raw vibration signals. Moreover, their model also considered the temporal coherence of signals. Convolutional DBN (CDBN) was constructed by stacking several Convolutional Restricted Boltzmann Machines (CRBM) together [91]. This type of deep structure was proved to have better performance compared to typical CNN and DBN [92]. An improved model of CBDN was proposed by Shao et al. [93]. In their method, two advantages were included as follows. First, compressed sensing was exploited to compress the bearing vibration data. Second, exponential moving average technique was employed to modify the learning process.

4. Conclusion

Through this paper, a systematic overview of bearing fault diagnosis with DL algorithms has been presented, includes AE, RBM, and CNN. For a long time, conventional ML has been widely applied in bearing fault diagnosis. However, the performance of this approach highly requires hand-craft feature extraction, expert knowledge, and human labor. That makes the task of creating an automatic fault diagnosis model impossible. DL with the ability of automatically learning multiple levels of features and data abstraction has been considered to possess the potential to solve the drawback of conventional ML. Until now, AE has been the most popular algorithm of DL applied for bearing fault diagnosis. The reason is that AE has a simple structure as a typical NN. Moreover, even the deep models of AE such as SAE and SDA can be simply constructed and well trained by the greedy layer-wise training method. RBM is also an unsupervised learning algorithm like AE in the family DL. RBM has become popular since the invention of CD algorithm. RBM based deep models can also be trained by the greedy layer-wise training. AE and RBM based deep models were used as unsupervised learning, using unlabeled data to learn high-level salient features from fault data. CNN, a supervised learning algorithm is the third DL architecture widely applied in bearing fault diagnosis. Using CNN models, there was a novel approach when bearing vibration signals are presented in 2-D form beside the approach working directly with 1-D signals. CNN models are integrated both feature extractors and feature classifiers in their structures. Their training process is supervised learning which requires labeled data. In cases of AE and RBM, theirs deep models are used as feature extractors. The training processes of AE and RBM based deep models are unsupervised learning which has two phases: pre-training phase and fine-turning phase. RNN is also an important deep structure in the DL family. RNN has memory and can process arbitrary sequences of input patterns, in a sense RNN is the deepest model [94]. RNN models have been applied in many areas. However, until now there are not much publications of RNN on bearing fault diagnosis. That is the reason we did not mention about RNN in the previous review section. Signals from bearings are time series data in nature, so RNN is also a promising tool for bearing fault diagnosis.

Through the previous overview sections, we can see that there are two approaches to process raw signals before feeding into the DL models. The first approach uses deep models with raw signals in time domain directly, while the second one employs signal processing technique to extract low-level features or transforming raw signals into other forms before feeding into a deep model. The low-level features can be extracted from the time domain, frequency domain, and time-frequency domain. In some publications using models constructed from CNN, raw signals were transformed into 2-D forms, such as square matrix or image form. Obviously, this approach still requires signal processing techniques, expert knowledge, and human labor. In other words, this approach cannot create an automatic process for diagnosing bearing fault, although DL architectures were exploited and the performance is better than traditional ML based methods. The first approach which deals with raw signals successfully created the end-to-end systems which can automatically diagnose bearing fault.

Obviously, the effectiveness of DL in bearing fault diagnosis is undeniable. A vast number of deep models have been proposed and verified with various types of bearing fault signals. However, all the above publications stopped at the problem of a single fault, while the more challenging case, the multiple fault bearing diagnosis has not been verified yet. We can see that most of the experiments were conducted with the public bearing data of Case Western Reverse University [95] and the dataset provided by the University of Cincinnati [96], some others were carried out with self-established test-bed in laboratory scope, all of them are single fault bearing test-bed. In real applications many types of bearing faults may appear, a wide variety of bearing faults occur, some faults may appear at the same time. Thus, more efforts should be done to verify the performance of DL in multiple fault bearing diagnosis.

Compared with conventional ML architectures, DL models can learn salient features from data more easily, but it is not easy to design an appropriate deep model for any specific diagnosis task. Since each hyper-parameter has a reasonable effect on the model performance. Through the above-reviewed publications, we can see that hyper-parameters were almost chosen by trial and error method. This is a time-consuming work and requires experiences. Designing a deep model is still a challenge in DL research when there is still no standard way to select appropriate hype-parameters.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF-2016R1D1A3B03930496) funded by the Ministry of Education.

References

[1] L. Deng, A tutorial survey of architectures, algorithms, and applications for deep learning, APSIPA Trans. Sig. Inf. Process. 3 (2014) e2.
[2] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2007, pp. 153–160.
[3] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.
[4] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[5] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Proceedings of the AAAI, 2017, pp. 4278–4284.
[6] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, S. Khudanpur, Recurrent neural network based language model, in: Proceedings of the Interspeech, 2, 2010, p. 3.
[7] R. Collobert, J. Weston, A unified architecture for natural language processing: deep neural networks with multitask learning, in: Proceedings of the Twenty–Fifth International Conference on Machine Learning, ACM, 2008, pp. 160–167.
[8] H. Greenspan, B. van Ginneken, R.M. Summers, Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique, IEEE Trans. Med. Imag. 35 (5) (2016) 1153–1159.
[9] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, I. Eric, C. Chang, Deep learning of feature representation with multiple instance learning for medical image analysis, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2014, pp. 1626–1630.
[10] C. Chen, A. Seff, A. Kornhauser, J. Xiao, Deepdriving: learning affordance for direct perception in autonomous driving, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2722–2730.
[11] Q. Rao, J. Frtunikj, Deep learning for self-driving cars: chances and challenges, in: 2018 IEEE/ACM 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIAS), 2018, pp. 35–38.
[12] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Sig. Process. 115 (2019) 213–237.
[13] P. Tamilselvan, P. Wang, Failure diagnosis using deep belief learning based health state classification, Reliab. Eng. Syst. Saf. 115 (2013) 124–135.
[14] C. Lanham, Understanding the Tests that are Recommended for Electric Motor Predictive Maintenance, Baker Instrument Company, 2002.
[15] S. Nandi, H.A. Toliyat, X. Li, Condition monitoring and fault diagnosis of electrical motors-a review, IEEE Trans. Energy Convers. 20 (4) (2005) 719–729.
[16] D.-S. Huang, Systematic Theory of Neural Networks for Pattern Recognition, 201, Publishing House of Electronic Industry of China, Beijing, 1996.
[17] J. Zarei, M.A. Tajeddini, H.R. Karimi, Vibration analysis for bearing fault detection and classification using an intelligent filter, Mechatronics 24 (2) (2014) 151–157.
[18] J.L.F. Chacon, V. Kappatos, W. Balachandran, T.-H. Gan, A novel approach for incipient defect detection in rolling bearings using acoustic emission technique, Appl. Acoust. 89 (2015) 88–100.
[19] S. Singh, A. Kumar, N. Kumar, Motor current signature analysis for bearing fault detection in mechanical systems, Proced. Mater. Sci. 6 (2014) 171–177.
[20] Z.-Q. Zhao, D.-S. Huang, B.-Y. Sun, Human face recognition based on multi-features using neural networks committee, Pattern Recognit. Lett. 25 (12) (2004) 1351–1358.
[21] B. Samanta, K. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, Mech. Syst. Sig. Process. 17 (2) (2003) 317–328.
[22] A. Malhi, R.X. Gao, PCA-based feature selection scheme for machine defect classification, IEEE Trans. Instrum. Meas. 53 (6) (2004) 1517–1525.
[23] G.G. Yen, K.-C. Lin, Wavelet packet feature extraction for vibration monitoring, IEEE Trans. Ind. Electron. 47 (3) (2000) 650–667.
[24] X. Lou, K.A. Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mech. Syst. Sig. Process. 18 (5) (2004) 1077–1095.
[25] F. Hemmati, W. Orfali, M.S. Gadala, Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation, Appl. Acoust. 104 (2016) 101–118.
[26] M. Van, H.-J. Kang, Two-stage feature selection for bearing fault diagnosis based on dual-tree complex wavelet transform and empirical mode decomposition, Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. 230 (2) (2016) 291–302.

[27] C. Li, V. Sanchez, G. Zurita, M.C. Lozada, D. Cabrera, Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time–frequency ridge enhancement, ISA Trans. 60 (2016) 274–284.

[28] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Netw. 13 (4) (2000) 411–430.

[29] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (Mar) (2003) 1157–1182.

[30] P. Baraldi, F. Cannarile, F. Di Maio, E. Zio, Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions, Eng. Appl. Artif. Intell. 56 (2016) 1–13.

[31] L.S. Dhamande, M.B. Chaudhari, Bearing fault diagnosis based on statistical feature extraction in time and frequency domain and neural network, Int. J. Vehicle Struct. Syst. 8 (4) (2016) 229–240.

[32] D.-S. Huang, A constructive approach for finding arbitrary roots of polynomials by neural networks, IEEE Trans. Neural Netw. 15 (2) (2004) 477–491.

[33] Y. Li, M. Xu, Y. Wei, W. Huang, A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree, Measurement 77 (2016) 80–94.

[34] D.-S. Huang, J.-X. Du, A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks, IEEE Trans. Neural Netw. 19 (12) (2008) 2099–2115.

[35] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R.E. Vásquez, Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis, Neurocomputing 168 (2015) 119–127.

[36] F. Jia, Y. Lei, J. Lin, X. Zhou, N. 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 (2016) 303–315.

[37] L. Deng, D. Yu, et al., Deep learning: methods and applications, Found. Trends Sig. Process. 7 (3–4) (2014) 197–387.

[38] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[39] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A.Y. Ng, Multimodal deep learning, in: Proceedings of the Twenty-eighth International Conference on Machine Learning (ICML-11), 2011, pp. 689–696.

[40] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Cognit. Model. 323 (6088) (1986) 533.

[41] P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol, Extracting and composing robust features with denoising autoencoders, in: Proceedings of the Twenty-fifth International Conference on Machine Learning, ACM, 2008, pp. 1096–1103.

[42] P. Smolensky, Information processing in dynamical systems: foundations of harmony theory, 1986. Technical Report

[43] G. Hinton, A practical guide to training restricted Boltzmann machines, Momentum 9 (1) (2010) 926–947.

[44] G.E. Hinton, Deep belief networks, Scholarpedia 4 (5) (2009) 5947.

[45] R. Salakhutdinov, G. Hinton, Deep Boltzmann machines, in: Proceedings of the Artificial Intelligence and Statistics, 2009, pp. 448–455.

[46] R. Salakhutdinov, H. Larochelle, Efficient learning of deep Boltzmann machines, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 693–700.

[47] T. Amaral, L.M. Silva, L.A. Alexandre, C. Kandaswamy, J.M. Santos, J.M. de Sá, Using different cost functions to train stacked auto-encoders, in: Proceedings of the Twelfth Mexican International Conference on Artificial Intelligence (MICAI), IEEE, 2013, pp. 114–120.

[48] H. Shao, H. Jiang, H. Zhao, F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mech. Syst. Sig. Process. 95 (2017) 187–204.

[49] W. Ma, H. Qu, G. Gui, L. Xu, J. Zhao, B. Chen, Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments, J. Frankl. Inst. 352 (7) (2015) 2708–2727.

[50] 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.

[51] M. Xia, T. Li, L. Liu, L. Xu, C.W. de Silva, Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder, IET Sci. Meas. Technol. 11 (2017) 687–695.

[52] C. Lu, Z.-Y. Wang, W.-L. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, Sig. Process. 130 (2017) 377–388.

[53] H. Shao, H. Jiang, F. Wang, H. Zhao, An enhancement deep feature fusion method for rotating machinery fault diagnosis, Knowl. Based Syst. 119 (2017) 200–220.

[54] S. Rifai, P. Vincent, X. Muller, X. Glorot, Y. Bengio, Contractive auto-encoders: explicit invariance during feature extraction, in: Proceedings of the Twenty-eighth International Conference on Machine Learning (ICML-11), 2011, pp. 833–840.

[55] H. Liu, L. Li, J. Ma, Rolling bearing fault diagnosis based on STFT-deep learning and sound signals, Shock Vib. 2016 (2016) 1–12.

[56] T. Junbo, L. Weining, A. Juneng, W. Xueqian, Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder, in: Proceedings of the twenty-seventh Chinese Control and Decision Conference (CCDC), IEEE, 2015, pp. 4608–4613.

[57] R. Thirukovalluru, S. Dixit, R.K. Sevakula, N.K. Verma, A. Salour, Generating feature sets for fault diagnosis using denoising stacked auto-encoder, in: Proceedings of the IEEE International Conference on Prognostics and Health Management (ICPHM), IEEE, 2016, pp. 1–7.

[58] L. Guo, H. Gao, H. Huang, X. He, S. Li, Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring, Shock Vib. 2016 (2016) 1–10.

[59] W. Mao, J. He, Y. Li, Y. Yan, Bearing fault diagnosis with auto-encoder extreme learning machine: a comparative study, Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. 231 (8) (2017) 1560–1578.

[60] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1) (2006) 489–501.

[61] J. Tang, C. Deng, G.-B. Huang, Extreme learning machine for multilayer perceptron, IEEE Trans. Neural Netw. Learn. Syst. 27 (4) (2016) 809–821.

[62] L.L.C. Kasun, H. Zhou, G.-B. Huang, C.M. Vong, Representational learning with extreme learning machine for big data, IEEE Intell. Syst. 28 (6) (2013) 31–34.

[63] A. Makhzani, B.J. Frey, Winner-take-all autoencoders, in: Proceedings of the Advances in Neural Information Processing Systems, 2015, pp. 2791–2799.

[64] C. Li, W. Zhang, G. Peng, S. Liu, Bearing fault diagnosis using fully-connected winner-take-all autoencoder, IEEE Access 6 (2018) 6013–6115.

[65] F. Jia, Y. Lei, L. Guo, J. Lin, S. Xing, A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines, Neurocomputing 272 (2018) 619–628.

[66] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the twenty-seventh International Conference on Machine Learning (ICML-10), 2010, pp. 807–814.

[67] K. Konda, R. Memisevic, D. Krueger, Zero-bias autoencoders and the benefits of co-adapting features, Stat 1050 (2014) arXiv preprint arXiv:1402.3337.

[68] F. Zhou, Y. Gao, C. Wen, A novel multimode fault classification method based on deep learning, J. Control Sci. Eng. 2017 (2017) 1–14.

[69] Z. Chen, S. Deng, X. Chen, C. Li, R.-V. Sanchez, H. Qin, Deep neural networks-based rolling bearing fault diagnosis, Microelectron. Reliab. 75 (2017) 327–333.

[70] S. Deng, Z. Cheng, C. Li, X. Yao, Z. Chen, R.-V. Sanchez, Rolling bearing fault diagnosis based on deep boltzmann machines, in: Proceedings of the Prognostics and System Health Management Conference (PHM-Chengdu), IEEE, 2016, pp. 1–6.

[71] C. Li, R.-V. Sánchez, G. Zurita, M. Cerrada, D. Cabrera, Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning, Sensors 16 (6) (2016) 895.

[72] M. Gan, C. Wang, et al., Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings, Mech. Syst. Signal Process. 72 (2016) 92–104.

[73] H. Shao, H. Jiang, X. Zhang, M. Niu, Rolling bearing fault diagnosis using an optimization deep belief network, Meas. Sci. Technol. 26 (11) (2015) 115002.

[74] H. Shao, H. Jiang, F. Wang, Y. Wang, Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet, ISA Trans. 69 (2017) 187–201.

[75] J. Tao, Y. Liu, D. Yang, Bearing fault diagnosis based on deep belief network and multisensor information fusion, Shock Vib. 2016 (2016) 1–9.

[76] M. Dong, D. He, Hidden semi-markov model-based methodology for multi-sensor equipment health diagnosis and prognosis, Eur. J. Operat. Res. 178 (3) (2007) 858–878.

[77] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, S. Van Hoecke, Convolutional neural network based fault detection for rotating machinery, J. Sound Vib. 377 (2016) 331–345.

[78] X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement 93 (2016) 490–502.

[79] X. Ding, Q. He, Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis, IEEE Trans. Instrum. Meas. 66 (2017) 1926–1935.

[80] P. Sermanet, Y. LeCun, Traffic sign recognition with multi-scale convolutional networks, in: Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, 2011, pp. 2809–2813.

[81] L. Eren, Bearing fault detection by one-dimensional convolutional neural networks, Math. Probl. Eng. 2017 (2017) 1–9.

[82] T. Ince, S. Kiranyaz, L. Eren, M. Askar, M. Gabbouj, Real-time motor fault detection by 1-d convolutional neural networks, IEEE Trans. Ind. Electron. 63 (11) (2016) 7067–7075.

[83] C. Farabet, C. Poulet, J.Y. Han, Y. LeCun, Cnp: an fpga-based processor for convolutional networks, in: Proceedings of the International Conference on Field Programmable Logic and Applications, FPL, IEEE, 2009, pp. 32–37.

[84] F. Wang, H. Jiang, H. Shao, W. Duan, S. Wu, An adaptive deep convolutional neural network for rolling bearing fault diagnosis, Meas. Sci. Technol. 28 (2017) 095005.

[85] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, Sensors 17 (2) (2017) 425.

[86] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Signal Process. 100 (2018) 439–453.

[87] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting., J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.

[88] G. Brown, Ensemble learning, in: Encyclopedia of Machine Learning, Springer, 2011, pp. 312–320.
[89] Z. Chen, W. Li, Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network, IEEE Trans. Instrum. Meas. 66 (7) (2017) 1693–1702.
[90] R. Zhang, Z. Peng, L. Wu, B. Yao, Y. Guan, Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence, Sensors 17 (3) (2017) 549.
[91] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: Proceedings of the Twenty-sixth Annual International Conference on Machine Learning, ACM, 2009, pp. 609–616.
[92] M. Norouzi, M. Ranjbar, G. Mori, Stacks of convolutional restricted Boltzmann machines for shift-invariant feature learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2009, pp. 2735–2742.
[93] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, S. Wu, Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing, Mech. Syst. Signal Process. 100 (2018) 743–765.
[94] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.
[95] K.A. Loparo, Bearing Data Center, Case Western Reserve University, 2013.
[96] J. Lee, H. Qiu, G. Yu, J. Lin, et al., Rexnord Technical Services, 'Bearing Data Set', IMS, University of Cincinnati, NASA Ames Prognostics Data Repository, 2007.

Duy-Tang Hoang received the B.S. degree in Electrical Engineering from Hanoi University of Science and Technology, Vietnam in 2010. He is currently working toward the Ph.D. degree in Electrical Engineering, University of Ulsan, Korea. His research interests include machine learning, deep learning based fault diagnosis.

Hee-Jun Kang received the B.S. degree in Mechanical Engineering from Seoul National University, Korea in 1985. He received the M.S. and Ph.D. degrees in Mechanical Engineering from The University of Texas at Austin, USA in 1988 and 1991, respectively. Since March in 1992, he has been a Professor of Electrical Engineering in University of Ulsan. His current research interests are Sensor based robotic application, Robot calibration, Robot fault diagnosis and Robot fault tolerant control.