Neurocomputing 305 (2018) 1–14

Contents lists available at ScienceDirect

Neurocomputing

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

Adaptive deep feature learning network with Nesterov momentum and its application to rotating machinery fault diagnosis

Shenghao Tang[a], Changqing Shen[a,*], Dong Wang[b], Shuang Li[a], Weiguo Huang[a], Zhongkui Zhu[a]

[a] School of Rail Transportation, Soochow University, Suzhou 215131, PR China
[b] College of Science and Engineering, City University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 20 July 2017
Revised 4 January 2018
Accepted 1 April 2018
Available online 4 May 2018

Communicated by Dr Qinfeng Shi

Keywords:
Fault diagnosis
Adaptive learning rate
Feature learning
Deep belief network
Nesterov momentum

ABSTRACT

The effective fault diagnosis of rotating machinery is critical to ensure the continuous operation of equipment and is more economical than scheduled maintenance. Traditional signal processing-based and artificial intelligence-based methods, such as wavelet packet transform and support vector machine, have been proved effective in fault diagnosis of rotating machinery, which prevents unexpected machine breakdowns due to the failure of significant components. However, these methods have several disadvantages that make them unable to automatically and effectively extract valid fault features for the effective fault diagnosis of rotating machinery. A novel adaptive learning rate deep belief network combined with Nesterov momentum is developed in this study for rotating machinery fault diagnosis. Nesterov momentum is adopted to replace traditional momentum to enable declining in advance and to improve training performance. Then, an individual adaptive learning rate method is used to select a suitable step length for accelerating descent. To confirm the utility of the proposed deep learning network architecture, two examinations are implemented on datasets from gearbox and locomotive bearing test rigs. Results indicate that the method achieves impressive performance in fault pattern recognition. Comparisons with existing methods are also conducted to demonstrate that the proposed method is more accurate and robust.

1. Introduction

Rotating machinery is applied on a large scale in modern industries, but this machinery inevitably experiences a variety of faults under complex working conditions that can lead to serious economic losses and human casualties [1]. In rotating machinery, 30% of faults are caused by rolling bearings and 10% by gears. Therefore, effective and accurate fault diagnosis of the vital components of rotating machinery has considerable significance in ensuring safe operation and in avoiding disasters. Accordingly, the fault diagnosis of key parts of rotating machinery has attracted extensive research attention in recent decades [2].

The fault diagnosis methods for rotating machinery currently include two main types: signal processing based and artificial intelligence (AI) based fault diagnosis methods. In the former, signal processing methods such as empirical mode decomposition, morphological filter, and wavelet packet transform [3–5] are typically applied to extract the fault characteristics from signals. Guo et al. [6] combined spectral kurtosis with ensemble empirical mode decomposition to develop a hybrid signal processing method to extract impulses generated by bearing faults. Huo et al. [7] proposed a new multiple-speed fault diagnostic approach based on self-adaptive wavelet transform components to identify four conditions of rolling bearings. Hu et al. [8] adopted a simple harmonic wave for the construction of structuring elements to improve morphological filter and obtained fault characteristics from signals with low signal-to-noise ratios. Although these methods have been proven to be remarkable and powerful in rotating machine fault diagnosis, they still face various challenges. For example, empirical mode decomposition suffers from modal aliasing and endpoint effects. For morphological filter and wavelet packet transform, the selection of structuring elements and base functions requires machinery expertise and comprehensive mathematical skills, which significantly influence the diagnostic result.

An AI-based fault diagnosis method can be regarded as a pattern recognition problem by using the features extracted from the collected signals [9]. Examples of these methods include artificial neural networks, support vector machines (SVMs), and fuzzy inference [10–12]. Xia et al. [13] built a neural network by using a key kernels-particle swarm optimization method to identify different

* Corresponding author.
E-mail address: cqshen@suda.edu.cn (C. Shen).

https://doi.org/10.1016/j.neucom.2018.04.048
0925-2312/

health statuses via the kernels of the Volterra series. Zhang et al. [14] proposed a multivariable ensemble-based incremental SVM to detect multiple faults. Lei et al. [15] applied an adaptive neuro-fuzzy inference system to identify the health status of a planetary gearbox. Although these studies verified the effectiveness of AI-based methods, these methods still exhibit three disadvantages. (1) The extraction of valid features relies primarily upon advanced and complex signal processing technologies and expertise in diagnosis. (2) The selected features require considerable time and manpower because the quality of the features determines the classification quality of an AI-based method. In addition, the selected features are only suitable for a specific issue and require reselection. (3) SVMs and artificial neural networks are shallow architectures that experience difficulty in learning complex non-linear relationships in fault diagnosis [16]. The design of deep architectures with the ability to automatically extract valid fault features for fault diagnosis is encouraged.

As a new tool in machine learning, deep learning [17], which includes convolutional neural networks [18], deep neural networks [19], and deep belief networks (DBNs) [20], can overcome the aforementioned limitations of AI-based fault diagnosis methods. Deep learning adopts architectures composed of multiple non-linear learning layers to obtain highly representative features from data and achieve good performance in pattern recognition [21,22]. In recent years, deep learning has been successfully adopted in fault diagnosis. Sun et al. [23] used a sparse auto-encoder to implement feature learning in the identification of induction motor faults. Janssens et al. [24] presented a feature learning model based on convolutional neural networks for bearing fault detection, and Chen et al. [25] used convolutional neural networks to identify faults in gearboxes. Tran et al. [26] used DBN and Teager–Kaiser energy operator to recognize faults in compressor valves, and Shao et al. [27] combined DBN with particle swarm to design a novel optimization DBN for diagnosis of rolling bearing fault.

Although deep learning models can automatically learn valid features from signals, signals are frequently high-dimensional and massive, which results in poor performance and considerable training time. Numerous optimization methods, such as the adaptive subgradient method [28] and adaptive moment estimation (Adam) [29], have been proposed to accelerate the training speed and improve the classification performance in image identification. However, these methods may lead to the vanishing gradient problem, and a few researchers have proven the availability of these optimization techniques for fault diagnosis.

In this study, an adaptive learning rate DBN (ADDBN) with Nesterov momentum (NM) is proposed for diagnosis of rotating machinery fault diagnosis. The data from gearbox and locomotive bearings are used with the proposed method for fault diagnosis. The advantages of the proposed method include: (1) better performance in updating the proper gradient of the learning rate to ensure satisfactory generalization ability; (2) the ability to automatically extract sensitive deep features without artificial feature selection; and (3) greater accuracy than several existing methods.

The rest of this paper is organized as follows. Section 2 presents the theoretical background of DBN, and Section 3 explains the theory of the proposed method. Section 4 validates the effectiveness of the proposed DBN model in the fault diagnosis of gearbox and locomotive bearings, and Section 5 presents the conclusions of the study.

2. Theoretical context of DBNs

2.1. Restrict Boltzmann machines (RBMs) architecture

DBN is a deep learning model that stacks multilayer RBMs. As is shown in Fig. 1, each RBM is composed of a visible layer that contains visible neurons $v = \{v_1, v_2, \dots, v_i\}$ and a hidden layer that contains hidden neurons $h = \{h_1, h_2, \dots, h_j\}$.

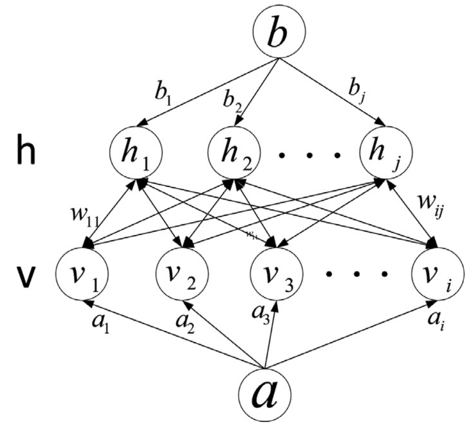

Fig. 1. Schematic architecture of RBM.

The visible layer with bias vector $\boldsymbol{a}$ is connected to the hidden layer with bias vector $\boldsymbol{b}$ by weights $\boldsymbol{W}$. No connection exists between the neurons in the visible and hidden layers. $v_i \in \{0, 1\}$ and $h_j \in \{0, 1\}$ are binary and discrete stochastic variables that indicate an active or inactive state. Then, the energy function for a set of determined states (v, h) is defined as follows:

$$E(v, h; \theta) = -\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} v_i h_j - \sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j \quad (1)$$

where $\theta = \{W, a, b\}$, v_i is the status of the ***i***th visible neuron, and h_j is the status of the ***j***th hidden neuron. w_{ij} is the connecting weight between the ***i***th visible neuron and the ***j***th hidden neuron. The joint probability distribution for visible and hidden vectors is calculated using Eq. (1) as follows:

$$p(v, h; \theta) = \exp(-E(v, h; \theta))/Z(\theta) \quad (2)$$

where $Z(\theta)$ is a normalizing factor and is defined as:

$$Z(\theta) = \sum_{v}\sum_{h} \exp(-E(v, h; \theta)) \quad (3)$$

No intra-layer connection exists between visible and hidden layers; hence, the conditional probabilities of visible and hidden neurons that are conditional independent are given via Eqs. (4) and (5), respectively, as follows:

$$P(h|v; \theta) = P(v, h; \theta)/P(v; \theta) = \prod\nolimits_j P\left(h_j|v; \theta\right) \quad (4)$$

$$P(v|h; \theta) = P(v, h; \theta)/P(h; \theta) = \prod\nolimits_i P(v_i|h; \theta) \quad (5)$$

When binary neurons are considering, the individual activation probabilities are expressed as:

$$P(h_j = 1|v; \theta) = \sigma\left(\sum\nolimits_{i=1}^{m} w_{ij} v_i + b_j\right) \quad (6)$$

$$P(v_i = 1|h; \theta) = \sigma\left(\sum\nolimits_{j=1}^{n} w_{ij} h_j + a_i\right) \quad (7)$$

where $\sigma = 1/(1 + e^{-x})$ is a sigmoid function. To maximize the fitting of input data, the stochastic gradient descent (SGD) is applied to maximize the logarithmic likelihood function of RBM to obtain the optimal parameters, which are denoted as θ^* and expressed in the following equation:

$$\theta^* = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \sum\nolimits_{v} \log P(v|\theta) \quad (8)$$