Measurement 127 (2018) 246–255

Contents lists available at ScienceDirect

Measurement

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

Multi-level wavelet packet fusion in dynamic ensemble convolutional neural network for fault diagnosis

Yan Han, Baoping Tang*, Lei Deng

The State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400030, China

A R T I C L E I N F O

Keywords:
Convolutional neural network
Dynamic ensemble
Fault diagnosis
Multi-level wavelet packet

A B S T R A C T

Due to the complicated structure and varying operating conditions of machinery in various applications, intelligent identification of the health state based on the vibration data is still a great challenge in fault diagnosis. In this paper, a variant of the convolutional neural network, named dynamic ensemble convolutional neural network was proposed for fault diagnosis by intelligent fusion of the multi-level wavelet packet. First, wavelet packet transform was employed to construct multi-level wavelet coefficients matrixes for representing the nonstationary vibration signal comprehensively. Then, several paralleled convolutional neural networks with shared parameters were built, not only to learn the multi-level fault features automatically, but also to restrain the overfitting of the deep learning partially. At last, a dynamic ensemble layer was applied to fuse multi-level wavelet packet by assigning weights dynamically. The validation on two experimental datasets of the planetary gearbox under varying speed demonstrated that the developed method can fuse the fault features in multi-level wavelet packet thoroughly, and improve the effectiveness and robustness for fault diagnosis of gearbox under whether the sufficient or limited fault data conditions.

1. Introduction

Due to ever-complicated structure and varying working conditions, the high ratio of the machine failure and the long maintenance time impairs the human safety and economic benefits of enterprise seriously, intelligent identification of the health conditions of machinery base on the vibration data is still a great challenge in fault diagnosis [1,2].

Recently, deep learning raises a new climax of artificial intelligence, which has been widely used in various pattern recognition applications [3–6]. Through the multiple layers of nonlinear transformation, deep learning can learn discriminative features automatically, which not only make up the deficiency of inferior representational ability of the traditional shallow models, but also get rid of the tedious process of feature extraction and selection.

Because of its incomparable superiorities in pattern recognition filed, different kinds of deep learning techniques, such as deep belief networks, the stacked autoencoder and the convolutional neural network (CNN) have been applied in various fault diagnosis applications during recent years. In some studies, the deep learning was only used as a classifier to obtain the better performance than shallow models [7–9], however, this usage of deep learning still requires plenty of professional knowledge for feature extraction, and the depending of feature selection on experience influences the diagnostic accuracy severely. By contrast, another scholars fed the raw vibration data or frequency spectral into deep learning models directly for fault diagnosis [10–12], which took full advantages of the end-to-end learning ability of deep learning to obtain the discriminative features. Nevertheless, while considering complex structures and variable operating conditions of the machinery, the fault information concealed in time or frequency domain cannot reveal the non-stationary dynamic characteristics of the vibration signal. In order to solve this problem, different kinds of time-frequency analysis methods, such as short-time Fourier transform (STFT), wavelet packet transform (WPT) and empirical mode decomposition (EMD) have been combined with deep learning for fault diagnosis [13–17]. These methods transformed 1D vibration signals to 2D representations by time-frequency analysis and utilized deep learning methods to extract discriminative features from the time-frequency representations instead of the time or frequency domain, which have provided a new approach for fault diagnosis under varying operating conditions.

Comparing to STFT and EMD, WPT has the characteristics of multi-resolution analysis and a solid theoretical basis. Therefore, wavelet packet was selected as the input of deep learning model for fault diagnosis in this paper. However, the following issues need to be considered while combining wavelet packet and deep learning methods for fault diagnosis: First, it is difficult to determine which level of the

* Corresponding author.
E-mail address: bptang@cqu.edu.cn (B. Tang).

https://doi.org/10.1016/j.measurement.2018.05.098
Received 28 February 2018; Received in revised form 7 May 2018; Accepted 24 May 2018
Available online 25 May 2018

wavelet packet contains the most useful fault features, and the fault information in one single level of wavelet packet is limited. Second, the contributions of each level in wavelet packet are unbalanced, inappropriate weights allocation will impede the promotion of the performance. At last, as the fault data in some engineering applications are limited, the overfitting will degrade the learning ability of deep learning badly. Therefore, the applicability of the deep learning method under limited data scenario should be taken into consideration.

In view of the problems mentioned above, a novel intelligent fault diagnosis method for gearbox is developed in this paper, which utilized the dynamic ensemble convolutional neural network (DECNN) to extract and fuse discriminative features from multi-level wavelet packet. Besides, the verification of two experimental datasets has demonstrated that the proposed method was more effective and robust than other methods for fault diagnosis of gearbox under whether the sufficient or limited fault data conditions. The main contributions of this paper can be summarized as follows.

(1) Multi-level wavelet packet was used as the input of the deep learning model to enrich fault information. It can not only represent nonstationary vibration signal comprehensively, but also get rid of the process of selecting the optimum level of wavelet packet.
(2) A dynamic ensemble layer was applied to fuse useful fault features at different time-frequency resolution by assigning weights to different level of wavelet packet adaptively, the effectiveness and robustness of the proposed model were further improved.
(3) The proposed model with shared parameters in convolutional layers helps restrain the overfitting of the deep learning to some degree, especially in limited fault data scenario.

The remainder of this paper is organized as follows. Section 2 introduces the basic theory of the CNN. In Section 3, the proposed method is described in details. In Section 4, the experimental datasets of a gearbox are introduced, and the effectiveness of the developed method is verified by the comparison with other diagnosis methods. Conclusion and future work are given in Section 5.

2. A brief introduction of the CNN

The CNN is an important subtype of deep learning methods, which has shown prominent performance in various pattern recognition applications due to its powerful feature extraction ability. The basic structure of a CNN is composed of the input layer, convolution layers, pooling layers, also known as sampling layers, fully connected layers and output layer. The front end of a CNN is usually stacked by a few convolutional layers and pooling layers alternately, while the back-end of the architecture consists of a fully connected layer and a traditional classification model before the output layer.

Convolutional layers are used to obtain the discriminative features of the input data for pattern recognition. In a convolutional layer, a series convolutional kernels are used to generate various feature maps by sliding different local filters over the whole input neurons at previous layer; in addition, all the neurons share the weights in the same kernel, this helps to reduce the number the parameters and accelerate the convergence time of the model. The convolution operation between the input neurons and the learnable convolution kernels can be expressed as [18]:

$$x_j^l = f\left(\sum_i x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{1}$$

where x_j^l is j^{th} feature map at the lth layer; x_i^{l-1} denotes the i^{th} input feature map at $(l-1)^{th}$ layer, k_{ij}^l denotes the convolutional kernel which connected i^{th} input feature map with j^{th} feature map, b_j^l denotes the bias, and $*$ denotes the 2d convolutional operation. $f(\bullet)$ is an activation function, where $f(\bullet)$ includes sigmoid function, hyperbolic tangent function, rectified linear units (ReLU) and so on. Nowadays, ReLU activation function [19] is widely applied to avoid the vanishing gradient problems and accelerate the convergence of the CNN, the ReLU function is described as:

$$f(x) = \max(x, 0) \tag{2}$$

Pooling layers usually follows convolution layers alternately, which are used to reduce the parameters of the whole network and achieve low resolution of feature maps. The commonly used pooling methods include max-pooling, average-pooling, stochastic-pooling and so on [20]. After multi-stage of convolutional layers and pooling layers, a fully connected layer is added to integrate the discriminative local information of category and improve the performance of the CNN, in the full connection layer, dropout technology is often used as a regularization method to restrain overfitting. At last, a classifier is used to connect the fully connected layer and output layer to complete the task of classification.

For a specific task of classification, the cost function is significant to the performance of the CNN. As the cross-entropy can accelerate the update speed of weights and convergence speed of the whole model, in this paper, cross-entropy cost function is applied, defined as [21]:

$$L(p(x), q(x)) = -\sum_{i=1}^{N} p_i(x)\log(q_i(x)) \tag{3}$$

where N is the number of the classes, $p_i(x)$ and $q_i(x)$ denote the real probability and estimated probability of the observation x belonging to i^{th} class, respectively.

3. The proposed method

Due to the strong representational ability of WPT for nonstationary vibration signal and the powerful learning ability of the CNN, the combination of WPT and the CNN has employed in applications of fault diagnosis of machinery under the variable operating conditions. However, in various fault diagnosis applications, there is no general consensus that which level of the wavelet packet contains the most useful fault features, and it is prone to lose effective fault information if it rely on only one single level wavelet packet; meanwhile, one single CNN cannot extract discriminative features from the multi-level wavelet packet directly.

To solve the problems listed above, a variant of the CNN called DECNN was proposed for fault diagnosis by intelligent fusion of multi-level wavelet packet in this paper. The structure of the developed fault diagnosis method is shown in Fig. 1, at the front end, multi-level wavelet packet was used to construct the multi-level coefficients matrixes (MWCMs), which provide rich fault features at different scales. In the middle part of the proposed network, multiple sub CNNs with shared parameters are used to run discriminative features extraction from the MWCMs. Finally, a dynamic ensemble layer is added for fusing the result of the sub CNNs dynamically at the back end. In the following of this section, the proposed method is introduced in three parts: input of the DECNN, feature learning and fusion in DECNN, and the general procedure of the proposed method.

3.1. Input of DECNN

Due to the varying operating conditions, the vibration based fault diagnosis methods for the strong nonstationary signal usually call for time-frequency analysis. Wavelet transform (WT) is an effective time-frequency method which has been extensively used in nonstationary signal processing [22]. However, WT only decomposes the low frequency bands and ignores the high frequency bands in which the resonance information always exists. As an extension of WT, WPT inherits the merits of prefect time-frequency localization of WT, in addition, WPT has detailed decomposition, stronger reconstruction ability and