Applied Soft Computing Journal 73 (2018) 898–913

Contents lists available at ScienceDirect

Applied Soft Computing Journal

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

Roller bearing fault diagnosis using stacked denoising autoencoder in deep learning and Gath–Geva clustering algorithm without principal component analysis and data label

Fan Xu, Wai tai Peter Tse *, Yiu Lun Tse

Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China

HIGHLIGHTS

- Reduce the dimension of extracted feature using SDAE directly without PCA.
- Fulfilling the bearing fault diagnosis by using SDAE and clustering model without data label.
- Using the SDAE reduce the high dimension to 2 or 3 directly from frequency domain feature directly after FFT decomposition.

ARTICLE INFO

Article history:
Received 19 February 2018
Received in revised form 24 September 2018
Accepted 26 September 2018
Available online xxxx

Keywords:
Deep learning
Stacked denoising autoencoder
Gath–Geva clustering algorithm
Roller bearing fault diagnosis

ABSTRACT

Most deep learning models such as stacked autoencoder (SAE) and stacked denoising autoencoder (SDAE) are used for fault diagnosis with a data label. These models are applied to extract the useful features with several hidden layers, then a classifier is used to complete the fault diagnosis. However, these fault diagnosis classification methods are only suitable for tagged datasets. Actually, many datasets are untagged in practical engineering. The clustering method can classify data without a label. Therefore, a method based on the SDAE and Gath–Geva (GG) clustering algorithm for roller bearing fault diagnosis without a data label is proposed in this study. First, SDAE is selected to extract the useful feature and reduce the dimension of the vibration signal to two or three dimensions direct without principal component analysis (PCA) of the final hidden layer. Then GG is deployed to identify the different faults. To demonstrate that the feature extraction performance of the SDAE is better than that of the SAE and EEMD with the FE model, the PCA is selected to reduce the dimension of eigenvectors obtained from several previously hidden layers, except for the final hidden layer. Compared with SAE and ensemble empirical mode decomposition (EEMD)-fuzzy entropy (FE) models, the results show that as the number of the hidden layers increases, all the fault samples under different conditions are separated better by using SDAE rather than those feature extraction models mentioned. In addition, three evaluation indicators such as PC, CE, and classification accuracy are used to assess the performance of the method presented. Finally, the results show that the clustering effect of the method presented, and its classification accuracy are superior to those of the other combination models, including the SAE-fuzzy C-means (FCM)/Gustafson–Kessel (GK)/GG and EEMD-fuzzy entropy FE-PCA-FCM/GK/GG.

1. Introduction

The roller bearing is one of the important components of a rotating machinery system or equipment, and its operating status affects the reliability of the whole system direct. It is an urgent task to feature extraction and fault identification, since the characteristics of rolling bearing fault signals are nonlinear and have non-flat stability [1–4]. For the information extraction, vibration-based feature extraction is a common and useful method. As the vibration signals are unstable, the fault diagnosis is challenging in the mechanical society.

For machinery condition monitoring, feature extraction based on nonlinear analysis methods, such as empirical mode decomposition (EMD), and ensemble empirical mode decomposition (EEMD), has been widely applied in the field of mechanical fault diagnosis. Due to the characteristics of rolling bearing fault vibration, signals are nonlinear and have nonflat stability, but a self-adaptively method, named EMD, can decompose a complicated signal into some intrinsic mode functions (IMFs), based on the local characteristic time scale of the signal [5]. However, there is a mode-mixing problem in EMD. Moreover, to overcome the problems of mode mixing in EMD, EEMD [6], an improved version

* Corresponding author.
E-mail address: Peter.W.Tse@cityu.edu.hk (W.t.P. Tse).

https://doi.org/10.1016/j.asoc.2018.09.037
1568-4946/

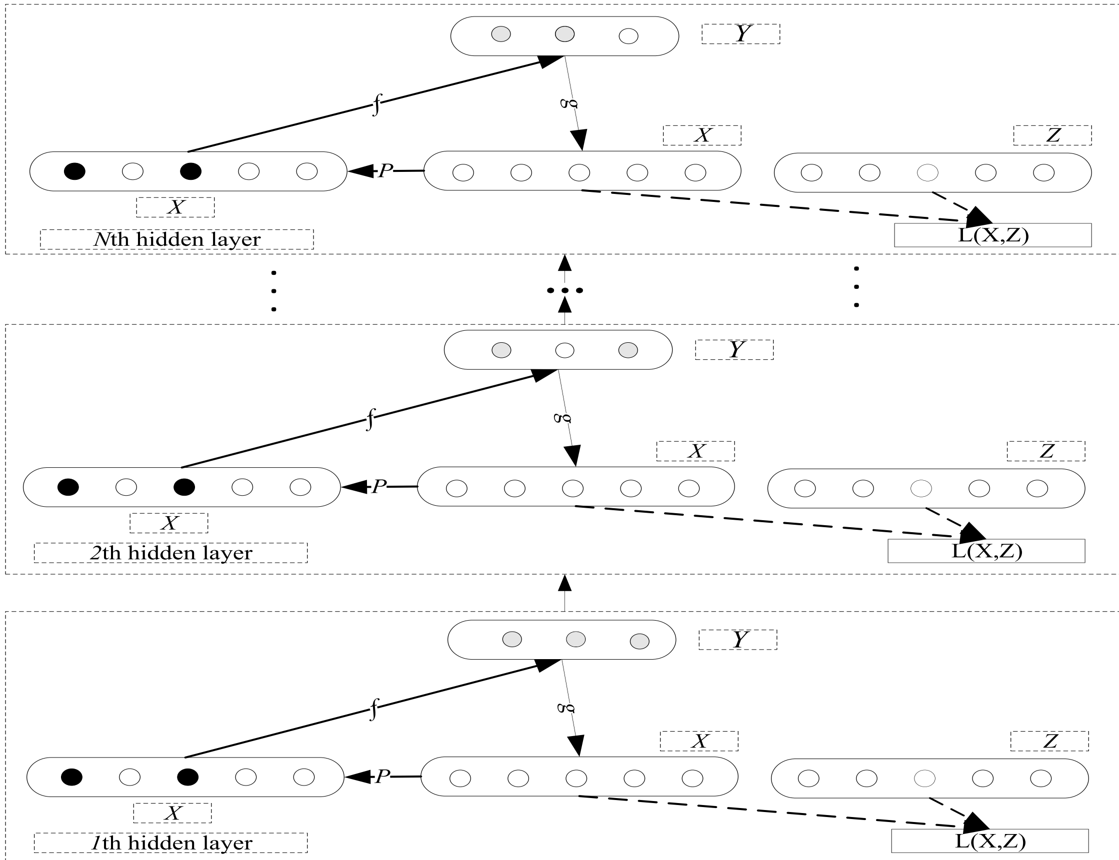

Fig. 1. The structure of SDAE.

of EMD, was proposed by Huang et al. To eliminate the mode-mixing phenomenon, the EEMD introduced a random white noise signal to calculate the ensemble means of each IMF. Recently, EMD and EEMD have been widely applied in fault diagnosis [7–11].

However, the aforementioned vibration signal feature extraction methods generally require complex mathematical operations and people need to have an extensive experiment to understand the vibration signal. For some complex systems, such as ambient interference and the internal structure interacting with each other, it is difficult to extract the useful vibration signal features from the measured vibration signals automatically and effectively [12,13].

In recent years, many scholars have gradually paid attention to deep learning due to its strong feature extraction capabilities. Compared with the EMD and EEMD, a stacked autoencoder (SAE) in deep learning can extract the useful information automatically and reduce the dependence on expert troubleshooting experience with signal processing technology. In addition, SAE has been successfully applied in fault diagnosis. Feng et al. presented a method based on a stacked autoencoder (SAE) in DL for roller bearing fault diagnosis [14]. In this paper, SAE was used to extract the vibration signals' features, then the pre-training, fine-tuning network parameters, and eigenvectors through the final hidden layer were regarded as the input of the output layer to distinguish the different fault categories. Tan et al. [15] denoised signals with a digital wavelet frame (DWF) and performed fault diagnosis based on a stacked autoencoder (SAE). It combined low-level features to form more abstract high-level features to represent data-distributed characteristics. Some scholars have already employed the SAE successfully for fault feature extraction and fault diagnosis [16–19].

However, SAE is only a simple reconstruction of the input data. The features learnt do not have good generalization capabilities. By adding the "damage noise" to the raw data, based on the structure of the SAE network, that is, a part of the input data is randomly zeroed, and the "pure raw input" is reconstructed from "noisy data", a stacked denoising autoencoder (SDAE) can obtain more robust expression features for the original input information than SAE [20,21]. The input is corrupted by randomly setting some of the input to zero, which is called dropout noise. This denoising process helps the autoencoders to learn robust representation. In addition, each autoencoder layer is intended to learn an increasingly abstract representation of the input [16]. SDAE has been extensively used in different types of application [15,22–24].

For fault diagnosis, many machine learning models have been widely applied in fault diagnoses, such as support vector machine (SVM), random forest (RF), and artificial neural networks (ANN) [8,25,26]. However, most of the deep learning models with a classifier output layer and the above different machine learning classification models for fault diagnosis are only suitable for a tagged dataset. Sometimes, the researchers also rely on the knowledge about faults, such as the fault characteristic frequency, and extract the related frequency band energy as features, and then establish the relationship between the feature vectors and their labels (fault type or healthy condition) [27]. But marking data requires large amounts of time and material resources when the amount of data is large. A clustering method can classify data without a label. Fuzzy c-means (FCM) is one of the most common clustering methods. Zhang et al. have developed a method based on FCM for fault diagnosis [28], but the FCM is only suitable for a homogeneous data structure and handles the spherical distance data with standard specification. Gustafson–Kessel (GK) is an improved algorithm based on FCM. GK can handle any direction data with subspace dispersion by introducing the adaptive distance norm and covariance matrix [29]. Wang et al. applied the GK in roller bearing fault diagnosis [30]. However, FCM and GK are