Neurocomputing 272 (2018) 619–628

Contents lists available at ScienceDirect

Neurocomputing

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

A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines

Feng Jia, Yaguo Lei*, Liang Guo, Jing Lin, Saibo Xing

State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China

ARTICLE INFO

Article history:
Received 28 May 2017
Revised 9 July 2017
Accepted 14 July 2017
Available online 21 July 2017

Communicated by Hongli Dong

Keywords:
Normalized sparse autoencoder
Deep learning
Intelligent fault diagnosis
Local connection network

ABSTRACT

In traditional intelligent fault diagnosis methods of machines, plenty of actual effort is taken for the manual design of fault features, which makes these methods less automatic. Among deep learning techniques, autoencoders may be a potential tool for automatic feature extraction of mechanical signals. However, traditional autoencoders have two following shortcomings. (1) They may learn similar features in mechanical feature extraction. (2) The learned features have shift variant properties, which leads to the misclassification of mechanical health conditions. To overcome the aforementioned shortcomings, a local connection network (LCN) constructed by normalized sparse autoencoder (NSAE), namely NSAE-LCN, is proposed for intelligent fault diagnosis. We construct LCN by input layer, local layer, feature layer and output layer. When raw vibration signals are fed to the input layer, LCN first uses NSAE to locally learn various meaningful features from input signals in the local layer, then obtains shift-invariant features in the feature layer and finally recognizes mechanical health conditions in the output layer. Thus, NSAE-LCN incorporates feature extraction and fault recognition into a general-purpose learning procedure. A gearbox dataset and a bearing dataset are used to validate the performance of the proposed NSAE-LCN. The results indicate that the learned features of NSAE are meaningful and dissimilar, and LCN helps to produce shift-invariant features and recognizes mechanical health conditions effectively. Through comparing with commonly used diagnosis methods, the superiority of the proposed NSAE-LCN is verified.

1. Introduction

With the development of industry, machines have been more automatic and efficient, and their components are linked to each other inseparably [1]. Once a component has a fault, this fault would quickly produce chain reaction and lead to the damage of other components. Such unexpected faults would make machines break down, resulting in economic loss and even person safety threat [2]. Therefore, the fault diagnosis of machines has received lots of attention.

Intelligent fault diagnosis is one of the powerful tools in the field of fault diagnosis [3]. Based on massive monitored signals of the machines, it is able to replace diagnosticians with artificial intelligent techniques like neural networks to rapidly process these signals and automatically recognize mechanical health conditions [4–6]. Thus, intelligent fault diagnosis plays an irreplaceable role in modern industries especially when massive vibration signals are available. As we know, traditional intelligent fault diagnosis has two main steps: feature extraction and fault recognition [7]. Based on these steps, lots of effort has been taken on intelligent fault diagnosis. Georgoulas et al. [8] designed the features of motor faults based on time-frequency methods and employed Mahalanobis Distance classifier to recognize motor health conditions. Prieto et al. [9] proposed a method using statistical features and hierarchical networks to classify bearing health conditions. Amar et al. [10] proposed a feature enhancement procedure to obtain features from vibration spectra and applied neural networks to diagnose the bearing faults. Wang [11] designed a feature extraction algorithm that extracts redundant statistical features from different wavelet decomposition levels, and applied K-nearest neighbor algorithm to identify gear health conditions. Lei et al. [12] designed two features for gearboxes specifically and used these features and relevance vector machine to recognize the health conditions.

Although the studies above achieved good results, they may suffer the weakness as follows. In these methods, plenty of the actual effort is taken for the manual design of feature extraction algorithms since traditional classifiers cannot extract the representative features from raw signals [13]. Such feature designing processes should make full use of human knowledge in signal processing and diagnostic expertise, which costs much human labor and makes the methods less automatic. Among deep learning

* Corresponding author.
E-mail address: yaguolei@mail.xjtu.edu.cn (Y. Lei).

http://dx.doi.org/10.1016/j.neucom.2017.07.032
0925-2312/

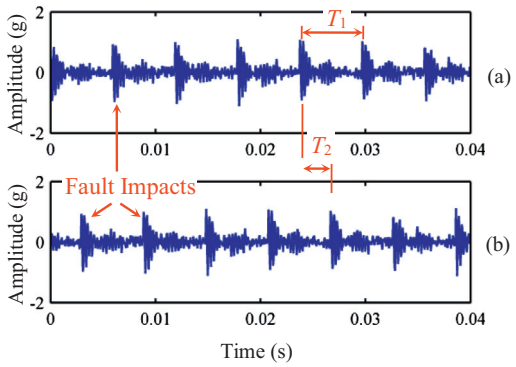


Fig. 1. Two simulated samples for mechanical fault signals: (a) the first sample and (b) the second sample.

techniques, autoencoders may help fault diagnosis to handle the weakness above since their basic motivation is to be fed with raw signals and accomplish the task of feature extraction automatically [14–16]. Currently, autoencoders have attracted attentions in the field of fault diagnosis. Thirukovalluru et al. [17] employed denoising autoencoder to extract high-level features from manual features and two classifiers to recognize mechanical faults. Jia et al. [18] used frequency spectra as the input of a deep network based on autoencoders to recognize mechanical health conditions. Chen and Li [19] applied sparse autoencoder (SAE) to get representative features from statistical values of bearing signals and recognized the health conditions using deep belief network. Mao et al. [20] proposed a fault diagnosis method using frequency spectra and autoencoder extreme learning machines. It can be seen that most of these studies, however, still used manual features as the input of the neural networks, which may deviate from the basic motivation of the autoencoders.

The following two shortcomings of the autoencoders are the main reasons why they are not easily used to learn features well from raw vibration signals of machines. (1) The autoencoders cannot be ensured to get various meaningful features from the vibration signals. The feature extraction process of an autoencoder can be regarded as the dot produce results between its weight matrix composed by a set of basis vectors and the vibration signals. So obtaining good features of raw data depends on the weight matrix of the autoencoder. In the well trained weight matrix, its basis vectors should not only have the own patterns like acting as Gabor bases to produce the meaningful features, but also be different from each other so as to produce various features. Traditional constraints applied to autoencoders, such as sparse regularization and weight decay, could force the basis vectors of the weight matrix to learn patterns but cannot force them to be different. So autoencoders learn too many similar features and prevent their applications in intelligent fault diagnosis of machines. (2) The autoencoders cannot be directly used for feature learning when the data have shift variant properties. Unfortunately, the vibration signals of a faulty machine always show such properties. In Fig. 1, we use two samples that simulate the vibration signals of a mechanical fault to illustrate the properties. It can be seen that when a fault occurs in the machines, the periodic fault impacts are excited by the contact of the fault component and other components, and the contact period is T_1. Such impulse-like vibration behavior of a vibration signal is an important characteristic for fault recognition. When machines operate, their components contact with each other in a time-varying way. So the fault impacts of the first sample and the second sample would shift by T_2. Once we use the autoencoders to extract features from these samples, the features also have shift variant properties, leading to the misclassification of mechanical fault samples.

We propose a local connection network (LCN) constructed by normalized sparse autoencoder (NSAE), namely NSAE-LCN, to overcome the shortcomings of autoencoders. LCN is constructed by four layers, i.e., input layer, local layer, feature layer and output layer, where the local layer is trained by NSAE. So when raw vibration signals are fed to the input layer, LCN first uses NSAE in the local layer to locally learn various meaningful features from the vibration signals, then obtains shift-invariant features from the learned features in the feature layer and finally recognizes mechanical health conditions in the output layer. The proposed NSAE-LCN is validated by a gearbox dataset and a bearing dataset respectively, both involving different health conditions under various operating conditions. And its superiority is verified by comparing with commonly used diagnosis methods.

The contributions of this paper can be summarized as follows.

(1) Based on sparse autoencoder, NSAE is proposed for automatic feature extraction from the vibration signals of machines. Since an orthonormality constraint is used in NSAE, the weight matrix trained by NSAE can be viewed as a set of basis showing time-frequency properties, which encourages the learned features of NSAE to not only have meaningful patterns but also be dissimilar. Thus, NSAE performs well in mechanical feature extraction.
(2) We propose NSAE-LCN for intelligent fault diagnosis. It incorporates the processes of feature extraction and fault recognition into a general-purpose learning procedure. Therefore, the proposed method can be used to directly learn features from raw vibration signals and recognize the health conditions of machines for various diagnosis tasks.

The rest of this paper is organized as follows. In Section 2, sparse autoencoder is briefly described. Section 3 details the proposed NSAE-LCN. In Section 4, the diagnosis cases of a gearbox dataset and a bearing dataset are studied separately using NSAE-LCN. Finally, conclusions are drawn in Section 5.

2. Sparse autoencoder

SAE is a widely used autoencoder that attempts to learn features from raw data. It has symmetrical neural network with an input layer, a hidden layer and an output layer [21]. The input layer and the hidden layer constitute the encoder of SAE, which transforms the input data into features. And the hidden layer and the output layer constitute the decoder of SAE, which reconstructs the input data from the corresponding features.

Given unlabeled data $\{\boldsymbol{x}_m\}_{m=1}^{M}$ where $\boldsymbol{x}_m \in \Re^{N\times 1}$, the encoder uses a mapping function f_{SAE} to calculate the feature $\boldsymbol{h}^m$ from $\boldsymbol{x}_m$ and $\boldsymbol{h}^m$ has K dimensions.

$$\boldsymbol{h}^m = f_{\mathrm{SAE}}(\boldsymbol{x}_m) = \sigma_s(\boldsymbol{W}_{\mathrm{SAE1}}\boldsymbol{x}_m + \boldsymbol{b}_1) \tag{1}$$

where $\sigma_s(z)$ is the sigmoid function, $\boldsymbol{W}_{\mathrm{SAE1}}$ is the weight matrix of the encoder and $\boldsymbol{b}_1$ is the bias vector. The decoder uses a mapping function g_{SAE} for reconstruction

$$\hat{\boldsymbol{x}}_m = g_{\mathrm{SAE}}(\boldsymbol{h}^m) = \boldsymbol{W}_{\mathrm{SAE2}}\boldsymbol{h}^m + \boldsymbol{b}_2 \tag{2}$$

where $\boldsymbol{W}_{\mathrm{SAE2}}$ is the weight matrix of the decoder and $\boldsymbol{b}_2$ is the bias vector. The reconstruction error can be calculated by

$$J_{\mathrm{SAE}} = \frac{1}{2M}\sum_{m=1}^{M}\left\|\boldsymbol{x}_m - \hat{\boldsymbol{x}}_m\right\|^2. \tag{3}$$

To get better features of the input data, sparse representation is used in SAE. So Kullback–Leibler (KL) divergence function is applied to encourage the sparsity of the feature $\boldsymbol{h}^m$ and learn the

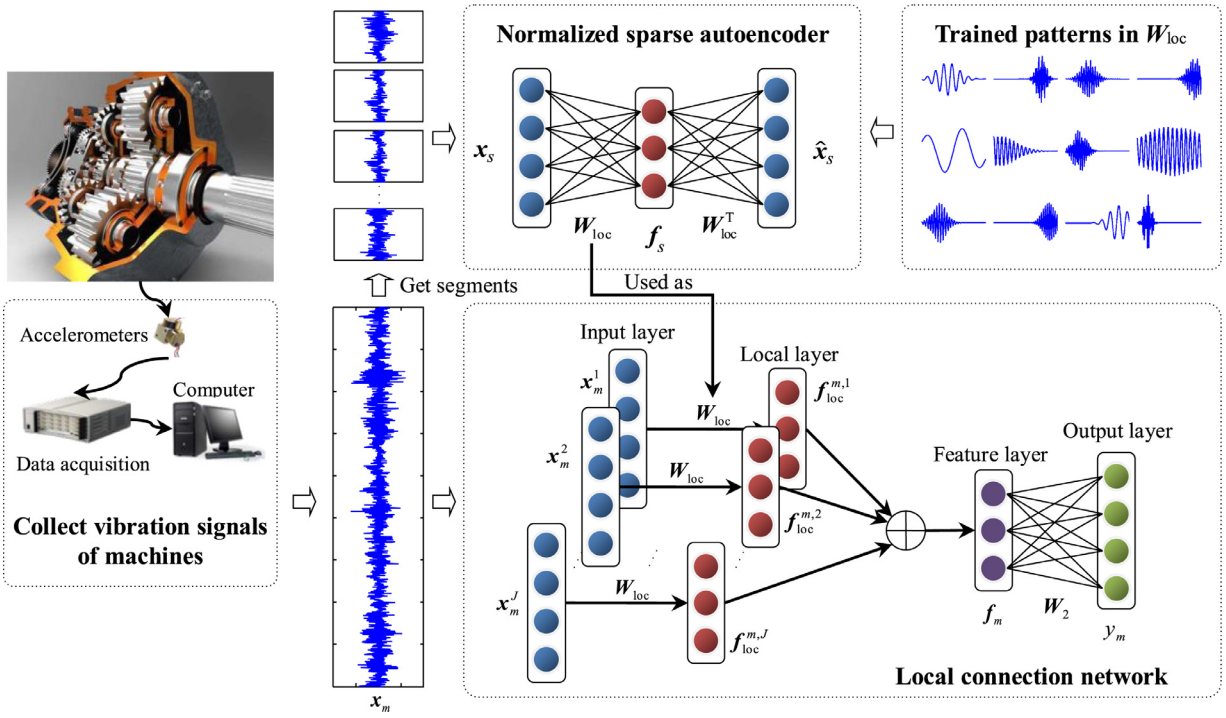


Fig. 2. The flow chart of the proposed NSAE-LCN.

sparse representation. The average activation of $\boldsymbol{h}^m$ is

$$\hat{p}_k = \frac{1}{M}\sum_{m=1}^{M} h_k^m \tag{4}$$

where $k = 1, 2, \ldots, K$.

The sparsity penalty term $\mathrm{KL}(p\|\hat{p}_k)$ based on the KL divergence function can be used to penalize $\hat{p}_k$ deviating from the sparsity parameter p.

$$\mathrm{KL}(p\|\hat{p}_k) = p\log\frac{p}{\hat{p}_k} + (1-p)\frac{1-p}{1-\hat{p}_k} \tag{5}$$

SAE is finally trained by the optimization problem in Eq. (6).

$$\min_{\boldsymbol{W},\boldsymbol{b}} \; J_{\mathrm{SAE}} + \beta\sum_{k=1}^{K}\mathrm{KL}(p\|\hat{p}_k) \tag{6}$$

where β is the regular parameter controlling the weight between the reconstruction error and the sparsity penalty term.

3. The proposed method

This section details the proposed NSAE-LCN method for intelligent fault diagnosis of machines, as shown in Fig. 2. In this method, NSAE is proposed to overcome the shortcomings of autoencoders in learning various features, and LCN is developed to deal with the shift-variance classification problem in intelligent fault diagnosis.

3.1. Normalized sparse autoencoder

Based on SAE, we propose NSAE for various meaningful features in intelligent diagnosis of machines. The differences between NSAE and SAE are listed as follows. Firstly, we use rectified linear units (ReLU) as the activation function in NSAE since ReLU allows for more efficient training for a network than sigmoid function and encourages sparse activation [22]. Secondly, bias is not used in NSAE because the research results [23] show that bias is nonessential in feature learning. Thirdly, KL divergence function is replaced with L1 norm to find the sparse solution of the autoencoder. The two functions are both commonly used for sparse representation, but KL divergence function has two parameters and L1 norm only needs one parameter. Finally, a soft orthonormality constraint is added in the cost function of NSAE so as to force NSAE to learn dissimilar features from mechanical vibration signals.

Given unlabeled data $\{\boldsymbol{x}_m\}_{m=1}^{M}$, the encoder of NSAE uses a mapping function f to calculate the feature $\boldsymbol{h}_{\mathrm{NSAE}}^m$ from $\boldsymbol{x}_m$.

$$\boldsymbol{h}_{\mathrm{NSAE}}^m = f(\boldsymbol{x}_m) = \sigma_{\mathrm{r}}(\boldsymbol{W}_{\mathrm{NSAE1}}\boldsymbol{x}_m) \tag{7}$$

where σ_{r} is ReLU and it is described in Eq. (8).

$$\sigma_{\mathrm{r}}(z) = \begin{cases} 0, & if\ z < 0 \\ z, & if\ z \geq 0 \end{cases} \tag{8}$$

Then the decoder of NSAE reconstruct $\hat{\boldsymbol{x}}_m$ by a mapping function g_{NSAE}.

$$\hat{\boldsymbol{x}}_m = g_{\mathrm{NSAE}}(\boldsymbol{h}_{\mathrm{NSAE}}^m) = \boldsymbol{W}_{\mathrm{NSAE2}}\boldsymbol{h}_{\mathrm{NSAE}}^m \tag{9}$$

To learn various features from vibration signals, an orthonormality constraint is added to the weight matrix of NSAE. The cost function of NSAE using a hard orthonormality constraint is shown in Eq. (10).

$$\min_{\boldsymbol{W}} \; \frac{1}{2M}\sum_{m=1}^{M}\left\|\hat{\boldsymbol{x}}_m - \boldsymbol{x}_m\right\|_2^2 + \lambda\sum_{m=1}^{M}\|\boldsymbol{h}_{\mathrm{NSAE}}^m\|_1 \tag{10}$$

$$s.t. \quad \boldsymbol{W}_{\mathrm{NSAE1}}\boldsymbol{W}_{\mathrm{NSAE2}} = \boldsymbol{I}$$

where λ is the regular parameter of NSAE.

Traditionally, it is difficult to solve this problem with the hard orthonormality constraint. To loose this hard constraint, we use a soft constraint based on the tied-weight strategy. We replace $\boldsymbol{W}_{\mathrm{NSAE1}}$ and $\boldsymbol{W}_{\mathrm{NSAE2}}$ with $\boldsymbol{W}$ and $\boldsymbol{W}^{\mathrm{T}}$ in Eq. (10), respectively, and the optimization problem becomes minimizing the following cost function

$$J_{\mathrm{NSAE}} = \frac{1}{2M}\sum_{m=1}^{M}\left\|\boldsymbol{W}^{\mathrm{T}}\sigma_{\mathrm{r}}(\boldsymbol{W}\boldsymbol{x}_m) - \boldsymbol{x}_m\right\|_2^2 + \lambda\sum_{m=1}^{M}\|\sigma_{\mathrm{r}}(\boldsymbol{W}\boldsymbol{x}_m)\|_1 \tag{11}$$

In Eq. (11), the nonlinear activation function ReLU forces the learned features to be non-negative, which means that the negative values in $\boldsymbol{W}\boldsymbol{x}_m$ would be zeros and some information of the input data would be lost in the training process. However, the first term in J_{NSAE} forces NSAE to reconstruct the input data as well as possible. Thus, NSAE actually encourages the results of $\boldsymbol{W}\boldsymbol{x}_m$ to be positive, and the first term could degenerate into $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{x}_m - \boldsymbol{x}_m$. When the value of the first term is close to zero in the training process, the result of $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}$ becomes $\boldsymbol{I}$.

Since L-BFGS is an adaptive optimization algorithm and free of learning rate selection, it is used to minimize the cost function in Eq. (11). We calculate the gradient of J_{NSAE} with respect to $\boldsymbol{W}$:

$$\nabla \boldsymbol{J} = \frac{\partial J_{\mathrm{NSAE}}}{\partial \boldsymbol{W}} = \left[\left(\frac{1}{M}\boldsymbol{W}\boldsymbol{D} + \lambda \cdot \boldsymbol{sgn}\right) \cdot \sigma_{\mathrm{r}}'(\boldsymbol{W}\boldsymbol{x})\right]\boldsymbol{x}^{\mathrm{T}} + \frac{1}{M}\sigma_{\mathrm{r}}(\boldsymbol{W}\boldsymbol{x})\boldsymbol{D}^{\mathrm{T}} \tag{12}$$

where $\boldsymbol{D} = \boldsymbol{W}^{\mathrm{T}}\sigma_{\mathrm{r}}(\boldsymbol{W}\boldsymbol{x}) - \boldsymbol{x}$, $\boldsymbol{sgn}$ is the result of the sign function of $\sigma_{\mathrm{r}}(\boldsymbol{W}\boldsymbol{x})$, σ_{r}' is the derivative function of ReLU, and $\boldsymbol{x}$ is the matrix form of $\boldsymbol{x}_m$.

Based on L-BFGS algorithm, the update process of $\boldsymbol{W}$ can be described as

$$\boldsymbol{W}^{(i+1)} = \boldsymbol{W}^{(i)} - \eta^{(i)}\boldsymbol{H}^{(i)}\nabla \boldsymbol{J}^{(i)} \tag{13}$$

where i is the ith interval of the update process, $\boldsymbol{H}^{(i)}$ is the inverse of Hessian matrix, and η is the step size. The computation of $\boldsymbol{H}^{(i)}$ is detailed in Ref. [24], and η can be automatically determined by

$$\eta^{(i)} = \arg\min_{\eta>0} J(\boldsymbol{W}^{(i)} - \eta\boldsymbol{H}^{(i)}\nabla \boldsymbol{J}^{(i)}) \tag{14}$$

After the training process of NSAE, its weight matrix is actually normalized by the soft orthonormality constraint. This is the reason why we call such a method normalized sparse autoencoder.

3.2. Local connection network constructed by NSAE

As shown in Fig. 1, although the first sample and the second sample share the shift variant properties, the local properties of the first sample and the second sample are similar. Thus, to deal with the shift-variance classification problem, we can first use NSAE to extract the local features from the fault signals, and then averaged these local features into the features. In this way, the obtained features can enhance local properties of the fault signals and suppress random noise carried by the signals and thus they can present the shift-invariant properties. Finally, the obtained features are input into the softmax classifier for classification. We model these processes as a form of neural network and call it LCN. In this section, we detail the construction of LCN for intelligent fault diagnosis using NSAE, namely NSAE-LCN. As shown in Fig. 2, LCN has an architecture containing four layers: input layer, local layer, feature layer and output layer. Firstly, raw vibration signals are fed into the input layer. Then, in the local layer, NSAE is used to extract local features from the vibration signals. Thirdly, the feature layer is used to average the local features so as to learn the shift-invariant features of the vibration signals. Finally, the shift-invariant features are input into the output layer and the health conditions of the vibration signals are determined. The details of these layers are described as follows.

Input layer: Use raw vibration signals as its inputs.

The raw vibration signals are collected from machines, and are used to compose the training set $\{\boldsymbol{x}_m, y_m\}_{m=1}^{M}$. $\boldsymbol{x}_m \in \Re^{N\times 1}$ is the mth sample having N data points, and y_m is the mth label indicating the health condition of $\boldsymbol{x}_m$. We alternately divide $\boldsymbol{x}_m$ into J segments and input the segments of $\boldsymbol{x}_m$ into LCN.

$$\boldsymbol{x}_m = \left[\boldsymbol{x}_m^1,\ \boldsymbol{x}_m^2,\ \ldots,\ \boldsymbol{x}_m^J\right] \tag{15}$$

$$J = N/N_{\mathrm{loc}}, \qquad \exists J \in \mathrm{N}^+ \tag{16}$$

where $\boldsymbol{x}_m^j \in \Re^{N_{\mathrm{loc}}\times 1}$ and $j = 1, 2, \ldots, J$.

Local layer: Learn the local features of the signals.

In this layer, the local weight matrix $\boldsymbol{W}_{\mathrm{loc}} \in \Re^{K_{\mathrm{loc}}\times N_{\mathrm{loc}}}$ of LCN is used to get the local features from $\boldsymbol{x}_m^j$. To train $\boldsymbol{W}_{\mathrm{loc}}$, we first select the segment $\boldsymbol{x}_s \in \Re^{N_{\mathrm{loc}}\times 1}$ from the training samples $\{\boldsymbol{x}_m\}_{m=1}^{M}$ randomly. These selected segments compose the local training set $\{\boldsymbol{x}_s\}_{s=1}^{N_s}$ where N_s is the number of the segments. Then, NSAE is employed to calculate $\boldsymbol{W}_{\mathrm{loc}}$. After the segment $\boldsymbol{x}_s$ is preprocessed by whitening [25], the $\boldsymbol{W}_{\mathrm{loc}}$ is used to transform $\boldsymbol{x}_s$ from original space into the feature $\boldsymbol{f}_s$, and $\boldsymbol{W}_{\mathrm{loc}}^{\mathrm{T}}$ is used to reconstruct $\boldsymbol{x}_s$ as $\hat{\boldsymbol{x}}_s$. Based on Eq. (11), this objective function is given as

$$\min_{\boldsymbol{W}_{\mathrm{loc}}} \sum_{s=1}^{N_s} \left\|\boldsymbol{W}_{\mathrm{loc}}^{\mathrm{T}}\sigma_{\mathrm{r}}(\boldsymbol{W}_{\mathrm{loc}}\boldsymbol{x}_s) - \boldsymbol{x}_s\right\|_2^2 + \lambda \sum_{s=1}^{N_s}\sum_{k=1}^{K_{\mathrm{loc}}} \left\|\sigma_{\mathrm{r}}(\boldsymbol{W}_{\mathrm{loc}}^{k}\boldsymbol{x}_s)\right\|_1 \tag{17}$$

where $\boldsymbol{W}_{\mathrm{loc}}^{k}$ is the kth row vector of $\boldsymbol{W}_{\mathrm{loc}}$. By minimizing the objective function in Eq. (17), $\boldsymbol{W}_{\mathrm{loc}}$ can be obtained. LCN can get the local features of the signal $\boldsymbol{x}_m$ using $\boldsymbol{W}_{\mathrm{loc}}$. Thus, the local feature $\boldsymbol{f}_{\mathrm{loc}}^{m,j} \in \Re^{K_{\mathrm{loc}}\times 1}$ of the segment $\boldsymbol{x}_m^j$ is calculated as

$$\boldsymbol{f}_{\mathrm{loc}}^{m,j} = \sigma_{\mathrm{r}}\left(\boldsymbol{W}_{\mathrm{loc}} \cdot \boldsymbol{x}_m^j\right). \tag{18}$$

In order to express the process concisely, we rewrite it using the matrix form. The weight matrix $\boldsymbol{W}_1 \in \Re^{K\times N}$ of the local layer is composed by J local weight matrix $\boldsymbol{W}_{\mathrm{loc}}$.

$$\boldsymbol{W}_1 = \begin{bmatrix} \boldsymbol{W}_{\mathrm{loc}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_{\mathrm{loc}} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{W}_{\mathrm{loc}} \end{bmatrix} \tag{19}$$

$$J = N/N_{\mathrm{loc}} = K/K_{\mathrm{loc}} \tag{20}$$

Feature layer: Get the shift-invariant features.

We use an average strategy to deal with the shift variant properties of the sample $\boldsymbol{x}_m$ and produce its shift-invariant features. In the local layer, we have obtained the local feature $\boldsymbol{f}_{\mathrm{loc}}^{m,j}$ of the segment $\boldsymbol{x}_m^j$ of $\boldsymbol{x}_m$. So the shift-invariant feature $\boldsymbol{f}_m$ of $\boldsymbol{x}_m$ is obtained by averaging J local feature $\boldsymbol{f}_{\mathrm{loc}}^{m,j}$. Thus, the feature $\boldsymbol{f}_m$ of $\boldsymbol{x}_m$ can be obtained by

$$\boldsymbol{f}_m = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{f}_{\mathrm{loc}}^{m,j} \tag{21}$$

Output layer: Determine the labels of input signals.

Softmax layer is used as the output layer of LCN since it is easy to be implemented and computed fast. Suppose that we have the training set $\{\boldsymbol{f}_m\}_{m=1}^{M}$ with its label set $\{y_m\}_{m=1}^{M}$ where $y_m \in \{1, 2, \ldots, R\}$. For each $\boldsymbol{f}_m$, the softmax layer attempts to estimate the probability $p(y_m = r|\boldsymbol{f}_m)$ for each label of $r = 1, 2, \ldots, R$. Thus, the softmax layer is trained by minimizing the objective function:

$$J(\boldsymbol{W}_2) = -\frac{1}{M}\left[\sum_{m=1}^{M}\sum_{r=1}^{R} 1\{y_m = r\} \log \frac{e^{(\boldsymbol{W}_2^r)^{\mathrm{T}}\boldsymbol{f}_m}}{\sum_{l=1}^{R} e^{(\boldsymbol{W}_2^l)^{\mathrm{T}}\boldsymbol{f}_m}}\right] \tag{22}$$

where $1\{\cdot\}$ denotes the indicator function returning 1 if the condition is true, and 0 otherwise, $\boldsymbol{W}_2$ is the weight matrix of the softmax layer, and $\boldsymbol{W}_2^r$ is the row vectors of $\boldsymbol{W}_2$.

In the previous training process of NSAE-LCN, the weight matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are trained separately. After this training process, fine-tuning strategy can be chosen to simultaneously update the weight matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ of NSAE-LCN so as to improve the classification accuracies.

Table 1
The classification accuracies of NSAE-LCN using various parameters in the gearbox diagnosis case.

N_{loc}	K_{loc} (%)				
	50	75	100	150	200
50	98.48	99.05	98.99	99.22	99.35
75	94.08	98.86	99.24	99.20	99.31
100	87.56	97.34	99.20	99.25	99.29
150	55.32	84.58	94.86	99.16	99.26
200	36.42	64.36	79.85	95.69	98.71

Table 2
The cost time of NSAE-LCN using various parameters in the gearbox diagnosis case.

N_{loc}	K_{loc} (min)				
	50	75	100	150	200
50	1.31	1.80	2.34	3.37	4.27
75	1.53	1.94	2.53	3.63	4.98
100	1.87	2.40	2.79	4.01	5.04
150	2.40	3.07	3.52	4.64	5.59
200	2.64	3.26	3.71	5.02	6.13

4. Fault diagnosis using NSAE-LCN

4.1. Case study 1: fault diagnosis of planetary gearboxes

4.1.1. Data collection

The data were acquired from a two-stage planetary gearbox. This planetary gearbox operated under 7 health conditions, involving a flaked needle on the planet bearing of the first stage, a pitted tooth on the sun gear of the first stage, a cracked tooth on the planetary gear of the first stage, a bearing inner race fault of a planetary gear of the first stage, a missing tooth on the sun gear of the second stage, a chipped tooth on the sun gear of the second stage, and the normal condition. The gearbox under the first health condition operated under 6 operating conditions and the gearbox in other health conditions dividedly operated under 8 operating conditions: 4 different drive motor speeds (2100 rpm, 2400 rpm, 2700 rpm and 3000 rpm) combined with 2 loading conditions (no load and the maximum load). The accelerometer mounted on the gearbox was employed to collect data with a sampling frequency of 5120 Hz. There are 335 samples collected for each health condition under an identical operating condition, and each sample is a signal containing 1800 data points. So the gearbox dataset contains 18,090 samples and is a seven-class classification problem.

4.1.2. Parameter study of the proposed NSAE-LCN

There are totally three parameters in the proposed LCN, i.e., N_{loc}, K_{loc}, and λ. So we investigate the classification results of NSAE-LCN using various values of these parameters. To classify the health conditions of the planetary gearbox using NSAE-LCN, 25% of samples are randomly selected for training, 40,000 segments are selected from these samples to train NSAE and other samples are used for testing. It is noted that each of the following experiments is conducted in 15 trials so as to reduce the effect of randomness, and the displayed results are averaged by the results of these 15 trials.

N_{loc} and K_{loc} determine the architecture of NSAE-LCN. N_{loc} denotes the dimension of a signal segment and measures how many segments a signal would be divided into. K_{loc} determines the dimension of the local feature, and it denotes how many bases $\boldsymbol{W}_{loc}$ contains. Fine-tuning is not used in this section since we pay attention to self-learning ability of NSAE under different parameters. Table 1 shows the classification accuracies of the NSAE-LCN based on various N_{loc} and K_{loc} when λ is fixed as 0.25. It can be seen that the diagnosis accuracies are relatively high when $N_{loc} \leq K_{loc}$. So NSAE-LCN is able to robustly obtain the good results if we set K_{loc} not less than N_{loc}. However, as shown in Table 2, the larger K_{loc} is, the larger the architecture of NSAE-LCN is and the more the time is spent. Thus, it is actually a tradeoff between higher accuracies and less time when we choose the parameter N_{loc} and K_{loc}.

Another parameter, λ, is the regular parameter used in the NSAE-LCN. The diagnosis accuracies using different values of λ are shown in Fig. 3. In this figure, N_{loc} and K_{loc} equal 100 and the accuracy distributions of the 15 trials using different λ are clearly displayed. It can be seen that when λ equals 0.1–0.4, the mean accuracy of the proposed NSAE-LCN is over 98% and varies little. Thus, λ can be selected loosely and is advised to be from 0.1 to 0.4.

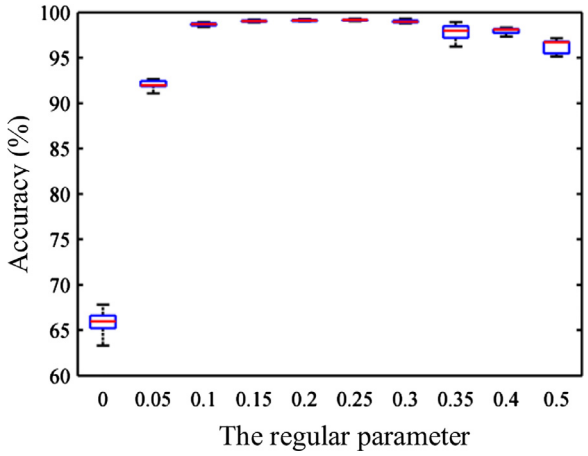


Fig. 3. Classification accuracies of NSAE-LCN for the gearbox dataset using different regular parameter.

Table 3
The comparison of classification accuracies (mean accuracy ± standard deviation) of the gearbox dataset.

Methods	Description	Training accuracy	Testing accuracy
1	Features + Softmax	86.20% ± 0.609%	85.87% ± 0.404%
2	Features + SVM	97.14% ± 0.451%	95.54% ± 0.216%
3	PCA + SVM	100%	41.04% ± 1.165%
4	Stacked SAE + Softmax	48.63% ± 2.086%	34.75% ± 0.436%
5	SAE-LCN	98.89% ± 0.471%	94.41% ± 0.537%
6	NSAE-LCN	**100%**	**99.43% ± 0.105%**

4.1.3. Diagnosis results

Based on the analysis in the previous section, N_{loc} and K_{loc} are both set as 100 and λ is selected as 0.25. As shown in Table 3, the training accuracy of the proposed NSAE-LCN is 100% and the testing accuracy is 99.43%, and the corresponding standard deviation values are 0% and 0.105%, respectively. The results indicate that NSAE-LCN is able to directly classify the health conditions of the gearbox from the raw vibration signals and obtains high and stable classification accuracies.

To show the effectiveness of NSAE-LCN, we compare it with several diagnosis methods. In Method 1, five features specifically designed for gearboxes [26], i.e., FM0, FM4, energy ratio (ER), sideband index (SI) and sideband level factor (SLF), five features commonly used in gearbox diagnosis, i.e., root mean square, energy operator, skewness, kurtosis and crest, and four energy ratios of wavelet coefficients are combined to characterize the health conditions of the planetary gearbox. Based on these manual features, the health conditions are classified by the softmax layer using these fourteen normalized features. In Method 2, the health conditions

are classified by support vector machine (SVM) using these manual features and the two parameters of SVM are optimized by the grid search method. As displayed in Table 3, the diagnosis accuracies of Method 1 are 86.20% for training and 85.87% for testing, and the standard deviation values are 0.609% for training and 0.404% for testing. The diagnosis accuracies of Method 2 are 97.14% for training and 95.54% for testing, and the standard deviation values are 0.451% for training and 0.216% for testing. Comparing Methods 1 and 2 with Method 6, it is illustrated that the features automatically extracted by NSAE can help NSAE-LCN obtain higher and more stable diagnosis accuracies in classifying the gearbox dataset. In addition, the two compared methods require the manual feature extraction using human knowledge on the vibration signals of planetary gearboxes, whereas NSAE learns the features automatically from the raw vibration signals using less prior knowledge. Thus, NSAE releases us from the tough work of designing feature extraction algorithms.

In Method 3, principal component analysis is used for learning features and SVM is used for classification. As shown in Table 3, the training accuracy of the compared method using Method 3 is 100%, and the testing accuracy is 41.04% with 1.165% standard deviation. It is seen that a serious overfitting phenomenon occurs in Method 3. The main reason is that the features extracted by PCA are linear, so they have limited abilities in characterizing the health conditions of machines. In Method 4, stacked SAE and softmax layer constitute a deep neural network, whose training accuracy is 48.63% with 2.086% standard deviation and the testing accuracy is 34.75% with 0.436% standard deviation. It indicates that autoencoders are unsuitable for learning features from the raw vibration signals directly because of the shift variant properties of the signals. So in Method 5, SAE combined with the architecture of LCN, namely SAE-LCN, is used for intelligent diagnosis of the gearboxes. The accuracies of SAE-LCN are 98.89% and 94.41%, and the standard deviation values are 0.471% and 0.537%. It indicates that the LCN helps to produce shift-invariant features of vibration signals and improve the diagnosis accuracies. By comparing Methods 5 with 6, it shows that a certain over-fitting phenomenon occurs in the training processes of SAE-LCN and the testing accuracies of NSAE-LCN are higher than SAE-LCN, which demonstrates that NSAE is able to learn better features than SAE.

We also compare the receiver operating characteristic (ROC) curves of Methods 1–6. ROC curves [27] are used to visualize the performance of the classifier since they measure the ability of a classifier to produce good relative sample scores that serve to discriminate positive and negative samples. Fig. 4 shows the ROC curves of the methods. The best classifier will yield the point in the upper left corner of the ROC space, representing no false negatives and no false positives. An untrained classifier will achieve the dashed line in the ROC space, representing the random performance of the classifier. To quantitatively measure the performance, we calculate the area under the ROC curve (AUC). So the bigger AUC of a classifier is, the better it will perform. As we can see, the AUCs of NSAE-LCN are larger than other methods. It shows the superiority of the proposed NSAE-LCN in classifying the health conditions of the planetary gearbox under various operation conditions.

4.1.4. Properties of NSAE-LCN

Here, we try to understand the feature learning process of NSAE-LCN. As we know, a neural network is always viewed as a black box. And in the field of fault diagnosis, few papers attempt to discover the patterns in the weights of autoencoders. So understanding the properties of these patterns is extremely essential. As shown in Fig. 5, we rewrite the encoder part of NSAE in Eq. (18) as the linear transformation form

$$f_k^{m,j} = \boldsymbol{W}_{\text{loc}}^k \cdot \boldsymbol{x}_m^j \tag{23}$$

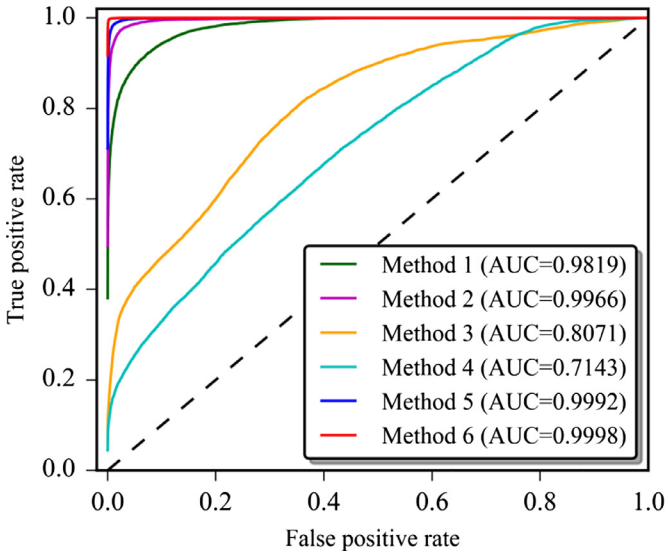


Fig. 4. ROC curves and AUCs of various methods for the gearbox dataset.

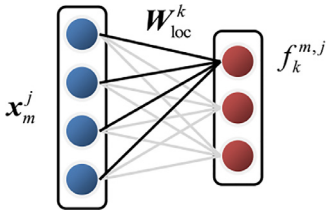


Fig. 5. The encoder of NSAE.

where $k = 1, 2, \ldots, K_{\text{loc}}$, $\boldsymbol{W}_{\text{loc}}^k$ is the kth row vector of $\boldsymbol{W}_{\text{loc}}$, and $f_k^{m,j}$ is the kth element of $\boldsymbol{f}_{\text{loc}}^{m,j}$. Eq. (23) indicates that the value of the kth output neuron is obtained by the inner product operation of the signal segment $\boldsymbol{x}_m^j$ and the row vector $\boldsymbol{W}_{\text{loc}}^k$. And the features in Eq. (18) are just the nonlinear transformation form of the features in Eq. (23). In this view, $\boldsymbol{W}_{\text{loc}}$ can be regarded as a set of normal orthogonal bases, and thus the local feature learning in the local layer of LCN can be interpreted as decomposing the signal segment in the normal orthogonal system $\boldsymbol{W}_{\text{loc}}$.

We randomly select 15 row vectors of $\boldsymbol{W}_{\text{loc}}$ trained by NSAE and plot them in Fig. 6(a). The corresponding frequency spectra of these row vectors are displayed in Fig. 6(b). It can be seen that the row vectors show local striking in the time domain and they have narrow spectral bandwidth in the frequency domain. So these row vectors have the time-frequency properties and show similarities to one-dimensional Gabor functions, which serve as good bandpass bases for mechanical signals. For comparison, we randomly plot 15 row vectors of $\boldsymbol{W}_{\text{loc}}$ trained by SAE in Fig. 6(c) and (d). It can be seen that these vectors also have some time-frequency properties but the properties are too similar.

We use $\boldsymbol{W}_{\text{loc}}\boldsymbol{W}_{\text{loc}}^{\text{T}}$ to measure the similarity of the row vectors of NSAE and SAE and display the results in Fig. 7. It shows that the inner product between the same row vector of NSAE approaches to 1 and the inner product between the two different vectors is close to 0, which means that NSAE is able to encourage the learned weight matrix to have clear and different patterns. However, the inner product results show that the different vectors of the weight matrix of SAE have unclear and similar patterns. Such properties prevent SAE from learning the various meaningful features. So SAE performs worse than NSAE. Considering the accuracies of Methods 5 and 6 in Table 3, the results demonstrate that the clearer and more different the time-frequency properties of the trained

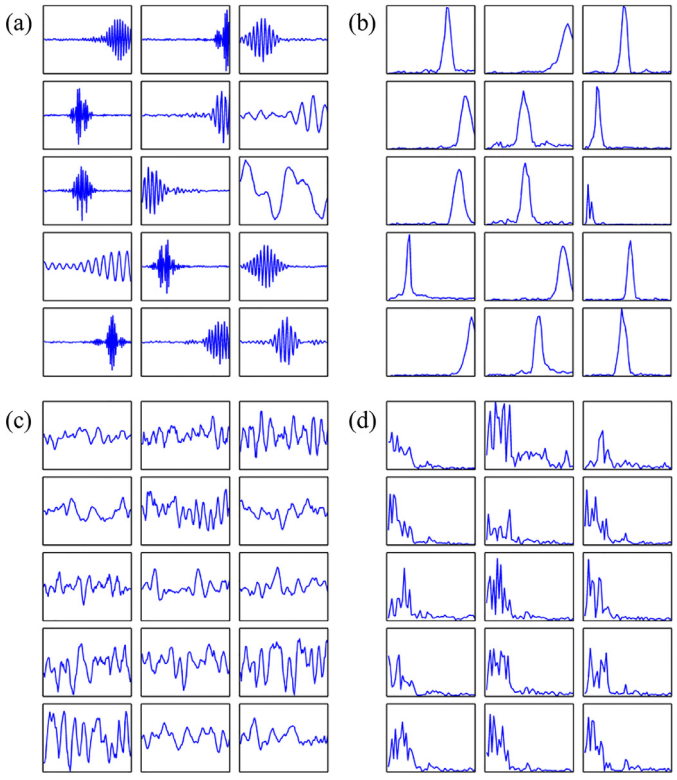


Fig. 6. Row vectors of $\boldsymbol{W}_{\text{loc}}$ for the gearbox dataset: (a) the vectors of NSAE in the time domain, (b) the vectors of NSAE in the frequency domain, (c) the vectors of SAE in the time domain, and (d) the vectors of SAE in the frequency domain.

weight matrix are, the better the methods would perform in the intelligent fault diagnosis of machines. Therefore, NSAE encourages learned features to be meaningful and dissimilar so as to improve the classification accuracies.

One way to understand the time-frequency properties of the weight matrix of NSAE is to tile them in a time-frequency plane. As we know, in the time-domain, frequency discrimination is completely ignored to get the maximum temporal discrimination. And in the frequency domain, temporal discrimination is sacrificed to maximize the information about the frequency composition. Wavelet or other time-frequency schemes are between the two extremes. For wavelet bases, the temporal width becomes narrower as frequency increases, giving the bases the ability to capture the self-similar structure of signals [28]. To analyze the time-frequency properties of the row vectors of the weight matrix of NSAE, we use the method described in Ref. [29] to tile these row vectors in the time-frequency space, as shown in Fig. 8(a). It can be seen that the distribution of these vectors is similar to the wavelet scheme: the temporal width decreases as the bandwidth increases. This perspective can give us more insight to understand NSAE-LCN in the fault diagnosis of machines and may inspire the researches to employ an alternative perspective on other neural networks used in the fault diagnosis. Besides, in the previous analysis, we conclude that K_{loc} should be equal or larger than N_{loc}. In other words, complete or overcomplete representation [30] is better to be used in LCN. Here, we use the tiling of the time-frequency plane to explain the reasons. In Fig. 8(a), the learned row vectors cover all the range of times and frequencies and fulfill the time-frequency plane when K_{loc} equals N_{loc}, producing the complete representation of the input signals. When K_{loc} is smaller than N_{loc}, the corresponding time-frequency plane is shown in Fig. 8(b). It can be seen that although the learned row vectors are approximately orthogonal, their distribution cannot cover the ranges of low frequencies in the time-frequency plane, leading to the lower diagnosis accuracies. So the complete representation strategy is chosen in this paper.

4.2. Case study 2: fault diagnosis of motor bearings

4.2.1. Data description

Here, we apply the proposed NSAE-LCN to diagnose the motor bearing dataset provided by Case Western Reserve University [31]. In the bearing dataset, the vibration signals were collected from the drive end of a motor in the test rig under four different conditions: normal condition, outer race fault, inner race fault, and roller fault. For each fault case, three different severity levels (0.18, 0.36, and 0.53 mm) were introduced. The test rig was operated under four load conditions (0, 1, 2 and 3 hp) and the sampling frequency was 12 kHz. Therefore, this bearing dataset totally contains ten bearing health conditions under the four loads. There are 100 samples for each health condition under one load and each

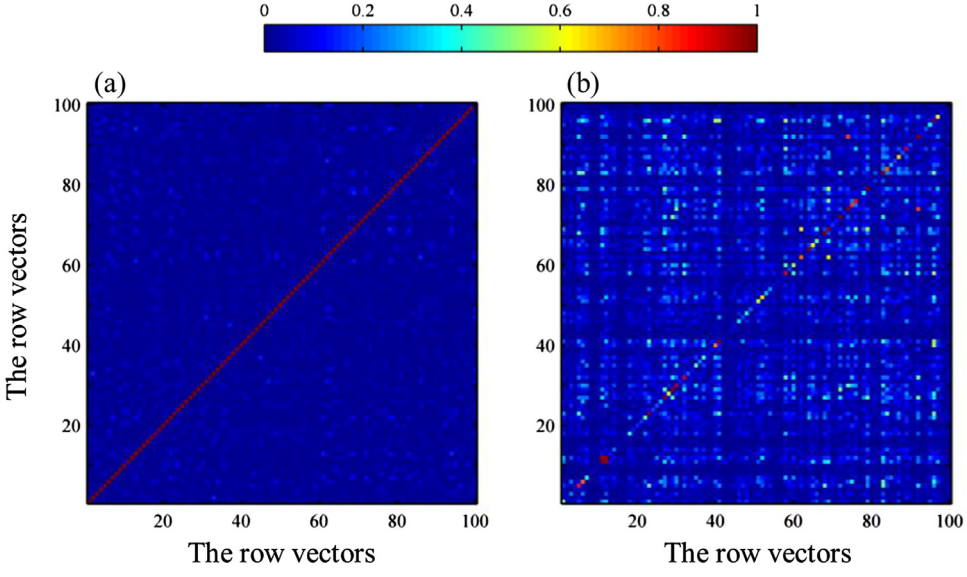


Fig. 7. The results of $\boldsymbol{W}_{\text{loc}}\boldsymbol{W}_{\text{loc}}^{\text{T}}$ for the gearbox dataset: (a) NSAE and (b) SAE.

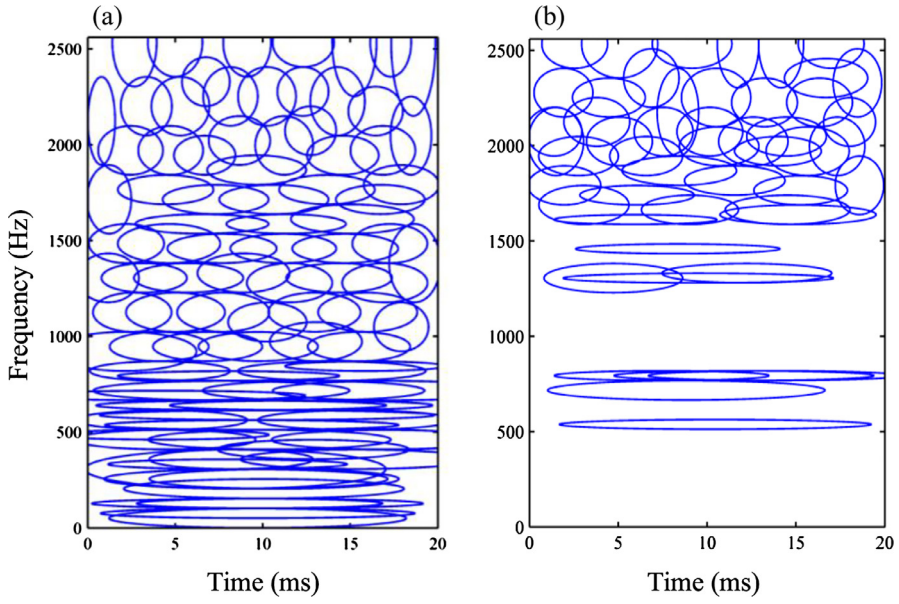


Fig. 8. Time-frequency analysis of $\boldsymbol{W}_{\text{loc}}$ of NSAE for gearbox dataset: (a) K_{loc} equals 100 and (b) K_{loc} equals 50.

Table 4
Classification comparison of the bearing dataset.

Methods	Description	No. of classes	Training samples	Testing accuracy
[32]	EEMD features + SVM	6	25%	97.04%
[33]	EMD-WKLFDA + SVM	10	40%	98.80%
[34]	Bi-spectrum features + SVM	4	50%	96.98%
[35]	MFDFA features + MDC	12	*N/A*	96.67%
[36]	SAX based features + NN	10	50%	99.43%
[37]	Multiple domain features + ensemble NNs	7	50%	99.07%
[38]	Spectra + Deep neural networks	4	*N/A*	94.73%
Proposed	NSAE-LCN	10	25%	**99.92%**

sample is a vibration signal of bearings containing 1200 data points. Therefore, the dataset totally contains 4000 samples and is a ten-class classification problem.

4.2.2. Diagnosis results and comparisons

We use 25% of samples of the bearing dataset to train the proposed NSAE-LCN and use the rest to test. The parameters of NSAE-LCN used for the bearing dataset are the same as those used for the gearbox dataset. The testing accuracy of NSAE-LCN is shown in Table 4. It can be seen that NSAE-LCN obtains 99.92% testing accuracy in classifying the bearing dataset. Fig. 9 shows the inner product results of the weight matrix, and it can be seen that the majority of the row vectors of the weight matrix are approximately orthogonal.

Since the bearing dataset is a benchmark in mechanical fault diagnosis, we compare the proposed NSAE-LCN with the methods using the same dataset in the published papers, where all those methods are thoroughly developed. The comparisons are shown in Table 4. In Ref. [32], ensemble empirical mode decomposition (EEMD) was used to extract features and an optimized SVM was applied to classify the six bearing faults and finally 97.04% accuracy was achieved. Van and Kang [33] integrated empirical mode decomposition (EMD), wavelet kernel local fisher discriminant analysis (WKLFDA) and SVM to distinguish ten health conditions and got 98.80% testing accuracy. Saidi et al. [34] used bi-spectrum features to get features and SVM to classify four bearing health conditions, and a classification accuracy of 96.98% was obtained. Lin and Chen [35] proposed a method using multifractal detrended fluctuation analysis (MFDFA) to extract features and Mahalanobis Distance classifier (MDC) to diagnose faults, and they got 96.67% accuracy.

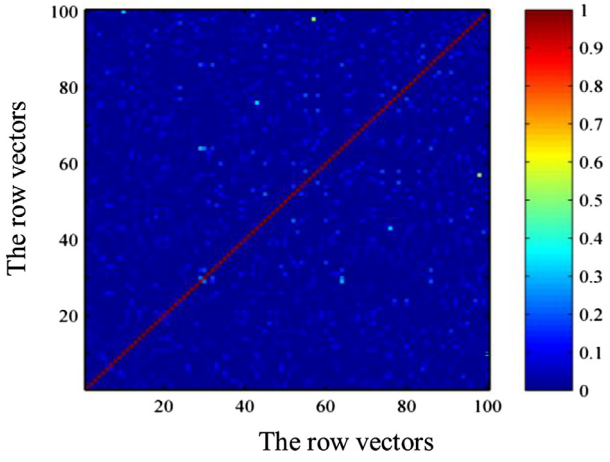


Fig. 9. The results of $\boldsymbol{W}_{\text{loc}}\boldsymbol{W}_{\text{loc}}^{\text{T}}$ of NSAE for the bearing dataset.

Based on symbolic aggregate approximation (SAX) features and a neural network (NN), Georgoulas et al. [36] got 99.43% accuracy in classifying ten bearing health conditions. Xu et al. [37] used multiple domain features to characterize the health conditions of bearings and applied ensemble fuzzy ARTMAP neural networks to recognize the health conditions, and they obtained 99.07% accuracy. Lu et al. [38] applied deep neural networks for feature learning and fault classification from the spectra of vibration signals and 94.73% accuracy was obtained. In the most of the methods above, plenty of researchers took their effort on manual feature extraction

algorithms using signal processing techniques, like Fourier transform, EEMD, high order spectrum and wavelet analysis. Although such processes make full use of human knowledge on fault diagnosis, too much human labor would be spent in analyzing mechanical signals and grasping their properties. As the collected data become larger, it is increasingly difficult to manually design the feature extraction algorithm. Compared with these methods, the proposed NSAE-LCN integrates the feature learning and fault diagnosis in one network so as to extract features and classify the faults automatically, which would make the proposed method less dependent on prior knowledge and human labor. So the proposed method performs more intelligently compared with these methods and obtains higher classification accuracies in the bearing diagnosis case.

5. Conclusions

To release us from the task of designing feature extraction algorithms, a method using NSAE-LCN is proposed for intelligent fault diagnosis of machines. In this method, LCN first uses NSAE to locally learn various meaningful features from vibration signals, and then LCN produces shift-invariant features based on the learned features and classifies the health conditions of machines. Since NSAE-LCN uses vibration signals as the input and classifies mechanical health conditions automatically, this method does not rely on much human labor and can be used for the different diagnosis cases directly. A gearbox dataset and a bearing dataset are used to verify the proposed NSAE-LCN. The results show that the learned features of NSAE are meaningful and dissimilar, and LCN helps to produce shift-invariant features and recognizes mechanical health conditions effectively. Through comparing with commonly used diagnosis methods, the superiority of the proposed NSAE-LCN is verified.

In this paper, the parameters of NSAE-LCN are determined manually based on the study of parameter selection. The automatic selection of these parameters may make NSAE-LCN easily to be utilized in intelligent fault diagnosis of machines and even raise its performance. Thus, the optimization techniques such as PSO [39,40] will be used to adaptively adjust the parameters of NSAE-LCN in future work.

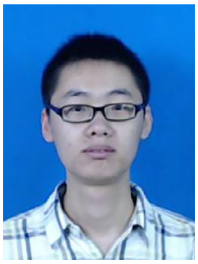
Acknowledgments

This research was supported by National Natural Science Foundation of China (51475355 and 61673311) and National Program for Support of Top-notch Young Professionals.

References

[1] P.K. Wong, J. Zhong, Z. Yang, C.M. Vong, Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis, Neurocomputing 174 (2016) 331–343.
[2] L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings, Neurocomputing 240 (2017) 98–109.
[3] F. Pacheco, J.V. de Oliveira, R.-V. Sánchez, M. Cerrada, D. Cabrera, C. Li, G. Zurita, M. Artés, A statistical comparison of neuroclassifiers and feature selection methods for gearbox fault diagnosis under realistic conditions, Neurocomputing 194 (2016) 192–206.
[4] J. Liu, W. Wang, F. Golnaraghi, An enhanced diagnostic scheme for bearing condition monitoring, IEEE Trans. Instrum. Meas. 59 (2010) 309–321.
[5] P. Henriquez, J.B. Alonso, M.A. Ferrer, C.M. Travieso, Review of automatic fault diagnosis systems using audio and vibration signals, IEEE Trans. Syst. Man Cybern. A: Syst. 44 (2014) 642–652.
[6] Y. Lei, F. Jia, J. Lin, S. Xing, S.X. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, IEEE Trans. Ind. Electron. 63 (2016) 3137–3147.
[7] C. Shen, D. Wang, F. Kong, P.W. Tse, Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier, Measurement 46 (2013) 1551–1564.
[8] G. Georgoulas, V. Climente, J.A. Antonino-Daviu, I. Tsoumas, C. Stylios, A. Arkkio, G. Nikolakopoulos, The use of a multilabel classification framework for the detection of broken bars and mixed eccentricity faults based on the start-up transient, IEEE Trans. Ind. Inform. 13 (2016) 625–634.
[9] M.D. Prieto, G. Cirrincione, A.G. Espinosa, J.A. Ortega, H. Henao, Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks, IEEE Trans. Ind. Electron. 60 (2013) 3398–3407.
[10] M. Amar, I. Gondal, C. Wilson, Vibration spectrum imaging: a novel bearing fault classification approach, IEEE Trans. Ind. Electron. 62 (2015) 494–502.
[11] D. Wang, K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: revisited, Mech. Syst. Signal Process. 70 (2016) 201–208.
[12] Y. Lei, Z. Liu, X. Wu, N. Li, W. Chen, J. Lin, Health condition identification of multi-stage planetary gearboxes using a mRVM-based method, Mech. Syst. Signal Process. 60 (2015) 289–300.
[13] N. Zeng, Z. Wang, H. Zhang, W. Liu, F.E. Alsaadi, Deep belief networks for quantitative analysis of gold immunochromatographic strip, Cogn. Comput. 8 (2016) 684–692.
[14] Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 1798–1828.
[15] N. Zeng, H. Zhang, Y. Li, J. Liang, A.M. Dobaie, Denoising and deblurring gold immunochromatographic strip images via gradient projection algorithms, Neurocomputing 247 (2017) 165–172.
[16] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F.E. Alsaadi, A survey of deep neural network architectures and their applications, Neurocomputing 234 (2017) 11–26.
[17] 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 IEEE Conference on Prognostics and Health Management, 2016, pp. 1–7.
[18] 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.
[19] Z. Chen, W. Li, Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network, IEEE Trans. Instrum. Meas. 66 (2017) 1693–1702.
[20] 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 (2017) 1560–1578.
[21] E. Li, P. Du, A. Samat, Y. Meng, M. Che, Mid-level feature representation via sparse autoencoder for remotely sensed scene classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (2017) 1068–1081.
[22] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: Proceedings of International Conference on Machine Learning, 2010, pp. 807–814.
[23] R. Memisevic, D. Krueger, Zero-bias autoencoders and the benefits of co-adapting features, in: Proceedings of International Conference on Learning Representations, 2015, pp. 1–11.
[24] D.C. Liu, J. Nocedal, On the limited memory BFGS method for large scale optimization, Math. Program. 45 (1989) 503–528.
[25] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Netw. 13 (2000) 411–430.
[26] P.D. Samuel, D.J. Pines, A review of vibration-based techniques for helicopter transmission diagnostics, J. Sound Vib. 282 (2005) 475–508.
[27] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006) 861–874.
[28] B.A. Olshausen, K.N. O'Connor, A new window on sound, Nat. Neurosci. 5 (2002) 292–294.
[29] M.S. Lewicki, Efficient coding of natural sounds, Nat. Neurosci. 5 (2002) 356–363.
[30] B.A. Olshausen, D.J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1? Vis. Res. 37 (1997) 3311–3325.
[31] X. Lou, K.a Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mech. Syst. Signal Process. 18 (2004) 1077–1095.
[32] X. Zhang, J. Zhou, Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines, Mech. Syst. Signal Process. 41 (2013) 127–140.
[33] M. Van, H-J. Kang, Bearing defect classification based on individual wavelet local fisher discriminant analysis with particle swarm optimization, IEEE Trans. Ind. Inform. 12 (2016) 124–135.
[34] L. Saidi, J.B. Ali, F. Fnaiech, Application of higher order spectral features and support vector machines for bearing faults classification, ISA Trans. 54 (2015) 193–206.
[35] J. Lin, Q. Chen, Fault diagnosis of rolling bearings based on multifractal detrended fluctuation analysis and Mahalanobis distance criterion, Mech. Syst. Signal Process. 38 (2013) 515–533.
[36] G. Georgoulas, P. Karvelis, T. Loutas, C.D. Stylios, Rolling element bearings diagnostics using the symbolic aggregate approximation, Mech. Syst. Signal Process. 60 (2015) 229–242.
[37] Z. Xu, Y. Li, Z. Wang, J. Xuan, A selective fuzzy ARTMAP ensemble and its application to the fault diagnosis of rolling element bearing, Neurocomputing 182 (2016) 25–35.
[38] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, T. Zhang, Deep model based domain adaptation for fault diagnosis, IEEE Trans. Ind. Electron. 64 (2017) 2296–2305.
[39] N. Zeng, H. Zhang, W. Liu, J. Liang, F.E. Alsaadi, A switching delayed PSO optimized extreme learning machine for short-term load forecasting, Neurocomputing 240 (2017) 175–182.

[40] N. Zeng, Z. Wang, H. Zhang, F.E. Alsaadi, A novel switching delayed PSO algorithm for estimating unknown parameters of lateral flow immunoassay, Cogn. Comput. 8 (2016) 143–152.



Feng Jia is currently working toward the Ph.D. degree in mechanical engineering at the State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, P.R. China. He received the B.S. and M.S. degree in mechanical engineering from Taiyuan University of Technology, P.R. China, in 2011 and 2014, respectively. His research interests include machinery condition monitoring and fault diagnosis, intelligent fault diagnostics of rotating machinery.

Yaguo Lei received the B.S. and Ph.D. degrees in mechanical engineering from Xi'an Jiaotong University, Xi'an, P.R. China, in 2002 and 2007, respectively. He is currently a Full Professor of mechanical engineering at Xi'an Jiaotong University. Prior to joining Xi'an Jiaotong University in 2010, he was a Postdoctoral Research Fellow with the University of Alberta, Edmonton, AB, Canada. He was also an Alexander von Humboldt Fellow with the University of Duisburg-Essen, Duisburg, Germany. His research interests focus on machinery condition monitoring and fault diagnosis, mechanical signal processing, intelligent fault diagnostics, and remaining useful life prediction. Dr. Lei is a member of the editorial boards of more than ten journals, including Mechanical System and Signal Processing and Neural Computing & Applications. He is also a member of ASME and a member of IEEE. He has pioneered many signal processing techniques, intelligent diagnosis methods, and remaining useful life prediction models for machinery.

Liang Guo received the B.S. and Ph.D. degrees in mechanical engineering from Southwest Jiaotong University, Chengdu, P.R. China, in 2011 and 2016, respectively. He is currently working as a Postdoctoral researcher at the State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an, P.R. China. His current research interests include machinery condition monitoring, intelligent fault diagnostics and remaining useful life prediction.

Jing Lin received his B.S., M.S. and Ph.D. degrees from Xi'an Jiaotong University, P.R. China, in 1993, 1996 and 1999, respectively, all in mechanical engineering. He is currently a Professor with the State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University. From July 2001 to August 2003, he was a Postdoctoral Fellow with the University of Alberta, Edmonton, AB, Canada, and a Research Associate with the University of Wisconsin–Milwaukee, Milwaukee, WI, USA. From September 2003 to December 2008, he was a Research Scientist with the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, under the sponsorship of the Hundred Talents Program. His current research directions are in mechanical system reliability, fault diagnosis, and wavelet analysis. Dr. Lin was a recipient of the National Science Fund for Distinguished Young Scholars in 2011.

Saibo Xing is currently working for the Ph.D. degree in mechanical engineering from Xi'an Jiaotong University, P.R. China and received the B.S. degree in material science and engineering from Xi'an Jiaotong University, in 2015. He graduated from Hsue-shen Tsien Experimental Class majoring in material science and engineering, and became the graduated student without examination majoring in mechanical engineering. His research interests focus on intelligent fault diagnostics and prognostics of rotating machinery.