The Optimized Deep Belief Networks With Improved Logistic Sigmoid Units and Their Application in Fault Diagnosis for Planetary Gearboxes of Wind Turbines

Yi Qin, *Member, IEEE*, Xin Wang, and Jingqiang Zou

***Abstract*—Efficient and accurate planetary gearbox fault diagnosis is the key to enhance the reliability and security of wind turbines. Therefore, an intelligent and integrated approach based on deep belief networks (DBNs), improved logistic Sigmoid (Isigmoid) units, and impulsive features is proposed in this paper. The vanishing gradient problem is an inherent drawback of conventional Sigmoid units, and it usually occurs in the backpropagation process of DBNs, resulting in that the training is considerably slowed down and the classification rate is reduced. To solve this problem, Isigmoid units are designed to combine the merits of unsaturation from leaky rectified linear (LReL) units. The results of handwritten digit recognition experiments show the superiority of Isigmoid over Sigmoid on convergence speed and classification accuracy. Since impulses contain much useful fault information, especially for early failures, an integrated approach using the optimized Morlet wavelet transform, kurtosis index, and soft-thresholding is applied to extract impulse components from original signals to improve the diagnosis accuracy. Then, the features extracted from original signals and impulsive signals are employed to train and test the DBNs with Isigmoid, Sigmoid, and LReL units for comparison. Finally, the results of planetary gearbox fault diagnosis show that Isigmoid has higher comprehensive performance than conventional sigmoid and LReL.**

***Index Terms*—Deep belief networks, improved logistic Sigmoid, impulsive feature, vanishing gradient problem, wind turbine.**

I. Introduction

WIND power is the world's fastest growing renewable energy source, and with the development of wind power industry, fault diagnosis and maintenance for wind turbines become increasingly important [1], [2]. Planetary gearboxes are one of the most important components in mechanical transmission systems of wind turbines. However, harsh operating conditions caused by high speed, heavy load, and high–low temperature give rise to a high probability of machinery failure, and to make matters worse, their complex structures make the fault detection and maintenance more difficult [3]. Once the failures of planetary gearboxes are worsened, it will lead to a serious halt of the whole power transmission chain, and even catastrophic economic losses and casualties [4], [5]. Therefore, in order to prevent the faults from deteriorating, it is necessary to diagnose the faults as early as possible. Various planetary gearbox fault diagnosis methods have been studied. For instance, Wang *et al*. [6] developed an improved envelope analysis method to detect the sun gear crack fault in planetary gearbox. In order to achieve intelligent fault diagnosis for planetary gearboxes, Jia *et al*. [7] proposed a local connection network constructed by normalized sparse autoencoder.

Conventional diagnosis methods are based on techniques of signal processing such as ensemble empirical mode decomposition [8], multicomponent signal decomposition [9], and autocorrelation-based time synchronous averaging [10], but their procedures are complicated and tedious. Besides, the accuracies of conventional diagnosis methods are not satisfactory when they are used by the people with poor experience and knowledge. Therefore, what engineers need most urgently is an accurate and intelligent diagnosis method. By improving the classical deep belief networks (DBNs), this study explores a novel and integrated diagnosis approach to meet these requirements.

Compared with shallow neural networks, such as naive Bayes networks [11], backpropagation neural networks [12], and support vector machine [13], DBNs have better extensive adaptability and mapping capability that are helpful for classification. DBNs have been successfully applied to diagnose various equipments, including vehicle on-board equipment of high-speed trains [14], aircrafts [15], electric locomotive bearing [16], and valves in reciprocating compressors [17], etc., thus DBNs can be qualified for the diagnosis of planetary gearbox.

It is well known that the DBNs learning process includes two stages: the first one is the greedy layer-wise unsupervised

Manuscript received December 2, 2017; revised May 8, 2018 and June 10, 2018; accepted July 3, 2018. Date of publication July 20, 2018; date of current version December 28, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 51675065, in part by Chongqing Research Program of Basic Research and Frontier Technology under Grant cstc2017jcyjAX0459, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018CDQYJX0011. *(Corresponding author: Yi Qin.)*

The authors are with the State Key Laboratory of Mechanical Transmission, College of Mechanical Engineering, Chongqing University, Chongqing 400044, China (e-mail: qy_808@cqu.edu.cn; wx-wangxin@foxmail.com; binggehe@qq.com)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2018.2856205

learning of restricted Boltzmann machines (RBMs) by contrastive divergence (CD) algorithm, and the second one is a supervised finetuning process based on backpropagation algorithm. During the finetuning process, the vanishing gradient problem usually occurs. The magnitude of the derivative of conventional logistic Sigmoid function is less than 1 in the whole range and decreases rapidly with the growth of the absolute values of the inputs. Because the backpropagation algorithm is a gradient-based method, this property of Sigmoid decelerates dramatically the finetuning, especially when the inputs fall into the intervals with small derivatives. To make matter worse, backpropagation computes gradients by the chain rule in a multilayer network, consequently the gradients decrease exponentially with the number of layers, and this results in that the front layers learn slowly and in some cases the DBN converges to a poor local minimum [18].

To overcome the vanishing gradient problem, several methods such as multilevel hierarchy [19], long short-term memory [20], and using rectified linear (ReL) units [21] were proposed. ReL function never suffers from the vanishing gradient problem because of its rectified linearity and bigger mapping range.

Inspired by the property of ReL function, a new transfer function Isigmoid is proposed in this study to relax the problem. Isigmoid combines conventional Sigmoid with ReL function by replacing the gradual gradients with a constant gradient when the absolute values of the inputs are greater than the preset threshold. The Mixed National Institute of Standards and Technology database (MNIST database) of handwritten digits is applied to train and test the DBNs that use Isigmoid and Sigmoid in backpropagation processes, respectively, and the results show that Isigmoid has higher convergence speed and classification rate, which means that the proposed Isigmoid function may be better applied to wind turbine fault diagnosis than the conventional activation functions.

In general, when the gear of a wind turbine planetary gearbox has a damage fault, such as crack, pitting, and wear, the impacts generated by tooth mesh usually occur, it then follows that we can diagnose the gear damage by detecting impulses excited at a specific rate. Thus, a transient detection approach proposed in [22] will be first used to extract the impulsive component from the original signal. Then, various features of both the original signal and the impulsive signals will be applied to train the DBNs. Impulsive features are sensitive to early failures, so they can make up the deficiencies of original features. Finally, the experimental results show that the DBNs based on the Isigmoid function are superior to those based on the conventional Sigmoid and leaky rectified linear (LReL) [18] function not only in test accuracy and convergence speed but also in generalization capacity.

Generally, the flow diagram of the proposed diagnosis approach is shown in Fig. 1. During the training process, vibration signals of gears with artificial faults are collected first, then impulsive signals are extracted from original signals. Various features are extracted from original and impulsive signals respectively to construct samples, and finally they are applied to train DBNs with different transfer functions. In the diagnosis process, the samples that have not been used in the training process are applied to test the trained DBNs.

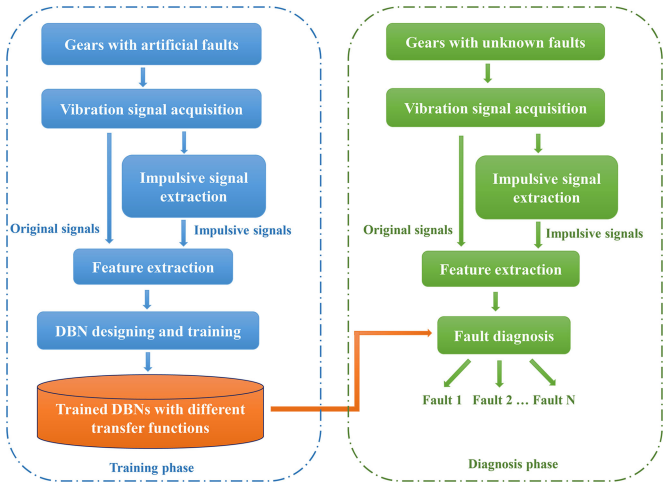

Fig. 1. Flow diagram of the diagnosis approach for planetary gearboxes based on DBNs.

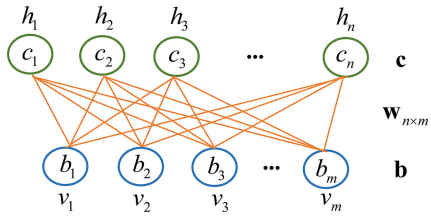

Fig. 2. Architecture of a RBM with m visible neurons and n hidden neurons.

According to Fig. 1, the rest of this paper is organized as follows. Section II briefly presents the architectures and learning rules of DBNs and RBMs. The design of the Isigmoid function and the handwritten digit recognition experiment are described in Section III. Section IV introduces the signal acquisition, impulsive signals extraction, feature extraction, and sample construction. Section V discusses the fault diagnosis experiments and the results. Finally, some conclusions are addressed in Section VI.

II. Deep Belief Network Model

A. Restricted Boltzmann Machines

An RBM is a specific energy-based stochastic model [23] that has two layers, in which visible neurons $\mathbf{v} = (v_1, v_2, \ldots, v_m)$ are connected fully to hidden neurons $\mathbf{h} = (h_1, h_2, \ldots, h_n)$ via symmetrically weighted connections [24]. The architecture of an RBM is shown in Fig. 2.

The energy of the joint configuration $(\mathbf{v}, \mathbf{h})$ is given by the following:

$$E(\mathbf{v}, \mathbf{h}) = -\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} v_i h_j - \sum_{i=1}^{m} b_i v_i - \sum_{j=1}^{n} c_j h_j \quad (1)$$

where w_{ij} is the symmetric weight; v_i, h_j are the binary states, and b_i, c_j are their biases, respectively.

The joint probability for every possible pair of visible and hidden state vector is assigned by the RBM via the energy

function

$$P(\mathbf{v},\mathbf{h}) = \frac{e^{-E(\mathbf{v},\mathbf{h})}}{\sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h})}}. \tag{2}$$

The unbiased samples of v_i and h_j given a hidden vector $\mathbf{h}$ and a visible vector $\mathbf{v}$ can be calculated respectively by logistic Sigmoid functions

$$\begin{aligned} p(v_i = 1|\mathbf{h}) &= \text{Sigmoid}\left(\sum_{j=1}^{n} {w_{ij}}^T h_j + b_i\right) \\ p(h_j = 1|\mathbf{v}) &= \text{Sigmoid}\left(\sum_{i=1}^{m} {w_{ij}}^T v_i + c_j\right) \end{aligned} \tag{3}$$

where the Sigmoid function is defined as follows:

$$f_S(x) = \text{Sigmoid}(x) = 1/\left(1+e^{-x}\right). \tag{4}$$

The derivative of the log-likelihood with respect to the weight w_{ij} is as follows:

$$\frac{\partial \ln P(\mathbf{v})}{\partial w_{ij}} = \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}} \tag{5}$$

where $\langle v_i h_j \rangle_{\text{data}}$ and $\langle v_i h_j \rangle_{\text{model}}$ represent expectation with respect to the data distribution and the model distribution, respectively.

Obtaining an unbiased sample of $\langle v_i h_j \rangle_{\text{model}}$ is computationally intractable, so Hinton proposed the CD algorithm to crudely approximate the gradient [25]

$$\Delta w_{ij} = \eta(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}}) \tag{6}$$

where $\langle v_i h_j \rangle_{\text{recon}}$ represents the expectation of reconstruction states calculated by Gibbs sampler (3) that is initialized by the input data, and η is the learning rate.

B. Finetuning of DBNs

A deep belief network is a kind of generative model, and it consists of multiple RBMs stacked with each other and a top-layer classifier acting as the output layer, whose structure is shown in Fig. 3. After the bottom-up pretraining for the stack of RBMs, the DBN is then finetuned top-down by using the back-propagation algorithm [26], [27]. Back-propagation is a classical method commonly based on gradient descent optimization algorithm to update the weights and bias, which is widely used in various machine learning models. The backpropagation of a pretrained DBN with four hidden layers is illustrated in Fig. 4.

In this paper, the top classifier f_2 is set as Softmax, and f_1 is the activation function of hidden layers, and the cost function C is cross entropy, which is given by the following [28]:

$$\text{C} = -\frac{1}{k}\sum_{j=1}^{k}\left[t_k \log(a_k) + (1-t_k)\log(1-a_k)\right] \tag{7}$$

where k is the number of neurons in the output layer (for simplicity, $k=1$), t is the expected output.

The update of the weight $\mathbf{w}_4$ in the output layer is given by the following:

$$\Delta \mathbf{w}_4 = \eta \frac{\partial \text{C}}{\partial \mathbf{w}_4} = \eta(\mathbf{a}_4 - \mathbf{t})\mathbf{a}_3. \tag{8}$$

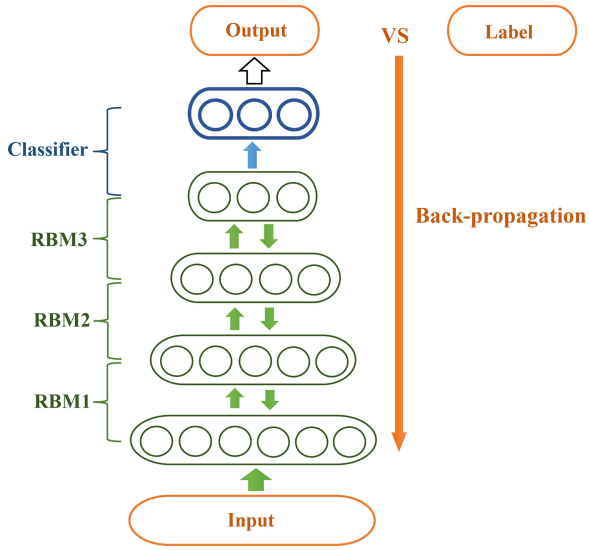

Fig. 3. DBN with three hidden layers.

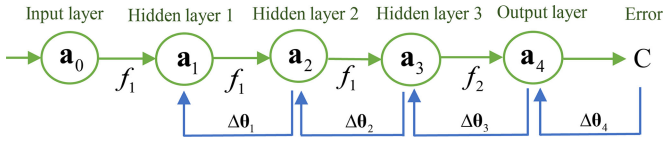

Fig. 4. DBN backpropagation process (where the $\mathbf{a}_i$ is the output of the layer i ($i = 1, 2, 3, 4$), and parameters $\Delta\theta(\Delta\mathbf{w}, \Delta\mathbf{b})$ are the updates of weights $\mathbf{w}$ and bias $\mathbf{b}$. For simplicity, it can be hypothesized that there is only one neuron in each layer.

Then, based on the chain rule, the weights of the hidden layers can be updated by the following:

$$\Delta \mathbf{w}_3 = \eta \frac{\partial \text{C}}{\partial \mathbf{w}_3} = \eta(\mathbf{a}_4 - \mathbf{t})\mathbf{w}_4 f_1'(\mathbf{z}_3)\mathbf{a}_2 \tag{9}$$

$$\Delta \mathbf{w}_2 = \eta \frac{\partial \text{C}}{\partial \mathbf{w}_2} = \eta(\mathbf{a}_4 - \mathbf{t})\mathbf{w}_4 f_1'(\mathbf{z}_3)\mathbf{w}_3 f_1'(\mathbf{z}_2)\mathbf{a}_1 \tag{10}$$

$$\Delta \mathbf{w}_1 = \eta \frac{\partial \text{C}}{\partial \mathbf{w}_1} = \eta(\mathbf{a}_4 - \mathbf{t})\mathbf{w}_4 f_1'(\mathbf{z}_3)\mathbf{w}_3 f_1'(\mathbf{z}_2)\mathbf{w}_2 f_1'(\mathbf{z}_1)\mathbf{a}_0 \tag{11}$$

where $\mathbf{z}_i$ ($i = 1, 2, 3, 4$) is the input $\mathbf{w}_i \times \mathbf{a}_i + \mathbf{b}_i$ of layer i.

It is obvious that the backpropagation learning slows down exponentially as the depth of the network increases, owing to the chain rule and the saturation regime of Sigmoid. Similarly, the bias $\mathbf{b}$ is calculated by the chain rule, which also suffers from the vanishing gradient problem.

III. Isigmoid and Handwritten Digits Recognition

A. Isigmoid

With (4), the derivative of the conventional Sigmoid is calculated by the following:

$$f_S'(x) = e^{-x}/\left(1+e^{-x}\right)^2. \tag{12}$$

The waveforms of Sigmoid and its derivative are shown in Fig. 5.

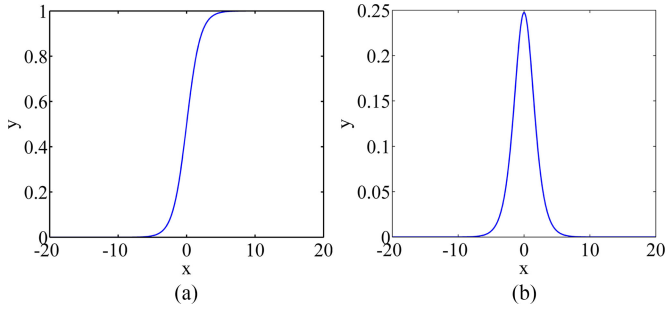

Fig. 5. Waveforms of Sigmoid and its derivative. (a) Sigmoid. (b) The derivative of Sigmoid.

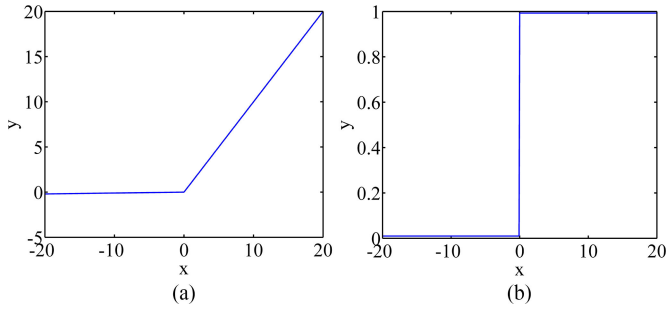

Fig. 6. Waveforms of LReL function and its derivative. (a) LReL. (b) Derivative of LReL.

As can be seen in Fig. 5, the derivative of Sigmoid is almost close to zero beyond the interval of $[-6, 6]$ with the maximum of 0.25, which is the inherent drawback and will cause vanishing gradient problem in the finetuning process.

Therefore, ReL function ($f_R(x) = \mathrm{Max}(0, x)$) is proposed, and it has been validated that the ReL function is superior to the Sigmoid function in the convolutional neural networks, deep neural networks [29], [30]. Moreover, the LReL function [18] is proposed by optimizing the ReL function to achieve better robustness during finetuning [31]. The LReL function and its derivative are mathematically given by (13) and (14), and their waveforms are illustrated in Fig. 6

$$f_{\mathrm{LR}}(x) = \mathrm{Max}(0.01x, x) \tag{13}$$

$$f'_{\mathrm{LR}}(x) = \begin{cases} 1 & x > 0 \\ 0.01 & x < 0 \end{cases}. \tag{14}$$

However, both ReL and LReL functions are not applied to DBNs, because they cannot make full use of the pretraining effects of RBMs. Therefore, in order to solve the vanishing gradient problem during the backpropagation of DBNs, an improved Sigmoid (Isigmoid) that absorbs partly the advantage of LReL function is proposed in this study. Isigmoid function and its derivative are mathematically given by (15) and (16), and their waveforms are illustrated in Fig. 7

$$f_I(x) = \begin{cases} \alpha(x - a) + \mathrm{Sigmoid}(a) & x >= a \\ \mathrm{Sigmoid}(x) & -a < x < a \\ \alpha(x + a) + \mathrm{Sigmoid}(a) & x <= -a \end{cases} \tag{15}$$

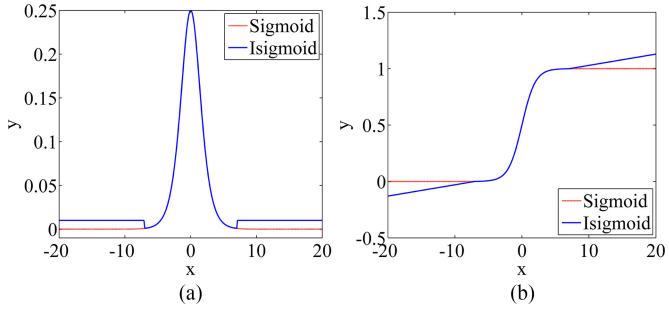

Fig. 7. Waveforms of Sigmoid (red) and Isigmoid (blue) and their derivatives, where the parameters of Isigmoid are set as $\alpha = 7$ and $a = 0.01$. (a) Isigmoid. (b) Derivative of Isigmoid.

$$f'_I(x) = \begin{cases} \alpha & |x| >= a \\ \mathrm{Sigmoid}'(x) & |x| < a \end{cases} \tag{16}$$

where a is the threshold and α is the slope, and both of them are preset.

The second derivative of the conventional logistic Sigmoid function can be calculated by the following:

$$f''_S(x) = \left(e^{-2x} - e^{-x}\right) / \left(1 + e^{-x}\right)^3. \tag{17}$$

It follows from (17) that $f''_S(x) > 0$ when $x < 0$, and $f''_S(x) < 0$ when $x > 0$. Then, we can easily know that $f'_S(x)$ monotonically increases in an interval $(-\infty, 0)$, while it monotonically decreases in an interval $(0, +\infty)$. In other words, given a positive real number a_c, f'_S satisfies the following inequalities:

$$f'_S(a_c) > f'_S(x) \qquad x > a_c \tag{18}$$

$$f'_S(-a_c) > f'_S(x) \qquad x < -a_c. \tag{19}$$

Suppose that the threshold a is set at $a = a_c$, the derivative of Isigmoid satisfies the following:

$$f'_I(x) > f'_S(a) \qquad x > a \tag{20}$$

$$f'_I(x) > f'_S(-a) \qquad x < -a. \tag{21}$$

Then, it follows from (9)–(11) that the update obtained by Isigmoid is larger than that by Sigmoid

$$\Delta \mathbf{w}_{Ii} > \Delta \mathbf{w}_{Si}. \tag{22}$$

According to (22), it is easy to note that the Isigmoid function has higher convergence speed than the sigmoid, and Isigmoid will train the lower layers more completely.

Note that in order to let (20) and (21) hold, the slope α and the threshold a must satisfy the following relationship:

$$\alpha > \alpha_{\min} = e^{-a} / \left(1 + e^{-a}\right)^2 \tag{23}$$

where $\alpha_{\min}$ is the minimum slope to make Isigmoid work.

The above mathematical derivations prove that Isigmoid is capable to relax the vanishing gradient problem in theory. Comparing Sigmoid, LReL, and Isigmoid comprehensively, Sigmoid is the most common transfer function used in the backpropagation process of DBNs, while it usually leads to vanishing gradient problem. LReL have better gradient propagation without the vanishing gradient problem, but it cannot utilize most of

the effects from RBM pretraining. By integrating the advantages of Sigmoid and LReL, Isigmoid is proposed by mitigating saturation to alleviate the vanishing gradient problem, and Isigmoid can be completely suitable for the data reconstruction model obtained by pretraining of RBMs. Thus, Isigmoid will have better performance than Sigmoid and LReL in DBNs.

B. Parameter Presetting

The parameters of Isigmoid function greatly influence the performance of optimized DBNs. In some cases, the DBNs with improper parameters may not get correct recognition results, therefore it is necessary to explore how to set the parameters of DBNs, especially for the parameters of the Isigmoid function. The method for the parameter setting of Isigmoid is proposed in this study and the concrete steps are summarized as follows.

Step 1: Select arbitrarily a part of training samples to build and train a DBN with conventional Sigmoid units, and then tune the parameters till the best performance.

Step 2: Rough estimation for the threshold: The threshold a can be preset roughly according to the distributions of inputs $w_{ij} \times a + b_j$ of the conventional DBN with Sigmoid units. The inputs that are greater than the threshold will be put into linear regions at first, which reflects to a great extent the effects of Isigmoid function. Usually, at least 20% of the inputs should be larger than the smallest a.

Step 3: Optimize the slope based on the rough threshold: Generally, with a given threshold a, as the slope α increases, the classification accuracy of DBN with Isigmoid will increase first and then decrease. For the searching range $[\alpha_{\text{upper}}, \alpha_{\text{lower}}]$, initialize α_{upper} as $\alpha_{\min}$ and set a reasonable α_{lower} that is around 10 times of the α_{upper}, thus the rough searching region is $[\alpha_{\min}, 10\alpha_{\min}]$. In order to narrow the searching interval, it can be divided into p subintervals $[\alpha_{\text{upper}_i}, \alpha_{\text{lower}_j}]$, $i, j = 1, 2, \ldots, p$, then the α_{upper_i} and α_{lower_j} with the highest accuracy respectively is the new searching interval $[\alpha_{\text{upper\_new}}, \alpha_{\text{lower\_new}}]$. If $|\alpha_{\text{upper\_new}} - \alpha_{\text{lower\_new}}|$ is sufficiently small, the best slope can be selected from two endpoints $\alpha_{\text{upper\_new}}$, $\alpha_{\text{lower\_new}}$ and the midpoint $(\alpha_{\text{upper\_new}} + \alpha_{\text{lower\_new}})/2$ by comparing accuracies. If the interval can be divided further, we continue to further divide the new interval. In this process, the test accuracy is the main selection criteria.

Step 4: Optimize the threshold further based on the obtained best slope: Centered on the rough threshold, several thresholds could be used to search the best threshold with the highest accuracy.

As an empirical fact, the learning rate and momentum of DBNs with Isigmoid function should be set as smaller values for better stability.

C. Handwritten Digits Recognition

The MNIST database, a special database of handwritten digits with general applicability and widespread use, is applied in this study to certify the feasibility of the parameter presetting method and the superiority of Isigmoid units.

In this study, 1500 samples are selected to determine the parameters. According to the distribution of the conventional DBN, the threshold a can be preset roughly as 4, and the minimal slope $\alpha_{\min}$ can be calculated by the following:

$$\alpha_{\min} = e^{-a}/\left(1 + e^{-a}\right)^2 = 0.0177. \tag{24}$$

The optimal slope can be found by the proposed searching method, and the results are listed in Fig. 8. It can be easily known from Fig. 8 that the Isigmoid can obtain the highest classification accuracy when $\alpha = 0.05$. Then several thresholds around the rough threshold are chosen to search the best threshold further and the results are shown in Fig. 9. According to Fig. 9, $a = 4$ is the best choice. Therefore, $a = 4$ and $\alpha = 0.05$ are the optimal parameters of Isigmoid in handwritten digits recognition experiment.

To demonstrate the superiority of the proposed activation function, the DBN based on the optimized Isigmoid is compared with the conventional DBN with Sigmoid under the same architecture and relevant parameters. In this process, a total of 6000 samples will be used to train the DBNs, and the average test accuracies of 10 repeated runs are listed in Table I and the training accuracy curves of backpropagation process are shown in Fig. 10.

We can easily see from Table I that the DBN based on the Isigmoid function has a higher classification rate than the DBN based on the conventional Sigmoid function. Moreover, Fig. 10 shows that the DBN with Isigmoid converges more rapidly to a higher accuracy by the same training samples. The comparative results show that Isigmoid can speed up the learning for the whole DBN, especially it can enhance the learning for the lower layers so as to improve the identification accuracy. Thus, Isigmoid can reduce the vanishing gradient problem.

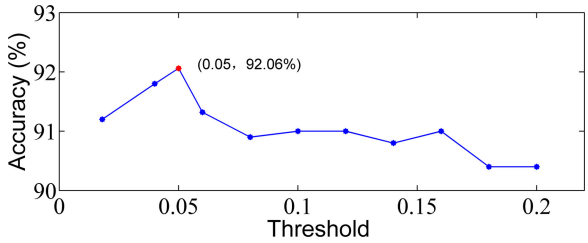

Fig. 8. Classification accuracies for Isigmoid with different slopes.

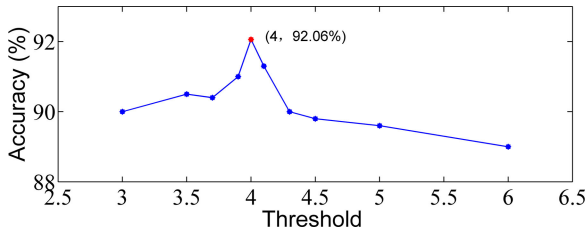

Fig. 9. Classification accuracies for Isigmoid with different thresholds.

TABLE I
CLASSIFICATION RATES FOR SIGMOID AND ISIGMOID FUNCTIONS

Function	Sigmoid	Isigmoid
Accuracy	93.73%	95.56%

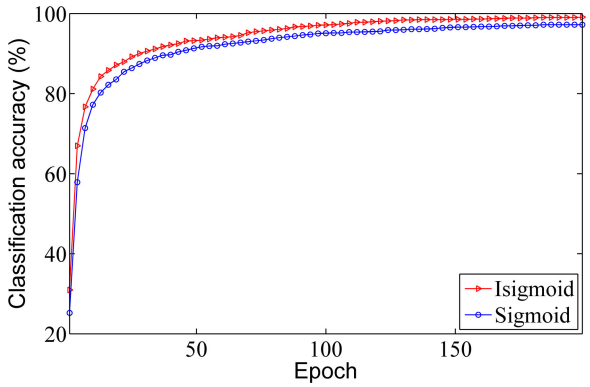

Fig. 10. Training classification accuracy curves for DBNs with Isigmoid and Sigmoid.

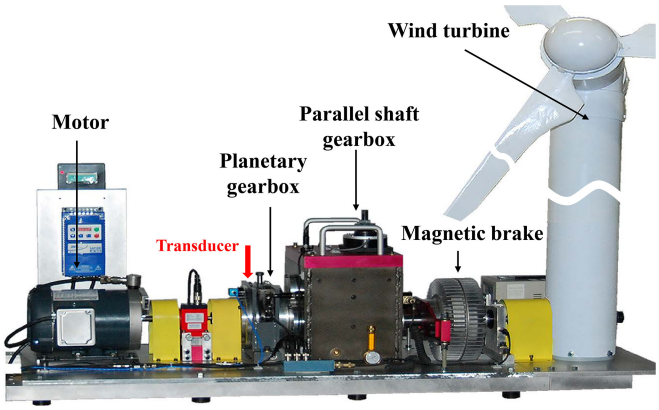

Fig. 11. Wind turbine drivetrain diagnostics simulator.

IV. Data Acquisition and Feature Extraction

A. Experimental Setup

As shown in Fig. 11, the wind turbine drivetrain diagnostics simulator (WTDS) designed by SpectraQuest Inc.[1] is applied to acquire data for planetary gearbox fault diagnosis. The test rig mainly consists of a driving motor, a two-stage planetary gearbox, a two-stage parallel-axis gearbox, a programmable magnetic brake, and a wind turbine. Via controlling the torque of the brake, different load conditions can be simulated. In this study, we focus on the secondary sun gear since it is more failure-prone, and four most common gear faults including surface wear, crack tooth, chipped tooth, and missing tooth are discussed. Meanwhile, a normal gear is used for comparison.

The vibration signals are collected under four different load conditions (0, 1.4, 2.8, and 25.2 N·m, respectively), so that more diverse fault information can be obtained. The acceleration transducer LC0103T produced by Lance Technologies Inc. is fixed on the box of the planetary gearbox (marked with a red arrow in Fig. 11) to collect vibration signals.

The time-domain waveforms, frequency spectra, and envelope spectra of the original signals under the load of 1.4 N·m are shown in Fig. 12.

[1]The company website has more detailed information on WTDS. [Online]. Available: http://www.pinxuntech.com/en/index.htm

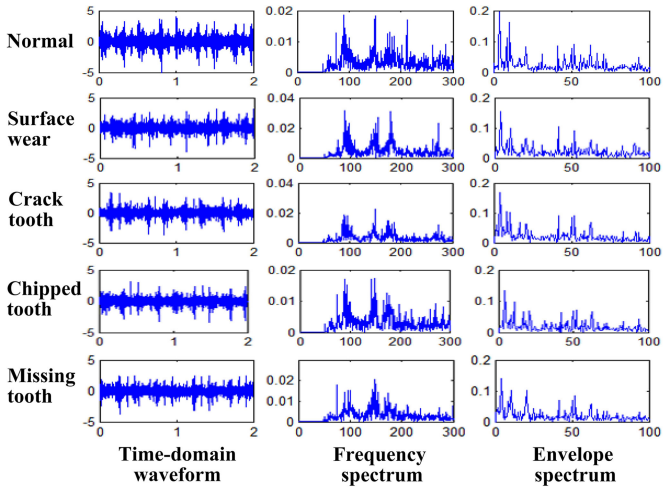

Fig. 12. Time-domain waveforms, frequency spectra, and envelope spectra of the original signals under the load of 1.4 N·m.

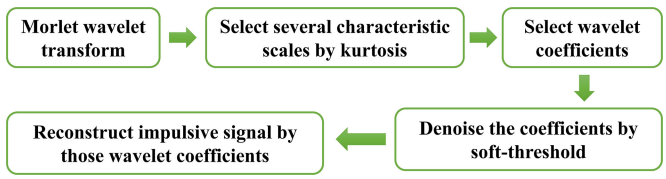

Fig. 13. Flow diagram of the proposed impulsive signals extraction approach.

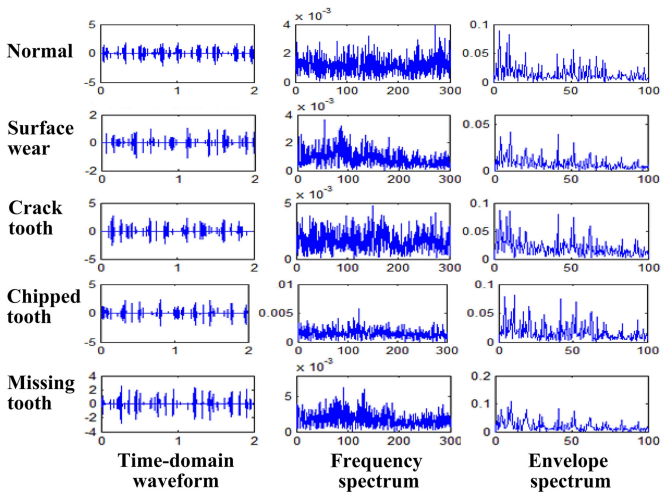

Fig. 14. Time-domain waveforms, frequency spectra, and envelope spectra of the impulsive signals under the load of 1.4 N·m.

B. Impulsive Signal Extraction

In particular, vibration transients excited by faults at specific rates contain more information of early faults. Consequently, an integrated approach proposed in [22] is employed to extract the impulsive signals, whose features will be input into the DBNs.

The used impulsive signals extraction method can be seen in Fig. 13. First, the Morlet wavelet transform is performed on the original signal; then, the kurtosis index is applied to select impulsive components (wavelet coefficients) at characteristic

TABLE II
STATISTICAL FEATURES

Name	Formula	Name	Formula	Name	Formula
Peak-to-peak (Pk-pk)	$t_1=\mathrm{MAX}\ \lvert x(i)\rvert-\mathrm{MIN}\ \lvert x(i)\rvert$	Variance (Var)	$t_8=\frac{1}{N}\sum_i^N(x(i)-\bar{x})^2$	Impulsive Factor (IF)	$t_{15}=\mathrm{MAX}\ \lvert x(i)\rvert\Big/\frac{1}{N}\sum_i^N\lvert x(i)\rvert$
Peak	$t_2=\mathrm{MAX}\ \lvert x(i)\rvert$	Standard Deviation (SD)	$t_9=\sqrt{\frac{1}{N}\sum_i^N(x(i)-\bar{x})^2}$	Clearance Factor (CF)	$t_{16}=\mathrm{MAX}\ \lvert x(i)\rvert\Big/\left(\frac{1}{N}\sum_i^N\lvert x(i)\rvert\right)^{1/2}$
Mean	$t_3=\frac{1}{N}\sum_i^N x(i)$	Skewness (Ske)	$t_{10}=\frac{1}{N}\sum_i^N x(i)^3$	Waveform Factor (WF)	$t_{17}=\sqrt{\frac{1}{N}\sum_i^N x(i)^2}\Big/\frac{1}{N}\sum_i^N\lvert x(i)\rvert$
Mean Square (MS)	$t_4=\frac{1}{N}\sum_i^N x(i)^2$	Kurtosis (Kur)	$t_{11}=\frac{1}{N}\sum_i^N\lvert x(i)\rvert^4$	Mean Frequency (MF)	$f_1=\frac{1}{N}\sum_j^N X(j)$
Root Mean Square (RMS)	$t_5=\sqrt{\frac{1}{N}\sum_i^N x(i)^2}$	Skewness Factor (SF)	$t_{12}=\frac{1}{N}\sum_i^N\lvert x(i)\rvert^3\Big/\left(\sqrt{\frac{1}{N}\sum_i^N x(i)^2}\right)^3$	Frequency Center (FC)	$f_2=\sum_j^N(f(j)\times X(j))\Big/\sum_j^N X(j)$
Mean Amplitude (MA)	$t_6=\frac{1}{N}\sum_i^N\lvert x(i)\rvert$	Kurtosis Factor (KF)	$t_{13}=\frac{1}{N}\sum_i^N\lvert x(i)\rvert^4\Big/\left(\sqrt{\frac{1}{N}\sum_i^N x(i)^2}\right)^4$	RMS Frequency (RMSF)	$f_3=\sqrt{\sum_j^N(f(j)^2X(j))\Big/\sum_j^N X(j)}$
Square Mean Root (SMR)	$t_7=\left(\frac{1}{N}\sum_i^N\lvert x(i)\rvert^{1/2}\right)^2$	Peak Factor (PF)	$t_{14}=\mathrm{MAX}\ \lvert x(i)\rvert\Big/\sqrt{\frac{1}{N}\sum_i^N x(i)^2}$	Standard Deviation Frequency (SDF)	$f_4=\sqrt{\sum_j^N((f(j)-f_c)^2X(j))\Big/\sum_j^N X(j)}$

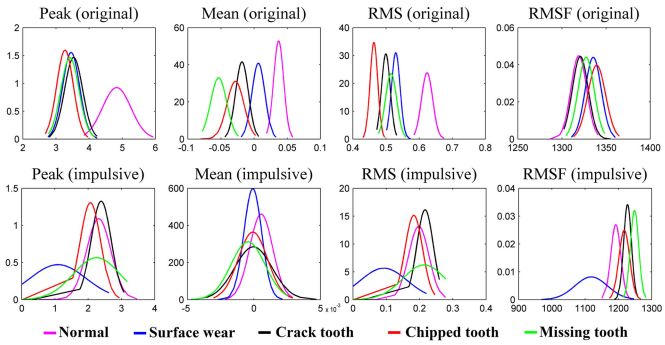

Fig. 15. Probability density functions of original and impulsive features under the load of 1.4 N·m.

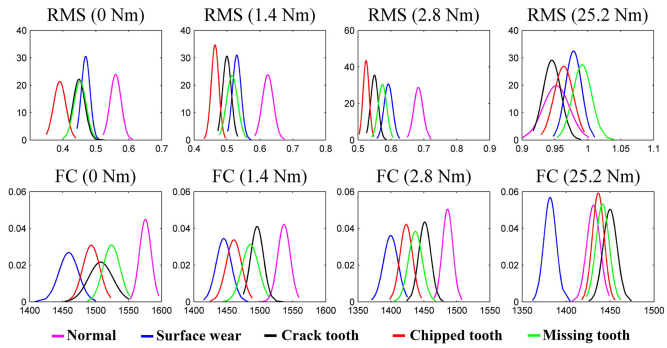

Fig. 16. Probability density functions of original RMSs and FCs under four different loads.

scales; finally, an adaptive soft-thresholding method is used to denoise the characteristic wavelet coefficients, and the impulsive signal is reconstructed by the denoised coefficients.

By the use of this method, the time-domain waveforms, frequency spectra, and envelope spectra of the impulsive signals extracted from the original signals under the load of 1.4 N·m are illustrated in Fig. 14.

Actually, ensemble empirical mode decomposition (EEMD) is usually used to extract impulsive components from original signals [8]. Therefore, for choosing better impulsive signals, EEMD will be compared with the proposed method.

C. Feature Extraction and Distribution Analysis:

Statistical features, such as mean and peak, reflect the fault information quantitatively, so in this study these features are calculated to train DBNs. These features are listed in Table II. As shown in Table II, 25 features will be calculated from original signals, where t_1–t_{17} are time-domain features, and f_1–f_4 are frequency-domain features that are extracted from frequency spectra and envelope spectra, respectively. Similarly, the same features will be extracted from the impulsive signals extracted by the proposed approach. Therefore, 50 features can be obtained in total.

Probability density functions of several typical features under the load of 1.4 N·m are illustrated in Fig. 15 to analyze and compare intuitively the feature sensitivities to different faults. It is obvious that the feature distributions of different faults are different, and these differences will be learned by DBNs to identify the faults. If a feature has less overlapping area among the curves, such as the mean of original signals, it has higher between-class scatter and lower within-class scatter, which means it is more sensitive to faults, and vice versa [32], [33]. It can be seen from Fig. 15 that the features such as the mean of original signals and the peak of impulsive signals extracted by the proposed method have better classification capacities, while the other features such as the peak of the original signal and the mean of the impulsive signal are less sensitive.

Comparing original and impulsive features comprehensively, we can know that some impulsive features such as the RMSF (rms frequency) and the peak have stronger classification capability than original ones. For example, the RMSF and rms of impulsive signal are very sensitive to surface wear failure, while the original features of surface wear failure are always severely mixed with those of crack tooth and tooth missing. Therefore, impulsive features can play a supplementary role in fault diagnosis, which will be further verified in the next section.

In order to analyze the influence of load conditions on feature distributions, the RMS and FC (frequency center) of original signals under different loads are listed together in Fig. 16. As is shown, rms values under loads of 0, 1.4, and 2.8 N·m are

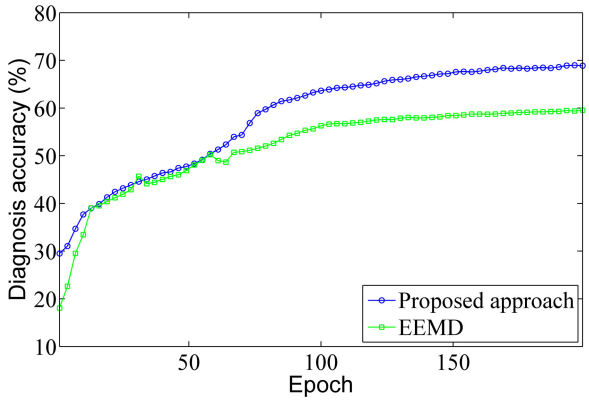

Fig. 17. Training diagnosis accuracy curves of DBNs trained with impulsive features obtained by the proposed approach and EEMD.

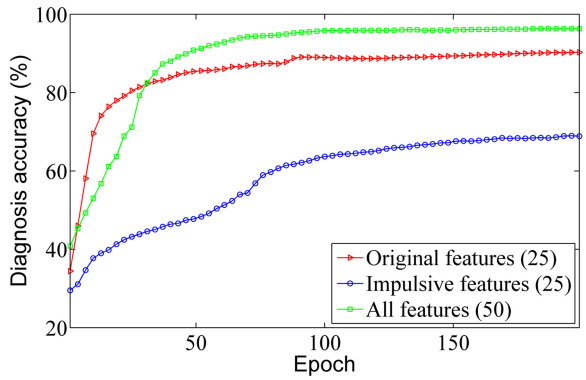

Fig. 18. Training diagnosis accuracy curves of DBNs trained with different sample sets.

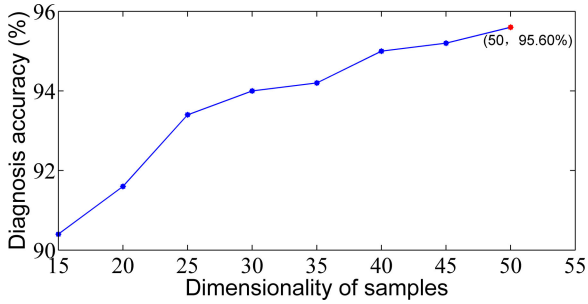

Fig. 19. Diagnosis accuracies of DBNs trained with samples of different dimensionalities.

more sensitive than that under load of 25.2 N·m. It then follows that under heavy load, the vibration excited by normal rotation became larger and will mask some of the characteristic components caused by gear faults, then it may seriously affect the feature distribution and worsen their sensitivities.

D. Sample Construction

For comparison, the impulsive features obtained by the proposed approach and EEMD are independently taken to train the DBNs with the same architectures and parameters, and the training accuracy curves, as shown in Fig. 17. It can be clearly noted from this figure that the impulsive features obtained by the proposed method have better performance than those extracted by EEMD. In consequence, the impulsive signals extracted by the proposed method will be employed in the following fault diagnosis experiments rather than those by EEMD.

For the construction of samples, if the computers used to train the DBNs have powerful computing capacity and the training is not time-consuming, the basic concern for sample construction would be how to maximize the diagnosis accuracy by using those features, not taking computing efficiency into consideration.

Original features and impulsive features extracted by the proposed method are respectively taken to train and test the DBNs, besides they are grouped together for comparison. The training accuracy curves are shown in Fig. 18, and it is obvious that the 50-dimension samples consisting of both original and impulsive features reach to better accuracy than 25-dimension samples with original features and 25-dimension samples with impulsive features. Therefore, impulsive features are more sensitive to some faults and can improve the diagnosis performance, which coincides with the conclusion of feature distribution analysis.

It then follows that impulsive features are applied to construct samples. Besides, to avoid the adverse effect caused by less sensitive features, feature selection should be focused on. With the index proposed by Xie and Beni in [34], between-class scatter and within-class scatter are used to express quantifiably the sensitivities of features, and the index can be calculated by the following:

$$XB(c) = \frac{\sum_{k=1}^{c}\sum_{i=1}^{n}\mu_{ik}^{2}||x_i - v_k||}{n\min(||v_i - v_j||)} \tag{25}$$

where x_i is a sample for certain feature set $\mathbf{X} = \{x_1, x_2, \ldots, x_n\}$ of $c = 5$ kinds of faults. v_k, v_i, and v_j represent the centers of the features of fault k, i, and j, respectively. Note that $||\bullet||$ is the usual Euclidean norm, andμ_{ik} is the fuzzy membership and could be crudely set as $\mu_{ik} = 1$ in this study for simplicity. The numerator of (25) represents the compactness of each cluster and the denominator indicates the separation between clusters. So, $XB(c)$ could be used to evaluate the sensitivities of features, and the smaller the $XB(c)$ is, the more sensitive the feature is.

Then, 50 original and impulsive features are used to construct samples with various dimensionalities, such as 45-dimension, 40-dimension, and so on (taking 45-dimension samples, for example, the most sensitive features are kept and the least sensitive five are abandoned). In total, eight kinds of samples are obtained for comparison and their diagnosis accuracies are shown intuitively in Fig. 19. It is obvious that the 50-dimension samples consisting of all original and impulsive features are superior to low-dimensional samples. Besides, the training processes of the DBNs using 50-dimension samples take only 98 s on average, which is acceptable. Therefore the 50-dimension samples are the best choice not only in diagnosis accuracy but also in computing time.

TABLE III
THRESHOLDS AND SLOPES OF ISIGMOID FOR VARIOUS LOAD CONDITIONS

	Load condition (Nm)				
	0	1.4	2.8	25.2	Mixed
Threshold (a)	5	6	3.5	4.5	5
Slope (α)	0.15	0.15	0.15	0.20	0.16

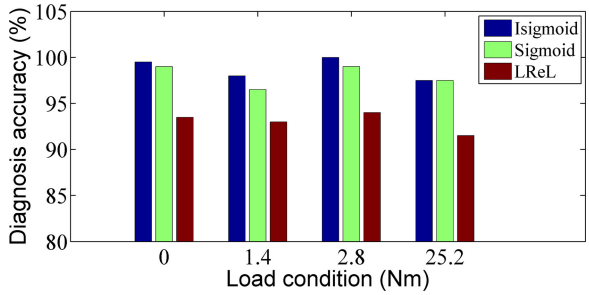

Fig. 20. Diagnosis accuracies for three functions with subsets under different loads.

In total, 4000 50-dimension samples could be obtained, in which subsets under each load (0, 1.4, 2.8, and 25.2 N·m) contain 1000 samples, respectively.

V. FAULT DIAGNOSIS

In this section, some samples are first applied to preset parameters of Isigmoid units. With the method proposed in Section III, the optimal parameters for the subsets under various loads can be seen in Table III, and all the slopes and thresholds satisfy (23).

Then, to compare the performances of three different functions under four different load conditions (0, 1.4, 2.8, and 25.2 N·m), training subsets that contain 800 samples (the rest 200 for testing) collected under each load will be used to train the DBNs independently. In the end, the subsets under each load will be chaotically combined into a mixed training set to train the DBNs for further comparison.

A. Experimental Results Under Different Single Loads:

By the use of different transfer functions in the backpropagation process, the diagnosis accuracies of DBNs trained by the samples under different loads are compared, and the average accuracies of ten repeated runs obtained by Isigmoid, Sigmoid, and LReL are shown in Fig. 20. As shown in Fig. 20, all the diagnosis results obtained by Isigmoid are better than those obtained by conventional Sigmoid and LReL. It implies that the Isigmoid function really play a positive role in vanishing gradient problem. Furthermore, for three functions, the DBNs trained by the sample subset under the load of 25.2 N·m perform worse than the ones trained with the sample subsets under the loads of 0, 1.4, and 2.8 N·m. This is mainly due to the fact that the vibration excited by normal rotation under heavy load becomes larger and masks some characteristic components caused by gear faults. These experimental results coincide well with the analysis results with respect to the influence of load conditions on feature distributions.

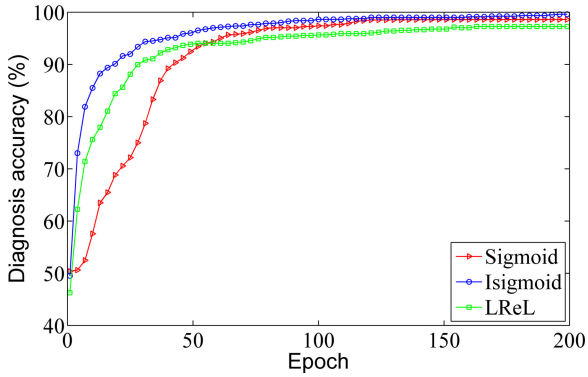

Fig. 21. Training diagnosis rate curves of different functions trained with the training set under the load of 2.8 N·m.

TABLE IV
DIAGNOSIS ACCURACIES FOR THREE FUNCTIONS WITH THE MIXED SET

	Isigmoid	Sigmoid	LReL
Train accuracy	97.49%	97.03%	94.60%
Test accuracy	96.32%	95.60%	92.55%
Difference	1.17%	1.43%	2.05%

Then, the sample subsets under the load of 2.8 N·m are applied to further study the performances of three different functions during the training process. The diagnosis accuracy curves of Isigmoid, Sigmoid, and LReL functions trained by the samples under loads of 2.8 N·m are shown in Fig. 21. It can be seen from Fig. 21 that all the accuracy curves are composed of two stages: The first one is a rapid learning stage with a large slope, and the second one is a slow learning stage with an almost constant training accuracy. Isigmoid and LReL have almost the same slopes that are greater than Sigmoid at the first stage, and they move into the second stage at the epoch of around 50, which is earlier than Sigmoid. Besides, the Isigmoid function converges to a final training accuracy of 99.62%, which is better than the Sigmoid and LReL function.

From the above analysis, we can conclude that the proposed Isigmoid function has inherited the speed advantage of LReL function, thus Isigmoid has faster convergence rate than Sigmoid. In addition, Isigmoid can solve the vanishing gradient problem partly, therefore Isigmoid have higher diagnosis accuracies than Sigmoid. It is also worth noting that the training and test accuracies of LReL function is worse than the other two functions, this is mainly because LReL cannot utilize most of the effects from RBM pretraining.

B. Experimental Results by the Mixed Set

To validate the performance of the proposed DBNs under mixed loads, the whole set consisting of 4000 samples will be used to train and test the DBNs with different activation functions, where 3500 samples are used for training and the rest 500 samples are used for testing. The accuracies obtained by three activation functions are listed in Table IV, and the

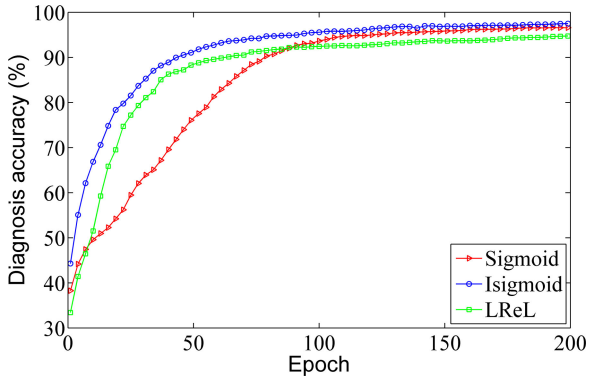

Fig. 22. Training diagnosis rate curves of different functions trained with the mixed training set.

corresponding training diagnosis accuracy curves are shown in Fig. 22.

As listed in Table IV, the training and test accuracies of DBNs with Isigmoid units are 97.49% and 96.32%, respectively, which are higher than those obtained by Sigmoid and LReL units. And the difference between the training and the test accuracy of Isigmoid is 1.17%, which indicates that Isigmoid has stronger generalization capability than Sigmoid (1.43%) and LReL (2.05%). Furthermore, Fig. 22 shows that the DBNs with Isigmoid units converge to a higher diagnosis accuracy with a faster speed than the DBNs with Sigmoid and LReL units.

It can be known from the experiments that the improvement of Isigmoid function successfully solves the vanishing gradient problem to some extent, and the Isigmoid function is superior to the conventional Sigmoid function and LReL function. Consequently, the DBN fault diagnosis approach based on the Isigmoid function is effective and advanced.

VI. Conclusion

This paper presented a novel intelligent fault diagnosis approach based on optimized DBNs for planetary gearboxes of wind turbines. Vanishing gradient problem is an inherent shortcoming of conventional logistic Sigmoid function, therefore the Isigmoid function was proposed based on LReL and Sigmoid, and the mathematical derivations demonstrated its feasibility theoretically. Then, the handwritten digit recognition experiment further validated that the improvement is successful. Owing to the strong sensitivities of impulsive features to early faults, both the original features and impulsive features were extracted to train DBNs, so as to improve the fault classification accuracy. In the end, the experimental results proved again that the proposed Isigmoid function has superiority over the conventional Sigmoid function and the LReL function on convergence speed and accuracy.

The two main contributions of this paper are as follows: 1) A creative transfer function Isigmoid is constructed to substitute the conventional Sigmoid for overcoming the vanishing gradient problem occurring in the DBN backpropagation process; and 2) by combining the optimized DBNs with Isigmoid units with the impulsive feature extraction, an integrated, accurate, and intelligent fault diagnosis approach is proposed. Besides, the fault diagnosis accuracy is up to 96.32%, thus the proposed fault diagnosis approach for planetary gearboxes of wind turbines is feasible in practical engineering.

In the coming research, the parameter optimization of Isigmoid can be performed on each layer respectively to further improve the DBNs.

References

[1] B. Lu, Y. Li, X. Wu, and Z. Yang, "A review of recent advances in wind turbine condition monitoring and fault diagnosis," in *Proc. IEEE Power Electron. Mach. Wind Appl.*, 2009, pp. 1–7.

[2] J. Yoon, D. He, and B. V. Hecke, "On the use of a single piezoelectric strain sensor for wind turbine planetary gearbox fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6585–6593, Oct. 2015.

[3] Y. Qin, "A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2716–2726, Mar. 2018.

[4] P. D. Samuel and D. J. Pines, "A review of vibration-based techniques for helicopter transmission diagnostics," *J. Sound Vibration*, vol. 282, nos. 1/2, pp. 475–508, Apr. 2005.

[5] Z. Feng, M. J. Zou, and F. L. Chu, "Application of regularization dimension to gear damage assessment," *Mech. Syst. Signal Process.*, vol. 24, no. 4, pp. 1081–1098, May 2010.

[6] L. Wang, Y. Shao, and Z. Cao, "Optimal demodulation subband selection for sun gear crack fault diagnosis in planetary gearbox," *Measurement*, vol. 125, pp. 554–563, Sep. 2018.

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

[8] Z. Feng *et al.*, "Fault diagnosis for wind turbine planetary gearboxes via demodulation analysis based on ensemble empirical mode decomposition and energy separation," *Renewable Energy.*, vol. 47, pp. 112–126, 2012.

[9] Y. Qin, Y. Mao, and B. Tang, "Multicomponent decomposition by wavelet modulus maxima and synchronous detection," *Mech. Syst. Signal Process.*, vol. 91, pp. 57–80, Jul. 2017.

[10] J. M. Ha *et al.*, "Autocorrelation-based time synchronous averaging for condition monitoring of planetary gearboxes in wind turbines," *Mech. Syst. Signal Process.*, vol. 70/71, pp. 161–175, Mar. 2016.

[11] J. Yu, M. Bai, G. Wang, and X. Shi, "Fault diagnosis of planetary gearbox with incomplete information using assignment reduction and flexible naive Bayesian classifier," *J. Mech. Sci. Technol.*, vol. 32, no. 1, pp. 37–47, Jan. 2018.

[12] M. A. Kramer and J. A. Leonard, "Diagnosis using back-propagation neural networks—Analysis and criticism," *Comput. Chem. Eng.*, vol. 14, no. 12, pp. 1323–1338, Dec. 1990.

[13] J. Qu, Z. Liu, M. J. Zuo, and H.-Z. Huang, "Feature selection for damage degree classification of planetary gearboxes using support vector machine," *Proc. Inst. Mech. Eng. Part C, Mech. Eng. Sci.*, vol. 225, pp. 2250–2264, Jun. 2011.

[14] J. Yin and W. Zhao, "Fault diagnosis network design for vehicle on-board equipments of high-speed railway: A deep learning approach," *Eng. Appl. Artif. Intell.*, vol. 56, pp. 250–259, Nov. 2016.

[15] P. Tamilselvan, P. Wang, and B. D. Youn, "Multi-sensor health diagnosis using deep belief network based state classification," in *Proc. ASME Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, 2011, pp. 749–758.

[16] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.

[17] V. T. Tran, F. Althobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief net-works," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4113–4122, Jul. 2014.

[18] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1–6.

[19] J. Schmidhuber, "Learning complex, extended sequences using the principle of history compression," *Neural Comput.*, vol. 4, no. 2, pp. 234–242, Mar. 1992.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] R. H. Hahnloser *et al.*, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit," *Nature*, vol. 405, no. 6789, pp. 947–951, Jun. 2010.
[22] Y. Qin, J. Xing, and Y. Mao, "Weak transient fault feature extraction based on an optimized Morlet wavelet and kurtosis," *Meas. Sci. Technol.*, vol. 27, no. 8, Aug. 2016, Art. no. 085003.
[23] R. R. Salakhutdinov and I. Murray, "On the quantitative analysis of deep belief networks," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 872–879.
[24] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. UTML TR 2010-003, 2010.
[25] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.
[26] X. S. Liu, B. W. Li, and X. Y. Yang, "Engine components fault diagnosis using an improved method of deep belief networks," in *Proc. IEEE Int. Conf. Mech. Aerosp. Eng.*, London, U.K., 2016, pp. 454–459.
[27] Y. Li, Y. Fu, H. Li, and S. W. Zhang, "The improved training algorithm of back propagation neural network with self-adaptive learning rate," in *Proc. IEEE Int. Conf. Comput. Intell. Natural Comput.*, 2009, vol. 1, pp. 73–76.
[28] B. K. Humpert, "Improving back propagation with a new error function," *Neural Netw.*, vol. 7, no. 8, pp. 1191–1192, 1994.
[29] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach, Learn.*, Haifa, Israel, 2010, pp. 807–814.
[30] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[31] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *ICML Deep Learn.*, pp. 1–5, 2015.
[32] Y. Liu *et al.*, "Feature fusion using kernel joint approximate diagonalization of eigen-matrices for rolling bearing fault identification," *J. Sound Vibration*, vol. 385, pp. 389–401, Dec. 2016.
[33] H. Shao *et al.*, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Nov. 2015, Art. no. 115002.
[34] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.

Yi Qin (M'17) received the B.Eng. and Ph.D. degrees in mechanical engineering from Chongqing University, Chongqing, China, in 2004 and 2008, respectively.

Since January 2009, he has been with Chongqing University, where he is currently a Professor with the College of Mechanical Engineering. His current research interests include signal processing, fault prognosis, mechanical dynamics and smart structure.

Dr. Qin is a Member of SPIE.

Xin Wang received the B.Eng. degree in vehicle engineering in 2017 from Chongqing University, Chongqing, China, where he is currently working toward M.S. degree in automotive engineering with the School of Automotive Engineering, Chongqing University.

His research interests mainly include signal processing, mechanical fault diagnosis, and artificial intelligence.

Jingqiang Zou received the B.Eng. degree in mechanical engineering in 2016 from Chongqing University, Chongqing, China, where he is currently working toward M.S. degree in mechanical engineering with the College of Mechanical Engineering, Chongqing University.

His research interests mainly include signal processing and mechanical fault diagnosis.