Research on Fault Diagnosis of Highway Bi-LSTM Based on Attention Mechanism

Indexed by:
Web of Science Group

Xueyi Li[a], Kaiyu Su[a], Qiushi He[a], Xiangkai Wang[a], Zhijie Xie[a,*]

[a] College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin, 150040, China.

Highlights

- The AHBi-LSTM method proposed can simultaneously process the bearing raw vibration signals according to the positive and inverse time-domain sequences, which is more conducive to practical industrial applications.

- The Attention mechanism allows the network to pay attention to essential features in different time steps, improving the fault diagnosis accuracy for deep groove ball bearing.

- The AHBi-LSTM method introduces an adaptive gating mechanism to manage the information flow in the network. The method can effectively solve multi-layer networks that are difficult to train.

Abstract

Deep groove ball bearings are widely used in rotary machinery. Accurate for bearing faults diagnosis is essential for equipment maintenance. For common depth learning methods, the feature extraction of inverse time domain signal direction and the attention to key features are usually ignored. Based on the long short term memory(LSTM) network, this study proposes an attention-based highway bidirectional long short term memory (AHBi-LSTM) network for fault diagnosis based on the raw vibration signal. By increasing the Attention mechanism and Highway, the ability of the network to extract features is increased. The bidirectional LSTM network simultaneously extracts the raw vibration signal in positive and inverse time-domains to better extract the fault features. Six deep groove ball bearings with different health conditions were used to validate the AHBi-LSTM method in an experiment. The results showed that the accuracy of the proposed method for bearing fault diagnosis was over 98%, which was 8.66% higher than that of the LSTM model. The AHBi-LSTM model is also better than other relevant models for bearing fault diagnosis.

Keywords

Fault diagnosis, Bi-LSTM, Attention, Highway, Deep learning, Ball Bearing.

1. Introduction

Deep groove ball bearings are widely used in rotating machinery, and the fault diagnosis of bearings is critical to ensure high-performance transmission[8]. The fault of the transmission system will lead to the suspension of production and affect the whole production process[5]. Finding before failure occurs is an effective means to ensure the regular operation of equipment and avoid economic losses. Bearing fault diagnosis has always been regarded as a research hotspot in prognosis and health management[18,13].

There are mainly two kinds of fault diagnosis methods for bearings, one is model-based methods, such as physical model [28], Kalman filter [19], strong tracking estimator [31], radial basis function neural network [6], and so on. The other is data-based methods, such as feature extraction[10], support vector machine[7], backpropagation neural network[35], and deep learning[1]. Model-based methods, such as Li et al.[12] proposed an approach based on frequency band entropy (FBE) to optimize the intrinsic mode function (IMF) of variational mode decomposition (VMD) with rich fault information. To find the best description of the fault signal, Bayesian optimization is used to infer the structure of the formal specification [11]. Model-based methods usually need prior knowledge as a research basis, and the cost of learning and application is high. Besides, some researchers diagnose bearing faults by fusing the features of different sensors [25]. Wang et al. [24] proposed combining multi-mode sensor signals to realize a more accurate

and reliable bearing fault diagnosis.

In recent years, more and more researchers have applied the deep learning method to the fault diagnosis of bearings. Compared with traditional signal feature extraction methods, which require a lot of prior knowledge, deep learning methods such as convolutional neural networks (CNN) can automatically extract features [17]. Huang *et al*. [9] proposed a multi-scale cascade convolutional neural network (MC-CNN) to enhance the input classification information. Tao *et al*. [21] proposed a fault diagnosis method using multi-vibration signals and deep belief networks(DBN). This method can adaptively fuse multi-feature data and identify various bearing faults. Wu *et al*. [27] put forward an adaptive architecture by integrating the idea of deep adaptation networks (DAN) with the simplified lightweight model, aiming at enhancing the generalization ability of the model. Mao *et al*. [14] proposed a semi-random subspace method with bidirectional gate recurrent units to use fusion features for bearing fault diagnosis. Deep learning methods based on CNNs and recurrent neural networks (RNN) and semi-supervised methods have been widely used in bearing fault diagnosis [30]. Although these methods have made particular progress in bearing fault diagnosis, they can only extract part of the feature information. The inverse time-domain sequence features are not extracted. Besides, problems such as insufficient attention to critical features, and too many training layers are difficult to converge need to be solved.

To make full use of positive and inverse time-domain sequence features, this study employs bidirectional long short term memory networks (Bi-LSTM) to extract the signal features of deep groove ball bearings. Taking the raw vibration signal as input, Bi-LSTM can not only use the positive time-domain features of the signals but also make full use of the inverse time-domain features.

Inspired by the long short term memory networks (LSTM), Srivastava proposed Highway Networks to solve multi-layer networks difficult to training [20]. The Highway introduces an adaptive gating mechanism to manage the information flow when training multi-layer networks in deep learning. Zilly et al.[34] proposed recurrent Highway networks, which extends the LSTM architecture. Zia et al. put forward the residual recurrent Highway network and the hierarchical recurrent Highway network, which included the network structure, alleviating the gradient disappearance problem [32,33]. Recurrent Highway Networks has also been well applied in speech synthesis and machine translation [26,16]. All these results show that the deep neural network combined with the Highway network can obtain higher accuracy.

In 2014, the Google Mind team published a paper that made the Attention mechanism famous [15]. Then, in the published article by Xu et al. [29], the Attention mechanism was applied in the image caption. Since then, the Attention mechanism has been widely used in various deep learning tasks [23]. In 2017, Google proposed self-Attention mechanisms were used in machine translation to learn text representation [22]. Chen et al.[3] offered a network based on the Attention mechanism autoencoder framework to predict the remaining useful life (RUL) value. For the automatic drive, the Attention mechanism has also made significant progress in recent research work [4]. At present, the Attention mechanism is not widely used in the fault diagnosis of bearings. The Attention mechanism allows the LSTM network to pay more attention to different features at different time steps.

To make full use of the inverse time-domain sequence features, the attention-based Highway bidirectional long short term memory (AHBi-LSTM) method proposed in this paper uses the Bi-LSTM networks to extract the bearing signals. Taking raw vibration signals as input, the Bi-LSTM is used to extract the signal features of deep groove ball bearings. Simultaneously, the Highway network is used to optimize features, which alleviates the deep neural network is difficult to train. Besides, the Attention mechanism allows the Bi-LSTM network to pay attention to different features at different time steps. The highlight of this study are as follows:

1) The AHBi-LSTM method proposed can simultaneously process the bearing raw vibration signals according to the positive and inverse time-domain sequences.

2) The Attention mechanism allows the network to pay attention to essential features in different time steps, improving the fault diagnosis accuracy of deep groove ball bearing.

3) The AHBi-LSTM method introduces an adaptive gating mechanism to manage the information flow in the network. This method can effectively solve multi-layer networks difficult to training.

4) The method proposed directly uses the raw vibration signals of bearings to extract features without time-frequency conversion, which is more conducive to practical industrial application.

The rest of this paper is organized as follows. In Section 2, the AHBi-LSTM method was proposed in this paper. The theories of the Bi-LSTM, Highway network, and Attention mechanism were introduced. In Section 3, taking the deep groove ball bearing as an example, experiments are designed to validate the effectiveness of the model. In Section 4, the experimental results were analyzed and discussed, and the effectiveness and superiority were verified by comparing them with other advanced neural network methods. The conclusion and future research direction are drawn in Section 5.

2. The Method

In this study, a hybrid model of Highway Bi-LSTM based on Attention mechanism is proposed for fault diagnosis of deep groove ball bearings. The traditional LSTM method can only extract the signal in the positive time-domain direction. In the AHBi-LSTM model, Bi-LSTM simultaneously extracts the signal features in the positive and inverse time-domain directions. The Attention mechanism is used to enhance the attention to essential fault features. Finally, the Highway network is used to further optimize the features and improve the bearing fault diagnosis effect. The model was calculated recursively under Softmax to optimize the fault diagnosis result. The overall framework of the AHBi-LSTM method proposed is shown in Figure 1
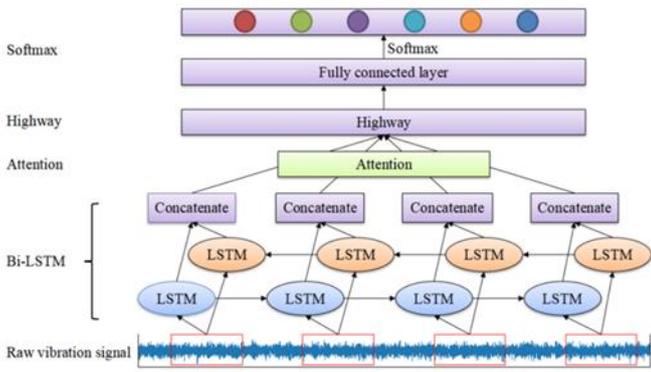
Figure 1. General frame diagram of the AHBi-LSTM method.

2.1. Bidirectional Long Short Term Memory Networks

The Bi-LSTM uses two LSTM networks simultaneously, one forward and one backward, and the two networks are connected to the same output layer. The raw bearing vibration signals were extracted by the Bi-LSTM network in both positive and negative directions. As shown in Figure 2, this structure is characterized by the fact that each node of the output layer can fully utilize the raw vibration signal information of the deep groove ball bearing. Based on this idea, the raw vibration signals are trained to diagnose the faults of bearings and judge the severity of the faults.
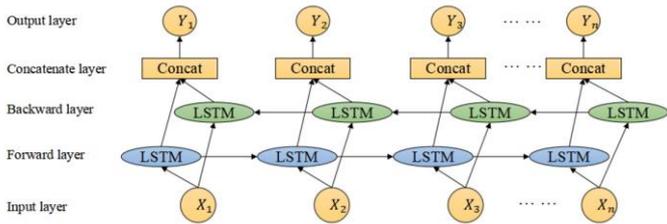

Figure 2. Schematic diagram of Bi-LSTM.

The LSTM is improved and optimized based on RNN, and a long span information transmission channel is added. The gating mechanism is introduced to keep the gradient value stable, which is convenient for training. First, LSTM receives the current moment input$x_t$and the previous moment state value h_{t-1}as the total input. After training, four states can be obtained, which can be expressed as follow:

$$Z_f = \sigma\big(W_f \times [h_{t-1}, x_t] + b_f\big) \tag{1}$$
$$Z_i = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{2}$$
$$Z = tanh(W \times [h_{t-1}, x_t] + b) \tag{3}$$
$$Z_o = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4}$$

where,Z_f,Z_i,$Z_o$are the total input vector multiplied by the weight matrixW_f,W_i,W_o, and then mapped to the value between 0 and 1 through the sigmoid activation function, as a gated state.Z is the total input vector multiplied by the weight matrix W and converted into a value between -1 and 1 by the tanh activation function as the new input data. The internal operation process of LSTM can be further expressed as follow:

$$c_t = Z_f \odot C_{t-1} + Z_i \odot Z \tag{5}$$
$$h_t = Z_o \odot tanh(c_t) \tag{6}$$
$$y_t = \sigma\big(W_y h_t\big) \tag{7}$$

where, $\odot$ represents the multiplication operation of corresponding matrix elements. In the LSTM cell,$Z_f$is the forget gate, which controls what information is forgotten; $Z_i$is the input gate, which controls what information is stored in the cell state;$Z_o$is the output gate, which controls the information that needs to be output at this moment, and the final output $y_t$can be obtained by changing h_t. The output $y_t$of Bi-LSTM at one moment is not only related to the hidden state $h_{t-1}$of the previous moment, but also related to the hidden state of the last moment. Bi-LSTM can achieve more accurate results than the LSTM [2].

2.2 Highway Unit

To obtain more accurate fault diagnosis result, the layer number of networks becomes deeper and deeper, resulting in network training becoming more difficult. The Highway is a learnable gate mechanism that divides the data input into two parts. One part needs to go through a nonlinear transformation, and the other part can be directly crossed through the layer without transformation. What data can pass through the network is determined by the weight matrix and input data. Under this mechanism, some information selectively passes through some layers, which reduces the number of training parameters. This is the reason why Highway can solve the difficult problem of deep network training. The main goal of Highway is to learn the proportion of original information that should be retained. It's like an information highway. The Highway is trained by the stochastic gradient descent (SGD). In a forward neural network with L layer, the input $x_i$can use nonlinear mapping transformation $H$with parameters $W_H$to generate output y_i, which can be expressed as follow:

$$y = H(x, W_H) \tag{8}$$

Two nonlinear mapping functions T and $\boldsymbol{C}$ are added to the Highway based on the above forward neural network, and the output y of the Highway is calculated as follow:

$$y = H(x, W_H) \times T(x, W_T) + x \times C(x, W_C) \tag{9}$$

where,$T$is the transform gate;C is the carry gate. To facilitate calculation and simplify the model, $\boldsymbol{C} = \boldsymbol{1} - \boldsymbol{T}$is defined, then the modified Highway can be expressed as follow:

$$y = H(x, W_H) \cdot T(x, W_T) + x \cdot \big(1 - T(x, W_{1-T})\big) \tag{10}$$

where, $\boldsymbol{y}$ is the final output of Highway. In (10), the dimensions of $\boldsymbol{x}$,$\boldsymbol{y}$,H,and Tmust be the same. The SGD algorithm is used to adjust network parameters as follow:

$$W_{t+1} \leftarrow W_t + \eta \frac{1}{N} \sum_{n=1}^{N} \theta(-y_n W_t x_n)(y_n x_n) \tag{11}$$

where, $W$is the weight parameter, η is the learning rate, N is the number of samples for one training, θ is the network parameter, $y_n$is the output, and $x_n$is the input.

2.3 Attention mechanism

The Attention mechanism in deep learning is similar to that in the human brain. Its core goal is to give more attention to more critical information in the general information. The advantages of the Attention mechanism are as follows: on the one hand, it helps to enhance the ability of the neural network to focus on features, that is, to select features that are more critical to target output; On the other hand, Attention mechanism can also be used as a resource allocation scheme, so that more computing resources can be obtained for more essential tasks.

The Attention mechanism allocates more attention to essential points. In the fault diagnosis of rotating machinery, different signal features have different influences. The Attention

mechanism is introduced to identify which information is essential. In this study, the Attention mechanism is added to the next step of the first Bi-LSTM layer, and the input of the Attention mechanism is the output of the Bi-LSTM layer, so:

$$u_i = tanh(W_A H_t + b_A) \tag{12}$$

$$\alpha_i = \frac{exp(u_i^T u_A)}{\sum_{j=1}^n u_j^T u_A} \tag{13}$$

$$v_i = \sum_t \alpha_i H_t \tag{14}$$

where, W_A is the weight matrix, b_A is the bias value, u_i is the hidden layer representation of H_t, and the weight a_i is the similarity of u_i with adjacent features u_A. u_A uses random initialization and dynamically updates during training. v is the output vector of the Attention mechanism.

The self-Attention used in this study calculated the correlation between each feature and all other features by using the Attention mechanism. The features associated with it had a high attention score. Attention scores can be used to get a weighted representation and then put them into a feedforward neural network to get a new representation, which considers the information of the features very well.

3. Bearing test experiments

To validate the effectiveness of the AHBi-LSTM method on groove bearing fault diagnosis and whether different degrees of fault can be effectively distinguished, an experiment of deep groove ball bearings was designed. The bearing fault diagnosis experiment rig is shown in Figure 3. The test rig comprises a 0.75kW three-phase asynchronous motor, motor controller, support device, shaft, and bearing seat. The maximum speed of the motor is 8000rpm. In this test, the speed was set at 1200rpm. The sampling frequency was 19.2kHz, and the sampling time was set at 110 seconds with every different health states. In the process of vibration signal acquisition, acceleration sensors are arranged in the vertical direction of the bearing seat to collect vibration signals.
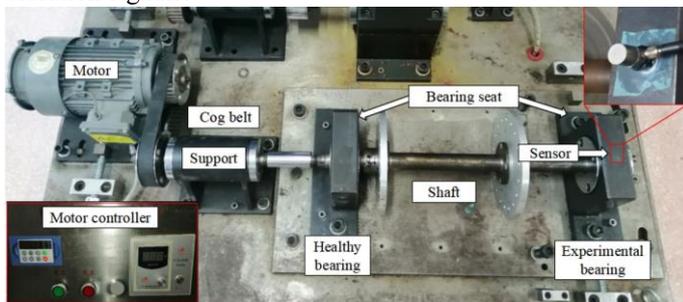

Figure 3. Deep groove ball bearing fault diagnosis experiment rig.

In this study, the SKF6205 deep groove ball bearing model was used to validate the AHBi-LSTM method. The specific parameters of SKF6205 are shown in Table 1. Local faults of bearings are produced by wire cutting. In this study, five types of deep groove ball bearings with different faults were selected for testing, among which two kinds of inner ring fault and three kinds of outer ring fault. The pictures of the normal bearing and bearings with local faults are shown in Figure 4.The groove depth of the inner ring faults is 2mm, and the groove width of inner ring faults is 2mm and 3mm, respectively. The groove depth of the outer ring faults is 1mm, and the groove width of the outer ring fault is 3mm, 2mm, and 0.5mm, respectively. The specific fault description of deep groove ball bearings is shown in Table 2.
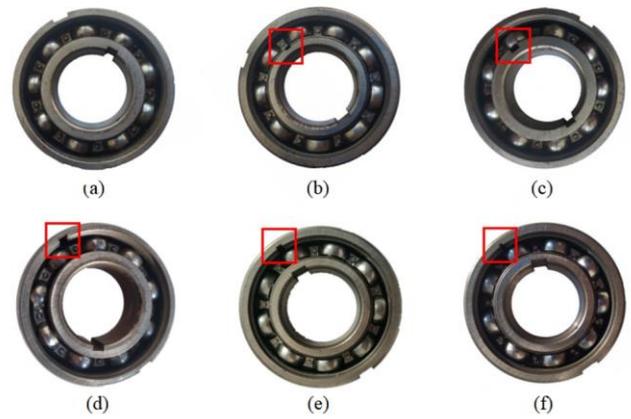

Figure 4. The picture of local faults of SKF6205 deep groove ball bearings. (a) Normal state bearing (NOR), (b) Inner ring fault 1 (IRF1), (c) Inner ring fault 2 (IRF2), (d) Outer ring fault 1 (ORF1), (e) Outer ring fault 2 (ORF2), (f) Outer ring fault 3 (ORF3).

Table 1. Parameters of deep groove ball bearing of SKF6205.

parameter	value
Outer ring diameter/mm	52
Inner ring diameter/mm	25
Pitch diameter/mm	39
Rolling body diameter/mm	7.94
Number of the rolling body	9

Table 2. Fault description of the deep groove ball bearing.

Bearing health states	Groove width (mm)	Groove depth (mm)	Category label
Normal (NOR)	-	-	1
Inner ring fault 1 (IRF1)	2	2	2
Inner ring fault 2 (IRF2)	3	2	3
Outer ring fault 1 (ORF1)	3	1	4
Outer ring fault 2 (ORF2)	2	1	5
Outer ring fault 3 (ORF3)	0.5	1	6

The raw vibration signals of deep groove ball bearings are shown intuitively in Figure 5. For deep groove ball bearings with six different health states, the raw vibration signals of bearings in health state 1 are close to those in health state 2, and the vibration signals in health state 3 are noticeable compared with the first two types. By comparing the raw vibration signals in the healthy states 2 and 3, it is shown that the narrower the groove width of the inner ring fault is, the more stable the raw vibration signals are. For healthy state 4, the impact of grooves and balls is more serious, and pronounced impact signals can be seen. The raw vibration signals in health states 5 and 6 can also see obvious impact characteristics. On the one hand, it shows that the larger the groove width of the outer ring fault is, the more unstable the bearing signal is. On the other hand, it also shows that the outer ring fault has a more significant influence on the raw vibration signal than the inner ring fault.
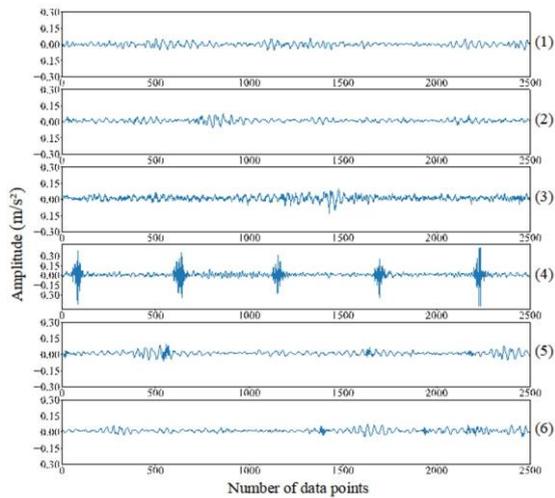
Figure 5. Raw vibration signals of deep groove ball bearings with different health states

By observing the spectrum chart of different bearing health states, the influence of faults on bearing vibration signals can be intuitively visualization from the frequency domain. The spectrum charts of the 6 different health states are shown in Figure 6. The abscissa is the normalized frequency. Normalized frequency means that the sampling frequency is set as 1, and the other frequencies are expressed as percentages. That is, after normalization, frequencies are converted to between [0,1]. Normalized frequency enables a uniform standard to be used to compare the distribution of various frequencies. Yellow means spectral energy is higher, blue means spectral energy is lower. It can be seen from the spectrum chart that the larger spectral values of normal deep groove ball bearing are more concentrated around 0.1. For the fault bearing, the energy is shifted to the high frequency. The yellow color on the left side in Figure 6(b) becomes lighter because the energy is shifted to the right, resulting in a lighter color. The more obvious the shock is, the more concentrated in the high-frequency value. For health state 4, the vibration signal of bearing with noticeable impact can be seen, and the normalized frequency value is more concentrated around 0.8.
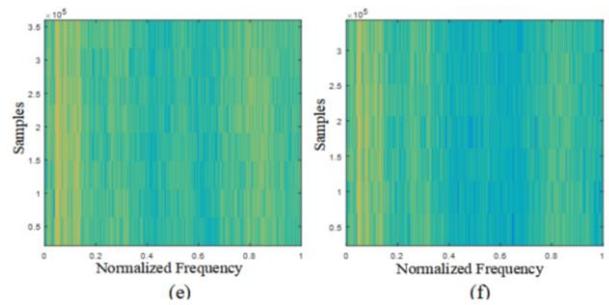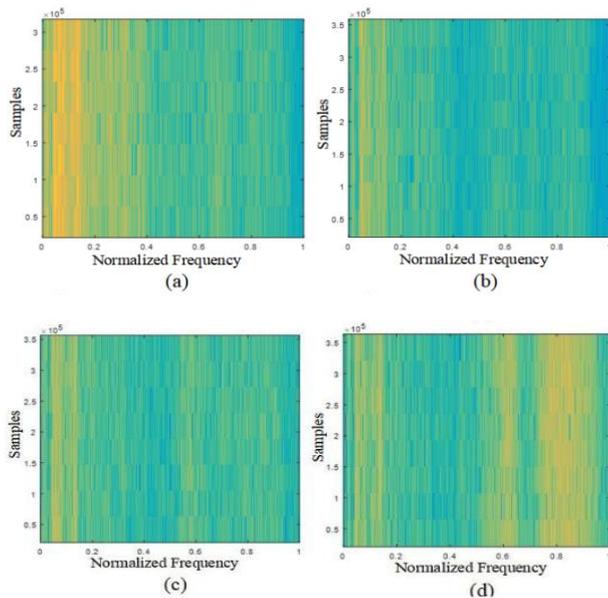

Figure 6. Spectrum chart of six different health conditions. (a) Normal state bearing (NOR), (b) Inner ring fault 1 (IRF1), (c) Inner ring fault 2 (IRF2), (d) Outer ring fault 1 (ORF1), (e) Outer ring fault 2 (ORF2), (f) Outer ring fault 3 (ORF3).

The AHBi-LSTM method superimposed two Bi-LSTM neural networks, with the cell size is 32 and 64, respectively. Dropout is used to process the output parameter behind the first Bi-LSTM layers to avoid overfitting. Meanwhile, the method uses the Attention mechanism to identify essential features. Due to the low complexity of the Highway, a multi-layer Highway is used to optimize the output feature of Bi-LSTM continuously. In this study, the layer number of the Highway network is three selected by the experiments. The dense and softmax layers are added behind Highway to diagnose the fault severity of deep groove ball bearings. The details of the network are shown in Table 3. A total of 100 thousand samples were used as the training set, and 10 thousand samples as the testing set. There is no reused data between the training set and the testing set. The test was based on an Asus desktop computer, equipped with an i7-10700 CPU and a memory frequency of 2133 MHz, and the NVIDIA graphics card was used to increase the calculation speed.

Table 3. Network details of the AHBi-LSTM method.

Layer	Output Shape	Parameters
Input	960, 1	0
Bidirectional LSTM	960, 64	8704
Batch Normalization	960, 64	256
Dropout	960, 64	0
Attention	960, 64	12288
Bidirectional LSTM	128	66048
Highway	128	33024
Highway	128	33024
Highway	128	33024
Batch Normalization	128	512
Dense	64	8256
Softmax	6	390

4. Discussions

The accuracy values in the training process were extracted, and the accuracy curves of the training set and the validation set were drawn. After smoothing, the curves were shown in Figure 7. The accuracy rate of the training set rose rapidly before the 20th epochs, rising to more than 95%, and then rose steadily and gradually stabilized. The accuracy rate of the verification set increased slowly at the beginning and gradually stabilized after

the 50th epochs. The AHBi-LSTM method has a good effect on the fault diagnosis of deep groove ball bearings.
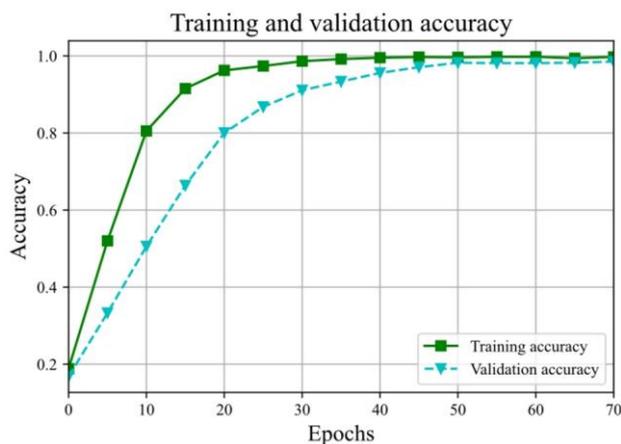

Figure 7. Training and validation accuracy curves.

The confusion matrix is drawn to see the fault diagnosis effect of the AHBi-LSTM method for deep groove ball bearings with different health states. As shown in Figure 8, the fault diagnosis accuracy of deep groove ball bearings in healthy state 4 reaches 100%, the accuracy of healthy state 3 exceeds 99%, and that of other healthy states is all above 97%. It shows that the AHBi-LSTM method proposed can effectively judge the bearing faults and distinguish different fault degrees.
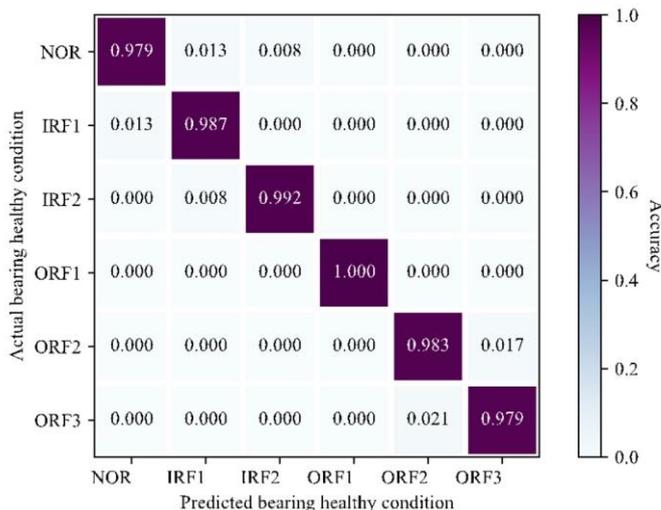

Figure 8. Confusion matrix for deep groove ball bearings with different health states.

To more clearly represent the predicted results of every type of deep groove ball bearings, 235 samples of each healthy state of bearing were selected for testing. The test results are shown in Figure 9. For bearing healthy state 1, three and two samples were misjudged as healthy states 2 and 3, respectively. Three samples in healthy state 2 were misjudged as healthy state 1. Two samples in healthy 3 were misjudged as healthy 2. The bearings in health state 4 were entirely judged correctly, which may be due to the apparent signal characteristics. Four samples and five samples of bearings with health status 5 and 6 were misjudged as each other, respectively. The error may be because the bearing raw vibration signals of the adjacent healthy states are close. The misjudgment rate of the AHBi-LSTM method is low, which indicates that the method is effective for the fault diagnosis and can distinguish different fault degrees of deep groove ball bearings.
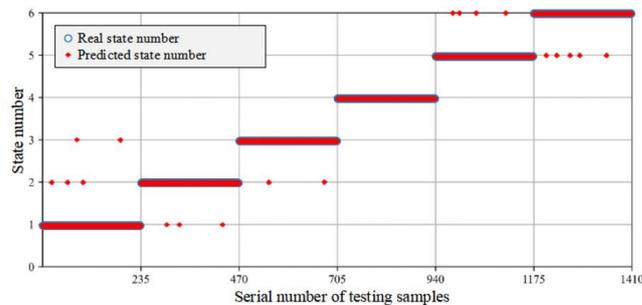

Figure 9. Fault diagnosis results of deep groove ball bearings with different health states.

To more clearly demonstrate the fault diagnosis results of the AHBi-LSTM method proposed in this study on deep groove ball bearings, 50 bearings signals in each health state were selected for clustering visualization shown in Figure 10. As shown from the figure, each health state bearings are relatively concentrated, and clear boundaries can be seen. Individual IRF2 samples are closer to IRF1, and this may be due to environmental noise or other factors that lead to sample points anomaly. Overall, the AHBi-LSTM method has an excellent fault diagnosis effect on the faults of deep groove ball bearings.
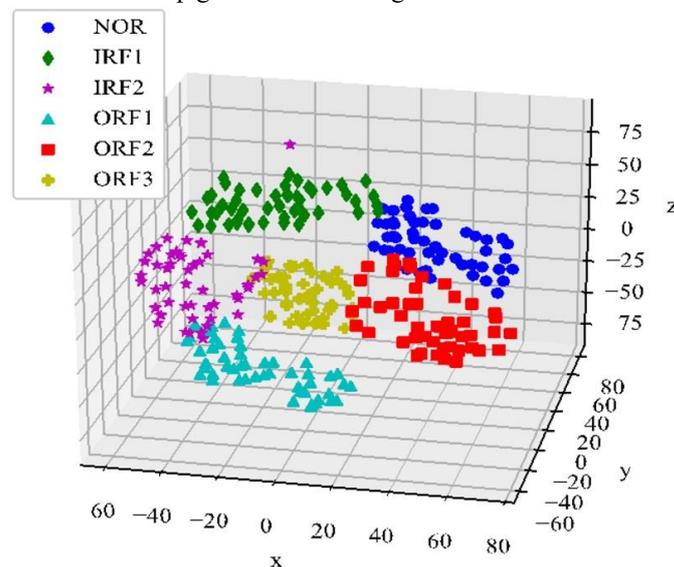

Figure 10. Clustering visualization of deep groove ball bearings with different health states.

To further validate the effectiveness of the AHBi-LSTM method proposed, the method was compared with other relevant deep learning methods. In addition to the LSTM networks and Bi-LSTM networks, the Attention mechanism and the Highway network on Bi-LSTM only compared. The experimental results are shown in Table 4.

Table 4. The accuracy of the AHBi-LSTM method and the other methods.

Methods	1st-fold	2nd-fold	3rd-fold	4th-fold	5th-fold	Average
LSTM	89.46%	89.32%	89.21%	89.72%	90.16%	89.57%
Bi-LSTM	91.73%	91.70%	91.98%	91.52%	91.43%	91.60%
Bi-LSTM+Highway	93.54%	93.16%	93.66%	93.75%	93.08%	93.44%
Bi-LSTM+Attention	95.30%	95.31%	95.38%	95.81%	95.27%	95.41%
AHBi-LSTM	98.65%	97.79%	98.28%	98.38%	98.03%	98.23%

The results show that both of them have certain improvements in the accuracy of the bearing fault diagnosis.

Compared with these approximate models, the AHBi-LSTM model achieves better fault diagnosis performance for deep groove ball bearings.

5. Conclusions

To sum up, the AHBi-LSTM method was proposed in this study to solve the feature of inverse time-domain signal direction was ignored in the fault diagnosis. The Attention mechanism is introduced to focus on effective information. The Highway can selectively process data to reduce computation. The experimental results show that the model is effective for fault diagnosis of deep groove ball bearing and judge the fault severity. Compared with other methods, the results show that the AHBi-LSTM method is more effective for bearing fault diagnosis. The author's future research focus on the interpretability of fault feature extraction.

References

1. Barcelos A S, Cardoso A J M. Current-based bearing fault diagnosis using deep learning algorithms[J]. Energies, 2021, 14(9): 2509, https://doi.org/10.3390/en14092509.
2. Chadha G S, Panambilly A, Schwung A, et al. Bidirectional deep recurrent neural networks for process fault classification[J]. ISA transactions, 2020, 106: 330-342, https://doi.org/10.1016/j.isatra.2020.07.011.
3. Chen Y, Peng G, Zhu Z, et al. A novel deep learning method based on attention mechanism for bearing remaining useful life prediction[J]. Applied Soft Computing, 2020, 86: 105919, https://doi.org/10.1016/ j.asoc.2019.105919.
4. Chikhaoui B, Ruer P, Vallières É F. A Deep Learning Method for Automatic Visual Attention Detection in Older Drivers[C]//International Conference on Smart Homes and Health Telematics. Springer, Cham, 2019: 49-60. https://doi.org/10.1007/978-3-030-32785-9_5
5. Duan Z, Wu T, Guo S, et al. Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: a review[J]. The International Journal of Advanced Manufacturing Technology, 2018, 96(1-4): 803-819, https://doi.org/10.1007/s00170-017-1474-8.
6. Fang S, Zijie W. Rolling bearing fault diagnosis based on wavelet packet and RBF neural network[C]//2007 Chinese Control Conference. IEEE, 2007: 451-455. https://doi.org/10.1109/CHICC.2006.4346979
7. Gryllias K C, Antoniadis I A. A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments[J]. Engineering Applications of Artificial Intelligence, 2012, 25(2): 326-344, https://doi.org/10.1016/j.engappai.2011.09.010.
8. He M, He D. Deep learning based approach for bearing fault diagnosis[J]. IEEE Transactions on Industry Applications, 2017, 53(3): 3057-3065, https://doi.org/10.1109/TIA.2017.2661250.
9. Huang W, Cheng J, Yang Y, et al. An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis[J]. Neurocomputing, 2019, 359: 77-92, https://doi.org/10.1016/ j.neucom.2019.05.052.
10. Kaplan K, Kaya Y, Kuncan M, et al. An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis[J]. Applied Soft Computing, 2020, 87: 106019, https://doi.org/10.1016/ j.asoc.2019.106019.
11. Kolar D, Lisjak D, Pająk M, et al. Intelligent fault diagnosis of rotary machinery by convolutional neural network with automatic hyper-parameters tuning using Bayesian optimization[J]. Sensors, 2021, 21(7): 2411, https://doi.org/10.3390/s21072411.
12. Li H, Liu T, Wu X, et al. An optimized VMD method and its applications in bearing fault diagnosis[J]. Measurement, 2020, 166: 108185, https://doi.org/10.1016/j.measurement.2020.108185.
13. Li X, Li J, Zhao C, et al. Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1D separable convolution with residual connection[J]. Mechanical Systems and Signal Processing, 2020, 142: 106740, https://doi.org/10.1016/j.ymssp.2020.106740.
14. Mao X T, Zhang F, Wang G, et al. Semi-Random Subspace with Bi-GRU: Fusing Statistical and Deep Representation Features for Bearing Fault Diagnosis[J]. Measurement, 2020: 108603, https://doi.org/10.1016/ j.measurement.2020.108603.
15. Mnih V, Heess N, Graves A. Recurrent models of visual attention[C]//Advances in neural information processing systems. 2014: 2204-2212.
16. Parmar M, Devi V S. Neural Machine Translation with Recurrent Highway Networks[C]//International Conference on Mining Intelligence and Knowledge Exploration. Springer, Cham, 2018: 299-308. https://doi.org/10.1007/978-3-030-05918-7_27
17. Plakias S, Boutalis Y S. Fault detection and identification of rolling element bearings with Attentive Dense CNN[J]. Neurocomputing, 2020, 405: 208-217, https://doi.org/10.1016/j.neucom.2020.04.143.
18. Singh J, Azamfar M, Ainapure A, et al. Deep learning-based cross-domain adaptation for gearbox fault diagnosis under variable speed conditions[J]. Measurement Science and Technology, 2020, 31(5): 055601, https://doi.org 10.1088/1361-6501/ab64aa.
19. Singleton R K, Strangas E G, Aviyente S. Extended Kalman filtering for remaining-useful-life estimation of bearings[J]. IEEE Transactions on Industrial Electronics, 2014, 62(3): 1781-1790, https://doi.org 10.1109/ TIE.2014.2336616.
20. Srivastava R K, Greff K, Schmidhuber J. Highway networks[J]. arXiv preprint arXiv:1505.00387, 2015, https://doi.org/10.48550/arXiv.1505.00387.
21. Tao J, Liu Y, Yang D. Bearing fault diagnosis based on deep belief network and multisensor information fusion[J]. Shock and Vibration, 2016, 2016, https://doi.org/10.1155/2016/9306205.
22. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017: 5998-6008.
23. Wang H, Xu J, Yan R, et al. Intelligent bearing fault diagnosis using multi-head attention-based CNN[J]. Procedia Manufacturing, 2020, 49: 112-118, https://doi.org/10.1016/j.promfg.2020.07.005.
24. Wang J, Mo Z, Zhang H, et al. A deep learning method for bearing fault diagnosis based on time-frequency image[J]. IEEE Access, 2019, 7: 42373-42383, https://doi.org /10.1109/ACCESS.2019.2907131.
25. Wang X, Mao D, Xiaodong L I. Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network[J]. Measurement, 2020: 108518, https://doi.org/10.1016/j.measurement.2020.108518.
26. Wang X, Takaki S, Yamagishi J. Investigating very deep highway networks for parametric speech synthesis[J]. Speech Communication,

2018, 96: 1-9, https://doi.org/10.1016/j.specom.2017.11.002.

27. Wu J, Tang T, Chen M, et al. A study on adaptation lightweight architecture based deep learning models for bearing fault diagnosis under varying working conditions[J]. Expert Systems with Applications, 2020, 160: 113710, https://doi.org/10.1016/j.eswa.2020.113710.

28. Xu G, Hou D, Qi H, et al. High-speed train wheel set bearing fault diagnosis and prognostics: A new prognostic model based on extendable useful life[J]. Mechanical Systems and Signal Processing, 2020, 146: 107050.

29. Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention[C]//International conference on machine learning. 2015: 2048-2057.

30. Xueyi L I, Jialin L I, Yongzhi Q U, et al. Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning[J]. Chinese Journal of Aeronautics, 2020, 33(2): 418-426, https://doi.org/10.1016/j.cja. 2019.04.018.

31. Yin Z, Li G, Du C, et al. An adaptive speed estimation method based on a strong tracking extended Kalman filter with a least-square algorithm for induction motors[J]. Journal of Power Electronics, 2017, 17(1): 149-160, https://doi.org/10.6113/JPE.2017.17.1.149.

32. Zia T, Razzaq S. Residual recurrent highway networks for learning deep sequence prediction models[J]. Journal of Grid Computing, 2020, 18(1): 169-176, https://doi.org/10.1007/s10723-018-9444-4.

33. Zia T. Hierarchical recurrent highway networks[J]. Pattern Recognition Letters, 2019, 119: 71-76. https://doi.org/10.1016/j.patrec.2018.06.023

34. Zilly J G, Srivastava R K, Koutnık J, et al. Recurrent highway networks[C]//International Conference on Machine Learning. PMLR, 2017: 4189-4198.

35. Zuo L, Zhang L, Zhang Z H, et al. A spiking neural network-based approach to bearing fault diagnosis[J]. Journal of Manufacturing Systems, 2020, https://doi.org/10.1016/j.jmsy.2020.07.003.