Research on Bearing Fault Diagnosis Method based on Deep Residual Long-term and Short-term Network

Chuanqing Ma[1,2], Sanpeng Deng[1,2, *]

[1] Institute of Robotics and Intelligent Equipment Tianjin University of Technology and Education, Tianjin 300222, China

[2] Tianjin Key Laboratory of Intelligent Robot Technology and Application, Tianjin 300350, China

Abstract

Aiming at the problems that the classification of noisy data relies on expert knowledge and the processing is cumbersome, a fault diagnosis model based on the combination of Deep Residual Contraction Network (DRSN-CW)and Long Short-Term Memory Neural Network (LSTM) is proposed-Deep Residual Long Short-Term Network(DRSN-CW-LSTM). On the one hand, DRSN-CW extracts the spatial features of the signal while noise reduction, and on the other hand, LSTM extracts the temporal features of the signal, so that the signal can be fully extracted without the need for pre-noise reduction, and experimental analysis shows that the DRSN-CW-LSTM model has a better diagnostic performance in a strong noise environment.

Keywords

Fault Diagnosis; Data Acquisition; Data Noise Reduction; Deep Residual Contraction Network; Long and Short Term Memory Network.

1. Introduction

In the bearing fault diagnosis process, due to the complex and changing working environment around, mechanical equipment operation process noise and other reasons, the bearing vibration signal acquisition process will inevitably be collected together with the noise data, if the information containing noise is not appropriately processed, the algorithmic model of the final data classification accuracy will be greatly reduced. Deep learning can freely set the parameters of each layer, and the features in the data can be effectively extracted through autonomous learning, so it has been widely used in the field of equipment health detection and fault diagnosis.

Zhao et al [1]proposed a sparse filtering combined with variational modal decomposition (VMD) method for gearbox fault diagnosis, firstly, the signal is decomposed into a set of unidirectional quantities by VMD, and then the instantaneous amplitude energy is computed for them, and the features with excellent performance are extracted by sparse filtering to improve the diagnostic performance of the algorithm in a noisy environment. Fan Yuxue[2]et al. solved the problem of low diagnostic accuracy of LSTM under small batch samples by a two-way LSTM to accomplish intelligent fault diagnosis for small batch samples. Li Junxing et[3]alproposed a bearing fault diagnosis method based on EEMD and CNN-BiLSTM, firstly, the signal is decomposed and reconstructed by EEMD to eliminate the noise features of the signal, then the retained feature information is inputted into CNN-BiLSTM for feature extraction, and finally, it is compared with other deep learning methods, and the results show that the method has better performance. Ke Wei et al[4]proposed a fault diagnosis method based on similar segmented collaborative filtering (SSCF) and temporal rescheduling synchronization (TSET) strategy for the problem of noise interference

Volume 10 Issue 2, 2024
DOI: 10.6919/ICJE.202402_10(2).0002

during bearing diagnosis, and the experiments show that it has strong robustness for bearing fault diagnosis under strong noise environment.Zhao et al[5]proposed a bearing fault diagnosis method based on deep residual shrinkage network, which for the first time integrates the noise reduction method with neural network, and the results show that the method has better performance under the noise environment. noise reduction method and neural network fusion to form an integrated method of noise reduction and feature extraction, and validated it on a public dataset.

The above methods only focus on one of the data spatial features or temporal features, while in the process of fault diagnosis of bearings, while focusing on the spatial relationship of the data, it is also necessary to focus on the changes of the data in the temporal dimension, for this reason, this paper combines the combination of the DRSN-CW and the LSTM to excavate more potential data features, in order to improve the model's accuracy of the fault diagnosis in the presence of strong noise.

2. Overall Model Architecture and Principles

Aiming at the limitations of the single network model to identify faults, the insufficient utilization of the dynamic time-series characteristics of the data, and the cumbersome processing of strongly noisy data, the DRSN-CW-LSTM fault diagnosis method is proposed, which contains three modules, namely, data preprocessing, feature extraction, and data visualization. The first part is the data preprocessing module, which slices the samples by non-overlapping sampling. The second part is the feature extraction module, which consists of the DRSN-CW and LSTM modules, using the small sub-network of the residual contraction block in the DRSN-CW to adaptively set the threshold for the feature map, carry out the noise reduction of the signal, and complete the spatial feature extraction of the signal through the convolution, pooling, and batch normalization operations in the deep residual contraction module, and then take advantage of the processing of the temporal data in the LSTM, to complete the extraction of the signal timing features. Then use the advantage of LSTM to process temporal data to complete the extraction of signal temporal features. Finally, the results are output and visualized. The overall architecture of the model is shown in Fig. 1.

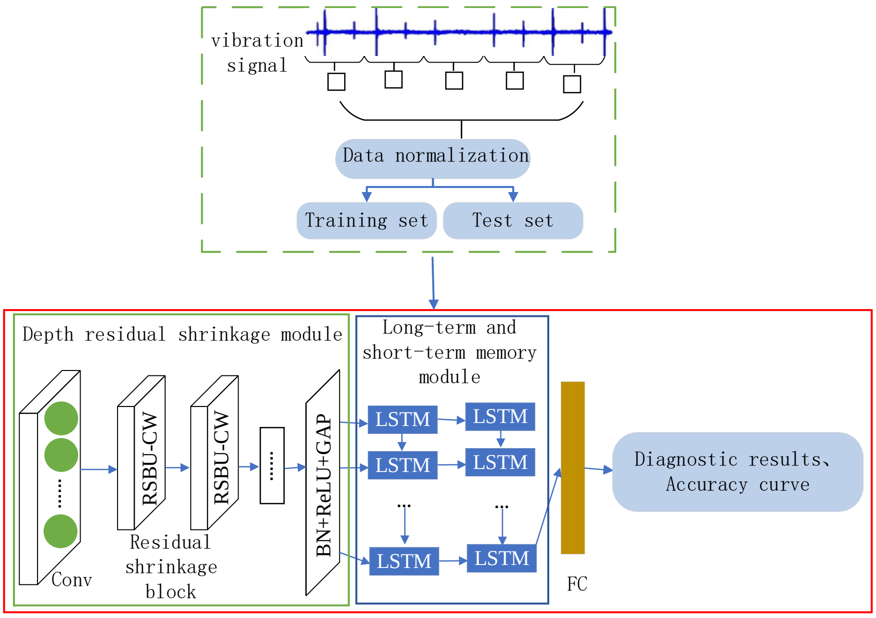

Fig.1 Overall Model Architecture

2.1 Spatial Feature Extraction for Deep Residual Contraction Networks

DRSN is a deep learning algorithm that combines signal soft-threshold noise reduction method, residual learning method and deep neural network proposed by Minghang Zhao et al [5] in 2020, including DRSN-CW and DRSN-CS. DRSN-CW performs feature shrinkage by adaptively setting different thresholds for the feature maps of different channels, so for the signals contaminated by noise, DRSN-CW has a In addition, the residual contraction module in DRSN-CW reduces one step of averaging operation than that in DRSN-CS, which makes the residual module in DRSN-CW faster.The residual contraction in DRSN-CS and DRSN-CW are respectively the shared threshold residual contraction unit (RSBU-CS) and the different threshold residual contraction unit (RSBU-CS) among channels and the different threshold residual contraction unit (RSBU-CS) among channels. threshold residual contraction unit (RSBU-CW), and the structure of RSBU-CW is shown in Fig. 2. Therefore, the article adopts DRSN-CW network with better average accuracy and less computing time for signal spatial feature extraction.

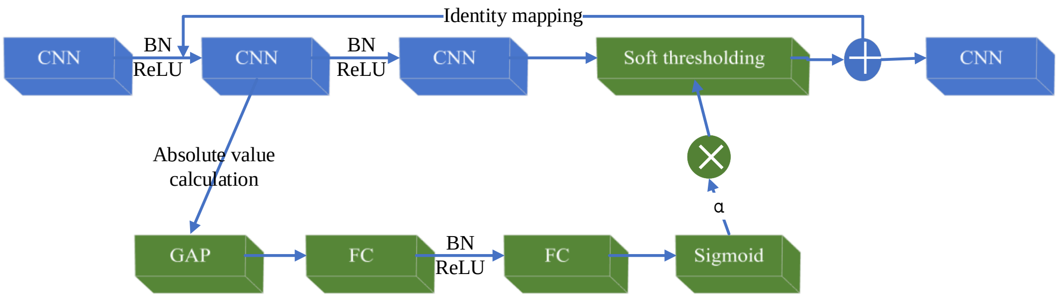

Fig.2 RSBU-CW structure diagram

RSBU-CW working principle: assuming that the input shape of the feature map is C*W*1, where C and W*1 are the number of channels and the size of the input feature map, the feature map is input into the residual shrinkage module and then two convolution operations are performed to get the output feature map after convolution, and then the absolute value of the output feature map is obtained by global average pooling (GAP) to get C one-dimensional feature vectors, i.e., the feature map of C*1*1; in another path, the GAP features are input into a small fully connected subnetwork, and a scaled parameter is obtained by nonlinear transformation with the fully connected and activation functions. of the feature map; in the other path, the features after GAP are input to a small fully connected sub-network, which is nonlinearly transformed by the fully connected and activation functions to obtain a scaled parameter, and the output is adjusted to a number between 0 and 1 by a Sigmoid function denoted as α; and then a number between 0 and 1 is multiplied by the average of the absolute values of the feature maps, i.e., A * α to obtain the final threshold value. Finally, the signal noise reduction features obtained from soft thresholding are associated with the input features through residual linking to complete residual learning. The difference with RSBU-CS is that RSBU-CW has no demand averaging operation and the rest of the operations are the same.The computation of the threshold value in RSBU-CW is calculated as shown in equation (1):

$$\tau_c = \frac{1}{1+e^{-z_c}} * |x_{i,j,c}| \tag{1}$$

Where $\tau_c$represents the result of soft thresholding, $z_c$is the output of the cth neuron of the second fully connected layer, $1/(1+e^{-z_c})$ is the result of the scaling of the second fully connected layer by the Sigmoid activation function, i, j, and c represent the length and width and the number of channels of the feature maps of the second convolutional layer, respectively, $x_{i,j,c}$ is the output of the cth channel of the second convolutional layer.

2.2 Temporal Feature Extraction for Long and Short-term Memory Networks

LSTM, as a variant of recurrent reach-in network (RNN), solves the gradient vanishing and gradient exploding problems of the original RNN through the unique gate mechanism and memory cells to learn the long-term dependent information at the same time, LSTM judges the input data through its gating unit, leaving the data that conforms to the rules, and forgetting the data that doesn't conform to the rules, and the internal neuron structure of the LSTM is shown in Fig. 3.

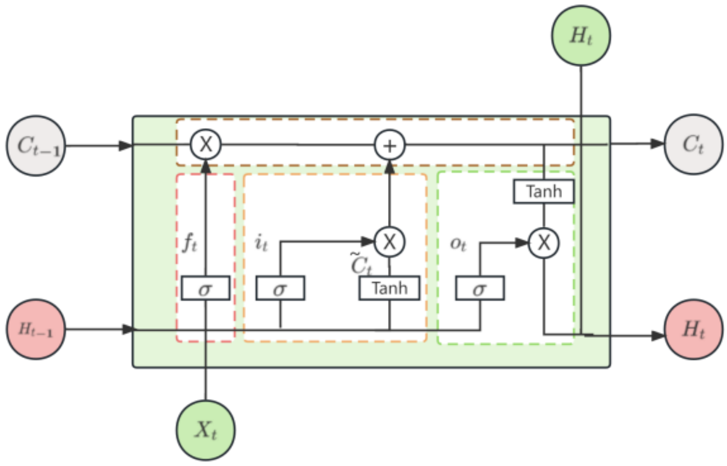

Fig.3 Internal neuronal structure of LSMT

LSTM updates the state of the current "cell" through three gating information: input gate, forgetting gate and output gate. Firstly, LSTM selectively forgets the output information of the previous cell through the forgetting gate, and decides the information to be discarded by calculating Ft, and finally outputs the final state of the current neuron under the calculation of the output gate. result, the mathematical expression is shown in equation (2).

$$H_t = O_t \cdot tanh(C_{t-1} \cdot F_t + \tilde{C}_{t-1} \cdot I_t) \tag{2}$$

Where F_t, I_t, O_t, C_t, $\tilde{C}_t$represent the forgetting gate, the input gate, the output gate, the output of the candidate "cell" and the memory "cell" after computation, respectively, and t and t-1 represent the current and previous moments, respectively. The feature map to be computed is input to the LSTM according to a certain format, and the temporal features of the data are extracted step by step by forward propagation of the LSTM.

3. Experimentation and Analysis

3.1 Data Description

Table 1. data set component

Fault type	SNR			
	-10	-8	-6	0
normal	488	488	488	488
Bearing inner ring	488	488	488	488
Bearing outer ring	488	488	488	488
Rolling body failure	488	488	488	488

In order to verify the performance of the improved DRSN-CW algorithm, it is verified by the data set of Jiangnan University, which is divided into four categories of faults, namely: normal, bearing inner ring faults, bearing outer ring faults, and rolling body faults. The data points of each fault category

ISSN: 2414-1895
DOI: 10.6919/ICJE.202402_10(2).0002

are half a million, and the data points are sampled by non-overlapping sampling, with 1024 data points as a set of samples, which are divided into training samples and test samples according to the ratio of 7:3. The data set components are shown in Table 1.

In the experiment, in order to restore the noise information generated during the operation of the machine and to verify the noise immunity of the model, so different levels of Gaussian noise are introduced into the signal to generate noise signals with different signal-to-noise ratios, and the signal-to-noise ratio formula and the probability density expression of the Gaussian distribution are shown in Equation (3)and(4):

$$SNR = 10log_{10}(\frac{P_{signal}}{P_{noise}}) \tag{3}$$

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{4}$$

Where, SNR is the calculated signal-to-noise ratio, and P_signal and P_noise represent the power of signal and noise, σ, μ, and x represent the standard deviation, mean, and random variable, respectively. Here five different groups of noise samples are generated on the original signal, and the signal-to-noise ratios of the signals are -6dB, -8dB and -10dB, respectively. Taking the vibration signals of the bearing inner ring failure as an example, 20,000 vibration sample points are selected, and the signal pairs of their different signal-to-noise ratios are shown in Fig. 4. It can be seen that, with the increase of noise content, the vibration signal of the multiple impact signal is flooded by impulse noise, it is difficult to identify the periodic components of the vibration signal, resulting in difficulties in signal feature extraction.

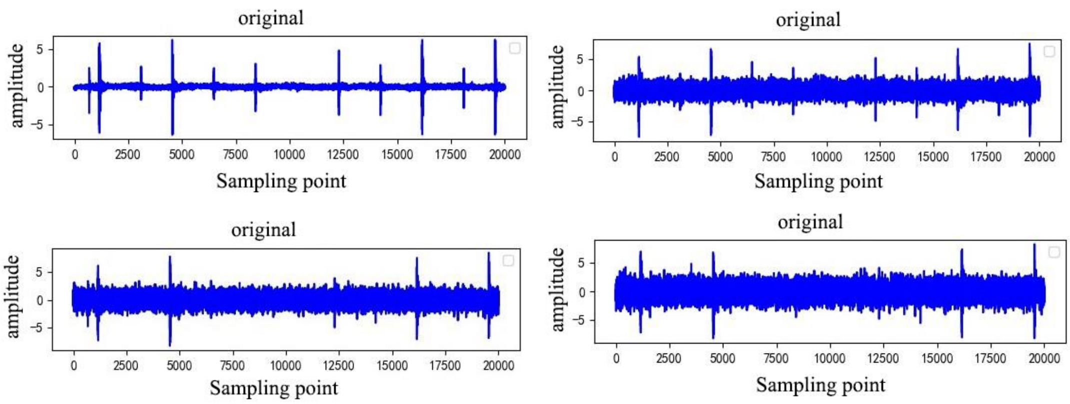

Fig.4 Effect of different signal-to-noise ratios on the signal

3.2 Comparative Analysis of Experiments

In order to verify the advantages of the algorithms in strong noise environments, the article selects the following three data classification algorithms to compare ResNet34, LSTM and DRSN-CW. A single model is used to compare the algorithmic effects of model fusion in order to validate the advantages of model fusion in data classification.

(1) Diagnostic results of each model in different noise environments

The diagnostic results of the DRSN-CW-LSTM model and the existing algorithm model at different SNRs are shown in Table 3-2. The accuracy is obtained by repeating the experiment five times to take its average value, in the case of the original data to SNR=-10dB, that is, the noise content gradually increases, the average accuracy of the DRSN-CW-LSTM model in the training set decreases from 99.51% to 95.47%, and with the increase of the noise content, the model's accuracy

Volume 10 Issue 2, 2024
DOI: 10.6919/ICJE.202402_10(2).0002

also decreases, but the result is still satisfactory, and the noise reduction in the signal of the signal while extracting the features of the signal, so the model has a stronger feature extraction ability in strong noise environments.

Table 2. Average Accuracy of Test Sets for Different Models at Different Signal-to-Noise Ratios(%)

algorithmic model	Origin data	SNR=-6	SNR=-8	SNR=-10
ResNet34	94.66	89.57	87.55	84.55
LSTM	92.72	84.22	84.54	82.32
DRSN-CW	97.35	94.57	93.68	93.15
DRSN-CW-LSTM	99.51	98.26	96.54	95.47

(2) Visualization results of different models in strong noise environment

The diagnostic accuracies of the different models in the training and test sets in a strong noise environment with a noise of -10dB are shown in Fig. 5. It can be seen that the accuracy of the DRSN-CW-LSTM model is higher than the other three models in both the training and test sets in a strong noise environment, and the diagnostic effect of DRSN-CW is only second to that of DRSN-CW-LSTM because of the soft-thresholding operation contained in DRSN-CW itself, so the robustness of the DRSN-CW-LSTM model for the feature extraction of noisy signals is better.

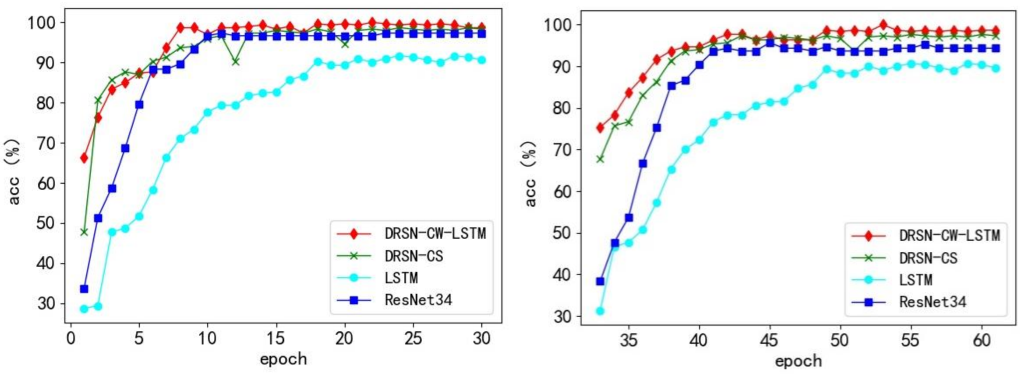

Fig.5 Test and training set accuracy curves for different models with S/N ratio = -10

4. Generalize

In this paper, for the problem of fault diagnosis under strong noise environment, we propose a servo motor fault diagnosis model that can be fully extracted with adaptive noise reduction + features under strong noise environment. Finally, the effectiveness of the algorithm model is verified by experimentally comparing and analyzing the data recognition accuracy and misclassification of different models under different noise, different models under the same level of noise, and the method of this paper under strong noise.

Acknowledgments

Tianjin Municipal Education Commission Scientific Research Plan Natural Science Key Projects (2022ZD026, 2022ZD032).

References

[1] Zhao C, Feng Z. Application of multi-domain sparse features for fault identification of planetary gearbox[J]. Measurement, 2017, 104: 169-179.

[2] Fan Yuxue,Wang Jiangwen,Mei Guiming et al. Research on small sample rolling bearing fault diagnosis method based on BI-LSTM[J]. Noise and vibration control,2020,40(04):103-108.

[3] Li Junxing,XU Xing,Jia Xianzhao et al. Research on rolling bearing fault prediction method based on EEMD and CNN-BiLSTM noise environment[J/OL]. Bearing:1-10[2023-11-14].

[4] Ke Wei,Jin Zhongping,Dong Lingjun et al. A rolling bearing impact fault diagnosis method based on SSCF and TSET[J]. Mechanical and electrical engineering,2023,40(06):835-843.

[5] ZHAO M H, ZHONG S S,FU X Y, et al.Deep residual shrinkage networks for fault diagnosis [J]. IEEE Transactions on Industrial Informatics, 2020, 99: 1.