Academic Journal of Engineering and Technology Science
ISSN 2616-5767 Vol.5, Issue 3: 5-12, DOI: 10.25236/AJETS.2022.050302

Bearing fault diagnosis under class unbalanced data based on deep learning

Jialin Ma[1,a,*]

[1] *College of Computer and Communication, Lanzhou University of Technology, Lanzhou, 730050, China*
[a] *jialinm@lut.edu.cn*
** Corresponding author*

Abstract: *There is a huge difference in the number of operating samples and failure samples in industrial production. If the diagnostic model is trained through deep learning under an unbalanced data set, it will make the model recognize the faulty samples as normal samples. Aiming at this problem, an adaptive focus loss function mechanism is proposed. It can avoid over-learning large-scale samples during small-batch imbalance training. At the same time, to improve the generalization ability of fault samples, a pre-training data enhancement mechanism is proposed. By using the rolling element bearing fault diagnosis data set, the effectiveness of the two mechanisms is verified. The two mechanisms can not only complete training tasks under unbalanced data and improve generalization ability.*

Keywords: *Bearing fault diagnosis, Class imbalance dataset, Adaptive focus loss function, Pre training data enhancement*

1. Introduction

Fault diagnosis is an important technology to ensure safe production in the industrial field. From the perspective of diagnostic methods, it can be divided into model-based methods such as gray fault diagnosis [1], Kalman filter [2], etc.; data-driven methods such as SVM [3], deep learning [4]-[7], extreme learning machine [8], etc. In recent years, with the rapid development of deep learning [9], [10], the diagnosis results have been greatly improved by acquiring sensor data and then performing pattern recognition.

When used for fault diagnosis of rolling element bearings, vibration signals are generally obtained through vibration sensors for diagnosis. The data acquired by the sensor is a continuous signal, which is input to the deep learning model to extract feature vectors, and finally gives the diagnosis result through the classifier. The model structure of deep learning includes multi-layer perceptron [11]-[13], convolutional neural network [14]-[16], etc. The researchers improve the model’s ability to express feature vectors for signal extraction by optimizing the network structure. Dibaj et al. [14] proposed a method of extracting temporal features of signals through Variational Mode Decomposition (VMD), and then further extracting feature maps for classification through convolutional neural networks. Jing et al. [15] proposed a one-dimensional convolutional neural network structure applied to gearboxes for fault diagnosis of gearbox signals. Huang et al. [17] proposed a fault diagnosis method under complex working conditions based on the combination of long-short-term memory network and convolutional neural network, and achieved extremely high diagnostic results under different working conditions. [18] proposed a multi-scale convolutional neural network for fault diagnosis by extracting feature maps at different scales. However, in real industrial production, the faulted data accounts for a very small proportion of the total data, so we need to construct class-imbalanced data to improve deep learning for fault diagnosis results when the number of fault data is much smaller than the normal data.

Deep learning is usually trained in the form of mini-batch. Under the class imbalance data, a mini-batch may not have any small-scale samples at all, so this training will shorten the small-scale samples and large-scale samples. The distance of the labels will eventually lead to inaccurate diagnostic results for small samples. Pei et al. [19] proposed a CS-Boosting method to solve the bearing fault diagnosis method with less data under class imbalance. Shi and Zhang et al. [20] proposed to optimize the Support Vector Machine (SVM) algorithm by Grey Wolf Optimizer (GWO), which was applied to unbalanced datasets in the context of autonomous driving. Xu et al. [21] proposed a reproducible fusion fault diagnosis network for fault diagnosis problems under variable speed, quasi-imbalanced data.

When the small-scale samples are further reduced and the class imbalance intensifies, some problems

ISSN 2616-5767 Vol.5, Issue 3: 5-12, DOI: 10.25236/AJETS.2022.050302

will become more obvious: 1) When there are no samples of a certain category in multiple consecutive mini-batches, the extracted feature vectors will be biased towards other results as a whole, reducing the performance of this category. precision. 2) The accuracy of the overall data will be very high, but the data of the fault category under fault diagnosis cannot be identified, so such evaluation is irresponsible. Aiming at these two problems, the main work of this paper is as follows:

1) An adaptive loss class imbalance loss function is proposed, which can significantly improve the training performance of small-scale samples during mini-batch training.

2) Design a class-imbalanced data augmentation pre-training framework, which effectively improves the generalization ability of small-scale classes.

3) The diagnostic model was evaluated by means of confusion matrix evaluation.

2. Related work

2.1. Class Imbalanced Data for Fault Diagnosis

Rolling element bearings in working state usually show different vibration characteristics under different working conditions and health conditions. The vibration signal is generally obtained continuously through the vibration sensor fixed on the equipment, and the sensor is defined to obtain a signal with a length of health status. According to the structure of the bearing, it includes normal, rolling element failure, inner ring failure, and outer ring failure. In fact, the bearing rarely fails under normal operation, and it is difficult to accurately mark the fault when it occurs. Therefore, the researchers use the electric spark to create the fault to collect the data of the corresponding position fault. Define normal, rolling element fault, inner ring fault, and outer ring fault four kinds of label data are projected onto the two-dimensional plane and the collected domain is shown in Figure 1.

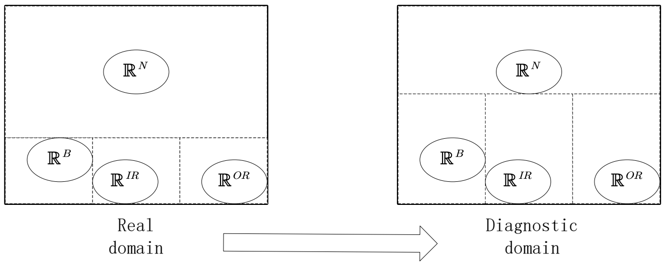


Figure 1: Data distribution diagram.

Where $\mathbb{R}^N, \mathbb{R}^B, \mathbb{R}^{IR}, \mathbb{R}^{OR}$ represent the domain of the corresponding data respectively, and within the boundary of the outer dotted line is the corresponding real distribution domain. On the left is the real distribution domain of the four labels, but under limited data, such as the wrong diagnostic domain on the right, it will also be considered to completely identify the data under the fault. Fortunately, the normal operation data can be collected indefinitely, that is, expanding $\mathbb{R}^N$ to divide the real domain under normal and fault conditions as much as possible in the diagnostic domain.

Although this method seems to maximize the generalization ability of the diagnostic model, under deep learning, since the data volume of large-scale samples is much larger than that of small-scale samples, it will bring huge problems to the training task of the model.

2.2. 1D convolutional neural networks and imbalanced datasets

One-dimensional convolutional neural networks [22] can extract feature vectors from time-domain signals and define the mapping of one-dimensional convolution $f^{\mathrm{CNN}}: x \rightarrow \mathrm{y}$ such as formula (1) shown.

$$y = W{*}x + \boldsymbol{b} \quad (1)$$

Where $W, \boldsymbol{b}$ represents the weight, which contains two key parameters in the convolution operation, the size and stride of the convolution kernel. The weight in the convolution kernel is dot-multiplied with a part of the signal in the sliding window to obtain this the output of a layer, the step size determines the span over which the sliding window moves. In the convolution operation, the number of convolution kernels can control the number of output channels. After there are too many parameters, down sampling is performed through the pooling layer to reduce the number of features, including the maximum and average pooling layers, that is, the feature vector in a sliding window. Take the largest, average operation to down sample as an eigenvalue.

After the feature vector is obtained through the convolutional neural network, the dimension is reduced to the vector with the same number of categories through the fully connected layer, and the fully connected mapping $f^{\mathrm{MLP}}: x \rightarrow y$ is defined as formula (2) Show.

$$y = Wx + \boldsymbol{b} \quad (2)$$

Generally, after obtaining the same feature vector as the number of categories, the Softmax classifier outputs the probability corresponding to each category, as shown in formula (3).

$$P(y' = j) = \frac{e^{x^j W + b}}{\sum_{k \in |K|} e^{x^k W + b}} \quad (3)$$

The output of formula (3) represents the probability that the output y' of the model is j, and $|K|$ represents the case of all classifications. When evaluating the performance of the diagnostic model, the accuracy rate is generally used as the evaluation index, as shown in formula (4).

$$ACC = \mathop{\mathbb{E}}_{y' \in Y', y \in Y} (1_{y' = y}) \quad (4)$$

$$\mathcal{L} = -\sum_{i=1}^{|K|} \sum_{c \in C} P(y_i = c) \log P(y_i{}' = c) \quad (5)$$

Where Y' represents the diagnostic result of the model on the input set X, and Y represents the actual label of the input set. The accuracy statistics in formula (4) correspond to the probability that the diagnosis result is the same as the actual label. In fact, deep learning aims to maximize the ACC evaluation index by minimizing the cross-entropy loss function as formula (5).

Assuming that when large-scale samples: small-scale samples are 49:1, even if all samples are fully identified as labels for large-scale samples, there is still a 98% accuracy rate, and the evaluation at this time is positive. However, as a matter of fact, in the context of fault diagnosis, it can be stated that this diagnostic model does not identify fault samples at all. Therefore, under the class imbalanced data, more evaluation methods are needed to correctly evaluate the model. On the other hand, the original training direction of maximizing the overall classification loss, mini batch and other methods will still maximize the accuracy, while ignoring the accuracy of small-scale samples.

2.3. Focal loss function

The focal loss function is a method to deal with the training of class-imbalanced datasets. When the number of large-scale samples is too large and easy to classify, the optimization direction of the model, that is, the optimization direction of the loss function, will be biased towards large-scale samples. At this time, 1) the loss generated by large-scale samples is large; 2) the loss generated by large-scale samples accounts for a large proportion. The researchers [23] proposed the amount function in the binary classification task, as shown in formula (6)

$$\mathcal{L} = \begin{cases} -(1 - y')^{\gamma} \log y' & y = 1 \\ -y'^{\gamma} \log(1 - y') & y = 0 \end{cases} \quad (6)$$

Where $\gamma > 0$ reduces the contribution of large-scale samples to the loss function. Compared with the cross-entropy loss, the loss of large-scale samples is reduced by γ, and the loss of effective samples is expanded, thereby highlighting the performance of small-scale sample accuracy.

However, γ is a hyperparameter that needs to be set manually. In addition, there are still small-scale samples in each training batch in the task, and the fault diagnosis task may not have small-scale samples in the mini batch under the extremely unbalanced dataset. sample size.

ISSN 2616-5767 Vol.5, Issue 3: 5-12, DOI: 10.25236/AJETS.2022.050302

3. Weighted training method based on small-scale sample data augmentation

3.1. Adaptive Focal Loss Function

Four output types are included in bearing fault diagnosis, so the focal loss function is first extended to multi-classification, as shown in formula (7).

$$\mathcal{L}=\sum_{i=1}^{|K|}\sum_{c\in C}P(y_i=c)^{\gamma^i}\log P(y_i{}'=c) \tag{7}$$

Among them, γ^i represents the discount parameter of the i category. In the focal loss function, when the number of class samples is simple and large, the γ is smaller; otherwise, the γ is larger. However, in the mini batch task, in each epochs, the mini batch randomly samples from the overall samples, and the number of samples in the four categories in the mini batch is defined as N^N, N^B, N^{IR}, N^{OR}. However, since each mini batch is randomly sampled, the initial hyperparameter setting cannot determine the proportion of each category of samples in the mini batch, and it is difficult to set γ. To this end, this paper proposes a mini-batch adaptive focus loss function as shown in formula (8).

$$\mathcal{L}=\sum_{i=1}^{|K|}\sum_{c\in C}P(y_i=c)^{1/\frac{N^c}{\sum_{d\in C}N^d}}\log P(y_i{}'=c) \tag{8}$$

When calculating the loss during the training process, the discount parameter is set adaptively with the distribution of the number of samples. The larger the number of class samples, the smaller the class loss function, and vice versa to achieve an adaptive effect. In the limit state, when all the large-scale samples in a mini batch, the loss calculated during training will be very small, and the updated gradient will be slow, which will not cause the lack of small-scale samples to be stretched out. However, it is still difficult to train small-scale samples only through the adaptive focal loss function mechanism. For this reason, this paper gives the model prior knowledge through the pre-data enhancement mechanism, and fine-tunes the model through the adaptive focal loss function mechanism under the prior model. diagnosis results.

3.2. Class Imbalanced Data Augmentation Pretraining

During diagnosis, the bearing vibration signal is divided into each data segment as a sample by window cutting. Generally, researchers use a non-overlapping form to divide the sample, as shown in Figure 2.

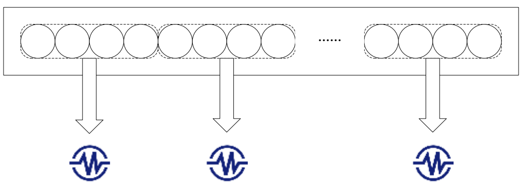


Figure 2: Schematic diagram of non-overlapping data segmentation.

The non-overlapping segmentation method ensures no duplication between data, however, overlapping data in the convolutional neural network can better train the convolution kernel weights, as shown in Figure 3.

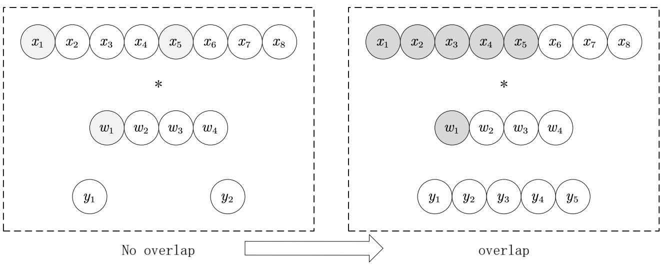


Figure 3: Effect of overlap on training's Schematic diagram.

ISSN 2616-5767 Vol.5, Issue 3: 5-12, DOI: 10.25236/AJETS.2022.050302

In the non-overlapping training, the weight w_1 is only fitted with x_1, x_5, but in fact when a signal starts is uncertain, and overlapping training can make w_1 and $x_1 \sim x_5$ both Fitting is performed to fit the signal segment at each starting point. More researchers adopt the method of non-overlapping segmentation in order to avoid the correlation between samples. In the fault diagnosis problem of class imbalance, since small-scale samples have more severe problems, overlapping sampling has to be considered to expand the class sample set.

After overlapping sampling, the number of small and medium-sized samples in fault diagnosis can be rapidly expanded to a large number. However, in fact, the similarity between samples is high, and it is difficult to improve the generalization ability of the model to small-scale samples. There is a very significant feature in the rolling element bearing signal, usually the amplitude of abnormal health samples is significantly different from normal. In this way, improving the generalization ability by adding perturbation becomes a feasible means.

The signal of fault class improves the generalization ability of the model to small-scale samples by adding random disturbances that obey the Gaussian distribution. Small convolution kernels in convolutional neural networks are more sensitive to disturbances in a small range. Different convolution kernels obtain the feature vectors of the signal from different scales (different fields of view), and more comprehensively extract the time domain features of the signal. As explained above, overlapping sampling may reduce the generalization ability of the class, so the overall training method is shown in Figure 4.

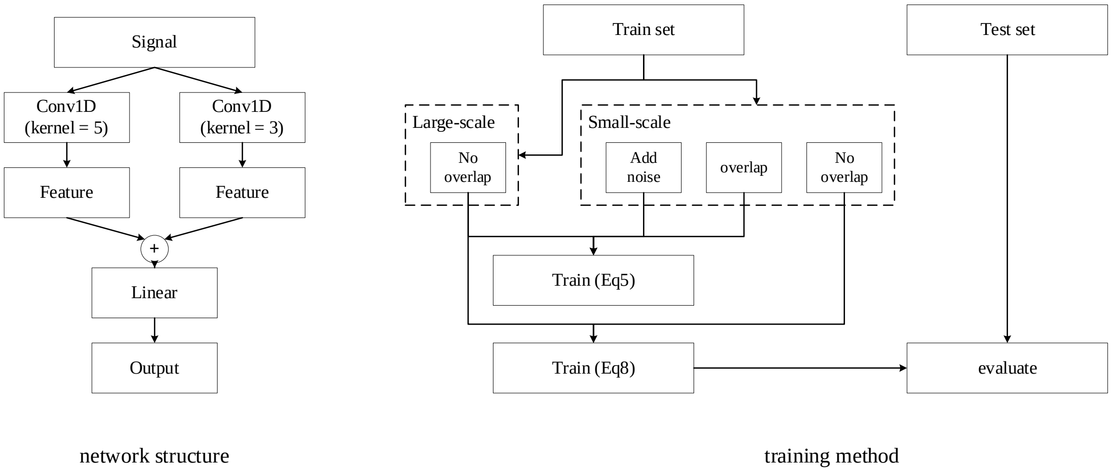


Figure 4: Model training method diagram.

In the training set, small-scale classes are enhanced by adding noise and overlapping sampling. At this time, the class imbalance dataset is balanced by data augmentation, so formula 5 is used for training. After the training, the model is further trained through the adaptive focal loss function through non-overlapping samples, and finally the model is evaluated through the test set.

4. Experimental and discussion

4.1. Experiment preparation

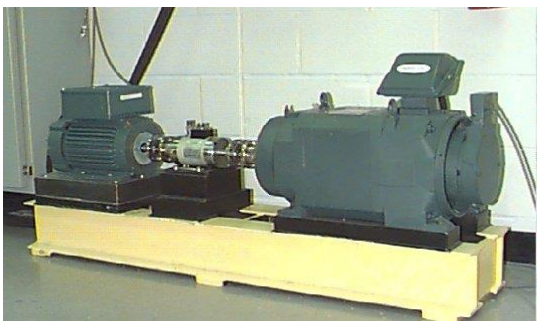


Figure 5: Bearing test bench.

ISSN 2616-5767 Vol.5, Issue 3: 5-12, DOI: 10.25236/AJETS.2022.050302

In this paper, the bearing data collected by Case Western Reserve University is used. The data set is collected by placing the experimental bearing on a metal test bench, and the vibration sensor is magnetically adsorbed under the experimental bench to obtain the vibration signal under the operating state of the bearing [24]. As shown in Figure 5.

In this experiment, four kinds of data of 0, 1, 2, and 3 are used for the experiment under 12000 samples/second. The sampling points of each set in the sample include 120,000. In order to create an unbalanced data set, after every 400 sampling points are divided without overlap, the sample distribution of the training set and the test set is shown in Table 1 (the same under the four loads).

Table 1: The number of samples in the different categories.

	Train set	Test set
Normal	240	60
Ball	15	285
Inner	15	285
Outer	15	285

Artificially, the ratio of normal (large-scale samples) to each fault (small-scale samples) in the training set is 16:1, and the other fault samples are used as the test set to test the generalization ability.

In the experiment, the detailed structure of the network structure is shown in Table 2. Compared with the method in this paper, a large convolution kernel network with a convolution kernel of 5 and a small convolution kernel network with a convolution kernel of 3 are selected in the experiment. , without using the adaptive focal loss function method.

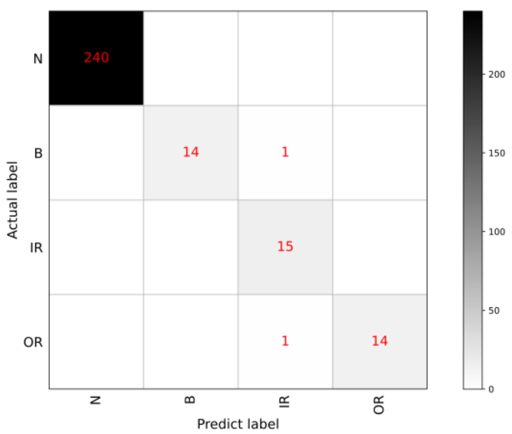
Table 2: Network structure.

<table>
<tr><th>kernel=5</th><th>kernel=3</th></tr>
<tr><td>Conv1D (1, 4)</td><td>Conv1D (1, 4)</td></tr>
<tr><td>Conv1D (4, 16)</td><td>Conv1D (4, 16)</td></tr>
<tr><td>Pooling layer</td><td>Pooling layer</td></tr>
<tr><td>Conv1D (16, 4)</td><td>Conv1D (16, 4)</td></tr>
<tr><td>Conv1D (4, 1)</td><td>Conv1D (4, 1)</td></tr>
<tr><td colspan="2">Linear (197, 64)</td></tr>
<tr><td colspan="2">Linear (64, 4)</td></tr>
<tr><td colspan="2">Softmax classifier</td></tr>
</table>

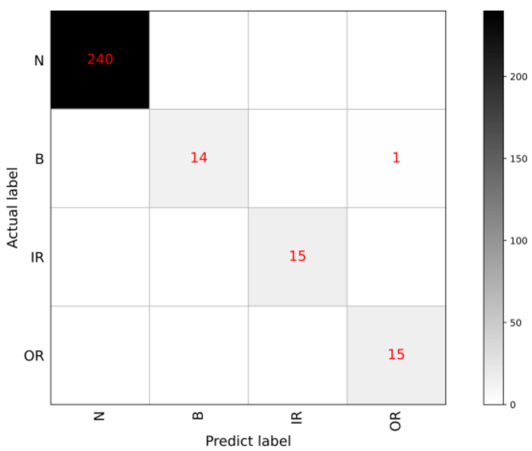
To fairly evaluate the diagnostic results under class imbalance, this paper adopts a confusion matrix representation of each class accuracy.

4.2. Experimental results

Training is carried out in the form of Figure 4, and the number of training iterations is set by early stop [25], which is a way to stop training when the model loss no longer decreases. After adding the pre-trained model through Equation 5, save the trained model and continue training through Equation 8. The results obtained under the four loads are shown in Figure 6(A-D).

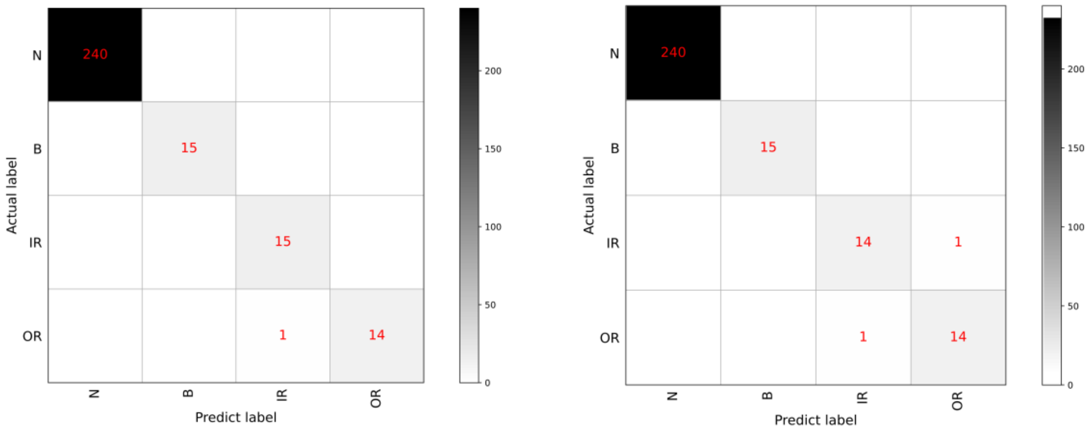


(A) Training set load 1.



(B) Training set load 2.

ISSN 2616-5767 Vol.5, Issue 3: 5-12, DOI: 10.25236/AJETS.2022.050302



(C) Training set load 3. (D) Training set load 4.

Figure 6: Diagnostic confusion matrix.

It can be seen from Figure 6 that the samples of the normal category are completely recognized through the two mechanisms, and the abnormal samples are not recognized as normal samples, and there is a certain misjudgment between the fault samples. This phenomenon will be discussed in this paper. Discuss analysis.

In order to avoid the chance of the experiment, the results of different methods on the test set were compared under 20 groups of experiments, as shown in Table 3.

Table 3: Accuracy of different methods.

Load	0	1	2	3
CNN (3)	0.1879 ±0.0527	0.2207 ±0.0491	0.1989 ±0.0739	0.1770 ±0.0813
CNN (5)	0.1825 ±0.0360	0.2180 ±0.1031	0.2153 ±0.0880	0.1934 ±0.0303
No enhanced pre-training	0.8224 ±0.0198	0.8252 ±0.0141	0.8115 ±0.0233	0.8098 ±0.0103
No focus loss	0.7421 ±0.0162	0.7475 ±0.0203	0.7240 ±0.0239	0.7347 ±0.0151
Our method	0.9748 ±0.0141	0.9705 ±0.0129	0.9530 ±0.0175	0.9639 ±0.0098

It can be seen from the results that both the focal loss training mechanism and the augmentation pre-training mechanism can be effective, and the combination of the two mechanisms can greatly improve the generalization ability of class-imbalanced datasets. Traditional convolutional neural network training through cross entropy will make small-scale samples largely identified as normal samples under the class-imbalanced data set. On the other hand, the generalization ability of small-scale samples cannot be guaranteed. The two mechanisms can well improve the diagnostic ability of class imbalanced data.

5. Conclusions

A kind of unbalanced data environment is proposed for the background that it is difficult to collect fault data and easy to collect non-fault data in fault diagnosis. This paper improves the diagnosis results for small-scale samples by proposing an adaptive focal loss function mechanism and a class-imbalanced data augmentation pre-training mechanism.

In the experimental verification link, this paper found that the diagnostic model can distinguish between normal and faults well, but there is still a certain error between faults. Further improving the accuracy of the test set is our next goal.

Acknowledgements

The present work was funded by the Gansu Education Department (2021CXZX-517).

References

[1] L. Li, Y. Xie, L. Cen, and Z. Zeng, 'A novel cause analysis approach of grey reasoning Petri net based on matrix operations', Appl. Intell., 2021, doi: 10.1007/s10489-021-02377-4.

ISSN 2616-5767 Vol.5, Issue 3: 5-12, DOI: 10.25236/AJETS.2022.050302

[2] F. Naseri, E. Schaltz, K. Lu, and E. Farjah, 'Real-Time Open-Switch Fault Diagnosis in Automotive PMSM Drives Based on Kalman Filter', IET Power Electron., vol. 13, 2020, doi: 10.1049/iet-pel.2019.1498.
[3] X. Zhang, C. Li, X.-B. Wang, and H. Wu, 'A novel fault diagnosis procedure based on improved symplectic geometry mode decomposition and optimized SVM', Measurement, vol. 173, p. 108644, 2020, doi: 10.1016/j.measurement.2020.108644.
[4] C. Keleolu, H. Küük, and M. Demetgül, 'Fault Diagnosis of Bevel Gears Using Neural Pattern Recognition and MLP Neural Network Algorithms', Int. J. Precis. Eng. Manuf., vol. 21, no. 5, 2020.
[5] Y. Han, W. Qi, N. Ding, and Z. Geng, 'Short-Time Wavelet Entropy Integrating Improved LSTM for Fault Diagnosis of Modular Multilevel Converter', IEEE Trans. Cybern., vol. PP, no. 99, pp. 1–9, 2021.
[6] Z. Tang, L. Bo, X. Liu, and D. Wei, 'A semi-supervised transferable LSTM with feature evaluation for fault diagnosis of rotating machinery', Appl. Intell., pp. 1–15, 2021.
[7] J. Zhang, B. Xu, Z. Wang, and J. Zhang, 'An FSK-MBCNN based Method for Compound Fault Diagnosis in Wind Turbine Gearboxes', Measurement, 2020.
[8] G. Vashishtha and R. Kumar, 'Pelton Wheel Bucket Fault Diagnosis Using Improved Shannon Entropy and Expectation Maximization Principal Component Analysis', J. Vib. Eng. Technol., 2021, doi: 10.1007/s42417-021-00379-7.
[9] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning, vol. 1. MIT press Cambridge, 2016.
[10] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', nature, vol. 521, no. 7553, pp. 436–444, 2015.
[11] V. N. Ghate and S. V. Dudul, 'Design of optimal MLP and RBF neural network classifier for fault diagnosis of three phase induction motor', Int. J. Adv. Mechatron. Syst., vol. 2, no. 3, p. 204, 2010.
[12] T. Waqar and M. Demetgul, 'Thermal analysis MLP neural network based fault diagnosis on worm gears', Measurement, vol. 86, pp. 56–66, 2016.
[13] M. A. Hui, D. Che, Q. Niu, and S. Xia, 'Research on Fault Diagnosis of Hoisting Bearing Based on Deep Neural Network', Comput. Eng. Appl., 2019.
[14] A. Dibaj, M. M. Ettefagh, R. Hassannejad, and M. B. Ehghaghi, 'A hybrid fine-tuned VMD and CNN scheme for untrained compound fault diagnosis of rotating machinery with unequal-severity faults', Expert Syst. Appl., vol. 167, no. January, 2020.
[15] L. Jing, M. Zhao, P. Li, and X. Xu, 'A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox', Measurement, vol. 111, pp. 1–10, 2017.
[16] X. Li, _J_ Li, C. Zhao, Y. Qu, and D. He, 'Early Gear Pitting Fault Diagnosis Based on Bi-directional LSTM', 2019.
[17] T. Huang, Q. Zhang, X. Tang, S. Zhao, and X. Lu, 'A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems', Artif. Intell. Rev., pp. 1–27, 2021.
[18] D. Wang, Q. Guo, Y. Song, S. Gao, and Y. Li, 'Application of Multiscale Learning Neural Network Based on CNN in Bearing Fault Diagnosis', J. Signal Process. Syst. Signal Image Video Technol., vol. 91, no. 10, pp. 1205–1217, 2019.
[19] Y. Pei, Z. Wang, H. Jiang, and Z. Liu, 'Fault Diagnosis Method Based on CS-Boosting for Unbalanced Training Data', J. Vib. Meas. Diagn., 2013.
[20] Q. Shi and H. Zhang, 'Fault diagnosis of an autonomous vehicle with an improved SVM algorithm subject to unbalanced datasets', IEEE Trans. Ind. Electron., vol. PP, no. 99, pp. 1–1, 2020.
[21] Xu, S. Li, Jiang, Z. An, and T. Yu, 'A renewable fusion fault diagnosis network for the variable speed conditions under unbalanced samples', Neurocomputing, vol. 379, 2019.
[22] E. Levent, I. Turker, and K. Serkan, 'A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier', J. Signal Process. Syst., 2018.
[23] H. Wu, S. Luo, H. Lin, D. Shuangda, Y. Guan, and J. Rojas, Recovering from External Disturbances in Online Manipulation through State-Dependent Revertive Recovery Policies. 2018. doi: 10.1109/ROMAN.2018.8525771.
[24] W. Smith and R. B. Randall, 'Rolling Element Bearing Diagnostics Using the Case Western Reserve University Data: A Benchmark Study', Mech. Syst. Signal Process., vol. 64–65, 2015, doi: 10.1016/j.ymssp.2015.04.021.
[25] J. Cao, J. Ma, D. Huang, and P. Yu, 'Finding the optimal multilayer network structure through reinforcement learning in fault diagnosis', Measurement, p. 110377, 2021, doi: https://doi.org/10.1016/j.measurement.2021.110377.