Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P2106-cd

Bearing fault diagnosis based on lifelong learning under cross operating conditions

Ning Hu

China Electronic Product Reliability and Environmental Testing Research Institute, China. E-mail: huning@ceprei.com

Shixin Jiang

China Electronic Product Reliability and Environmental Testing Research Institute, China. E-mail: jiangshixin@ceprei.com

Guojian Nie

China Electronic Product Reliability and Environmental Testing Research Institute, China. E-mail: niegj@ceprei.com

Abstracts: Rolling bearing, a widely used core component in industry, will bring a serious threat to the safety of the machine and staff when it fails. At present, the time-varying operating conditions and catastrophic forgetting have brought great challenges to bearing fault diagnosis. One of the reasons is that good performance can only be maintained if the model is kept under the same conditions as the offline training phase. If the model is directly trained by using the data acquired from new operating condition, the model will suffer from catastrophic forgetting, resulting in poor performance of previous operating condition. In order to solve the above problems, a bearing fault diagnosis method based on lifelong learning is proposed in this paper, which is implemented based on Residual Network with Convolutional Block Attention Module(Res-CBAM) and Elastic Weight Consolidation (EWC). As the basic fault diagnosis model, Res-CBAM can adaptively extract fault features. The introduction of elastic weight consolidation can make the model retain the feature extraction ability of the past condition when learning the fault features of the new condition, so as to solve the catastrophic forgetting problem. The experimental results show that the proposed method has good performance in fault diagnosis under cross conditions.

Keywords: Bearing Fault Diagnosis, Lifelong learning, Elastic Weight Consolidation

1. Introduction

Rolling bearing is the core component of rotating machinery and has widely usage in industry (Cui L.L. et al. 2019). However, the bearing will be damaged easily due to the influence of complex working condition (Zhang W.L. et al. 2020). Once the rolling bearing fails, a serious threat to the safety of the machine and staff will occur. Therefore, it is of great significance and necessary to detect faults in time (Feng, J. 2021, Xing, S.B. 2022, Shao, H.D. 2021).

Due to the advantages of adaptive feature extraction and end-to-end learning, deep learning is widely used in the field of rolling bearing fault diagnosis. The most of proposed fault diagnosis methods based on deep learning (Kamat 2021, Xu, Y.2021, Peng, B. 2022) are trained under specific conditions, which means their generalization performance is insufficient, resulting in good performance can only be maintained if the model is kept under the same conditions as the offline training phase. Nevertheless, in practical applications, the continuous operations of mechanical equipment generate industrial streaming data, which consist of different operating conditions. In addition, if the model is directly trained by using the data acquired from new operating condition, the model will suffer from catastrophic forgetting, resulting in poor performance of previous operating condition. This brings great challenges to the application of fault diagnosis models. Therefore, it is necessary to find an incremental learning method to address this issue.

In order to solve the catastrophic forgetting, most studies resort to data replay, which moderately allows previous data rehearsal to consolidate old knowledge during the model updating. Russell (2024) proposed an adaptive online condition monitoring framework for machinery fault diagnosis, where the mixed-up enhanced data replay is introduced to mitigate the forgetting of old task knowledge. Chen (2024) introduced a generative rehearsal strategy to assist model updating when encountering online learning issues. Compared with direct data rehearsal, the generative-based strategy effectively addresses the class imbalance that occurred in the fault diagnosis task. The methods mentioned above are helpful for reducing catastrophic forgetting. However, the strategy of data replay requires additional space to store the data acquired from past operating condition. With the increase of operating conditions, the storing space occupied by storing data is unacceptable.

To solve the above problems, a lifelong learning method based on residual network with convolutional block attention module(Res-CBAM) and elastic weight consolidation is proposed in this paper. The Res-CBAM model is used as a usual fault diagnosis model. And the elastic weight consolidation is used as the model updating mechanism, which makes the model not forget the knowledge of past conditions when learning the fault diagnosis knowledge of new conditions.

2. Proposed Method

As is shown in Fig. 1,The bearing fault diagnosis based on lifelong learning under cross operating conditions is proposed. The method is consist of convolutional block attention module(Res-CBAM) and elastic weight consolidation. The Res-CBAM model is used as a usual fault diagnosis model. When new operating conditions come, elastic weight consolidation is used in the model updating. Compared with directly training, the usage of elastic weight consolidation makes the model not forget the knowledge of past conditions when learning the fault diagnosis knowledge of new conditions.

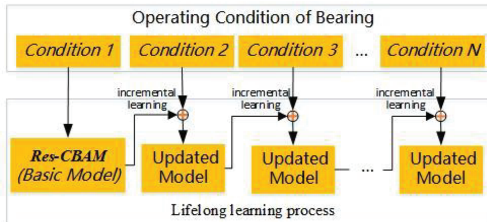

Fig. 1. Bearing fault diagnosis based on lifelong learning under cross operating conditions

2.1. *Res-CBAM*

Residual network(ResNet) (He, K., 2016) is a kind of artificial neural network widely utilized in the field of feature learning, especially in the field of image recognition and object detection. However, vibration data contains a lot of information that is not related to the fault. If only the residual network is used, bearing fault features cannot be accurately extracted. Therefore, convolutional block attention module(CBAM) is introduced into the ResNet to improve the ability of fault feature extraction. The architecture of Res-CBAM model is shown in Fig. 2.

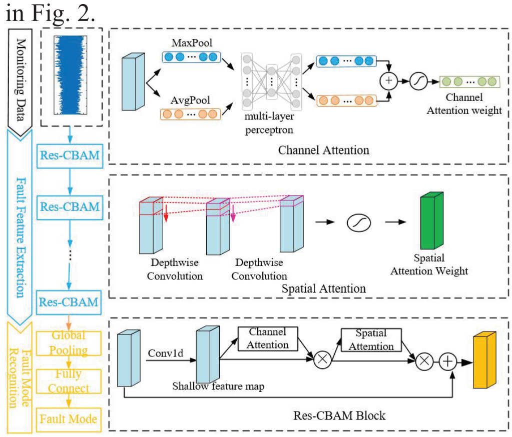

Fig. 2. The architecture of Res-CBAM model

The proposed model consists of two sub-network, fault feature extraction sub-network and fault mode recognition sub-network. The fault feature extraction sub-network is built by stacking Res-CBAM block. Each CBAM block consist of backbone and residual connection. The backbone is used to extract fault feature and the residual connection is used to reduce gradient explosion/vanish. The fault mode recognition sub-network consists of global average pooling and fully connection, which is used to establish mapping relationship between fault feature and fault mode.

In the backbone of Res-CBAM block, input vibration sequence data is processed by convolution operation to acquire shallow feature map. In one channel of feature map, there may be some bearing fault features such as periodic shocks or impulses. However, it is noted that the distribution of these features is discrete along the time axis and thus they merely can be found in some local locations of the channel of some of the feature map. Therefore, channel attention and spatial are introduced. These two attention mechanisms enable model to effectively learn “which” and “where” to attend in the channel and spatial (time axis) dimensions, facilitating the information flow within the network and enhancing the ability of fault feature extraction of the network (Wang, B., 2020). Fig. 2 shows details of these two attention mechanism.

The channel-wise attention is built by modeling the interrelationships between channels. Therefore, the global average pooling (GAP) and the global max pooling (GMP) are firstly used to aggregate the global information of each channel, generating two different channel descriptors. Then these two descriptors are forwarded to a multi-layer perceptron (MLP) with one hidden layer to capture the interchannel relationships and estimate the informativeness of every channel, respectively. After that, the outputs of two MLPs are merged by using element-wise summation and the channel attention weight can be got after sigmoid function. The calculation of channel attention weight can be described as:

$$W_{channel} = Sigmoid(MLP(GMP(input)) \oplus MLP(GAP(input))) \quad (1)$$

Finally, the channel-refined feature maps can be got by conducting channel-wise multiplication between the shallow feature map and the channel-wise attention weights.

The temporal attention captures informative locations by encoding the contextual relationships of each channel. Two Depthwise dilated convolutions layers are first employed to convolve the channel-refined feature maps, independently mapping the context of each channel. Then the hard sigmoid function is adopted to implement the nonlinear activation, resulting in the spatial attention weights. Finally, by performing element-wise multiplication between channel-refined feature maps and spatial weight, the refined spatial feature maps are got. The spatial weight can be calculated by:

$$W_{spatial} = Sigmoid(DConv_2(DConv_1(input))) \quad (2)$$

2.2. *Elastic Weight Consolidation*

In general, only some of the parameters in the neural network play a positive role in the prediction task but not all of them. Traditional training methods of networks change the important parameters which are important for previously task, resulting in catastrophic forgetting. In order to solve this problem, it is important to find a new training method to limit the update of these important parameters, such as elastic weight consolidation. Elastic weight consolidation (EWC) is a method to address the issue of neural networks forgetting previously learned knowledge when learning new tasks. The principle explanation of it is shown in Fig. 3.

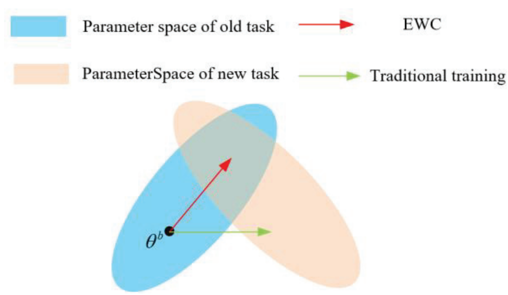

Fig. 3. The principle explanation of EWC

As is shown in Fig. 3, the blue area indicates the parameter space where make the model has good performance on previously task and the orange area indicates the parameter space where make the model has good performance on new task. EWC makes the model of the old task is adjusted towards the intersection space of old and new knowledge by introducing the penalty term, rather than towards the space of new knowledge only like traditional training. The loss function for training the new task with EWC is expressed as:

$$L_{new}(\theta) = L_{new}(\theta) + L_{old} \quad (3)$$

Where L_{old} is the penalty term mentioned before, and $L_{new}(\theta)$ is the loss function of new task. L_{old} can be calculated by:

$$L_{old} = \frac{\lambda}{2}\sum_i F_i(\theta_i - \theta^*_{A,i})^2 \quad (4)$$

Where θ_i represents the network parameters of the new task, $\theta^*_{A,i}$ represents the optimal network parameters of the old task. F_i represents the diagonal elements of the Fisher information matrix, which evaluates the importance of network parameters in the old task. λ is the regularization parameter to balance the importance of losses between new task and old task. F_i can be calculated by:

$$F_i = \frac{1}{|D|}\sum_{d \in D} \frac{\partial L(d,\theta)^2}{\theta^2} \quad (5)$$

3. *Case Study*

3.1. *Data Description*

The bearing fault datasets from the chair of design and drive technology, Paderborn University (Lessmeier C., 2016) are employed to verify the effectiveness of the proposed method. As shown in Fig. 4, the bearing test rig consists of five parts: 1) electric motor, 2) torque-measurement shaft, 3) rolling bearing module, 4) test flywheel, and 5) load motor. The vibration signals are acquired with a sampling rate of 64 kHz from acceleration sensors mounted on SKF6023 bearings. The operating parameters of bearing is listed as Tab. 1.

As is shown in Tab. 1, the bearing fault dataset includes three operating conditions. The detailed description of the operating conditions is as follows.

Operating condition 1(C1): Rotational speed 900 rpm, Load Torque 0.7 Nm, Radial force 1000 N.

Operating condition 2(C2): Rotational speed 1500 rpm, Load Torque 0.7 Nm, Radial force 1000 N.

Operating condition 3(C3):Rotational speed 1500 rpm, Load Torque 0.7 Nm, Radial force 400 N.

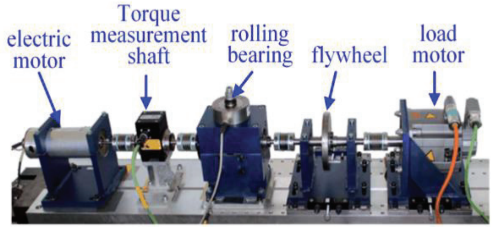

Fig. 4. Bearing test rig

Tab. 1. Operating parameters of bearing.

No.	Rotation-al speed (rpm)	Load Torque (Nm)	Radial force (N)	Categ-ories	Labe-ls	Sam-ples
C1	900	0.7	1000	N/IR/OR	0/1/2	150
C2	1500	0.7	1000	N/IR/OR	0/1/2	150
C3	1500	0.7	400	N/IR/OR	0/1/2	150

3.2. *Experimental Verification*

Three experiments are done to verify the performance of the proposed method. The first experiment is fault diagnosis under signal condition. The second experiment is fault diagnosis under cross condition without lifelong learning. The third experiment is fault diagnosis under cross condition with lifelong learning.

3.2.1. *Fault Diagnosis Under Signal Condition*

In this experiment, three Res-CBAM models proposed in Section 2.1 is constructed as the baseline fault diagnosis model. The data from different conditions is used to train the model separately, which means three baseline fault diagnosis models of different condition are acquired. These three models are denoted as *Model1*, *Model2* and *Model3*. The training and test datasets are split in the ratio 8:2. Cross entropy is used as the loss function of the network and Adam optimizer with a mini-batch size of 15 is used to update its weights and biases. The accuracy of the different condition is shown in Tab. 2.

Tab. 2. Accuracy of the diagnosis under signal condition

		Training dataset		
		C1	C2	C3
Accu -racy	C1	1.000	0.607	0.526
	C2	0.707	0.993	0.887
	C3	0.633	0.813	1.000

As is shown in Table 2. all the models have good performance on the test dataset corresponding to the conditions of their training dataset, which means the proposed model can effectively extract fault features of different fault modes. However, the performance on other condition is unsatisfactory, which causes by the different distribution of different condition. Specially, the performance on condition 3 of *Model2* and the performance on condition 2 of *Model3* better than them on condition 1. The main reason for this phenomenon is that the difference between conditions 2 and condition 3 is the radial forces, and the difference between condition 1 and others is rotational speed. The rotational speed has greater impression than the radial force on the bearing fault features, because the change of rotational speed directly leads to the change of fault characteristic frequency of vibration data. but the radial force does not.

3.2.2. *Fault Diagnosis Under Cross Condition Without Lifelong Learning*

In this experiment, the baseline models acquired from the first experiment are used for fault diagnosis under cross condition. These models are trained through the training dataset of other condition in order. After each training, the diagnosis accuracy is evaluated by the test dataset of current condition and past condition. Cross entropy and Adam are used as loss function and optimizer of the training process, respectively. The diagnosis accuracy of *Model1* is listed as Tab. 3.

Tab. 3. Diagnosis accuracy of *Model1* without lifelong learning

		Task Sequence		
		C1	C2	C3
Accu -racy	C1	1.000	0.567	0.427
	C2	-	0.993	0.907
	C3	-	-	1.000

As is shown in Table 3, the model has good performance on current condition. However, performance on past condition is unsatisfactory, which means the model suffers from serious catastrophic forgetting. Obviously, such model performance cannot be applied in real industrial scenarios.

3.2.3. *Fault Diagnosis Under Cross Condition With Lifelong Learning*

In this experiment, the baseline models acquired from the first experiment are used for fault diagnosis under cross condition. Different from section 3.2.2, these models are trained through the training dataset of other condition with EWC in order and the diagnosis accuracy is evaluated in all condition.As is shown in Table 4, the model has good performance on both condition 1 and condition 2 after training on condition 2. After training on condition 3, although the accuracy of condition 1 drops to 0.8, it still higher than the accuracy without lifelong learning, which means EWC can effectively reduce catastrophic forgetting.

Tab. 4. Diagnosis accuracy of Model1with lifelong learning

		Task Sequence		
		C1	C2	C3
Accu -racy	C1	1.000	0.927	0.800
	C2	0.707	0.927	0.907
	C3	0.633	0.799	0.980

4. Conclusion

A lifelong learning method based on residual network with convolutional block attention module(Res-CBAM) and elastic weight consolidation is proposed in this paper. It is shows by the experiments that the proposed method has good performance in bearing fault diagnosis under cross operating conditions. The introduction of EWC obviously reduces catastrophic forgetting.

Although the proposed method achieves a good fault diagnosis accuracy, the case of increasing fault types does not be considered, which becomes the future research direction.

Acknowledgement

This research was funded by the Key Research and Development Projects in Guangdong Province (grant number 2021B0101220005). This research was funded by the Key laboratory of Industrial Equipment Quality Big Data, MIIT. This research was funded by Development Projects of China Electronic Product Reliability and Environmental Test Institute (NO.24Z03).

References

Cui, L.L. Jin, Z. Huang, J.F. and Wang, H. (2019) Fault Severity Classification and Size Estimation for Ball Bearings Based on Vibration Mechanism. *IEEE Access 7*, 56107–56116.

Zhang, W.L. Xiang, J. Ma, X.D. Luo, H. and Xu, Z.L. (2020) Machinery fault diagnosis with imbalanced data using deep generative adversarial networks, *Measurement 152*,107377.

Feng, J. Yao, Y. Lu, S.X. and Liu, Y. (2021) Domain Knowledge-Based Deep-Broad Learning Framework for Fault Diagnosis. *IEEE Transactions on Industrial Electronics 68(4),* 3454–3464.

Xing, S.B. Lei, Y.G. Wang, S.H. Lu, N. and Li, N.P. (2022) A label description space embedded model for zero-shot intelligent diagnosis of mechanical compound faults. *Mechanical Systems and Signal Processing 162, 108036.*

Shao, H.D. Xia, M. Han G.J. Zhang, Y. and Wan, J.F. (2021) Intelligent Fault Diagnosis of Rotor-Bearing System Under Varying Working Conditions With Modified Transfer Convolutional Neural Network and Thermal Images. *IEEE Transactions on Industrial Informatics 17*, 3488–3496.

Kamat Pooja, Rekha Sugandhi and Satish Kumar.(2021) Data-driven bearing fault detection using hybrid autoencoder-LSTM deep learning approach. *International Journal of Modelling, Identification and Control 38(1),* 88-103.

Xu, Y. Cai W. Wang, L. and Xie, T. (2021) Intelligent Diagnosis of Rolling Bearing Fault Based on Improved Convolutional Neural Network and LightGBM. *Shock and Vibration* , 1205473.

Peng, B. Xia, H. Lv, X. et al. (2022) An intelligent fault diagnosis method for rotating machinery based on data fusion and deep residual neural network. *Applied Intelligence 52(3),* 3051-3065.

Russell, M. Wang, P. Liu, S. and I. S. Jawahir, (2024) Mixed-up experience replay for adaptive online condition monitoring. *IEEE Transactions on Industrial Electronics 71*, 1979-1986.

Chen, S. Ge, W. Liang, X. Jin, X. and Du, Z. (2024) Lifelong learning with deep conditional generative replay for dynamic and adaptive modelling towards net zero emissions target in building energy system. *Applied Energy 353, 122189*.

He, K. Zhang, X. Ren, S. and Sun, J..(2016) Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.

Wang, B. Lei, Y. Li, N. and Wang, W. (2020) Multiscale Convolutional Attention Network for Predicting Remaining Useful Life of Machinery. *IEEE Transactions on Industrial Electronics 68(2020)*, 7496-7504.

Lessmeier C., Kimotho J.K., Zimmer D. and Sextro W. (2016) Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In: PHM Society European Conference. Bilbao.