Cross-Domain Machinery Fault Diagnosis Using Adversarial Network with Conditional Alignments

Nan-Xi Xu
College of Sciences
Northeastern University
Shenyang, China
1800279@stu.neu.edu.cn

Xiang Li
College of Sciences, and Key Laboratory of Vibration and Control of Aero-Propulsion System Ministry of Education
Northeastern University
Shenyang, China
xiangli@mail.neu.edu.cn

Abstract—**The rolling element bearing is essential for rotating machines in real industries. The rolling element bearing faults are supposed to be detected and identified in time to reduce maintenance cost and enhance reliability. The traditional way of fault diagnosis depends on experts' knowledge and experience on the collected machinery vibration data. However, it is becoming more and more difficult to diagnose the health states of the complicated modern machines Using the conventional methods. Currently, with the rapid development of data-driven methods such as deep learning, more accurate diagnosis results have been obtained. While the data-driven methods have been widely used in the basic tasks where the concerned data are from the same distribution, the cross-domain diagnostic tasks are still challenging where distribution discrepancy exists between two different kinds of data. In this paper, a novel network for diagnostics is designed based on domain adversarial neural network, where the diagnostic knowledge learned from the domain for source data is transferred to the domain for the target data. Conditional domain adaptation is proposed to improve data alignments between domains in the learned sub-space. The experimental results on the popular CWRU dataset validate the effectiveness and efficiency of the cross-domain fault diagnosis network in this paper.**

Keywords-fault diagnosis; adversarial training; deep learning; transfer learning

I. INTRODUCTION

Nowadays, with the growth of the industrial scale, the rotating equipments in mechanical systems, aerospace industries *etc.* are becoming more and more complex. The operating performance of the machine is closely interrelated to the healthy state of the rolling element bearing [1]. The rolling element bearing faults should be detected in time so that the machine can operate in healthy state. Traditional methods such as wavelet analysis have been less effective in diagnosing faults of complex machine systems [2]. On the other hand, deep learning is promising to be used to solve those problems which is an emerging data-driven method. During recent decades, deep learning has great success in speech processing, image identification and text processing *etc*. Compared with fault diagnosis Using traditional methods, deep learning has powerful functions of extracting features from large amounts of data, and it does not rely much on expert knowledge [3]. By using deep learning, complex signals can be straightly interrelated to the health state of the rolling element bearing, thus facilitating industrial applications. Furthermore, deep learning is also promising to address the transfer learning problem, where the data used to train and data used to test come from different distributions. For instance, in fault diagnosis problems, the training and testing data can be under different rotating speeds, which is a challenging issue for data-driven methods [4]. In order to realize this idea, the methods based on transfer learning are widely proposed, and a considerable part of the studies are about machine fault diagnosis. A method for cross-domain fault diagnosis using deep generative neural networks was proposed in [5]. Pan *et al.* [6] proposed a feature-based domain adaptation method with neural networks. Li *et al.* [7] designed a neural network for the rolling element bearing fault diagnosis Using attention mechanism to solve the transfer learning problem.

In this study, a new neural network is designed for cross-domain fault diagnosis, where adversarial training is used to bridge the gaps between the target domain and the source domain. Conditional alignments between the training data and the testing data are employed to improve the model performance on extracting domain-invariant features.

In Section II, this paper introduces the proposed neural network. In Section III, experimental details are stated in detail, The conclusions are made in Section IV.

II. THE PROPOSED METHOD

A. Transfer Learning

The knowledge learned from one dataset can be applied to another dataset by transfer learning once the training data and the testing data are subject to different distributions. In the field of fault diagnosis, domain adaptation methods have been developed to accomplish the task of diagnosing cross-domain faults with the same label space. Thus, general diagnostic knowledge is extracted under different states of the same machine. Through the previous researches [8], general diagnostic knowledge is extracted by optimizing the data distribution. Gradually, deep learning has made great progress in extracting advanced features, which made transfer learning more efficient [9].

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

This paper aims to build an intelligent cross-domain fault diagnosis model Using sufficient supervised data in the source domain and insufficient supervised data in the target domain. The limited target-domain data fail to train an efficient target-domain fault diagnosis classifier on its own.

B. Neural Networks forDomain Adversarial Training

Domain adversarial training of neural networks has been extensively used in the recent years to solve the puzzle of domain adaptation [10]. Usually, classifier, feature extractor and domain discriminator make up domain adversarial neural network. They are represented by C, G, D and their parameters are θ_C , θ_G , θ_D respectively [11].

All the data is firstly fed into the feature extractor, then features picked up from the high level layers are sent into the classifier and the domain classifier. The extractor for features and the discriminator for domain formed a confrontational relationship with each other [12].

C. Proposed Network

In this paper, the task of cross-domain fault diagnosis on the rolling element bearing Using adversarial neural network with conditional alignments is investigated and the architecture is showed in Fig. 1. Firstly, vibration data x without pre-processing in the domain for source data and the domain for target data are directly fed into the proposed neural network, which requires little knowledge about fault diagnosis and signal preprocessing and thus facilitates industrial application. Samples of data pass through the feature extractor, and the advanced features are acquired. The feature extractor is constituted by one fully-connected layer and two convolutional layers. Next, multiple domain discriminators including both marginal and conditional ones are further adopted. The high-level representations coming from the feature extractor are regarded as the inputs of the next layer. The marginal domain discriminator aims to achieve fusion of marginal distributions, and the N_c conditional discriminators are used for conditional alignments of different health states. In each domain discriminator, one convolutional layer, a softmax function and two fully-connected layers are adopted. The classifier module is used for health condition classification, which consists of one convolutional layer, a softmax function and two layers which are fully-connected.

The ultimate goal of the proposed method is to use existing information to achieve high accuracy for fault diagnosis. Therefore, the primary goal to training the neural network is,

$$L_c = -[\frac{\alpha^s}{n_s}\sum_{i=1}^{n_s}\sum_{k=1}^{N_s} 1\{y_i^s = k\}\log\frac{e^{\eta_{s,k}^T x_{h,i}^s}}{\sum_{j=1}^{N_s} e^{\eta_{s,j}^T x_{h,i}^S}}]$$
$$-\frac{\alpha^t}{n_{lab}^t}\left[\sum_{i=1}^{n_{lab}^t}\sum_{k=1}^{N_t} 1\{y_i^t = k\}\log\frac{e^{\eta_{t,k}^T x_{h,i}^t}}{\sum_{j=1}^{N^t} e^{\eta_{t,j}^T x_{h,i}^t}}\right] \quad (1)$$

where L_c represents the cross-entropy loss function for classification of adequate supervised data in the source domain and weakly supervised data in the target domain. $x_{h,i}^s$ and $x_{h,i}^t$ are the high-level representations by the feature extractor of the samples in the two domains where the data not the same but related, respectively. y_i^s and y_i^t denote the corresponding rolling element bearing health condition labels. n_s represents the quantity of the samples which are labeled in the source domain, and n_{lab}^t is the quantity of the samples which are labeled in the target domain. N_s and N_t represent the quantities of the states of health conditions for rolling element bearing in the two domains where the data not the same but related, respectively. In this study, it is assumed $N_s = N_t$, $\eta_{s,k}$ and $\eta_{t,k}$ are the softmax parameters of the classifier for the two domains where the data not the same but related, respectively. α^s and α^t denote the penalty coefficients for the two domains where the data not the same but related. d_i represents the domain label. N_c denotes the numbers of conditional discriminators.

Equation (1) utilizes all the supervised data in both the source and target domains, and minimization of (1) makes sure that distinguishable features extracted from high-level representations denote different health states. Because of the difference in distributions, features extracted from the two domains may not be related to each other, which indicates that the knowledge of fault diagnosis learned from the source domain may not be applied to the classification of the target domain. Therefore, the domain adversarial training network is designed for this tricky problem. Representations $x_h = G(x)$ extracted from the feature extractor are inputs, multiple domain discriminators D_i are adopted, including both the marginal and conditional discriminators. The marginal discriminator aims to distinguish the two domains with respect to the marginal data distribution. Multiple conditional discriminators aim to distinguish domains for specific fault types respectively. The marginal and conditional domain classification error are defined as,

$$L_d^{marginal} = -\frac{1}{n_{all}}\sum_{i=1}^{n_{all}}\sum_{k=1}^{2} 1\{d_i = k\}\log\frac{e^{\eta_{d,k,m}^T x_{h,i,m}}}{\sum_{j=1}^{2} e^{\eta_{d,j,m}^T x_{h,i,m}}} \quad (2)$$

$$L_d^{conditional} = -\sum_{c=1}^{N_s}\left[\frac{1}{n_c}\sum_{i=1}^{n_c}\sum_{k=1}^{2} 1\{d_i = k\}\log\frac{e^{\eta_{d,k,c}^T x_{h,i,c}}}{\sum_{j=1}^{2} e^{\eta_{d,j,c}^T x_{h,i,c}}}\right] \quad (3)$$

where n_{all} represents the number of all the samples including samples in the two domains where the data not the same but related, and n_c represents the quantity of the samples in the *c*-th health state. $x_{h,i,m}$ is the learned high-level features in the marginal discriminator, and $x_{h,i,c}$ is the high-level representation in the *c*-th conditional discriminator for the samples belonging to the *c*-th class. *d* denotes the domain label. It is supposed that d_i=1 represents the source domain and

d_i=2 denotes the target domain. $\eta_{d,k}$ denotes the softmax parameters. The loss of domain classification is defined as follow,

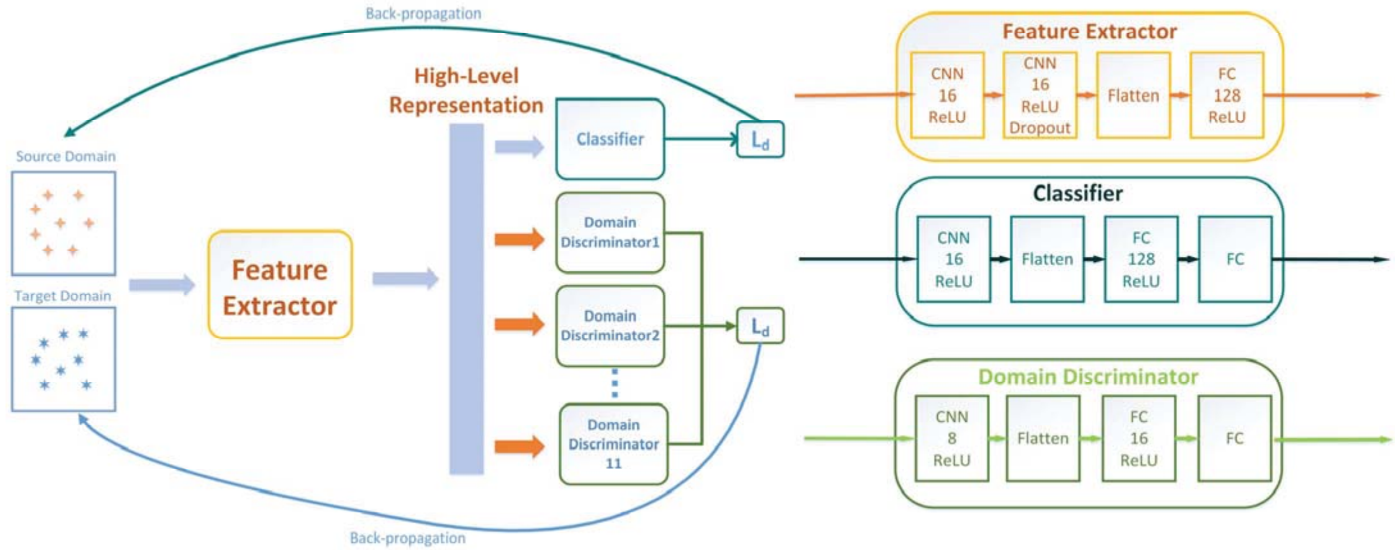

Figure 1 Proposed adversarial network architecture. GRL is a gradient reversal layer, which turns the symbol of the gradient round the in backpropagation.

$$L_d = \sum_{i=1}^{n_c} \alpha_i L_d^{conditional} + \alpha_m L_d^{marginal} \tag{4}$$

α_i and α_m denote the penalty for the conditional loss and the marginal loss respectively. The overall goal of the proposed method is,

$$L_{obj}\left(\theta_G, \theta_D, \theta_C\right) = L_c + \alpha_d L_d \tag{5}$$

$$\left(\hat{\theta}_G, \hat{\theta}_C\right) = \arg\min_{\hat{\theta}_G, \hat{\theta}_C} L_{obj}\left(\theta_G, \hat{\theta}_D, \theta_C\right) \tag{6}$$

$$\hat{\theta}_D = \arg\max_{\theta_D} L_{obj}\left(\hat{\theta}_G, \theta_D, \hat{\theta}_C\right) \tag{7}$$

where α_d is penalty coefficient for L_d .Through utilization of GRL (the gradient reversal layer), the neural network can be quickly optimized Using the prevalent stochastic gradient descent way, and the parameters are updated in the process of training as well.

Table I PARAMETER SET OF THE PROPOSED METHOD

Parameter	Value	Parameter	Value
Epochs	5000	α_d	1
Batch Sizes	16	α_m	10
Learning Rate	0.00001	α_i	1
Dropout Rate	0.5	Convolutional filter size	10

III. STUDY FOR EXPERIMENTAL

A. *Descriptions of CWRU Dataset*

The Bearing Data Center in Case Western Reserve University has accomplished an experiment on fault of rolling element bearings and obtained large amount data [13]. The experiment apparatus is presented in Fig. 2. The dataset from CWRU is public and has been extensively used in the machine fault diagnosis field. Trials were done using a two horsepower Reliance Electric motor. The vibration signal data is collected at two locations that are close to and far away from the bearings on the motor. In this paper, only two rotating speeds are used in all, i.e. 1797 rpm, and 1730 rpm, as well as four rolling element bearing health states, i.e. healthy that is without any faults (H), faults on outer race (O), faults on inner race (I) and faults on the ball (B) are included. There are three diverse fault severities, that are 7, 14 and 21 mils fault diameters, separately. There are ten bearing health states in all.

B. *Implementation Details*

Let $N_{s,train}$ and $N_{t,train}$ represent the quantities of the labeled samples for each health state in the two domains where the data not the same but related, respectively. $N_{t,test}$ denotes the quantity of the samples which are unlabeled for each health state to test in the domain which is the target data. In each domain, the size of mini-batch is 16, which is always used for optimization in neural network gradients.

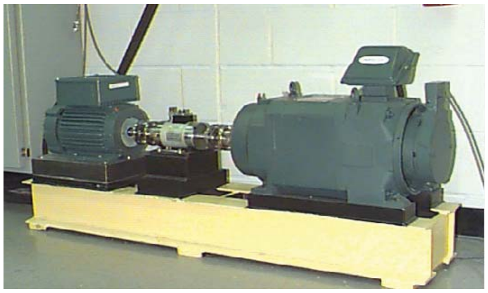

Figure 2 Experiment apparatus in the CWRU dataset.

In the course of neural network training, the initial weights of the neural network and biases are realized by the Xavier normal initializer [14]. The updates of all the parameters are realized by the back-propagation (BP) algorithm, and the Adam optimization method [15] is used. It is assumed $N_{s,train} = 200$ in the experiments. The final experimental result is obtained from the average of three trials. The specific parameters are presented in Table I.

C. Compared Approaches

For the purpose of further validating the effectiveness of the proposed neural network, three networks are applied to compare, which are using similar experimental setup and neural network with the neural network in this paper.

1) Baseline

First of all, a simple basic deep neural network is brought about, where only the labeled data in the domain composed of source data is adopted in the neural network in the process of training, and any transfer learning is not used.

2) TargetOnly

In the process of the model training, only the labeled data in the domain composed of target data is used and any transfer learning is not used. Through this method called TargetOnly, the efficient of the domain adversarial training program can be validate.

3) NoCondi

This study examines whether conditional distribution ameliorate the generalization of the neural network in the target domain. Thus, the NoCondi neural network is trained to compare with the existing neural network, where any conditional distribution is not used.

D. Results for Experimental

In this part, the great potential of the proposed neural network is presented. In Table II, it can be obviously known that the accuracy on average of the proposed method is up to 94.58% and the variance is 0.69, which performs best among all the methods mentioned in this paper. Baseline attains the lowest accuracy of 44.22 % and the variance is 2.33, because it cannot meet the requirement that reduces the distribution discrepancy of features come from diverse domains. TargetOnly attains the average accuracy of 78.30% with the variance of 8.73, which performs better than Baseline. The NoCondi method attains the average accuracy of 43.94% and the variance is 2.02.

In Fig. 3, it can be seen that the neural network in this paper is obviously superior to other three methods in most cases. While compared with the Baseline and No-Condi approaches, the TargetOnly method also achieves prospective results, but its performance is not the most efficient and effect in different scenarios of different tasks. In Fig. 4, it can be seen that the more training samples generally leads to higher testing accuracy.

Table II. TESTING ACCURACY FOR DIFFERENT METHODS

Method	Average accuracy	Variance
Proposed	94.58 %	0.69
Baseline	44.22 %	2.33
TargetOnly	78.30 %	8.73
No-Condi	43.94%	2.02

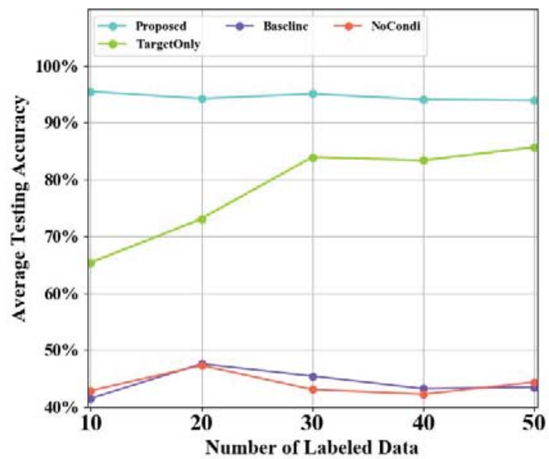

Figure 3 Diagnostic performances of different labeled data numbers in the target domain.

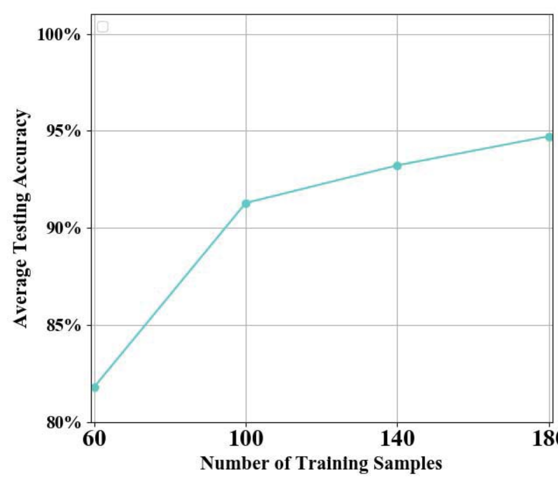

Figure 4 Diagnostic performances of different numbers of training samples in the source domain.

IV. CONCLUSION

In this paper, based on transfer learning, a novel neural network is designed for the fault diagnosis of rolling element bearing. Raw vibration data are directly fed into the neural network, facilitating industrial applications. Adequate labeled data in the source domain is used for the training of the proposed neural network. The learned knowledge could applied for the data in the target domain which could not train an efficient neural network independently. In this neural network, advanced features are obtained in the training process of unprocessed measured vibration signals, which are diagnosis results regard to several health states for two different rotating speeds. Domain adversarial training is brought about to reduce the distribution discrepancy of the learned features between the data from the source domain and the target domain.

Experiments are done using data under two different rotating speeds, including ten health cases. This proposed method contributes to existing knowledge of fault diagnosis by providing a novel neural network.

It should be noted that the generalisability of these results is subject to certain limitations. For instance, small amount of labeled data is required in the target domain. The study about transfer learning only using source-domain labeled data could be an understudied and important potential next frontier, which is more in line with actual industrial scenarios.

REFERENCES

[1] Y. Lecun, Y. Bengio, G. Hinton. "Deep learning," Nature, 2015, 521(7553):436.

[2] S. J. Pan, Y. Qiang. "A Survey on Transfer Learning," IEEE Transactions on Knowledge & Data Engineering, 2010, 22(10):1345-1359.

[3] M. Long , Y. Cao , J. Wang. "Learning Transferable Features with Deep Adaptation Networks," 2015.

[4] B. Gong, K. Grauman, F. Sha . "Connecting the Dots with Landmarks: Discriminatively Learning Domain-Invariant Features for Unsupervised Domain Adaptation," International Conference 7on International Conference on Machine Learning. JMLR.org, 2014.

[5] X. Li, W. Zhang, Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," IEEE Transactions on Industrial Electronics, 2019, 66 (7): 5525 - 5534.

[6] S. J. Pan , I. W. Tsang , J. T. Kwok. "Domain Adaptation via Transfer Component Analysis," IEEE Transactions on Neural Networks, 2011, 22(2):199-210.

[7] X. Li, W. Zhang, Q. Ding, "Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism," Signal Processing, 2019, 161: 136-154.

[8] Z. Wei, P. Gaoliang, L. Chuanhao. "A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals," Sensors, 2017, 17(2):425.

[9] Y. Ganin , V. Lempitsky, "Unsupervised Domain Adaptation by Backpropagation," 2014.

[10] L. Guo, Y. Lei, S. Xing, T. Yan and N. Li, "Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data," in IEEE Transactions on Industrial Electronics. 2018

[11] B. Yang, Y. Lei, F. Jia, S. Xing. "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," Mechanical Systems and Signal Processing, 2019, 122.

[12] X. Li, W. Zhang, Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," Neurocomputing, 2018, 310, 77-95.

[13] W. A. Smith, R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," Mechanical Systems and Signal Processing, 2015, 64-65:100-131.

[14] L. Zhu, Y. Chen, , A. Yuille, "Unsupervised learning of probabilistic grammar-markov models for object categories," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31, 114–128.

[15] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.