Proceedings of the 42nd Chinese Control Conference
July 24-26, 2023, Tianjin, China

Sparse Filtering with Joint Distribution Adaptation for Intelligent Fault Diagnosis

Pengtao Song[1], Qingyu Yang[1,2], Donghe Li[1], Hongtao Dang[1,3]

1. School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, 710049, P. R. China
E-mail: songpengtao@stu.xjtu.edu.cn; lidonghe2020@xjtu.edu.cn

2. State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an 710049, P. R. China
E-mail: yangqingyu@mail.xjtu.edu.cn

3. Rockwell Automation (China) Company Limited Xi'an Branch, Xi'an 710049, P. R. China
E-mail: hdang@rockwellautomation.com

Abstract: Most existing intelligent fault diagnosis schemes rely on the assumption that the training and test samples are independent and identically distributed, ignoring the domain distribution shift caused by diverse operating conditions, which may limit their flexible applications in practical diagnostic tasks. To address this problem, we propose a novel unsupervised transfer learning method, namely, sparse filtering with joint distribution adaptation (SFJDA) for mechanical fault diagnosis. Specifically, two sparse filters (SFs) are used to jointly extract features from each domain, and the final feature space is formed by stacking the subspaces obtained from double SFs. Then, the maximum mean difference (MMD) is introduced to measure the distribution discrepancy between different domains. By extending the marginal distribution adaptation (MDA) to joint distribution adaptation (JDA), the constructed framework can capture domain-invariant and class-separable features. Finally, the effectiveness of the proposed scheme is verified by a motor bearing dataset.

Key Words: Intelligent fault diagnosis, sparse filtering, transfer learning, domain adaptation, joint distribution adaptation.

1 Introduction

Intelligent fault diagnosis, with the ability to automatically extract features from raw data and identify fault patterns, has been seen as the key to improving the security and stability of industrial equipment [1, 2]. Over the past decade, many machine learning algorithms have been widely applied in the field of fault diagnosis, and have achieved superior diagnostic performance, such as convolutional neural network (CNN) [3, 4], autoencoder (AE) [5, 6] and sparse filtering (SF) [7–9]. Especially for SF, it can achieve the same performance as the deep network with a relatively simple structure and fewer data by constraining the mapped features.

However, most existing methods are mainly based on the strong assumption that the training and test data are independently and identically distributed [10]. In reality, the distributions between training and test data are different in practical situations due to the complex load conditions and working environment [11, 12]. Moreover, to ensure the generalization performance of diagnostic model, a large amount of labeled data is often needed to obtain distinguishing features. Note that, in real industrial sites, the data collected by the sensors is unlabeled raw data, so it is difficult to obtain data with labels for each load condition. These problems greatly restrict the extensive application of intelligent diagnosis schemes in industrial scenarios.

Domain adaptation [13–15], as a representative branch of transfer learning, aims to establish knowledge transfer from the source domain to the target domain by reducing the distribution discrepancy and exploring domain-invariant features. Generally, the basic strategies of domain adaptation can be divided into three types: 1) instance-based transfer aims to reweight labeled data to assist model training [13, 16]. 2) feature-based transfer reduces distribution differences through feature transformations [17, 18]. 3) model-based transfer identifies the fault patterns of target domain by sharing network parameters [14, 19]. Among them, feature-based transfer and model-based transfer are mainly applied in the learning problems with limited labeled samples. For example, Lu *et al.* [10] proposed a deep neural network for fault diagnosis, where the maximum mean discrepancy (MMD) was used to align the feature distribution. Han *et al.* [12] extended the marginal distribution adaptation to joint distribution adaptation with deep transfer networks and obtained a more accurate distribution matching. Jiao *et al.* [14] designed an adversarial adaptation network based on classifier discrepancy to learn domain-invariant features.

Deep transfer networks, by stacking multi-layer networks and transforming features, have achieved superior transfer performance in fault diagnosis [15]. However, deep networks often have numerous parameters to be optimized, which require long training times and large computing power. Faced with these challenges, how to use fewer layers to transfer knowledge and enhance the flexibility of diagnostic frameworks for diverse domain tasks is still a topic worth exploring, which is also the motivation of this paper.

To summarize the discussions made so far, this paper proposes a novel unsupervised method named sparse filtering with joint distribution adaptation (SFJDA) for intelligent diagnosis at the scenario of domain distribution shift. The main contributions are as follows:

1) Considering the distribution discrepancies between different domains, two sparse filters are constructed to jointly extract features, and the final feature space is formed by stacking the subspaces obtained from double SFs. Therefore, the proposed framework can mine rich information with a simple structure.
2) A strategy called sparse filtering with marginal dis-

This work is supported by Shaanxi Logistics Group – Logistics Science and Innovation Integration Development Research Center of Xi'an Jiaotong University under Grant SKH2023005.

tribution adaptation (SFMDA) is designed to achieve marginal distribution adaptation, where MMD is introduced to measure differences between domains. The proposed SFMDA can limit the distance between clustering centroids of different domains and thus capture class-separable features.

3) Based on SFMDA, the problem of joint distribution adaptation (JDA) is analyzed, and the final SFJDA network is established. It can extract domain-invariant features and effectively address the problem of performance degradation when using SF to directly diagnose unlabeled target samples.

The validity of these conclusions is finally verified by a motor bearing dataset.

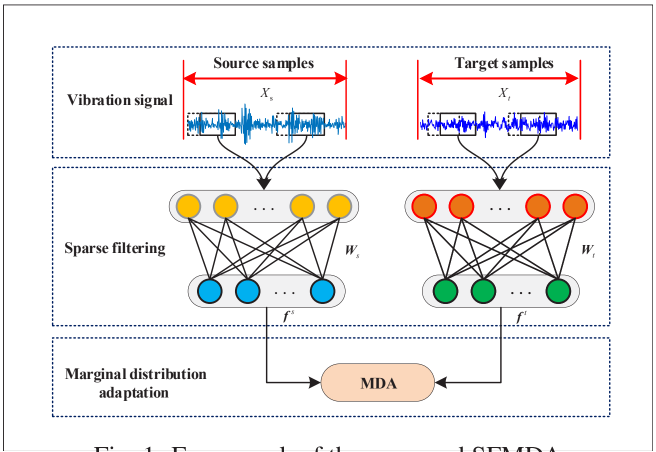

Fig. 1: Framework of the proposed SFMDA.

2 Fault Diagnosis Scheme Construction

2.1 Sparse Filtering with Marginal Distribution Adaptation

Sparse filtering has been proved to be able to extract discriminative features from raw data with a simple structure [9]. However, since the data generated under each load condition contains both fault and operational information, and is subject to distribution differences, diagnostic performance will be degraded when using models trained from source domain to directly predict target domain. Therefore, how to make full use of the information in each domain to obtain domain-invariant features is a key issue.

In this work, the problem of marginal distribution adaptation is analysed first, where two SFs are employed to extract features from the source and target domains respectively. The architecture of the proposed double sparse filtering with marginal distribution adaptation (SFMDA) is depicted in Fig. 1, in which $X_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ and $X_t = \{x_i^t\}_{i=1}^{n_t}$ denote the source domain dataset and target domain dataset. They are drawn from distributions $P_s(X)$ and $P_t(X)$ respectively, and $P_s(X) \neq P_t(X)$ due to the distribution discrepancy. But, the label space in different domains is same, i.e., $Y_s = Y_t$.

Sparse filtering maps the samples onto their features $\boldsymbol{f}_i^s$ and $\boldsymbol{f}_i^t$ using weight matrix $\boldsymbol{W}_s$ and $\boldsymbol{W}_t$ respectively. The features $\boldsymbol{f}_i^s$ and $\boldsymbol{f}_i^t$ are calculated by

$$\boldsymbol{f}_i^s = \sigma(\boldsymbol{W}_s x_i^s),\ \boldsymbol{f}_i^t = \sigma(\boldsymbol{W}_t x_i^t) \tag{1}$$

where $\boldsymbol{W}_s$ and $\boldsymbol{W}_t$ are parameters to be optimized. $\sigma(\cdot)$ is the activation function, and the soft-absolute function $\sigma(t) = \sqrt{t^2 + \varepsilon}$ with $\varepsilon = 10^{-8}$ is commonly recommended [7].

For each $\boldsymbol{f}_i$, including $\boldsymbol{f}_i^s$ and $\boldsymbol{f}_i^t$, it is composed of f_i^j ($j = 1, 2, \dots, N_{out}$, and N_{out} is the output dimension of the feature matrix). All feature values f_i^j form the feature matrix $\boldsymbol{F} = (f_i^j)_{N_{out} \times M}$ (M is the number of samples in datasets, i.e., n_s, n_t), whose i-th column and j-th row are separately denoted as $\boldsymbol{f}_i \in \Re^{1 \times M}$ and $\boldsymbol{f}^j \in \Re^{N_{out} \times 1}$. We first normalize each row of $\boldsymbol{F}$ by its l_2-norm across all samples, i.e., $\tilde{\boldsymbol{f}}^j = \boldsymbol{f}^j / \left\|\boldsymbol{f}^j\right\|_2$. Then, each column is normalized by its l_2-norm across all features

$$\hat{\boldsymbol{f}}_i = \tilde{\boldsymbol{f}}_i / \left\|\tilde{\boldsymbol{f}}_i\right\|_2 \tag{2}$$

Sequentially, the weight parameters can be solved by optimizing the cost function constraining l_1-norm for each sample, which is shown as follows

$$J_{\mathrm{SF}} = \sum_{i=1}^{n_s} \left\|\hat{\boldsymbol{f}}_i^s\right\|_1 + \sum_{i=1}^{n_t} \left\|\hat{\boldsymbol{f}}_i^t\right\|_1 \tag{3}$$

MMD is a criterion to measure the discrepancy of two distributions based on RKHS [20]. In contrast to many parametric criteria, e.g. Kullback–Leibler divergence, MMD allows the estimation of non-parametric distances between various distributions and avoids the calculation of intermediate densities, which is always a nontrivial task [10]. The definition of MMD is given by

$$L_{\mathrm{MMD}}(X^s, X^t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(\boldsymbol{x}_i^s) - \frac{1}{n_t} \phi(\boldsymbol{x}_i^t) \right\|_{\mathcal{H}}^2 \tag{4}$$

From (4), one can see that the empirical estimation of the discrepancy between distributions is considered as a distance of two domains in RKHS. A value of MMD close to zero implies that the two domains are marginal distribution matching. Rather than mapping them to a single space, dual SFs are used to map the samples between different domains to two different but similar spaces, and then apply the MMD term to estimate the distance between the source and target domains, which is expressed as

$$J_{\mathrm{MDA}} = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi_s(\boldsymbol{x}_i^s) - \frac{1}{n_t} \sum_{i=1}^{n_t} \phi_t(\boldsymbol{x}_i^t) \right\|_{\mathcal{H}}^2 \tag{5}$$

where ϕ_s and ϕ_t are feature transformations through SFs, n_s and n_t are the number of samples in source domain and target domain respectively, and $\mathcal{H}$ is the feature space after SFs.

By integrating (3) and (5) together, the final cost function of SFMDA is given by

$$J_{\mathrm{SFMDA}} = J_{\mathrm{SF}} + \lambda J_{\mathrm{MDA}} \tag{6}$$

where J_{SF} is used to extract sparse features, and J_{MDA} is employed to constrain the distance between the centroids of different domain clusters. $\lambda \geq 0$ controls the tradeoff between two terms. By jointly optimising these terms, the designed scheme is able to achieve domain-invariant features.

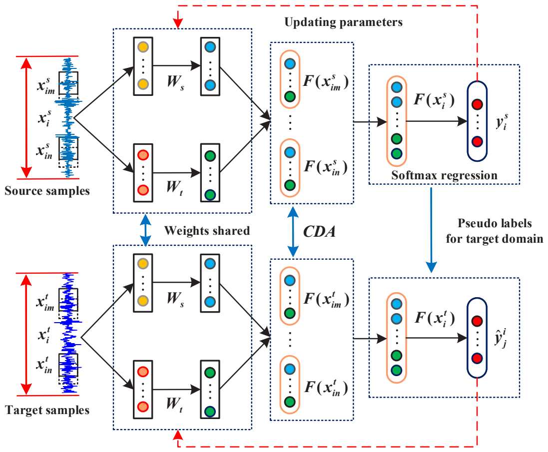

Fig. 2: Overall architecture of the proposed SFJDA for fault classification.

2.2 Sparse Filtering with Joint Distribution Adaptation

To further match the structure of the source and target domain differentiation and improve transfer performance, we introduced conditional distribution adaptation based on SFMDA. The whole architecture of SFJDA is shown in Fig. 2. After obtaining the initial parameters $\boldsymbol{W}_s$ and $\boldsymbol{W}_t$, the source samples are propagated forward through the double SFs to train the classifier, which is used to predict the pseudo labels for target domain with shared weight parameters. Then, the labeled source samples and target samples with pseudo labels are exploited to adjust the network parameters, aiming to reduce the distance between the same class samples in both domains and obtain class-separable features.

The discrepancy between the conditional probability distributions of two domains is defined as

$$D(P(Y^s\,|\mathcal{F}(X^s)),\, P(Y^t\,\left|\mathcal{F}(X^t)\right.)) \tag{7}$$

Based on Bayes' theorem, we rewrite it as

$$D(\frac{Q(\mathcal{F}(X^s)\,|Y^s)\cdot P(Y^s)}{P(\mathcal{F}(X^s))},\frac{Q(\mathcal{F}(X^t)\,|Y^t)\cdot P(Y^t)}{P(\mathcal{F}(X^t))}) \tag{8}$$

where $\mathcal{F}$ is the nonlinear transformation and stacked by two SFs. In this paper, we consider that the source domain and target domain have the same label space and assume $P(Y^s) = P(Y^t)$. When the marginal distribution is fixed, in order to reduce the difference of condition distribution, the optimization problem can be expressed as

$$\min D(Q(\mathcal{F}(X^s)\,|Y^s\,), Q(\mathcal{F}(X^t)\left|Y^t\right.)) \tag{9}$$

Although the target domain samples do not contain label information, we can use the pseudo-labels of target domain to handle the optimization problem (9), which have proved by previous studies that the pseudo labels can achieve a better prediction performance through continuous iterating and updating [12]. As before, the MMD is employed to measure the mismatch of conditional distributions. Then, we have

$$J_{\mathrm{CDA}} = \sum_{c=1}^{C}\left\|\frac{1}{n_s^c}\sum_{x_i^s\in X_s^c}\mathcal{F}(\boldsymbol{x}_i^s) - \frac{1}{n_t^c}\sum_{x_i^t\in X_t^c}\mathcal{F}(\boldsymbol{x}_i^t)\right\|_{\mathcal{H}}^2 \tag{10}$$

Algorithm 1 Learning Algorithm of SFJDA

Input: Datasets $X_s = \{(\boldsymbol{x}_i^s, y_i^s)\}_{i=1}^{n_s}$ and $X_t = \left\{\boldsymbol{x}_i^t\right\}_{i=1}^{n_t}$, two regular parameters λ and β

Output: Transferred network for target domain

1: Train a basic SFMDA network on the datasets X_s and X_t according to (6)
2: Train a classifier with the source dataset X_s and predict the pseudo labels $\hat{Y}_0\left\{y_i^t\right\}_{i=1}^{n_t}$ for target samples
3: **repeat**
4: $j = j+1$
5: Network optimization with respect to (11)
6: Update the classifier and pseudo labels $\hat{Y}_j$ with optimized network
7: **until** convergence or $\hat{Y}_j = \hat{Y}_{j-1}$

where C is the output category. X_s^c and X_t^c are the class c samples in source domain and target domain. n_s^c and n_t^c indicate the corresponding number of samples in each domain, and n_t^c is obtained by the pseudo labels $\hat{y}_c(\boldsymbol{x}_i^t)$. By optimizing (10), the discrepancy between samples of the same class in source and target domains can be reduced, so as to achieve conditional distribution adaptation.

By integrating (5) and (10) , the regularization term of JDA can be written as

$$J_{\mathrm{JDA}} = \lambda J_{\mathrm{MDA}} + \beta J_{\mathrm{CDA}} \tag{11}$$

where λ and β are positive constants to make a trade-off between different terms, and (11) is used to further fine tune $\boldsymbol{W}_s$ and $\boldsymbol{W}_t$ to capture class-separable features.

2.3 Training Strategy

The detailed training process is shown in Algorithm 1. In the first step, a SFMDA network is trained with the samples of each domian, hoping to reduce the discrepancy of marginal distribution between two domains. Secondly, the source samples are propagated forward through the double SFs to obtain the feature matrix, which is used to train the classifier. Then, the trained network and classifier are used to predict pseudo labels for target samples with the shared weight parameters. Finally, the labeled source samples and target samples with pseudo labels are exploited to adjust the network parameters, aiming to match the joint distribution of two domains.

By iterating step 2 and step 3 repeatedly, the network parameters will be updated until convergence. Meanwhile, the final transferred network, including the optimal parameters $\boldsymbol{W}_s$, $\boldsymbol{W}_t$ and softmax regression, can be obtained to diagnosis target samples.

3 Experiments and Results

3.1 Datesets preparation

The vibration signals of the motor bearing are provided by Case Western Reserve University (CWRU) [21], which were collected with 12 kHz sampling frequency under four loads (0, 1, 2 and 3 hp) from the drive-end bearing of the motor. Three different bearing conditions (roller, inner race, and outer race) with three severity levels (0.18, 0.36, and 0.53 mm) and a normal condition, are considered in this dataset, where the same condition at different loads is treated as one class. There are 100 samples for each condition under one

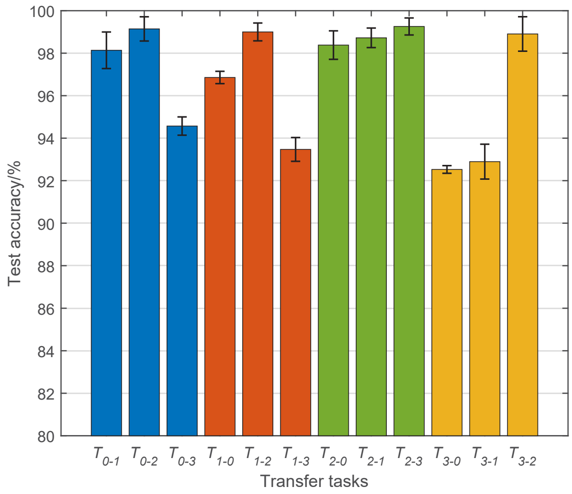

Fig. 3: Transfer accuracy under different tasks.

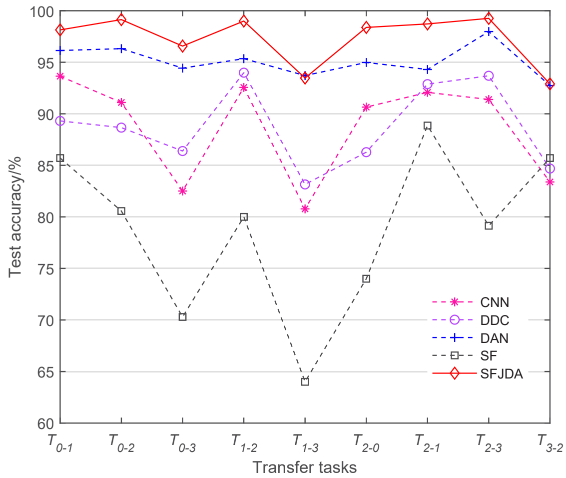

Fig. 5: Transfer accuracy under different methods.

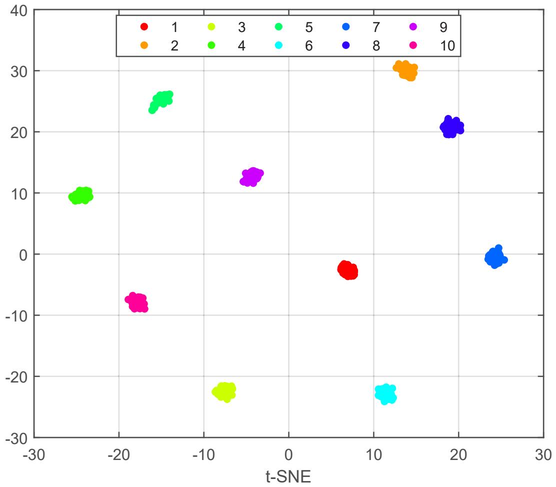

Fig. 4: Feature visualization under T_{1-2}.

load, and each sample contains 1200 data points. Therefore, the dataset totally contains 10 health conditions and 4000 (100 × 4 × 10) samples.

3.2 Experimental Setup

For model training, all weights are randomly initialized, then the network parameters are optimized by minimizing the objective functions (6) and (11). In this paper, L-BFGS algorithm [7] is applied to optimize the network parameters. For the designed SFJDA, the optimal parameters λ and β of (11) are set to 100 and 10 respectively, which are selected by empirically searching from the set of $\{0.01, 0.1, 10, 50, 100, 1000\}$. The output dimension of each SF is $N_{out} = N_{in}/4$ with $N_{in} = 100$, and the softmax regression is empirically fixed to 10^{-5}. In all experiments, 65% of target samples are randomly chosen for training and the remaining samples are used for testing, where each sample is preprocessed by segmenting with the overlap rate of 85% and whitening. To avoid the effects of randomness, all experiments are independently conducted for 10 times. The average classification accuracies and standard deviations on the test dataset are reported.

Table 1: Performance comparision between the proposed SFJDA and other methods

Methods	Transfer tasks	No. of training samples (%)	Testing accuracy (%)
[11]	8	70	91.92±1.30
[23]	9	75	81.20±3.81
[17]	9	75	83.10±3.12
DDC [24]	8	70	88.78±1.48
DAN	8	70	95.11±1.17
DANN [25]	8	70	95.78±1.59
SF [8]	12	65	76.14±3.26
SFMDA	12	65	90.14±2.12
SFJDA	**12**	**65**	**96.82±0.84**

3.3 Results and Analysis

1) *Classification accuracy analysis:* In this section, twelve different transfer tasks are constructed to evaluate the proposed method on motor bearing dataset. Fig. 3 shows the accuracies and standard deviation under different tasks, where T_{i-j} represents the transfer task from load i (source domain) to load j (target domain). As shown in Fig. 3, the transfer accuracy of the proposed method is over 92% in all tasks, and the average test accuracy is above 96%, which reveals that the designed scheme can learn domain-invariant features. In addition, feature visualization is performed to further evaluate the superiority of the extracted features. Specifically, the t-distributed stochastic neighbor embedding (t-SNE) technology [22] is employed to reduce dimensionality of feature representations for visualization. We randomly choose the transfer task from load 1 to load 2, and the clustering results of target domain are displayed in Fig. 4. One can see that the designed SFJDA can better cluster the same categories and separate different categories in the target domain. These results demonstrate the proposed strategy is able to capture domain-invariant and class-separate features.

2) *Comparison with other mainstream methods:* To further demonstrate the effectiveness of the proposed SFJDA, eight popular approaches are selected to handle same datasets for comparison, including full featur layers alignment based on MMD (FMMD) [11], transfer component analysis (TCA) [23], basic convolutional network without

domain adaptation (CNN) [17], deep domain confusion (DDC) [24], deep adaptation network (DAN), domain adversarial neural network (DANN) [25], etc. The experimental settings of these methods are determined according to the related literature. The comparison results are summarized in Table 1. One can see that compared with other schemes, the proposed SFJDA can achieve a better average diagnosis accuracy of 96.82±0.84% with fewer unlabeled training samples under twelve different transfer tasks, which also indicates that the designed method can obtain similar performance with deep transfer network. In addition, although SF has been proved to achieve remarkable diagnostic performance [7], the transfer accuracy of SF will decrease when there are no labeled samples available in the target domain and domain adaptation is not considered. By introducing joint distribution adaptation, the transfer accuracy is significantly improved compared to SF and SFMDA. Moreover, the test accuracy of SFJDA and other methods under nine transfer tasks is shown in Fig. 5, which can be seen that the proposed scheme has better stability and adaptability for various transfer tasks and is expected to meet more complicated diagnostic requirements.

4 Conclusions

In this paper, a novel SFJDA model is designed for intelligent fault diagnosis, which is formed by stacking the subspaces obtained by double SFs. Compared with some existing representation learning methods, the proposed SFJDA can capture domain-invariant features with simpler networks and fewer parameters. Moreover, it can solve the problem of performance degradation when using SF to diagnose unlabeled target samples directly. To optimize the network parameters, an easy-to-implement learning strategy is developed by continuously iterating and updating the pseudo labels. Experimental results have verified the effectiveness of the proposed method in fault diagnosis.

References

[1] Y. Lei, B. Yang, X. Jiang, F. Jia, *et al.*, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, no. 106587, 2020.

[2] H. Chen, B. Jiang, S. X. Ding, and B. Huang, "Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1700-1716, 2022.

[3] R. Zhao, R. Yan, Z. Chen, K. Mao, *et al.*, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213-237, 2019.

[4] T. Han, L. Zhang, Z. Yin, and A. Tan, "Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine," *Measurement*, vol. 177, no. 109022, 2021.

[5] X. Kong, X. Li, Q. Zhou, Z. Hu, and C. Shi, "Attention recurrent autoencoder hybrid model for early fault diagnosis of rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1-10, 2021.

[6] H. Shao, M. Xia, J. Wan, and C. W. de Silva, "Modified stacked autoencoder using adaptive morlet wavelet for intelligent fault diagnosis of rotating machinery," *IEEE/ASME Trans. Mech.*, vol. 27, no. 1, pp. 24-33, 2022.

[7] J. Ngiam, Z. Chen, S. Bhaskar, P. Koh, and A. Ng, "Sparse filtering," *in Proc. Adv. Neural Inf. Process. Syst.*, pp. 1125–1133, 2011.

[8] Y. Lei, F. Jia, J. Lin, S. Xing, and S. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137-3147, 2016.

[9] Z. Zhang, Q. Yang, and Z. Wu, "Sparse filtering with adaptive basis weighting: A novel representation learning method for intelligent fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 2, pp. 1019-1025, 2022.

[10] W. Lu, B. Liang, Y. Cheng, D. Meng, *et al.*, "Deep model based domain adaptation for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296-2305, 2017.

[11] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 692-706, 2019.

[12] T. Han, C. Liu, W. Yang, and D. Jiang, "Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application," *ISA Trans.*, vol. 97, pp. 269-281, 2020.

[13] L. Zhang, L. Guo, H. Gao, D. Dong, *et al.*, "Instance-based ensemble deep transfer learning network: A new intelligent degradation recognition method and its application on ball screw," *Mech. Syst. Signal Process.*, vol. 140, no. 106681, 2020.

[14] J. Jiao, M. Zhao, and J. Lin, "Unsupervised adversarial adaptation network for intelligent fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 67, no. 11, pp. 9904-9913, 2020.

[15] W. M. Kouw, and M. Loog, "A review of domain adaptation without target labels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 766-785, 2021.

[16] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Open-set domain adaptation in machinery fault diagnostics using instance-level weighted adversarial learning," *IEEE Trans. Ind. Informat.*,vol. 17, no. 11, pp. 7445-7455, 2021.

[17] 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," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316-7325, 2019.

[18] H. Shao, M. Xia, G. Han, Y. Zhang, and J. Wan, "Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer convolutional neural network and thermal images," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3488-3496, 2021.

[19] A. Rozantsev, M. Salzmann, and P. Fua, "Beyond sharing weights for deep domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 801-814, 2019.

[20] S. Danafar, P. M. V. Rancoita, T. Glasmachers, K. Whittingstall, and J. Schmidhuber, "Testing hypotheses by regularized maximum mean discrepancy," *Computer Science*, 2013.

[21] LK., "Case western reserve university bearing data center," 2013. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home.

[22] L. V. D. Maaten and G. Hinton, "Visualizing data using t-sne," *J. Mach. Learn. Res.*, pp. 2579–2625, 2008.

[23] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, 2011.

[24] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *Computer Science*, 2014.

[25] T. Han, C. Liu, W. Yang, and D. Jiang, "A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults," *Knowl. Based Syst.*, vol. 165, pp. 474–487, 2019.