

ORIGINAL RESEARCH

Open Access



A dynamic-model-based fault diagnosis method for a wind turbine planetary gearbox using a deep learning network

Dongdong Li, Yang Zhao and Yao Zhao*

Abstract

The planetary gearbox is a critical part of wind turbines, and has great significance for their safety and reliability. Intelligent fault diagnosis methods for these gearboxes have made some achievements based on the availability of large quantities of labeled data. However, the data collected from the diagnosed devices are always unlabeled, and the acquisition of fault data from real gearboxes is time-consuming and laborious. As some gearbox faults can be conveniently simulated by a relatively precise dynamic model, the data from dynamic simulation containing some features are related to those from the actual machines. As a potential tool, transfer learning adapts a network trained in a source domain to its application in a target domain. Therefore, a novel fault diagnosis method combining transfer learning with dynamic model is proposed to identify the health conditions of planetary gearboxes. In the method, a modified lumped-parameter dynamic model of a planetary gear train is established to simulate the resultant vibration signal, while an optimized deep transfer learning network based on a one-dimensional convolutional neural network is built to extract domain-invariant features from different domains to achieve fault classification. Various groups of transfer diagnosis experiments of planetary gearboxes are carried out, and the experimental results demonstrate the effectiveness and the reliability of both the dynamic model and the proposed method.

Keywords: Wind turbine planetary gearbox, Lumped-parameter dynamic model, Intelligent fault diagnosis, Convolutional neural network, Transfer learning theory

1 Introduction

Wind energy has become one of the vital energy sources in the world, while wind power generation systems have been widely studied and applied [1]. The planetary gearbox is one of the critical components in the transmission system of wind turbines (WTs) because of its advantages of compact structure, high power density and desirable transmission efficiency [2]. However, in operation, planetary gearboxes are prone to failure and have high maintenance costs under dynamic load and frequently changing operating conditions [3]. Therefore,

accurate gearbox fault diagnosis is of great significance to improve the safety, reliability and economy of WTs [4].

In recent years, many intelligent methods have been investigated for gearbox fault diagnosis [5–9], while the proposed methods have two assumptions: (1) the training and testing data are derived from the same probability distribution; (2) enough labeled history data with fault information can be obtained [10]. However, in industrial applications, it is impractical to satisfy those two assumptions because of operating condition change, equipment wear degradation, and environmental noise interference, leading to differences of data in the probability distribution [11] and unlabeled data collected from the diagnosed devices [12].

Therefore, to solve the above two disadvantages, some studies have introduced transfer learning into

*Correspondence: nihaozhayao@163.com

School of Electrical Engineering, Shanghai University of Electric Power, No. 2103, Pingliang Road, Yangpu District, Shanghai 200090, China

fault diagnosis of mechanical equipment [13–16]. The transfer learning tasks consist of two datasets, one from the source domain and the other from the target domain. The data in the target domain is distributed differently from the data in the source domain but contains relevant knowledge. Thus, the goal of transfer learning is to improve the property of the predictive model for the target domain by using the common knowledge of the source and target domains. With the theoretical research of deep learning, deep hierarchical models are applied to learn transferable features from the cross-domain data automatically [17–19].

Hence, transfer learning can use the learned common knowledge from the source domain to solve a related task in the target domain [20–22]. Accordingly, some transfer-learning-based methods that mainly concentrate on the transfer tasks between different operation conditions are applied in [11, 23, 24]. Also, the transfer fault diagnoses among different devices have been studied. Reference [12] proposes a transfer learning method for bearing fault diagnosis, and its effectiveness is verified by the datasets acquired from three different machines. In [25], a transfer method is presented and the health conditions of bearings used in actual devices are classified with the help of the diagnosis knowledge from those used in the laboratory. Based on such methods, the fault diagnosis model trained with labeled data obtained from one machine can be generalized to the unlabeled data obtained from other similar machines. However, in the fault diagnosis of WT gearboxes, the above methods will encounter the following two problems:

- (1) Labeled fault data from similar machines are hard to obtain. The planetary gearboxes in WTs will not be allowed to run to failure since such a fault could lead to the breakdown of a WT or even serious accidents. In addition, gearboxes often undergo a long degradation process from normal to failure. Therefore, the acquisition of fault data is time-consuming and laborious.
- (2) Experimental data acquisition of WT gearboxes is costly. WTs are usually large in size, so it is expensive to build experimental platforms similar to the actual ones, while in the laboratory, when the type and extent of the faults are changed, new components are required and this is costly.

Such problems lead to insufficient samples in the actual fault diagnosis task. As a result, the performance of the deep transfer learning models will deteriorate and even fail to complete the diagnostic task. To solve the problems, an easier method is needed to get signals containing actual fault features. To gain an insight into

the signal characteristics of the gearboxes, various dynamic models have been presented, including multi-body models and lumped parameter models [26–28]. Dynamic simulation of the planetary gearbox with different faults has been realized, and it proves that the dynamic model can show many features of the actual signals [29–33]. The above studies have a common approach of introducing the influences of gear faults into the dynamic models by changing the mesh stiffness function of the mesh pair. In order to get a more accurate vibration response, reference [32] constructs a vibration signal model that can express the effect of transmission path by using a modified Hamming function.

In this paper, a dynamic-model-based method for WT planetary gearbox fault diagnosis using a deep transfer learning network (DTLN) is proposed. A modified lumped-parameter dynamic model is established to simulate the vibration signals of a planetary gear train, and the resultant vibration response is analysed by considering the transmission path of the signals. Then, an optimized DTLN based on a one-dimensional deep convolutional neural network (1-D CNN) is built. The DTLN comprises three modules: health condition recognition module, domain classifier and distribution discrepancy metrics. With the proposed three modules, the DTLN can extract domain-invariant features from the simulation data and the actual data, and the fault classification of actual datasets is realized. The introduction of simulation datasets makes up for the possible influence of insufficient samples in fault diagnosis models. Finally, multiple transfer diagnosis experiments are performed to verify the feasibility of the proposed method.

The main insights and contributions of this paper are summarized as follows.

- (1) A novel fault diagnosis method combining transfer learning with the dynamic model is proposed. This aims to remove the difficulty in obtaining enough labeled fault samples in applications. The cross-domain-invariant features of the simulation signal and the actual signal are learned by a deep transfer learning network, so as to realize the fault diagnosis of the actual signals.
- (2) The optimized DTLN comprises three parts: health condition recognition module, domain classifier and distribution discrepancy metrics. The health condition recognition module is based on a 1-D CNN built to learn the deep features of the input data, while the domain classifier and distribution discrepancy metrics are applied to help the network learn more domain-invariant features.

- (3) The proposed diagnosis method is based on unsupervised transfer learning theory. The labeled samples are not necessarily needed in the target domain. In practical applications of fault diagnosis, the data obtained from the devices to be diagnosed are always unlabeled. Therefore, it is appropriate for actual real-time diagnostic scenarios.
- (4) A model and data-driven approach is proposed. The application of traditional artificial intelligence methods relies on a large number of labeled samples of devices to be diagnosed. Compared with the traditional artificial intelligence method which only relies on the data-driven, the proposed method requires only unlabeled samples to be diagnosed, and the required number of samples in the target domain is greatly reduced. Therefore, the require-

into the model by modifying the mesh stiffness function of the mesh pairs.

2.1 Lumped-parameter model for a single stage planetary gear train

The lumped-parameter model is shown in Fig. 1. The system consists of one ring gear ‘r’, one sun gear ‘s’, one carrier ‘c’ and N equally spaced planet gears ‘ p_n ’. Herein, Oxy is the coordinate system rotating at the speed of ω_c with the x axis going through the center of p_1 . The sun gear has three degrees of freedom, i.e., two lateral motions (x, y) and one torsional motion (u). The other components have only torsional motion (u). According to Newton’s second law, the motion of a planetary gear train can be written as several second-order differential equations:

$$\left\{ \begin{array}{l} m_s(\ddot{x}_s - 2\omega_c\dot{y}_s - \omega_c^2x_s) + k_{sx}x_s + c_{sx}\dot{x}_s - \sum_{n=1}^N [k_{spn}(t)x_{spn} + c_{spn}\dot{x}_{spn}] \sin \psi_{sn} = 0 \\ m_s(\ddot{y}_s - 2\omega_c\dot{x}_s - \omega_c^2y_s) + k_{sy}y_s + c_{sy}\dot{y}_s + \sum_{n=1}^N [k_{spn}(t)x_{spn} + c_{spn}\dot{x}_{spn}] \cos \psi_{sn} = 0 \\ (I_s/r_s^2)\ddot{u}_s + k_{su}u_s + c_{su}\dot{u}_s + \sum_{n=1}^N [k_{spn}(t)x_{spn} + c_{spn}\dot{x}_{spn}] = T_s(t)/r_s \\ (I_r/r_r^2)\ddot{u}_r + k_{ru}u_r + c_{ru}\dot{u}_r + \sum_{n=1}^N [k_{rpn}(t)x_{rpn} + c_{rpn}\dot{x}_{rpn}] = 0 \\ (I_{pn}/r_{pn}^2)\ddot{u}_{pn} + k_{spn}(t)x_{spn} + c_{spn}\dot{x}_{spn} - k_{rpn}(t)x_{rpn} - c_{rpn}\dot{x}_{rpn} = 0 \\ (I_c/r_c^2)\ddot{u}_c + k_{cu}u_c + c_{cu}\dot{u}_c - \sum_{n=1}^N [k_{spn}(t)x_{spn} + c_{spn}\dot{x}_{spn}] + \sum_{n=1}^N [k_{rpn}(t)x_{rpn} + c_{rpn}\dot{x}_{rpn}] = -T_c(t)/r_c \end{array} \right. \quad (1)$$

ment of the dataset is reduced and its value in practical application is increased.

The rest of this paper is organized as follows. The dynamic model is presented in Sect. 2, and the proposed fault diagnosis framework is described in Sect. 3. In Sect. 4, the proposed dynamic model is validated, and its feasibility is validated on various experimental scenarios. Section 5 draws the conclusions.

2 Dynamic model of planetary gearbox

In this section, a modified lumped-parameter dynamic model is established. The model has the following characteristics: (1) the horizontal and vertical displacements of ring, planet and carrier that have limited influence on the resultant vibration response are ignored, (2) the lumped virtual spring-damping units are adopted in the model, (3) the effects of planet gear faults are introduced

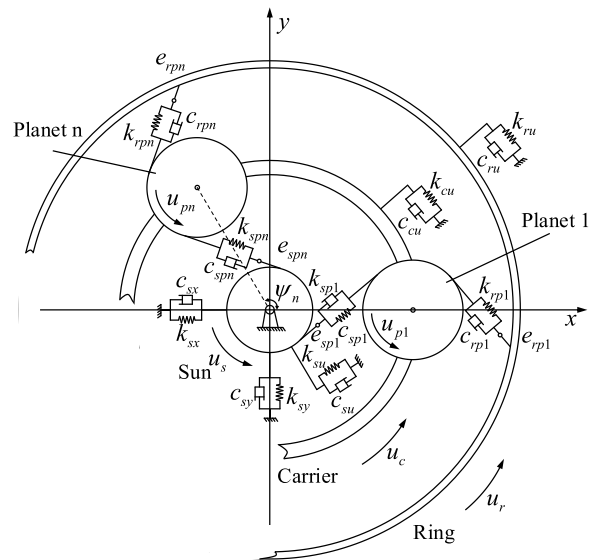


Fig. 1 Dynamic model of a single stage planetary gear train

with:

$$x_{spn} = -x_s \sin \psi_{sn} + y_s \cos \psi_{sn} + u_s + u_{pn} - e_{spn}(t) \tag{2}$$

$$x_{rpn} = u_r + u_{pn} - e_{rpn}(t) \tag{3}$$

$$\psi_{sn} = \psi_n - \alpha_s \tag{4}$$

$$\psi_n = 2\pi(n - 1)/N \tag{5}$$

The second-order nonlinear differential equations of motion can be solved by a fourth-order variable-step Runge–Kutta method after nondimensionalization.

2.2 Time-varying mesh stiffness

Time-varying mesh stiffness is one of the main sources of vibration response in a dynamic system. When the gearbox is free of any defects, the meshing stiffness of the gear is a function of its angular displacement and can be approximated by a square waveform. If a tooth is defective, partial contact loss will occur when the faulty tooth engages, leading to a local reduction of the mesh stiffness function [27]. Four planetary gear conditions are considered, including normal condition (NC), chipped tooth fault (CTF), surface wear fault (SWF) and missing tooth fault (MTF). The mesh stiffness losses denoted as ΔK are different under diverse faults. As the meshing stiffness is periodic, the meshing stiffness can be written as a Fourier series defined by (6) and (7), and the effects of the fault gears can then be introduced into the system.

$$k_{spn}(t) = k_{spnm} + \sum_{l=1}^q \left[k_{spna}^{(l)} \cos(l(\omega_m t - 2\pi\gamma_{spn} + \phi_{ek})) + k_{spnb}^{(l)} \sin(l(\omega_m t - 2\pi\gamma_{spn} + \phi_{ek})) \right] \tag{6}$$

$$k_{rpn}(t) = k_{rpnm} + \sum_{l=1}^q \left[k_{rpn a}^{(l)} \cos(l(\omega_m t - 2\pi\gamma_{rpn} + \phi_{ek})) + k_{rpn b}^{(l)} \sin(l(\omega_m t - 2\pi\gamma_{rpn} + \phi_{ek})) \right] \tag{7}$$

Because of the partial reduction of meshing stiffness, the amplitude and phase modulation effects appear in the vibration response spectrum in the form of sidebands, whose frequency locations depend on the fault location and fault type. These sidebands are also reflected in the vibration signals of actual planetary gearboxes. This is discussed in detail in Sect. 4.

2.3 Resultant signal model

In a planetary gearbox, the mesh vibrations along the torsional motion action lines are the main vibration

sources. Consequently, the mesh vibration acceleration signals of sun-planet and ring-planet mesh pairs are chosen to establish the resultant vibration signal model. The transmission path of a vibration signal is composed of two parts [29]: the first part is from the meshing vibration sources to the case, while the second part is the case to the transducer location. The influence of the first part on the vibration signals can be modeled by an attenuation coefficient, while the second part can be modeled by a modified Hamming function. Therefore, the resultant vibration signals of a planetary gearbox at the sensor location can be described as:

$$a(t) = \sum_n^N \exp^{\xi(\text{mod}(\omega_c t + \psi_n) - \pi)^2} W_n(S_{spn} a_{spn} \cos(\alpha_s - \omega_c t - \psi_n) + S_{rpn} a_{rpn} \cos(\alpha_r + \omega_c t + \psi_n)) \tag{8}$$

where the Hamming function $W_n = 0.54 - 0.46 \cos(\omega_c t + \psi_n)$. a_{spn} and a_{rpn} are the acceleration signals of sun-planet and ring-planet mesh pairs, respectively, and S_{spn} and S_{rpn} are the attenuation coefficients from the mesh pairs to the case. ξ is used to control the bandwidth of the Hamming function.

3 Proposed fault diagnosis framework

In this section, the proposed fault diagnosis framework based on the dynamic model and DTLN are introduced in detail.

3.1 Transfer learning problem definition

In order to clearly describe the problem, some concepts are introduced as follows. We take the source domain as $D_S = \{(xS, yS, i)\}$, where $xS, i \in \chi_S$ is a data sample and $yS, i \in Y_S$ is its corresponding label, and the target domain as $D_T = \{(xT, i)\}$, where $xT, i \in \chi_T$ is a data sample. D_S and D_T are drawn from distribution $P_S(X)$ and $P_T(X)$, and $P_S(X) \neq P_T(X)$ because of the domain bias. The same label space is used in different domains, i.e., $Y_T = Y_S$. In fault diagnosis, the goal of transfer learning is to improve the probabilistic prediction function of the domain D_T using the knowledge that can be learned in the domain D_S .

In this fault diagnosis, the target domain samples are the data obtained from the equipment to be diagnosed. In practical application, these data are unlabeled. The source domain samples are the available failure experimental data of similar equipment or the simulation data from the simulation model of the equipment to be diagnosed. These are labeled. The target task can be described as realizing the condition recognition of the target domain samples, that is, adding condition labels to the samples to be diagnosed.

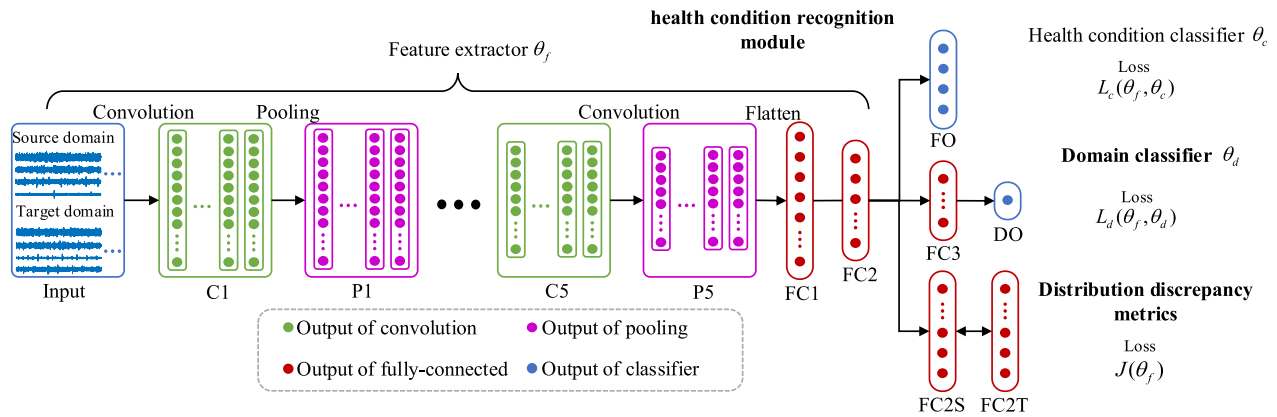


Fig. 2 Structural illustration of the DTLN

3.2 Structure and training process of the DTLN

As shown in Fig. 2, the optimized DTLN consists of three parts: health condition recognition module, domain classifier and distribution discrepancy metrics. These are briefly described below.

3.2.1 Health condition recognition module

The health condition recognition module is based on a 1-D CNN, which has the function of feature extraction and condition classification. In the 14-layer 1-D CNN, the first 13 layers are used for feature extraction and collation, and the last layer can be regarded as the condition classifier. In the convolutional layer, feature extraction is carried out, where the rectified linear unit is used as an activation function. Then a maximum pooling operation is introduced to reduce the feature dimension and enhance the feature robustly. The full connected layer and softmax regression are used at the end of the network to perform classification tasks. In summary, the output of the health condition recognition module can be defined as the output probability of the softmax function:

$$y = \frac{1}{\sum_{i=1}^K e^{(\mathbf{w}_i)^T f_2 + b}} \begin{bmatrix} e^{(\mathbf{w}_1)^T f_2 + b} \\ e^{(\mathbf{w}_2)^T f_2 + b} \\ \vdots \\ e^{(\mathbf{w}_K)^T f_2 + b} \end{bmatrix} \quad (9)$$

where f_2 is the output of the full connected layer FC2, \mathbf{w}_i denotes the weight matrix that concatenates to the i^{th} output neuron, b is the bias vector, and K is the number of health condition categories of the dataset.

3.2.2 Domain classifier

The domain classifier is a binary classifier that distinguishes source domains from target domains. As shown in Fig. 2, the domain classifier consists of a fully

connected layer and a binary output layer. The binary classifier setting with logistics regression is employed to distinguish between the source domain and target domain. The logistics regression is calculated as:

$$d = \frac{1}{1 + e^{-((\mathbf{w}_d)^T f_3 + b_d)}} \quad (10)$$

where \mathbf{w}_d is the weight matrix of the classifier, b_d is the corresponding bias vector and f_3 is the output of the layer FC3.

3.2.3 Distribution discrepancy metrics

In order to realize the extraction of domain-invariant features, a metric is required to represent the distribution difference between the features extracted from the source domain and those from the target domain. Here we use the Wasserstein distance to measure the distribution discrepancy between the two datasets. Let $P(f_2^{(S)})$ and $Q(f_2^{(T)})$ be the probability distributions where $f_2^{(S)}$ and $f_2^{(T)}$ are the features learned by 1-D CNN from the source domain and the target domain, respectively, according to the Kantorovich–Rubinstein dual theorem, The Wasserstein distance between the two distributions is computed as:

$$W = \sup_{\|G\|_L \leq 1} E_P[G(f_2^{(S)})] - E_Q[G(f_2^{(T)})] \quad (11)$$

where $\|G\|_L$ is the 1-Lipschitz function.

For the three components of the DTLN introduced above, each corresponds to an optimization object.

3.2.4 Object 1

Minimize the health condition classification error of the softmax classifier on source data. The objective function can be defined as the regression loss of a standard softmax classifier, as:

$$L_c = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k I[y_i = k] \log \frac{e^{((w_i)^T f_{2+b})}}{\sum_{l=1}^k e^{((w_l)^T f_{2+b})}} \right] \quad (12)$$

where m is the batch size of the data samples, k is the number of health condition categories, w_i denotes the weight matrix that concatenates to the i^{th} output neuron, and $I[\cdot]$ is an indicator function.

3.2.5 Object 2

Maximize the domain classification error on the source and target domain datasets. The loss function of the binary classifier can be represented as:

$$L = \frac{1}{m} \sum_{i=1}^m (l_i \log d(x_i) + (1 - l_i) \log (1 - d(x_i))) \quad (13)$$

where l_i denotes the real domain label, and $d(x_i)$ is a function that represents whether x_i comes from the source domain or the target domain. The objective function can be written as:

$$L_d = \frac{1}{n_s} \sum_{i=1}^{n_s} L(f_{2i}^{(S)}) + \frac{1}{n_t} \sum_{j=1}^{n_t} L(f_{2j}^{(T)}) \quad (14)$$

where $f_2(S)$ i and $f_2(T)$ j are the features learned from the source domain and the target domain, respectively.

3.2.6 Object 3

Minimize the Wasserstein distance between features extracted from the source and target domain datasets. Considering the gradient penalty item, the calculation formula is given as:

$$J = \frac{1}{n_s} \sum_{i=1}^{n_s} G(f_{2i}^{(S)}) - \frac{1}{n_t} \sum_{i=1}^{n_t} G(f_{2i}^{(T)}) - \gamma \left(\left\| \nabla_{\hat{H}} G(\hat{H}) \right\|_2 - 1 \right)^2 \quad (15)$$

where γ is the tradeoff parameter, n_s and n_t are the respective numbers of training samples from the source domain and target domain, and \hat{H} is a uniform sampling from the feature representations.

In conclusion, in order to extract as many cross-domain-invariant features as possible, the final optimization object can be combined as:

$$L(\theta_f^*, \theta_c^*, \theta_d^*) = \min_{\theta_f, \theta_c, \theta_d} L_c(\theta_f, \theta_c) - \lambda L_d(\theta_f, \theta_d) + \mu J(\theta_f) \quad (16)$$

where λ and μ are the hyperparameters, θ_f , θ_c , and θ_d are the parameters of the feature extractor, health condition classifier, and domain classifier, respectively.

Based on (16), in the back-propagation process, the parameters θ_f , θ_c , and θ_d are updated as:

$$\theta_f \leftarrow \theta_f - \varepsilon \left(\frac{\partial L_c}{\partial \theta_f} - \lambda \frac{\partial L_d}{\partial \theta_f} + \mu \frac{\partial J}{\partial \theta_f} \right) \quad (17)$$

$$\theta_c \leftarrow \theta_c - \varepsilon \frac{\partial L_c}{\partial \theta_c} \quad (18)$$

$$\theta_d \leftarrow \theta_d - \varepsilon \frac{\partial L_c}{\partial \theta_d} \quad (19)$$

where ε denotes the learning rate.

After training, the classifier can recognize the unlabeled samples from the target domain even if the learned domain-invariant features have equivocal domain categories and domain discrepancy. As shown in Fig. 3, DTLN uses labeled samples from the source domain and unlabeled samples from the target domain for training. The invariant features of the domain are learned first, and then the classifier determines the category based on the learned features. After the training, the trained network will be tested by the sample set from the target domain.

3.3 Proposed fault diagnosis framework

The framework of the proposed method is illustrated in Fig. 4. As shown, the method includes three parts, as introduced below.

3.3.1 Part 1: data acquisition and preprocessing

In this part, the source domain and target domain are constructed, where the target domain data samples are obtained from the gearbox to be diagnosed, and the source domain data samples are obtained by analyzing

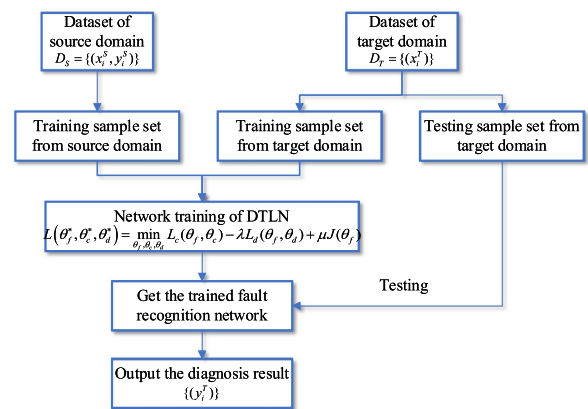


Fig. 3 Flowchart of the proposed DTLN

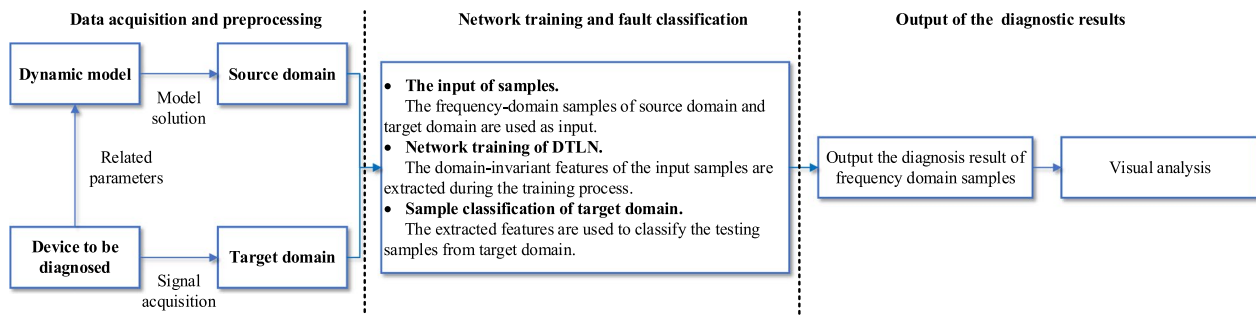


Fig. 4 Fault diagnosis framework of the proposed method

the dynamic model. It is worth noting that the relevant parameters of the dynamic model are taken from the device to be diagnosed. After acquiring the vibration signal, the samples are processed and the frequency-domain samples are used as the input of the DTLN. This is because frequency-domain samples are more robust to noise than time-domain samples and contain more domain-invariant features. This will be demonstrated in detail in Sect. 4.

3.3.2 Part 2: network training and fault classification

In order to extract more domain-invariant features of the source domain and target domain, frequency domain samples are used to train the DTLN, and the trained network can be obtained. The training process is based on (17–19). The trained network is tested by unlabeled testing samples from the target domain and outputs classified results.

3.3.3 Part 3: output of the diagnostic results

The trained diagnostic model is applied to the fault diagnosis of experimental equipment to output the diagnosis results. In order to show the feasibility of the proposed method, the above classification results are analyzed visually.

According to the above fault diagnosis framework, the dynamic model is used to construct the source domain in the transfer learning method. This is helpful for fault diagnosis of true devices. In the following, the rationality and advantages of the proposed method are verified.

4 Experimental results and comparisons

In this section, similarities between the simulation signal and the actual signal are analyzed. Multiple experiments are performed to validate the network and fault diagnosis framework.

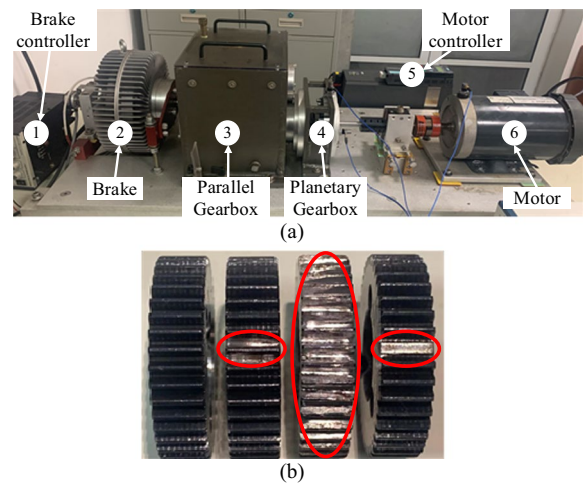


Fig. 5 Experimental setup for gearbox dataset. (a) DDS. (b) fault gears

Table 1 Parameters of the dynamic model

Nomenclature	s	p	r	c
Teeth number	20	40	100	–
Module (mm)	1	1	1	–
Pressure angle (deg)	20	20	20	–
Teeth width (mm)	10	10	10	–
Mass (kg)	0.069	0.057	0.422	0.709
Inertia (kg·mm ²)	6.415	11.791	1417.482	364.756
Young’s modulus (G Pa)	206	206	206	206
Poisson’s ratio	0.3	0.3	0.3	0.3
Pitch diameter (mm)	20	40	100	–
Torsional bearing stiffness (N·mm ⁻¹)	0	0	1000	0

4.1 Validation and analysis of the simulation model

4.1.1 Planetary gearbox fault experiment

The gearbox dataset is collected from the drivetrain dynamic simulator (DDS) shown in Fig. 5a. The planetary

gearbox has two stages, and the faults of planet gears in the first stage are studied. For vibration signal acquisition, an acceleration transducer is mounted. Experiments are carried out on planet gears in four healthy conditions, shown from left to right in Fig. 5b as NC, CTF, SWF and MTF, respectively. The sampling frequency of the transducer is set at 12 kHz.

4.1.2 Simulation parameters

The basic design parameters are listed in Table 1. The planetary gearbox has three planet gears ($N=3$) with a fixed ring gear. The mesh damping of sun-planet pair and ring-planet are set as 242.6 N·s/m and 410.3 N·s/m, respectively. The bearing stiffness and damping of sun gear are assumed to be 15 N·mm⁻¹ and 9.2 N·s·m⁻¹, respectively. The constant torque acting on the carrier is 1.26 Nm, and the sampling frequency of the simulation signal is 12 kHz. In the resultant signal model, according

to the structure of DDS, S_{spn} and S_{rpn} are set as 0.4 and 0.9, while ξ is -1.

To simplify the analysis, the order spectra are represented by normalizing with rotational frequency of the carrier. The mesh order $H_m=Z_r$ denotes the mesh frequency f_m ($H_m=f_m/f_c=Z_r f_c/f_c=Z_r$, where Z_r is the teeth of ring gear). The rotation period of the carrier T_c is equal to $Z_r T_m$, where T_m is the mesh period. For the planetary gear train, $Z_r=100$, $f_m=100f_c$, $T_c=100T_m$. With the above settings, the vibration response can be determined.

4.1.3 Spectrum analysis of simulation signal

As described in Sect. 2, modulation effects appear in vibration response because of the partial reduction of meshing stiffness, in the form of sidebands in the vibration spectrum. Therefore, the frequency spectra are analyzed.

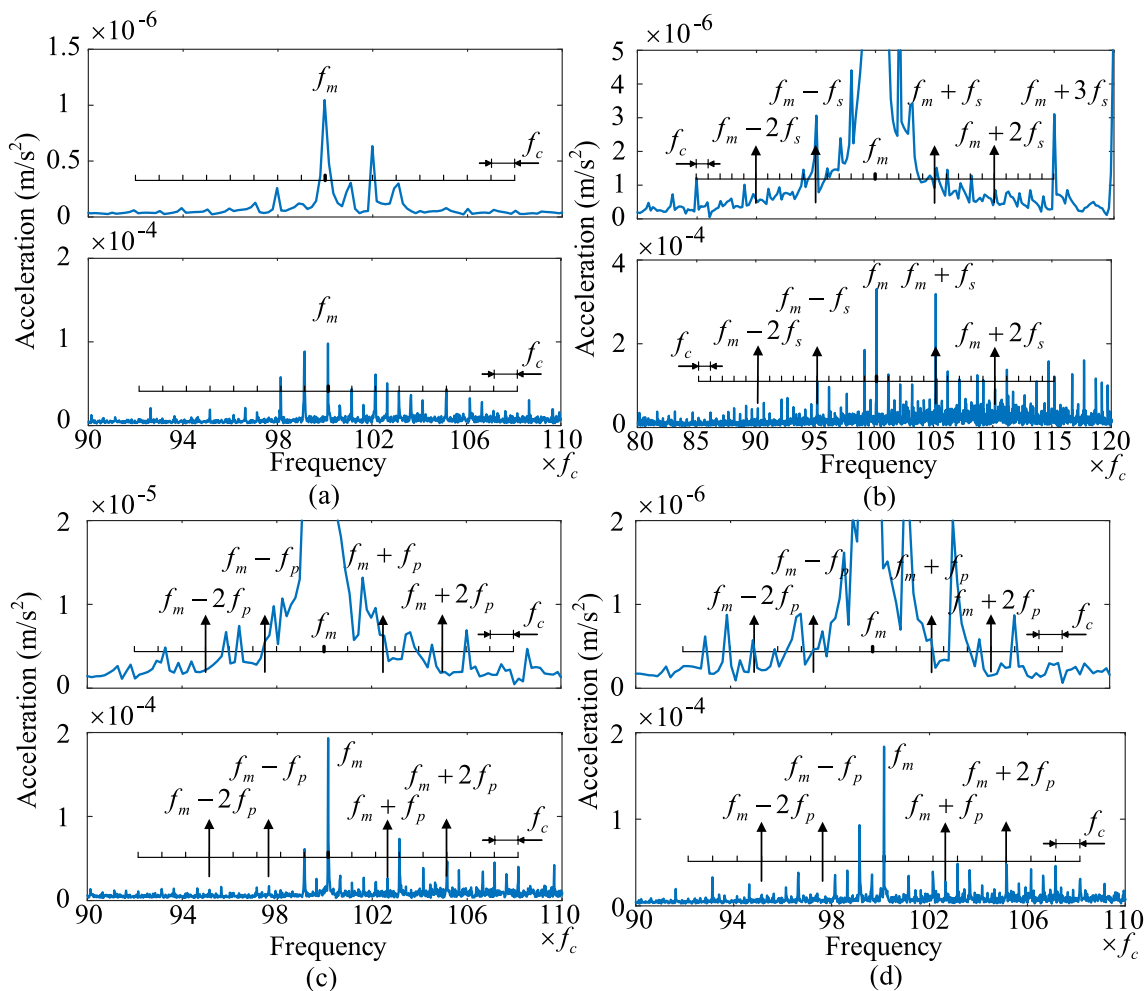


Fig. 6 Frequency spectra of vibration signals from dynamic model (above) and DDS (below). (a) NC. (b) SWF. (c) CTF. (d) MTF

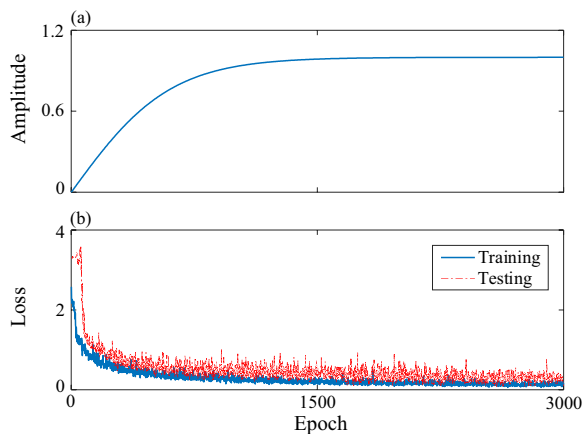


Fig. 7 Penalty parameter and loss of the proposed DTLN. (a) Penalty parameter. (b) Training and testing loss

When planet gear faults occur, fault features will appear near the meshing frequency, while the frequency locations depend on the fault location and type. As shown in Fig. 6a, when working under NC, the sidebands locate at $f_m \pm nf_c$ (n is an integer) because of the modulation of the transmission path. After introducing failures of the planet, some impulsive signals appear. As a result, the spectrum contains some additional frequency components. As shown in Fig. 7c and d, when the planet gear has local faults such as CTF and MTF, these sidebands are at the locations of $f_m \pm mf_p \pm nf_c$ (m is an integer), which also exist in the actual signal. This signifies that the signals are modulated by the fault of planet gear and the transmission path. It is worth noting that MTF causes more characteristic frequencies than CTF in both

simulation and experiment. As shown in Fig. 6b, the global fault causes the characteristic frequency f_s of the sun gear. The amplitudes of the order spectrum locate at the $f_m \pm kf_s \pm nf_c$ (k is an integer). As seen, the simulation and experimental results are consistent. From the above analyses, it is verified that the simulation results of the dynamic model contain some features of the actual signals. This is the basis for the applicability of transfer learning theory, i.e., the source domain and target domain contain common diagnostic knowledge. In addition, fault features can be detected from the order spectrum. Therefore, the fault identification of frequency domain signals can help the overall diagnosis decision.

4.2 Transfer fault diagnosis experiments

The three datasets required for validation and 16 diagnosis experiments are described in this section.

- (1) *A: Experimental Planetary Gearbox Dataset* The dataset is collected from DDS, which contains four working conditions with various motor speeds and a certain load: 1200r/min (A_1), 1800r/min (A_2). Each health condition, i.e., NC, CTF, MTF and SWF, has 800 samples. Thus, this dataset has a total of 3200 samples, each of which has 2000 data points.
- (2) *B: Planetary Gearbox Dataset Used by Another Group* The dataset is collected from a similar planetary gearbox under different working conditions. This was provided by Yan’s group [34]. Four healthy conditions in the dataset are selected, and the working conditions are investigated with the rotating speed system load set at 20 Hz–0 V (B_1) and 30 Hz–2 V (B_2). Similarly, 800 samples for each condition are intercepted, and each sample contains 2,000 points.
- (3) *C: Dataset Acquired by Simulation* The dataset is acquired by dynamic simulation, and rotation frequencies of the sun gear are set at 20 Hz (C_1) and 30 Hz (C_2), each of which also has four healthy conditions. The conditions and properties of the samples are the same as those of datasets A and B.

Table 2 Virus transfer task

Class	Transfer task	Source domain	Target domain	Healthy conditions
1	$A_1 \rightarrow B_1$	1200 rpm	1200 rpm	NC, CTF, MTF, SWF
	$B_1 \rightarrow A_1$			
2	$A_2 \rightarrow B_2$	1800 rpm	1800 rpm	
	$B_2 \rightarrow A_2$			
3	$A_1 \rightarrow B_2$	1200 rpm	1800 rpm	
	$B_1 \rightarrow A_2$			
4	$B_2 \rightarrow A_1$	1800 rpm	1200 rpm	
	$A_2 \rightarrow B_1$			
3	$C_1 \rightarrow A_1$	1200 rpm	1200 rpm	
	$C_1 \rightarrow B_1$			
4	$C_2 \rightarrow A_2$	1800 rpm	1800 rpm	
	$C_2 \rightarrow B_2$			
4	$C_1 \rightarrow A_2$	1200 rpm	1800 rpm	
	$C_1 \rightarrow B_2$			
4	$C_2 \rightarrow A_1$	1800 rpm	1200 rpm	
	$C_2 \rightarrow B_2$			

Sixteen transfer fault diagnosis experiments are shown in Table 2. Taking the task $A_1 \rightarrow B_1$ for example, A_1 is the source domain, and B_1 is the target domain. The standard assessment protocol for unsupervised transfer learning missions is adopted. In each transfer task, the training dataset consists of all labeled data samples from the source domain and half of the unlabeled data samples from the target domain, while the testing dataset is composed of the other half of samples from the target domain. Among the 16 experiments, the groups of Class

Table 3 Architecture of the 1-D CNN

Layer	Parameters	Stride	Output size
Input	/	/	2000×1
C1	64×1 conv	1	$16-[2000 \times 1]$
P1	2×1 max-pool	2	$16-[1000 \times 1]$
C2	16×1 conv	1	$32-[1000 \times 1]$
P2	2×1 max-pool	2	$32-[500 \times 1]$
C3	16×1 conv	1	$32-[500 \times 1]$
P3	2×1 max-pool	2	$32-[250 \times 1]$
C4	5×1 conv	1	$64-[250 \times 1]$
P4	2×1 max-pool	2	$64-[125 \times 1]$
C5	5×1 conv	1	$64-[125 \times 1]$
P5	2×1 max-pool	2	$64-[62 \times 1]$
FC1	Flatten	/	3968×1
FC2	Fully-connected	/	256×1
FO	Softmax	/	4×1

1 are the transfer diagnosis from one machine to another under similar working conditions, while those in Class 2 are under different working conditions. Class 3 is the transfer fault diagnosis between dynamic model and actual machines at the same speed, while Class 4 is the diagnostic experiments of the proposed method at different rotational speeds.

The detailed parameters of the DTLN can be found in Table 3, in which 64×1 conv denotes the size of the convolutional kernel, 2×1 max-pool stands for the size of max-pooling operation, and $16-[2000 \times 1]$ represents 16 feature maps of size 2000×1 . In order to restrain noise and extract useful knowledge, a wide kernel is used in C1. As shown in Fig. 7a, the hyperparameters λ and μ in (16) are set to gradually increase from 0 to 1, and the calculation formula is $2/(1 + \exp(-10 \times p)) - 1$, where p denotes the training progress. In order to minimize the loss function, the Adam is used as an optimization algorithm and

the learning rate is set as 0.001. The batch size is set as 512, and the epoch of the training is 3000. Taking the experiment $C_1 \rightarrow A_1$ for example, the loss function during the training process is drawn in Fig. 7b. It is clear that the loss function converges after about 1500 steps.

To reduce contingency and particularity of results, each transfer fault diagnosis experiment is carried out 10 times, and the results are shown in Table 4. It is worth noting that DTLN-T indicates that the input of the network is time-domain samples, while DTLN-F indicates that the input is frequency-domain samples. The figures in Table 4 represent the average accuracy rate and standard deviation of 10 repeated classification experiments. The accuracy rate reflects the reliability of the method, and the standard deviation reflects the stability of the method. In the transfer experiments between different devices, i.e., $A \rightarrow B$, $B \rightarrow A$, the diagnostic accuracies of the proposed method are over 91%. In addition, in the transfer experiments from dynamic model to actual devices, i.e., $C \rightarrow A$, $C \rightarrow B$, the average diagnostic accuracy of Class 3 is 90.9%, which indicates that the proposed method is feasible.

4.3 Results comparison and visual analysis

In order to demonstrate the effectiveness and the feasibility of the proposed method, two other networks are chosen for comparison. Among them, the basic convolutional network has the same structure and parameters as the 1-D CNN introduced above, and uses source data for training and then tries to classify target data. In addition, a domain adversarial neural network (DANN) which is a commonly used transfer learning method is also tested, and its parameter setting refers to [35]. The learning rate of the CNN is 0.01, and is 2×10^{-4} for the DANN. The Adam algorithm is used as optimization algorithm in both methods. It is worth noting that the inputs of the CNN, DANN and DTLN-F are all frequency-domain

Table 4 Recognition result of experiments

Method	CNN	DANN	DTLN-T	DTLN-F	Method	CNN	DANN	DTLN-T	DTLN-F
$A_1 \rightarrow B_1$	0.413 ± 0.028	0.888 ± 0.012	0.962 ± 0.009	0.961 ± 0.009	$A_1 \rightarrow B_2$	0.314 ± 0.024	0.77 ± 0.022	0.919 ± 0.022	0.911 ± 0.011
$B_1 \rightarrow A_1$	0.411 ± 0.015	0.885 ± 0.020	0.953 ± 0.015	0.964 ± 0.008	$B_1 \rightarrow A_2$	0.337 ± 0.031	0.798 ± 0.042	0.915 ± 0.023	0.906 ± 0.030
$A_2 \rightarrow B_2$	0.410 ± 0.015	0.879 ± 0.018	0.952 ± 0.010	0.951 ± 0.007	$B_2 \rightarrow A_1$	0.332 ± 0.031	0.781 ± 0.025	0.889 ± 0.018	0.918 ± 0.006
$B_2 \rightarrow A_2$	0.406 ± 0.016	0.891 ± 0.010	0.959 ± 0.016	0.949 ± 0.011	$A_2 \rightarrow B_1$	0.339 ± 0.019	0.783 ± 0.027	0.890 ± 0.014	0.906 ± 0.008
Average	0.410	0.886	0.957	0.956	Average	0.331	0.783	0.903	0.910
$C_1 \rightarrow A_1$	0.318 ± 0.027	0.723 ± 0.042	0.826 ± 0.010	0.904 ± 0.013	$C_1 \rightarrow A_2$	0.285 ± 0.030	0.658 ± 0.049	0.767 ± 0.017	0.892 ± 0.016
$C_1 \rightarrow B_1$	0.313 ± 0.031	0.702 ± 0.045	0.836 ± 0.008	0.914 ± 0.013	$C_1 \rightarrow B_2$	0.293 ± 0.023	0.656 ± 0.047	0.737 ± 0.027	0.882 ± 0.019
$C_2 \rightarrow A_2$	0.336 ± 0.027	0.688 ± 0.032	0.818 ± 0.011	0.899 ± 0.018	$C_2 \rightarrow A_1$	0.287 ± 0.025	0.673 ± 0.049	0.768 ± 0.027	0.898 ± 0.021
$C_2 \rightarrow B_2$	0.336 ± 0.029	0.676 ± 0.025	0.820 ± 0.016	0.918 ± 0.008	$C_2 \rightarrow B_2$	0.288 ± 0.028	0.696 ± 0.049	0.748 ± 0.032	0.885 ± 0.013
Average	0.326	0.697	0.825	0.909	Average	0.288	0.671	0.755	0.889

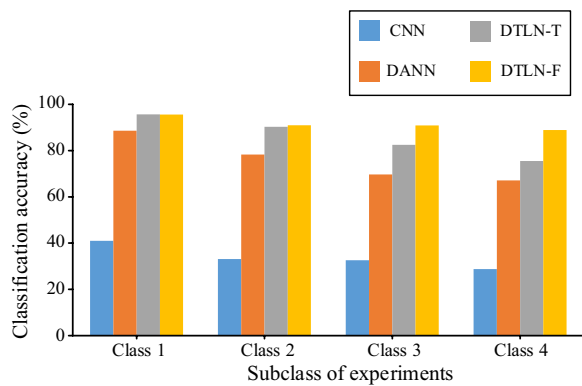


Fig. 8 Classification accuracy of average tenfold cross validation on the subclass of experiments

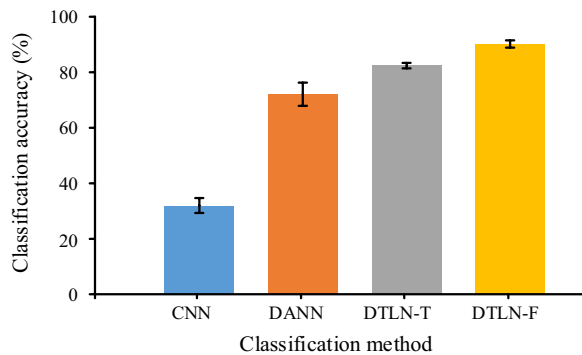


Fig. 9 Classification accuracy of average tenfold cross validation on the transfer experiments from C_1 (source domain) to A_1 (target domain)

samples. The accuracies and standard deviations on the 16 transfer fault diagnosis experiments are shown in Table 4. Classification accuracies of average tenfold cross

validation on the subclass in Table 2 are shown in Fig. 8. Taking the task $C_1 \rightarrow A_1$, for example, the classification accuracy and standard deviation of various methods are compared in Fig. 9. By comparing the results, three observations can be made:

- (1) For the transfer fault diagnosis missions where unlabeled data are retrievable in the target domain, the networks based on transfer learning are superior. It suggests that transfer learning can be an effective instrument to facilitate the practical application of intelligent diagnostics. Additionally, compared with DANN, DTLN obtains higher classification accuracies and lower standard deviations, as shown in Fig. 10. It indicates that DTLN reduces the distribution discrepancy between different domains more effectively and is relatively stable.
- (2) Compared with experiments where the networks are trained by data from machines, the accuracies of experiments that replace the actual labeled data with the simulation data are reduced. This indicates that the fault signals of similar equipment contain more domain-invariant features, but the simulated signals contain fewer. However, compared with the time-domain samples, the classification accuracy obtained when the frequency-domain samples are used as the input is higher, as shown in Fig. 8. Therefore, it can be inferred that when the samples are in the frequency domain, it is more favorable for the DTLN to extract the domain-invariant features. Therefore, the method proposed in this paper adopts the frequency-domain samples as the input. As a result, the transfer experiments from simulation model to actual devices can realize relatively high accuracy. When the set speed of the

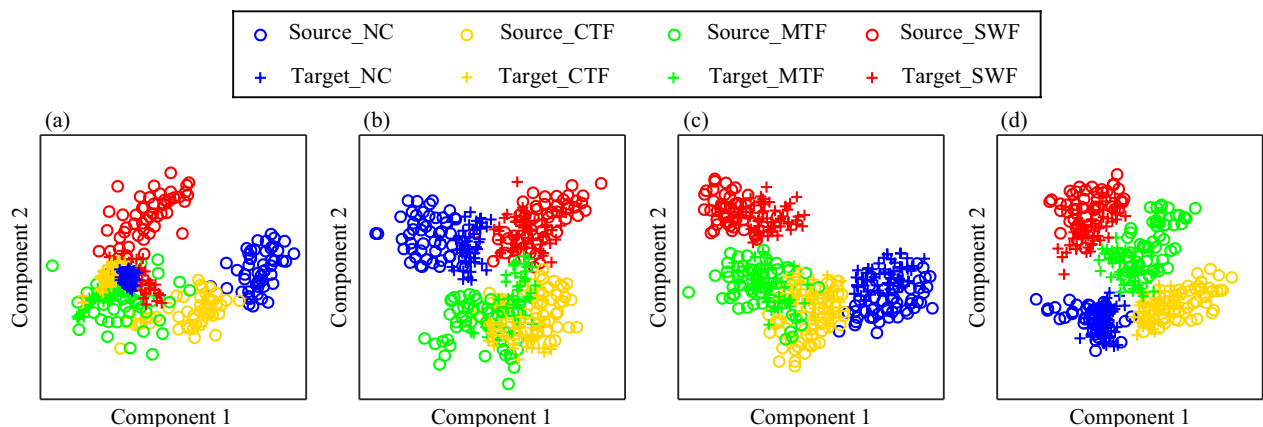


Fig. 10 The visualization of the learned features on the dataset C_1 (source domain) and the dataset A_1 (target domain): (a) CNN, (b) DANN, and (c) DTLN-T, (d) DTLN-F

dynamic model is the same as the actual speed, the average classification accuracy is 90.9%. Under different speed settings, the average classification accuracy is 88.9%. This proves that the diagnosis method combining the dynamic model with a deep transfer learning network has practical value. In similar cross-domain transfer diagnosis experiments, an 84.32% recognition rate for bearing faults in mechanical equipment is achieved in [25], while reference [12] achieves an 86.3% diagnosis accuracy. Therefore, compared with the existing research results, the proposed method has certain advantages.

- (3) The accuracies obtained from the transfer learning between datasets under similar working conditions are higher. This is exemplified by the fact that the accuracy of Class 2 is lower than that of Class 1, and Class 4 is lower than that of Class 3. Therefore, the working condition is one of the important influencing factors of transfer learning in practical application. In order to improve the accuracy of diagnosis, the rotational speed of the dynamic model can be adjusted to be the same as that of the actual equipment to be diagnosed, so as to obtain more accurate fault discrimination.

In order to intuitively show the classification effect, a t-distributed stochastic neighbor embedding (t-SNE) algorithm is introduced. This can map the high-dimensional features into 2-D space and the distribution of features can be plotted directly. Taking the task $C_1 \rightarrow A_1$ for example, the transferable features learned by CNN, DANN, DTLN-T and DTLN-F are shown in Fig. 10 via t-SNE. In addition, the confusion matrices for transfer results on dataset A_1 can be explored. These are shown in Fig. 11.

From Fig. 10a, the features learned by CNN have clear distribution discrepancy. As a result, when CNN is trained with C_1 , its recognition for A_1 is close to surmise. As for DANN, the cross-domain distribution discrepancy is amended to a certain extent as shown in Fig. 10b, so the accuracy of DANN for A_1 is much higher than that of CNN. From Fig. 10c and d, the proposed DTLN is able to amend the distribution discrepancy between the learned features of different datasets. However, because of the difference in transferability between the subclass samples of the source domain and target domain, the distribution discrepancy of the subclass samples is corrected asymmetrically. For example, in Fig. 10c, the distribution discrepancy of the cross-domain samples with CTF is still severe after the correcting of DTLN-T. To illustrate with confusion matrices, shown in Fig. 11c and d, the classification effects of DTLN-T and DTLN-F are defective, whereas the classification of DTLN-F is better.

5 Conclusion

In this paper, a dynamic-model-based transfer learning fault diagnosis method for WT planetary gearboxes is proposed. This method introduces a dynamic simulation dataset into the application of transfer learning and produces a diagnosis of unlabeled fault data obtained from actual machines. To verify the feasibility of the proposed method, spectrum analysis of the simulated and experimental signals is carried out, and 16 groups of transfer fault diagnosis experiments are completed. From the results, the following conclusions can be drawn.

- (1) Through the spectrum analysis, the vibration response solved by the dynamic model contains some features of the actual fault vibration signal, i.e., domain-invariant features required by transfer learning.

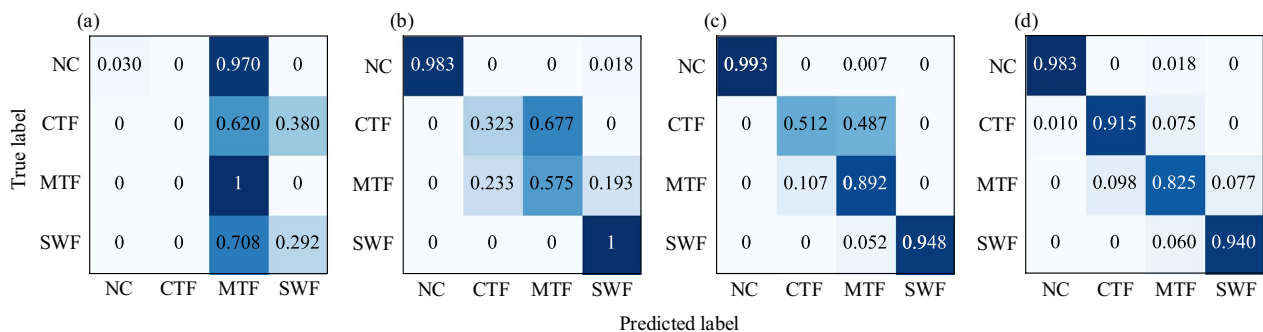


Fig. 11 The confusion matrices for the transfer results of the dataset A_1 : (a) CNN, (b) DANN, and (c) DTLN-T, (d) DTLN-F

- (2) The proposed DTLN can effectively realize the recognition of unlabeled fault data from the target domain. In the application of transfer fault diagnosis, the classification accuracy and stability of the DTLN are better than those of the DANN.
- (3) The proposed method combining a dynamic model with the deep transfer learning network can identify four kinds of faults of the planetary gearbox. When the set speed of the dynamic model is the same as the actual speed, the average classification accuracy is 90.9%.

The results indicate that the proposed method that combines the transfer learning theory with dynamic model is feasible, whereas the dynamic model proposed can be further optimized. After introducing the dynamic model, varied labeled fault data can be obtained, and the fault setting is more independent and convenient. This leads to practical application value.

There has been rapid development of artificial intelligence algorithms. However, because of the operating environment, working condition, data acquisition difficulty etc., artificial intelligence methods in the field of fault diagnosis are developing slowly. Therefore, how to combine artificial intelligence methods with practical applications of condition recognition to achieve higher accuracy is the direction of future research.

List of symbols

m_s : Mass of the sun gear; x_s, y_s : Translational displacement of x -axis and y -axis for sun gear; u_i : Torsional displacements for each component ($i = s, r, c, p_n$); k_{ij}, c_{ij} : Supporting stiffness and damping for each component at different motions ($j = x, y, u$); $k_{spn}(t), k_{rpn}(t)$: Stiffness function between sun-planet and ring-planet meshing pairs ($n = 1, \dots, M$); c_{spn}, c_{rpn} : Constant damping between sun-planet and ring-planet meshing pairs; $e_{spn}(t), e_{rpn}(t)$: Transmission errors of the n th sun-planet and ring-planet pairs; x_{spn}, x_{rpn} : Relative displacements along the torsional motion action lines of sun-planet and ring-planet; I_j : Inertias of each component; r_j : Base circle radii of each component; $T_s(t), T_c(t)$: External torques on the sun gear and the carrier; α_s : Meshing angles of the sun-planet gear mesh; ψ_n : Circumferential position angle of the n th planet gear; q : Number of harmonic terms; k_{spnm}, k_{rpnm} : Mean values of time-varying mesh stiffness of sun-planet and ring-planet meshing pairs; ω_m : Mesh frequency; γ_{spn} : Relative phase between the n th sun-planet mesh and the first sun-planet mesh; γ_{rpn} : Relative phase between the n th ring-planet mesh and the first sun-planet mesh; φ_{sk} : Phase difference between $e_{spn}(t)$ and $k_{spn}(t)$, and between $e_{rpn}(t)$ and $k_{rpn}(t)$; $K(l)_{spna}, K(l)_{spnb}, K(l)_{rpna}, K(l)_{rpnb}$: Harmonic coefficients of Fourier series.

Abbreviations

WT: Wind turbine; DTLN: Deep transfer learning network; 1-D CNN: One-dimensional deep convolutional neural network; NC: Normal condition; CTF: Chipped tooth fault; SWF: Surface wear fault; MTF: Miss tooth fault; DDS: Drivetrain dynamic simulator; DANN: Domain adversarial neural network; t-SNE: T-distributed stochastic neighbor embedding.

Acknowledgements

Not applicable.

Author's information

Dongdong Li received his B.S. and Ph.D. degrees from Zhejiang University and Shanghai Jiao Tong University both in electrical engineering in 1998 and 2005, respectively. He is currently a professor and dean of College of Electric Engineering in Shanghai University of Electric Power, Shanghai, China. His current research interests include analysis of electric power system, renewable energy system, smart grid and power electronization of power system.

Yang Zhao received his B.S. degree from Shanghai University of Electric Power in 2019. He is currently pursuing the M.S. degree in electrical engineering at Shanghai University of Electric Power. His current research interests include artificial intelligence algorithm, fault diagnosis of wind turbine planetary gearbox.

Yao Zhao received the B.S. degree in automation from Anhui University, Hefei, China, in 2009, the M.S. degree in electrical engineering from Shanghai Maritime University, Shanghai, China, in 2011, and the Ph.D. degree in from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2016. He is currently a Lecturer with the Shanghai University of Electric Power, Shanghai, China. His main research interests include electrical machines and power electronization of power system.

Author contributions

Not applicable.

Funding

Natural Science Foundation of Shanghai (21ZR1425400), Shanghai Rising-Star Program (21QC1400200), National Natural Science Foundation of China (51977128), Shanghai Science and Technology Project (20142202600).

Availability of data and materials

Data and materials are not public, but can be uploaded and made public if necessary.

Declarations

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 28 December 2021 Accepted: 28 May 2022

Published online: 10 June 2022

References

1. Nadour, M., Essadki, A., & Nasser, T. (2020). Improving low-voltage ride-through capability of a multimegawatt DFIG based wind turbine under grid faults. *Protection and Control of Modern Power Systems*, 5(4), 102–114.
2. Salameh, J. P., Cauet, S., Etien, E., Sakout, A., & Rambault, L. (2018). Gearbox condition monitoring in wind turbines: A review. *Mechanical Systems and Signal Processing*, 111, 251–264.
3. Leite, G. D. N. P., Araújo, A. M., & Rosas, P. A. C. (2018). Prognostic techniques applied to maintenance of wind turbines: A concise and specific review. *Renewable and Sustainable Energy Reviews*, 81, 1917–1925.
4. Desai, J. P., & Makwana, V. H. (2021). A novel out of step relaying algorithm based on wavelet transform and a deep learning machine model. *Protection and Control of Modern Power Systems*, 6(4), 500–511.
5. Jiang, G., He, H., Yan, J., & Xie, P. (2019). Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Transactions on Industrial Electronics*, 66(4), 3196–3207.
6. Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., & Lim, M. H. (2020). Gearbox fault diagnosis using a deep learning model with limited data sample. *IEEE Transactions on Industrial Informatics*, 16(10), 6263–6271.
7. Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108, 33–47.
8. Zhao, M., Kang, M., Tang, B., & Pecht, M. (2018). Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of

- planetary gearboxes. *IEEE Transactions on Industrial Electronics*, 65(5), 4290–4300.
9. Su, X., Shan, Y., Zhou, W., & Fu, Y. (2021). GRU and attention mechanism-based condition monitoring of an offshore wind turbine gearbox. *Power System Protection and Control*, 49(24), 141–149. (in Chinese).
 10. Ding, S., Li, X., Hang, J., Wang, Y., & Wang, Q. (2020). Deep learning theory and its application to fault diagnosis of an electric machine. *Power System Protection and Control*, 48(8), 172–187. (in Chinese).
 11. Jiao, J., Zhao, M., & Lin, J. (2020). Unsupervised adversarial adaptation network for intelligent fault diagnosis. *IEEE Transactions on Industrial Electronics*, 67(11), 9904–9913.
 12. Guo, L., Lei, Y., Xing, S., Yan, T., & Li, N. (2019). Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data. *IEEE Transactions on Industrial Electronics*, 66(9), 7316–7325.
 13. Mao, W., Liu, Y., Ding, L., Safian, A., & Liang, X. (2021). A New structured domain adversarial neural network for transfer fault diagnosis of rolling bearings under different working conditions. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–13.
 14. Chen, Z., He, G., Li, J., Liao, Y., Gryllias, K., & Li, W. (2020). Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery. *IEEE Transactions on Instrumentation and Measurement*, 69, 8702–8712.
 15. Shen, C., Wang, X., Wang, D., Li, Y., Zhu, J., & Gong, M. (2021). Dynamic joint distribution alignment network for bearing fault diagnosis under variable working conditions. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–13.
 16. Feng, L., & Zhao, C. (2021). Fault description based attribute transfer for zero-sample industrial fault diagnosis. *IEEE Transactions on Industrial Informatics*, 17, 1852–1862.
 17. Li, X., Zhang, W., Ding, Q., & Li, X. (2020). Diagnosing rotating machines with weakly supervised data using deep transfer learning. *IEEE Transactions on Industrial Informatics*, 16, 1688–1697.
 18. Xu, G., Liu, M., Jiang, Z., Shen, W., & Huang, C. (2020). Online fault diagnosis method based on transfer convolutional neural networks. *IEEE Transactions on Instrumentation and Measurement*, 69, 509–520.
 19. Wang, J., Zhao, R., & Gao, R. X. (2020). Probabilistic transfer factor analysis for machinery autonomous diagnosis cross various operating conditions. *IEEE Transactions on Instrumentation and Measurement*, 69, 5335–5344.
 20. Yang, B., Lei, Y., Jia, F., Li, N., & Du, Z. (2020). A polynomial Kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines. *IEEE Transactions on Industrial Electronics*, 67(11), 9747–9757.
 21. Persello, C., & Bruzzone, L. (2016). Kernel-based domain-invariant feature selection in hyperspectral images for transfer learning. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5), 2615–2626.
 22. Chen, Z., Gryllias, K., & Li, W. (2020). Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network. *IEEE Transactions on Industrial Informatics*, 16(1), 339–349.
 23. Lu, W., Liang, B., Cheng, Y., Meng, D., Yang, J., & Zhang, T. (2017). Deep model based domain adaptation for fault diagnosis. *IEEE Transactions on Industrial Electronics*, 64(3), 2296–2305.
 24. Xie, J., Zhang, L., Duan, L., and Wang, J. (2016). On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on Transfer Component Analysis. In *Proc. IEEE Int. Conf. Prognostics Health Manage*, pp. 1–6.
 25. Yang, B., Lei, Y., Jia, F., & Xing, S. (2019). An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing*, 122, 692–706.
 26. Inalpolat, M., & Kahraman, A. (2010). A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors. *Journal of Sound and Vibration*, 329(4), 371–393.
 27. Hong, L., Dhupia, J. S., & Sheng, S. (2014). An explanation of frequency features enabling detection of faults in equally spaced planetary gearbox. *Mechanism and Machine Theory*, 73, 169–183.
 28. Eritenel, T., & Parker, R. G. (2012). An investigation of tooth mesh nonlinearity and partial contact loss in gear pairs using a lumped-parameter model. *Mechanism and Machine Theory*, 56, 28–51.
 29. Liu, X., Yang, Y., & Zhang, J. (2018). Resultant vibration signal model-based fault diagnosis of a single stage planetary gear train with an incipient tooth crack on the sun gear. *Renewable Energy*, 122, 65–79.
 30. Luo, Y., Baddour, N., & Liang, M. (2019). Dynamical modeling and experimental validation for tooth pitting and spalling in spur gears. *Mechanical Systems and Signal Processing*, 119, 155–181.
 31. Parra, J., & Vicuña, C. M. (2017). Two methods for modeling vibrations of planetary gearboxes including faults: Comparison and validation. *Mechanical Systems and Signal Processing*, 92, 213–225.
 32. Liang, X., Zuo, M. J., & Liu, L. (2016). A windowing and mapping strategy for gear tooth fault detection of a planetary gearbox. *Mechanical Systems and Signal Processing*, 80, 445–459.
 33. Park, J., Ha, J. M., Oh, H., Youn, B. D., Choi, J., & Kim, N. H. (2016). Model-based fault diagnosis of a planetary gear: A novel approach using transmission error. *IEEE Transactions on Reliability*, 65(4), 1830–1841.
 34. Shao, S., McAleer, S., Yan, R., & Baldi, P. (2019). Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics*, 15(4), 2446–2455.
 35. Han, T., Liu, C., Yang, W., & Jiang, D. (2019). A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowledge-Based Systems*, 165, 474–487.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)