Computers in Industry 164 (2025) 104194

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.sciencedirect.com/journal/computers-in-industry

Maximum subspace transferability discriminant analysis: A new cross-domain similarity measure for wind-turbine fault transfer diagnosis

Quan Qian [a,b], Fei Wu [a,b], Yi Wang [a,b], Yi Qin [a,b,*]

[a] State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, PR China
[b] College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, PR China

ARTICLE INFO

Keywords:
Transferability discriminant analysis
Intrinsic feature extraction
Zero-label prior knowledge
Fault transfer diagnosis
Similarity measurement

ABSTRACT

In the field of fault transfer diagnosis, many approaches only focus on the distribution alignment and knowledge transfer between the source domain and target domain. However, most of these approaches ignore the precondition of whether this transfer task is transferable. Current mainstream transferability discrimination methods heavily depend on expert knowledge and are extremely vulnerable to the noise interference and variations in feature scale. This limits their applicability due to the intelligent requirements and complex industrial environment. To address the challenges mentioned previously, this paper introduces a novel cross-domain similarity measure called maximum subspace transferability discriminant analysis (MSTDA) with zero-label prior knowledge. MSTDA is comprised of a maximum subspace representation and a similarity measurement criterion. During the phase of maximum subspace representation, a new kernel-induced Hilbert space is designed to map the low-dimensional original samples into the high-dimensional space to maximize the separability of different faults and then solve the separable intrinsic fault features. Following that, a novel similarity measurement criterion that is resistant to variations in feature scale is developed. This criterion is based on the orthogonal bases of intrinsic feature subspaces. The mini-batch sampling strategy is used to ensure the timeliness of MSTDA. Finally, the experimental results on three cases, particularly in the actual wind turbine dataset, confirm that the proposed MSTDA outperforms other well-known similarity measure methods in terms of transferability evaluation. The related code can be downloaded from https://qinyi-team.github.io/2024/09/Maximum-subspace-transferability-discriminant-analysis.

1. Introduction

Owing to the rapid development of deep learning technology, there has been a significant focus on data driven intelligent fault diagnosis of mechanical equipment (Anvar and Mohammadi, 2023; Yu et al., 2023; Du et al., 2023; Qin et al., 2024). However, the assumption that the training dataset and testing dataset need to be independent and identically distributed hinders its practical application. To achieve this, transfer learning-based diagnosis methods provide a feasible solution. These methods can leverage the diagnosis knowledge from a source-domain to tackle the diagnosis task in a related but different target domain under distribution shift (Qian et al., 2023; Yang et al., 2023; He et al., 2023). Therefore, the diagnosis model, only containing the prior label knowledge of source-domain samples, can be directly applied to the unlabeled target domain.

Transfer learning technology has gained significant attention in recent years for its ability to diagnose distribution discrepancies, including the task of cross-load transfer diagnosis. Besides, the strong diagnosis performance also substantiates the superiority of transfer learning (Lei et al., 2023; Lian et al., 2024; Qian et al., 2024). The crucial aspect of deep transfer learning is to address the distribution discrepancy between two domains in order to extract domain-invariant and discriminative fault features. Methods for distribution alignment can be categorized into adversarial mechanism-based methods (Ganin et al., 2017) and distribution distance-based methods (Qian et al., 2023). In practice, the labeled fault samples in the collected historical database are extremely rare due to the following three reasons: First, actual mechanical equipment is not allowed to operate in a failure state to ensure security. Second, many industrial enterprises have been slow to adopt digital transformation, making it challenging to track label information for historical fault samples. Third, the process of annotating label information requires a significant amount of time and financial resources

* Corresponding author at: State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, PR China.
E-mail address: qy_808@cqu.edu.cn (Y. Qin).

https://doi.org/10.1016/j.compind.2024.104194
Received 22 July 2024; Received in revised form 2 September 2024; Accepted 12 September 2024
Available online 27 September 2024
0166-3615/

due to the need for disassembly inspection and the involvement of experts with specialized knowledge. To address this dilemma, transfer learning directly utilizes the entire historical dataset from different working conditions or mechanical equipment as the source domain to construct the diagnosis model. This includes approaches like multi-source transfer learning (Feng et al., 2023), partial-set transfer learning (Zhang et al., 2023), and open-set transfer learning (Yu et al., 2023). Nevertheless, most existing transfer diagnosis methods fail to consider whether the transfer tasks are transferable, i.e., without transferability analysis. If the selected source domain has a significant difference in distribution from the testing target domain, it can lead to a negative transfer phenomenon, resulting in poor diagnosis accuracy. Therefore, it is crucial to have a similarity measure that can evaluate the transferability of diagnosis knowledge between the source domain and the target domain. This measure can also assist in selecting the optimal source domain for a given target domain, ensuring the feasibility of the transfer diagnosis task.

Compared with distribution alignment, there is limited research dedicated to transferability. Existing similarity measures for cross-domain transferability discriminant analysis can be summarized into three types: feature visualization analysis, post-hoc diagnosis performance analysis, and distance metric analysis. The first approach is to map the original samples into a two-dimensional or three-dimensional feature space using dimensionality reduction technology such as autoencoder neural networks, principal component analysis, or other typical methods. Then, the overlapping ratio of feature points between two domains will be observed as the transferability criteria via human decision making. The huge dimension scaling ($R^n \to R^2/R^3$) may give rise to the loss of fault information, and the subjective decision-making process cannot ensure the reliability and normativity of discriminant results. Based on the agreement that a higher similarity score indicates improved diagnostic performance, the second approach determines the transferability of various transfer tasks by using the post-hoc diagnosis accuracy of basic classifiers such as support vector machines, convolutional neural networks, and others. While the second method removes the influence of human factors compared to the first one, it does require a significant amount of computational resources and time. Additionally, selecting the base classifier introduces another challenge. Unlike the previous two approaches, the third one is a method that measures similarity in advance and does not require auxiliary discrimination based on expert knowledge. It can be categorized as either an implicit distance metric or an explicit distance metric. The mainstream implicit distance metric mainly includes the single classifier-based A-distance (Ben-David et al., 2006), and dual classifiers “XOR” operation-based HΔH-distance (Ben-David et al., 2010), which also exist the selection issue of base classifier similar with the post-hoc diagnosis performance analysis and have a poor stability. The explicit distance metric is the most popular method to weigh the transferability between the given source domain and a specific target domain. Various methods, such as Euclidean distance, cosine distance, maximum mean discrepancy (MMD) (Gretton et al., 2012), correlation alignment (CORAL) (Sun et al., 2016), entropy-based (Belghazi et al., 2018; Jiao et al., in press), and optimal transport theory-based (Yang et al., 2021; Kolouri et al., 2019; Piccoli and Rossi, 2014), have been developed for this purpose. However, the following two disadvantages limit their application value.

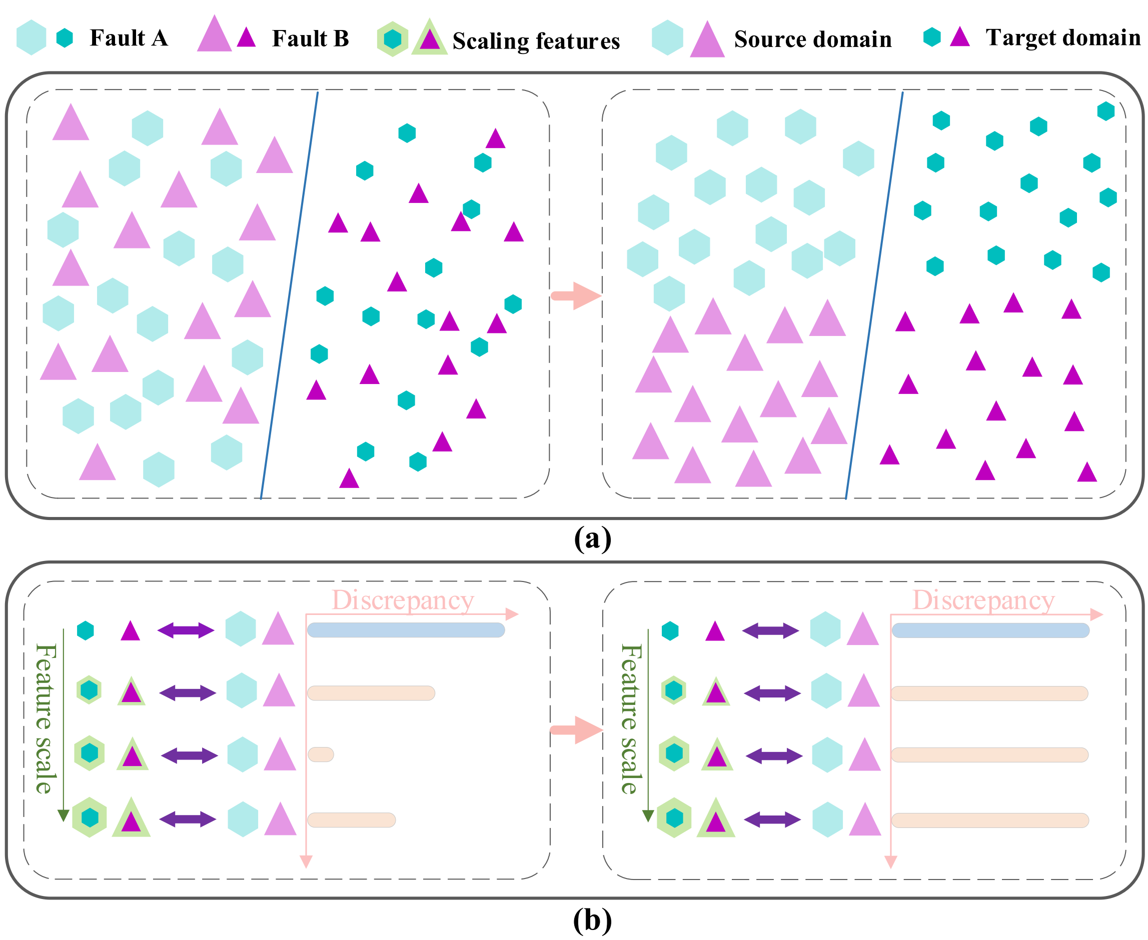

Fig. 1. Motivation demonstration: (a) intrinsic discrepancy representation; (b) similarity measure related to the feature scale. In each subfigure, the left side of the arrow indicates the disadvantages of current mainstream distance metrics in such case, and the right side represents the ideal state.

1) The prerequisite for accurate measurement of transferability is to ensure that different fault categories have good separability shown in the right side of Fig. 1(a). The better separability is conducive to measuring the intrinsic distribution discrepancy across domains more accurately. However, due to the impact of noise and other factors (Qian et al., 2023), the data samples of different fault categories are often mixed together with a poor separability, as shown in the left side of Fig. 1(a). Consequently, the current distance metrics cannot accurately assess the intrinsic discrepancy between the source domain and target domain and enhance in the left side of Fig. 1(a).
2) On the other hand, most distance metrics are sensitive to the fault feature scale. In fact, the use of nonlinear feature extractors in fault transfer diagnosis can lead to the unpredictable changes of feature scale in both source and target domains. In addition, various pre-processing methods and feature transforms can also increase the range of scale variation, as demonstrated by the federated transfer learning paradigm (Zhao et al., 2023). As shown in the left side of Fig. 1(b), if we assume that the feature scales of "Fault A" and "Fault B" in the target domain are increased, the current distance metrics will lead to the distortions of discrepancy evaluation, affecting the transferability discrimination results. Hence, a robust similarity measure is urgently needed, as illustrated in the right of Fig. 1(b).

Based on the previous discussion and analysis, it is evident that the crucial factor in obtaining the intrinsic discrepancy representation is to improve the separability between categories and establish a transferability criterion that is not dependent on the feature scale. To address these issues, this study proposes a novel similarity measure called maximum subspace transferability discriminant analysis (MSTDA). It is worth mentioning that MSTDA does not require any prior knowledge of the target-domain label, nor does it require knowledge of the source-domain label. MSTDA consists of two phases: maximum subspace representation and the similarity measurement criterion. In the first phase, a novel kernel-induced Hilbert space is developed to map the low-dimensional original samples into a high-dimensional space. This is done to maximize the separability between different faults. Then, the intrinsic low-dimensional embeddings of the high-dimensional features are resolved to achieve the most effective subspace representation. Following that, inspired by the Grassmann manifold in Ref. (Gopalan et al., 2011), the orthogonal bases of the source-domain and target-domain intrinsic feature subspaces are obtained using the singular value decomposition (SVD). Ultimately, a novel similarity measurement criterion that is resistant to variations in feature scale is developed. In addition, the mini-batch sampling strategy is employed to optimize computational efficiency and ensure timely results. The proposed MSTDA is successfully applied to actual wind turbines. The main contributions and innovations are outlined as follows:

1) A new kernel-induced high-dimensional Hilbert space is constructed for extracting the separable class features. It can ensure the final obtained maximum subspace representation possesses the intrinsic fault features.
2) A novel similarity measurement criterion is developed to eliminate the impact of feature scales. It is based on the orthogonal bases of intrinsic feature subspaces.
3) The proposed maximum subspace representation and similarity measurement criterion introduces a new similarity measure called MSTDA, which allows for transferability discriminant analysis in wind-turbine transfer tasks without any prior knowledge restrictions.

2. MSTDA similarity measure

As shown in Fig. 2, the proposed MSTDA similarity measure consists

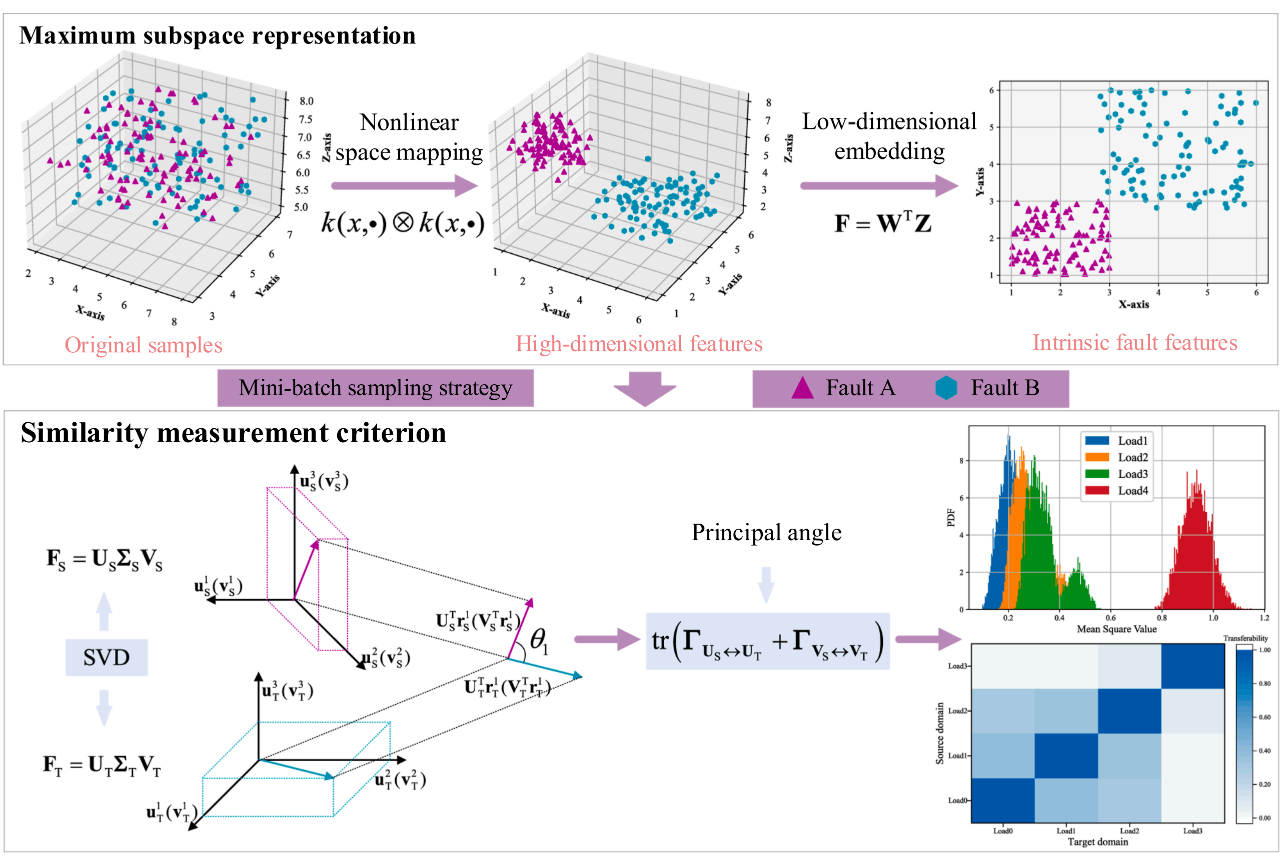

Fig. 2. Principle diagram of proposed MSTDA similarity measure.

of maximum subspace representation and a criterion for measuring similarity. The former is devoted to extract the separable intrinsic fault features, and the latter is the ultimate procedure to achieve transferability discrimination, which is robust to the feature scale. In the following subsections, we will provide a detailed introduction to these points.

2.1. Maximum subspace representation

The vibration monitoring signals, as the most common information carrier, are widely applied to the mechanical fault transfer diagnosis. The noise among them often hinders the mining of fault features, resulting in a significant decrease in the ability to distinguish between different types of faults. Therefore, ensuring the diagnosis performance and the effectiveness of transferability discrimination always depends on extracting the separable intrinsic fault features. To address the above issue, a potential solution is to consider the concept of nonlinear space mapping. This approach is supported by the pattern recognition theory (Cortes and Vapnik, 1995), which suggests that increasing the dimensionality of the space can improve the separability of features. As direct nonlinear high-dimensional space mapping will bring an explosion of computational complexity, thus the kernel-based ones attract lots of attention (Schölkopf et al., 1997).

First, the space mapping principle of kernel is introduced. Let $\mathscr{H}$ be a Hilbert space and $\mathscr{H}^*$ be its conjugate space. For each bounded function $T \in \mathscr{H}^*$, there is a unique $y_T \in \mathscr{H}$ such that (Reed, 2012):

$$T(h) = \langle h, y_T \rangle_{\mathscr{H}}, s.t. ||T|| = ||y_T|| \tag{1}$$

It can be known that the bounded linear functional in Hilbert space can be represented as the inner product of two vectors. Regarding y_T as the orthogonal bases of above Hilbert space, any functions in the space can been viewed as a projection on the bases. Now, how to find such a set of orthogonal bases becomes the key. Fortunately, the Mercer's theorem (Mercer, 1909) provides a feasible solution, whose formula is written as:

$$k\left(x, y\right) = \langle k(x, \cdot), k(\cdot, y)\rangle = \sum_{i=1}^{\infty} \lambda_i \psi_i(x)\psi_i(y) \tag{2}$$

$$\int k(x, y)\psi_i(y)dy = \lambda_i \psi_i(x) \tag{3}$$

where $k(x, y)$ is a continuous symmetric but non-negative function, λ_i denotes the non-negative eigenvalue, and $\psi_i(\cdot)$represents the orthogonal eigenfunction. Via the kernel function $\kappa(x, \cdot) = \left\{\sqrt{\gamma_i}\varphi_i(x)\right\}_{i=1,2,\cdots\infty}$, the low-dimensional data samples can be mapped to the high-dimensional Hilbert space, thereby enhancing the separability between different faults.

According to the mechanical vibration characteristics, the monitoring vibration signal is almost symmetric along the x-axis, it then follows that a new kernel is designed based on energy index (mean square value), i.e., $k(\boldsymbol{x},) \otimes k(\boldsymbol{x},)$, in which the tensor product form is employed for simplifying the calculation amount through the following theory:

$$\langle \phi_1 \otimes \phi_2, \varphi_1 \otimes \varphi_2 \rangle_{\mathscr{H}} = \langle \phi_1, \varphi_1 \rangle_{\mathscr{H}_1} \cdot \langle \phi_2, \varphi_2 \rangle_{\mathscr{H}_2} \tag{4}$$

where $\mathscr{H}$, $\mathscr{H}_1$, and $\mathscr{H}_2$ are three Hilbert spaces, $\phi_1, \varphi_1 \in \mathscr{H}_1$, and $\phi_2, \varphi_2 \in \mathscr{H}_2$. It can be concluded from Eq. (4) that the tensor product of two kernels ($k(\boldsymbol{x},)$,$k(\boldsymbol{x},)$) still possesses the property of kernel function. Then, using Ref. (Mercer, 1909), the kernel can be used to span Hilbert space $\mathscr{H}$:

$$\mathscr{H} = span\{k(\boldsymbol{x},) \otimes k(\boldsymbol{x},) | \boldsymbol{x} \in \mathbf{X}\} \tag{5}$$

where $\mathbf{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2 \cdots, \boldsymbol{x}_n) \in \mathbb{R}^{d \times n}$ is the sample set. The low-dimensional samples can be projected to the infinite-dimensional space $\mathscr{H}$ for maximumly enhancing the separability. Unfortunately, the high-dimensional representations pose a challenge due to the sparsity of data samples, making it difficult to measure similarity. For instance, assume that the sample number satisfying dense sampling is 100 in an attribute dimension, the feature space, including n attribution dimensions, needs 100 ^{n} samples. Therefore, it is crucial to have a subspace that contains the separable feature representation from the high-dimensional Hilbert space mentioned above, that is, solving an orthogonal feature transform matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2 \cdots, \mathbf{w}_3) \in \mathbb{R}^{d' \times d''}$, which is by:

$$\mathbf{f} = \mathbf{W}^T k(\boldsymbol{x},) \otimes k(\boldsymbol{x},) = \mathbf{W}^T k(\boldsymbol{x},)^{\otimes 2} \tag{6}$$

where the $k(\boldsymbol{x},)^{\otimes 2} \in \mathbb{R}^{d' \times 1}$ and $\mathbf{f} \in \mathbb{R}^{d'' \times 1}$ represent the high-dimension features and intrinsic fault features in Fig. 2, respectively. Using the maximum reconfigurability, the optimization objective can be written as follows:

$$\min_{\mathbf{W}} \sum_{i=1}^{n} \left\| \mathbf{W}\mathbf{f}_i - k(\boldsymbol{x}_i,)^{\otimes 2} \right\|_2^2 s.t. \mathbf{W}^T\mathbf{W} = \mathbf{I} \tag{7}$$

where $\mathbf{I} \in \mathbb{R}^{d'' \times d''}$ denotes the identity matrix. Then, the $\sum_{i=1}^{n} \left\| \mathbf{W}\mathbf{f}_i - k(\boldsymbol{x}_i,)^{\otimes 2} \right\|_2^2$ is simplified by:

$$\begin{aligned}
&\sum_{i=1}^{n} \left\| \mathbf{W}\mathbf{f}_i - k(\boldsymbol{x}_i,)^{\otimes 2} \right\|_2^2 \\
&= \sum_{i=1}^{n} \left(-\left(k(\boldsymbol{x}_i,)^{\otimes 2}\right)^T \mathbf{W}\mathbf{W}^T k(\boldsymbol{x}_i,)^{\otimes 2} + \left(k(\boldsymbol{x}_i,)^{\otimes 2}\right)^T k(\boldsymbol{x}_i,)^{\otimes 2} \right) \\
&= \sum_{i=1}^{n} -\left\| \mathbf{W}^T k(\boldsymbol{x}_i,)^{\otimes 2} \right\|_2^2 + \left(k(\boldsymbol{x}_i,)^{\otimes 2}\right)^T k(\boldsymbol{x}_i,)^{\otimes 2}
\end{aligned} \tag{8}$$

It can be seen from Eq. (7) that the optimization objective is related only to the term containing $\mathbf{W}$, Eq. (7) is rewritten as follows:

$$\begin{aligned}
&-\min_{\mathbf{W}} \; tr\left(\mathbf{W}^T k(\mathbf{X},)^{\otimes 2} \left(k(\mathbf{X},)^{\otimes 2}\right)^T \mathbf{W}\right) \\
&s.t. \mathbf{W}^T\mathbf{W} = \mathbf{I}
\end{aligned} \tag{9}$$

where the $k(\mathbf{X},)^{\otimes 2}$ is centralized:

$$k(\mathbf{X},)^{\otimes 2} \leftarrow k(\mathbf{X},)^{\otimes 2} \left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \tag{10}$$

where $\mathbf{1} \in \mathbb{R}^{n \times 1}$ is a column vector with all elements equal to 1.

Furthermore, Using Ref. (Anstreicher and Wolkowicz, 2000), the Lagrange function of the final optimization objective can be represented:

$$\begin{aligned}
&L(\mathbf{W}, \boldsymbol{\Delta}) = \\
&-tr\left(\mathbf{W}^T k(\mathbf{X},)^{\otimes 2} \left(k(\mathbf{X},)^{\otimes 2}\right)^T \mathbf{W}\right) + tr\left(\boldsymbol{\Delta}^T \left(\mathbf{W}^T\mathbf{W} - \mathbf{I}\right)\right)
\end{aligned} \tag{11}$$

The above parameter $\boldsymbol{\Delta} = diag(\eta_1, \eta_2 \cdots, \eta_{d''}) \in \mathbb{R}^{d'' \times d''}$ denotes the Lagrange multiplier matrix. Then, take the derivative with respect to $\mathbf{W}$:

$$\begin{aligned}
&\frac{\partial L(\mathbf{W}, \boldsymbol{\Delta})}{\partial \mathbf{W}} \\
&= \frac{\partial}{\partial \mathbf{W}} \left[-tr\left(\mathbf{W}^T k(\mathbf{X},)^{\otimes 2} \left(k(\mathbf{X},)^{\otimes 2}\right)^T \mathbf{W}\right) + tr\left(\boldsymbol{\Delta}^T \left(\mathbf{W}^T\mathbf{W} - \mathbf{I}\right)\right) \right] \\
&= -\frac{\partial}{\partial \mathbf{W}} tr\left(\mathbf{W}^T k(\mathbf{X},)^{\otimes 2} \left(k(\mathbf{X},)^{\otimes 2}\right)^T \mathbf{W}\right) + \frac{\partial}{\partial \mathbf{W}} tr\left(\boldsymbol{\Delta}^T \mathbf{W}^T \mathbf{W}\right) \\
&= -2k(\mathbf{X},)^{\otimes 2} \left(k(\mathbf{X},)^{\otimes 2}\right)^T \mathbf{W} + 2\mathbf{W}\boldsymbol{\Delta}
\end{aligned} \tag{12}$$

Making $\frac{\partial L(\mathbf{W}, \boldsymbol{\Delta})}{\partial \mathbf{W}} = 0$, the following equation can be obtained:

$$\begin{aligned}
&k(\mathbf{X},)^{\otimes 2} \left(k(\mathbf{X},)^{\otimes 2}\right)^T \mathbf{W} = \mathbf{W}\boldsymbol{\Delta} \\
&\Rightarrow k(\mathbf{X},)^{\otimes 2} \left(k(\mathbf{X},)^{\otimes 2}\right)^T \mathbf{w}_i = \eta_i \mathbf{w}_i, \quad i \in \{1, 2, \cdots d''\}
\end{aligned} \tag{13}$$

It is obvious from Eq. (13) that $\mathbf{w}_i \in \mathbf{W}$ and $\eta_i \in \boldsymbol{\Delta}$ are the eigenvector and eigenvalue of feature covariance matrix $k(\mathbf{X},)^{\otimes 2} \left(k(\mathbf{X},)^{\otimes 2}\right)^T$, respectively. Unfortunately, the feature covariance matrix cannot be obtained explicitly. In order to solve $k(\mathbf{X},)^{\otimes 2}\left(k(\mathbf{X},)^{\otimes 2}\right)^T$, Eq. (13) is transformed:

$$\left(k(\mathbf{X},)^{\otimes 2}\right)^T k(\mathbf{X},)^{\otimes 2}\boldsymbol{\beta}^i = \eta_i \boldsymbol{\beta}^i \tag{14}$$

where the $\boldsymbol{\beta}^j = (\beta_1^i, \beta_2^i \cdots \beta_n^i) \in \mathbb{R}^{n\times 1}$ with $\beta_j^i = \frac{1}{\eta_i}\left(k(\mathbf{x}_j,)^{\otimes 2}\right)^T \mathbf{w}_i$. According to Eq. (1) and the kernel tricks, Eq.(14) can be rewritten as:

$$\mathbf{K}^{\otimes 2}\boldsymbol{\beta}^i = \eta_i \boldsymbol{\beta}^i \tag{15}$$

where $\mathbf{K}^{\otimes 2}$ represents the kernel matrix with respect to $k^2(\cdot,\cdot)$, that is $\left(\mathbf{K}^{\otimes 2}\right)_{ij} = k^2(\mathbf{x}_i, \mathbf{x}_j), \quad i,j \in \{1,2,\cdots n\}$. Finally, $\boldsymbol{\beta}^i$and η_i can be easily solved through the eigenvalue decomposition. The intrinsic fault features on the maximum subspace derived from $\mathbf{W}$ are obtained:

$$\begin{aligned}\{\mathbf{f}_i\}_{i=1,2\cdots d''} &= \mathbf{w}_i^T k(\mathbf{x},)^{\otimes 2} = \sum_{j=1}^{n}\beta_j^i\left(k(\mathbf{x}_i,)^{\otimes 2}\right)^T k(\mathbf{x},)^{\otimes 2} \\ &= \sum_{j=1}^{n}\beta_j^i k^2\left(\mathbf{x}_i, \mathbf{x}\right)\end{aligned} \tag{16}$$

It can be seen that the dimension d'' of the obtained subspace will significantly affect the feature attributes, which indicates that too low or too high dimension will cause the loss of fault information or the curse of dimensionality. As η_i represents the information importance, the 80% threshold ratio is set to seek the optimal subspace dimension.

$$d'' = \min_{d^*}\left\{d^*\middle|\sum\nolimits_{i=1}^{d^*}\eta_i \Big/ \sum\nolimits_{i=1}^{n}\eta_i \geq 80\%\right\} \tag{17}$$

2.2. Similarity measurement criterion

Given the sourced-domain and target-domain samples, $\mathbf{X}_S \in \mathbb{R}^{d\times k}$ and $\mathbf{X}_T \in \mathbb{R}^{d\times m}$, the intrinsic fault features can be extracted using the above maximum subspace representation, i.e., $\mathbf{F}_S = MSR(\mathbf{X}_S) \in \mathbb{R}^{d''\times k}$ and $\mathbf{F}_T = MSR(\mathbf{X}_T) \in \mathbb{R}^{d''\times m}$, where $k$and m denote the sample number of source domain and target domain, respectively. Then, the orthogonal bases of intrinsic feature subspaces can be solved using the SVD algorithm:

$$\mathbf{F}_S = \mathbf{U}_S\boldsymbol{\Sigma}_S(\mathbf{V}_S)^T, \mathbf{F}_T = \mathbf{U}_T\boldsymbol{\Sigma}_T(\mathbf{V}_T)^T \tag{18}$$

where $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2\cdots)$ and $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2\cdots)$are the orthogonal bases of feature matrix space, and $\boldsymbol{\Sigma} = diag(\sigma_1, \sigma_2\cdots)$ denotes the singular value. It should be noted that the dimension of these source-domain components is related to the d'' and k:

$$\begin{aligned}&\mathbf{U}_S \in \mathbb{R}^{d''\times d''}, \boldsymbol{\Sigma}_S \in \mathbb{R}^{d''\times d''}, \mathbf{V}_S \in \mathbb{R}^{k\times d''}; d'' < k \\ &\mathbf{U}_S \in \mathbb{R}^{d''\times k}, \boldsymbol{\Sigma}_S \in \mathbb{R}^{k\times k}, \mathbf{V}_S \in \mathbb{R}^{k\times k}; d'' \geq k\end{aligned} \tag{19}$$

The dimension calculation of the target domain is the same as the above Eq. (19). The Frobenius norm is used to represent the scale of the feature matrix. Taking the source domain as an example, the following equation can be obtained with $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ and $d'' \geq k$:

$$\begin{aligned}\|\mathbf{F}_S\|_{\mathscr{F}}^2 &= tr\left(\mathbf{F}_S(\mathbf{F}_S)^T\right) = tr\left(\mathbf{U}_S\boldsymbol{\Sigma}_S(\mathbf{V}_S)^T\mathbf{V}_S(\boldsymbol{\Sigma}_S)^T(\mathbf{U}_S)^T\right) \\ &= tr\left(\mathbf{U}_S\boldsymbol{\Sigma}_S(\boldsymbol{\Sigma}_S)^T(\mathbf{U}_S)^T\right) = tr\left((\boldsymbol{\Sigma}_S)^T\boldsymbol{\Sigma}_S\right) = \sum_{i=1}^{k}\sigma_i^2\end{aligned} \tag{20}$$

It can be observed from Eq. (20) that the feature scale is only influenced by a singular value. Thus, an intuitional similarity measurement criterion, aiming to weigh the discrepancy between orthogonal bases of two domains, is designed, i.e., angle measure. Taking the $\mathbf{V}_S \leftrightarrow \mathbf{V}_T$ as an example with $d'' < k$, the cosine vector $\cos\boldsymbol{\Phi}_{\mathbf{V}_S\leftrightarrow\mathbf{V}_T} = (\cos\phi_1, \cos\phi_2\cdots \cos\phi_k)$ on each dimension of source-target feature subspace pair $(\mathbf{V}_S \leftrightarrow \mathbf{V}_T)$, can be defined as:

$$\begin{aligned}\cos\phi_1 &= \max_{\mathbf{v}_{S_1}\in\mathbf{V}_S, \mathbf{v}_{T_1}\in\mathbf{V}_T}\left(\mathbf{v}_{S_1}(\mathbf{v}_{T_1})^T\right)/(\|\mathbf{v}_{S_1}\|\|\mathbf{v}_{T_1}\|) \\ \cos\phi_2 &= \max_{\substack{\mathbf{v}_{S_2}\in\mathbf{V}_S, \mathbf{v}_{T_2}\in\mathbf{V}_T \\ \mathbf{v}_{S_2}\neq\mathbf{v}_{S_1}, \mathbf{v}_{T_2}\neq\mathbf{v}_{T_1}}}\left(\mathbf{v}_{S_2}(\mathbf{v}_{T_2})^T\right)/(\|\mathbf{v}_{S_2}\|\|\mathbf{v}_{T_2}\|) \\ &\;\;\vdots \\ \cos\phi_k &= \max_{\substack{\mathbf{v}_{S_k}\in\mathbf{V}_S, \mathbf{v}_{T_k}\in\mathbf{V}_T \\ \mathbf{v}_{S_k}\neq\mathbf{v}_{S_1}\cdots\neq\mathbf{v}_{S_{k-1}} \\ \mathbf{v}_{T_k}\neq\mathbf{v}_{T_1}\cdots\neq\mathbf{v}_{T_{k-1}}}}\left(\mathbf{v}_{S_k}(\mathbf{v}_{T_k})^T\right)/(\|\mathbf{v}_{S_k}\|\|\mathbf{v}_{T_k}\|)\end{aligned} \tag{21}$$

Similarly, the cosine vector $\cos\boldsymbol{\Phi}_{\mathbf{U}_S\leftrightarrow\mathbf{U}_T}$ on $\mathbf{U}_S \leftrightarrow \mathbf{U}_T$ can be calculated. When $\cos\boldsymbol{\Phi}_{\mathbf{U}_S\leftrightarrow\mathbf{U}_T}$ and $\cos\boldsymbol{\Phi}_{\mathbf{V}_S\leftrightarrow\mathbf{V}_T}$ are all one, the similarity between source and target domains is the highest, which also indicates that the current transfer task possesses a good transferability. It is obvious that the (18) has a high computation complexity. Thus, the principal angles $\boldsymbol{\Theta}_{\mathbf{U}_S\leftrightarrow\mathbf{U}_T}(\boldsymbol{\Theta}_{\mathbf{V}_S\leftrightarrow\mathbf{V}_T})$ between $\mathbf{U}_S(\mathbf{V}_S)$ and $\mathbf{U}_T(\mathbf{V}_T)$ is used to replace $\boldsymbol{\Phi}_{\mathbf{U}_S\leftrightarrow\mathbf{U}_T}(\boldsymbol{\Phi}_{\mathbf{V}_S\leftrightarrow\mathbf{V}_T})$, which can solved by SVD (Gong et al., 2012):

$$(\mathbf{U}_S)^T\mathbf{U}_S = \mathbf{R}_S\boldsymbol{\Gamma}_{\mathbf{U}_S\leftrightarrow\mathbf{U}_T}(\mathbf{R}_T)^T \tag{22}$$

where $\boldsymbol{\Gamma}_{\mathbf{U}_S\leftrightarrow\mathbf{U}_T} = diag(\cos\boldsymbol{\Theta}_{\mathbf{U}_S\leftrightarrow\mathbf{U}_T})$, and the $(\mathbf{U}_S)^T\mathbf{r}_S^i$ denotes the principal vector shown in Fig. 2. Finally, the similarity measurement criterion to weigh the transferability under a specific transfer task is defined as follows:

$$sim\left(\mathscr{D}_S, \mathscr{D}_T\right) = tr\left(\boldsymbol{\Gamma}_{\mathbf{U}_S\leftrightarrow\mathbf{U}_T} + \boldsymbol{\Gamma}_{\mathbf{V}_S\leftrightarrow\mathbf{V}_T}\right) \tag{23}$$

where $\mathscr{D}_S$ and $\mathscr{D}_T$ represent the source domain and target domain, respectively.

3. Algorithm overview

Returning to the initial issue, the task of selecting the most suitable diagnosis knowledge from multiple source domains for a specific target domain can be easily resolved using the similarity measure mentioned above. For a given target domain $\mathscr{D}_T$ and several source domains $\{\mathscr{D}_{S_1}, \mathscr{D}_{S_2}\cdots, \mathscr{D}_{S_p}\}$, the optimal source domain $\mathscr{D}_S{}^*$ can be defined as follows:

$$\mathscr{D}_S{}^* = \max_{\mathscr{D}_{S_i}}\left\{\mathscr{D}_{S_i}\middle|\left\{sim\left(\mathscr{D}_{S_i}, \mathscr{D}_T\right)\right\}_{i=1,2\cdots p}\right\} \tag{24}$$

To guarantee the real-time capability of transferability discrimination, the mini-batch sampling strategy is utilized. The algorithm overview of the proposed MSTDA is presented in Table 1.

4. Experiment study

In this section, three transferability discrimination experimental cases are conducted to test the universality of the proposed MSTDA. These cases include simulated transfer tasks, testbed transfer tasks, and actual wind turbine transfer tasks. In addition, MSTDA is evaluated against various distance metrics, such as MMD (Gretton et al., 2012), CORAL (Sun et al., 2016), A-distance (Ben-David et al., 2006), AHMM (Feng et al., 2023), and cosine distance, to evaluate its effectiveness. As some distances (MMD, CORAL, A-distance, and AHMM) have an opposite trend with MSTDA in similarity measurement, this is, the larger value means a smaller transferability, we use their reciprocals as the measurement criterion. All similarity measure methods are normalized into a range of [0,1] to ensure a consistent scale across multiple source-target pairs.

$$s_j = \frac{sim\left(\mathscr{D}_{S_j}, \mathscr{D}_T\right) - \min\left\{sim\left(\mathscr{D}_{S_i}, \mathscr{D}_T\right)\right\}_{i=1,2\cdots p}}{\max\left\{sim\left(\mathscr{D}_{S_i}, \mathscr{D}_T\right)\right\}_{i=1,2\cdots p} - \min\left\{sim\left(\mathscr{D}_{S_i}, \mathscr{D}_T\right)\right\}_{i=1,2\cdots p}} \tag{25}$$

Furthermore, three indexes, including correlation (CORR),

Table 1
The Algorithm overview of proposed MSTDA.

Input: Several source domains $\left\{\mathcal{D}_{S_1}, \mathcal{D}_{S_2} \cdots, \mathcal{D}_{S_p}\right\}$, a given target domain $\mathcal{D}_T$, mini-batch size n.

Training:

Split several source-target pairs $\left\{\left(\mathcal{D}_{S_i}, \mathcal{D}_T\right)\right\}_{i=1,2\cdots,p}$;

For $\left\{\left(\mathcal{D}_{S_i}, \mathcal{D}_T\right)\right\}_{i=1,2\cdots,p}$ **do**:

1. Take n samples from source domain and target domain of $\left(\mathcal{D}_{S_i}, \mathcal{D}_T\right)$ randomly and respectively, $\mathbf{X}_S = \left\{\boldsymbol{x}_{S_i}\right\}_{i=1,2\cdots n}$ and $\mathbf{X}_T = \left\{\boldsymbol{x}_{T_i}\right\}_{i=1,2\cdots n}$;
2. Calculate the kernel matrix $\mathbf{K}^{\otimes 2}$: $\left(\mathbf{K}^{\otimes 2}\right)_{ij} = k^2(\boldsymbol{x}_{S_i}, \boldsymbol{x}_{T_j})$, $\mathrm{i}, j \in \{1, 2, \cdots \mathrm{n}\}$;
3. Implement eigenvalue decomposition on kernel matrix: $\mathbf{K}^{\otimes 2}\boldsymbol{\beta}^i = \eta_i \boldsymbol{\beta}^i$;
4. Solve the dimension of feature intrinsic subspace: $\mathrm{d}'' = \min_{\mathrm{d}^*}\left\{\mathrm{d}^* \left| \sum_{i=1}^{\mathrm{d}^*}\eta_i \middle/ \sum_{i=1}^{n}\eta_i \geq 80\% \right.\right\}$;
5. Obtain the projection matrix $\mathcal{B} = (\boldsymbol{\beta}^1, \boldsymbol{\beta}^2 \cdots, \boldsymbol{\beta}^{\mathrm{d}''})$;
6. Extract the intrinsic fault features of source domain and target domain, $\mathbf{F}_S$ and $\mathbf{F}_T$;
7. Decompose the orthogonal bases of feature subspaces: $\mathbf{F}_S = \mathbf{U}_S\boldsymbol{\Sigma}_S\left(\mathbf{V}_S\right)^T$, $\mathbf{F}_T = \mathbf{U}_T\boldsymbol{\Sigma}_T\left(\mathbf{V}_T\right)^T$;
8. Solve the similarity between $\mathcal{D}_{S_i}$ and $\mathcal{D}_T$: $sim(\mathcal{D}_{S_i}, \mathcal{D}_T) = \mathrm{tr}\left(\boldsymbol{\Gamma}_{\mathbf{U}_{S_i} \leftrightarrow \mathbf{U}_T} + \boldsymbol{\Gamma}_{\mathbf{V}_{S_i} \leftrightarrow \mathbf{V}_T}\right)$;

end

Output: all similarity $\left\{sim(\mathcal{D}_{S_i}, \mathcal{D}_T)\right\}_{i=1,2\cdots,p}$ and the optimal source domain $\mathcal{D}_S^*$.

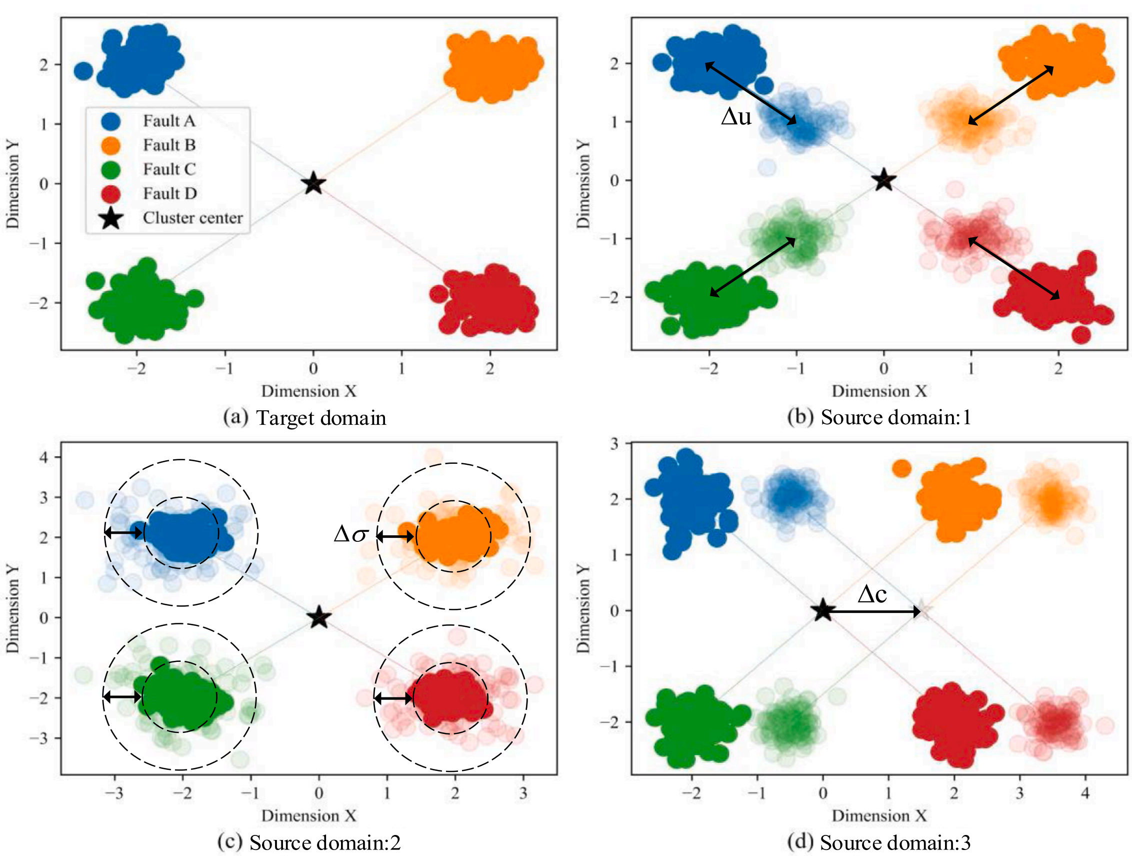

Fig. 3. Three source domain simulated schemes: (a) cluster scaling; (b) cluster variance change; (c) cluster center shift.

consistency (CONS), and monotonicity (MONO), are used to evaluate the comprehensive performance of each similarity measure method, where a larger value indicates a better transferability discriminant result. Given the reference benchmark of transferability discrimination for all source-target transfer tasks $\{\tau_i\}_{i=1,2\cdots,p}$, $\tau_i < \tau_{i+1}$ *or* $\tau_i > \tau_{i+1}$ and the corresponding similarity values of a specific method $\{s_i\}_{i=1,2\cdots,p}$, the above indexed can be defined as follows:

$$COR(\tau, s) = \frac{\sum_{i=1}^{p} (\tau_i - \overline{\tau})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{p} (\tau_i - \overline{\tau})^2 \sum_{i=1}^{p} (s_i - \overline{s})^2}} \tag{26}$$

$$CON(\tau, s) = \frac{\sum_{i=1}^{p} (\tau_i - \overline{\tau})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{p} (\tau_i - \overline{\tau})^2 (s_i - \overline{s})^2}} \tag{27}$$

$$MON(s) = \frac{1}{p-1} \left(\sum_{i=1}^{p} \delta(s_{i+1} - s_i) - \sum_{i=1}^{p} \delta(s_i - s_{i+1}) \right) \tag{28}$$

where $\delta(\cdot)$ denotes the simple unit step function. The reference benchmark is the preset variable parameter in a simulated experimental case, while the remaining cases focus on the diagnosis accuracy.

4.1. Transferability discrimination on simulated transfer tasks

As shown in Fig. 3, for a given target domain including four fault types, the two-dimension Gaussian distribution is employed to simulate three source domain schemes, such as cluster scaling (Δu), cluster variance change (Δσ), and cluster center shift (Δc). Five source domains are simulated in each scheme. In each transfer task, the sample number for both the source domain and target domain is set to 4 ×100, and the dimension is 2. To ensure the credibility of transferability results, the experiment on each transfer task is repeated five times. This helps to address any potential instability caused by the mini-batch sampling strategy. Additionally, all comparative methods maintain the same experimental setting to ensure fairness.

The experimental results of the three simulated schemes are listed in Tables 2–4, including the mean and standard deviation. It can be clearly observed that our proposed MSTDA similarity measure comprehensively outperforms other methods. In terms of cluster scaling and cluster center shift schemes, MSTDA has shown the best experimental performance among the three evaluation indexes. These comparative methods (MMD, CORAL, A-distance, and AHMM) are sensitive to the feature scale and lack the ability to extract separable intrinsic fault features. As a result, their transferability results are inferior to those of the MSTDA. Considering that the Cosine distance is easily affected by the vector orientation

Table 2
Transferability results of cluster scaling.

Variable (Δu)	**Methods**					
	MMD	CORAL	A-distance	L2-distance	Cosine	MSTDA
0.28	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.52 ± 0.37	1.00 ± 0.00
0.57	0.14 ± 0.04	0.43 ± 0.05	0.60 ± 0.15	0.54 ± 0.01	0.94 ± 0.08	0.85 ± 0.04
0.85	0.04 ± 0.00	0.18 ± 0.09	0.23 ± 0.10	0.22 ± 0.02	0.46 ± 0.39	0.61 ± 0.11
1.13	0.01 ± 0.00	0.11 ± 0.01	0.05 ± 0.02	0.07 ± 0.00	0.39 ± 0.42	0.31 ± 0.05
1.41	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.34 ± 0.46	0.00 ± 0.00
CORR	0.79	0.92	0.96	0.95	0.60	**1.00**
CONS	1.32	1.54	1.69	1.64	1.58	**1.73**
MONO	**1.00**	**1.00**	**1.00**	**1.00**	0.50	**1.00**

Table 3
Transferability results of cluster variance change.

Variable (Δσ)	**Methods**					
	MMD	CORAL	A-distance	L2-distance	Cosine	MSTDA
0.05	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	0.95 ± 0.04	1.00 ± 0.00
0.15	0.05 ± 0.02	0.35 ± 0.07	0.44 ± 0.10	0.44 ± 0.04	0.85 ± 0.17	0.91 ± 0.02
0.35	0.03 ± 0.02	0.17 ± 0.02	0.14 ± 0.06	0.16 ± 0.03	0.55 ± 0.41	0.71 ± 0.03
0.55	0.00 ± 0.00	0.20 ± 0.02	0.07 ± 0.03	0.07 ± 0.04	0.46 ± 0.30	0.41 ± 0.02
0.75	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.34 ± 0.46	0.00 ± 0.00
CORR	0.66	0.82	0.87	0.87	0.98	**0.99**
CONS	1.23	1.47	1.60	1.61	**1.80**	1.69
MONO	0.50	0.50	**1.00**	**1.00**	**1.00**	**1.00**

Table 4
Transferability results of cluster center shift.

Variable (Δc)	**Methods**					
	MMD	CORAL	A-distance	L2-distance	Cosine	MSTDA
0.5	1.00 ± 0.00	0.43 ± 0.32	1.00 ± 0.00	1.00 ± 0.00	0.33 ± 0.47	1.00 ± 0.00
1.0	0.09 ± 0.02	0.57 ± 0.42	0.19 ± 0.07	0.44 ± 0.04	0.41 ± 0.30	0.85 ± 0.02
1.5	0.02 ± 0.00	0.20 ± 0.10	0.08 ± 0.03	0.16 ± 0.03	0.50 ± 0.07	0.59 ± 0.01
2.0	0.04 ± 0.02	0.21 ± 0.17	0.00 ± 0.00	0.07 ± 0.04	0.65 ± 0.23	0.30 ± 0.01
2.5	0.00 ± 0.00	0.69 ± 0.44	0.04 ± 0.01	0.00 ± 0.00	0.33 ± 0.47	0.00 ± 0.00
CORR	0.67	0.91	0.73	0.80	−0.24	**1.00**
CONS	1.26	−0.27	1.35	1.58	−0.62	**1.73**
MONO	0.50	−0.50	0.50	**1.00**	−0.50	**1.00**

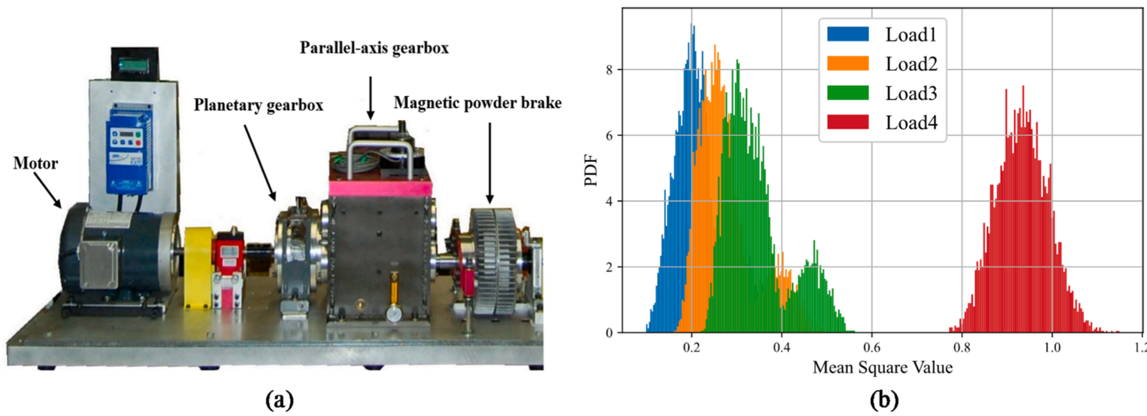

Fig. 4. DDS planetary gearbox testbed: (a) testbed structure; (b) data distribution of different load information.

Table 5
Transferability results of planetary gearbox testbed transfer tasks.

Transfer tasks	Accuracy benchmark	Methods					
		MMD	CORAL	A-distance	L2-distance	Cosine	MSTDA
1↔2	0.77	0.93 ± 0.06	0.82 ± 0.25	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
2↔3	0.71	0.99 ± 0.02	0.57 ± 0.05	0.70 ± 0.08	0.62 ± 0.02	0.78 ± 0.02	0.91 ± 0.00
1↔3	0.61	0.49 ± 0.02	0.81 ± 0.13	0.76 ± 0.19	0.81 ± 0.08	0.86 ± 0.02	0.85 ± 0.01
3↔4	0.42	0.97 ± 0.02	0.00 ± 0.00	0.49 ± 0.08	0.01 ± 0.01	0.00 ± 0.00	0.14 ± 0.01
2↔4	0.37	0.58 ± 0.03	0.04 ± 0.02	0.06 ± 0.08	0.01 ± 0.01	0.13 ± 0.01	0.10 ± 0.02
1↔4	0.34	0.00 ± 0.00	0.05 ± 0.02	0.05 ± 0.04	0.11 ± 0.01	0.22 ± 0.01	0.00 ± 0.00
CORR		0.60	0.91	0.93	0.93	0.93	**0.98**
CONS		1.34	2.25	1.98	2.14	2.23	**2.24**
MONO		0.20	−0.20	0.60	−0.20	−0.20	**1.00**

and the data sample used is randomly generated, thus its stability is significantly weaker than that of other measurement methods. As shown in Table 3, while the MSTDA consistency is slightly lower than the Cosine distance, its standard deviation is significantly higher. This indicates that MSTDA has better stability.

4.2. Transferability discrimination on testbed transfer tasks

The data samples are collected by the DDS planetary gearbox testbed. The testbed structure is shown in Fig. 4(a), consisting of a motor, a planetary gearbox, a parallel-axis gearbox, and a magnetic powder brake. The acceleration sensor is positioned on the shell of the planetary gearbox, with a sampling frequency of 5120 Hz. Five healthy types (normal condition, chipped tooth, missing tooth, surface wear, and root crack) are simulated in the second-stage sun gear of the planetary gearbox. Four different load information can be obtained by controlling the magnetic powder brake: 0 N·m (Load1), 1.4 N·m (Load2), 2.8 N·m (Load3), and 25.2 N·m (Load4). The output speed of the motor is 1500 min/r, and the sample dimension and sample number of each load are set to 3072 and 4 ×1000. Besides, the basic kernel function in

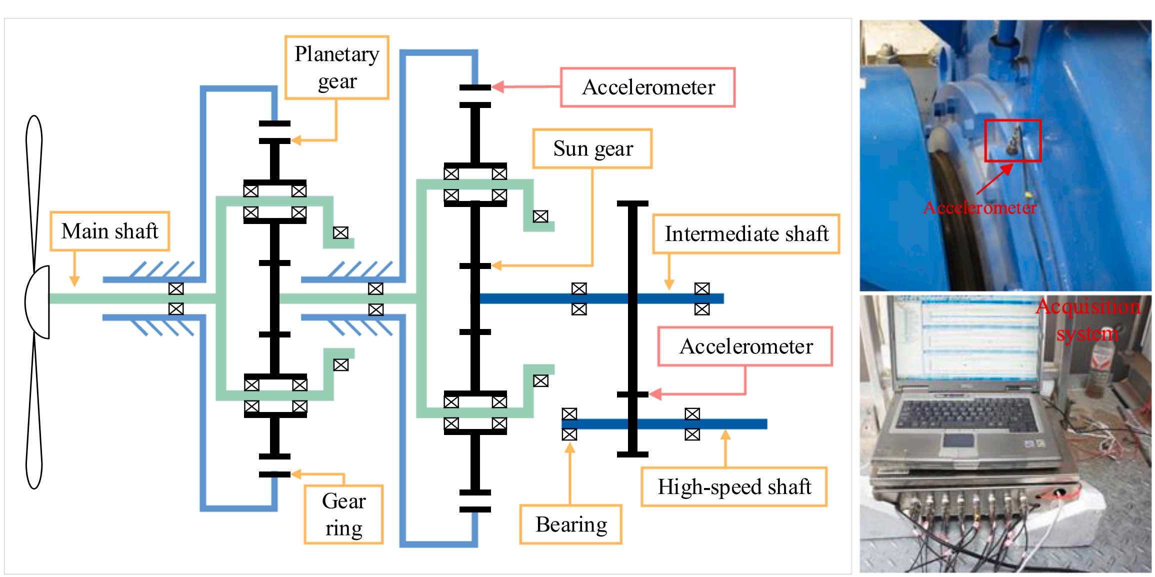

Fig. 5. Gearbox transmission structure of actual wind-turbine.

Table 1 is set to the Gaussian kernel by expertise and data characteristics.

Using the above collected data samples, six transfer tasks shown in Table 5 (1↔2, 2↔3, 1↔3, 3↔4, 2↔4 and 1↔4) are built. Unlike the simulated transfer tasks mentioned previously, the transferability discrimination of testbed transfer tasks poses greater difficulty due to various factors such as noise environment, transmission path, machining error, and assembly error. Since the vast majority of fault transfer diagnosis methods employ convolutional neural networks (CNN) as the basic classifier in the current fault diagnosis community, we also employ CNN to obtain the benchmark accuracy in Tables 5 and 7, in which the raw vibration signals are fed directly into the network without feature preprocessing. The structure parameters of CNN are same as Ref. (Qian et al., 2023). In order to better reflect the transferability, CNN is only trained by source-domain data samples without any transfer learning techniques, then the target-domain data samples are inputted into the trained CNN for testing the diagnosis accuracy. Moreover, as the accuracies of different diagnosis tasks are different, the average diagnosis accuracy of "1→2" and "2→1" diagnosis tasks is used as the final accuracy benchmark. Taking the "1↔2" as an instance, the former "1" and the latter "2" respectively denote the labeled source domain and unlabeled target domain in the "1→2" and the same goes for the "2→1". Considering the random volatility of deep neural networks, each transfer task is executed for ten times. From Fig. 4(b), it is evident that the data distribution of Load4 is significantly different from the other three types of loads. This suggests that the transferability of the transfer task involving Load4 is lower compared to the other transfer tasks. This point is verified by the accuracy benchmark listed in Table 5. The experimental setting for all similarity measure methods is consistent with Section 4.1. Finally, Table 5 clearly demonstrates that MSTDA outperforms other methods in terms of transferability performance through the three evaluation indexes.

4.3. Transferability discrimination on actual wind-turbine transfer tasks

The actual dataset is collected from several wind turbine gearboxes in a wind farm. The gearbox structure is shown in Fig. 5, consisting of a two-stage planetary transmission and a one-stage parallel-axis transmission. Two gear faults are present on both the second gear ring (SGR) and the high-speed shaft gear (HSG). The SGR is collected from two wind turbines, F34 and F47, while the HSG is collected from two other wind turbines, F14 and F38. The monitoring vibration signals derived from two acceleration sensors placed on the shells of the second gear ring and the high-speed shaft are used to carry out the transferability discrimination. The sampling frequencies of the two sensors are 12.8 kHz and 25.6 kHz, respectively. Since the minimum data points required to cover a fault period on the low-speed end are higher than those on the high-speed end, the minimum data points of the second gear ring are selected as the sample dimension for all fault types. This helps to reduce the calculation time significantly, with a set value of 20000. The sample number for each fault type is 200.

Table 6 displays the construction of eight domains using the actual wind turbine dataset. In this subsection, "A" is selected as the target domain, while the other seven domains are considered source domains. Table 7 provides a list of seven transfer tasks that can be derived from this setup. The distribution discrepancy of transfer tasks in the group (A, B, C, and D) is due to variations in speed ranges, while the remaining transfer tasks are influenced by a combination of different speed ranges and wind turbines. The accuracy benchmark setting is similar to Section 4.1. The comprehensive experimental result is presented in Table 7. It is evident that our proposed MSTDA outperforms other methods in terms of transferability discrimination performance in an actual wind turbine scenario. Specifically, the MSTDA has a unique advantage in terms of correlation and consistency. In contrast to the testbed dataset, the working condition of the actual wind turbine dataset varies non-linearly in real-time. This makes the actual wind turbine dataset more challenging compared to the testbed dataset.

From Table 7 it is evident that the performance of nearly all methods in the three evaluation metrics is inferior to that of Table 5. And the A-distance derived from the classifier tends to have a higher standard deviation compared to other similarity measure methods in most transfer tasks. Considering that MMD has become the most mainstream distribution discrepancy metric in the field of fault transfer diagnosis, therefore, three MMD-based similarity indexes (AHMM (Feng et al., 2023), DDM (Qian et al., 2023), and KMMD (Lu et al., 2024)) are supplemented as the comparative methods to further show the advantage of the proposed MSTDA. The experimental results are illustrated in Fig. 6. It can be obviously seen that the proposed MSTDA still possesses the best transferability discrimination capability. Finally, the comprehensive experimental results on the actual wind turbine dataset further demonstrate the superiority and universality of MSTDA.

To visually evaluate the effectiveness of MSTDA in extracting the separable fault features, we employ t-distributed stochastic neighbor (t-SNE) to project the original samples and the features obtained by MSTDA into a three-dimensional space. The resulting t-SNE mappings are shown in Fig. 7. It is evident that there are more overlapping data points between SGR and HSG in Fig. 7(a) than there are in Fig. 7(b). This indicates that MSTDA is effective in extracting separable features with zero-label prior knowledge. The extracted separable intrinsic fault features will contribute to dig the cross-domain intrinsic discrepancy in the phase of constructing the similarity measurement criterion, thus enhancing the transferability discrimination performance of MSTDA.

In addition, to demonstrate the feature scale robustness of different transferability discriminative methods, we take the fault features extracted from the cross-domain task "A↔C1" and the maximum subspace representation as the research objects. By varying the multiples of the feature values in the source domain "C1", we simulated different feature scales. The relative transferability values of all transferability discrimination methods are shown in Fig. 8. From the figure, it is clearly observed that: except for the MSTDA and Cosine, the results of other four methods fluctuate with the change of the preset multiples, and their fluctuation degrees are different, among which the fluctuation of A-distance is particularly severe. This indicates that MMD, CORAL, A-distance, and L2-distance are sensitive to the changes of feature scales and cannot accurately reflect the similarities between different cross-domain tasks. Although Cosine distance exhibits a strong feature scale robustness, its poor performance on transferability discrimination experiments still limits its practical application value. In contrast, the MSTDA ensures the excellent transferability discrimination performance while maintaining the robustness to feature scales.

Table 6
Details of actual wind-turbine dataset.

Information	Name							
	A	B	C	D	A1	B1	C1	D1
No. of wind turbine	F34 F14	F34 F14	F34 F14	F34 F14	F47 F38	F47 F38	F47 F38	F47 F38
Fault type	SGR HSG	SGR HSG	SGR HSG	SGR HSG	SGR HSG	SGR HSG	SGR HSG	SGR HSG
Speed range (r/min)	700–750	850–900	1000–1050	1200–1250	700–750	850–900	1000–1050	1200–1250

Table 7
Transferability results of actual wind-turbine transfer tasks.

Transfer tasks	Accuracy benchmark	Methods					
		MMD	CORAL	A-distance	L2-distance	Cosine	MSTDA
A↔A1	0.96	0.22 ± 0.13	1.00 ± 0.00	0.91 ± 0.12	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
A↔B	0.92	0.14 ± 0.05	0.27 ± 0.02	0.83 ± 0.12	0.35 ± 0.09	0.13 ± 0.12	0.73 ± 0.02
A↔B1	0.89	0.55 ± 0.55	0.58 ± 0.12	0.41 ± 0.30	0.92 ± 0.06	0.54 ± 0.02	0.48 ± 0.01
A↔C	0.85	0.17 ± 0.06	0.00 ± 0.00	0.60 ± 0.20	0.00 ± 0.00	0.00 ± 0.00	0.68 ± 0.03
A↔C1	0.81	1.00 ± 0.00	0.72 ± 0.04	0.58 ± 0.20	0.79 ± 0.04	0.55 ± 0.03	0.34 ± 0.04
A↔D	0.62	0.00 ± 0.00	0.42 ± 0.16	0.49 ± 0.21	0.94 ± 0.04	0.96 ± 0.04	0.30 ± 0.05
A↔D1	0.47	0.92 ± 0.01	0.49 ± 0.08	0.14 ± 0.11	0.77 ± 0.04	0.59 ± 0.03	0.00 ± 0.00
CORR		0.35	0.16	0.83	0.19	0.24	**0.89**
CONS		−0.82	0.56	1.49	−0.76	−0.76	**1.68**
MONO		0.00	0.00	**0.67**	0.00	0.00	**0.67**

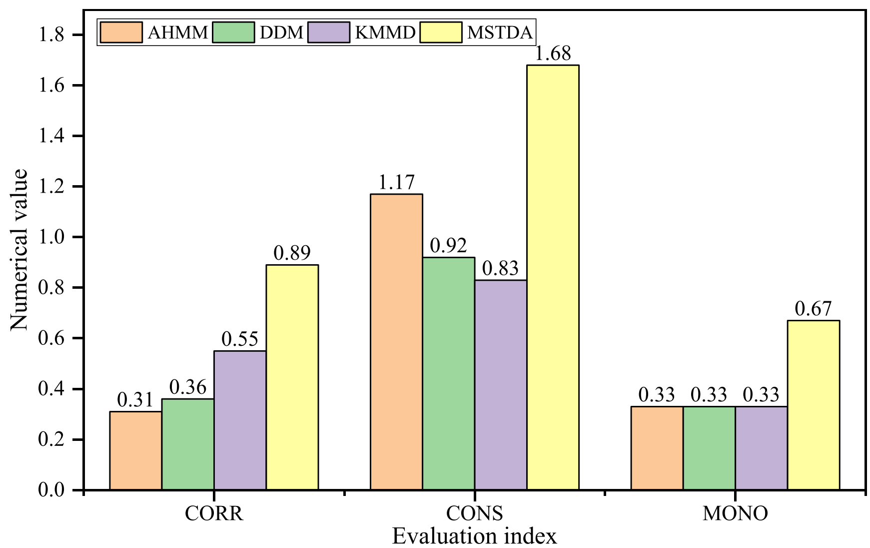

Fig. 6. The comparative experimental results between MSTDA and three MMD-based indexes.

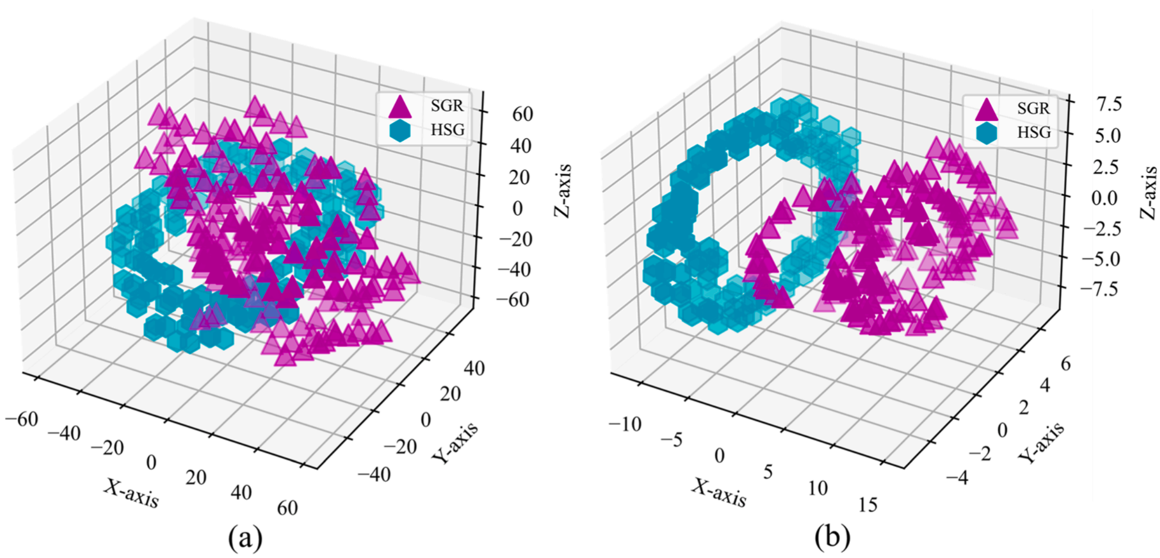

Fig. 7. Three-dimensional t-SNE mapping: (a) original samples; (b) extracted fault features.

4.4. Limitations and future works

It should be noted that the proposed MSTDA can only provide the relative transferability of different transfer tasks and still cannot answer the question of whether cross-domain diagnosis is possible for the corresponding transfer tasks. Secondly, the success of MSTDA also heavily relies on the existence of clear decision boundaries in different fault monitoring data. However, in practical engineering, the inherent separability may be weakened due to the complexity of working environment, thereby reducing the discriminant performance of MSTDA. In

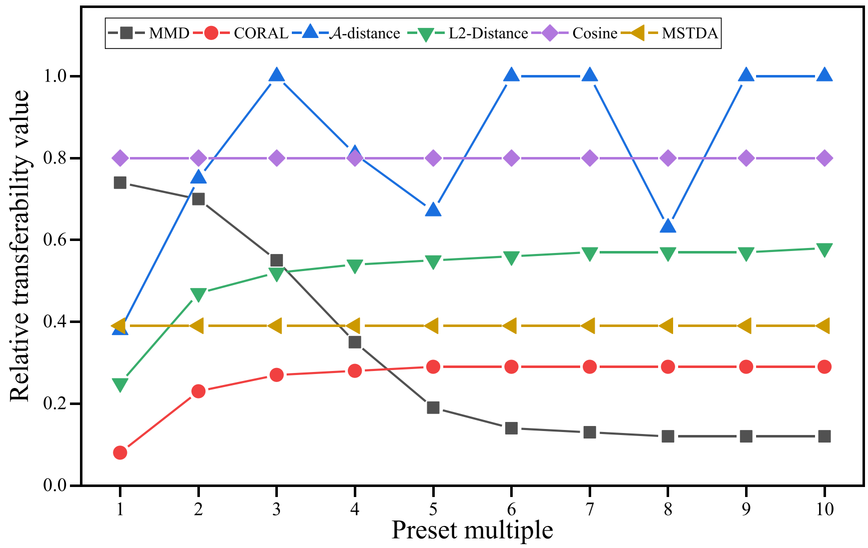

Fig. 8. Sensitivity of different transferability discrimination methods to feature scale changes.

future works, we will study the adaptive threshold decision rules that can discern whether specific cross-domain diagnostic tasks are transferable, and effective denoising methods.

5. Conclusions

This study proposes a novel cross-domain similarity measure called MSTDA, which aims to address transferability discrimination under zero-label prior knowledge. The approach incorporates the maximum subspace representation and similarity measurement criteria. In order to improve the separability of different fault signals, a Hilbert space is first constructed during the maximum subspace representation phase. This is achieved by using a newly designed kernel derived from the mean square statistic. The kernel maps the low-dimensional sample into a high-dimensional space. Then, the separable intrinsic fault feature is extracted to represent the distribution discrepancy between the source and target domains. Furthermore, a novel similarity measurement criterion is developed based on the orthogonal bases of the intrinsic feature subspaces from both the source and target domains, which are obtained through the SVD. The mathematical property that is robust to the feature scale is strictly proven. Ultimately, the comprehensive experimental results and discussions confirm that MSTDA has superior transferability discrimination performance compared to other well-known similarity measure methods.

CRediT authorship contribution statement

Yi Qin: Writing – review & editing, Formal analysis, Conceptualization. **Fei Wu:** Validation, Investigation. **Yi Wang:** Validation, Data curation. **Quan Qian:** Writing – original draft, Software, Methodology.

Declaration of Competing Interest

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.

Data Availability

The data that has been used is confidential.

Acknowledgment

This paper was supported by National Natural Science Foundation of China (no. 52175075).

References

Anstreicher, K., Wolkowicz, H., 2000. On Lagrangian relaxation of quadratic matrix constraints. SIAM J. Matrix Anal. Appl. 22, 41–55.
Anvar, A.A.T., Mohammadi, H., 2023. A novel application of deep transfer learning with audio pre-trained models in pump audio fault detection. Comput. Ind. 147, 103872.
Belghazi, M.I., Baratin, A., Rajeshwar, S., Ozair, S., Bengio, Y., Courville, A., Hjelm, D., 2018. Mutual information neural estimation. International Conference on Machine Learning. PMLR, pp. 531–540.
Ben-David, S., Blitzer, J., Crammer, K., Pereira, F., 2006. Analysis of representations for domain adaptation. Adv. Neural Inf. Process. Syst. 19, 1–8.
Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W., 2010. A theory of learning from different domains. Mach. Learn. 79, 151–175.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.
Du, N.H., Long, N.H., Ha, K.N., Hoang, N.V., Huong, T.T., Tran, K.P., 2023. Trans-Lighter: a light-weight federated learning-based architecture for Remaining Useful Lifetime prediction. Comput. Ind. 148, 103888.
Feng, Y., Chen, J., He, S., Pan, T., Zhou, Z., 2023. Globally localized multisource domain adaptation for cross-domain fault diagnosis with category shift. IEEE Trans. Neural Netw. Learn. Syst. 34, 3082–3096.
Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2017. Domain-adversarial training of neural networks. J. Mach. Learn. Res. 17, 2096-2030.
Gong, B., Shi, Y., Sha, F., Grauman, K., 2012. Geodesic flow kernel for unsupervised domain adaptation. IEEE Conf. Comput. Vis. Pattern Recognit., IEEE 2012, 2066–2073.
Gopalan, R., Li, R., Chellappa, R., 2011. Domain adaptation for object recognition: an unsupervised approach. 2011 International Conference on Computer Vision. IEEE, pp. 999–1006.
Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A., 2012. A kernel two-sample test. J. Mach. Learn. Res. 13, 723–773.
He, C., Shi, H., Li, J., 2023. IDSN: a one-stage interpretable and differentiable STFT domain adaptation network for traction motor of high-speed trains cross-machine diagnosis. Mech. Syst. Signal Process. 205, 110846.
J. Jiao, H. Li, J. Lin, H. Zhang, Entropy-Oriented Domain Adaptation for Intelligent Diagnosis of Rotating Machinery, IEEE Transactions on Systems, Man, and Cybernetics: Systems, in press (2023).
Kolouri, S., Nadjahi, K., Simsekli, U., Badeau, R., Rohde, G., 2019. Generalized sliced wasserstein distances. Adv. Neural Inf. Process. Syst. 32.
Lei, Z., Zhang, P., Chen, Y., Feng, K., Wen, G., Liu, Z., Yan, R., Chen, X., Yang, C., 2023. Prior knowledge-embedded meta-transfer learning for few-shot fault diagnosis under variable operating conditions. Mech. Syst. Signal Process. 200, 110491.
Lian, C., Zhao, Y., Shao, J., Sun, T., Dong, F., Ju, Z., Zhan, Z., Shan, P., 2024. CFI-LFENet: infusing cross-domain fusion image and lightweight feature enhanced network for fault diagnosis. Inf. Fusion 104, 102162.
Lu, F., Tong, Q., Jiang, X., Feng, Z., Xu, J., Huo, J., 2024. Deep multilayer sparse regularization time-varying transfer learning networks with dynamic kullback–leibler divergence weights for mechanical fault diagnosis. IEEE Trans. Ind. Inform.

Mercer, J., 1909. functions of positive and negative type, and their connection the theory of integral equations. Philos. Trans. R. Soc. Lond. Ser. A Contain. Pap. A Math. Or. Phys. Character 209, 415–446.
Piccoli, B., Rossi, F., 2014. Generalized Wasserstein distance and its application to transport equations with source. Arch. Ration. Mech. Anal. 211, 335–358.
Qian, Q., Luo, J., Qin, Y., 2024. Adaptive intermediate class-wise distribution alignment: a universal domain adaptation and generalization method for machine fault diagnosis. IEEE Trans. Neural Netw. Learn. Syst.
Qian, Q., Qin, Y., Luo, J., Wang, Y., Wu, F., 2023. Deep discriminative transfer learning network for cross-machine fault diagnosis. Mech. Syst. Signal Process. 186, 109884.
Qian, Q., Qin, Y., Luo, J., Xiao, D., 2023. Cross-machine transfer fault diagnosis by ensemble weighting subdomain adaptation network. IEEE Trans. Ind. Electron. 70, 12773–12783.
Qian, Q., Zhou, J., Qin, Y., 2023. Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions. IEEE Trans. Ind. Inform. 19, 9898–9908.
Qin, Y., Li, Q., Wang, S., Cao, P., 2024. Dynamics modeling of faulty planetary gearboxes by time-varying mesh stiffness excitation of spherical overlapping pittings. Mech. Syst. Signal Process. 210, 111162.
Reed, M., 2012. Methods of modern mathematical physics: Functional analysis. Elsevier.
Schölkopf, B., Smola, A., Müller, K.-R., 1997. Kernel principal component analysis. International Conference on Artificial Neural Networks. Springer, pp. 583–588.
Sun, B., Feng, J., Saenko, K., 2016. Return of Frustratingly Easy Domain Adaptation. Proc. AAAI Conf. Artif. Intell. 2058–2065.
Yang, B., Lei, Y., Li, X., Li, N., 2023. Targeted transfer learning through distribution barycenter medium for intelligent fault diagnosis of machines with data decentralization. Expert Syst. Appl., 122997
Yang, B., Lei, Y., Xu, S., Lee, C.-G., 2021. An optimal transport-embedded similarity measure for diagnostic knowledge transferability analytics across machines. IEEE Trans. Ind. Electron. 69, 7372–7382.
Yu, X., Zhao, Z., Zhang, X., Chen, X., Cai, J., 2023. Statistical identification guided open-set domain adaptation in fault diagnosis. Reliab. Eng. Syst. Saf. 232, 109047.
Yu, X., Zhao, Z., Zhang, X., Tian, S., Kwoh, C.-K., Li, X., Chen, X., 2023. A universal transfer network for machinery fault diagnosis. Comput. Ind. 151, 103976.
Zhang, Y., Ji, J., Ren, Z., Ni, Q., Gu, F., Feng, K., Yu, K., Ge, J., Lei, Z., Liu, Z., 2023. Digital twin-driven partial domain adaptation network for intelligent fault diagnosis of rolling bearing. Reliab. Eng. Syst. Saf. 234, 109186.
Zhao, K., Hu, J., Shao, H., Hu, J., 2023. Federated multi-source domain adversarial adaptation framework for machinery fault diagnosis with data privacy. Reliab. Eng. Syst. Saf. 236, 109246.