

DOI: 10.19650/j.cnki.cjsi.J2312241

# 一种基于条件度量迁移学习的机械故障 诊断可解释方法\*

路飞宇<sup>1</sup>, 佟庆彬<sup>1,2</sup>, 姜学东<sup>1</sup>, 徐建军<sup>1</sup>, 霍静怡<sup>1</sup>

(1. 北京交通大学电气工程学院 北京 100044; 2. 载运装备多源动力系统教育部重点实验室  
(北京交通大学) 北京 100044)

**摘要:** 迁移学习技术可以减小源域和目标域特征之间的分布差异。然而,在跨设备场景下,现有研究往往难以衡量并缩小不同设备间数据的条件分布差异,导致模型在源域获得的知识不能很好地迁移到目标域。此外,在实际的故障诊断场景中,决策者通常需要理解模型为何将某些数据归类为特定故障类型。由于深度学习模型的复杂性,其往往被看作是“黑匣子”,难以解释其内部工作机制。为了克服上述缺点,提出一种基于条件度量迁移学习的可解释故障诊断方法。首先利用 Hilbert 包络谱分析将时域信号转为频域信号;其次搭建深度孪生卷积神经网络和分类器,从频域中提取源域和目标域数据中的高维特征并做分类训练;然后将可解释的条件核 Bures 度量嵌入到无监督学习的损失函数中,提高条件分布下的特征适配能力和模型可解释性;最后利用博弈论中的 SHAP 方法对模型诊断结果做基于包络谱的事后可解释分析。在 3 种设备的 6 种跨设备轴承故障诊断任务中开展试验,对所提方法和其他相关对比方法进行评估,结果表明提出的方法可以有效地提高跨设备机械故障诊断精度,达到了平均 99.47% 的诊断精度。并解释了哪些频率点对模型的决策起到关键作用。

**关键词:** 条件度量;机械故障诊断;迁移学习;SHAP

中图分类号: TH17 文献标识码: A 国家标准学科分类代码: 510.40

## Interpretable method for mechanical fault diagnosis based on condition metric transfer learning

Lu Feiyu<sup>1</sup>, Tong Qingbin<sup>1,2</sup>, Jiang Xuedong<sup>1</sup>, Xu Jianjun<sup>1</sup>, Huo Jingyi<sup>1</sup>

(1. School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China; 2. Key Laboratory of Vehicular Multi-Energy Drive Systems (Beijing Jiaotong University), Ministry of Education, Beijing 100044, China)

**Abstract:** Transfer learning techniques can reduce the distribution difference between source and target domain features. However, in cross-device scenarios, existing research is often difficult to measure and reduce the differences in the conditions of data between different devices, resulting in the knowledge obtained by the model in the source domain cannot be migrated to the target domain. Additionally, in real-world failure diagnostic scenarios, decision-makers usually need to understand why the model classifies a specific type of fault. Due to the complexity of deep learning models, they are often seen as "black boxes," making it difficult to explain their internal workings. To address these issues, an interpretable fault diagnosis method based on conditional metric transfer learning is proposed. Firstly, Hilbert envelope spectrum analysis is used to convert time-domain signals into frequency-domain signals. Secondly, a deep twin convolutional neural network and classifier are built to extract high-dimensional features from both source and target domain data in the frequency domain and perform classification training. Then, the interpretable Conditional Kernel Bures is embedded into the loss function of unsupervised learning to enhance feature adaptation and model interpretability under conditional distribution. Finally, the SHAP method from game theory is used to conduct post-event interpretable analysis of the model diagnosis results based on the envelope spectrum. Tests were conducted on 12 cross-equipment bearing fault diagnosis tasks across three types of mechanical equipment, evaluating the proposed method against other related methods. The results show that the proposed method could effectively improve the

收稿日期: 2023-12-05 Received Date: 2023-12-05

\* 基金项目: 中央高校基本科研业务费专项资金(2023JBZY039)、北京市自然科学基金(L211010, 3212032)项目资助

accuracy of cross-equipment mechanical fault diagnosis, achieving an average diagnostic accuracy of 99.47%. It also identifies which frequency points played a crucial role in the model's decision-making process.

**Keywords:** condition metric; mechanical fault diagnosis; transfer learning; SHAP

## 0 引言

智能运维和健康管理对旋转机械设备的安全可靠运行起到了举足轻重的作用<sup>[1]</sup>。其中,轴承的故障诊断问题研究<sup>[2]</sup>为机械设备的智能维护提供重要的理论依据。

近年来,深度学习技术得到快速发展,基于深度学习的机械故障诊断模型具有智能化、简单化的优势。然而深度学习往往需要大量的有标签数据,同时训练集和测试集数据服从独立同分布,这限制了数据无标签下机械故障诊断的发展和应用<sup>[3]</sup>。为此,迁移学习被提出用来解决上述问题。迁移学习放宽了源域和目标域数据分布不一致的要求,并且可以在待测设备的目标域数据上开展无监督训练。Lei等<sup>[4]</sup>根据被研究的问题,将基于迁移学习的故障诊断任务划分为两大类:同一设备的迁移任务和不同设备间的迁移任务。

同设备迁移任务指的是有标签的源域和少量标签甚至是无标签的目标域数据来自同一个设备,但数据是在不同的运行工况下采集到的,如转速、负载、运行环境等。众多学者针对同设备迁移任务做了大量的探索和研究,并取得了不错的效果。陈祝云等<sup>[5]</sup>提出了一种增强迁移卷积神经网络,该模型将对抗思想嵌入到迁移模型中,在六种跨转速任务中得到了验证。Jiao等<sup>[6]</sup>面向预测性约束理论提出了一种新型的跨转速域自适应模型,与类似结构的对比方法相比,所提模型提高了行星齿轮和风机轴承中的故障诊断精度。Deng等<sup>[7]</sup>考虑到域内特征分布的差异性,提出了目标域内子标签学习机制,并结合联合分布损失函数提高了模型在跨转速故障诊断任务中的迁移性能。

不同设备迁移任务和同设备迁移相比,具有更为严苛的要求,源域和目标域数据来自于不同设备中被测元件,如不同的设备型号、不同的安装环境等,这无疑增大两种数据之间的分布差异。为此,许多跨设备故障诊断方法被提出。Guo等<sup>[8]</sup>在深度迁移卷积神经网络中引入对抗学习,实现了跨设备的轴承故障诊断。Liu等<sup>[9]</sup>为了解决跨转速和跨设备故障诊断问题,提出一种深度对抗子域自适应网络,并在两种不同的轴承数据集中得到了验证。Yang等<sup>[10]</sup>利用目标域中少量的标签信息,提出一种最优传输嵌入联合分布相似性度量,提高了诊断知识的可迁移性,实现了跨设备的机械故障诊断。

最近,基于可解释人工智能的深度学习模型被广泛

关注,机械故障诊断模型的可解释性有助于提高用户对模型决策的可信赖性,从而推广模型的使用范围。Pandey等<sup>[11]</sup>利用与模型无关的局部可解释(local interpretable model-agnostic explanations, LIME)实现了Lamb波下的轴承故障诊断。同样,Lu等<sup>[12]</sup>基于LIME提出了一种可解释的一维卷积神经网络,并开展了时变转速工况下的轴承事后可解释问题研究。Kim等<sup>[13]</sup>利用三轴传感器数据和基于频域类激活图可视化方法实现了对轴承和齿轮故障诊断的事后可解释分析。

上述研究结果表明,目前跨设备故障诊断问题研究往往忽略了源域和目标域数据之间的条件分布差异性导致迁移精度不高。可解释人工智能在故障诊断领域仍以同设备故障诊断问题为主,而跨设备故障诊断下的模型可解释研究相对较少,具有模型迁移机制可解释性弱的缺点。为此,受到条件度量和基于博弈论的事后可解释方法的启发,在跨设备场景下,提出一种基于条件度量迁移学习的机械故障诊断可解释方法。该方法首先利用包络谱分析从时域信号中提取信号的故障特征;然后在迁移学习架构的基础上,引入条件核Bures度量(conditional kernel bures, CKB),缩小源域和目标域数据之间的分布差异的同时增强迁移机制的可解释性;最后利用沙普利加法解释(shapley additive explanations, SHAP)值的大小来衡量模型预测结果和特征点之间的关联关系。

为了解决跨设备故障诊断中源域和目标域数据的条件分布差异和域偏移的问题,在跨设备诊断模型中创新地引入了包络频谱分析。该方法旨在减少频域级别不同设备生成的数据的差异,从而优化数据一致性并增强故障特性的表达。

采用了基于CKB的功能传输策略,该策略进一步降低了来自不同领域的数据的偏差,为高效且精准地训练故障诊断模型提供了坚实的基础。

为了提高模型的透明度和可解释性,实施SHAP技术,该技术可以和包络谱的特征值结合,实现基于包络谱特征值的模型可解释功能,保证了故障诊断过程的可信度。

为了验证本文所提方法的有效性,对两个公共轴承故障数据集和一个高速动车组列车轴承故障数据集进行了跨设备故障诊断研究。精心地设计了6种不同的诊断任务,并使用3个定量指标比较了5种先进的故障诊断方法,从消融实验、不同训练样本数量、抗噪性等角度开展分析。通过严格的实验设计和评估,所提方法证明了其在各种任务中的有效性和优越性。

## 1 跨设备机械故障诊断问题描述

定义源域数据  $D_s^{h_s} = \{(\mathbf{x}_s^{(i)}, \mathbf{y}_s^{(i)})\}_{i=1}^N$  来自于设备 A, 目标域数据  $D_T^{h_t} = \{(\mathbf{x}_t^{(i)})\}_{i=1}^M$  来自于设备 B,  $N, M$  分别为源域和目标域中的样本数量。  $h_s, h_t$  分别为相应的健康状态类别数。另外,  $D_s^{h_s} = \{(\mathbf{x}_s^{(i)})\}_{i=1}^N$  的标签  $\{\mathbf{y}_s^{(i)}\}_{i=1}^N$  已知, 而目标域数据  $D_T^{h_t} = \{(\mathbf{x}_t^{(i)})\}_{i=1}^M$  的标签未知。假设  $P, Q$  分别为源域和目标域的数据分布, 由于不同设备之间的数据分布具有较大的差异性, 导致域偏移, 即  $P \neq Q$ 。为此, 如何构建无监督的故障诊断模型, 实现设备 A 到设备 B (目标域) 跨设备故障诊断任务, 同时增强模型在诊断过程中的可解释性是本文要解决的问题。

条件分布差异和域偏移是指在迁移学习中, 源域和目标域数据在特征空间和标签空间上的统计分布不同, 导致从源域学习到的模型在目标域上性能下降。在机械故障诊断领域, 特别是跨设备场景下, 条件分布差异和域偏移主要体现在如下 4 个方面。

### 1.1 设备工作条件的差异

不同设备可能不同的工作环境中运行, 如温度、湿度、粉尘等因素都会影响振动信号的特征。负载条件的变化会直接影响机械部件的受力状态, 进而改变故障特征的表现形式。例如, 同样的轴承裂纹故障在高负载和低负载条件下, 其在振动信号中的表现可能有显著差异。运行速度的不同会导致故障特征频率的偏移。

### 1.2 设备结构和材料的差异

即使是相同型号的设备, 由于制造误差和材料批次的不同, 其固有频率和模态特性可能存在细微差异。轴承的安装位置和方式的差异会影响故障信号的传播路径, 从而改变信号在传感器处的表现。不同设备的减震装置和外壳结构可能不同, 这会导致相同故障在不同设备上产生的振动信号有所不同。

### 1.3 传感器系统的差异

传感器的型号、灵敏度和频率响应特性可能不同, 导致采集到的信号存在系统性偏差。传感器的安装位置和安装方式的差异会直接影响信号的采集质量。例如, 一个轻微的安装角度偏差可能导致某些频率成分被放大或衰减。信号采集系统的采样频率、滤波设置等参数的不同也会影响最终获得的频谱特征。

### 1.4 环境噪声的影响

不同的工业环境中存在各种不同的背景噪声, 这些噪声会叠加在故障信号上, 增加了跨设备诊断的难度。某些环境噪声可能与故障特征频率相近, 导致特征提取和故障识别的困难。

## 2 提出的机械故障诊断可解释方法

### 2.1 包络谱分析

跨设备故障诊断任务中, 由于源域和目标域数据从不同的运行工况中搜集而来, 如转速、负载、运行环境等, 导致数据之间的条件分布差异较大, 现有的无监督学习模型无法直接从原始振动信号中学习数据特征的相同特征分布。考虑到包络谱分析可以突出机械信号的故障特征<sup>[14]</sup>, 将包络谱分析作为信号的预处理, 期望提高后续迁移学习模型的特征提取能力。包络谱的原理非常简单, 在快速傅里叶变换 (FFT) 前执行 Hilbert 包络变换即可, 其计算公式如下:

$$E(t) = \text{FFT}(\sqrt{x^2(t) + \tilde{x}^2(t)}) \quad (1)$$

式中:  $x(t)$  为原始振动信号;  $\tilde{x}(t) = \text{Hilbert}(x(t))$  为  $x(t)$  的 Hilbert 解析信号;  $E(t)$  为包络谱结果。

### 2.2 CKB

CKB 是一种新型条件分布差异度量<sup>[15-16]</sup>, CKB 在最优传输 (optimal transport, OT) 框架下提供了一种统计和可解释的方法以探究知识在迁移学习模型中的迁移机制。CKB 的有效性和可解释性已经在计算机视觉和模式识别领域得到了验证。鉴于 CKB 优异的可解释性和域适配能力, 将其引入到跨设备机械故障诊断中, 提高迁移模型的特征可解释性并缩小不同设备之间的数据差异性。

假设源域的条件分布为  $P_s(X|Y)$ , 目标域数据的条件分布为  $P_T(X|Y)$ , 则 CKB 定义如下:

$$d_{\text{CKB}}^2(\mathbf{R}_{XX|Y}^s, \mathbf{R}_{XX|Y}^t) = \text{tr}(\mathbf{R}_{XX|Y}^s + \mathbf{R}_{XX|Y}^t - 2\mathbf{R}_{XX|Y}^s) \quad (2)$$

式中:

$$\mathbf{R}_{XX|Y}^s = \sqrt{\sqrt{\mathbf{R}_{XX|Y}^s} \mathbf{R}_{XX|Y}^t \sqrt{\mathbf{R}_{XX|Y}^s}} \quad (3)$$

$\mathbf{R}_{XX|Y}$  为条件协方差算子, 公式为:

$$\mathbf{R}_{XX|Y} = \mathbf{R}_{XX} - \mathbf{R}_{XY} \mathbf{R}_{YY}^{-1} \mathbf{R}_{YX} \quad (4)$$

$\mathbf{R}_{XY}$  为交叉协方差算子<sup>[17]</sup>, 定义如下:

$$\mathbf{R}_{XY} = \mathbb{E}_{XY}[(\phi(X) - \mu_X) \otimes (\psi(Y) - \mu_Y)] \quad (5)$$

式中:  $\mu_X$  和  $\mu_Y$  分别为源域和目标域数据的均值;  $\phi(X)$  和  $\psi(Y)$  分别为  $X$  和  $Y$  在可再生核希尔伯特空间 (reproducing kernel hilbert space, RKHS) 的非线性特征映射;  $\otimes$  表示张量积符号。当  $X = Y$  时,  $\mathbf{R}_{XX} = \mathbf{R}_{YY} = \mathbf{R}_{XY}$ 。

然而, 式(2)中  $\mathbf{R}_{XX|Y}$  是隐式公式, 无法直接求解, 为了等效为显示公式, 这里直接引用文献[16]中 CKB 经验估计公式:

$$\hat{d}_{\text{CKB}}^2(\hat{\mathbf{R}}_{XX|Y}^s, \hat{\mathbf{R}}_{XX|Y}^t) = \text{tr}[\mathbf{G}_X^s(\varepsilon n \mathbf{I}_n + \mathbf{G}_Y^s)^{-1}] + \varepsilon \text{tr}[\mathbf{G}_X^t(\varepsilon m \mathbf{I}_m + \mathbf{G}_Y^t)^{-1}] - \frac{2}{\sqrt{nm}} \|(\mathbf{H}_m \mathbf{C}_t)^T \mathbf{K}_{XX}^s (\mathbf{H}_n \mathbf{C}_s)\| \quad (6)$$

式中:  $n$ 、 $m$  分别为源域和目标域中的样本数量;  $\varepsilon > 0$  为正则化系数;  $\mathbf{H}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{I}_n \mathbf{I}_n^T$ ,  $\mathbf{I}_n$  为  $n$  阶对角矩阵,  $\mathbf{I}_n$  为取值全为 1 的  $n$  维向量,  $\mathbf{H}_m$  同理。  $(\mathbf{K}_{XX}^s)_{ij} = k_{\mathcal{X}}(\mathbf{x}_i^s, \mathbf{x}_j^s)$  为显式核矩阵;  $\|\cdot\|_*$  表示核范数;  $\mathbf{C}_s$  和  $\mathbf{G}_Y^s$  满足式(7)。

$$\varepsilon n (\mathbf{G}_Y^s + \varepsilon n \mathbf{I}_n)^{-1} = \mathbf{U}_s \mathbf{D}_s \mathbf{U}_s^T =$$

$$\mathbf{U}_s \sqrt{\mathbf{D}_s} (\mathbf{U}_s \sqrt{\mathbf{D}_s})^T = \mathbf{C}_s \mathbf{C}_s^T \quad (7)$$

式中:  $\mathbf{U}_s$  和  $\mathbf{D}_s$  分别为特征向量和特征值矩阵。

CKB 度量具有 OT 理论的可解释性<sup>[15]</sup>, 通过减少模型的泛化误差并提高可迁移性。文献[18]指出泛化误差同时受到特征上的边际差异和标签变量上的条件差异的影响。对于任何分类器  $\hat{Y} = h(Z)$ , 有如下不等式成立:

$$\varepsilon_t(h) \leq \varepsilon_s(h) + 2(K-1)\Delta_{\text{CE}}(\hat{Y}) + \|p_Y^s - p_Y^t\|_1 \cdot \text{BER}_s(\hat{Y} \| Y) \quad (8)$$

式中:  $\|\cdot\|_1$  为  $\ell_1$  范数。

$$\text{BER}_s(\hat{Y} \| Y) \triangleq \max_{j \in [K]} \mathbb{E}_{p_{\hat{Y}}^s | Y=j} [\mathbb{I}_{[\hat{Y} \neq Y]} | Y=j] \quad (9)$$

$\Delta_{\text{CE}}(\hat{Y})$  为条件误差间隙, 定义如下:

$$\Delta_{\text{CE}}(\hat{Y}) \triangleq \max_{i \neq j \in \mathcal{Y}^2} |p_{\hat{Y}=i | Y=j}^s - q_{\hat{Y}=i | Y=j}^s| \quad (10)$$

其中,  $\mathbb{I}$  为指示函数。

式(8)表明, 条件分布对齐的模型可以最小化上确界的  $\varepsilon_s(h) + 2(K-1)\Delta_{\text{CE}}(\hat{Y})$ , 而  $\|p_Y^s - p_Y^t\|_1 \cdot \text{BER}_s(\hat{Y} \| Y)$  表示目标域的偏移和源域上的分类性能。在所提方法中, 条件分布误差  $\Delta_{\text{CE}}(\hat{Y})$  和源域误差  $\text{BER}_s(\hat{Y} \| Y)$  足够小, 表明了条件分布对齐对知识迁移来说是足够的。

通过计算经验估计式(6)得到源域和目标域数据特征之间的条件分布距离。CKB 效果如图 1 所示。从图 1 可以看出, 源域和目标域数据首先通过特征提取器输出

高维特征, 并将高维特征映射到 RKHS 中, 然后, 计算源域和目标域的条件分布  $\mathbf{R}_{XX|Y}^s$  和  $\mathbf{R}_{XX|Y}^t$ , 并利用 CKB 评估二者的分布差异。其中, 分布曲线代表标签数据。经过 CKB 估计分布差异以及分布差异反馈过程, 模型自适应匹配相同标签下的数据, 从而实现特征匹配迁移的功能。

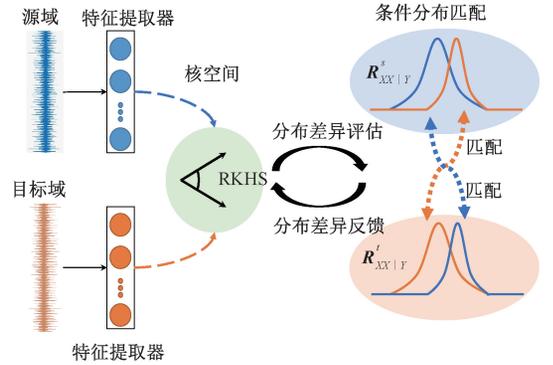


图 1 条件度量 CKB 效果

Fig. 1 The effect of conditional metric CKB

### 2.3 深度孪生迁移卷积神经网络模型

#### 1) 模型结构

狭义的孪生神经网络由 2 个参数共享结构相同的网络组成, 广义的孪生神经网络由两个任意的参数不共享的神经网络结合而成, 本文构建的是狭义的参数共享的孪生神经网络。

搭建的迁移学习模型基本结构如表 1 所示。该模型由特征提取器  $G_f$  和分类器  $G_c$  组成, 具体来说, 6 个卷积池化层构成特征提取器, 每层结构有都由不同的卷积核和最大池化组成, BN 为批归一化操作, 激活函数为 ReLU。分类器由 2 个全连接层构成, 输出节点数量分别为 256 和 4。

表 1 深度孪生迁移卷积神经网络模型参数

Table 1 Deep twin migration convolutional neural network model parameters

网络结构	层	模型参数
特征提取器	卷积池化 1	核大小 16-64×1, 步长 8, 填充 1; BN; ReLU; 最大池化 2×1, 步长 2
	卷积池化 2	核大小 32-3×1, 步长 1, 填充 1; BN; ReLU; 最大池化 2×1, 步长 2
	卷积池化 3	核大小 64-3×1, 步长 1, 填充 1; BN; ReLU; 最大池化 2×1, 步长 2
	卷积池化 4	核大小 64-3×1, 步长 1, 填充 1; BN; ReLU; 最大池化 2×1, 步长 2
	卷积池化 5	核大小 64-3×1, 步长 1, 填充 1; BN; ReLU; 最大池化 2×1, 步长 2
	卷积池化 6	核大小 1 024-3×1, 步长 1, ; BN; ReLU; 最大池化 2×1, 步长 2
分类器	全连接层 1	输出节点: 256
	全连接层 2	输出节点: 4; Softmax

注: 卷积核大小  $L-C \times D$  分别为卷积核的数量、大小和通道数

## 2) 模型训练

模型整体训练过程由3部分组成,源域数据的有监督训练,目标域数据的无监督训练以及源域和目标域数据的域适配训练。

在搭建的迁移孪生卷积神经网络基础上,利用源域中的标签数据做有监督的训练,该过程由交叉熵损失函数实现,具体公式如下:

$$L_C = \mathbb{E}_{(x_s^{(i)}, y_s^{(i)}) \in D_S^{k,s}} [-\log(\hat{y}_c^{(n)})] = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C y_c^{(n)} \log \frac{\exp(\hat{y}_c^{(n)})}{\sum_{c=1}^C \exp(\hat{y}_c^{(n)})} \quad (11)$$

式中: $N$ 为样本数量; $C$ 为类别数量; $y_c^{(n)}$ 为取值为0或1的符号函数; $\hat{y}_c^{(n)}$ 表示第 $n$ 个样本在全连接层2的第 $c$ 个节点的输出值。

采用输出熵约束目标域数据,以降低目标域数据的输出不确定性,从而实现无监督训练过程。其损失函数公式如下:

$$L_E = \mathbb{E}_{(x_t^{(i)}, y_t^{(i)}) \in D_T^{k,t}} [-\log(\hat{y}_c^{(m)})] = -\frac{1}{M} \sum_{m=1}^M \sum_{c=1}^C \hat{y}_c^{(m)} \log \frac{\exp(\hat{y}_c^{(m)})}{\sum_{c=1}^C \exp(\hat{y}_c^{(m)})} \quad (12)$$

式中: $M$ 为样本数量; $C$ 为类别数量; $y_c^{(m)}$ 为取值为0或1的符号函数; $\hat{y}_c^{(m)}$ 表示第 $m$ 个样本在全连接层2的第 $c$ 个节点的输出值。

当 $L_{CKB} = \hat{d}_{CKB}^2(\hat{R}_{XX|Y}^s, \hat{R}_{XX|Y}^t) = 0$ 时,表示源域和目标域数据的条件分布 $P(X|Y)$ 实现了对齐,为了实现联合分布 $P(XY) = P(X|Y)P(Y)$ 的对齐,需要将边际分布 $P(Y)$ 实现对齐,而边际分布可以近似为最大均值差异损失,其计算公式如下:

$$L_{MMD} = \hat{d}_{MMD}^2(X, Y) = \sup_{f \in \mathcal{F}} \left( \frac{1}{N} \sum_{i=1}^N f(x_s^{(i)}) - \frac{1}{M} \sum_{i=1}^M f(x_t^{(i)}) \right) \quad (13)$$

式中: $\mathcal{F}$ 是RKHS中的一个类别函数; $\sup(*)$ 为上确界。

因此,模型的整体损失函数可以表示为:

$$L_{all} = L_C + \alpha L_E + \beta (L_{MMD} + L_{CKB}) \quad (14)$$

式中: $\alpha$ 和 $\beta$ 为权重系数,本文取值均为1。

模型训练过程中,分类器的输出值作为源域有监督和目标域无监督训练的输入,特征提取器的输出值作为域适配训练的输入,通过最小化 $L_{all}$ 的值。

## 2.4 SHAP可解释方法

2017年Shrikumar等<sup>[19]</sup>证明了深度学习重要的特征(deep learning important features, DeepLIFT)可以选择近似的Shapley值,为此,实现了基于SHAP分析的深度学习

模型时候可解释功能。

SHAP分析方法可以简单理解为样本 $x_i$ 的特征 $j$ 对于模型 $f$ 的输出的重要性为特征对 $f(x_i)$ 的贡献在所有可能的特征组合上的加权后:

$$\phi_j^i(f) = \sum_{S \subseteq \{x^1, \dots, x^p\} \setminus \{x^j\}} \frac{|S|! (p - |S| - 1)!}{p!} (f(S \cup \{x^j\}) - f(S)) \quad (15)$$

式中: $S$ 是特征的子集; $p$ 是模型中特征的数量。

在诊断过程中,从轴承故障数据集中随机选取观察的剩余特征,通过代入 $\{x^1, \dots, x^p\} \setminus S$ 值得到相应的 $f(S)$ 。在计算观测 $f(x_i)$ 和背景观测 $f(z)$ 的模型输出时,每个SHAP值 $\phi_j^i(f)$ 是特征 $j$ 导致的差异 $f(x_i) - f(z)$ 的量。所得SHAP值可以直观反应出数据特征对诊断结果的影响程度,通过对各特征的SHAP值进行排序,得到关键性特征。

在提出的方法中,利用SHAP方法对模型的诊断结果进行了基于包络谱的事后可解释分析。SHAP旨在为每个输入特征分配一个贡献值,量化该特征对预测结果的影响。

1) 特征转换,将时域信号转换为包络谱对应的频域信号,提取频域特征。

2) 特征重要性计算,通过SHAP值计算每个频域特征对模型输出的重要性。对于从轴承故障数据集中随机选取的样本特征,计算其SHAP值,并通过代入式(15)得到相应的特征重要性。

3) 结果解释,在计算观测和背景观测的模型输出时,每个SHAP值是特征 $j$ 导致的差异的量,并与传统的故障诊断知识(故障特征频率)相结合,观测出不同特征点对诊断结果的影响程度。

## 2.5 模型诊断及可解释流程

迁移学习旨在识别并利用源域与目标域的共同特征表示,本质上是一种特征空间对齐策略。本文所提的基于CKB的迁移学习方法,不仅能减小源域与目标域间的特征分布差异,而且还能够保持内在的类别划分结构,从而提高跨设备诊断的精度。所提的机械故障诊断可解释方法的整体框架和流程分别如图2和3所示,步骤总结如下。

1) 从不同机械设备中获取轴承振动信号,并将源域和目标域数据以时间顺序划分为训练集和测试集。

2) 利用Hilbert包络谱分析分别从训练集和测试集得到频域数据。

3) 构建深度孪生卷积神经网络,并初始化模型参数。

4) 计算式(11)的整体损失函数,利用Ranger优化器执行反向传播算法,更新模型参数。

5) 反复训练模型,只到达到最大迭代次数,输出训

练好的模型。

6) 将目标域中的测试集数据输入到训练好的模型中,输出模型对健康状态的预测结果。

7) 利用 SHAP 方法对训练好的模型做事后可解释分析,判断哪些特征点对模型的预测起到了关键作用。

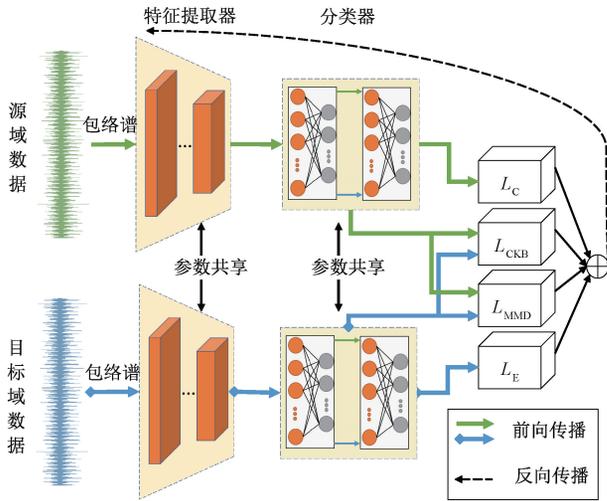


图 2 所提故障诊断可解释方法整体框架  
Fig. 2 Overall framework diagram of the proposed interpretable fault diagnosis method

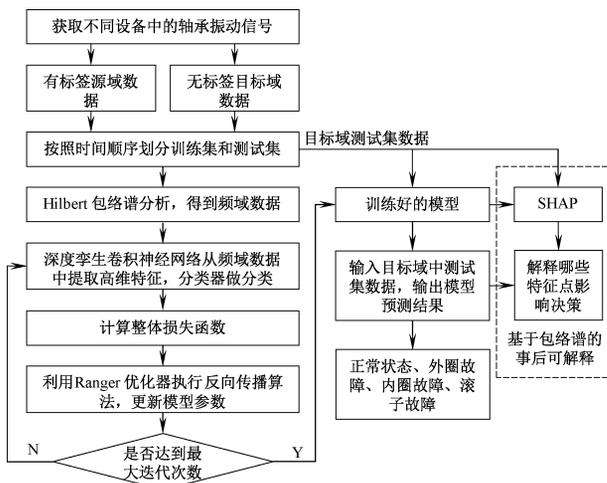


图 3 所提故障诊断可解释方法流程  
Fig. 3 Flowchart of the proposed interpretable method for fault diagnosis

### 3 故障诊断试验验证

#### 3.1 数据集描述

采用从 3 种不同机械设备中收集来的滚动轴承振动信号来制作试验数据集。对于每种数据集来说,均有

4×1 000 个样本,每个样本长度为 2 400 点。数据集如表 2 所示,设备来源如图 4 所示。

表 2 3 种轴承数据集的详细信息

Table 2 Detailed information of the three bearing datasets

数据集	轴承类型	健康状况	标签	采样频率/kHz	转速/(r·min <sup>-1</sup> ), 负载
A	电机轴承	正常	0	12	1 797,0 HP
		外圈故障	1		1 797,0 HP
		内圈故障	2		1 797,0 HP
		滚子故障	3		1 797,0 HP
B	轴支撑轴承	正常	0	20	2 000,26.6 kN
		外圈故障	1		2 000,26.6 kN
		内圈故障	2		2 000,26.6 kN
		滚子故障	3		2 000,26.6 kN
		正常	0		3 532,2.70 kN
C	高速牵引电机轴承	外圈故障	1	10	3 532,3.06 kN
		内圈故障	2		3 485,2.79 kN
		滚子故障	3		3 480,2.84 kN



图 4 数据集对应的机械设备

Fig. 4 Corresponding mechanical equipment diagram

数据集 A 是目前公认的基线测试数据集,来自凯斯西储大学,被测轴承来自于电机的驱动端,采样频率为 12 kHz,包含 4 种健康状态下的轴承振动数据,分别为正常、外圈故障、内圈故障和滚子故障,故障尺寸均为

0.177 8 mm。转速和负载分别为 1 797 r/min 和 0 HP (1 HP $\approx$ 745.7 W)。数据集 *B* 是来自于辛辛那提大学, 实验平台由一个交流驱动电机在 2 000 r/min 转速和 26.6 kN 的负载下经过全寿命周期实验得到的。采样频率为 20 kHz, 实验数据截取在故障发生时间段, 有 4 种健康状态。数据集 *C* 来自于北京交通大学, 实验平台为高速列车牵引电机仿真平台, 采样频率为 10 kHz, 同样有 4 种健康状态下的轴承数据。其中故障宽度为 0.1 mm, 深度为 0.15 mm。

### 3.2 对比方法和试验设置

为了充分证明所提方法的优越性, 采用最近的两种无监督跨设备故障诊断模型和 3 种域适配损失函数来构成对比方法。

1) 深度卷积迁移学习网络 (deep convolutional transfer learning network, DCTLN)。DCTLN 是 Guo 等<sup>[8]</sup>提出的一种无监督跨设备故障诊断方法, 该方法在卷积迁移学习的基础上, 利用最大均值差异损失函数 (maximum mean discrepancy, MMD) 和对抗学习测量, 实现了较好的跨设备故障诊断任务, 为此经常被用来做对比方法来研究。

2) 深度对抗子域自适应网络 (deep adversarial subdomain adaptation network, DASAN)。DASAN 是最新提出的一种跨域故障诊断方法<sup>[9]</sup>, 它结合子域自适应损失函数实现了较高精度的故障诊断。

3)~5) 在本文所提方法的框架下, 将域适配损失函数修改为 MMD、联合最大均值差异损失函数 (joint maximum mean discrepancy, JMMD) 以及子域最大值差异损失函数 (local maximum mean discrepancy, LMMD)。从而对比 CKB 损失函数在所提方法中的高效性。

根据目前的主流任务设定, 设计交叉跨设备故障诊断任务, 即数据集 *A*、*B*、*C* 之间的排列组合。一共有  $A_3^2=6$  种任务。具体来说, 一共有  $A\rightarrow B, A\rightarrow C, B\rightarrow A, B\rightarrow C, C\rightarrow A, C\rightarrow B$  6 种跨设备故障诊断任务, 以  $A\rightarrow B$  来说, 表示源域数据是来自于设备 *A*, 目标域数据来自于设备 *B*。实验数据集设定方面, 参考文献[20], 为了避免数据信息泄露, 以时间的先后顺序来划分训练集和测试集, 每个样本的长度为 1 200 个点。其中训练集和测试集的比例为 0.5, 也就是训练集和测试集的样本数量均为  $4\times 500$ 。在实验过程中, 为了降低单次实验引起的偏差, 每种方法或模型取 5 次实验的平均值和标准差作为最终的诊断结果。详细的参数配置如下: Ranger 优化器实现模型在反向传播过程中参数更新的功能的, 模型迭代次数为 200, 批量大小为 128, 其学习速率为 0.002, L2 重量衰减系数为 0.005。实验在笔记本电脑中进行, 处理器为 AMD Ryzen 95 900HX with Radeon Graphics, 3.30 GHz, GPU 为 3080, 16 G, python 3.9.13, pytorch 2.0.0。

### 3.3 跨设备诊断结果

分别验证提出的方法和上述 5 种对比方法在 6 种跨设备迁移任务中的诊断性能。采用了 3 种性能评估指标, 分别为诊断精度 (accuracy, ACC), F1-分数 (F1-score, F1) 和曲线下面积 (area under curve, AUC)。为了防止单次实验误差, 采用 5 次实验求平均值和标准差, 故障诊断精度统计结果如表 3 所示。

从表 3 可以看出, 提出的方法实现了最高的诊断精度, 平均诊断精度为 99.47%,  $A\rightarrow C, B\rightarrow C, C\rightarrow A$  任务上的诊断精度为 100%, 标准差为 0。表明了所提方法在跨设备故障诊断中的高性能。

DCTLN 方法在 6 种迁移任务上的平均诊断精度为 71.10%, 明显低于其他方法。原因如下: DCTLN 的特征提取能力不足, 无法从原始振动信号中相似的高维特征向量; 在部分迁移任务中, 对抗学习策略对模型起到了负迁移作用, 如任务  $C\rightarrow A$  的诊断精度为 50%。

DASAN 方法的平均诊断精度为 91.70%, 明显要高于 DCTLN 的诊断精度, 这是因为 DASAN 在 DCTLN 的基础上增加了子域适配和目标域伪标签学习策略, 这在数据偏差较大的无监督的跨设备故障诊断任务中起到了关键作用。

对比方法 3)~5) 的诊断精度分别为 96.78%、94.70%、94.39%, 诊断精度要高于 DCTLN 和 DASAN, 这是因为方法 3)~5) 的输入是包络谱信号, 这表明了通过简单的包络谱分析, 可以大幅提高跨设备故障诊断的精度。针对跨设备的复杂迁移针对任务, 与其反复调节模型的参数并为每种迁移任务定制恰当的模型, 不如将模型的时域输入信号转化为故障信息丰富的频域信号。

为了直观地呈现出本文所提方法的优势以及迁移学习的效果, 采用 *t* 分布随机邻居嵌入 (*t*-distributed stochastic neighbor embedding, *t*-SNE) 可视化方法来对每种故障诊断方法的输出特征做分析, 结果如图 5 所示。可以看出, 和其他对比方法相比, 本文方法可以很好的将源域和目标域中相同的标签数据聚类到一起, 类内间隔最小、类间间隔最大, 这是由于 CKB 考虑了条件分布下的判别信息, 表明了本文方法的特征对齐性能最优。

为了定量分析所提方法在条件分布上的具体表现, 对于诊断任务  $A\rightarrow B$  来说, 采用 Wasserstein 距离和条件最大平均差异 (conditional maximum mean discrepancy, cMMD) 来衡量算法的优越性。结果如表 4 所示。可以看出, 本文方法的 EMD 和 cMMD 值最小, 表明了该方法在源域和目标域数据之间具有更小的边缘分布和条件分布差异。再次表明了所提方法在缩小不同域数据之间条件分布差异的性能。

另外, 在参数共享的两个卷积神经网络的作用下, 源域和目标域数据被映射到相似的特征空间中, 这在一定程度上体现了孪生神经网络的特性。而对比其他方法和

表 3 各种方法在跨设备故障诊断任务中的测试结果

Table 3 Test results of various methods in cross-device fault diagnosis tasks

跨设备迁移任务	结果指标	A→B	A→C	B→A	B→C	C→A	C→B	平均值
DCTLN	Acc	84.86±6.79	63.62±10.12	58.31±7.74	76.87±18.43	50.00±0.00	92.91±14.25	71.10
	F1	0.82±0.10	0.59±0.07	0.47±0.10	0.75±0.19	0.47±0.04	0.92±0.17	0.67
	AUC	0.90±0.05	0.76±0.07	0.72±0.05	0.85±0.12	0.67±0.00	0.95±0.10	0.81
DASAN	Acc	97.39±2.50	99.30±0.70	76.82±6.60	98.96±1.53	79.74±12.15	99.63±0.83	91.97
	F1	0.97±0.03	0.99±0.01	0.70±0.10	0.99±0.02	0.79±0.14	1.00±0.01	0.91
	AUC	0.98±0.02	1.00±0.00	0.85±0.04	0.99±0.01	0.86±0.08	1.00±0.01	0.95
MMD	Acc	91.70±4.33	100.00±0.00	99.66±0.38	99.89±0.25	92.30±12.06	97.14±1.95	96.78
	F1	0.91±0.05	1.00±0.00	1.00±0.00	1.00±0.00	0.91±0.16	0.97±0.02	0.97
	AUC	0.94±0.03	1.00±0.00	1.00±0.00	1.00±0.00	0.95±0.08	0.98±0.01	0.98
JMMD	Acc	87.50±5.60	100.00±0.00	91.11±10.06	100.00±0.00	93.76±8.46	95.84±1.97	94.70
	F1	0.86±0.06	1.00±0.00	0.90±0.12	1.00±0.00	0.94±0.09	0.96±0.02	0.94
	AUC	0.92±0.04	1.00±0.00	0.94±0.07	1.00±0.00	0.96±0.06	0.97±0.01	0.97
LMMD	Acc	87.97±5.92	94.05±13.30	94.35±4.45	100.00±0.00	99.97±0.04	89.99±6.56	94.38
	F1	0.87±0.07	0.93±0.15	0.94±0.05	1.00±0.00	1.00±0.00	0.90±0.07	0.94
	AUC	0.92±0.04	0.96±0.09	0.96±0.03	1.00±0.00	1.00±0.00	0.93±0.04	0.97
本文方法	Acc	97.61±2.09	100.00±0.00	99.39±1.36	100.00±0.00	100.00±0.00	99.82±0.14	99.47
	F1	0.98±0.02	1.00±0.00	0.99±0.01	1.00±0.00	1.00±0.00	1.00±0.00	0.995
	AUC	0.98±0.01	1.00±0.00	1.00±0.01	1.00±0.00	1.00±0.00	1.00±0.00	0.997

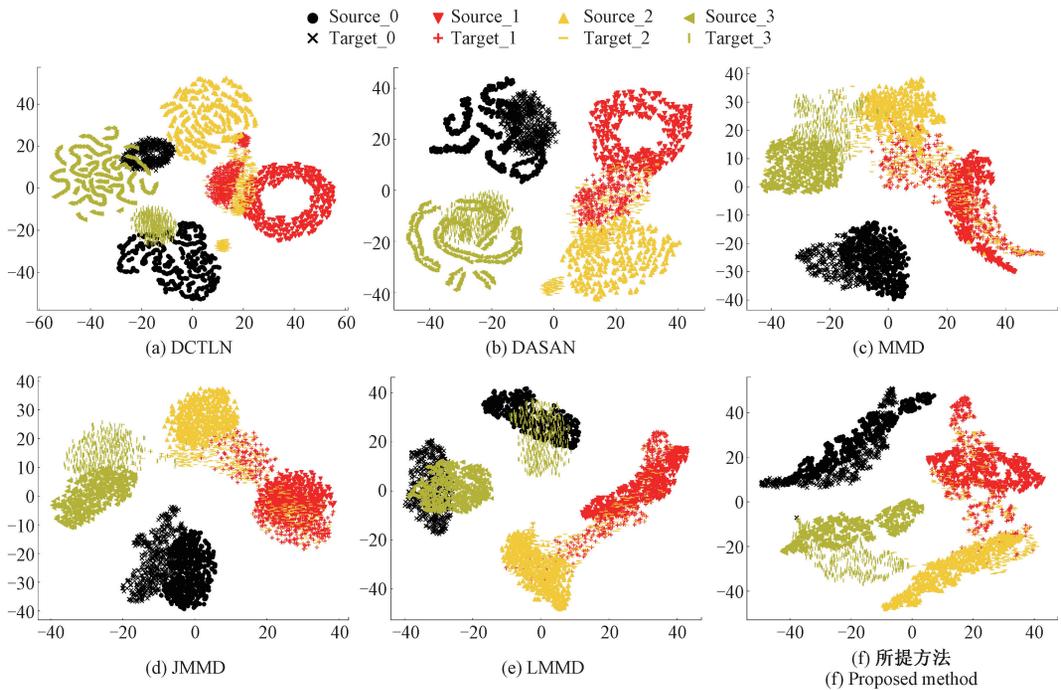


图 5 t-SNE 可视化结果

Fig. 5 Graphs of t-SNE visualization results

表4 不同方法的 Wasserstein 距离和条件最大平均差异结果汇总表

Table 4 Summary table of Wasserstein and cMMD results for different method

方法	Wasserstein 距离	条件最大平均差异
DCTLN	0.165 506 806	0.000 494 510
DASAN	0.214 376 925	0.001 308 235
MMD	1.177 033 163	0.084 701 472
JMMD	0.835 665 316	0.018 384 900
LMMD	0.126 338 644	0.000 482 013
本文方法	0.055 603 957	0.000 156 746

本文方法可以看出,CKB 在模型的特征输出层面缩小了源域和目标域数据间的条件分布差异,提高了模型的迁移和诊断性能。

上述结果和分析表明,所提方法在 6 种跨设备故障诊断任务中取得了最好的诊断精度,充分证明了所提方法的有效性和优越性。

### 3.4 消融试验

为了充分验证所提方法中每个操作步骤的合理性和有效性,开展消融试验。在所提方法的基础上,分别删除包络谱分析、CKB+MMD 损失函数、目标域的输出熵约束,然后采用上述 3 种结构测试 6 种跨设备故障诊断任务中的诊断精度,并与所提方法做比较,对比结果如图 6 所示。可以看出,在删除包络谱和 CKB+MMD 后,模型在不同迁移任务中的诊断效果产生了波动,而在删除熵约束后,模型的诊断精度有所降低。不管怎么说,所提方法的每个步骤都是必不可少的,因此,证明了方法结构安排的合理性。

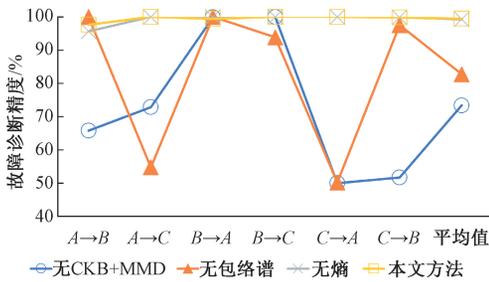


图6 消融试验结果

Fig. 6 Graph of ablation test results

### 3.5 不同的训练集和测试集数量对模型诊断性能的影响

为了探究不同数量的训练集和测试集对不同模型的性能影响规律,在不同的训练集和测试集比例下测试所提方法,MMD、JMMD、LMMD 在 6 种跨设备迁移任务的诊断效果,诊断精度结果如图 7 所示。其中,总共有 6 种数据集比例,即每种健康状态下的训练集和测试集数据

量分别为 50/950、100/900、200/800、300/700、400/600、500/500,可以看出,本文所提方法的精度最高且性能最稳定,在 50/950 的比例下还能保持较好的诊断性能。表明了所提方法的有效性和优越性。

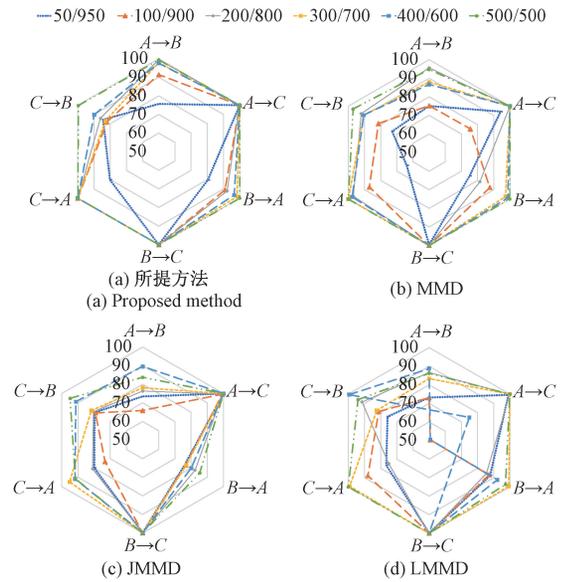


图7 不同数据样本量下模型的性能结果

Fig. 7 Graph of model performance for different data sample sizes

### 3.6 抗噪性能分析和限定条件

考虑到机械设备经常运行在强背景噪声的工作环境,执行抗噪性分析。本文将源域和目标域数据中同时添加不同噪声强度的高斯噪声,噪声强度从 20 dB 降到 -10 dB,间隔 5 dB 取值。所提方法在不同噪声下的平均诊断效果如图 8 所示。可以看出,随着噪声的增加,模型的性能逐渐降低,在 -10 dB 的强背景噪声下,平均诊断精度为 77.55%。从 AUC 的性能评估阈值来看,在 -10 dB 下的 AUC 达到了 0.85,表明了所提方法的分类效果很好。这说

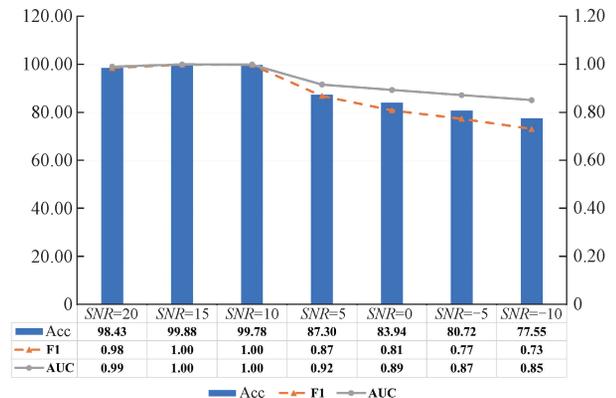


图8 不同信噪比下模型的性能结果

Fig. 8 The graph of the performance results of the model under different signal-to-noise ratios

明了所提方法具有较好的抗噪性。

为了量化所提方法在不同条件分布差异下的适用范围和限定条件,从实验设计和结果分析两方面来进行研究。

### 1) 实验设计

(1) 数据集构建,基于现有的轴承故障数据集,通过添加不同程度的人工噪声,构建一系列具有不同程度条件分布差异的源域-目标域训练集,如图 8 所示。通过调整噪声强度,可以精确控制数据分布差异的程度。

(2) 差异度量,使用两种统计距离(如 Wasserstein 距离和条件最大平均差异)来量化源域和目标域之间的条件分布差异。综合 6 种跨设备任务下的数据分布差异,并衡量不同噪声强度下的数据对应的诊断效果。

(3) 性能评估,对不同噪声强度下的源域-目标域数据,将评估我们方法的故障诊断准确率。

### 2) 结果分析

条件分布差异和诊断准确率变化趋势,如图 9 所示。可以看出,随着噪声强度的增加,数据之间的条件分布差异逐渐增大。通过观察图 9 的诊断性能趋势,设定 80% 的诊断精度为最低阈值,从而确定噪声的临界差异阈值为 -5 dB,当噪声强度高于 -5 dB 时,即 Wasserstein 距离超过 1.11 或条件最大平均差异超过 0.86 时,所提方法的迁移性能开始显著下降,无法准确地识别出故障位置。

### 3.7 事后可解释分析

为了对所提迁移学习模型实现事后可解释,采用基于博弈论的 SHAP 技术对  $C \rightarrow B$  迁移任务做局部样本可解释分析。首先,从测试数据中 4 种健康状态下分别随

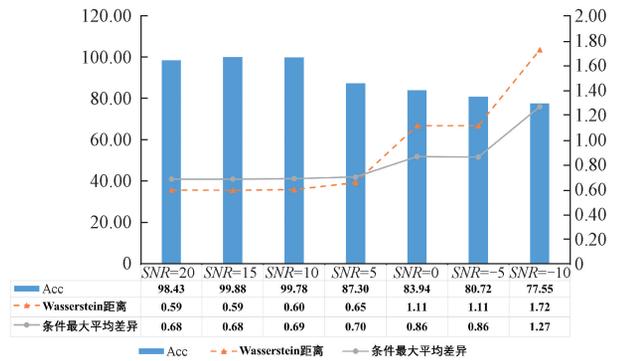


图 9 模型的限定条件分析

Fig. 9 Analysis diagram of the limiting conditions of the proposed model

机抽取的 10 个样本数据,得到  $10 \times 4$  个数据结果,每个结果的大小与包络谱数据大小相同,均为 1 200 点,说明由于包络谱分析内有快速傅里叶变换,故 2 400 点的原始数据转化为 1 200 点的频域数据。然后将上述测试数据输入到训练好的模型中,最后采用 SHAP 方法对该预测过程做分析。得到结果如图 10 所示。灰度较浅的信号为输入的包络谱,正负相间的信号为每个样本点的 SHAP 值。从图 10 可以直观的看出,正常状态下 SHAP 值构成的曲线与包络谱的相似度不高,而故障状态下的 SHAP 曲线与包络谱相似度很高。

另一方面,在正常、外圈故障、内圈故障及滚子故障 4 种健康状态下,采用单个样本做分析,试验结果如图 11 所示。图 11(a) 为输入的 4 种健康状态下的样本信号,为包络谱信号。图 11(b) ~ (e) 构成类似于模型输出的

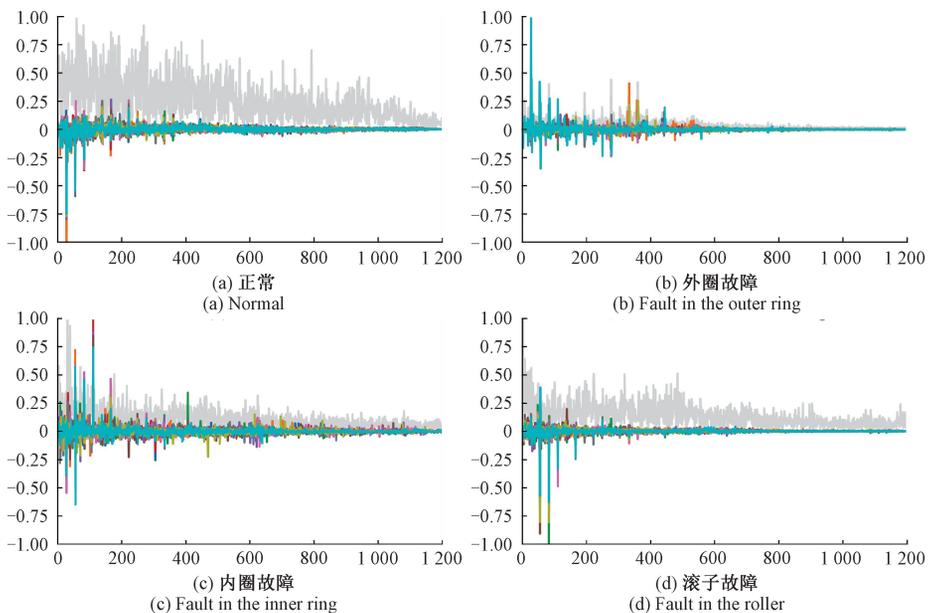


图 10 SHAP 方法在  $C \rightarrow B$  迁移任务上的局部样本可解释而分析结果

Fig. 10 Local sample interpretable while analyzing results of SHAP method on  $C \rightarrow B$  migration task

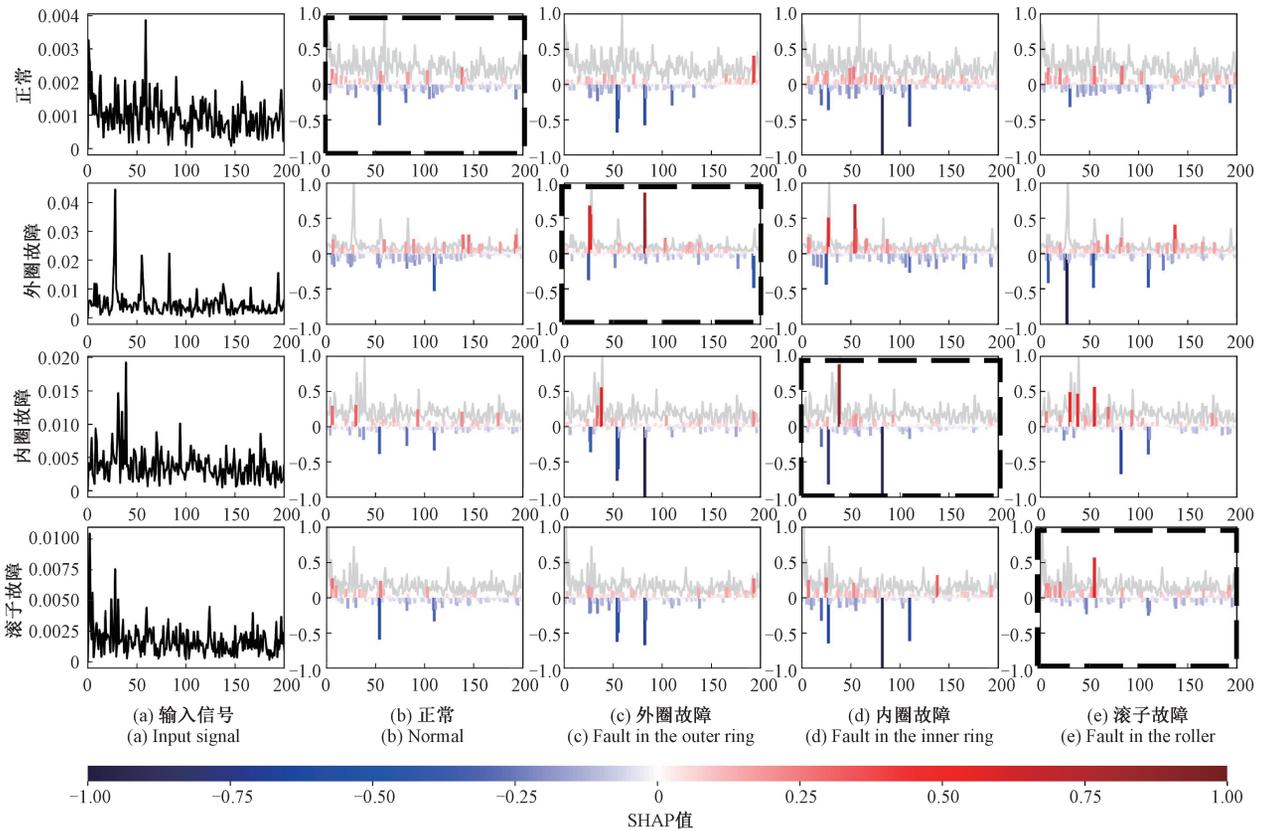


图 11 SHAP 分析方法在单样本上的结果

Fig. 11 Results of SHAP analysis method on single sample

4×4 的混淆矩阵,该矩阵的每一行中的浅灰色背景曲线为真实信号,不同程度的灰度曲线为每个数据点对应的 SHAP 值,对角线(黑色虚线框)对应的 4 幅图为模型在预测值下的可解释结果。例如,图 11(c) 的外圈故障的输出结果可以解释如下:外圈故障信号输入后,经过包络谱分析,得到频域信号,随后该频域信号经过迁移学习模型的预测,在第 2 个神经元节点的位置(即外圈故障处)得到了最大的概率预测结果,被预测为外圈故障信号,随后 SHAP 分析方法对该过程做了定性和定量的解释,SHAP 值越高,代表此处的数据点对模型的决策起到了更为重要的作用,从图 11 可以看出,模型的预测值与输入谱线的高低具有某种非线性关系,而包络谱线恰好是故障特征的间接反映,因此,证明了在本文所提方法中,迁移学习模型与谱线之间具有很强的相关性。

## 4 结 论

提出了一种基于条件度量迁移学习的机械故障可解释方法,解决了跨设备故障诊断任务中存在的诊断精度低、迁移机制可解释性弱的问题。与基于 MMD 的域自

适应方法相比,所提方法在平均诊断精度上提高了 2.69%。通过试验和分析,得出以下结论。

试验结果表明,包络谱分析可以将原始振动信号转为故障信息突出的频域信号,从而提高数据分布偏差较大情况下模型的诊断精度。通过引入 CKB 度量,可以提高模型的迁移能力,缩小了源域和目标域数据之间的条件分布差异,在 Wasserstein 距离和条件最大平均差异上仅有 0.055 603 957 和 0.000 156 746 的差异结果值。同时使得模型的迁移机制具有一定的可解释性。通过 SHAP 事后可解释分析方法揭示了模型预测结果与输入的频域数据之间的关联性。总之,本文所提方法和试验结果为跨设备迁移模型的落地提供了理论参考。

在未来,将继续探索条件核 Bures 度量的更多变种,以进一步减少不同域之间的差异,同时保持模型的可解释性。为了检验所提方法的泛化能力,将在更广泛的故障类型和更多的故障诊断任务下进行试验验证,例如,不平衡样本问题、与齿轮故障相结合的复合故障诊断任务等。

## 参考文献

- [1] 沈飞,陈超,徐佳文,等. 基于时间迁移模型的旋转机械实时故障诊断[J]. 仪器仪表学报, 2019,

- 40(10): 84-94.
- SHEN F, CHEN CH, XU J W, et al. Time transfer model based rotating machine real-time fault diagnosis[J]. Chinese Journal of Scientific Instrument, 2019, 40(10): 84-94.
- [ 2 ] DI MARTINO F, DELMASTRO F. Explainable AI for clinical and remote health applications: A survey on tabular and time series data [ J ]. Artificial Intelligence Review, 2023, 56(6) : 5261-5315.
- [ 3 ] 雷亚国, 杨彬, 李乃鹏, 等. 跨设备的机械故障靶向迁移诊断方法 [ J ]. 机械工程学报, 2022, 58(12) : 1-9.
- LEI Y G, YANG B, LI N P, et al. Targeted transfer diagnosis method across different machines [ J ]. Journal of Mechanical Engineering, 2022, 58(12) : 1-9.
- [ 4 ] LEI Y, YANG B, JIANG X, et al. Applications of machine learning to machine fault diagnosis: A review and roadmap [ J ]. Mechanical Systems and Signal Processing, 2020, 138: 106587.
- [ 5 ] 陈祝云, 钟琪, 黄如意, 等. 基于增强迁移卷积神经网络的机械智能故障诊断 [ J ]. 机械工程学报, 2021, 57(21) : 96-105.
- CHEN ZH Y, ZHONG Q, HUANG R Y, et al. Intelligent fault diagnosis for machinery based on enhanced transfer convolutional neural network [ J ]. Journal of Mechanical Engineering, 2021, 57(21) : 96-105.
- [ 6 ] JIAO J Y, LIANG K X, DING CH C, et al. Towards prediction constraints: A novel domain adaptation method for machine fault diagnosis [ J ]. IEEE Transactions on Industrial Informatics, 2022, 18(10) : 7198-207.
- [ 7 ] DENG M Q, DENG AI D, SHI Y W, et al. A novel sub-label learning mechanism for enhanced cross-domain fault diagnosis of rotating machinery [ J ]. Reliability Engineering & System Safety, 2022, 225 : 108589.
- [ 8 ] GUO L, LEI Y G, XING S B, et al. Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data [ J ]. IEEE Transactions on Industrial Electronics, 2019, 66(9) : 7316-25.
- [ 9 ] LIU Y, WANG Y, CHOW T W S, et al. Deep adversarial subdomain adaptation network for intelligent fault diagnosis [ J ]. IEEE Transactions on Industrial Informatics, 2022, 18(9) : 6038-46.
- [ 10 ] YANG B, LEI Y G, XU S C, et al. An optimal transport-embedded similarity measure for diagnostic knowledge transferability analytics across machines [ J ]. IEEE Transactions on Industrial Electronics, 2022, 69(7) : 7372-82.
- [ 11 ] PANDEY P, RAI A, MITRA M. Explainable 1-D convolutional neural network for damage detection using Lamb wave [ J ]. Mechanical Systems and Signal Processing, 2022, 164: 108220.
- [ 12 ] LU F Y, TONG Q B, FENG Z W, et al. Explainable 1DCNN with demodulated frequency features method for fault diagnosis of rolling bearing under time-varying speed conditions [ J ]. Measurement Science and Technology, 2022, 33(9) : 095022.
- [ 13 ] KIM M S, YUN J P, PARK P. Deep learning-based explainable fault diagnosis model with an individually grouped 1-D convolution for three-axis vibration signals [ J ]. IEEE Transactions on Industrial Informatics, 2022, 18(12) : 8807-8817.
- [ 14 ] 崔玲丽, 刘银行, 王鑫. 基于改进奇异值分解的滚动轴承微弱故障特征提取方法 [ J ]. 机械工程学报, 2022, 58(17) : 156-169.
- CUI L L, LIU Y H, WANG X. Feature extraction of weak fault for rolling bearing based on improved singular value decomposition [ J ]. Journal of Mechanical Engineering, 2022, 58(17) : 156-169.
- [ 15 ] REN CH X, LUO Y W, DAI D Q. BuresNet: Conditional bures metric for transferable representation learning [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4) : 4198-4213.
- [ 16 ] LUO Y W, REN CH X. Conditional bures metric for domain adaptation [ C ]. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 13984-13993.
- [ 17 ] SONG L, HUANG J, SMOLA A, et al. Hilbert space embeddings of conditional distributions with applications to dynamical systems [ C ]. Proceedings of the 26th International Conference on Machine Learning-ICML/09, 2009: 961-968.
- [ 18 ] TACHET R, ZHAO H, WANG Y X, et al. Domain adaptation with conditional distribution matching and generalized label shift [ J ]. ArXiv preprint arXiv: 2003.

04475, 2020.

- [19] SHRIKUMAR A, GREENSIDE P, KUNDAJE A. Learning important features through propagating activation differences [C]. Proceedings of the 34th International Conference on Machine Learning, 2018, 7(8): 4844-4866.
- [20] ZHAO ZH B, LI T F, WU J Y, et al. Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study [J]. ISA Transactions, 2020, 107: 224-255.

### 作者简介



路飞宇, 2018 年于北华大学获得学士学位, 2021 年于石家庄铁道大学获得硕士学位, 现为北京交通大学博士研究生, 主要研究方向为旋转机械设备状态检测和智能故障诊断。

E-mail: 21117039@bjtu.edu.cn

**Lu Feiyu** received his B. Sc. degree from Beihua University in

2018, and M. Sc. degree from Shijiazhuang Tiedao University in 2021. He is currently a Ph. D. candidate at Beijing Jiaotong University. His main research interests include the status detection and intelligent fault diagnosis of rotating machinery equipment.



佟庆彬 (通信作者), 2009 年于哈尔滨工业大学获得博士学位, 现为北京交通大学教授、博士研究生导师, 主要研究方向为轨道交通、电力电子和能源领域的人工智能和智能测试。

E-mail: qbtong@bjtu.edu.cn

**Tong Qingbin** (Corresponding author) obtained his Ph. D. degree from Harbin Institute of Technology in 2009. He is currently a professor and Ph. D. supervisor at Beijing Jiaotong University. His main research interests include artificial intelligence and intelligent testing in the fields of rail transit, power electronics, and energy.