

面向旋转机械装备的智能故障诊断 通用基础模型研究

李响, 徐宜销, 雷亚国, 李熹伟, 李乃鹏, 杨彬

(西安交通大学现代设计及转子轴承系统教育部重点实验室, 710049, 西安)

摘要: 针对现有旋转机械智能故障诊断方法普适性较差, 通常仅适用于特定型号、结构、工况、测点、载荷等场景的问题, 建立了面向旋转机械装备的智能故障诊断通用基础模型。通过挖掘多种类型旋转机械装备海量状态监测数据, 提出了多源数据结构多尺度自适应对齐方法; 构建了多层级状态融合的智能诊断模型; 建立了对典型旋转机械装备普适性强的智能故障诊断通用基础模型, 并提出了诊断通用基础模型的个体定制化适配方法。在广泛使用的旋转机械装备状态监测数据集上对所提方法进行了验证。实验结果表明: 所提智能诊断通用基础模型能够直接对未知实测装备进行异常检测, 在未进行监督微调的情况下, 模型整体诊断准确率达到 88.5%; 在少量数据微调下, 能够迅速适应实测装备, 实现高达 98.6% 的诊断准确率。此外, 所提数据预处理方法在实现跨设备健康状态信号幅值尺度归一化的同时, 可保持同一设备内健康与故障状态信号间的幅值相对分布不变, 有效保留了关键幅值特征差异。所提出的方法具有良好的工程应用潜力, 有望在工程场景下推广应用。

关键词: 智能故障诊断; 通用基础模型; 旋转机械; 定制化适配

中图分类号: TH17 **文献标志码:** A

DOI: 10.7652/xjtuxb202507001 **文章编号:** 0253-987X(2025)07-0001-12

Research on General Foundation Model for Intelligent Fault Diagnosis for Rotating Machinery

LI Xiang, XU Yixiao, LEI Yaguo, LI Xiwei, LI Naipeng, YANG Bin

(Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System,

Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Given that existing intelligent fault diagnosis methods for rotating machinery often lack generalizability and are typically limited to specific models, structures, operating conditions, measurement points, and load scenarios, a universal fundamental model for intelligent fault diagnosis tailored to rotating machinery is developed. By mining massive volumes of state monitoring data from various types of rotating machinery, a multi-source data structure with a multi-scale adaptive alignment method is proposed. A multi-level state fusion intelligent diagnosis model is constructed, and a universal fundamental model with strong applicability to typical rotating machinery is established. Additionally, a method for individualized customization and adaptation of the diagnosis model is introduced. The proposed method is validated on extensive state monitoring datasets for rotating

收稿日期: 2024-09-30。 作者简介: 李响(1990—), 男, 教授, 博士生导师; 雷亚国(通信作者), 男, 教授, 博士生导师。

基金项目: 国家重点研发计划资助项目(2022YFB3402100); 国家自然科学基金资助项目(52475125, 52005086)。

网络出版时间: 2025-01-02

网络出版地址: <https://link.cnki.net/urlid/61.1069.T.20250102.0812.003>

machinery. Experimental results show that the universal intelligent diagnosis model can directly detect anomalies in unknown measured equipment, achieving an overall diagnosis accuracy of 88.5% without any supervised fine-tuning. With minor fine-tuning using a small amount of measured data, the model rapidly adapts to new equipment and achieves a diagnosis accuracy of up to 98.6%. Furthermore, the proposed data preprocessing method enables cross-equipment signal amplitude normalization while preserving the relative amplitude distribution between healthy and faulty states within the same equipment, effectively retaining key amplitude-based feature differences. These findings demonstrate the strong engineering potential of the proposed method and its promise for widespread application in real-world industrial scenarios.

Keywords: intelligent fault diagnosis; general foundation model; rotating machine; customized adaptation

近年来,随着人工智能技术的突飞猛进,大数据驱动的机械装备智能故障诊断方法取得了显著进步^[1]。以深度学习为代表的智能诊断方法能够对机械时序振动信号的自动处理,完成高效的故障特征提取,形成原始监测数据到故障诊断结果“端到端”的智能诊断范式,在整体取得较高诊断精度的同时显著降低了对专家经验知识的需求,便于工程应用^[2]。

在当前文献中,智能诊断方法在实验场景下取得了较好的诊断效果。然而,当前主流方法往往假设数据服从独立同分布,即有健康状态标注的训练数据与未知状态的测试数据采集于相同(或相似)装备、工况、测点、载荷等条件。但在工程场景下,旋转机械装备型号多样、工况频繁变化、测点难以统一、载荷差异显著,导致监测数据存在明显的分布差异,严重降低了智能诊断模型的实测精度,造成诊断模型通用性较低^[3]。近年来,迁移学习^[4]被成功引入智能诊断问题中并提升了模型的泛化能力。Chen等^[5]提出一种领域对抗迁移网络,通过最小化领域之间的特征分布差异来学习源域与目标域的共性知识,实现了高准确率的跨领域故障诊断。Zhao等^[6]提出了条件加权的 Wasserstein 自编码器迁移诊断方法,通过量化不同领域之间的相似性,解决了传统迁移诊断方法对源域与目标域数据分布差异过大时诊断效果不佳的问题。然而,当下迁移诊断方法主要聚焦于有限场景下的迁移适配,对多场景的普适能力稍显不足。因此,智能诊断模型在工程场景下尚未广泛应用。

近年来,以 ChatGPT、Sora 为代表的通用基础模型在自然语言处理、计算机视觉等领域逐渐兴起^[7-8]。通用基础模型往往以深度神经网络为建模框架,使用海

量数据训练具有大规模参数结构的网络,实现对常规任务的全面覆盖,形成普适性极强的模型基底。2023年,OpenAI 公司发布了大语言模型 GPT-4^[9],该模型在自然语言处理领域表现出色,具备强大的情感分析、语言理解及生成等功能。同年,谷歌发布了基于 Transformer 解码器架构的 Gemini 多模态通用基础模型^[10],该模型能够基于多模态数据生成文本、音频、图像等多模态输出,在广泛的数据处理任务中表现出了强大的多模态数据泛化能力。

受通用基础模型思想启发,基于旋转机械动力学机理,本文通过深入挖掘不同类型、不同装备、不同工况、不同测点、不同载荷等多种场景的状态监测数据,希望建立对典型旋转机械装备及其关键零部件普适性强的智能故障诊断通用基础模型,实现对实测新装备的快速适配,克服当前大数据驱动的建模方法对实测装备标注数据需求量大的限制,提升智能诊断模型的工程适用性。

目前,针对机械装备智能故障诊断的通用基础模型研究处于起步阶段,现有文献主要围绕装备健康管理梳理了通用基础模型构建路线图,讨论数据、模型、算法等方面的适用性。Li等^[11]系统地概述了预测与健康管理领域通用基础模型的关键组成部分和最新发展趋势,针对数据集、安全性、可信度、可迁移性等多方面挑战提出了相应的技术路线。Zhang等^[12]阐述了通用基础模型在智能制造领域的发展现状及其优势,对当前通用基础模型在工程应用中遇到的挑战给出了解决方案,同时介绍了智能制造场景下相关的应用案例。目前文献对具体智能诊断通用基础模型构建方法的研究尚不成熟。

尽管智能诊断通用基础模型的建立预期能够形成通用性强的模型基底,然而现有建模方法难以在

机械故障诊断问题中直接应用,主要存在数据、模型与适配三方面的研究难题。

(1)数据:“大数据”是通用基础模型的重要支撑。在机械装备故障诊断问题中,振动数据往往能够充分反映装备健康状态。然而,在不同装备、工况、测点等条件下,振动数据模式通常存在显著差异,难以直接用于通用基础模型的构建,具体主要表现在振动幅值、采样频率、运行工况不一致等方面。振动数据来源于不同机械装备,其幅值往往差异较大;不同振动传感器采样频率不同,导致时序数据样本结构不一致;不同旋转机械转速工况通常不同,在相同采样时间内包含的故障成分存在差异。因此,振动数据模式的不一致为智能诊断通用基础模型的建立带来了明显阻碍。

(2)模型:当前,多层感知机、卷积神经网络、循环神经网络、长短时记忆等模型已被广泛用于智能故障诊断^[13-17],以 Transformer 结构^[18]为框架的自注意力机制能够捕捉时序数据片段的关键信息,适合对振动数据进行处理。在模型输出结构方面,由于通用基础模型对机械装备有不同层级的诊断结果,如异常检测的二元分类、多部件故障定位、精细

化故障诊断等,因此需要对输出标签及模型结构进行定制化设计。

(3)适配:不同工程场景下,机械装备、工况、数据采集方式等差异显著,监测数据间不可避免地存在分布偏差,导致智能诊断通用基础模型往往难以直接应用。当前,以领域自适应^[19]为代表的迁移学习方法能够提升模型适配性,然而其通常对不同场景数据的可获取性有较高要求,在考虑多场景海量监测数据的通用基础模型构建问题中难以满足。因此,如何基于实测装备的少量数据完成对通用基础模型的快速适配应用是亟需解决的难题。

针对以上难题,本文在数据方面提出了多源数据结构多尺度自适应对齐方法,通过对不同场景振动数据进行时频信息对齐,解决了由振幅、采频、工况等差异引起的信息结构不一致问题;在模型方面,提出了多层级状态融合的智能诊断模型,实现对不同层级健康状态标注信息的高效利用;在适配方面,建立了通用基础模型的个体定制化适配方法,提升了通用基础模型对存在数据分布偏差的实测场景适应性。本文所建立的智能故障诊断通用基础模型框架如图 1 所示。

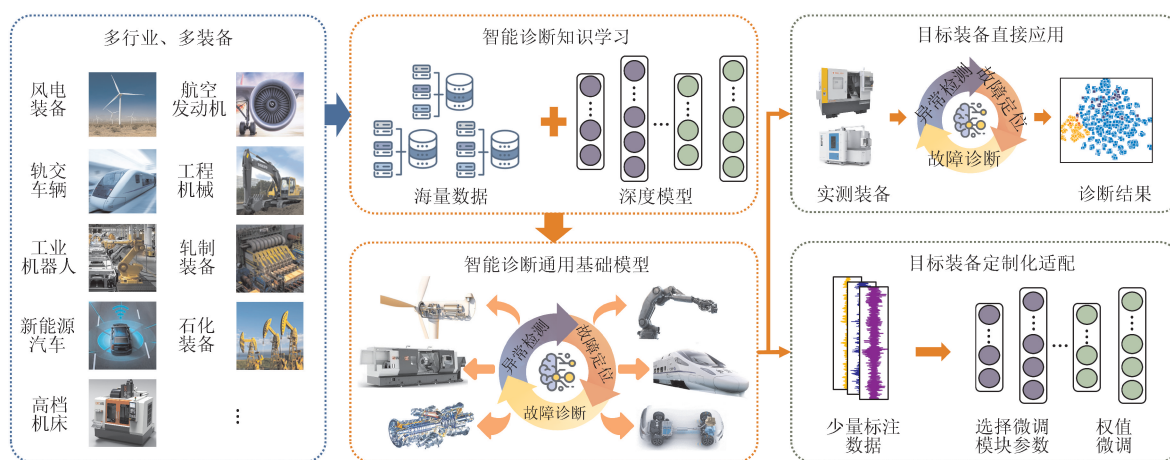


图 1 智能故障诊断通用基础模型框架

Fig. 1 Framework of the intelligent fault diagnosis general foundation model

1 智能诊断通用基础模型构建方法

本文主要面向典型旋转机械及其关键零部件,建立基于振动信号的智能故障诊断通用基础模型。

1.1 方法介绍

在建模过程中,离线可获取的多源旋转机械装备状态监测数据集定义为 $D_i^{\text{Known}} = \{(X_{i,j}, Y_{i,j})\}_{j=1}^{N_i^{\text{Known}}}$, $i = 1, 2, \dots, N_{\text{Dataset}}$, 其中 $X_{i,j}$ 表示第 i 个

装备的第 j 个数据样本, $Y_{i,j}$ 为与 $X_{i,j}$ 对应的健康状态标签, N_i^{Known} 表示第 i 个装备的健康状态数量, N_{Dataset} 表示离线装备数据源的数量。目标实测装备数据集定义为 $D^{\text{Target}} = \{(X_k, Y_k)\}_{k=1}^{N^{\text{Target}}}$, X_k 表示目标装备 D^{Target} 的第 k 个数据样本, N^{Target} 表示目标装备的健康状态数量。

对多源数据集 D_i^{Known} 和目标装备数据集 D^{Target} 进行预处理后的数据集可以分别表示为 $d_i^{\text{Known}} =$

$\{(x_{i,m}, y_{i,m})\}_{m=1}^{N_{\text{Sample}}^{\text{Known},i}}$ 与 $d^{\text{Target}} = \{(x_n, y_n)\}_{n=1}^{N_{\text{Sample}}^{\text{Target}}}$, $i = 1, 2, \dots, N_{\text{Dataset}}$, 其中 $x_{i,m} \in \mathbf{R}^{1 \times L_{\text{Sample}}}$ 表示第 i 个装备的第 m 个样本, $y_{i,m}$ 为与 $x_{i,m}$ 对应的健康状态标签, $N_{\text{Sample}}^{\text{Known},i}$ 表示第 i 个装备数据集的样本数量, $x_n \in \mathbf{R}^{1 \times L_{\text{Sample}}}$ 表示实测装备的第 n 个样本, y_n 为与 x_n 对应的健康状态, $N_{\text{Sample}}^{\text{Target}}$ 表示实测装备的样本数量, L_{Sample} 表示样本维度。

本文通过联合多装备、多测点、多工况的旋转机械装备状态监测数据 d_i^{Known} , 首先对多层级状态融合的智能诊断模型进行预训练得到通用基础模型 $y = f_{\text{Pre-train}}(d_i^{\text{Known}})$; 对于实测场景, 使用少量目标装备数据 d^{Target} 对通用基础模型进行自适应微调得到适配模型 $y = f_{\text{Fine-tune}}(d^{\text{Target}})$, 进而完成测试应用。

1.2 多源数据结构多尺度自适应对齐

1.2.1 多源数据时频域结构对齐

不同场景的旋转机械装备振动数据由于传感器采频、转速工况等因素不同, 存在时频域数据结构不一致的问题。对此, 本文提出多源数据时频域结构对齐方法, 通过将多源数据自适应转化为基准结构, 可以消除不同场景下的数据结构差异, 实现规范化统一建模。方法具体流程如下:

(1) 设定数据基准采样频率 $f_{s, \text{Base}}$ 和旋转机械基准转频 $f_{n, \text{Base}}$;

(2) 比较不同装备传感器的实际采样频率与基准采样频率的差异, 对多源数据进行插值重采样及低通滤波处理, 实现数据时域结构对齐;

(3) 基于多源数据各自转频与基准转频的差异计算伸缩系数 R_{i, f_n} , 进而利用该伸缩系数对多源数据进行时域拉伸与压缩处理, 从而解决转频差异引起的频域结构不一致的问题, 实现转频数据结构对齐。具体计算如下

$$R_{i, f_n} = \frac{f_{n,i}}{f_{n, \text{Base}}} \quad (1)$$

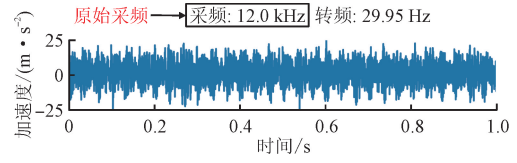
$$X_{i,j, f_n} = f_{\text{Interp}}(X_{i,j}, R_{i, f_n}) \quad (2)$$

$$t_{i,j, f_n} = \frac{t_{i,j}}{R_{i, f_n}} \quad (3)$$

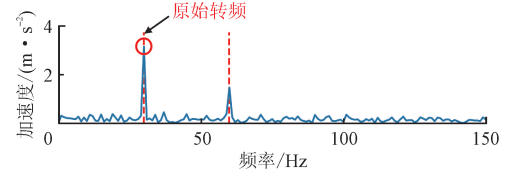
式中: $f_{n,i}$ 表示数据源 i 的转速; $f_{\text{Interp}}(\cdot)$ 表示线性插值计算, 根据伸缩系数 R_{i, f_n} 返回指定信号转频对齐后的信号 X_{i,j, f_n} ; $t_{i,j}$ 、 t_{i,j, f_n} 分别为转频对齐前、后的时间步。

图 2 为采用本文所提方法对某一数据源进行时频域结构对齐的结果。其中, 原始信号采样频率(采频)为 12.0 kHz, 旋转频率(转频)为 29.95 Hz; 目标信号采频为 12.8 kHz, 转频为 20 Hz。图 2(a)、(c)、(e)分别为原始采样信号、时域结构对齐信号、时/频

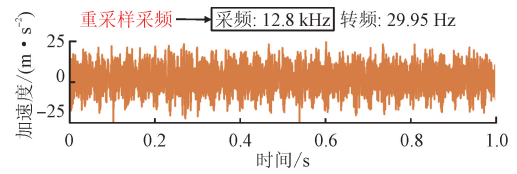
域结构对齐信号的时域波形; 图 2(b)、(d)、(f)分别为原始采样信号、时域结构对齐信号、时/频域结构对齐信号的包络谱。



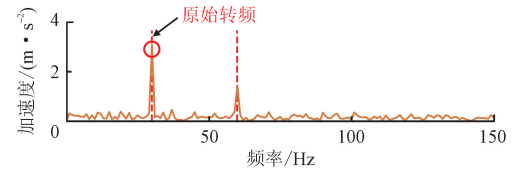
(a) 原始采样信号的时域波形



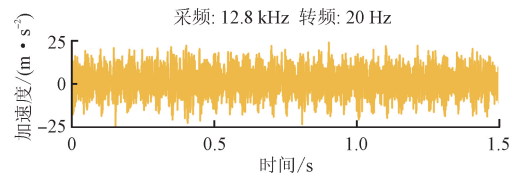
(b) 原始采样信号的包络谱



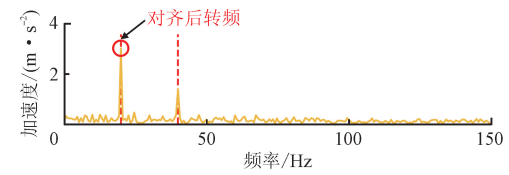
(c) 时域结构对齐信号的时域波形



(d) 时域结构对齐信号的包络谱



(e) 时/频域结构对齐信号的时域波形



(f) 时/频域结构对齐信号的包络谱

图 2 振动数据的时频域结构对齐结果

Fig. 2 Results of time-frequency domain structure alignment of the vibration data

从图 2 可以看出, 本文所提方法在对原始信号进行重采样时对时域波形无明显影响, 能够较好地还原原始信号, 此外频域部分保留了低频信息, 实现了采频与基准采频的对齐。同时, 本文所提方法在

对转频进行对齐时,能够在保持整体时域波形的情
况下,实现转频与基准转频的自适应对齐。

1.2.2 多源数据幅值归一化

不同装备的机械结构、运行环境等因素往往存
在显著差异,导致状态监测振动信号的幅值通常不
一致。例如,大型重载装备在正常状态下的振幅往
往也会高于小型轻载装备在故障状态下的振幅。因
此,本文提出多源数据幅值归一化方法以实现振幅
自适应统一。具体流程如下:

(1)对不同时序振动数据段进行去均值处理;

(2)计算不同装备正常状态数据的特征指标 I

$$I = f_I(X) \quad (4)$$

式中: X 表示输入信号; $f_I(\cdot)$ 为特征指标计算公
式。所提方法能够适用于多种特征指标,在本文实
验中以均方根误差(RMS)特征为例。

(3)从多源数据中选取某一装备的正常状态数
据作为基准数据,并使用基准数据与其他装备数据
特征的比对其余装备数据进行振动幅值缩放,具体
计算式为

$$R_{i, \text{Base}} = \frac{f_I(X_{i, \text{Healthy}} - \mu_{i, \text{Healthy}})}{f_I(X_{\text{Base}, \text{Healthy}} - \mu_{\text{Base}, \text{Healthy}})} \quad (5)$$

$$X_{i, j, \text{Norm}} = R_{\text{Scale}} \frac{X_{i, j} - \mu_{i, j}}{R_{i, \text{Base}}} \quad (6)$$

式中: $X_{i, \text{Healthy}}$ 、 $X_{\text{Base}, \text{Healthy}}$ 分别表示第 i 个装备实体、
基准数据的正常状态信号; $\mu_{i, \text{Healthy}}$ 、 $\mu_{\text{Base}, \text{Healthy}}$ 分别为
 $X_{i, \text{Healthy}}$ 、 $X_{\text{Base}, \text{Healthy}}$ 的均值; $R_{i, \text{Base}}$ 表示第 i 个装备的
幅值比例因子; $\mu_{i, j}$ 表示第 i 个装备第 j 种健康状态
 $X_{i, j}$ 的均值, $i = 1, 2, \dots, N_{\text{Dataset}}, j = 1, 2, \dots, N_i$, 其
中 N_i 表示第 i 个装备的健康状态数量; $X_{i, j, \text{Norm}}$ 表示
 $X_{i, j}$ 归一化后的信号; R_{Scale} 为缩放比例因子。

本文提出的数据归一化方法一方面缓解了多
源数据振动幅值差异大导致的模型过拟合以及单
位不统一问题;另一方面未对同一装备的正常与
故障状态数据进行过多处理,保持了装备原有的
正常与故障状态数据特性。此外,引入了缩放比
例因子,避免了数据归一化后幅值过低导致模型
难以拟合的现象。

图 3 展示了多源数据幅值归一化方法对数据
集 A、B 的处理效果。由图 3 (a)、(e) 可以看出,不
同装备原始波形幅值差异较大,不同装备正常状态
的振动幅值差异较为明显。由图 3 (b)、(f) 可以看
出,经过所提方法预处理后,不同装备正常状态振
动幅值范围相近,同时保留了各个装备正常与故
障信号的幅值差异信息。

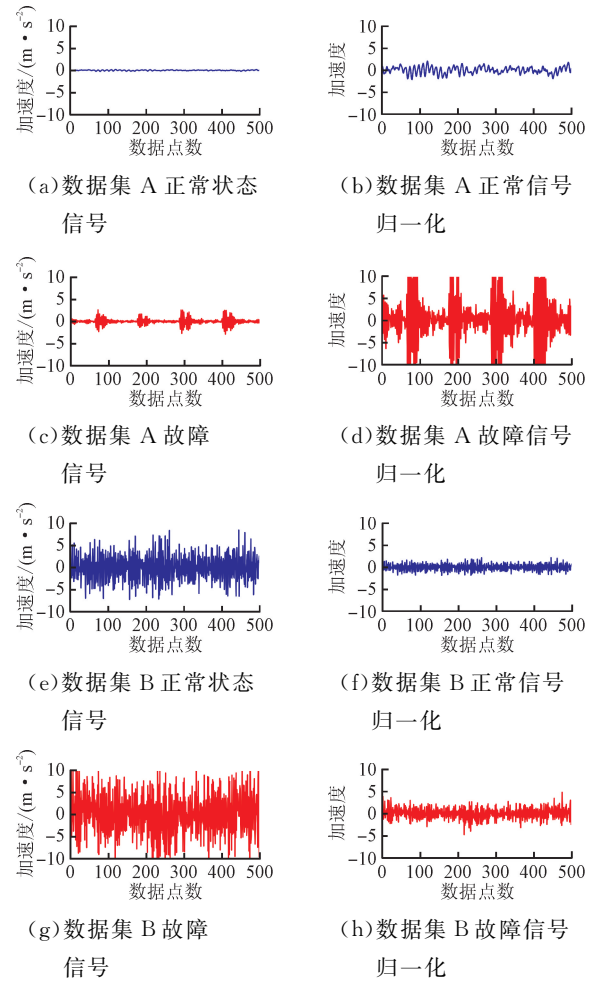


图 3 多源数据幅值归一化方法效果

Fig. 3 Results of the amplitude normalization method for multi-source data

1.3 多层次状态融合的智能诊断模型

图 4 展示了多层次状态融合的智能诊断模型结
构。采用适合时序数据特征提取的 Transformer 深度
神经网络作为智能诊断通用基础模型基础结构,主
要包括数据段嵌入模块、多头注意力特征提取模块
与多层次状态融合的诊断模块。其中,Tokenizer 模
块为一个卷积层, GELU、LeakyReLU 均为激活函数。

(1)数据段嵌入模块主要完成数据样本的预处
理与结构分割,由卷积、位置编码及分类词元构成。
卷积操作的目的是将原始输入数据词元化,生成
长度相同的数据段,使模型能够更好地捕捉时序数
据中的局部与全局特征;位置编码模块生成不同
位置的编码向量,将由卷积操作后的数据段与相
应的编码向量相加,对输入数据段赋予时序位置
信息,提升特征提取效果;分类词元是人为注入的
特殊标记,用于获取原始序列的全局信息,最终用
其状态向量作为诊断的依据。

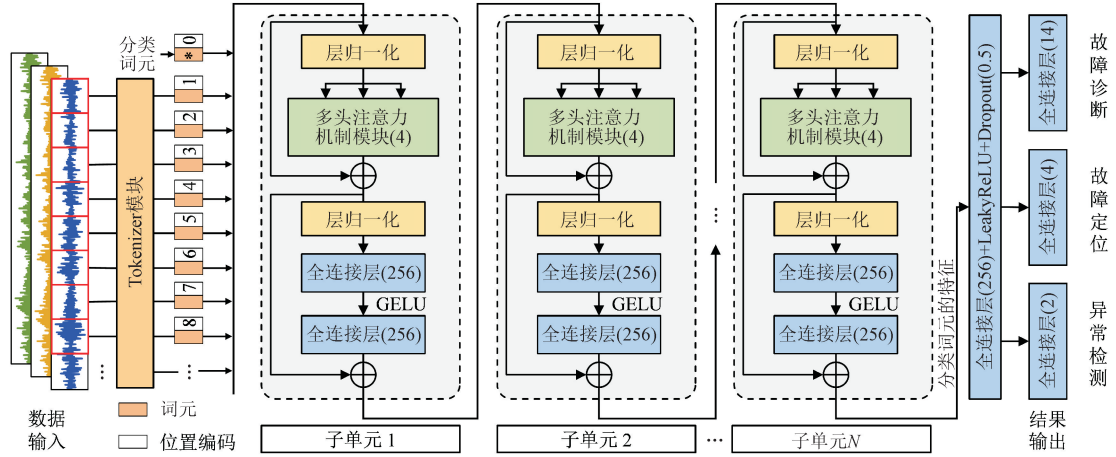


图 4 多层次状态融合的智能诊断通用基础模型框架

Fig. 4 Framework for intelligent diagnostic general foundation model with multi-level state fusion

(2)多头注意力特征提取模块主要由多个重复堆叠的注意力机制子单元构成,其输入与输出数据的维度始终保持一致。单一子单元由两个残差模块串联而成,第一个残差模块由原始输入数据与其标准化后再经过多头注意力特征提取的数据组成,第二个残差模块由输入数据及其经过标准化和全连接层的输出结果组成。

采用残差模块的目的是防止在模型参数较多的情况下训练过程中出现梯度消失等现象,标准化操作通过减小数据间的数值差异来防止出现梯度爆炸现象,多头注意力机制通过在训练过程中对不同时序数据段进行加权聚合并迭代优化,从而实现了对包含故障信息的数据段赋予更大的权重,提升智能诊断效果。

(3)多层次状态融合的诊断模块包含异常检测分类器、故障定位分类器与故障诊断分类器。异常检测分类器对机械健康状态进行低层级的状态识别,通过正常与故障状态二元标签的模型训练,实现对低层级健康状态特征的自动提取,完成异常检测。故障定位分类器考虑旋转机械关键零部件层级的状态识别,针对轴承、齿轮等部件,通过多故障位置标签的模型训练,实现对故障定位特征的智能提取与识别。故障诊断分类器进一步考虑故障类型深层次状态识别,通过将数据状态标签细化,实现对不同部件、不同故障类型的准确诊断。

基于多层次状态融合的智能诊断模型,最终损失函数 L_{All} 及模型参数 $\theta_{f,i}$ 更新过程如下

$$L_1 = -\frac{1}{N_{Data}} \left[\sum_{i=1}^N \sum_{k=1}^{k_1} 1\{y_i = k_1\} \ln \frac{\exp(x_{i,k_1}^{fc_1})}{\sum_{j=1}^{k_1} \exp(x_{i,j}^{fc_1})} \right] \quad (7)$$

$$L_2 = -\frac{1}{N_{Data}} \left[\sum_{i=1}^N \sum_{k=1}^{k_2} 1\{y_i = k_2\} \ln \frac{\exp(x_{i,k_2}^{fc_2})}{\sum_{j=1}^{k_2} \exp(x_{i,j}^{fc_2})} \right] \quad (8)$$

$$L_3 = -\frac{1}{N_{Data}} \left[\sum_{i=1}^N \sum_{k=1}^{k_3} 1\{y_i = k_3\} \ln \frac{\exp(x_{i,k_3}^{fc_3})}{\sum_{j=1}^{k_3} \exp(x_{i,j}^{fc_3})} \right] \quad (9)$$

$$L_{All} = \epsilon_1 L_1 + \epsilon_2 L_2 + \epsilon_3 L_3 \quad (10)$$

$$\theta'_{f,i} \xleftarrow{\text{更新}} \theta_{f,i} - \eta \left(\epsilon_1 \frac{\partial L_1}{\partial \theta_{f,i}} + \epsilon_2 \frac{\partial L_2}{\partial \theta_{f,i}} + \epsilon_3 \frac{\partial L_3}{\partial \theta_{f,i}} \right) \quad (11)$$

式中: L_1 、 L_2 与 L_3 分别为故障诊断分类器、故障定位分类器、异常检测分类器的损失函数; $x_{i,k_1}^{fc_1}$ 、 $x_{i,k_2}^{fc_2}$ 、 $x_{i,k_3}^{fc_3}$ 分别表示第 i 个样本 x_i 在上述 3 个分类器全连接层 fc_1 、 fc_2 和 fc_3 的第 k_1 、 k_2 和 k_3 个神经元的个数; y_i 为与 x_i 相对应的健康状态标签; N_{Data} 表示输入的样本数量; ϵ_1 、 ϵ_2 、 ϵ_3 分别为损失函数 L_1 、 L_2 与 L_3 的权重; $\theta_{f,i}$ 是模型第 i 层的参数权重; η 为学习率。

上述 3 个层次的诊断分类器均由相应的全连接层构成,同时接收主干网络所提取的高级特征为输入进行融合训练与测试。通过这种方式,多源数据的多层级健康状态共性特征能够被高效提取。同

时,由于装备故障诊断在不同工程场景下有不同层级的诊断需求,因而该方法使通用基础模型能够更鲁棒地适用于广泛的应用场景。

1.4 通用基础模型的个体定制化适配

针对实测场景装备监测数据与训练数据分布差异较大,导致智能诊断模型难以适配应用的难题,提出通用基础模型的装备个体定制化适配方法。建立预训练-微调框架,基于通用基础模型,利用实测装备有限数据进行微调,具体步骤如下。

首先,利用多源数据 d_i^{Known} 训练得到智能诊断通用基础模型 $y = f_{\text{Pre-train}}(d_i^{\text{Known}})$;然后,根据实测需求设定所需微调的结构模块 $\theta_{f, \text{Adapt}}$ 并固定模型其余参数,采用实测装备数据调整模型;最后,将个体定制化适配后的智能诊断模型 $y = f_{\text{Fine-tune}}(d^{\text{Target}})$ 用于目标装备的测试应用。微调过程损失函数 $L_{\text{All, New}}$ 的计算及模型权重更新过程如下

$$L_{\text{All, New}} = \varepsilon_1 L_1 + \varepsilon_2 L_2 + \varepsilon_3 L_3 \quad (12)$$

$$\theta'_{f, \text{Adapt}} \leftarrow \theta_{f, \text{Adapt}} - \eta \left(\varepsilon_1 \frac{\partial L_1}{\partial \theta_{f, \text{Adapt}}} + \varepsilon_2 \frac{\partial L_2}{\partial \theta_{f, \text{Adapt}}} + \varepsilon_3 \frac{\partial L_3}{\partial \theta_{f, \text{Adapt}}} \right) \quad (13)$$

本文所建立的智能诊断通用基础模型具有对实测装备数据直接进行异常检测与故障定位的能力,在有少量数据支持的条件下,经过模型微调能够对实测装备形成较高精度的个体定制化智能诊断模型,因而具有良好的工程适用性。

1.5 所提方法整体流程

图5为本文提出的旋转机械装备智能故障诊断通用基础模型建模方法整体流程,详细步骤如下。

(1)设定超参数:依据多源监测数据设定基准采样频率、基准转频等超参数并初始化深度神经网络模型。

(2)多源数据结构多尺度自适应对齐:针对多源数据执行式(1)~(6)所示的时频域结构对齐与幅值归一化,完成数据预处理。

(3)健康状态标签设定:根据不同装备健康状态标注信息,设定多层级标签构建训练样本。

(4)建立通用基础模型:采用多源数据样本训练智能诊断通用基础模型,直至模型收敛。

(5)个体定制化适配:针对实测装备获取有限数据,根据式(13)对通用基础模型进行微调适配。

(6)实测验证:利用适配的智能诊断模型在实测装备上进行测试验证。

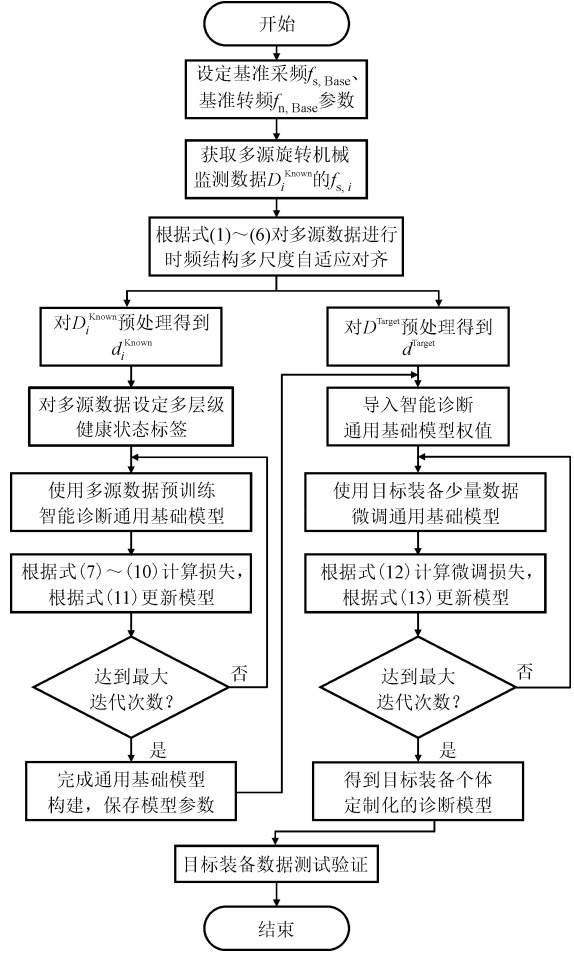


图5 本文所提方法流程图

Fig. 5 Flow chart of the proposed method

本文在构建预训练模型和个体化定制诊断模型时都已采取大量实验,尝试不同实验参数设置。其中,最大迭代次数的设定是基于模型训练过程中达到收敛状态所需的具体迭代次数。通过实验观察模型的损失函数的变化趋势,当损失函数在某一迭代次数后趋于平稳,且后续增加迭代次数并未显著降低损失时,认为模型已经达到收敛状态。

2 智能诊断实验验证

2.1 旋转机械装备状态监测数据集

旋转机械状态监测“大数据”是构建通用基础模型的重要基础,本文利用当前文献中被普遍采用的旋转机械故障模拟公开数据与研究团队实验室故障模拟数据进行研究,构建了广泛的旋转机械装备状态监测数据集,涵盖轴承、齿轮等多种关键零部件的常见故障类型,同时考虑不同转速工况以及不同振动传感器采样频率等条件,详细描述如表1。

表 1 旋转机械装备状态监测数据信息

Table 1 Information of rotating machinery equipment condition monitoring data

代号	数据集名称	主要故障类型	转频/Hz	采样频率/kHz
A	凯斯西储大学轴承故障数据集 ^[20]	(轴承)内圈故障、外圈故障、滚动体故障	28.83,29.17,29.53,29.95	12.0
B	帕德博恩大学轴承故障数据集 ^[21]	(轴承)内圈故障、外圈故障、内外圈复合故障	15,25	64.0
C	江南大学轴承故障数据集 ^[22]	(轴承)内圈故障、外圈故障、滚动体故障	10.00,13.33,16.67	50.0
D	东南大学轴承故障数据集 ^[23]	(轴承)内圈故障、外圈故障、滚动体故障、内外圈复合故障	20,30	23.8
E	都灵理工大学故障轴承数据集 ^[24]	(轴承)内圈故障、滚动体故障	100,200,300,400,500	51.2
F	机泵循环和故障模拟试验台	(轴承)内圈故障、外圈故障、滚动体故障、滚动体与内圈复合故障、滚动体与外圈复合故障、内外圈复合故障、滚动体与内外圈复合故障	28,30,35,40	12.8
G	RTS 轴承故障数据集	(轴承)内圈故障、外圈故障、滚动体故障	25,30,35,40	12.8
H	东南大学齿轮故障数据集 ^[23]	(齿轮)缺损、缺齿、齿根裂纹、齿面磨损	20,30	23.8
I	单级齿轮故障数据集	(齿轮)行星轮缺齿、行星轮磨损、行星轮裂纹、太阳轮缺齿、太阳轮磨损、太阳轮裂纹、	20,30,40,50	25.6
J	锥齿轮故障数据集	(齿轮)大齿轮剥落、小齿轮剥落、大小齿轮均剥落	35,40,45	12.8
K	混合轮系齿轮故障数据集	(轴承、齿轮)太阳轮剥落、太阳轮断齿、太阳轮磨损、行星轮剥落、行星轮断齿、行星轮磨损、中间轮裂纹、中间轮断齿、中间轮磨损、轴承内圈故障、轴承外圈故障、轴承滚动体故障	20,30,40,50	12.8
L	电机故障数据集	(轴承、电机)滚动体故障、外圈故障、电机缺相、电机断条、电机短路、电机电压不平衡	15,35,45	25.6
M	XJTU-SY 轴承全寿命周期数据集 ^[25]	(轴承)内圈故障、外圈故障、保持架故障、内外圈复合故障、内圈保持架滚动体复合故障	35.0,37.5,40.0	25.6

2.2 智能诊断任务

为了验证所提出智能诊断通用基础模型的有效性与优越性,本文考虑了多种故障诊断任务,如表 2 所示。智能诊断任务涵盖了广泛的验证场景,包括对不同装备数据的测试、采用不同数量的训练样本、不同数量的实测数据等,能够较为充分地对所提方法进行验证。本文涉及 13 个数据集,在每个任务中,除非另有说明,预训练阶段使用 12 个数据集,微调和实测阶段使用另外剩余单一目标装备数据集。

本文所采用的超参数如表 3 所示,此类超参数主要基于智能诊断任务 T1 的验证效果选取,并在测试中直接用于其他诊断任务。

表 2 故障诊断任务信息

Table 2 Information of fault diagnostic task

任务	目标装备数据集
T1	凯斯西储大学轴承故障数据集
T2	帕德博恩大学轴承故障数据集
T3	江南大学轴承故障数据集
T4	东南大学轴承故障数据集
T5	都灵理工大学故障轴承数据集
T6	机泵循环故障模拟试验台
T7	RTS 轴承故障数据集
T8	东南大学齿轮故障数据集
T9	单级齿轮故障数据集
T10	锥齿轮故障数据集
T11	混合轮系齿轮故障数据集
T12	电机故障数据集
T13	XJTU-SY 轴承全寿命周期数据集

表 3 本文所提方法采用的超参数

Table 3 Hyperparameters of the proposed method

参数	数值	参数	数值
N_{Dataset}	13	N_{Patch}	16
$N_{\text{Sample}}^{\text{Known}}$	100	N_{Head}	16
$N_{\text{Sample}}^{\text{Target}}$	50	$f_{s,\text{Base}}/\text{kHz}$	12.8
L_{Sample}	1 024	$f_{n,\text{Base}}/\text{Hz}$	20
ϵ_1	1	R_{Scale}	10
ϵ_2	1	η	3×10^{-4}
ϵ_3	1		

2.3 对比方法

本文对比了当下主流的其他智能诊断方法在相应诊断任务中的效果。为了降低模型随机性影响,本文给出的主要诊断结果均为相应 5 次实验的平均结果。

Base^N :基准方法。仅采用实测装备的有限数据进行模型训练与测试,不考虑多源数据的利用与通用基础模型,该方法是当前主流的智能诊断建模方法。

No-Adapt_M^N :在建立智能诊断通用基础模型后直接对目标装备数据进行测试,从而评估所提个体定制化适配方法的有效性。

Adapt-Few_M^N :基于通用基础模型,在适配阶段微调多层级诊断任务分类器参数,评估适配不同结构对模型性能的影响。

Proposed_M^N :即本文所提方法。

2.4 实验结果

2.4.1 整体智能诊断结果

图 6 展示了 2 种方法在不同实验设置下的诊断结果,输入数据均经过所提数据预处理。可以看出,在大多数情况下(如任务 T1~T7),基准方法 Base^N 的诊断准确率均明显低于本文所提方法,且在少量微调数据样本支持下所提方法诊断准确率的提升更为明显。在 T10、T11 等部分实验中,少量样本的支持难以实现通用基础模型对目标装备的高效适配,在样本量较为充足情况下,本文所提方法的诊断结果相较于对比方法有一定提升。以上实验结果说明,本文所提方法在经过预训练后初步具有通用故障特征提取能力,能够对新装备实测数据进行快速适配,但在数据分布差异较大的情况下效果欠佳。这一问题的解决方法包括增加微调样本数量以及在预训练阶段扩充多源数据集的数量等,从而提升通用基础模型的泛化能力。

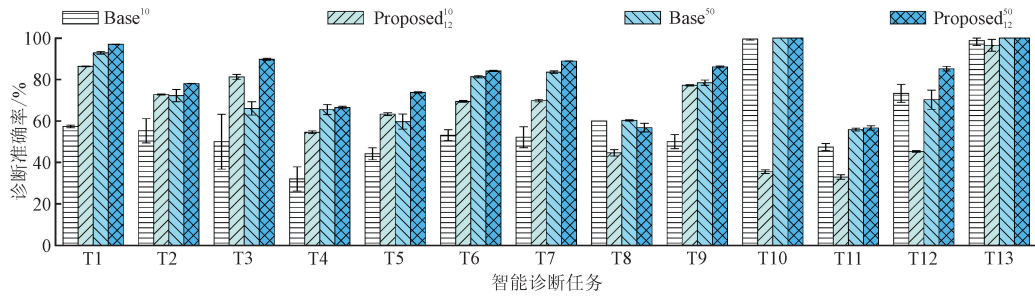


图 6 2 种方法对不同智能诊断任务的测试结果

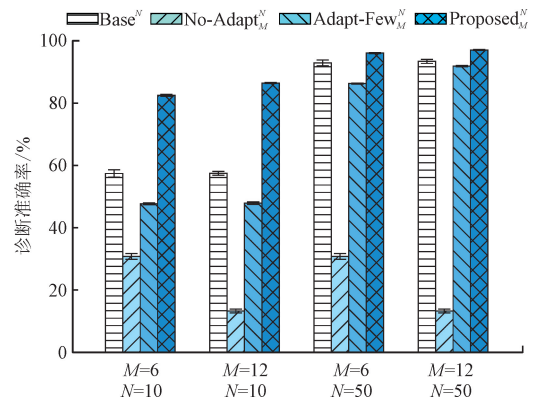
Fig. 6 Test results of two methods for different intelligent diagnostic tasks

2.4.2 多源数据集及微调样本数量对结果影响

本文开展了对比实验,研究多源数据集数量和微调样本量对于诊断结果的影响。图 7(a)、(b)分别为诊断任务 T1 和 T3 在不同预训练数据集及不同微调样本数量下的诊断结果。

从图 7(a)、(b)可以看出,诊断任务 T1 的基准诊断精度相对较高,任务较为简单,而诊断任务 T3 在样本量充足的情况下基准诊断精度也较低,因此诊断任务 T3 较为困难。从图 7(a)可以看出,在较为简单的数据集开展诊断任务,本文所提方法相较于基准方法在少量数据条件下提升明显,在数据量充足情况下也有一定提升。从图 7(b)可以看出,在处理更为困难的诊断任务 T3 时,当预训练数据集数量 M 保持不变时,随着微调样本数量 N 的增多,所提方法诊断

准确率有所提升;当微调样本数量不变时,随着预训练数据集的增加,所提方法诊断准确率也有明显提升。因此,预训练阶段包含的数据集越丰富,通用基



(a) 诊断任务 T1

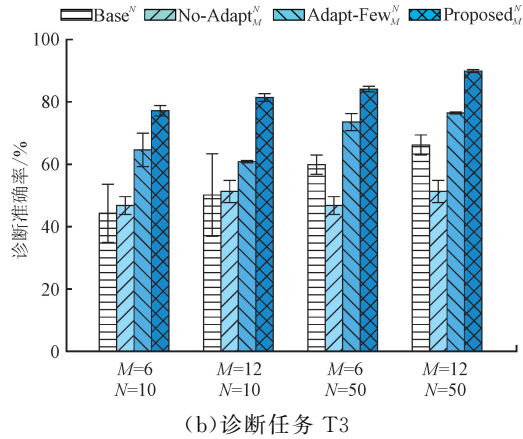


图 7 预训练数据集和微调样本数量对故障多分类诊断结果的影响

Fig. 7 Impact of pre-training dataset, number of fine-tuned samples on multi-type fault diagnostic results

基础模型的通用性就越强,微调阶段数据量的提升也能够进一步增强通用基础模型对实测装备的适配性。

2.4.3 数据预处理对诊断结果的影响

为了进一步探究所提多源数据结构多尺度自适应对齐数据预处理方法对通用基础模型性能的影响,开展了对比实验,结果如图 8 所示。

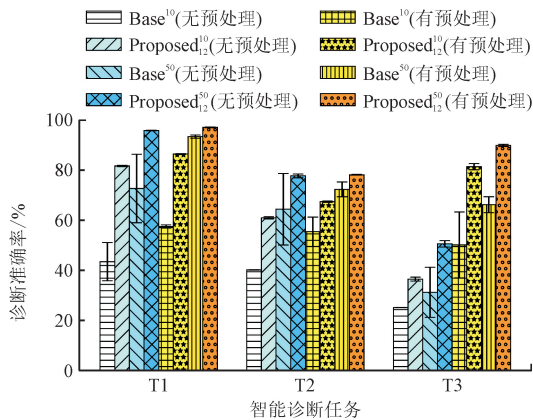


图 8 不同诊断任务中多源数据结构多尺度自适应对齐数据预处理对结果的影响

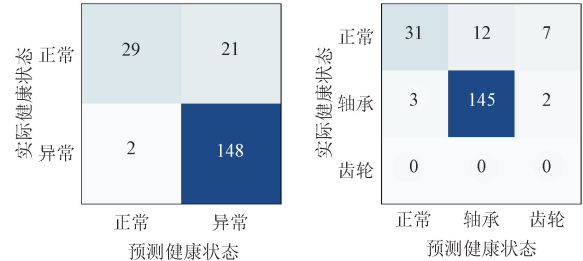
Fig. 8 Impact of multi-source data structure multi-scale adaptive alignment data preprocessing on results in different diagnostic tasks

图 8 表明:①与不做预处理相比,本文所提方法在不同任务中的诊断准确率均有一定的提升,体现了所提预处理方法的有效性;②与基准方法相比,所提方法在不同数据预处理条件下的诊断准确率也有明显提升,说明了所提定制化适配方法的有效性。

2.4.4 通用基础模型对实测数据直接诊断性能评估

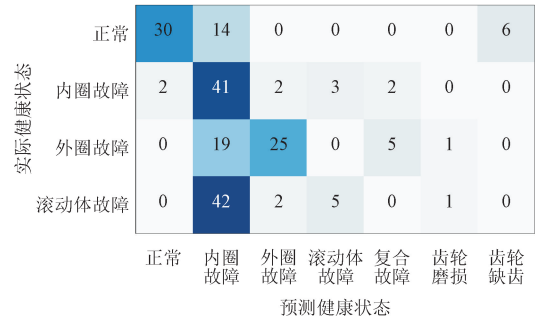
使用通用基础模型对诊断任务 T3 直接进行测试

的混淆矩阵如图 9 所示。对 T3 的异常检测分类器的高层特征进行可视化结果如图 10。其中,对异常检测分类器的高层特征进行可视化,采用 t -SNE 方法进行降维至二维,比较预训练与实测数据正常和异常状态特征的分布情况。



(a) 异常检测分类结果

(b) 故障定位分类结果



(c) 故障诊断分类结果

图 9 T3 任务的各分类器的混淆矩阵

Fig. 9 Confusion matrix of each classifier for diagnostic task T3

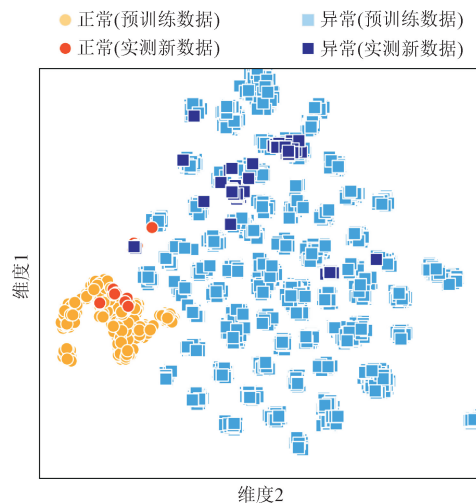


图 10 通用基础模型对新装备实测数据直接测试的异常检测模块高层特征可视化结果

Fig. 10 High-level feature visualization results of anomaly detection module for direct testing of new equipment real-world data by base model substrate

从各分类器的混淆矩阵和特征可视化结果可以看出,经过预训练阶段构建的展现了对未见实测数据的诊断能力,在异常检测与故障定位任务中能够对实测数据进行初步诊断。

在更为具体的故障诊断任务中,存在部分滚动体故障误分类现象,原因是在没有针对性微调的情况下,模型只能依赖于预训练阶段的权重来进行故障分类,因此模型未能充分学习到 T3 任务特定的数据特征和健康状态信息。可以从以下两点解决上述误分类现象:一方面在预训练阶段,尽可能扩大数据集的多样性和覆盖范围,特别是包含不同设备、不同故障类型的样本,通过引入更多的数据,帮助模型学习到更广泛的旋转机械故障特征,从而提高模型的通用诊断能力;另一方面,在获得目标设备的带标签数据后,采用该装备的带标签数据集进行有监督微调,提高模型对于该装备的适配能力;当目标设备的数据缺少标签时,则采用迁移学习的策略,使用如领域自适应、对抗学习等方法对源域与目标域数据进行特征对齐,以提高模型对目标装备的适配能力。

实验结果证明了所构建的智能诊断通用基础模型对于异常检测具有较强的泛化能力,有望在工程场景下进行应用。

3 结论与讨论

本文提出了面向旋转机械装备的智能故障诊断通用基础模型建模方法;建立了多源数据结构多尺度自适应对齐方法,解决了数据时频结构不一致的问题;提出了多层级状态融合的智能诊断模型,实现了对装备健康状态标注信息的充分利用;构建了通用基础模型的个体定制化适配方法,提升了模型对实测场景适应性。

采用广泛的旋转机械装备状态监测数据对所提方法进行了充分验证,实验结果表明智能诊断通用基础模型能够有效提升对实测装备的适用性,显著增强诊断模型的泛化应用能力。

尽管通过实验室模拟数据,在智能诊断通用基础模型的初步探索中取得了积极的实验结果,但单纯依赖实验室数据进行建模存在一定的局限性。后续将进一步丰富和扩展数据集的来源,不仅采用实验室模拟数据,还将积极引入工业现场的实际测量数据(包括变转速、不同载荷等多工况数据),以提高模型的泛化能力。

此外,在机械装备状态监测问题中,温度、电流、声音等传感数据已被广泛用于故障诊断,本文针对

振动数据建立了面向旋转机械装备的智能诊断通用基础模型。后续还会考虑引入温度、压力、电流等多模态监测数据,构建多模态监测数据融合的智能诊断通用基础模型,通过融合不同类型的工况参数,以增强模型对设备健康状态的全面感知能力,并考虑跨类型机械(如旋转机械与往复机械)的智能故障诊断通用基础模型建模方法。

参考文献:

- [1] LEI Yaguo, YANG Bin, JIANG Xinwei, et al. Applications of machine learning to machine fault diagnosis: a review and roadmap [J]. Mechanical Systems and Signal Processing, 2020, 138: 106587.
- [2] YU Yaoxiang, GUO Liang, GAO Hongli, et al. Fed-CAE: a new federated learning framework for edge-cloud collaboration based machine fault diagnosis [J]. IEEE Transactions on Industrial Electronics, 2024, 71(4): 4108-4119.
- [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 Yaguo, YANG Bin, LI Naipeng, et al. Targeted transfer diagnosis method across different machines [J]. Journal of Mechanical Engineering, 2022, 58(12): 1-9.
- [4] LI Weihua, HUANG Ruyi, LI Jipu, et al. A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: theories, applications and challenges [J]. Mechanical Systems and Signal Processing, 2022, 167(Part A): 108487.
- [5] CHEN Zhuyun, HE Guolin, LI Jipu, et al. Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(11): 8702-8712.
- [6] ZHAO Ke, JIA Feng, SHAO Haidong. A novel conditional weighting transfer Wasserstein auto-encoder for rolling bearing fault diagnosis with multi-source domains [J]. Knowledge-Based Systems, 2023, 262: 110203.
- [7] WU Tianyu, HE Shizhu, LIU Jingping, et al. A brief overview of ChatGPT: the history, status quo and potential future development [J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(5): 1122-1136.
- [8] LI Xuan, MIAO Qinghai, LI Lingxi, et al. Sora for scenarios engineering of intelligent vehicles: V&V, C&C, and beyonds [J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(2): 3117-3122.
- [9] WEN Licheng, YANG Xueming, FU Daocheng, et al.

- On the road with GPT-4V(ision): explorations of utilizing visual-language model as autonomous driving agent [C]//ICLR 2024 Workshop on Large Language Model (LLM) Agents. New York, USA: ICLR, 2024: 1-31.
- [10] ANIL R, BORGEAUD S, ALAYRAC J B, et al. Gemini: a family of highly capable multimodal models [EB/OL]. (2024-06-17)[2024-09-01]. <https://arxiv.org/abs/2312.11805>.
- [11] LI Yanfu, WANG Huan, SUN Muxia. ChatGPT-like large-scale foundation models for prognostics and health management: a survey and roadmaps [J]. Reliability Engineering & System Safety, 2024, 243: 109850.
- [12] ZHANG Haotian, SEMUJJU S D, WANG Zhicheng, et al. Large scale foundation models for intelligent manufacturing applications: a survey [J/OL]. Journal of Intelligent Manufacturing. (2025-01-04)[2025-01-10]. <https://doi.org/10.1007/s10845-024-02536-7>.
- [13] 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 Zhuyun, ZHONG Qi, HUANG Ruyi, et al. Intelligent fault diagnosis for machinery based on enhanced transfer convolutional neural network [J]. Journal of Mechanical Engineering, 2021, 57(21): 96-105.
- [14] 卞文彬, 邓艾东, 刘东川, 等. 基于改进深度残差收缩网络的风电机组滚动轴承故障诊断方法 [J]. 机械工程学报, 2023, 59(12): 202-214.
- BIAN Wenbin, DENG Aidong, LIU Dongchuan, et al. Fault diagnosis method of wind turbine rolling bearing based on improved deep residual shrinkage network [J]. Journal of Mechanical Engineering, 2023, 59(12): 202-214.
- [15] 王志颖, 李天福, 许文纲, 等. 降噪混合注意力变分自编码器及其在轴向柱塞泵故障诊断中的应用 [J]. 机械工程学报, 2024, 60(4): 167-177.
- WANG Zhiying, LI Tianfu, XU Wengang, et al. Denoising mixed attention variational auto-encoder for axial piston pump fault diagnosis [J]. Journal of Mechanical Engineering, 2024, 60(4): 167-177.
- [16] 康涛, 段蓉凯, 杨磊, 等. 融合多注意力机制的卷积神经网络轴承故障诊断方法 [J]. 西安交通大学学报, 2022, 56(12): 68-77.
- KANG Tao, DUAN Rongkai, YANG Lei, et al. Bearing fault diagnosis using convolutional neural network based on a multi-attention mechanism [J]. Journal of Xi'an Jiaotong University, 2022, 56(12): 68-77.
- [17] 贾峰, 李世豪, 沈建军, 等. 采用深度迁移学习与自适应加权的滚动轴承故障诊断 [J]. 西安交通大学学报, 2022, 56(8): 1-10.
- JIA Feng, LI Shihao, SHEN Jianjun, et al. Fault diagnosis of rolling bearings using deep transfer learning and adaptive weighting [J]. Journal of Xi'an Jiaotong University, 2022, 56(8): 1-10.
- [18] DING Yifei, JIA Minping, MIAO Qiuhua, et al. A novel time-frequency transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings [J]. Mechanical Systems and Signal Processing, 2022, 168: 108616.
- [19] ZHANG Yongchao, JI J C, REN Zhaohui, et al. Digital twin-driven partial domain adaptation network for intelligent fault diagnosis of rolling bearing [J]. Reliability Engineering & System Safety, 2023, 234: 109186.
- [20] SMITH W A, RANDALL R B. Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study [J]. Mechanical Systems and Signal Processing, 2015, 64-65: 100-131.
- [21] LESSMEIER C, KIMOTH O J K, ZIMMER D, et al. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification [C]//PHM Society European Conference. New York, USA: PHM, 2016: 1577.
- [22] ZHAO Chao, ZIO E, SHEN Weiming. Domain generalization for cross-domain fault diagnosis: an application-oriented perspective and a benchmark study [J]. Reliability Engineering & System Safety, 2024, 245: 109964.
- [23] CHEN Mingzhi, SHAO Haidong, DOU Haoxuan, et al. Data augmentation and intelligent fault diagnosis of planetary gearbox using ILoFGAN under extremely limited samples [J]. IEEE Transactions on Reliability, 2023, 72(3): 1029-1037.
- [24] DAGA A P, FASANA A, MARCHESIELLO S, et al. The Politecnico di Torino rolling bearing test rig: description and analysis of open access data [J]. Mechanical Systems and Signal Processing, 2019, 120: 252-273.
- [25] 雷亚国, 韩天宇, 王彪, 等. XJTU-SY 滚动轴承加速寿命试验数据集解读 [J]. 机械工程学报, 2019, 55(16): 1-6.
- LEI Yaguo, HAN Tianyu, WANG Biao, et al. XJTU-SY rolling element bearing accelerated life test datasets: a tutorial [J]. Journal of Mechanical Engineering, 2019, 55(16): 1-6.

(编辑 亢列梅)