

基于改进 MHGAT 的小样本滚动轴承故障诊断方法^{*}

王玉静¹, 姚丽洋¹, 康守强¹, 王庆岩¹, 梁欣涛¹, 谢金宝²

(1. 哈尔滨理工大学测控技术与通信工程学院 哈尔滨, 150080)

(2. 海南师范大学物理与电子工程学院 海口, 571158)

摘要 针对实际工程中滚动轴承带标签故障样本获取困难, 以及不同工况下滚动轴承样本特征分布差异大的问题, 提出一种改进多头图注意力网络 (muti-head graph attention network, 简称 MHGAT) 的小样本滚动轴承故障诊断方法。首先, 利用快速傅里叶变换构建轴承频域样本集, 根据元学习训练策略将样本集按多个任务划分, 每个任务内的样本集细分为支持集和查询集; 其次, 引入高斯原型模块和斯皮尔曼算法改进多头图注意力网络, 完成支持集、原型集和查询集样本间的连接, 实现轴承频域样本有向图的构造; 然后, 将有向图输入多头图注意力网络的图注意力层, 以提取轴承状态特征; 最后, 在任务内外循环过程中分别引入仿射变换和任务辅助损失, 完成循环多次交替迭代, 得到滚动轴承故障诊断模型。实验结果表明, 所提方法在 16 组不同规格小样本实验任务和 3 个数据集小样本实验任务下, 平均诊断准确率分别为 98.75% 和 95.77%, 性能均优于对比方法。

关键词 小样本; 多头图注意力网络; 元学习; 滚动轴承; 故障诊断

中图分类号 TH165; TP18

0 引言

滚动轴承广泛应用于航空航天、风电机组等领域, 其故障大约占机械设备故障的 1/3^[1]。由于滚动轴承多处于正常运行状态, 故障样本稀缺, 导致出现小样本故障诊断的难题^[2]。同时, 不同工况下滚动轴承故障样本特征分布存在差异, 进一步增加了诊断难度。因此, 实现不同工况下小样本滚动轴承的故障诊断具有重要意义。

随着人工智能技术的发展, 深度学习因具备提取深层次特征的能力, 被广泛用于轴承故障诊断。文献[3-4]利用卷积神经网络实现了轴承故障诊断。虽然经典的深度学习方法取得了良好效果, 但这些方法的可解释性较弱^[5]。图神经网络 (graph neural network, 简称 GNN) 作为新兴的深度学习网络, 具有可解释性更强的拓扑结构, 可通过节点间的信息传递与特征聚合, 充分挖掘样本的深层关联信息。文献[6]综合观测数据与先验知识建立相关图, 并将其输入图卷积神经网络 (graph convolution networks, 简称 GCN), 用于识别牵引系统的故障类型。文献[7]将多头图注意力网络引入轴承故障诊断, 得到了较好的分类效果。

GNN 的特征提取与分类能力优于经典的深度学习方法, 但单独使用依然无法解决小样本问题。为了解决轴承带标签故障样本不足的问题, 小样本学习受到关注。由于元学习具备通过先验知识快速泛化新任务的优势, 因此将元学习框架应用到滚动轴承故障诊断领域, 成为有效解决轴承小样本问题的主流方法之一^[8]。然而, 通用的元学习框架未考虑样本间的关系, 若能充分利用支持集样本与查询集样本之间的关系, 则可更好地解决小样本问题^[9]。GNN 可以实现样本间的信息交互, 将元学习框架与 GNN 相结合来解决小样本问题已成为研究热点。Garcia 等^[10]将 GNN 与元学习框架相结合, 解决了小样本图像分类问题。文献[11]将 GNN 及其变体与元学习框架相结合, 在小样本场景中实现了轴承故障诊断。

上述元学习方法虽能有效解决小样本故障诊断问题, 但在不同工况下模型的泛化能力较弱。针对源域与目标域特征分布差异的问题, 迁移学习有效弥补了该缺陷。文献[12]提出基于 Laplace 小波卷积和双向门控循环单元的轴承故障诊断方法, 利用模型迁移的思想解决小样本跨域故障诊断问题。文献[13]改进域对抗网络, 引入 Wasserstein 距离度量

^{*} 国家自然科学基金资助项目 (52375533); 山东省自然科学基金资助项目 (ZR2023ME057); 哈尔滨市制造业科技创新人才资助项目 (2023CXRC017)

收稿日期: 2024-03-17; 修回日期: 2024-07-17

不同域之间的差异,提高小样本下滚动轴承跨域诊断的精度。由于元学习通过支持集微调模型以适配新任务,其核心机理与迁移学习思想相近,因此诸多研究基于元学习方法进行优化改进,用以解决变工况下的小样本轴承故障诊断问题。文献[14]提出一种新型元学习算法,利用二次梯度规则更新不同任务下的模型参数,提升了变工况下小样本轴承故障诊断模型的泛化能力。文献[15]将元学习框架与孪生网络相结合,通过源域训练得到性能良好的孪生网络与对比样本集,以实现小样本轴承跨域故障诊断。

对于不同工况下的小样本轴承故障诊断问题,现有元学习方法虽然取得了较好的识别效果,但仍存在以下问题:①在现有的元学习微调操作中,使用低复杂度的神经网络避免出现过拟合现象,限制了模型的特征提取能力;②在不同工况下,元学习模型各元任务间的差异会影响模型最优参数的确定,从而导致模型性能降低。

笔者提出一种改进的元学习框架,并结合改进的多头图注意力网络,构建全新的滚动轴承小样本故障诊断方法。该方法引入高斯原型模块与斯皮尔曼算法对多头图注意力网络进行优化,增强模型特征提取能力;同时引入仿射变换与任务辅助损失,对元学习框架中任务内外循环的损失函数进行改进,减少元任务间差异对模型最优参数的影响,最终实现不同工况下小样本滚动轴承的故障诊断。

1 图神经网络基本理论

1.1 图神经网络

GNN是一种使用图结构完成特征表示的网络结构,图1为图神经网络的特征表示。可以看出,图神经网络是一种连接模型,其构建的图由节点矩阵和邻接矩阵组成。其中,节点矩阵和邻接矩阵分别表示节点的信息及节点间的连接关系。

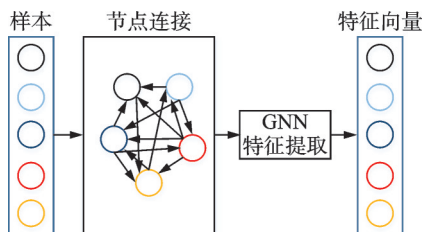


图1 图神经网络的特征表示

Fig.1 Feature representation graphs of GNN

1.2 多头图注意力网络

MHGAT引入了一种可学习的注意力机制,通过在每个源节点与目标节点之间分配权重,使模型在聚合邻居节点特征信息时,能够自适应判别不同邻居节点的重要程度。图2为MHGAT示意图,包括图构建层、具有多头功能的注意力层和全连接层。

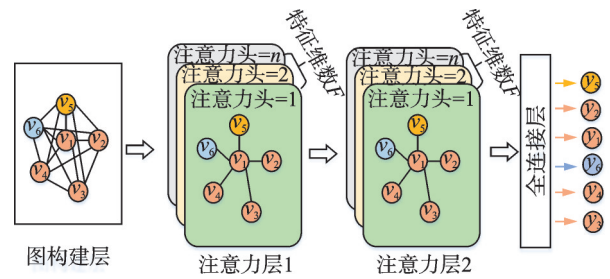


图2 MHGAT示意图

Fig.2 The diagram of the MHGAT

注意力层的计算机制是MHGAT算法的核心,包括以下2个步骤。

1) 通过计算相邻节点之间的注意力系数来表示不同节点之间的相关性。节点 v_i 与邻居节点 v_j 之间的注意力系数 a_{ij} 定义为

$$a_{ij} = \frac{e^{(\text{LeakyReLU}(\delta[\mathbf{W}h_i \parallel \mathbf{W}h_j]))}}{\sum_{k \in N_i} e^{(\text{LeakyReLU}(\delta[\mathbf{W}h_i \parallel \mathbf{W}h_k]))}} \quad (1)$$

其中: $\mathbf{h}_i, \mathbf{h}_j, \mathbf{h}_k$ 分别为节点 v_i, v_j 和 v_k 的特征向量; N_i 为顶点 v_i 的所有邻居节点; $\mathbf{W}$ 为边缘权重矩阵; δ 为计算原始相似度的函数; $[\cdot \parallel \cdot]$ 为向量的拼接;LeakyReLU为激活函数。

2) 运用多头注意机制捕捉相邻节点对中心节点的影响,并对节点信息进行加权和聚合,完成特征提取过程,即

$$\mathbf{h}_i' = \parallel_{l=1}^L \Phi \left(\sum_{j \in N_i} a_{ij}' \mathbf{W}^l \mathbf{h}_j \right) \quad (2)$$

其中: $\mathbf{h}_i'$ 为MHGAT输出节点 v_i 的聚合后特征向量; $\parallel$ 为拼接; Φ 为激活函数; a_{ij}' 为第 l 头节点间的注意力系数; $\mathbf{W}^l$ 为第 l 头节点的边缘权重矩阵。

2 多头图注意力网络模型的改进

2.1 高斯原型模块的构建

在小样本问题中,MHGAT易受离群样本的影响,导致样本信息不足,而原型可以最大程度地保留样本集中相应类别所蕴含的信息。因此,在MHGAT

构图前引入高斯原型模块,计算各类别的原型样本用于训练,可有效捕捉数据特征和类别信息,减弱离群样本的干扰。图3为高斯原型计算过程示意图。

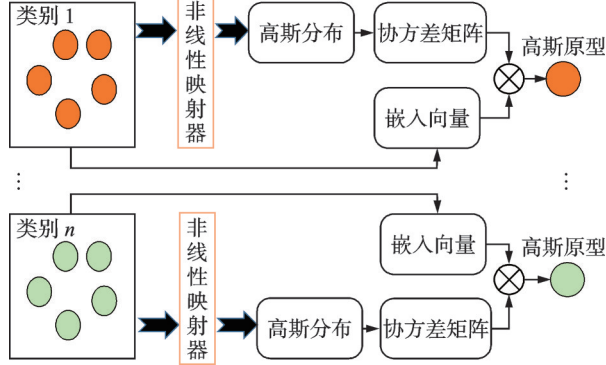


图3 高斯原型计算过程示意图

Fig.3 The diagram of Gauss prototype calculation process

为保证轴承振动信号严格服从高斯分布,将每个类别的样本数据送入非线性映射器,得到对应的高斯分布,变换公式为

$$y_{gi}(\lambda) = \begin{cases} \frac{x_i^\lambda - 1}{\lambda} & (\lambda \neq 0) \\ \ln x_i & (\lambda = 0) \end{cases} \quad (3)$$

其中: $y_{gi} = [y_{g1}, y_{g2}, \dots, y_{gk}]$ 为变换后的数据($i=1, 2, \dots, k$); $x_i = [x_1, x_2, \dots, x_k]$ 为原始轴承数据; k 为同类样本数量; λ 为变换参数。

采用协方差矩阵对角化后的对角矩阵与原始样本向量进行加权求和,得到各类别的原型。各类别样本的权重与原型的计算式分别为

$$\varphi_i(y_{gi}) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y_{gi} - \mu)^2}{2\sigma^2}} \quad (4)$$

$$c_f = \sum_{i=1}^K \varphi_i x_i \quad (5)$$

其中: φ_i 为第*i*个样本的高斯权重; σ 为高斯分布下同类样本的标准差; μ 为高斯分布下同类样本的均值; $f=1, 2, \dots, N, N$ 为样本类别数; c_f 为类别*f*的高斯原型,具有更好的代表性。

高斯原型模块的输出与原始样本共同构成MHGAT的节点矩阵,在图构建层中,设定类别原型节点与同类别样本节点之间的连接权重取最大值。

2.2 图构建层的改进

传统构图方法易受轴承数据分布的影响,导致样本节点的连接具有随机性,信息传递效果受限。斯皮尔曼相关系数将向量变换为等级变量并进行排序,通过观测不同变量间等级的差值进行相关性判

断,能够提供更准确的相关性度量。因此,笔者利用斯皮尔曼算法的思想,对MHGAT图构建层中的构图方法进行改进。

1) 在构图过程中,将所有样本作为节点矩阵,利用斯皮尔曼算法计算不同节点之间的相似程度,得到节点间的相似矩阵。斯皮尔曼算法的计算式为

$$\rho = 1 - \frac{6 \sum_m (R(V_{im}) - R(V_{jm}))^2}{m(m^2 - 1)} \quad (6)$$

其中: m 为节点维数; $R(V_{im})$ 、 $R(V_{jm})$ 分别为节点向量 V_{im} 、 V_{jm} 的秩; ρ 为节点间的相似程度, ρ 值越大,表示两个信号越相似。

2) 根据相似矩阵,采用阈值*T*控制故障样本图的稀疏性,邻接矩阵 E_{ij} 的计算式为

$$E_{ij} = \begin{cases} 1 & (\rho \geq T) \\ 0 & (\rho < T) \end{cases} \quad (7)$$

3) 根据节点间的相似程度,构成不同节点间的权重矩阵。

4) 根据节点矩阵、邻接矩阵和权重矩阵完成构图,得到轴承频域样本的有向图。

3 改进的元学习框架

3.1 元学习训练策略

元学习是一种机器学习框架,通过从多个任务中学习相关经验,从而适应新任务。现有的通用元学习框架是与模型无关的元学习(model-agnostic meta-learning,简称MAML),该框架适用于与深度学习模型相结合的方法。图4为通用元学习算法示意图。可以看出,MAML在内循环中提供初始化参数训练各任务模型,在外循环中寻找普适的初始化方法,从而快速构建新任务模型,解决小样本和跨域等问题。

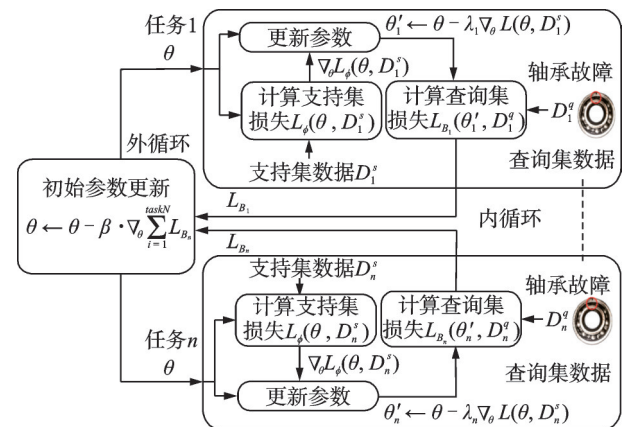


图4 通用元学习算法示意图

Fig.4 The diagram of universal meta-learning algorithm

3.2 改进的内外循环损失函数

当源域与目标域的工况不同时,元学习框架中各训练任务间的差异会影响模型最优参数的确定,

使得模型无法有效应对新的任务。因此,笔者对元学习算法的内外循环损失函数进行改进,减小元任务间的差异,提高模型的泛化能力。图 5 为改进后的元学习算法示意图。

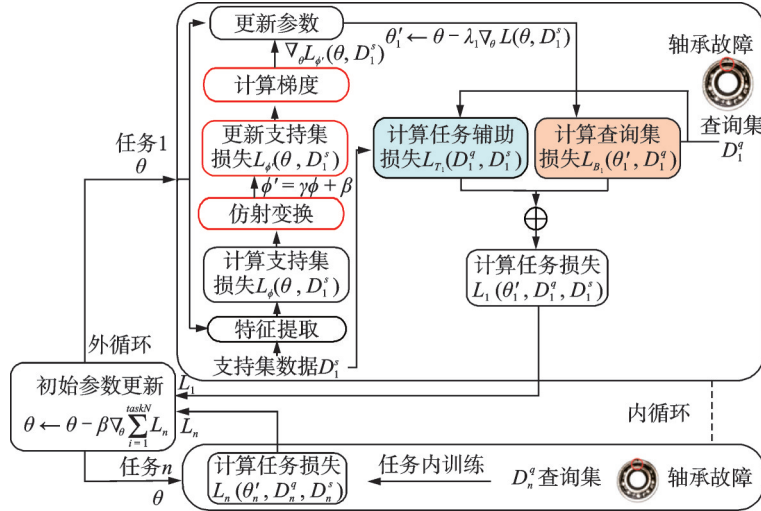


图 5 改进后的元学习算法示意图

Fig.5 The diagram of improved meta-learning algorithm

在任务内循环部分,为了保证每个任务训练好的模型能够更适用于该任务,采用仿射变换更新支持集损失函数的权重参数,使模型在参数更新时能够根据任务目标进一步调整和优化。参数更新过程为

$$\theta'_n \leftarrow \theta - \lambda_n \nabla_{\theta} L_{\phi'}(\theta, D_n^s) \quad (8)$$

其中: θ'_n 为任务 n 更新后的模型参数, θ 为更新前的模型参数; λ_n 为任务 n 下内循环学习率; ϕ' 为优化后的支持集损失参数; $L_{\phi'}(\theta, D_n^s)$ 为优化参数后的支持集损失函数。

根据式(8),利用更新后的参数计算查询集损失函数 $L_{B_n}(\theta'_n, D_n^q)$ 。

在任务外循环部分,为减小各个元任务之间的分布差异,引入任务辅助损失。通过最大限度地保留查询集与支持集节点表示之间的互信息,在查询集信息匹配阶段引入支持集信息,使外循环模型更新时能够考虑支持集的差异影响。

互信息最大化的优化问题可表述为

$$\begin{aligned} \max_{\bar{\theta}} I(Q; S) = & \max_{\bar{\theta}} \sum_{i=1}^Q \sum_{j=1}^N p(q_i, c_j; \bar{\theta}) \log \frac{p(q_i | c_j; \bar{\theta})}{p(q_i; \bar{\theta})} = \\ & \max_{\bar{\theta}} \sum_{i=1}^Q \sum_{j=1}^N p(q_i | c_j) p(c_j) \log \frac{p(c_j | q_i)}{p(c_j)} \end{aligned} \quad (9)$$

其中: Q 为查询集样本合集; S 为支持集样本合集; θ 为框架的优化参数; q_i 为第 i 个查询集样本向量; c_j 为

第 j 类支持集样本的原型。

每个元任务包含 N 个类别,假设 $p(c_j)$ 的先验概率服从均匀分布。若 q_i 属于 c_j 所代表的类别,则 $p(q_i | c_j) = 1$; 否则 $p(q_i | c_j) = 0$ 。式(9)可简化为

$$I(Q; S) = \sum_{i=1}^Q \log(p(c_i | q_i; \bar{\theta})) \quad (10)$$

其中: c_i' 为与 q_i 同类别的支持集原型。

根据欧式距离 L_2 范数准则,利用 Softmax 函数对概率进行归一化处理,定义 $p(c_i | q_i; \bar{\theta})$,则式(10)可表示为

$$\max_{\bar{\theta}} I(Q; S) = \min_{\bar{\theta}} - \sum_{i=1}^Q \log \frac{e^{(q_i \cdot c_i')}}{\sum_{j=1}^N e^{(q_i \cdot c_j)}} \quad (11)$$

任务辅助损失函数为

$$L_{T_n} = - \sum_{i=1}^Q \log \frac{e^{(q_i \cdot c_i')}}{\sum_{j=1}^N e^{(q_i \cdot c_j)}} \quad (12)$$

元训练任务 n 的任务损失为

$$L_n = L_{B_n} + L_{T_n} \quad (13)$$

综合所有任务损失进行外循环训练,完成参数更新,得到普适性模型,用于应对元测试集的新分类任务。图 6 为改进后的元学习训练流程图。

4 小样本下滚动轴承故障诊断方法

在不同工况的小样本下,基于改进多头图注意力

网络模型的滚动轴承故障诊断方法流程如图7所示。

其具体步骤如下。

1) 数据选取及预处理。对滚动轴承原始振动信号进行快速傅里叶变换,得到频域幅值序列,分别构成元训练集和元测试集。从元训练集中随机选取 N 个类别,每类选取 K 个样本作为支持集,从同类别剩余样本中选取 Q 个样本作为查询集,构成一个 N 类 K 样本(N -way K -shot)任务,每次迭代选取不同的任务。例如,5-way 5-shot表示每个任务包含5个类别,每个类别提供5个带标签的样本用于模型对当前任务的快速适配。元测试集以同样方式构建任务进行测试。

2) 改进的多头图注意力网络模型构建。各个任务的支持集通过高斯原型模块得到原型集。支持集、查询集和原型集利用改进的图构建层完成连接,得到轴承频域样本的有向图。之后,将有向图输入搭建好的多头图注意力层,提取各个样本的特征。

3) 改进的元学习框架构建。改进的多头图注意力网络模型构建完成后,计算内循环过程中各任务更新后的支持集损失,迭代完成参数更新,得到仅适合各任务的模型。之后,综合所有任务的损失完成外循环训练,得到训练好的轴承故障诊断模型。

4) 元测试阶段。将元测试集分任务输入到训练好的诊断模型中,利用高斯原型模块计算原型集,使用支持集样本进行内循环训练,实现模型微调。

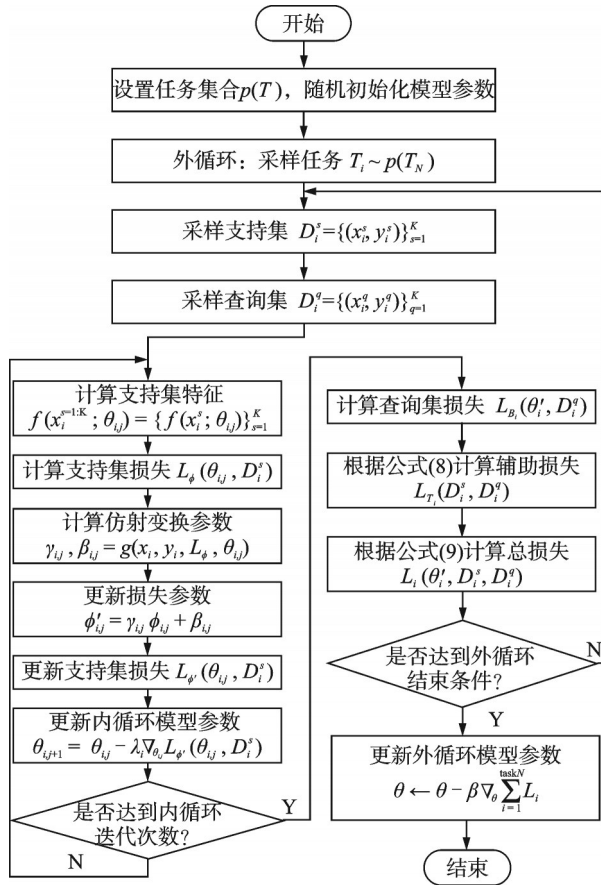


图6 改进后的元学习训练流程图

Fig.6 Flowchart of training process for improved meta-learning

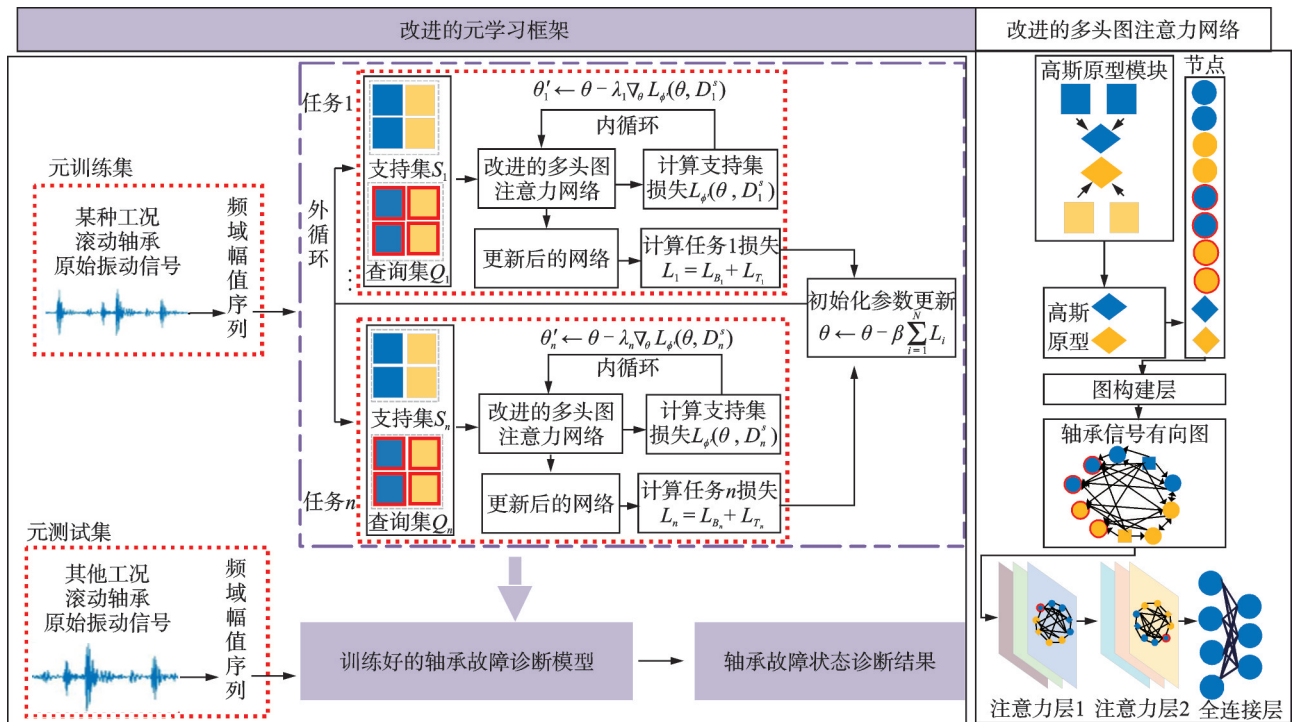


图7 滚动轴承故障诊断方法流程图

Fig.7 Flow diagram of rolling bearing fault diagnosis method

使用微调好的模型对查询集样本进行分类,直接得到轴承故障分类结果。

5 应用与分析

5.1 实验数据集

本研究的实验数据来自美国凯斯西储大学(CWRU)的公开轴承数据集。图 8 为轴承实验台示意图^[16]。实验轴承由驱动端轴承(SKF6205)和风扇端轴承(SKF6203)组成。使用加速度传感器采集轴承数据,采样频率为 12 kHz,此外,驱动端还包含采样频率为 48 kHz 的振动数据。

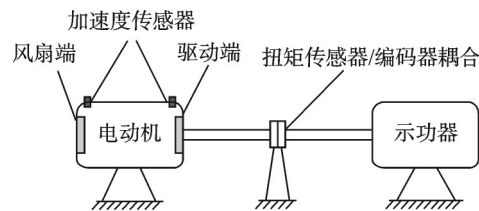


图 8 轴承实验台示意图
Fig.8 Bearing test bench diagram

轴承故障由电火花加工形成的单点故障产生。实验在 0、0.75、1.5 和 2.25 kW 负载条件下,采集 2 种不同规格的滚动轴承信号,得到 10 种故障类型,分为正常状态(N)以及各带有 3 种损伤程度的内圈故障(IR)、外圈故障(OR)和滚动体故障(B)。3 种损伤程度的故障直径分别为 0.178、0.356 和 0.533 mm,分别以数字 07、14 和 21 表示。实验数据集表示方法如表 1 所示。其中,0、0.75、1.5 和 2.25 kW 对应的转速分别为 1 797、1 772、1 750 和 1 730 r/min。

表 1 实验数据集表示方法			
Tab.1 Representation of experimental data			
滚动轴承规格	多种状态	负载/kW	转速/(r·min ⁻¹)
SKF6205	N	0	1 797
	IR07、IR4、IR21	0.75	1 772
	OR07、OR4、OR21	1.50	1 750
	B07、B14、B21	2.25	1 730
	N	0	1 797
SKF6203	IR07、IR4、IR21	0.75	1 772
	OR07、OR4、OR21	1.50	1 750
	B07、B14、B21	2.25	1 730
	N	0	1 797
	IR07、IR4、IR21	0.75	1 772

根据不同工况将实验数据划分为 8 种数据集。表 2 为 CWRU 轴承实验数据集。

将数据集集中的振动信号截断成长度为 2 048 点

表 2 CWRU 轴承实验数据集
Tab.2 Experimental data set settings for CWRU bearings

负载/kW	转速/(r·min ⁻¹)	数据集设置	
		SKF6205	SKF6203
0	1 797	A	E
0.75	1 772	B	F
1.50	1 750	C	G
2.25	1 730	D	H

的样本,元训练集和元测试集每类各 30 个样本。根据元学习训练策略划分 10 个元任务数据,考虑到每个元任务中网络分开训练,对每个元任务进行 N 类 K 样本的小样本设置。通常将每类样本数小于 30 的情况称为小样本,因此在小样本设置中 N 设为 3、5、 K 设为 3、5、10、20 和 30。

基于 CWRU 数据集设置了 16 组不同工况的实验任务,表 3 为 CWRU 不同实验任务设置表。其中:任务 A-D 到任务 D-C 选取 CWRU 数据集中滚动轴承规格为 SKF6205 的不同转速作为源域和目标域;任务 B-G 到任务 H-D 选取 CWRU 数据集中不同规格数据作为源域和目标域。

表 3 CWRU 不同实验任务设置表
Tab.3 The settings of CWRU different tasks

任务	源域	目标域	源域样本个数	目标域样本个数
A-D	A	D	300	300
B-C	B	C	300	300
B-D	B	D	300	300
C-B	C	B	300	300
C-D	C	D	300	300
D-B	D	B	300	300
D-C	D	C	300	300
B-G	B	G	300	300
B-H	B	H	300	300
C-F	C	F	300	300
C-H	C	H	300	300
D-F	D	F	300	300
D-G	D	G	300	300
B-G	B	G	300	300
G-C	G	C	300	300
H-D	H	D	300	300

5.2 实验与分析

5.2.1 斯皮尔曼构图阈值确定实验及实验验证

为了探究构图时阈值对故障诊断准确率的影响,采取不同大小的阈值 T 进行实验。 T 越小,表示构图时对样本间相似程度要求越低。 T 为 0 时,代表所有样本都进行连接; T 为 0.1 时,代表样本的斯皮尔曼相似度大于等于 0.1 时才进行连接。不同阈

值下故障诊断准确率如表4所示。

表4 不同阈值下故障诊断准确率

阈值	任务			
	B-C	B-D	G-C	H-D
0	89.43	89.65	90.94	89.14
0.1	90.47	89.97	91.56	90.06
0.2	94.07	93.25	96.27	92.26
0.3	94.22	93.74	95.56	93.47
0.4	95.79	95.22	95.47	94.48
0.5	97.99	96.69	98.65	97.32
0.6	99.53	99.27	99.34	98.76
0.7	98.47	97.32	98.37	96.15
0.8	94.57	93.97	95.06	93.08
0.9	95.45	94.98	95.55	94.64

由表4可知,阈值为0时,模型的故障准确率并不是最高。为了直观地展示阈值对故障诊断模型性能的影响,绘制不同阈值下4个迁移任务平均故障诊断准确率的点线图。图9为不同阈值下分类准确率对比结果。可以看出,平均准确率在阈值为0.6时达到最高,故采用斯皮尔曼构图时阈值设为0.6。

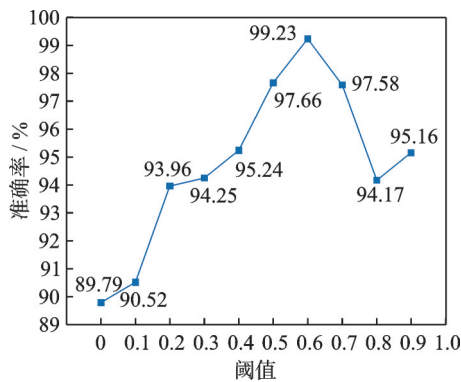


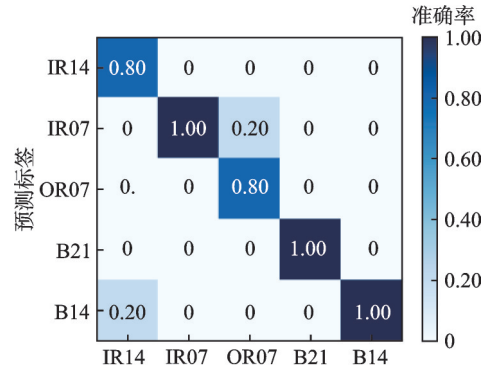
图9 不同阈值下分类准确率对比结果

Fig.9 The comparison result of the classification accuracy for different thresholds

在确定最优阈值后,为了验证斯皮尔曼构图方法的优越性,笔者增设对比实验,与欧氏距离构图方法进行对比。为了保证实验的公平性,欧氏距离也采取上述实验选择合适的阈值,最终确定欧氏距离阈值为0.2时诊断效果最佳。

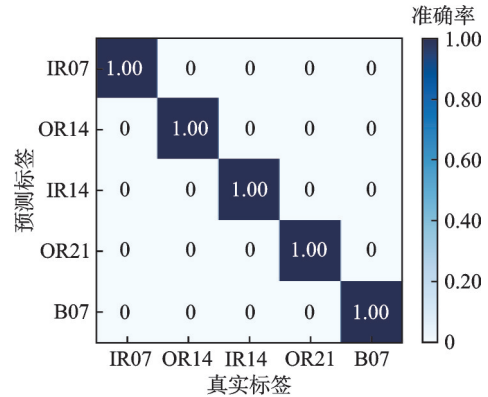
为了直观地观察斯皮尔曼构图方法在故障诊断问题中的有效性,引入多分类混淆矩阵对诊断结果进行分析。在最优阈值下,分别使用斯皮尔曼构图方法和欧氏距离构图方法,进行实验任务为B-C、B-D、G-C、H-D下的5-way 5-shot分类对比实验。

以任务B-C的分类结果为例,不同构图方法的混淆矩阵如图10所示。



(a) 欧氏距离构图方法

(a) Euclidean distance composition method



(b) 斯皮尔曼构图方法

(b) Spearman composition method

图10 不同构图方法的混淆矩阵

Fig.10 Confuse matrix of different composition methods

从图10可以看出,在5-way 5-shot样本设置中,使用欧氏距离构图方法时,IR14、OR07这2类故障被错分,而使用斯皮尔曼构图方法不存在错分。因此,采用斯皮尔曼构图方法可有效提高模型的特征提取能力,从而提升分类诊断效果。

5.2.2 高斯原型模块的消融实验

为验证加入高斯原型模块能够减弱离群样本对相邻样本特征的影响,笔者对加入高斯原型模块前后的分类效果进行对比实验。利用t-分布邻域嵌入算法直观证明高斯原型模块的有效性,选取支持集和查询集每类各一个样本,随机添加信噪比(signal-to-noise ratio,简称SNR)为0、3和6 dB的高斯白噪声来模拟离群样本。其中,信噪比为有用信号的功率与噪声功率的比值,表达式为

$$\delta_{\text{SNR}} = 10 \lg \left(\frac{P_s}{P_n} \right) \quad (14)$$

以实验任务G-C为例,在5-way 20-shot和5-way 30-shot小样本设置下,对查询集样本特征进行可视化处理。图11为加入高斯原型模块前后特

征可视化效果。

图11中,不同数字代表不同类别的样本。从图11(a)和图11(c)可以看出,未加高斯原型模块时,

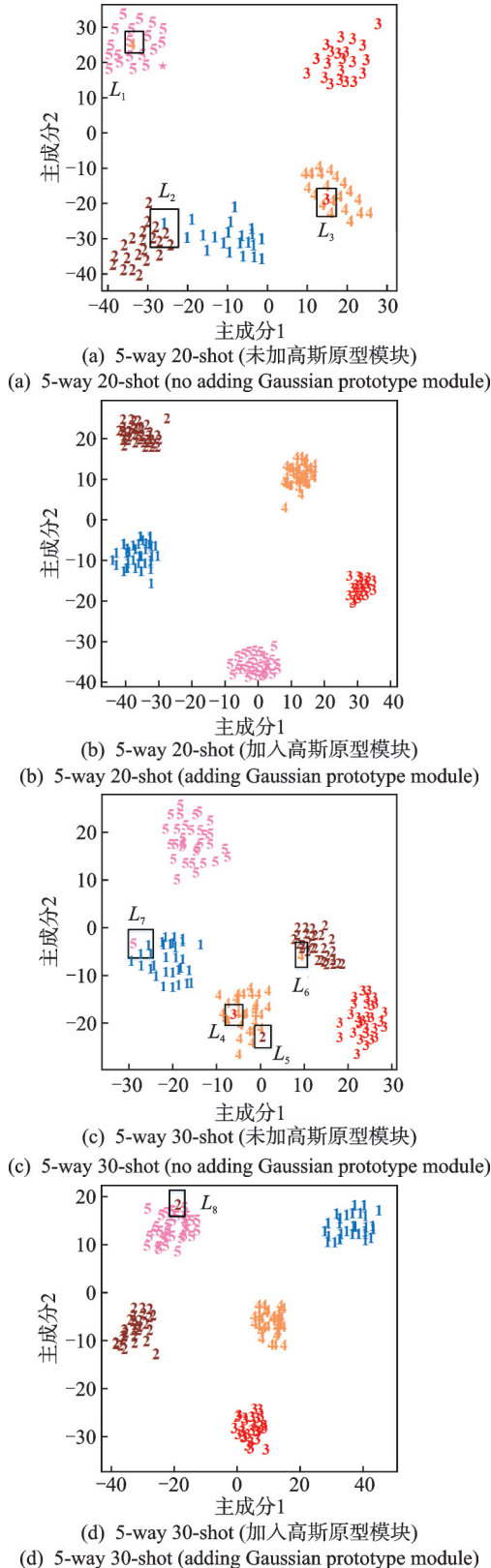


图11 加入高斯原型模块前后特征可视化效果

Fig.11 Feature visualization rendering before and after adding Gaussian prototype module

$L_1 \sim L_7$ 区域均存在混叠,且类间距离较小,分类效果较差。图11(b)和图11(d)显示,加入高斯原型模块后,仅 L_8 区域仍存在错分,说明离群样本对分类的影响减小,类间距离增大,类内距离减小,分类边界明显。因此,高斯原型模块与多头图注意力网络的结合,能够减弱离群样本对相邻样本特征的影响,起到增强特征能力的作用。

5.2.3 所提方法与其他方法对比实验

为了验证所提方法能够用于不同工况下小样本滚动轴承故障诊断问题,笔者将其与文献[12-15]及文献[17]方法进行对比。其中,文献[12]、[13]为迁移学习方法,其余均为元学习方法。所有实验中,源域和目标域每类样本数量为30个,共10类。不同诊断任务下不同方法的对比实验结果如图12所示。不同诊断任务下各方法分类结果汇总如表5所示。

可以看出,所提方法未能在所有实验任务设置下取得最佳分类效果,但在5-way 10-shot设置下所提方法的平均准确率最高,为98.75%。在5-way 5-shot设置下,所提方法的准确率高其他元学习方法。实验任务为D-B时,所有方法的准确率普遍较低,但所提方法在5-way 10-shot设置下的准确率仍达到97.47%,比文献[15]在5-way 5-shot设置下的准确率高9.82%。对比不同转速下7组小样本实验任务,在5-way 10-shot设置下,所提方法的准确率最低为97.47%,最高为99.78%,上下浮动不超过3%。可见,在应对不同转速工况时,所提方法模型具有更好的泛化性。

为了进一步证明所提方法在不同工况下的优越

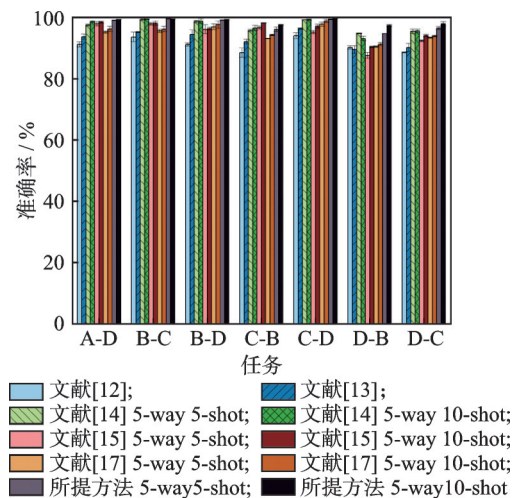


图12 不同诊断任务下不同方法的对比实验结果

Fig.12 Comparative experimental results of different methods under different diagnosis tasks

性,在不同转速的基础上,使用不同规格的数据集进行对比实验验证。源域和目标域分别为每类 30 个样本,共 10 类。不同规格和不同转速设置下,不同方法的分类结果如图 13 所示。可以看出,所提方法

的准确率最高,表现出明显的优越性。其中,在 D-G 实验设置下,所提方法的准确率接近 100%,且所有实验设置下所提方法的准确率均在 90% 以上,表明模型具有良好的泛化性。

表 5 不同诊断任务下各方法分类结果汇总

Tab.5 Summary of classification results of different methods under different diagnosis tasks

%

方法	任务							平均 准确率
	A-D	B-C	B-D	C-B	C-D	D-B	D-C	
文献[12]	91.27	93.66	91.21	88.54	94.11	90.24	88.64	91.10
文献[13]	93.76	95.23	94.52	92.13	96.43	89.56	90.23	93.12
文献[14] 5-way 5-shot	97.56	99.45	98.75	95.78	99.36	94.85	95.50	97.32
文献[14] 5-way 10-shot	98.69	99.58	98.88	96.61	99.40	93.22	95.64	97.43
文献[15] 5-way 5-shot	97.98	97.95	96.21	96.80	95.24	87.65	92.38	94.89
文献[15] 5-way 10-shot	98.56	98.21	96.43	98.19	97.24	90.47	94.16	96.18
文献[17] 5-way 5-shot	95.23	95.64	97.03	93.21	97.85	90.55	93.46	94.71
文献[17] 5-way 10-shot	96.31	96.27	97.75	94.34	98.97	91.35	93.98	95.57
所提方法 5-way 5-shot	99.15	99.68	99.25	96.17	99.47	94.75	96.63	97.87
所提方法 5-way 10-shot	99.44	99.63	99.37	97.62	99.78	97.47	97.97	98.75

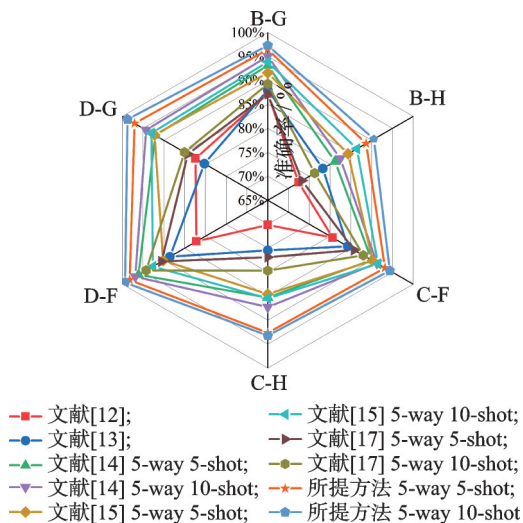


图 13 不同方法的分类结果

Fig.13 The classification results of different methods

5.2.4 数据集泛化实验

为了进一步验证所提方法的有效性和优越性,笔者补充 MFPT 滚动轴承数据集和实验台采集数据集作为实验数据集,进行不同工况下滚动轴承故障诊断的泛化实验。

MFPT 轴承数据规格为 NICE,转速为 25 r/s,采样频率为 48 828 Hz,4 个负载分别为 222.4、444.8、667.2 和 889.6 N。每种负载包含 IR、OR 及 N 这 3 种状态的数据,每种状态有 30 个样本。实验台轴承数据规格为 6307E,转速为 680 r/s,采样频率为 8 192 Hz,包含 IR、OR 及 N 共 3 种状态数据,每种状态有 30 个样本。

对不同负载的样本进行数据集划分,实验数据集不同工况表示方法如表 6 所示。

表 6 实验数据集不同工况表示方法

Tab.6 Representation of different operating conditions in the experimental dataset

负载/N	轴承规格	数据集
222.4	NICE	O
444.8	NICE	P
667.2	NICE	Q
889.6	NICE	R
0	6307E	L

根据 CWRU、MFPT 和实验台数据集设置了 12 个不同工况的实验任务。不同实验任务设置如表 7 所示。

根据表 7 的任务设置进行跨数据集、不同工况的小样本滚动轴承故障分类实验,所有实验均在 3-way

表 7 不同实验任务设置

Tab.7 The settings of different tasks

任务	源域	目标域	源域样本个数	目标域样本个数
B-O	B	O	300	90
C-P	C	P	300	90
D-Q	D	Q	300	90
R-F	R	F	90	300
O-G	O	G	90	300
P-H	P	H	90	300
A-L	A	L	300	90
E-L	E	L	300	90
R-L	R	L	90	90
L-D	L	D	90	300
L-H	L	H	90	300
L-R	L	R	90	90

3-shot、3-way 5-shot和3-way 10-shot的小样本设置下进行。图 14为跨数据集不同工况下的泛化实验结果。表 8为不同工况下跨数据集泛化分类结果汇总。可以看出,所提方法在所有实验设置中的故障准确率均在 90% 以上。在实验设置 B-O、小样本设置为 3-way 10-shot的实验中,故障准确率高达 99.32%。虽然在 6307E 与 CWRU 数据集的跨域实验任务 L-H 中,小样本设置为 3-way 3-shot时的故障准确率最低,但仍达到 90.23%。

综合上述实验结果可知,在 3 种数据集、12 种不同工况下,所提方法的故障分类平均准确率为 95.77%,表现出良好的分类效果。在跨数据集、不同工况的实验任务中,由于规格和负载的不同,振动信号幅值和频谱特征的差异较大,在小样本情况下很难获得高准确率的迁移模型,但所提方法的小样本模型准确率仍达到 90% 以上。经实验验证,所提方法能够有效解决跨数据集下不同规格、不同负载的轴承故障诊断问题,证明了该方法具有很强的泛化性和普适性。

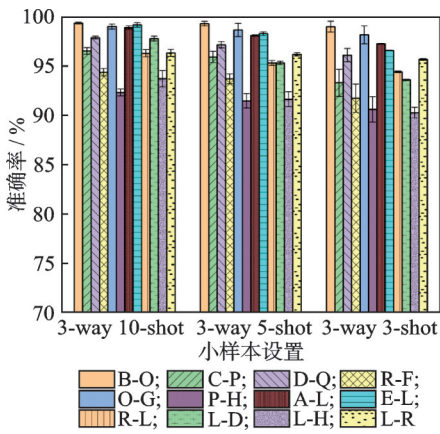


图 14 跨数据集不同工况下泛化实验结果

Fig.14 Generalization experiment results under different working conditions across datasets

表 8 不同工况下跨数据集泛化分类结果汇总

Tab.8 Summary of cross-dataset generalization classification results under different working condition %

任务	3-way 10-shot	3-way 5-shot	3-way 3-shot
B-O	99.32	99.26	98.96
C-P	96.49	95.88	93.26
D-Q	97.84	97.11	96.06
R-F	94.32	93.65	91.69
O-G	98.97	98.62	98.12
P-H	92.27	91.43	90.55
A-L	98.87	98.05	97.21
E-L	99.13	98.24	96.53
R-L	96.25	95.27	94.37
L-D	97.74	95.28	93.56
L-H	93.66	91.59	90.23
L-R	96.28	96.14	95.63

6 结 论

1) 提出在多头图注意力网络中引入高斯原型模块,挖掘类别原型信息,缓解了离群样本带来的错误影响。特征可视化实验结果表明,引入高斯原型模块后,错分现象减少,模型的特征提取能力得到增强。

2) 提出在图构建层中引入斯皮尔曼算法,使构图过程不受轴承信号幅值变化的影响,加强了轴承样本间有效信息的交换。通过不同阈值下 4 种工况的小样本对比实验,确定阈值为 0.6 时模型分类效果最佳,平均准确率高达 99.23%。与经典构图方法相比,在 5-way 5-shot 样本设置下,利用混淆矩阵证明了斯皮尔曼构图能得到性能更优的故障分类模型。

3) 提出一种基于改进多头图注意力网络的小样本轴承故障诊断方法,改进了元学习内外循环过程中的损失函数,将改进后的元学习框架与多头图注意力网络相结合,能够在不同工况的小样本条件下,建立有效的故障诊断模型。实验结果表明,所提方法在 16 组不同规格和 12 组跨数据集的小样本实验任务中,平均诊断准确率分别为 98.75% 和 95.77%,性能优于对比方法。

参 考 文 献

[1] 王玉静,夏林,康守强,等. 基于多源域异构模型迁移的滚动轴承故障诊断方法[J]. 振动与冲击, 2023, 42(24): 257-266.

WANG Yujing, XIA Lin, KANG Shouqiang, et al. A fault diagnosis method of rolling bearings based on multi-source domain heterogeneous model transfer[J]. Journal of Vibration and Shock, 2023, 42 (24) : 257-266. (in Chinese)

[2] 柳雅倩,蔡浩原,李文宽,等. 小样本条件下轴承故障的 DCGAN 诊断方法[J]. 振动、测试与诊断, 2023, 43(4): 817-823.

LIU Yaqian, CAI Haoyuan, LI Wenkuan, et al. Research on deep convolutional generative adversarial networks diagnosis method of bearing fault under small sample condition[J]. Journal of Vibration, Measurement & Diagnosis, 2023, 43(4): 817-823. (in Chinese)

[3] HUANG X F, XIE T L, WU J H, et al. Deep continuous convolutional networks for fault diagnosis[J]. Knowledge-Based Systems, 2024, 192: 111623.

[4] ZHU H Z, FANG T. Compound fault diagnosis of rolling bearings with few-shot based on DCGAN-Re-

- pLKNet[J]. Measurement Science and Technology, 2024, 35: 066105.
- [5] 窦慧, 张凌茗, 韩峰, 等. 卷积神经网络的可解释性研究综述[J]. 软件学报, 2024, 35(1): 159-184.
DOU Hui, ZHANG Lingming, HAN Feng, et al. Survey on convolutional neural network interpretability[J]. Journal of Software, 2024, 35(1): 159-184. (in Chinese)
- [6] CHEN Z W, XU J M, PENG T, et al. Graph convolutional network-based method for fault diagnosis using a hybrid of measurement and prior knowledge[J]. IEEE Transactions on Cybernetics, 2021, 99: 1-13.
- [7] JIANG L, LI X J, WU L, et al. Bearing fault diagnosis method based on a multi-head graph attention network [J]. Measurement Science and Technology, 2022, 33(7): 075012.
- [8] 李凡长, 刘洋, 吴鹏翔, 等. 元学习研究综述[J]. 计算机学报, 2021, 44(2): 422-446.
LI Fanchang, LIU Yang, WU Pengxiang, et al. A survey on recent advances in meta-learning[J]. Chinese Journal of Computers, 2021, 44(2): 422-446. (in Chinese)
- [9] CHENG Y, YU M, GUO X X, et al. Few-shot learning with meta metric learners[DB/OL]. (2019-01-26) [2023-12-10]. <https://arxiv.org/abs/1901.09890>.
- [10] GARICA V, BRUNA J. Few-shot learning with graph neural networks[DB/OL]. (2018-02-20) [2023-12-10]. <https://arxiv.org/pdf/1711.04043>.
- [11] GAO Y Y, CHEN M, YU D J. Semi-supervised graph convolutional network and its application in intelligent fault diagnosis of rotating machinery[J]. Measurement, 2021, 186: 110084.
- [12] 罗浩, 何超, 陈彪, 等. 基于Laplace小波卷积和BiGRU的少量样本故障诊断方法[J]. 振动与冲击, 2022, 41(24): 41-50.
LUO Hao, HE Chao, CHEN Biao, et al. Small sample fault diagnosis based on Laplace wavelet convolution and BiGRU[J]. Journal of Vibration and Shock, 2022, 41(24): 41-50. (in Chinese)
- [13] 张金汇, 谢林柏. 基于改进域对抗网络的轴承故障迁移诊断方法[J]. 电子测量与仪器学报, 2022, 36(11): 107-115.
ZHANG Jinhui, XIE Linbai. Bearing fault migration diagnosis method based on improved domain adversarial network[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(11): 107-115. (in Chinese)
- [14] CHEN J J, HU W H, CAO D, et al. A meta-learning method for electric machine bearing fault diagnosis under varying working conditions with limited data [J]. IEEE Transactions on Industrial Informatics, 2022, 19(3): 2552-2564.
- [15] ZHANG Y Z, LI S B, ZHANG A S, et al. A novel bearing fault diagnosis method based on few-Shot transfer learning across different datasets[J]. Entropy, 2022, 24(9): 1295.
- [16] LOPARO K A. Bearing data center[DS/OL]. Cleveland: Case Western Reserve University (2013) [2023-09-20]. <https://engineering.case.edu/bearingdatacenter>
- [17] JI Z, CHAI X L, YU Y L, et al. Improved prototypical networks for few-shot learning[J]. Pattern Recognition Letters, 2020, 140: 81-87.



第一作者简介:王玉静,女,1983年3月生,博士、教授。主要研究方向为非平稳信号处理,故障诊断、状态评估与预测技术。曾发表《Unsupervised fault diagnosis method for rolling bearings based on federated universal domain adaptation (《Engineering Applications of Artificial Intelligence》2025, Vol.162)等论文。
E-mail:mirrorwyj@163.com

通信作者简介:康守强,男,1980年12月生,博士、教授。主要研究方向为非平稳信号处理,故障诊断、状态评估与预测技术。
E-mail:kangshouqiang@163.com