

混合图卷积网络的轴承故障诊断方法*

杨培义¹, 李功准², 张天彤¹, 张 龙²

(1. 郑州铁路职业技术学院机车车辆学院 郑州, 450046)

(2. 华东交通大学机电与车辆工程学院 南昌, 330013)

摘要 针对现有轴承故障诊断方法难以兼顾振动特征的低阶局部拓扑关系与高阶跨频带关联,且多层特征融合过程中易引入冗余信息的问题,提出基于双向交叉注意力的渐进式混合图-超图卷积网络(progressive hybrid graph-hypergraph convolutional network,简称PHGHCN)方法。首先,利用多尺度K近邻构建渐进式图和超图,设计双分支网络分别提取特征;其次,通过双向交叉注意力机制实现特征对齐交互与自适应融合;最后,在凯斯西储大学(Case Western Reserve University,简称CWRU)数据集及JL-501数据集上进行了验证。结果表明:PHGHCN准确率分别达100%和98.29%;相较于图卷积网络(graph convolutional network,简称GCN)等模型,所提方法通过协同建模低阶拓扑依赖与高阶群体关联,增强了特征互补性,提升了故障模式的可分性,在准确率与稳定性上具有整体优势。

关键词 故障诊断;轴承;图神经网络;双向交叉注意力机制

中图分类号 TH165;TH133

0 引言

滚动轴承的健康状态对维持连续的工业生产和制造具有重要作用^[1]。在外部环境波动的工况下,设备运行中发生故障的可能性随之提高。对于重要的旋转机械而言,若未能及时识别与处理轴承缺陷,不仅会引发其性能衰减,还可能导致突发停机甚至灾难性损坏^[2-3]。因此,轴承故障诊断具有极其重要的工程意义^[4-6]。

深度学习为故障特征识别提供了更高效的建模方式,推动了智能诊断方法的进一步发展^[7]。Qiu等^[8]提出了基于信道筛选、特征图融合与深度学习的动态自适应诊断方法,为滚动轴承故障识别提供了数据参考。Xu等^[9]提出基于深度稳定ResNet34网络与随机傅里叶方法提取故障特征,实现了对不平衡轴承数据的高精度诊断,并在多工况下表现出优于多种基线模型的泛化能力。Liang等^[10]提出一种结合卷积神经网络-长短期记忆网络与增量迁移学习的诊断方法,可在不同加工条件之间高效迁移知识,提升了故障识别精度并减少了训练数据需求。这些研究提升了深度网络对复杂工况的适应能力,但其特征学习仍主要依赖网络自身的卷积或序列结

构,难以直接表达设备部件及信号间的结构关系。因此,以图神经网络(graph neural networks,简称GNN)为代表的结构化建模方法逐渐成为提升故障诊断能力的重要方向,其能够针对部件间的复杂关系高效建模^[11-14]。Zhang等^[15]通过因果加权构图实现了噪声干扰下高精度故障识别。Wang等^[16]利用超图神经网络提取了高阶耦合故障特征。Li等^[17]提出一种结合可见性图与增强图神经网络的端到端故障诊断方法,提升了机械设备时间序列数据的诊断精度和映射效率。这些单一视角的图或超图模型在特定场景下表现优异,但往往难以兼顾“节点-节点”的低阶局部拓扑与“节点-超边”的高阶群体关联的互补性。

为突破单一图结构的表征局限,学者们开始探索图与超图的联合建模^[18-19]。然而,在离散样本的故障诊断中,现有联合框架仍存在以下不足:①现有方法多依赖单一样本内部的时序进行建模,难以刻画样本间复杂故障模式的差异;②在信息融合策略上,多数方法采用“独立提取-末端拼接”的浅层融合方式,使得表征局部瞬态特征的低阶拓扑与反映群体关联的高阶结构缺乏有效交互;③仅依赖末端拼接与线性映射,难以实现异构特征空间的充分融合,

*国家自然科学基金资助项目(52565011)

收稿日期:2026-01-15;修回日期:2026-04-27

从而限制了特征的判别能力。

针对上述问题,本研究提出一种基于双向交叉注意力的渐进式混合图-超图卷积网络。首先,采用渐进式多尺度K近邻(k-nearest neighbors,简称KNN)策略,将快速傅里叶变换后的振动信号构建为混合图与超图结构,使网络拓扑能够同时刻画局部瞬态冲击特征与跨样本同频调制关系,其中低阶图结构用于捕捉样本间的局部相似性,高阶超图结构用于表征多样本之间的群体关联,从而实现复杂故障模式的协同建模;其次,在特征融合阶段引入双向交叉注意力机制,通过跨拓扑结构的查询-键映射实现高阶与低阶特征的交互更新,促进异构特征空间的有效融合;最后,在CWRU数据集及工程实测数据集上进行了实验验证。

1 图卷积及超图卷积简介

GNN是一类面向图数据结构的学习模型,常用 $G=(V,E)$ 表示一张图。其中: $V=\{v_1, v_2, \dots, v_n\}$,为节点集合; E 为边的集合。GCN作为GNN中的典型网络,通过聚合相邻节点的特征来生成新的节点表示。由于这种聚合通常涉及对邻居节点信息的加权平均,使节点能够融合来自其直接连接邻居的局部信息,因此GCN可用于捕捉低阶、成对的局部拓扑特征,通过堆叠多个GCN层,节点逐步捕获更深层的局部结构信息,并提取用于分类的语义特征。其计算过程为

$$X' = \text{GCNConv}(X) = \text{ReLU}(D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}} X W) \quad (1)$$

其中: X' 为更新后的节点特征;Relu为激活函数; D 为节点的度矩阵,且为对角矩阵, $D_{ii} = \sum A_{ij}$; $\tilde{A} = A + I$ 表示节点的自环过程; I 为单位矩阵; W 为用于节点特征变换的可学习权重矩阵。

相较于传统图只能描述节点间的二元关系,超图能够更灵活地捕捉数据中的多元关联^[20],因此本研究引入超图。在超图中,一条超边可以连接多个节点,且一个图常存在多条超边,能够表示节点之间复杂的高阶关系。一个超图被定义为 $\mathcal{G}=(\mathcal{V}, \mathcal{E})$,其中: $\mathcal{V}$ 为节点集; $\mathcal{E}$ 为超边集。每条超边 $\hat{e}_j \in \mathcal{E}$,可以连接任意数量的节点,而非仅限于2个。为刻画复杂的连接模式,超图的拓扑关系常用关联矩阵 $H \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{E}|}$ 表示,若节点 v_i 属于超边 $\hat{e}_j \in \mathcal{E}$,则 $H_{ij} = 1$;否则为0。

基于超图结构,超图卷积遵循“节点-超边-节点”的传播范式。超图的拉普拉斯矩阵 L_H 定义为

$$L_H = I - D_v^{-\frac{1}{2}} H W_e D_e^{-1} H^T D_v^{-\frac{1}{2}} \quad (2)$$

其中: D_v 、 D_e 分别为节点度矩阵和超边度矩阵; W_e 为超边权。

超图卷积(hypergraph convolution,简称HG-Conv)通过将特征从节点传递到超边,然后将特征从超边返回到节点来聚集和传播信息^[21]。超图卷积算子表示为

$$X^{(l+1)} = \text{ReLU}(D_v^{-\frac{1}{2}} H W_e D_e^{-1} H^T D_v^{-\frac{1}{2}} X^{(l)} \Theta^{(l)}) \quad (3)$$

其中: $D_v^{-\frac{1}{2}} H W_e D_e^{-1} H^T D_v^{-\frac{1}{2}}$ 表示 H 的对称归一化; $\Theta^{(l)}$ 为第 l 层的滤波器参数。

2 混合图卷积网络

滚动轴承在复杂工况下运行时,其振动信号往往呈现非平稳性强、噪声干扰严重等特点,仅依赖单一结构建模的方法难以兼顾振动特征间的低阶局部拓扑关系与跨频带、多特征耦合的高阶群体关联,限制了故障特征的充分表达。为此,本研究提出一种用于轴承故障诊断的渐进式混合图-超图卷积网络,通过构建多尺度图与超图结构分别刻画低阶邻域依赖与高阶群体关系,采用双分支网络实现特征并行提取,并引入双向交叉注意力机制对不同结构信息进行自适应融合,从而形成判别性更强的故障特征表示。

2.1 图结构与超图结构的构建

在数据结构化阶段,每个振动信号样本被抽象为一个节点 $v_i \in V$ 。节点 v_i 的特征向量 $x_i \in \mathbb{R}^d$,直接来源于其经过快速傅里叶变换的频域信号,其中 d 为信号的维度。所有节点的特征汇聚为特征矩阵 $X = [x_1, x_2, \dots, x_N]^T$, $X \in \mathbb{R}^{N \times d}$ 。

为捕获节点间精细的多粒度低阶局部拓扑关系,本研究采用多尺度KNN方法同时构建图与超图。节点间的相似性通过其特征向量的欧氏距离进行衡量,欧氏距离为

$$\text{dist}(v_i, v_j) = \|x_i - x_j\|^2 \quad (4)$$

对于每个节点 v_i ,选择与其距离最近的 k 个节点作为邻居。若 v_j 属于 v_i 的邻居节点,则在 v_i 和 v_j 之间建立一条无向边。通过定义不同的K近邻数 $K \in \{k_1, k_2, k_3\}$,可构建多个不同尺度的邻接矩阵 $A^{(k)}$ 。每个 $A^{(k)}$ 描述了在特定尺度下节点间的二元连接关系,其中: $A_{ij}^{(k)} = 1$,表示节点 v_i 和 v_j 存在边

连接;否则为0。

为引入连接强度信息,边的权重基于高斯核函数计算,得到加权邻接矩阵 A_{ij} 为

$$A_{ij} = \exp[-\text{dist}(v_i, v_j)^2 / 2\sigma^2] \quad (5)$$

其中: σ 为高斯核的带宽。

最终得到一组带权重的多尺度邻接矩阵 A^{k_1} 、 A^{k_2} 和 A^{k_3} , 将其作为渐进式 GCN 算子的输入, 提取低阶局部特征。

为保持低阶图结构与高阶超图结构之间的空间一致性, 超图的构建同样基于多尺度 KNN 机制。具体而言, 在每个尺度 k 下, 将节点 v_i 及其特征向量获得的 KNN 邻域 $\mathcal{N}_k(i) = \{v_j | v_j \in \text{KNN}(v_i, k)\}$ 作为一条高阶超边 $\hat{e}_i^{(k)}$, 即

$$\hat{e}_i^{(k)} = \{v_j | v_j \in \mathcal{N}_k(i)\} \quad (6)$$

所有节点对应的超边构成 k 个尺度的超边集

合 $\mathcal{E}^{(k)}$, 即

$$\mathcal{E}^{(k)} = \{\hat{e}_i^{(k)}\}_{i=1}^N \quad (7)$$

与图结构一致, 每条超边内部的连接强度仍采用同样的高斯核权重, 即节点 v_i 在 k 尺度超边 $\hat{e}_i^{(k)}$ 的权重, 表示为

$$w_{ie}^{(k)} = \exp[-\text{dist}(v_i, v_j)^2 / 2\sigma^2] \quad (8)$$

在得到节点与邻居的权重后, 将每个节点对应的邻域作为超边, 并构建超图的关联矩阵, 即

$$H^{(k)}(i, \hat{e}) = \begin{cases} w_{ie}^{(k)} & (v_i \in \hat{e}) \\ 0 & (\text{其他}) \end{cases} \quad (9)$$

基于关联矩阵 H , 利用式(3)定义的超图卷积算子进行高阶特征传播。通过这一过程, 超图卷积能够逐层整合高阶节点关联信息, 为后续特征提取与分类打下基础。图结构与超图结构的构建如图1所示。

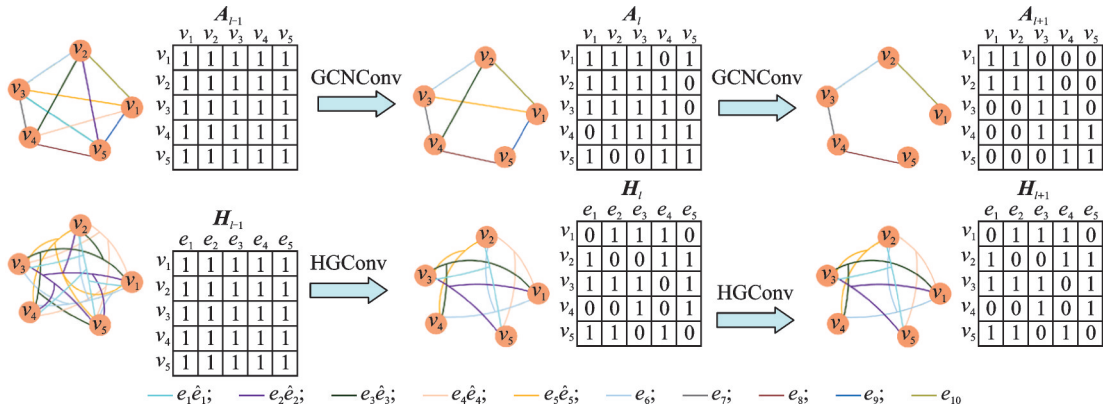


图1 图结构与超图结构的构建

Fig.1 Construction of graph structures and hypergraph structures

本研究提出的渐进式 KNN 构图策略在建模机制上与面向时序寿命预测的方法^[18]存在本质差异。文献[18]以同一信号通道的不同时间步作为图节点, 用于刻画单一样本的时序退化演化过程; 本研究面向离散样本的故障诊断任务, 节点特征由振动信号的频域表征直接构成, 从而更有效地刻画出轴承故障的动力学特性。此外, 两者在特征交互模式上也存在差异: 文献[18]采用“独立提取-末端拼接”的浅层融合方式; 本研究引入双向交叉注意力机制, 建立低阶局部瞬态冲击与高阶同频调制特征之间的双向信息交互通路, 实现了异构语义空间的有效对齐与深层融合。

2.1.1 低阶图拓扑与瞬态冲击的映射

局部缺陷会在特定频带内引发能量突变, 表现为瞬态冲击特征。本研究通过式(4)计算频域节点之间的距离并构建二元边, 本质上是识别主频能量

分布相似的样本对, 使低阶图能够刻画同类故障在瞬态冲击特征上的局部一致性。频域特征驱动的混合图物理意义映射如图2所示。由图可知, 低阶图依赖主频能量对齐, 虽然能够连接节点 v_1 与 v_2 , 但在高能量干扰峰下可能将干扰样本 x_3 误判为相似节点并连接至 v_3 , 同时遗漏因噪声遮蔽使主峰减弱的真实故障节点 v_4 。

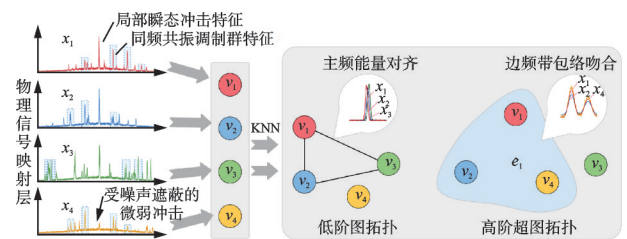


图2 频域特征驱动的混合图物理意义映射

Fig.2 Frequency-domain feature-driven mapping of the physical meaning of hybrid graphs

2.1.2 高阶超图与同频共振的映射

复杂轴承损伤通常伴随显著的幅值调制与谐波共振,其频域表现为以故障特征频率为基频、两侧分布多组边频带的谱系结构。在此情况下,传统二元边易出现连接不稳定或断裂的现象,通过式(6)构建的超边能够将具有相似边频带调制模式的多个样本聚合于同一高阶集合。由图2可知,高阶超图中的超边 e_1 通过刻画边频带包络在不同样本间的一致性,可有效排除缺乏共振特征的干扰节点 v_3 ,并重新关联主峰受噪声遮蔽的真实故障节点 v_4 ,从而实现跨频带、跨样本同频共振模式的表征。

尽管图与超图结构均基于同构KNN以确保物理邻域的空间一致性,但两者在特征传播机制与语义表征上存在本质互补性。图卷积基于邻接矩阵 A 执行成对特征聚合,侧重于捕获局部节点间的直接相似性与信号连续性,从而有效平滑高频噪声;超图卷积基于关联矩阵 H ,遵循“节点-超边-节点”的集合传播范式,通过映射邻域节点至公共超边原型,提取群组内的高阶共性而非简单的两两混合。在轴承诊断中,前者保障了频域信号的局部平滑度,后者则能有效捕捉跨频段的调制关联模式。因此,两者在同一物理邻域下构建了正交的互补特征视图,而非信息的简单冗余。

2.2 混合图特征提取模块

本研究构建PHGHCN模型框架对低阶与高阶关系进行联合编码,以实现特征并行提取与深度融合。模型分两路传播:低阶分支通过3层KNN图卷积捕获局部模式;高阶分支利用超图卷积提取群组结构。特征经双向交叉注意力融合实现互补,最终经全连接层完成故障识别,并以交叉熵损失作为目标函数,即

$$\text{Loss}(y, \bar{y}) = -\lg \left(\frac{e^{y_j}}{\sum_{j=1}^C e^{y_j}} \right) \quad (10)$$

其中: C 为类别; y 和 $\bar{y} \in \{1, 2, \dots, C\}$,分别为真实标签和预测标签。

2.3 双向交叉注意力机制

在基于振动信号的轴承故障诊断中,图结构与超图结构对于信号的结构化表达具有显著差异性。图-超图联合框架通常采用“独立特征提取、末端拼接”的浅层融合方式。由于低阶图拓扑刻画局部瞬态冲击(记为特征 $F \in \mathbb{R}^{N \times d_f}$,其中: N 为节点数量; d_f 为图特征维度),而多尺度高阶超图关联则侧重于表

征跨频段的同频共振关系(记为特征 $T \in \mathbb{R}^{N \times d_t}$,其中: d_t 为超图特征维度),二者处于不同语义空间。由于简单拼接难以实现有效语义交互,易引入信息冗余,因此本研究引入双向交叉注意力机制实现特征对齐与协同建模^[22]。双向交叉注意力融合机制如图3所示。结合图2和图3可知:低阶查询可从高阶特征中获取同频调制信息,以强化微弱故障表征;高阶查询通过低阶特征获得局部结构约束,可有效抑制GNN深层传播中易出现的过度平滑现象。通过该机制,模型能够充分利用异构拓扑的互补特性,构建出判别性更强的特征表达。

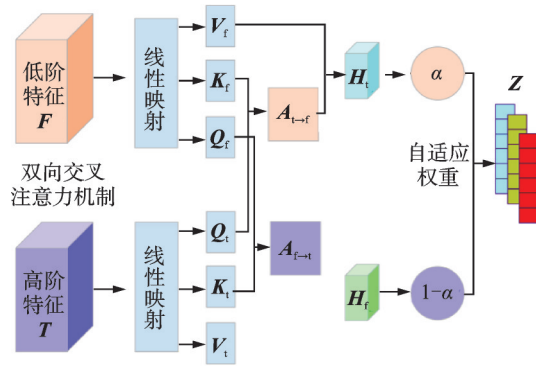


图3 双向交叉注意力融合机制

Fig.3 Bidirectional cross-attention fusion mechanism

为增强模型在不同表示子空间中捕获图与超图复杂交互关系的能力,本研究在双向交叉注意力模块中引入了多头注意力机制。首先,对图特征 F 与超图特征 T 分别构造查询(query, Q)、键(key, K)和值(value, V)表示,用以刻画两类结构在注意力空间中的映射。设 $W_f^Q \in \mathbb{R}^{d_f \times d_q}$ 、 $W_f^K \in \mathbb{R}^{d_f \times d_k}$ 、 $W_f^V \in \mathbb{R}^{d_f \times d_v}$ 为图特征到注意力空间的线性映射矩阵, $W_t^Q \in \mathbb{R}^{d_t \times d_q}$ 、 $W_t^K \in \mathbb{R}^{d_t \times d_k}$ 、 $W_t^V \in \mathbb{R}^{d_t \times d_v}$ 为超图特征的线性映射矩阵,其中: d_q 、 d_k 和 d_v 分别为查询、键和值的投影维度。为避免训练初期的梯度异常并加速异构特征空间的对齐,上述所有可学习的投影权重矩阵均采用Xavier均匀初始化策略进行赋值。低阶特征生成的查询向量 Q_f 定义为 $Q_f = F W_f^Q$,高阶特征生成的键 K_t 与值 V_t 分别定义为 $K_t = T W_t^K$ 和 $V_t = T W_t^V$ 。上述映射用于构造低阶特征对高阶特征的注意力关系,其核心交互原理见图3。在具体的网络实现中,实验将多头注意力的头数设置为4,各注意力头的输出拼接后进一步通过层归一化进行处理,以稳定融合后的特征分布。

相应地,为构建反向关注机制,高阶特征的查询向量定义为 $Q_t = T W_t^Q$,低阶特征的键和值定义为

$K_i = FW_i^K$ 和 $V_i = FW_i^V$ 。在完成查询、键和值的构建后,通过缩放的点积注意力对特征间的交互关系进行建模。

该机制通过双向交互实现特征的空间对齐与语义增强。以低阶关注高阶为例,其注意力权重矩阵 $A_{i \rightarrow t}$ 由低阶特征查询 Q_i 与高阶特征键 K_t 的缩放点积得到,即

$$A_{i \rightarrow t} = \text{softmax}\left(Q_i K_t^T / \sqrt{d_k}\right) \quad (11)$$

其中: d_k 为键的维度,用于缩放点积值以稳定梯度。

将 $A_{i \rightarrow t}$ 作用于高阶特征的值,得到融合后的低阶特征 H_i 更新表示,即

$$H_i = A_{i \rightarrow t} V_t \quad (12)$$

其中: $H_i \in \mathbb{R}^{N \times d_v}$, 为低阶特征在高阶结构指引下更新后的表示,用于补偿图结构难以捕获的跨频段群体结构特征。

同理,高阶特征关注低阶结构的注意力权重矩阵表示为

$$A_{t \rightarrow i} = \text{softmax}\left(Q_t K_i^T / \sqrt{d_k}\right) \quad (13)$$

对应的高阶特征更新表示 H_t 为

$$H_t = A_{t \rightarrow i} V_i \quad (14)$$

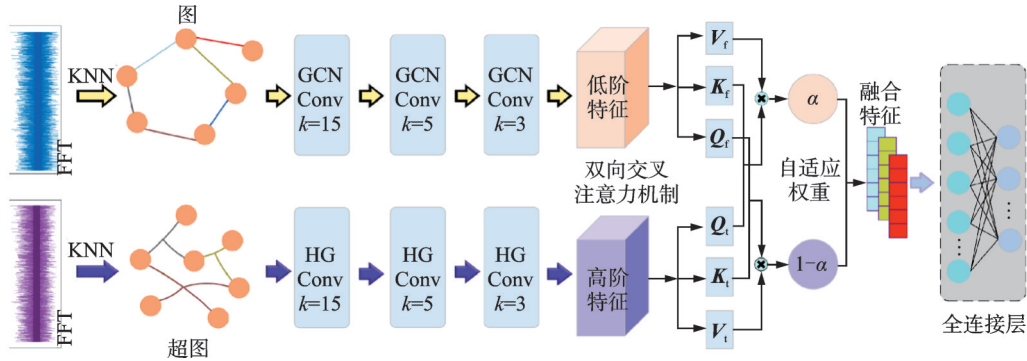


图4 PHGHCN模型框架

Fig.4 PHGHCN model framework

3 实验介绍

实验选用 CWRU 及南昌铁路局 JL-501 数据。为验证所提方法的优越性,引入主流 GNN 模型作为对比,包括超图卷积网络(hypergraph convolution network, 简称 HGCN)、稀疏图卷积网络(sparse graph convolution network, 简称 SGCN)、GCN 以及验证注意力机制的图注意力网络(graph attention network, 简称 GAT)。此外,为进一步对比先进图网络算法,本研究引入了内积超图注意力网络(inner product hypergraph attention network, 简称

其中: $H_i \in \mathbb{R}^{N \times d_v}$, 为在局部结构信息的指导下得到的高阶特征修正结果,使超图结构更贴近局部频率点的实际变化模式。

获得双向更新特征后,为平衡结构贡献,引入可学习权重 $\alpha \in [0, 1]$ 进行自适应融合。 α 通过反向传播进行优化,使模型能够自适应地调整高阶和低阶关注度的权重,从而提高特征融合的有效性。融合特征 Z 定义为

$$Z = (1 - \alpha) H_i + \alpha H_t \quad (15)$$

其中: $Z \in \mathbb{R}^{N \times d_v}$, 为经双向注意力融合的混合特征。

Z 兼具局部敏感性与高阶稳健性,可有效结合局部冲击响应与跨频段一致性,增强了故障的区分度。

2.4 诊断流程

PHGHCN 模型框架如图 4 所示。该模型包含特征提取、融合与分类。首先,预处理信号并构建多尺度图与超图;其次,经双分支编码器并行提取低阶局部与高阶群体特征;然后,利用双向交叉注意力实现跨结构交互对齐,并经自适应加权融合;最后,输入全连接层完成故障分类。

IPHGAT)与图元素级变换器(graph element-wise transformer, 简称 GraphEWT)^[16, 23]。同时,设置无交叉注意力的渐进图-超图网络(progressive graph-hypergraph convolution network, 简称 PGHCN)为消融基准,其融合方式采用拼接操作。所有模型采用 KNN 构图以确保一致性。PHGHCN 网络参数和训练参数分别如表 1、2 所示。本研究构图中近邻个数 k 与中间层 k 一致,数据集样本长度为 1 024,均通过快速傅里叶变换后作为输入构建 KNN 图或超图。训练集与测试集比例为 6:4。

表 1 PHGHCN 网络参数					
Tab.1 PHGHCN network parameters					
层次结构	网络结构	输入 维度	输出维 度	池化 方式	说明
Layer1	GCNConv1	512	512	自适应 步长	$k=15$
	HGConv1	512	512	—	$k=15$
Layer2	GCNConv2	512	512	自适应 步长	$k=5$
	HGConv2	512	512	—	$k=5$
Layer3	GCNConv3	512	512	自适应 步长	$k=3$
	HGConv3	512	512	—	$k=3$
特征融合	Attention Fusion	512×2	512	—	双向交叉 注意力权 重融合
分类器 1	Linear+ Dropout	512	256	—	Dropout $=0.5$
分类器 2	Linear	256	类别数	—	—

表 2 PHGHCN 训练参数			
Tab.2 PHGHCN training parameters			
超参数	值	超参数	值
训练轮次	100	初始学习率	1×10^{-3}
批量大小	64	优化器	SGD
KNN 中 k 值	15/5/3	训练轮次=40 时衰减	0.1 倍
动量系数	0.9	权重衰减系数	5×10^{-4}

3.1 CWRU 数据集

案例 1 采用 CWRU 轴承数据集进行验证。实验选取 1 797 r/min 工况下的驱动端轴承振动信号,采样频率为 12 kHz。通过电火花加工技术在轴承的内圈、外圈及滚动体上分别植入单点损伤,以模拟实际运行中的故障。故障损伤直径包括 0.177 8、0.355 6 及 0.533 4 mm 等 3 种尺寸,结合正常运行状态,共构建 10 类健康状态(标签记为 $C_0 \sim C_9$),每类状态提取 117 个样本。CWRU 数据集样本划分及各模型分类准确率分别如表 3、4 所示。

由表 4 可知,不同网络模型在结构建模与特征表达能力上存在显著差异。GCN 与 SGCN 的分类准确率分别为 96.00% 和 96.80%,表明基于局部邻域的特征聚合能够有效刻画低阶局部拓扑依赖,其中稀疏邻域有助于抑制冗余信息传播;HGCN 的分类准确率为 95.24%,表明在以局部聚集特征为主的数据中,单纯进行高阶跨频带关联建模易引发特征的过度平滑与扩散,导致性能受限;GAT 的分类准确率仅为 92.80%,其注意力机制未能有效突出关键节点的贡献。

表 3 CWRU 数据集样本划分			
Tab.3 CWRU dataset sample segmentation			
健康状态	故障尺寸/mm	样本数量	标签
正常	—	117	C_0
内圈	0.177 8	117	C_1
	0.355 6	117	C_2
	0.533 4	117	C_3
外圈	0.177 8	117	C_4
	0.355 6	117	C_5
	0.533 4	117	C_6
滚动体	0.177 8	117	C_7
	0.355 6	117	C_8
	0.533 4	117	C_9

表 4 CWRU 数据集各模型分类准确率			
Tab.4 Experimental results of various models on the CWRU dataset			
模型	分类准确率/%	模型	分类准确率/%
GCN	96.00	PGHCN	97.62
SGCN	96.80	IPHGAT	97.59
HGCN	95.24	GraphEWT	100.00
GAT	92.80	PHGHCN	100.00

相比之下,引入先进机制的图网络在特征挖掘能力上表现出更优性能。IPHGAT 取得 97.59% 的分类准确率,说明通过注意力机制自适应建模跨频带高阶关联特征,在一定程度上缓解了特征扩散的问题,但其主要依赖单一高阶拓扑结构,缺乏低阶局部信息协同,性能仍有提升空间;GraphEWT 通过扩张残差图拓扑增强节点间的局部关联建模能力,并结合元素级注意力机制捕获全局交互信息,从而实现局部与全局特征的协同表征,实现了 100% 的分类准确率。

本研究提出的图-超图混合结构模型同样表现出良好的诊断性能。基准模型 PGHCN 的准确率为 97.62%,优于多数传统图模型,验证了图-超图联合建模在刻画低阶局部关系与高阶群体结构方面的有效性。在此基础上,引入双向交叉注意力机制后,PHGHCN 达到 100% 的分类准确率。

CWRU 数据集上 PHGHCN 的损失曲线和准确率曲线分别如图 5、6 所示。由图可知,模型在训练初期展现出高效的收敛速率,随着网络权重的不断优化,训练集与测试集的损失值同步大幅下降,准确率迅速上升;进入训练后期,测试集曲线整体趋于平稳并与训练集保持高度贴合,未出现发散等过拟合迹象,表明模型具备优异的泛化能力。PHGHCN 在测试集上实现了 100% 的故障识别精度,验证了所提方法在特征挖掘与状态识别上的高效性。

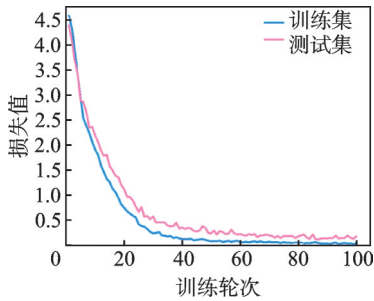


图5 CWRU数据集上PHGHCN的损失曲线
Fig.5 Loss curves of PHGHCN on CWRU dataset

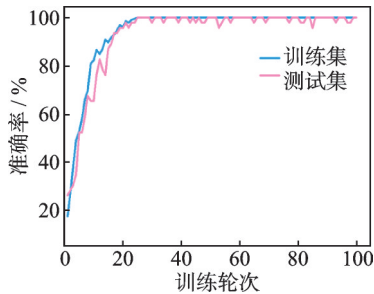


图6 CWRU数据集上PHGHCN的准确率曲线
Fig.6 Accuracy curves of PHGHCN on CWRU dataset

3.2 南昌铁路局工程数据集

案例2的数据源自JL-501型机车轴承故障实验台,如图7所示。测试对象为NJ2232WB型圆柱滚子轴承,该批实验轴承均来源于DF4型机车实际服役轮对。为验证所提模型对故障的定量识别性能,本研究构建了融合人工受控与真实故障的混合数据集。针对实际损伤形貌的随机性,引入“等效损伤当量”替代绝对物理尺寸,将故障分为轻、中、重3个等级:①内圈采用人工打磨长约4、7和19 mm的缺陷作为定量基准,分别对应轻、中、重度等级;②外圈与滚动体选取实际服役的自然疲劳磨损样本,通过综合评估其损伤面积与振动响应强度,将其等效映射至上述基准当量(即等效尺寸约为7、19 mm)以统一分类尺度,该混合数据集兼顾了定量评估基准与真实工业非规则损伤特征。

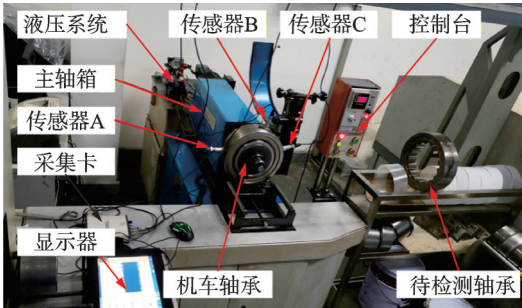


图7 JL-501型机车轴承故障实验台
Fig.7 JL-501 locomotive bearing failure test bench

不同健康状态下的轴承图片如图8所示。实验采用三通道加速度传感器以20 kHz采样率获取振动信号。本研究选取传感器B记录的7种典型工况(标签为 $N_0 \sim N_6$),每类包含390个独立样本。

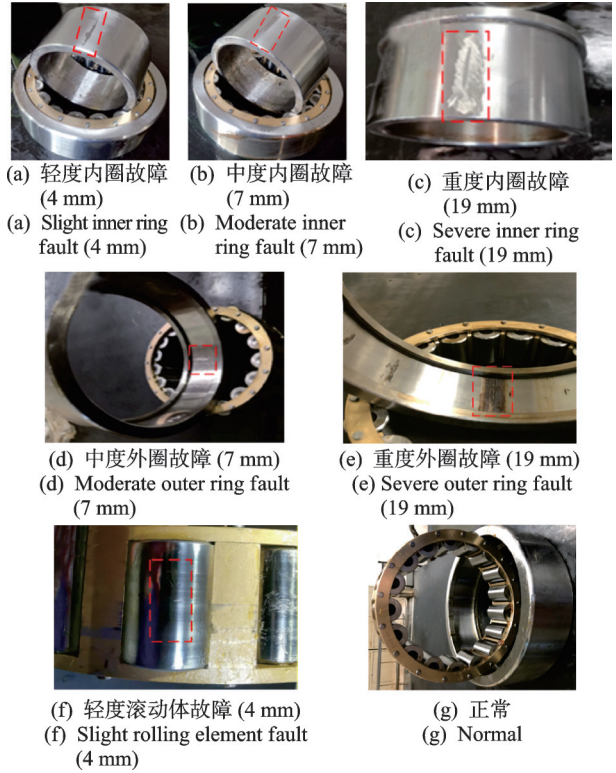


图8 不同健康状态下的轴承图片
Fig.8 Photographs of bearings under different health conditions

工程数据集样本划分和诊断结果分别如表5、6所示。由表可知,所提方法在10次独立实验中取得98.29%的平均准确率,显著优于对比模型。测试时间为0.380 s,体现了性能与效率的良好平衡。该方法通过双向交叉注意力协同建模低阶图结构与高阶超图关系,自适应强化局部频域关联与跨频带群体特征,从而可在非平稳、强噪声振动信号中构建更具判别性的特征表示。相比之下,GCN与GAT受限于低阶邻域聚合,对远程关联刻画不足;SGCN在邻域稀疏化过程中削弱了关键信息;HGCN虽具备高阶建模能力,但易放大噪声;PGHCN融合结构后性能有所提升,但仍缺乏有效的结构调控机制;对比模型IPHGAT与GraphEWT在测试时间上表现出高效率(分别为0.110与0.180 s),但在处理非平稳、强噪声的工程信号时,其诊断精度相较于PHGHCN仍有明显差距。这表明单纯引入超图注意力或图变换器结构虽能加速局部特征聚合,但在复杂跨频带关联挖掘上仍显不足。

表 5 工程数据集样本划分

Tab.5 Sample segmentation of engineering data sets

健康状况	故障程度	样本数	标签
正常	—	390	N_0
内圈故障	轻度	390	N_1
	中度	390	N_2
	重度	390	N_3
外圈故障	中度	390	N_4
	重度	390	N_5
滚动体故障	轻度	390	N_6

表 6 工程数据集诊断结果

Tab.6 Diagnostic results on the engineering dataset

模型	准确率/%	测试 时间/s	参数量/ ($\times 10^6$)	计算量/ ($\times 10^9$)
GCN	91.62 ± 4.08	0.482	0.332 0	1.448 2
SGCN	86.95 ± 4.38	0.500	0.199 4	0.727 4
HGCN	68.8 ± 8.01	0.220	0.331 3	1.448 2
GAT	91.79 ± 1.80	0.481	0.333 6	1.456 6
IPHGAT	94.44 ± 1.99	0.110	0.857 9	1.084 8
GraphEWT	93.49 ± 2.11	0.180	0.728 1	3.606 4
PGHCN	95.61 ± 0.98	0.340	0.791 6	2.896 7
PHGHCN	98.29 ± 2.37	0.380	0.792 6	2.907 9

此外,常规单分支网络(如GCN等)虽参数规模较小,但特征表达能力受限。相比消融模型PGHCN,本研究方法仅增加约0.001 M参数,在保持计算时间可控的同时,实现了诊断精度的显著提升,既避免了IPHGAT参数冗余与GraphEWT计算负担,又实现了模型复杂度与性能间的有效平衡。

相较于CWRU数据集的100%准确率,PHGHCN在JL-501工程集上准确率略微降至

98.29%,这是因为实际工况中强轮轨冲击与背景噪声对早期微弱故障的掩盖。但对比HGCN降至68.8%的准确率,本模型精度损耗极小,充分印证了双向交叉注意力机制在强噪声环境下的优异鲁棒性。

工程数据集各方法下的混淆矩阵如图9所示。由图可知:PHGHCN的预测结果高度集中于对角线,全局判别力最优;单一高阶聚合的HGCN易引发特征扩散;稀疏建模的SGCN存在结构信息不足的问题;GCN与GAT在区分相似故障模式时能力有限。当故障演化程度接近时,部分模型出现明显误判:IPHGAT将内圈中度故障(N_2)高比例错分为内圈轻度故障(N_1),GraphEWT在外圈中度故障(N_4)与重度故障(N_5)之间存在混淆;PGHCN简单拼接双图特征后类间边界仍不清晰。针对这些问题,PHGHCN通过双向交叉注意力深度协同低阶拓扑与高阶关系,强化了不同损伤等级间的频域特征差异,实现了更稳定的故障程度区分。

工程数据集各方法下的T-SNE可视化如图10所示。由图可知,不同模型在特征分布上存在显著差异:HGCN受高阶超边聚合影响易产生特征扩散;GCN与SGCN主要依赖低阶邻域建模,在相近工况下出现特征重叠;GAT在局部区分性上有所提升,但仍易受噪声干扰;IPHGAT虽在多数类别上形成一定聚集,但由于缺乏低阶局部拓扑约束, N_1 与 N_2 特征簇仍出现明显混叠;GraphEWT受限于成对消息传递机制且缺乏高阶关联,导致 N_4 与 N_5 特征分布高度重叠;PGHCN融合低阶与高阶结构后聚类效果改善,但相邻类别仍存在贴合;PHGHCN通

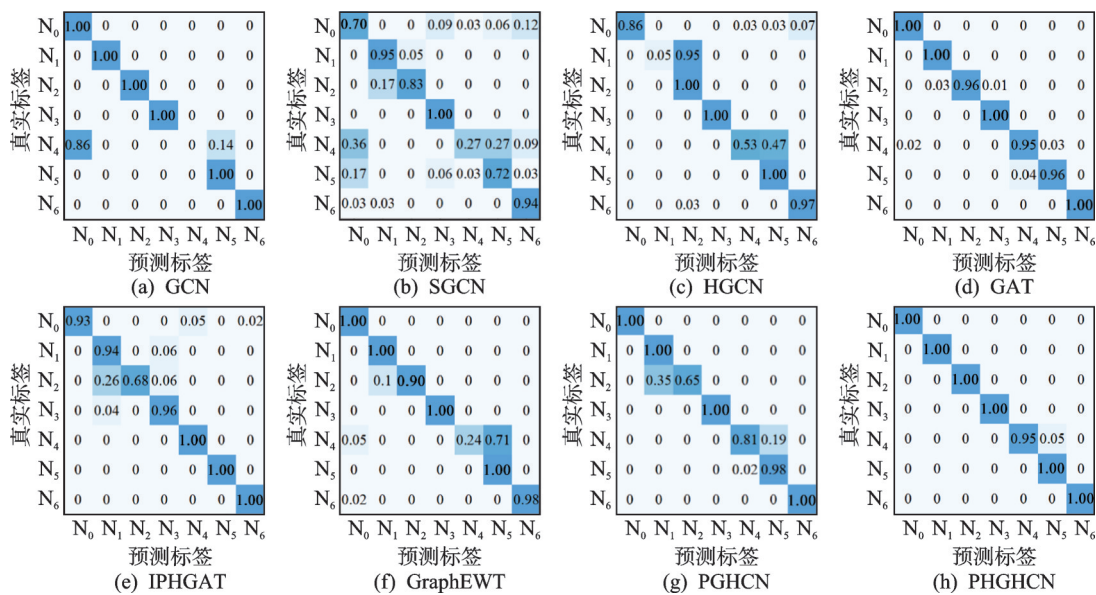


图9 工程数据集各方法下的混淆矩阵

Fig.9 Confusion matrix under each method for engineering datasets

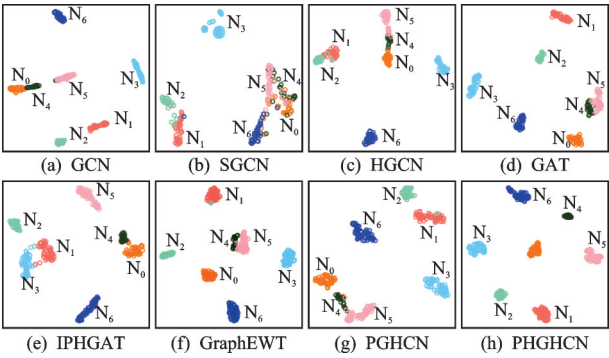


图 10 工程数据集各方法下的 T-SNE 可视化

Fig.10 T-SNE visualization under each method for engineering datasets

过双向交叉注意力实现低阶拓扑与高阶关联的协同建模,构建了更具可分性的特征空间,有效提升了故障识别的稳定性。

PHGHCN 与对比模型显著性检验结果见表 7。为排除随机误差,本研究基于多次实验准确率,将 PHGHCN 与 IPHGAT、PGHCN 进行独立样本 t 检验与 Wilcoxon 秩和检验。结果显示, p 值均严格小于 0.01。这在统计学上证实了 PHGHCN 的极显著性能优势,并充分印证了双向交叉注意力机制在特征对齐与诊断精度提升上的有效性。

表 7 PHGHCN 与对比模型显著性检验结果

Tab.7 Significance test results of PHGHCN and comparison models

对比模型	t 检验 p 值	Wilcoxon 检验 p 值	显著性水平
PHGHCN 和 IPHGAT	$< 0.000\ 1$	0.000 52	极显著
PHGHCN 和 PGHCN	$< 0.000\ 1$	0.000 36	极显著

3.3 k 值敏感性分析

KNN 中的 k 值决定了模型在图数据中捕获局部至全局信息的能力。本研究对中间层 k 值开展敏感性实验,不同 k 值下所提方法的准确率如图 11 所示。由图可知,准确率呈现非单调波动: $k=5$ 时达

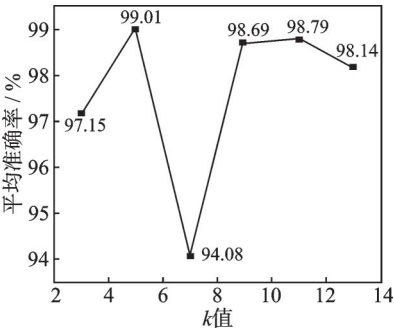


图 11 不同 k 值下所提方法的准确率

Fig.11 Accuracy of the proposed method at different k values

到最优(99.01%); $k=7$ 时显著骤降,随后回升平稳。这表明中间层 k 值的选择对特征聚合至关重要,不当的邻域范围易引入噪声或导致信息丢失。

4 结 论

1) 针对单一结构特征表达不全的问题,提出 PHGHCN 方法。该方法利用多尺度 KNN 图与超图分别捕获低阶与高阶特征,并通过双向交叉注意力机制实现两者的深度交互与自适应融合,增强了特征的判别性与稳健性。

2) 在 CWRU 及工程数据集实验中,PHGHCN 分别实现了 100% 和 98.29% 的故障分类准确率,较 HGCN 及消融模型 PGHCN 分别提升了 29.49% 和 2.68%,充分验证了交叉注意力机制在提升诊断精度与鲁棒性方面的有效性。

3) 本研究通过协同建模局部拓扑与高阶群体结构,并引入双向交叉注意力融合机制,增强了故障模式判别能力。实验结果表明,所提方法在强噪声工程数据中不仅能准确定位故障部件,还能精确判别不同故障程度,具有巨大的工程应用潜力。

参 考 文 献

[1] 张龙,胡燕青,赵丽娟,等.采用递归图编码技术与残差网络的滚动轴承故障诊断[J].西安交通大学学报,2023,57(2):110-120.
ZHANG Long, HU Yanqing, ZHAO Lijuan, et al. Fault diagnosis of rolling bearings using recurrence plot coding technique and residual network[J]. Journal of Xi'an Jiaotong University, 2023, 57(2): 110-120. (in Chinese)

[2] ZHANG L, WANG J B, XIAO Q, et al. A bearing fault diagnosis method with implementation of multi-scale time-frequency feature fusion and bidirectional information interaction[J]. IEEE Sensors Journal, 2025, 25(13): 23740-23756.

[3] 官源林,刘贵林,于春雨,等.基于 CNN 与 BLS 的滚动轴承故障诊断方法[J].振动、测试与诊断,2024,44(5):1017-1022.
GUAN Yuanlin, LIU Guilin, YU Chunyu, et al. Fault diagnosis method of rolling bearing based on CNN and BLS[J]. Journal of Vibration, Measurement & Diagnosis, 2024, 44(5): 1017-1022. (in Chinese)

[4] 张顺,邓艾东,徐硕,等.基于 DFT 与 ECA 的滚动轴承故障诊断[J].振动、测试与诊断,2024,44(4):754-760.
ZHANG Shun, DENG Aidong, XU Shuo, et al. Fault

- diagnosis of rolling bearing based on discrete Fourier transform and efficient channel attention[J]. *Journal of Vibration, Measurement & Diagnosis*, 2024, 44(4): 754-760.(in Chinese)
- [5] CAO Y, CHEN Y Y, ZHANG D, et al. A dynamic spatial-temporal subgraph convolutional network for noncontact fault diagnosis[J]. *IEEE Transactions on Instrumentation and Measurement*, 2025, 74: 3504711.
- [6] 张龙, 赵丽娟, 王朝兵, 等. 自适应频域字典的机车轮对轴承稀疏诊断方法[J]. *铁道科学与工程学报*, 2023, 20(4): 1456-1468.
ZHANG Long, ZHAO Lijuan, WANG Chaobing, et al. Adaptive frequency domain dictionary for sparse diagnosis of locomotive wheelset bearing[J]. *Journal of Railway Science and Engineering*, 2023, 20(4): 1456-1468.(in Chinese)
- [7] 张龙, 肖逸文, 周神赐, 等. 一种双阻尼小波赋能的可解释卷积神经网络在轴承故障诊断中的应用[J]. *西安交通大学学报*, 2025, 59(3): 210-221.
ZHANG Long, XIAO Yiwen, ZHOU Shenci, et al. Application of interpretable convolutional neural network enabled by bi-damped wavelets in bearing fault diagnosis[J]. *Journal of Xi'an Jiaotong University*, 2025, 59(3): 210-221.(in Chinese)
- [8] QIU B B, LIU S Q, LI W D, et al. Dynamic adaptive fault diagnosis using multi-channel image fusion and deep learning in channel failure occasions on rolling bearings[J]. *Advanced Engineering Informatics*, 2025, 68: 103631.
- [9] XU Z B, LEE C K M, WONG C N. A novel fault diagnosis method based on deep stable learning for bearings with imbalanced data samples[J]. *Expert Systems with Applications*, 2025, 281: 127634.
- [10] LIANG Y C, WANG Y Q, LI W D, et al. Adaptive fault diagnosis of machining processes enabled by hybrid deep learning and incremental transfer learning[J]. *Computers in Industry*, 2025, 167: 104262.
- [11] LI T F, ZHAO Z B, SUN C, et al. Domain adversarial graph convolutional network for fault diagnosis under variable working conditions[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 3515010.
- [12] 谢锋云, 王玲岚, 宋明桦, 等. 基于GCN的多源变工况滚动轴承故障诊断[J]. *铁道科学与工程学报*, 2024, 21(5): 2109-2118.
XIE Fengyun, WANG Linglan, SONG Minghua, et al. Multi-source rolling bearing fault diagnosis under variable working conditions based on GCN[J]. *Journal of Railway Science and Engineering*, 2024, 21(5): 2109-2118. (in Chinese)
- [13] LI T F, ZHOU Z, LI S N, et al. The emerging graph neural networks for intelligent fault diagnostics and prognostics: a guideline and a benchmark study[J]. *Mechanical Systems and Signal Processing*, 2022, 168: 108653.
- [14] 袁科研, 兰媛, 黄家海, 等. 基于GCN的轴向柱塞泵故障诊断方法[J]. *振动、测试与诊断*, 2025, 45(1): 140-145.
YUAN Keyan, LAN Yuan, HUANG Jiahai, et al. Fault diagnosis of axial piston pump based on graph convolutional neural network [J]. *Journal of Vibration, Measurement & Diagnosis*, 2025, 45(1): 140-145.(in Chinese)
- [15] ZHANG Z W, WU L F. Graph neural network-based bearing fault diagnosis using Granger causality test[J]. *Expert Systems with Applications*, 2024, 242: 122827.
- [16] WANG S L, JIAO X X, JING B, et al. Coupling fault diagnosis of bearings based on hypergraph neural network[J]. *Sensors*, 2024, 24(19): 6391.
- [17] LI X M, WANG Y X, ZHAO S C, et al. Adaptive convergent visibility graph network: an interpretable method for intelligent rolling bearing diagnosis[J]. *Mechanical Systems and Signal Processing*, 2025, 222: 111761.
- [18] ZHANG Y R, SU C, HE X L, et al. Progressive hybrid hypergraph attention network with channel information fusion for remaining useful life prediction of rolling bearings[J]. *Mechanical Systems and Signal Processing*, 2025, 236: 112987.
- [19] WANG Y, ZHU D. A hypergraph-based hybrid graph convolutional network for intracity human activity intensity prediction and geographic relationship interpretation [J]. *Information Fusion*, 2024, 104: 102149.
- [20] OUVARD X. Hypergraphs: an introduction and review [EB/OL]. (2020-02-12). <https://arxiv.org/abs/2002.05014>.
- [21] FENG Y F, YOU H X, ZHANG Z Z, et al. Hypergraph neural networks [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 3558-3565.
- [22] LIU Y P, ZHENG W N, DU Y, et al. Bearing fault diagnosis based on cross image multi-attention mechanism [J]. *Scientific Reports*, 2025, 15: 21100.
- [23] SU S Z, CHEN Y Y, ZHU Y M, et al. Fault diagnosis method based on multi-modal hypersensing graph element-wise Transformer[J]. *Measurement Science and Technology*, 2025, 36(8): 086110.



第一作者简介:杨培义,男,1991年11月生,硕士、讲师。主要研究方向为城轨车辆技术、故障诊断、图像采集和职业教育。

E-mail:771576680@qq.com