

元类增量迁移学习驱动的跨域终身智能诊断方法

林翠颖¹, 陈科^{1,4}, 吕宇藩¹, 孔运^{1,2,3}, 董明明¹, 刘辉¹, 褚福磊⁵

(1. 北京理工大学机械与车辆学院, 北京 100081; 2. 重庆大学高端装备机械传动全国重点实验室, 重庆 400044;
3. 北京理工大学唐山研究院, 河北唐山 063015; 4. 内蒙古第一机械集团股份有限公司, 内蒙古包头 014032;
5. 清华大学机械工程系, 北京 100084)

摘要: 机械装备在长期服役过程中将持续新增故障模式, 这对故障诊断模型的持续学习与智能诊断能力提出了更高要求。类增量学习驱动的终身智能诊断技术为高端装备全寿命安全服役保障提供了一种途径, 但现有类增量学习方法难以解决跨工况条件下高效增量迁移诊断的难题。为此, 本文提出元类增量迁移学习驱动的跨域终身智能诊断方法。通过集成深度残差网络与卷积块自注意力特征融合模块, 设计了增强型特征提取器, 实现通道和空间维度的深度特征提取与融合; 结合特征级与决策级知识蒸馏机制, 构建了多级知识蒸馏策略, 解决增量迁移诊断场景下的灾难性遗忘难题; 将元学习思想融入类增量学习框架, 提出了元类增量参数学习机制, 提高模型的增量迁移诊断泛化性能。开展了列车传动系统故障试验验证, 结果表明不同增量迁移诊断场景下所提方法的平均诊断精度为 94.96%, 平均遗忘率为 3.85%, 优于前沿类增量学习方法, 为实现高端装备全寿命周期健康管理的终身智能故障诊断提供了解决方案。

关键词: 终身智能诊断; 类增量学习; 元学习; 知识蒸馏; 迁移学习

中图分类号: TH165⁺.3 **文献标志码:** A **文章编号:** 1004-4523(2025)06-1270-10

DOI: 10.16385/j.cnki.issn.1004-4523.2025.06.015

Meta-class-incremental transfer learning method for cross-domain lifelong intelligent diagnosis

LIN Cuiying¹, CHEN Ke^{1,4}, LYU Yufan¹, KONG Yun^{1,2,3}, DONG Mingming¹, LIU Hui¹, CHU Fulei⁵

(1. School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China; 2. State Key Laboratory of Mechanical Transmission for Advanced Equipment, Chongqing University, Chongqing 400044, China; 3. Tangshan Research Institute, Beijing Institute of Technology, Tangshan 063015, China; 4. Inner Mongolia First Machinery Group Co., Ltd., Baotou 014032, China; 5. Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China)

Abstract: New fault modes will continuously emerge in the long-term operation and service process of machinery equipment, which poses higher requirement of the continual learning and lifelong diagnosis capability for intelligent diagnostic models. Lifelong intelligent diagnosis technology driven by class-incremental learning provides new approaches to ensure the full lifecycle safe operation of high-end equipment. However, existing class-incremental learning methods cannot address the problem of efficient incremental transfer diagnosis under the circumstance of cross-operating conditions. To this end, this paper proposes a cross-domain lifelong intelligent diagnostic method driven by meta-class-incremental transfer learning. An enhanced feature extractor is developed via integrating deep residual networks with a convolutional block attention feature fusion module to achieve deep feature extraction and fusion across channel and spatial dimensions. A multi-level knowledge distillation strategy is constructed through combining feature-level and decision-level knowledge distillation mechanisms to effectively address catastrophic forgetting issues in incremental transfer diagnostic scenarios. A meta-class-incremental parameter learning mechanism is proposed by innovatively incorporating the idea of meta-learning into class-incremental learning framework, thus improving the model generalization ability for incremental transfer diagnosis. Experiment validations were conducted on subway train transmission system test rig. Results show that the proposed method achieves an average diagnostic accuracy of 94.96% and an average forgetting rate of 3.85% across different incremental transfer diagnostic scenarios, and outperforms state-of-the-art class-incremental learning methods, offering insights for

收稿日期: 2025-04-28; **修订日期:** 2025-06-02

基金项目: 北京市自然科学基金资助项目(3252008); 国家自然科学基金资助项目(52105108); 中国科协青年人才托举工程资助项目(2023QNRC001); 北京市科协青年人才托举工程资助项目(BYESS2024294); 高端装备机械传动全国重点实验室开放基金资助项目(SKLMT-MSKFKT-202304); 河北省自然科学基金面上项目(E2023105039)

achieving lifelong intelligent fault diagnosis in full lifecycle health management of high-end equipment.

Keywords: lifelong intelligent diagnosis; class-incremental learning; meta-learning; knowledge distillation; transfer learning

随着大数据与深度学习赋能的智能运维技术快速发展,智能故障诊断技术已成为保障机械装备可靠安全服役的重要支撑,同时有助于降低装备运行的维护成本^[1-2]。深度学习驱动的智能诊断技术无需依赖故障先验知识,通过深层特征学习即可识别故障状态,故而受到学术与工业界的广泛关注^[3]。

然而,现有深度学习诊断模型多依赖静态、固定的数据分布假设,即训练与测试阶段的故障模式一致,这与高端装备长期服役的真实场景有差异^[4]。在复杂服役工况下,高端机械设备全寿命周期长期运行过程中可能持续出现新的故障模式,致使传统智能诊断模型在离线训练阶段难以覆盖所有潜在的故障类别^[5]。随着新故障模式的持续出现,若直接在现有模型基础上进行训练,往往会导致灾难性遗忘现象,即模型在学习新故障类别后,先前已学故障类别的诊断性能会明显退化,从而影响智能诊断模型的持续诊断效果^[6]。虽然通过持续保存所有完整历史数据并重新训练模型在一定程度上可缓解灾难性遗忘问题,但受限于存储能力、计算资源及数据安全等因素,这一方式在工业实际应用中不仅代价高昂且难以实现。因此,研究具备抗遗忘性能与持续学习能力的终身智能诊断模型,已成为当前智能故障诊断领域的研究重点与挑战之一。

近年来,类增量学习为智能诊断模型提供了一种可扩展性强的解决思路。该方法在适应新故障模式的同时,还具备保留已有故障类别知识的能力,从而有效应对深度学习模型在增量诊断场景中普遍存在的灾难性遗忘问题^[7-8]。现有类增量学习驱动的智能故障诊断研究可粗略分为基于知识蒸馏的方法、基于样本回放的方法以及基于扩展的方法。基于知识蒸馏的方法通过引导当前模型保留旧模型对先前健康状态类别的输出响应或特征分布,有效缓解了灾难性遗忘问题。ZHANG等^[9]提出了一种监督对比知识蒸馏方法,实现了小样本下旋转机械的有效类增量故障诊断。LI等^[10]设计了一种类知识蒸馏方法,有效增强了跨机械的类增量迁移诊断性能。DING等^[11]通过引入知识蒸馏损失增强图神经网络,实现了列车转向架的系统级类增量故障诊断。基于样本回放的方法通过保留少量旧类别的代表性样本或生成旧类别样本的合成数据,以缓解灾难性遗忘问题。YAN等^[12]提出了一种面向风力发电机故障诊断的小样本类增量学习方法,融合正向虚拟样本生成与反向记忆库机制,以缓解灾难性遗忘问题。HU等^[13]提出了集成样本回放策略的自适应增

量模型,有效增强了类增量故障诊断性能。GU等^[14]开发一种用于化工故障诊断的增量不平衡网络,通过扩展不平衡数据集提升类增量学习性能。基于扩展的方法通过增加新的网络结构或参数来提升模型性能,不仅增强了模型对新任务的适应能力,也有效避免了灾难性遗忘。SHI等^[15]将广义类增量网络与域自适应策略融合于类增量学习框架,实现了旋转机械在跨工况场景下的故障诊断。XU等^[16]将多源信息融合与广义学习理论引入类增量学习框架,有效提升了类增量故障诊断的性能。ZHANG等^[17]提出了融合原型网络与类增量学习的故障诊断框架,实现了旋转机械半监督类增量故障诊断。HUANG等^[18]提出了结合数字孪生与增量学习策略的多保真信息融合方法,有效提升了旋转机械的类增量故障诊断性能。

尽管类增量学习方法在故障诊断领域已取得初步应用,但现有研究仍多建立在训练集与测试集服从同一分布的假设基础上,针对类增量迁移诊断场景的研究相对匮乏。因此,实际工业场景下频繁变化的工况条件以及新故障模式的不断涌现,极大限制了传统类增量学习方法实现增量迁移诊断的适用性,甚至可能因存在严重的域偏移进一步加重灾难性遗忘,最终导致方法的完全失效。与此同时,传统类增量学习模型在变工况条件下的泛化能力与适应性较弱,难以有效执行持续学习任务,进而制约了其在工业场景中的实际推广和应用。因此,开展具备更强泛化能力的跨域终身智能诊断方法研究,提升类增量学习模型在复杂变工况下的增量迁移诊断性能,具有重要的研究意义与应用价值。

为此,本文建立了一种元类增量迁移学习(meta-class-incremental transfer learning, MCITL)模型,提出元类增量迁移学习驱动的跨域终身智能诊断方法,以应对高端机械装备在变工况场景下的增量迁移故障诊断挑战。所提方法设计了增强型特征提取器和多级融合知识蒸馏策略,并提出了新型元类增量参数学习机制,旨在解决现有类增量学习模型在增量迁移诊断场景中诊断性能受限的难题。通过列车牵引传动系统故障数据集试验验证,表明所提方法可实现变工况条件下的类增量迁移诊断,并且表现出优越的抗遗忘性能与泛化诊断性能。

1 元类增量迁移诊断问题描述

本文将元迁移学习策略融入增量学习框架,以

增强先验知识保留能力并缓解灾难性遗忘,实现增量迁移诊断场景下的持续学习与跨域终身智能诊断。所解决的元类增量迁移诊断问题如图 1 所示。

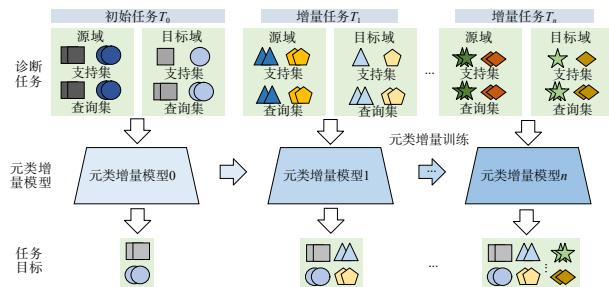


图 1 元类增量迁移诊断问题描述

Fig. 1 Problem description of the meta-class-incremental transfer diagnosis

假定元类增量迁移诊断场景中,考虑 n 个类别互不重叠的增量任务。第 k 个增量任务的源域训练数据 $D_{S,k} = \{S_{S,k}, Q_{S,k}\}_{k=0}^n$ 由源域支持集 $S_{S,k}$ 和源域查询集 $Q_{S,k}$ 构成,二者具体可表示为:

$$S_{S,k} = \{x_{S,k,j}, y_{S,k,j}\}_{j=1}^{N_S}, Q_{S,k} = \{x_{S,k,j}, y_{S,k,j}\}_{j=1}^{N_Q} \quad (1)$$

式中, $x_{S,k,j}$ 和 $y_{S,k,j}$ 分别表示增量任务 T_k 中第 j 个源域训练样本及其对应标签; N_S 和 N_Q 分别表示源域支持集和查询集样本数量。

目标域数据可以记为 $D_{T,k} = \{S_{T,k}, Q_{T,k}\}_{k=0}^n$ 。其对应的目标域支持集与目标域查询集的定义如下:

$$S_{T,k} = \{x_{T,k,j}, y_{T,k,j}\}_{j=1}^{N_{T,S}}, Q_{T,k} = \{x_{T,k,j}\}_{j=1}^{N_{T,Q}} \quad (2)$$

式中, $x_{T,k,j}$ 和 $y_{T,k,j}$ 分别表示增量任务 T_k 中第 j 个目标

域训练样本及其对应标签; $N_{T,S}$ 和 $N_{T,Q}$ 分别表示目标域支持集和查询集中的样本数量。需指出 $D_{S,k} \cap D_{T,k} = \emptyset$, $N_{T,S}$ 远小于 $N_{T,Q}$ 。在元类增量迁移诊断中,任务 T_k 的训练过程包括两个阶段:元训练和快速自适应。在元训练阶段,利用源域支持集 $S_{S,k}$ 训练第 k 个元类增量模型,进而利用训练好的模型在源域查询集 $Q_{S,k}$ 上测试评估其泛化性能;在快速自适应阶段,在目标域支持集 $S_{T,k}$ 上对预训练的元类增量模型进行微调。最终,在目标域查询集 $Q_{T,k}$ 上测试快速自适应微调后的元类增量模型,获取任务 T_k 对应的元类增量迁移诊断结果。

2 元类增量迁移学习模型

所提元类增量迁移学习方法包含增强型特征提取器、多级知识蒸馏以及元类增量参数学习,其架构如图 2 所示。首先,本文提出了一种增强型特征提取器,可实现跨通道和空间维度的深层特征提取与融合,有效突出关键特征表示,抑制无关信息。其次,设计了一种多级融合的知识蒸馏策略,综合考虑特征级和决策级知识蒸馏,旨在缓解增量迁移诊断场景下的灾难性遗忘问题。最后,提出一种元类增量参数学习策略,通过将改进的元学习方法集成到类增量学习框架中,并结合多步梯度更新的元训练与快速适应机制,以提升模型在类增量迁移诊断任务中的泛化能力。

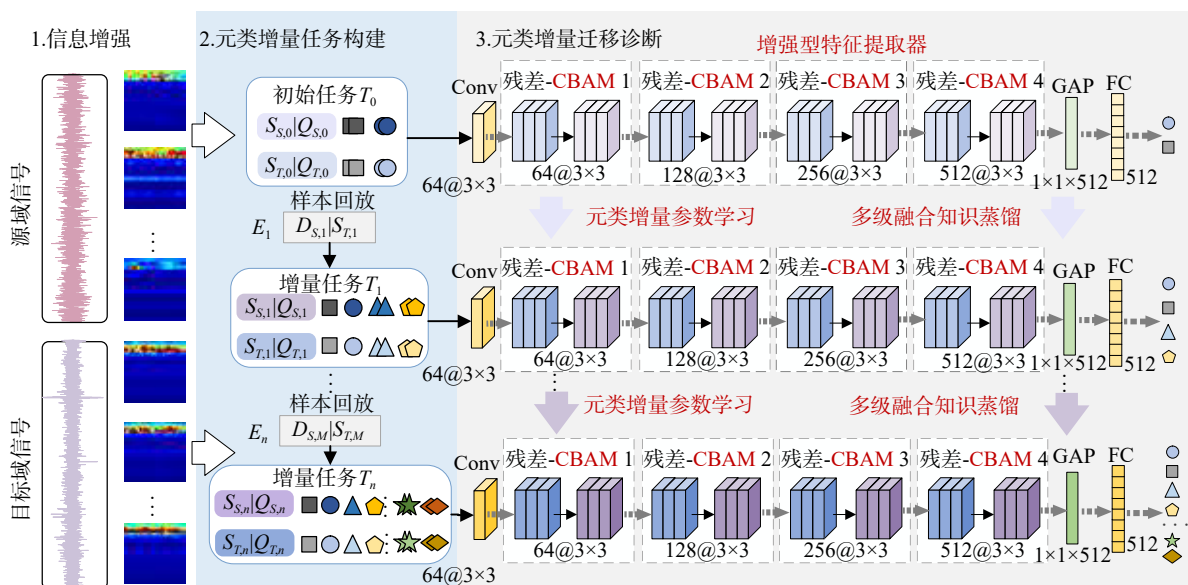


图 2 元类增量迁移学习模型架构

Fig. 2 The architecture of meta-class-incremental transfer learning model

2.1 增强型特征提取器

为有效提取增量迁移诊断场景的故障特征,本文所提元类增量学习迁移方法采用深度残差网络

ResNet 并集成卷积块注意力机制 (convolutional block attention mechanism, CBAM)^[19] 特征融合模块,设计了增强型特征提取器。该增强型特征提取器的结构如

图 2 所示, 包括一层 2D 卷积层 Conv、四个残差-CBAM 模块、一层全局平均池化层 GAP 以及一层全连接层 FC。表 1 展示了所提增强型特征提取器的架构。图 3(a) 展示了所提残差-CBAM 模块的架构, 由两个二维卷积层、一个 CBAM 特征融合模块和一个残差加法操作 Add 组成。

表 1 所提增强型特征提取器的架构

Tab. 1 Structure of the proposed enhanced feature extractor

网络层	参数
卷积层1	$C = 64, K = 3 \times 3, S = 1$
残差-CBAM模块1	$C = 64, K = 3 \times 3, S = 1$
残差-CBAM模块2	$C = 128, K = 3 \times 3, S = 2$
残差-CBAM模块3	$C = 256, K = 3 \times 3, S = 2$
残差-CBAM模块4	$C = 512, K = 3 \times 3, S = 2$
全局平均池化层	$K = 4 \times 4$
全连接层	诊断类别数

注: C 为通道数, K 为卷积核, S 为步长。

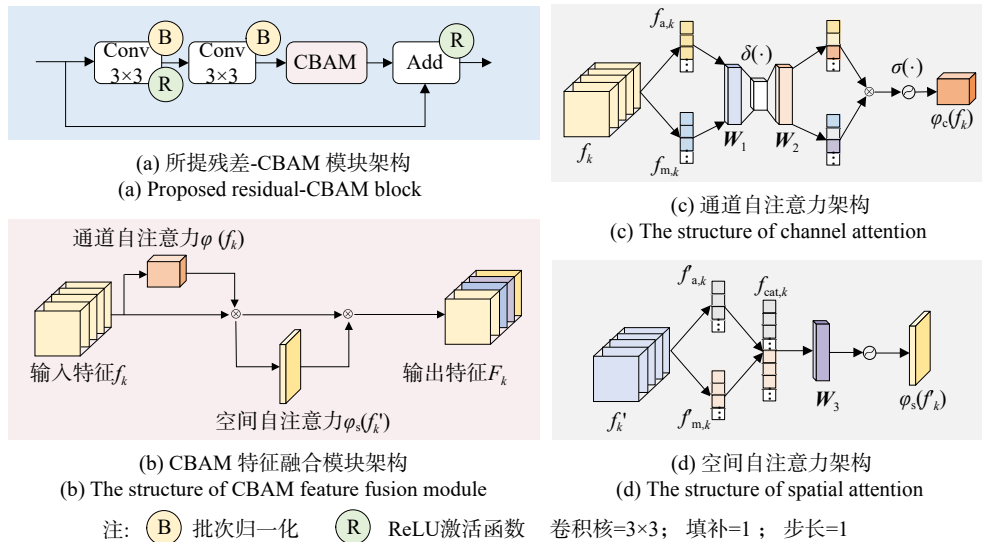


图 3 残差-CBAM 模块和 CBAM 特征融合模块架构

Fig. 3 The architecture of the proposed residual-CBAM block and the CBAM feature fusion module

2.2 多级融合知识蒸馏策略

所提方法中的多级融合知识蒸馏策略可缓解增量迁移诊断场景下的灾难性遗忘问题。所提策略融合了特征级与决策级的知识蒸馏机制, 旨在保留更丰富且深层次的故障结构信息, 以实现第 k 个至第 $k+1$ 个元类增量模型的高效知识迁移。

2.2.1 特征级知识蒸馏

所提特征级知识蒸馏策略通过在多个任务间共享元类增量模型的通用特征, 增强模型对潜在新故障模式与结构的理解能力, 从而在特征层面实现有效的知识保留与迁移, 具体过程阐述如下:

首先, 分别计算第 $k+1$ 个元类增量模型以及第 k 个元类增量模型的输出特征余弦相似度矩阵 S_k

和 S_{k+1} , 公式如下:

$$\varphi_c(f_k) = \sigma(W_2 \delta(W_1 f_{a,k}) + W_2 \delta(W_1 f_{m,k})) \quad (3)$$

$$\varphi_s(f'_k) = \sigma(W_3 f_{cat,k}) \quad (4)$$

$$f'_k = \varphi_c(f_k) \otimes f_k, F_k = \varphi_s(f'_k) \otimes f'_k \quad (5)$$

式中, $f_{a,k}$ 和 $f_{m,k}$ 分别表示在第 k 个增量任务 T_k 中, 通过通道维度的全局平均池化和全局最大池化提取得到的全局特征和显著特征; W_1 、 W_2 和 W_3 分别表示三个不同卷积层的权重矩阵; 函数 $\delta(\cdot)$ 和 $\sigma(\cdot)$ 分别表示 ReLU 激活函数和 Sigmoid 激活函数; $f_{cat,k}$ 为第 k 个增量任务 T_k 中空间维度的全局特征 $f'_{a,k}$ 和显著特征 $f'_{m,k}$ 的串联; f_k 为通道自注意力模块的输出特征; f_k 和 F_k 分别表示第 k 个任务 T_k 中 CBAM 特征融合模块的输入与输出特征; φ_c 和 φ_s 分别表示通道自注意力和空间自注意力的操作。

和 S_{k+1} , 公式如下:

$$\tilde{F}'_k = \frac{F'_k}{\|F'_k\|}, \tilde{F}'_{k+1} = \frac{F'_{k+1}}{\|F'_{k+1}\|} \quad (6)$$

$$S_k = \cos(\tilde{F}'_k, \tilde{F}'_k^T), S_{k+1} = \cos(\tilde{F}'_{k+1}, \tilde{F}'_{k+1}^T) \quad (7)$$

式中, F'_k 和 F'_{k+1} 分别表示第 k 个和第 $k+1$ 个元类增量模型特征提取器提取的特征; $\|\cdot\|$ 表示特征向量的欧几里德范数; \tilde{F}'_k 和 \tilde{F}'_{k+1} 分别表示第 k 个和第 $k+1$ 个元类增量模型特征提取器的归一化特征; \tilde{F}'_k^T 与 \tilde{F}'_{k+1}^T 分别表示 \tilde{F}'_k 与 \tilde{F}'_{k+1} 的转置; $\cos(\cdot)$ 表示余弦相似度运算。

其次, 将第 k 个和第 $k+1$ 个元类增量模型的相似度矩阵 S_k 和 S_{k+1} 转换为概率分布。由于余弦取值范围为 $[-1, 1]$, 因此需要将其映射至 $[0, 1]$, 公式如下:

$$\bar{S}_k = \frac{S_k + 1}{2}, \bar{S}_{k+1} = \frac{S_{k+1} + 1}{2} \quad (8)$$

式中, \bar{S}_k 和 \bar{S}_{k+1} 分别表示 S_k 和 S_{k+1} 的转换相似度矩阵。

然后, 将第 k 个和第 $k+1$ 个元类增量模型的转换相似度矩阵 \bar{S}_k 和 \bar{S}_{k+1} 逐行归一化为概率分布, 公式如下:

$$\bar{P}_k = \frac{\bar{S}_k(q)}{\sum_p \bar{S}_k(q, p)}, \bar{P}_{k+1} = \frac{\bar{S}_{k+1}(q)}{\sum_p \bar{S}_{k+1}(q, p)} \quad (9)$$

式中, $\bar{S}_k(q)$ 和 $\bar{S}_{k+1}(q)$ 分别表示转换相似度矩阵 \bar{S}_k 和 \bar{S}_{k+1} 的第 q 行相似度值; $\sum_p \bar{S}_k(q, p)$ 和 $\sum_p \bar{S}_{k+1}(q, p)$ 分别表示转换相似度矩阵 \bar{S}_k 和 \bar{S}_{k+1} 的第 q 行相似度值的总和; \bar{P}_k 和 \bar{P}_{k+1} 分别表示第 k 个和第 $k+1$ 个元类增量模型的增强预测概率。

最后, 特征级知识蒸馏损失 L_{FKD} 如下式所示:

$$L_{FKD} = \frac{1}{N_b} \sum_{j=1}^{N_b} \bar{P}_{k,j} \ln \left(\frac{\bar{P}_{k,j} + \varepsilon}{\bar{P}_{k+1,j} + \varepsilon} \right) \quad (10)$$

式中, N_b 为输入模型的批样本数量; ε 为极小正数。

2.2.2 决策级知识蒸馏

为进一步增强元类增量模型的类别预测能力, 本文提出一种决策级知识蒸馏方法。该方法通过有效捕获样本间的类内与类间关系, 增强模型在预测输出层的知识迁移效果, 具体过程阐述如下:

首先, 采用蒸馏温度 T_e 缩放第 k 个和第 $k+1$ 个元类增量模型的预测概率 P_k 和 P_{k+1} , 公式如下:

$$P_k = C_S \left(\frac{a_k \cdot F'_k + b_k}{T_e} \right), P_{k+1} = C_S \left(\frac{a_{k+1} \cdot F'_{k+1} + b_{k+1}}{T_e} \right) \quad (11)$$

式中, $\theta_k = (a_k, b_k)$ 和 $\theta_{k+1} = (a_{k+1}, b_{k+1})$ 分别表示第 k 个和第 $k+1$ 个元类增量模型的训练参数; $C_S(\cdot)$ 表示 Softmax 分类器。

然后, 计算类间关系损失 L_{inter} 和类内关系损失 L_{intra} , 以对齐类分布并保持特征关系, 公式如下:

$$L_{inter} = 1 - \frac{\sum_{j=1}^{N_b} (P_{k+1,j}^T - \bar{P}_{k+1}^T)(P_{k,j}^T - \bar{P}_k^T)}{\sqrt{\sum_{j=1}^{N_b} (P_{k+1,j}^T - \bar{P}_{k+1}^T)^2} \cdot \sqrt{\sum_{j=1}^{N_b} (P_{k,j}^T - \bar{P}_k^T)^2} + \varepsilon} \quad (12)$$

$$L_{intra} = 1 - \frac{\sum_{j=1}^{N_b} (P_{k+1,j} - \bar{P}_{k+1})(P_{k,j} - \bar{P}_k)}{\sqrt{\sum_{j=1}^{N_b} (P_{k+1,j} - \bar{P}_{k+1})^2} \cdot \sqrt{\sum_{j=1}^{N_b} (P_{k,j} - \bar{P}_k)^2} + \varepsilon} \quad (13)$$

式中, P_k^T 和 P_{k+1}^T 分别表示 P_k 和 P_{k+1} 的转置; \bar{P}_k 、 \bar{P}_{k+1} 、 \bar{P}_k^T 和 \bar{P}_{k+1}^T 分别表示 P_k 、 P_{k+1} 、 P_k^T 和 P_{k+1}^T 的均值。

最终, 决策级知识蒸馏损失 L_{DKD} 如下式所示:

$$L_{DKD} = \eta T_e^2 L_{inter} + \gamma T_e^2 L_{intra} \quad (14)$$

式中, η 和 γ 为调节 L_{inter} 和 L_{intra} 相对权重的超参数。

2.2.3 多级融合知识蒸馏

基于上述所提特征级知识蒸馏与决策级知识蒸

馏损失, 本文提出多级融合知识蒸馏策略, 其对应的损失函数 L_{MKD} 定义如下:

$$L_{MKD} = \lambda_1 L_{FKD} + \lambda_2 L_{DKD} \quad (15)$$

式中, λ_1 和 λ_2 为调节特征级知识蒸馏损失 L_{FKD} 与决策级知识蒸馏损失 L_{DKD} 之间相对权重的超参数。

2.3 元类增量参数学习机制

为解决增量迁移诊断场景中模型泛化能力低的难题, 本文进一步将改进的元训练和快速自适应策略与增量学习^[20] 框架相融合, 提出元类增量参数学习机制。该机制通过内外循环训练策略, 元训练能够有效捕获泛化特征, 从而提高模型在不同工况条件下的适应性, 进而增强类增量模型的泛化能力。

除初始任务的初始参数 $\theta_0^{(0)}$ 采用随机初始化外, 第 $k+1$ 个元类增量模型的初始参数均继承自第 k 个元类增量模型。关于元训练, 在增量任务 T_{k+1} 的内循环阶段, 首先基于源域支持集 $S_{S,k+1}$ 进行参数多轮次梯度下降迭代更新, 公式如下:

$$\theta_k^{(i)} = \theta_k^{(i-1)} - \alpha (\nabla_{\theta_k^{(i-1)}} L_{clf} + \nabla_{\theta_k^{(i-1)}} L_{MKD}) \quad (16)$$

式中, $\theta_k^{(i-1)}$ 表示增量任务 T_k 在第 $i-1$ 个内循环的训练参数, 用于更新参数 $\theta_k^{(i)}$; $\nabla_{\theta_k^{(i-1)}}$ 表示对 $\theta_k^{(i-1)}$ 求导; L_{clf} 表示分类器的损失; α 表示元类增量模型的动态学习率。

在外循环阶段, 增量任务 T_{k+1} 的元训练参数 θ_{k+1} 在经过 R 轮内循环迭代后继续迭代更新, 并且重复执行 t 轮外循环迭代, 公式如下:

$$\theta_{k+1}^{(t)} = \theta_k^{(t-1)} - \alpha \sum_{i=1}^R \nabla_{\theta_k^{(t-1,i-1)}} (L_{clf} + L_{MKD}) \quad (17)$$

式中, $\theta_k^{(t-1)}$ 表示增量任务 T_k 在第 $t-1$ 个外循环的训练参数, 用于更新增量任务 T_{k+1} 的元训练参数 $\theta_{k+1}^{(t)}$; $\theta_k^{(t-1,i-1)}$ 为第 $t-1$ 个外循环下第 $i-1$ 个内循环中的参数; $\nabla_{\theta_k^{(t-1,i-1)}}$ 表示对 $\theta_k^{(t-1,i-1)}$ 求导。

执行元训练阶段后可以获得增量任务 T_{k+1} 的模型训练参数 θ_{k+1} 。随后, 评估验证元类增量模型在源域查询集 $Q_{S,k}$ 的泛化能力, 并动态调整学习率 α 。在快速自适应阶段, 利用目标域支持集 $S_{T,k}$ 对元训练参数 θ_{k+1} 进行 e 轮微调, 公式如下:

$$\theta_{k+1}^{(e)} = \theta_{k+1}^{(e-1)} - \beta (\nabla_{\theta_{k+1}^{(e-1)}} L_{clf} + \nabla_{\theta_{k+1}^{(e-1)}} L_{MKD}) \quad (18)$$

式中, $\theta_{k+1}^{(e)}$ 表示增量任务 T_{k+1} 中第 e 次迭代后的训练参数, 最终训练参数记为 θ_{k+1} ; $\theta_{k+1}^{(e-1)}$ 表示模型在第 $e-1$ 轮的训练参数; $\nabla_{\theta_{k+1}^{(e-1)}}$ 表示对 $\theta_{k+1}^{(e-1)}$ 求导; β 为快速自适应阶段的学习率。由于不存在先验模型, 初始任务 T_0 的初始元类增量模型训练过程不包含多级知识蒸馏损失 L_{MKD} 。所提元类增量迁移学习 MCITL 模型的诊断算法流程如表 2 所示。

表 2 所提 MCITL 模型的算法流程

Tab. 2 Algorithm procedures of proposed MCITL model

算法: 元类增量迁移学习方法

输入: 信息融合源域数据 $D_{S,k}$, 信息融合目标域支持集 $S_{T,k}$, 初始任务循环数 e_{it} , 增量任务循环数 e_{in} , 每类记忆样本数 M , 批次大小 b , 快速自适应循环数 e_f , 蒸馏温度 T_e , 学习率 α , β , η , γ .

输出: 训练参数 θ_{k+1} .

1: **If** $k=0$ (初始任务) **then**

2: 随机初始化参数 $\theta_0^{(0)}$.

3: **For** each epoch **do**

4: 根据 $D_{S,0}$ 更新元训练参数:

$$5: \theta_0^{(i)} \leftarrow \theta_0^{(i-1)} - \alpha \nabla_{\theta_0^{(i-1)}} L_{\text{clf}}, \theta_1^{(i)} \leftarrow \theta_0^{(i-1)} - \alpha \sum_{j=1}^R \nabla_{\theta_0^{(i-1,j-1)}} L_{\text{clf}}.$$

6: 根据 $S_{T,0}$ 快速自适应微调:

$$7: \theta_1^{(e)} \leftarrow \theta_1^{(e-1)} - \beta \nabla_{\theta_1^{(e-1)}} L_{\text{clf}}.$$

8: **End For**

9: 构建样本回放数据集 E_1

10: **Else** (增量任务 T_{k+1})

11: 初始化参数 $\theta_k^{(0)} = \theta_k'$.

12: **For** each epoch **do**

13: 根据 $D_{S,k+1}$ 更新元训练参数:

$$14: \theta_k^{(i)} \leftarrow \theta_k^{(i-1)} - \alpha (\nabla_{\theta_k^{(i-1)}} L_{\text{clf}} + \nabla_{\theta_k^{(i-1)}} L_{\text{MKD}}),$$

$$15: \theta_{k+1}^{(i)} \leftarrow \theta_k^{(i-1)} - \alpha \sum_{j=1}^R \nabla_{\theta_k^{(i-1,j-1)}} (L_{\text{clf}} + L_{\text{MKD}}).$$

16: 根据 $S_{T,k+1}$ 快速自适应微调:

$$17: \theta_{k+1}^{(e)} \leftarrow \theta_{k+1}^{(e-1)} - \beta (\nabla_{\theta_{k+1}^{(e-1)}} L_{\text{clf}} + \nabla_{\theta_{k+1}^{(e-1)}} L_{\text{MKD}}).$$

18: **End For**

19: 构建样本回放数据集 E_{k+1}

20: **End**

2.4 元类增量迁移学习方法流程

本文所提元类增量迁移学习方法的流程包括下述 3 个关键步骤。

步骤 1: 信息增强。 采用连续小波变换对源域和目标域振动信号样本进行特征增强, 以生成信息增强样本。

步骤 2: 元增量任务构建。 构造源域和目标域包括支持集和查询集的增量任务, 同时为进一步缓解灾难性遗忘, 在第 $k+1$ 个增量任务中引入样本回放策略^[21], 以辅助元类增量模型的训练。

步骤 3: 元类增量迁移诊断。 对于初始任务, 首先通过元类增量参数学习获取训练参数, 并利用目标域查询集 $Q_{T,0}$ 评估训练后的元类增量模型, 从而获得初始任务的诊断结果。对于第 $k+1$ 个增量任务, 结合多级融合知识蒸馏和元类增量参数学习更新模型参数。最后, 利用目标域查询集 $Q_{T,k+1}$ 评估更新后的元类增量模型, 生成最终迁移诊断结果。

3 列车牵引传动系统试验验证

本节分析列车牵引传动系统故障数据集, 验证所提元类增量迁移学习方法的有效性与优越性。

3.1 对比方法和参数设置

本节选取 4 种主流的类增量学习和基于微调的方法作为所提方法的对比方法, 详细阐述如下。

(1) iCaRL^[21]: 增量分类器与表示学习 (incremental classifier and representation learning, iCaRL) 方法是所提方法的对比基准, 其核心为第 $k+1$ 个类增量模型通过样本回放与知识蒸馏^[22] 在第 k 个类增量模型指导下实现知识保留与迁移。

(2) WA^[23]: 权重对齐 (weight aligning, WA) 方法通过引入权重对齐策略提高数据不平衡场景增量学习模型的知识迁移性能。

(3) Replay^[24]: 第 $k+1$ 个增量模型仅通过样本回放在第 k 个增量模型指导下实现知识保留迁移。

(4) DER^[25]: 动态可扩展表示 (dynamically expandable representation, DER) 方法通过设计辅助任务损失指导类增量学习。

(5) Finetune^[24]: 该方法是基于参数微调的智能迁移诊断方法, 它使用当前训练的模型作为起始点, 并在新的任务上进行模型训练微调。

前述对比方法除 Finetune 方法外, 其余类增量学习方法均采用样本回放策略, 且每一类别的样本回放存储容量设为 30。对于所提元类增量迁移学习方法, 初始任务与增量任务的训练迭代次数分别为 20 和 10, 批次大小为 32, 快速适应阶段的优化迭代次数为 20。对于对比方法, 初始任务与增量任务的迭代次数分别为 40 和 30。所有方法的初始学习率均为 0.001, 初始任务与增量任务的权重衰减参数分别为 0.0005 和 0.0002, 优化器均采用 Adam。除所提 MCITL 方法使用增强型特征提取器外, 其余所有对比方法均采用 ResNet 作为特征提取器。所有方法均基于长度为 1024 的样本进行预处理, 并基于 cmor 小波作为母小波的连续小波变换实现样本信息增强, 信息增强样本的输入形状为 $32 \times 32 \times 3$ 。试验配置为 PyTorch 2.4.0 与 RTX 4090D GPU 显卡。为避免随机因素, 所有方法采用 5 次独立重复试验平均结果。

3.2 数据集概述

列车牵引传动系统试验台, 如图 4(a) 所示^[11, 26]。试验数据集由安装在齿轮箱输入轴旁的加速度计 (通道序号 11) 采集, 采样频率为 64 kHz。试验数据集涵盖 4 种工况, 即电机转速 20、40、60 Hz 以及电



图4 地铁列车牵引传动系统试验台与九种故障状态

Fig. 4 Experimental platform of subway train transmission system and nine faulty states

机转速为 60 Hz 且分别施加横向负载+10 kN 和-10 kN (即 60 Hz/+10 kN 与 60 Hz/-10 kN)。试验数据集考虑 10 种健康状态, 包括正常状态和 9 种不同故障状态, 故障状态如图 4(b)~(j) 所示。

列车传动系统数据集的类增量任务设置如表 3 所示。为模拟工业场景持续出现新故障模式, 10 种健康状态划分为 1 个初始任务和 4 个增量任务。其中每种健康状态的训练样本为 520 个, 包括 500 个源域训练样本和 20 个目标域训练样本; 每种健康状态

表3 地铁列车牵引传动系统数据集的类增量任务设置

Tab. 3 Class-incremental task settings of subway train transmission system dataset

任务	健康状态	标签
初始任务	正常	0
	齿根裂纹	1
增量任务1	齿面磨损	2
	齿轮缺齿	3
增量任务2	齿轮断齿	4
	轴承内圈故障	5
增量任务3	轴承外圈故障	6
	轴承滚动体故障	7
增量任务4	轴承保持架故障	8
	电机短路	9

表4 4 种类增量迁移诊断场景中不同方法的诊断精度 (单位: %)

Tab. 4 Diagnosis accuracies across different class-incremental transfer diagnosis scenarios (Unit: %)

方法	类增量迁移诊断场景				平均诊断精度
	20 Hz→60 Hz	40 Hz→60 Hz	60 Hz /+10 kN→60 Hz	60 Hz/-10 kN→60 Hz	
MCITL	92.10±2.02	95.96±1.21	95.74±1.16	96.02±1.16	94.96±1.65
iCaRL	87.98±3.31	92.81±4.53	90.20±4.71	92.03±3.87	90.76±1.86
WA	86.66±2.87	94.36±1.89	92.20±3.57	93.34±4.75	91.64±2.97
Replay	84.58±2.74	93.00±1.61	94.07±1.36	92.59±4.06	91.06±3.78
DER	87.54±2.87	91.11±3.96	93.97±2.72	94.88±1.97	91.86±2.85
Finetune	45.59±0.03	45.72±0.08	45.72±0.13	45.67±0.07	45.68±0.05

的目标域测试样本为 480 个。关于对比方法, 每种健康状态的 520 个训练样本直接用于模型训练, 480 个目标域测试样本用于测试, 从而完成增量迁移诊断。相比之下, 所提 MCITL 方法首先开展元训练, 分别采用每种健康状态的 400 和 100 个样本构建源域支持集与源域查询集; 随后在快速适应阶段, 利用每种健康状态 20 个目标域支持集样本进行微调。最后, 将 480 个目标域测试样本输入训练好的元类增量模型, 实现元类增量迁移诊断。

3.3 试验结果分析

首先, 评估所提元类增量迁移学习方法在 5 个增量任务上的诊断性能优越性。在此基础上, 分析 4 种典型类增量迁移诊断场景 (即 20 Hz 至 60 Hz、40 Hz 至 60 Hz、60 Hz/+10 kN 至 60 Hz 以及 60 Hz/-10 kN 至 60 Hz 的增量迁移诊断场景)。4 种类增量迁移诊断背景下不同方法的诊断精度和平均诊断精度分别如表 4 和图 5 所示。

所提 MCITL 方法在 4 种类增量迁移诊断场景中展示了卓越的诊断性能。具体而言, 所提 MCITL 方法取得 94.96% 的平均诊断精度, 超出 iCaRL 方法 4.20%、WA 方法 3.32%、Replay 方法 3.90%、DER 方

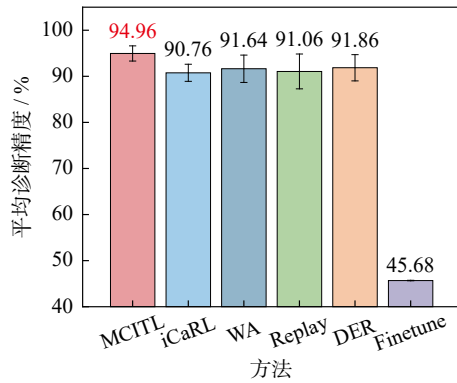


图 5 4 种类增量迁移诊断场景不同方法的平均诊断精度
Fig. 5 Average diagnosis accuracies of various methods across four class-incremental transfer diagnosis scenarios

法 3.10% 和 Finetune 方法 48.89%。这主要得益于所提 MCITL 方法能够在类增量迁移诊断场景中高效学习更全面且泛化能力更好的故障特征。所有方法中, Finetune 方法的诊断精度最低, 仅为 45.68%。这一显著差异突显了样本回放策略在类增量迁移诊断模型中的关键作用。

为评估所提 MCITL 方法在不同增量任务中的诊断性能, 以 20 Hz 至 60 Hz 的类增量迁移诊断场景为例, 不同方法的跨任务诊断精度如图 6 所示。结果表明, 所提 MCITL 方法在 5 个不同增量任务中均展现稳定的诊断性能, 在初始任务中取得 100% 的最高

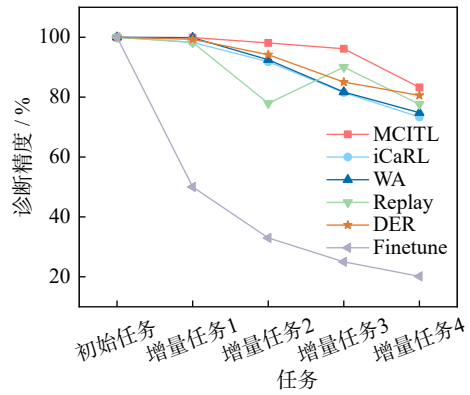


图 6 20 Hz 至 60 Hz 迁移诊断场景不同方法的跨任务诊断精度
Fig. 6 Cross-task diagnosis accuracies of various approaches in the transfer diagnosis scenario from 20 Hz to 60 Hz

精度, 最具挑战性的增量任务 4 中, 仍保持 83.25% 的较高精度, 领先于所有对比方法。值得注意的是, 在初始任务之后, Finetune 微调方法的性能显著下降, 表明其无法有效完成类增量迁移诊断任务。

为评估不同类增量学习方法的抗遗忘性能, 进一步采用遗忘率为模型稳定性度量指标。四种类增量迁移诊断情景下不同方法的遗忘率和平均遗忘率分别如表 5 和图 7 所示。结果表明, 本文所提 MCITL 方法的平均遗忘率最低为 3.85%, 相较 iCaRL、WA、Replay 和 DER 对比方法分别低 0.62%、0.33%、0.85% 和 0.73%。

表 5 4 种类增量迁移诊断场景下不同方法的遗忘率 (单位: %)

方法	类增量迁移诊断场景				平均遗忘率
	20 Hz→60 Hz	40 Hz→60 Hz	60 Hz /+10 kN→60 Hz	60 Hz/-10 kN→60 Hz	
MCITL	5.33±0.87	2.84±0.68	3.88±1.17	3.35±0.99	3.85±0.93
iCaRL	6.14±0.50	3.56±1.04	5.59±2.07	2.60±0.55	4.47±1.45
WA	6.39±1.08	3.26±0.74	4.39±1.39	2.69±1.25	4.18±1.41
Replay	7.36±1.70	3.79±1.45	4.15±0.86	3.51±1.52	4.70±1.55
DER	6.24±2.21	4.93±1.32	3.65±1.57	3.51±1.09	4.58±1.11

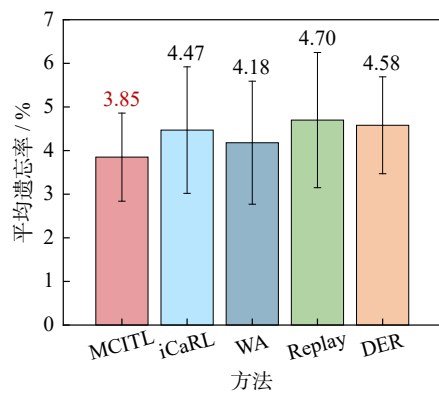


图 7 4 种类增量迁移诊断场景下不同方法的平均遗忘率
Fig. 7 Average forgetting rates of different methods across four class-incremental transfer diagnosis scenarios

图 8 显示了 20 Hz 至 60 Hz 迁移场景中, 不同类增量学习方法的任务遗忘率指标。所提 MCITL 方

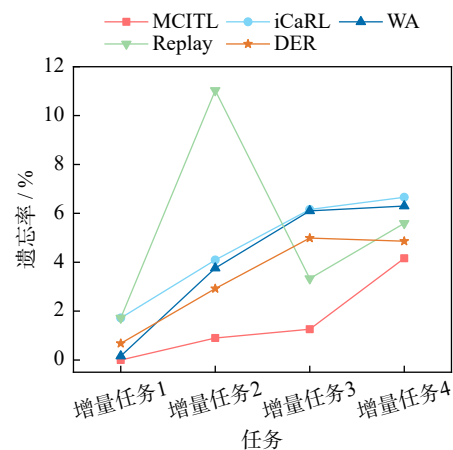


图 8 20 Hz 至 60 Hz 迁移诊断场景下不同方法的任务遗忘率
Fig. 8 Task forgetting rates of different methods in the transfer diagnosis scenario from 20 Hz to 60 Hz

法在 4 个类增量任务中,始终保持最低且稳定的遗忘率。相比之下,其他增量学习对比方法随着增量任务的增加,遗忘率较大且逐渐上升,表明对比方法存在灾难性遗忘问题。综合上述诊断精度和遗忘率分析结果,验证了所提 MCITL 方法的有效性和优越性。

3.4 消融试验分析

为评估所提 MCITL 方法各模块的贡献,开展消融试验分析,将所提方法与去除 CBAM 特征融合模块(方法 1)、去除元类增量参数学习模块(方法 2)以

及去除多级融合知识蒸馏模块(方法 3)进行对比。表 6 和 7 分别列出了 4 种类增量迁移诊断场景下消融试验的诊断精度与遗忘率。

消融试验结果表明,所提 MCITL 方法在各类增量迁移诊断场景中始终取得最佳诊断准确率和最低遗忘率。其中,元类增量参数学习模块对提升模型性能的作用最为关键。元学习策略显著增强了类增量学习在动态环境中的泛化性能与抗遗忘特性,有效提升了类增量迁移诊断性能。上述消融试验结果验证了所提 MCITL 方法在类增量迁移诊断场景中的优势。

表 6 4 种类增量迁移诊断场景下消融试验的诊断精度(单位: %)

Tab. 6 Diagnosis accuracies for ablation experiment across four class-incremental transfer diagnosis scenarios(Unit: %)

方法	类增量迁移诊断场景				平均诊断精度
	20 Hz → 60 Hz	40 Hz → 60 Hz	60 Hz /+10 kN → 60 Hz	60 Hz /-10 kN → 60 Hz	
MCITL	92.10±2.02	95.96±1.21	95.74±1.16	96.02±1.16	94.96±1.65
方法1	88.65±3.09	93.53±3.50	92.94±4.21	94.26±1.17	92.35±2.18
方法2	76.47±2.77	87.01±3.75	90.26±3.34	89.21±4.83	85.74±5.48
方法3	89.97±1.95	93.02±1.54	93.09±3.09	94.54±1.48	92.66±1.66

表 7 4 种类增量迁移诊断场景下消融试验的遗忘率(单位: %)

Tab. 7 Forgetting rates for ablation experiment across four class-incremental transfer diagnosis scenarios(Unit: %)

方法	类增量迁移诊断场景				平均遗忘率
	20 Hz → 60 Hz	40 Hz → 60 Hz	60 Hz /+10 kN → 60 Hz	60 Hz /-10 kN → 60 Hz	
MCITL	5.33±0.87	2.84±0.68	3.88±1.17	3.35±0.99	3.85±0.93
方法1	7.33±1.99	2.99±0.86	4.43±1.20	4.06±1.03	4.70±1.61
方法2	8.98±1.63	5.16±1.21	4.79±1.08	4.04±1.09	5.74±1.91
方法3	7.25±0.74	3.56±0.68	4.77±0.86	3.76±0.74	4.84±1.47

4 结 论

本文提出了元类增量迁移学习驱动的跨域终身智能诊断方法,以解决新故障模式不断涌现与跨运行工况条件下的类增量迁移诊断难题,为跨域持续诊断提供解决方案。首先,设计了增强型特征提取器,将残差网络与卷积块自注意力机制相结合,旨在有效提取关键故障特征。然后,提出了多级融合知识蒸馏策略,通过结合特征级与决策级蒸馏,有效缓解增量迁移诊断中的灾难性遗忘问题。最后,提出了元类增量参数训练策略,提升模型在类增量迁移诊断中的泛化能力。列车牵引传动系统的试验结果充分证明了所提元类增量迁移学习方法实现增量迁移诊断的有效性与优势,且与现有类增量诊断方法相比,所提方法在不同工况的表现更稳健。

尽管所提元类增量迁移学习方法展现出优越的增量迁移诊断性能,但其计算复杂度仍相对较高,导致模型训练的时间成本较高。因此,未来研究将重点开发计算高效的无监督类增量迁移学习方法,突破计算时间成本高昂与目标域标签样本稀缺的限制。同时,将探索方法在其他重大装备领域(如风电

机组、航空发动机等)中的增量迁移诊断能力,以提升模型在多种实际工业场景的诊断性能。

参考文献:

- [1] 陈钱, 陈康康, 董兴建, 等. 一种面向机械设备故障诊断的可解释卷积神经网络[J]. 机械工程学报, 2024, 60(12): 65-76.
CHEN Qian, CHEN Kangkang, DONG Xingjian, et al. Interpretable convolutional neural network for mechanical equipment fault diagnosis[J]. Journal of Mechanical Engineering, 2024, 60(12): 65-76.
- [2] 杨诚, 何清波, 贾民平, 等. 流形正则化支持高阶张量机及其在行星齿轮箱半监督故障诊断中的应用[J]. 振动工程学报, 2025, 38(1): 78-87.
YANG Cheng, HE Qingbo, JIA Mingping, et al. Manifold regularized support higher-order tensor machines for semi-supervised fault diagnosis of planetary gearboxes[J]. Journal of Vibration Engineering, 2025, 38(1): 78-87.
- [3] 袁静, 任港星, 蒋会明, 等. 基于多元提升核神经网络的机械故障诊断方法及其特征提取可解释性研究[J]. 机械工程学报, 2024, 60(12): 51-64.

- YUAN Jing, REN Gangxing, JIANG Huiming, et al. Neural network driven by multiple lifting kernels for mechanical fault diagnosis and its interpretability research of feature extraction[J]. *Journal of Mechanical Engineering*, 2024, 60(12): 51-64.
- [4] 张洁, 孔运, 黄国昱, 等. 面向高可信机械故障诊断的多模态数据插补与融合方法 [J/OL]. *振动工程学报*, 1-12 [2025-04-28]. <http://kns.cnki.net/kcms/detail/32.1349.TB.20250214.1644.004.html>.
- ZHANG Jie, KONG Yun, HUANG Guoyu, et al. Multi-modal data imputation and fusion method for highly trustworthy machine fault diagnosis[J/OL]. *Journal of Vibration Engineering*, 1-12[2025-04-28]. <http://kns.cnki.net/kcms/detail/32.1349.TB.20250214.1644.004.html>.
- [5] HE H X, ZHUANG C B, XIONG H. A novel incremental method for bearing fault diagnosis that continuously incorporates unknown fault types[J]. *Mechanical Systems and Signal Processing*, 2024, 216: 111524.
- [6] LIU Y, CHEN B J, WANG D, et al. A lifelong learning method based on generative feature replay for bearing diagnosis with incremental fault types[J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1-11.
- [7] XU X F, BAO S, LIANG P F, et al. A broad learning model guided by global and local receptive causal features for online incremental machinery fault diagnosis[J]. *Expert Systems with Applications*, 2024, 246: 123124.
- [8] MASANA M, LIU X L, TWARDOWSKI B, et al. Class-incremental learning: survey and performance evaluation on image classification[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(5): 5513-5533.
- [9] ZHANG H R, YAO Y F, WANG Z X, et al. Class incremental fault diagnosis under limited fault data via supervised contrastive knowledge distillation[J]. *IEEE Transactions on Industrial Informatics*, 2025, 21(6): 4344-4354.
- [10] LI J P, YUE K, WU Z Q, et al. KDN: a class-added continual learning framework for cross-machine fault diagnosis with limited samples[J]. *Mechanical Systems and Signal Processing*, 2025, 227: 112379.
- [11] DING A, QIN Y, WANG B, et al. Evolvable graph neural network for system-level incremental fault diagnosis of train transmission systems[J]. *Mechanical Systems and Signal Processing*, 2024, 210: 111175.
- [12] YAN S, SHAO H D, WANG X D, et al. Few-shot class-incremental learning for system-level fault diagnosis of wind turbine[J/OL]. *IEEE/ASME Transactions on Mechatronics*, 1-10 [2025-04-28]. <https://ieeexplore.ieee.org/document/10759278>.
- [13] HU K, HE Q B, CHENG C M, et al. Adaptive incremental diagnosis model for intelligent fault diagnosis with dynamic weight correction[J]. *Reliability Engineering & System Safety*, 2024, 241: 109705.
- [14] GU X H, ZHAO Y L, YANG G, et al. An imbalance modified convolutional neural network with incremental learning for chemical fault diagnosis[J]. *IEEE Transactions on Industrial Informatics*, 2022, 18(6): 3630-3639.
- [15] SHI M K, DING C C, CHANG S Y, et al. Cross-domain class incremental broad network for continuous diagnosis of rotating machinery faults under variable operating conditions[J]. *IEEE Transactions on Industrial Informatics*, 2024, 20(4): 6356-6368.
- [16] XU X F, BAO S, SHAO H D, et al. A multi-sensor fused incremental broad learning with D-S theory for online fault diagnosis of rotating machinery[J]. *Advanced Engineering Informatics*, 2024, 60: 102419.
- [17] ZHANG Y, HAN D Y, SHI P M. Semi-supervised prototype network based on compact-uniform-sparse representation for rotating machinery few-shot class incremental fault diagnosis[J]. *Expert Systems with Applications*, 2024, 255: 124660.
- [18] HUANG X F, XIE T L, LUO S Y, et al. Incremental learning with multi-fidelity information fusion for digital twin-driven bearing fault diagnosis[J]. *Engineering Applications of Artificial Intelligence*, 2024, 133: 108212.
- [19] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[M]//*Computer Vision—ECCV 2018*. Cham: Springer International Publishing, 2018: 3-19.
- [20] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[EB/OL]. (2017-03-09)[2025-04-28]. <https://arxiv.org/abs/1703.03400v3>.
- [21] REBUFFI S A, KOLESNIKOV A, SPERL G, et al. iCaRL: incremental classifier and representation learning[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017: 5533-5542.
- [22] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015-03-09)[2025-04-28]. <https://arxiv.org/abs/1503.02531>
- [23] ZHAO B W, XIAO X, GAN G J, et al. Maintaining discrimination and fairness in class incremental learning[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020: 13205-13214.
- [24] ZHOU D W, WANG F Y, YE H J, et al. PyCIL: a Python toolbox for class-incremental learning[J]. *Science China-Information Sciences*, 2023, 66(9): 197101.
- [25] YAN S P, XIE J W, HE X M. DER: dynamically expandable representation for class incremental learning[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2021: 3013-3022.
- [26] LIN C Y, KONG Y, HAN Q K, et al. IF-EDAAN: an information fusion-enhanced domain adaptation attention network for unsupervised transfer fault diagnosis[J]. *Mechanical Systems and Signal Processing*, 2025, 224: 112180.

第一作者: 林翠颖(1998—), 女, 博士研究生。

E-mail: cuiying.lin@bit.edu.cn

通信作者: 孔运(1993—), 男, 博士, 副教授。

E-mail: kongyun@bit.edu.cn