

小样本条件下基于改进原型网络的 电力变压器故障诊断方法^{*}

韦强宇¹ 吐松江·卡日¹ 张紫薇² 李振恩¹ 黄 瑞¹

(1.新疆大学电气工程学院 乌鲁木齐 830049; 2.清华四川能源互联网研究院 成都 610213)

摘要:为解决小样本条件下变压器故障诊断准确率低的问题,提出了一种基于改进原型网络的变压器故障诊断方法。首先,采用基于残差结构的多层感知机对变压器油色谱数据进行特征提取,借助深层网络结构获取更有判别力的特征表示;然后,提出了一种任务与样本双层注意力驱动的原型构建机制,通过任务条件注意力机制动态调整支持集的表示,并结合样本注意力机制对支持样本进行加权以突出判别特征,从而生成更有区分性的类原型,解决了原型网络在跨任务适应性和类内判别性方面的不足;最后,通过计算样本与各类别原型之间的欧氏距离实现变压器故障诊断。实验结果表明,在 IEEE Dataport 数据集上本文方法相比原型网络、匹配网络和关系网络的诊断准确率分别提高了 15.14%、21.49% 和 18.68%,在 IEC TC 10 Dataport 上分别提高了 6.43%、6.90% 和 8.65%。为小样本条件下变压器故障诊断方法提供新的思路与参考。

关键词: 电力变压器;原型网络;故障诊断;小样本学习;样本注意力;任务条件注意力

中图分类号: TN713;TM411 **文献标识码:** A **国家标准学科分类代码:** 470.40

Fault diagnosis method of power transformer based on an improved prototypes network under few-shot conditions

Wei Qiangyu¹ Tusongjiang·Kari¹ Zhang Ziwei² Li Zhenen¹ Huang Rui¹

(1. School of Electrical Engineering, Xinjiang University, Urumqi 830049, China;

2. Sichuan Energy Internet Research Institute, Chengdu 610213, China)

Abstract: To address low accuracy in transformer fault diagnosis under small-sample conditions, we propose a method based on an improved prototypical network. First, a residual multi-layer perceptron extracts features from transformer oil chromatography data to obtain more discriminative representations. Second, we introduce a dual-layer attention-driven prototype construction mechanism; Task-conditioned attention dynamically adjusts support-set representations, and sample attention weights support samples to emphasize discriminative features. This yields more distinguishable class prototypes and addresses prototypical networks' limitations in cross-task adaptability and intra-class discriminability. Finally, diagnosis is performed by measuring the Euclidean distance between samples and class prototypes. Experiments on the IEEE Dataport and IEC TC 10 Dataport datasets show significant improvements: On the IEEE Dataport dataset our method outperforms the prototypical network, matching network and relation network by 15.14%, 21.49% and 18.68%, respectively; on the IEC TC 10 Dataport the gains are 6.43%, 6.90% and 8.65%, respectively. The study provide new insights and references for transformer fault diagnosis under small-sample conditions.

Keywords: power transformer; prototype network; fault diagnosis; few shot learning; sample attention; task-conditional attention

0 引 言

变压器作为电力系统的关键设备,其运行状态直接关系

到电网的安全性与稳定性。因此,建立高效准确的变压器故障诊断方法有助于及时发现其潜在故障并制定合理、有效的应对策略,从而最大程度地保证电网的安全稳定运行^[1]。

收稿日期:2025-10-18

^{*}基金项目:新疆维吾尔自治区重大专项(2025A01006-3)、国家自然科学基金(52067021,52207165)、新疆维吾尔自治区重点研发项目(2024B04002-1)资助

随着在线监测与传感器技术的普及,研究学者可通过油中溶解气体、振动、温度以及电流、电压波形等多源信号对变压器运行状态进行实时感知与分析,这些监测手段为故障特征提取提供了丰富的信息。深度学习方法凭借强大的非线性特征学习与表示能力,能够自适应地提取并利用关键信息而备受关注^[2]。研究人员采用卷积神经网络^[3]、Transformer神经网络^[4]以及残差神经网络^[5]等深度学习模型实现故障诊断,并取得了较高的准确率,但其训练通常依赖大量带标签样本。实际上,变压器在发生故障前或故障早期就被检修或更换,很难获取足够的带标签故障样本^[6]。因此,如何在小样本情况下实现变压器故障准确识别受到广泛关注。

近年来,针对小样本故障诊断问题,已有部分研究并取得一定的成效。Ding等^[7]利用生成式对抗网络通过随机噪声合成各类标签样本,以扩充数据集,解决标签样本过少问题。王欢等^[8]将梅尔声谱图与改进 Wasserstein 生成对抗网络相结合构建的声纹诊断模型,显著提升了小样本条件下变压器声纹故障识别性能。生成式对抗网络虽能有效提升变压器故障诊断性能,但存在模型训练耗时长、收敛性差和容易发生模式崩溃等问题。

元学习^[9]算法的核心思想是让模型能够在有限新任务数据下快速适应和泛化,该方法在小样本文本分类、图像分类以及故障诊断等领域表现出良好效果^[10-11]。He等^[12]提出基于匹配网络的混合方法,缓解域间差异问题,实现在有限训练数据和复杂工作环境下的机械故障诊断。Xue等^[13]提出注意力加权关系网络,通过突出关键样本特征实现小样本条件下机械故障诊断。Tang等^[14]提出轻量化关系网络变体用于强调跨域泛化与效率,实现跨域少样本故障诊断。Chen等^[15]将残差结构与关系网络结合,以增强特征提取与分类能力,实现小样本条件下工业设备故障识别。这些研究表明,关系网络、匹配网络等元学习方法能够显著缓解工业场景中“多数正常、少数故障”的数据不平衡问题。但这些方法仍依赖相对稳定、特征明显的数据场景,在处理复杂工况、微弱故障特征方面仍存在局限,难以满足小样本条件下变压器故障诊断实际需求。

在众多元学习方法中,原型网络(prototypical networks, PN)^[16]因结构简单、度量清晰、收敛稳定而备受关注,该方法通过构建每个类别在嵌入空间中1个原型向量,然后计算查询样本与各类别原型的距离,最后将查询样本划归距离最小的原型对应类别来完成分类,其凭借结构简单、泛化能力强、计算效率高等优势被广泛应用于小样本分类。

目前,一些研究工作将其引入到小样本故障诊断任务,实现在小样本条件下有效识别故障。Wang等^[17]提出基于时域和频域融合的小波原型网络,利用并行双通道卷积与小波层强化信号特征提取以实现小样本故障诊断。余浩帅等^[18]设计混合自注意力原型网络通过引入自注意力模块

获取更强的判别性特征,实现风电齿轮箱故障识别。Wang等^[19]提出一种基于改进原型网络的滚动轴承故障诊断方法,通过对原型加权来增强类别表征,提升小样本条件下的诊断性能。Li等^[20]提出一种加权正则化原型网络,通过类内重加权与正则化来减小噪声对原型估计的影响,从而提升小样本故障诊断稳定性。这些方法能够在有限样本条件下实现较高的诊断准确率,但在弱故障或早期故障特征识别方面表现仍有限,类内差异大的情况下容易导致原型计算偏差。虽然原型网络及其改进方法在小样本工业故障诊断中具有良好的分类性能和泛化能力,但多数研究集中在振动类机械设备,对于特征复杂的电力变压器适用性仍不明确。

为解决小样本条件下变压器故障样本不足且难以支撑深度模型有效训练的问题,本文提出了一种基于改进原型网络的变压器故障诊断方法。首先,采用残差多层感知机(residual multi-layer perceptron, ResMLP)提取油色谱数据的深层判别特征;然后,利用任务条件注意力机制(task-conditioned attention, TCA)实现支持集在不同任务下的动态调节,增强模型的跨任务适应性;同时,引入样本注意力机制(sample attention, SA)以突出关键支持样本对原型表示的主导作用,从而提升类内判别性,增强模型在微弱或复杂故障征兆下的识别能力;最后,通过样本与类别原型的度量实现故障分类。实验结果表明,本文方法通过任务与样本双层注意力增强原型构建,弥补了原型网络在跨任务适应性和类内判别性方面的不足,有效提高了小样本条件下变压器故障诊断准确率。

1 相关原理

1.1 特征提取

多层感知机(multi-layer perceptron, MLP)是一种前馈神经网络,能够有效地从原始气体浓度数据中学习反映设备内部故障状态的复杂特征模式。为了增强其在原型网络中的特征提取能力,提出了一种基于残差结构的多层感知机编码器,该结构通过堆叠多个残差单元,能有效缓解深层模型的梯度衰减问题,并提升训练稳定性,具体结构如图1所示。

首先,将原始输入特征 x 映射到隐藏维度空间,并使用ReLU激活函数进行处理,得到第1层输出如式(1)所示。

$$z_0 = \text{ReLU}(W_0 x + b_0) \quad (1)$$

式中: W_0 为线性变换的权重; b_0 为线性变换的偏置。

然后,在网络的中间层堆叠残差块(residual block)。每个残差块由一个MLP子网络与残差连接构成,残差块内部包含一个非线性映射 F_i ,并使用短路连接将输入与映射结果相加形成输出,该结构能够有效地引导信息跨层传递,提升深层MLP的表达能力,具体计算公式如式(2)所示。

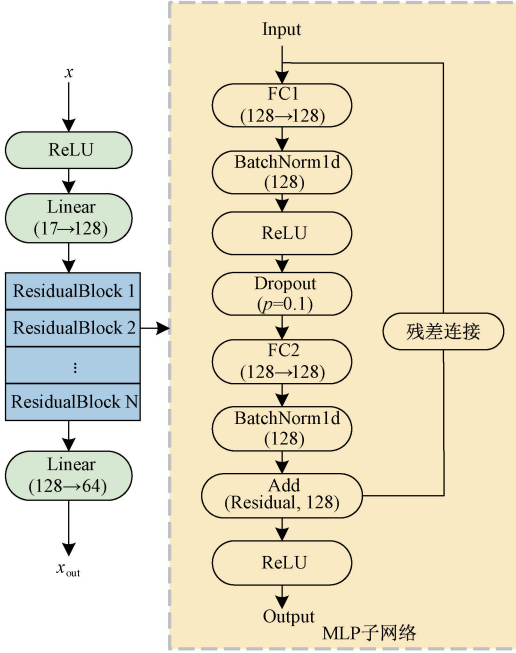


图 1 ResMLP 结构

Fig. 1 ResMLP architecture

$$\begin{cases} z_{i+1} = \text{ReLU}(F_i(z_i) + z_i) \\ i = 0, 1, \dots, N-1 \end{cases} \quad (2)$$

式中: $F_i(\cdot)$ 由 1 层 BatchNorm、1 层 ReLU 以及 1 层线性变换构成。

最后,使用 1 个线性映射将堆叠后的特征 z_N 投影到目标嵌入维度,并再次使用 ReLU 激活以获得最终的嵌入表示如式(3)所示。

$$x_{\text{out}} = \text{ReLU}(W_{\text{out}} z_N + b_{\text{out}}) \quad (3)$$

式中: W_{out} 为输出层权重; b_{out} 为输出层偏置。

1.2 原型网络

原型网络是一种基于度量学习的元学习方法。其核心思想是通过在度量空间中构建各类别原型表示,并衡量查询样本与各类原型之间的距离,将查询样本划归距离最小的原型对应类别来完成分类^[21]。原型网络结构如图 2 所示,与传统深度学习不同,原型网络将训练数据集划分为支持集 D^S 与查询集 D^Q 用于模型训练,类原型为每类信号样本的平均特征向量,第 k 类信号的原型可表示为:

$$c_k = \frac{1}{|S_k|} \sum_{x_i \in S_k} f_{\varphi}(x_i) \quad (4)$$

式中: f_{φ} 为特征提取网络; φ 为网络参数; S_k 为第 k 类支持样本集合; x_i 为第 k 类第 i 个样本。

假设 x 为查询集 D^Q 中的待识别样本,通过计算样本与各类原型间的欧氏距离 d ,距离公式如式(5)所示。

$$\text{dist}[f_{\varphi}(x_i), c_k] = \sqrt{[f_{\varphi}(x_i) - c_k]^2} \quad (5)$$

式中: c_k 为每个类原型点。

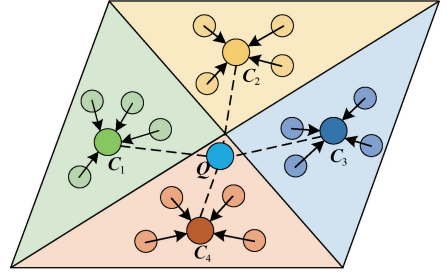


图 2 原型网络

Fig. 2 Prototypical Networks

将 d 取负,通过 Softmax 层得到各类别的预测概率,概率最大的类作为网络输出结果,样本 x 属于第 k 类样本的概率计算公式如式(6)所示。

$$P_{\varphi}(y = k | x) = \frac{\exp(-d(f_{\varphi}(x), c_k))}{\sum_{k'} \exp(-d(f_{\varphi}(x), c_{k'}))} \quad (6)$$

训练过程中将 $\log$ 概率和真实类别标签 y_i 计算负对数似然损失,具体计算公式如式(7)所示。

$$\mathcal{L} = -\frac{1}{N \cdot Q} \sum_{i=1}^{N \cdot Q} \log P(y_i | x_i) \quad (7)$$

式中: N 诊断为类别数; Q 为每类的查询样本数; $P(y_i | x_i)$ 是模型对第 i 个查询样本预测为其真实类别的概率。

1.3 任务条件注意力机制

在小样本场景中,不同任务的类别空间和数据分布通常存在显著差异。传统的原型网络仅通过支持集构造类中心,忽略了当前查询样本对任务的整体语义描述。为了增强支持集嵌入对当前任务的适应性,引入了任务条件注意力模块^[22]。该模块通过查询集的全局特征对支持集表示进行条件建模,动态调整支持集表示,从而提升原型的判别性和任务特异性。

首先,从查询样本中获取任务嵌入 t ,将其作为该任务的整体语义特征,其计算公式如式(8)所示。

$$t = \frac{1}{q} \sum_{i=1}^q x_i^q \quad (8)$$

式中: x_i^q 为第 i 个查询样本的嵌入; q 为每个任务中的查询样本数。

然后,将该任务向量通过非线性变换后复制 k 次,与支持集嵌入 S 一起输入多头注意力模块,具体计算公式如式(9)所示。

$$\begin{cases} \tilde{T} = \text{ReLU}(W_t t), \quad \tilde{T} = \text{repeat}(\tilde{T}, k) \\ S' = \text{MultiHead}(Q = \tilde{T}, K = S, V = S) \end{cases} \quad (9)$$

式中: W_t 为可学习的线性变换矩阵; $\tilde{T}$ 为变换后的任务表示; k 为每个类别中支持样本的数量; S 为该类型支持样本的嵌入集合; MultiHead 为多头注意力机制。

最后,将注意力结果与原始支持相加并输入 Transformer 编码器进一步建模局部依赖性,形成任务条件支持如式(10)所示。

$$\hat{S} = \text{TransformerEncoder}(S + S') \quad (10)$$

1.4 样本注意力机制

1) 支持样本注意力

支持样本注意力^[23]用于为支持集中的每个样本分配权重,其作用是自动识别并强调关键样本,从而得到更具代表性的原型,以提高小样本分类的区分能力。

由式(4)可知,在传统原型网络中,每个类的原型是该类支持集中所有样本的平均,但是简单的平均计算会存在两个问题:1)可解释性差,所有样本被等权重对待,难以判断哪个支持样本对预测起关键作用;2)鲁棒性差,错误或离群点也会被平均,影响原型表示。因此本文引入支持样本注意力机制,对支持集中的样本进行加权,从而更好地构造原型,支持集中的样本特征不再是简单的平均,而是加权平均。在任务条件支持的基础上,计算样本注意力权重如式(11)所示。

$$\begin{cases} \alpha_j = \frac{\exp(g(\hat{x}_j^s, t))}{\sum_{x_j^s \in S_k} \exp(g(\hat{x}_j^s, t))} \\ \sum_{x_j^s \in S_k} \alpha_j = 1 \end{cases} \quad (11)$$

式中: $\hat{x}_j^s \in \hat{S}$; $g(\cdot)$ 为衡量支持样本与任务嵌入之间相关性函数。

最后,得出任务条件样本注意力原型计算公式如式(12)所示。

$$c_k = \sum_{x_j^s \in S_k} \alpha_j \cdot \hat{x}_j^s \quad (12)$$

2) 查询样本注意力

在传统的原型网络中,通常通过欧氏距离或余弦相似度来衡量查询样本与各类原型的相似程度,并通过 Softmax 转换为概率。然而该方法对所有查询样本采用统一策略,忽略了每个查询样本对不同类别的注意力偏好。

为此,本文引入查询样本注意力机制^[24],用于对查询集中的每个样本进行加权,强调更具判别性的查询特征,从而更好地匹配原型。与支持样本注意力不同,它在查询样

本层面上做注意力,但不聚合查询样本,通过对查询特征逐通道、逐样本重标定,从而增强模型对不同类别分布差异的适应能力。

基础原型网络的分类概率计算方法如式(13)所示。

$$P_\varphi(y = k | x) = \frac{\exp(-d(f_\varphi(x), c_k))}{\sum_{k'} \exp(-d(f_\varphi(x), c_{k'}))} \quad (13)$$

引入注意力机制,定义查询样本对第 k 类原型的注意力权重,具体计算公式如式(14)所示。

$$\begin{cases} \alpha_k(x) = \frac{\exp(-d(f_\varphi(x), c_k))}{\sum_{k'} \exp(-d(f_\varphi(x), c_{k'}))} \\ \sum_k \alpha_k(x) = 1 \end{cases} \quad (14)$$

式中: d 为各类原型间的欧氏距离。

最后,得到引入查询注意力的分类概率计算公式,如式(15)所示。

$$P_{\text{att}}(y = k | x) = \frac{\alpha_k(x) \cdot \exp(-d(f_\varphi(x), c_k))}{\sum_{k'} \alpha_{k'}(x) \cdot \exp(-d(f_\varphi(x), c_{k'}))} \quad (15)$$

该形式在计算分类概率时,不再仅依赖距离,而是综合考虑了每个查询样本自身对各类别原型的关注度。

2 基于改进原型网络的故障诊断方法

2.1 故障诊断模型

原型网络通过判断查询样本与各类原型之间的相似度来判定类别,但在小样本场景下,支持集样本稀少且易受异常样本干扰,导致原型表示准确性和类内判别性不足。为了解决上述问题,本文对原型网络的特征提取模块进行了改进,并在类原型构建模块中引入了任务条件注意力和样本注意力机制,根据任务的上下文信息动态调整各类别特征的关注重点,衡量支持集中各样本对类原型构成的重要性,有效抑制异常样本的干扰。

改进原型网络变压器故障诊断模型如图3所示,模型主要包括特征提取模块、注意力增强原型构建模块与距离度量模块。在特征提取模块中,使用 ResMLP 来提取油色

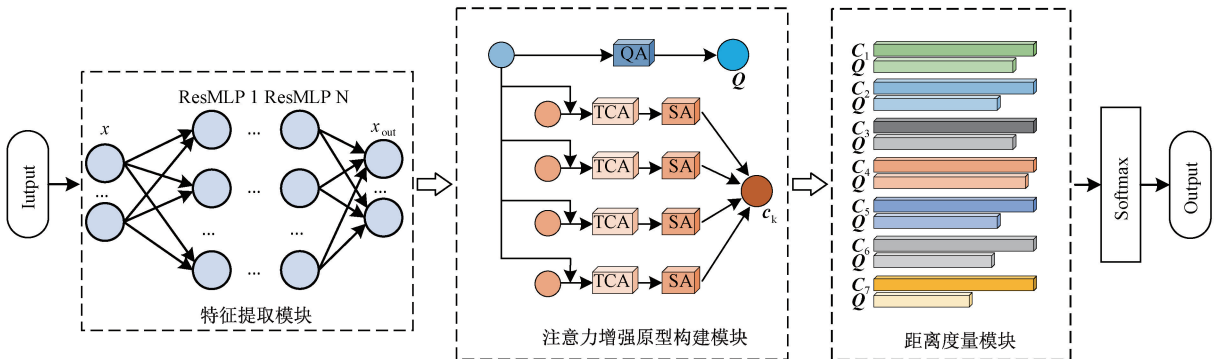


图3 改进原型网络变压器故障诊断模型

Fig. 3 Improved prototypical networks for transformer fault diagnosis

谱样本深层特征;在注意力增强原型构建模块中,使用任务与样本双层注意力增强原型构建,获取具有更强判别能力的支持样本原型;在距离度量模块中,计算查询样本与类原型之间的欧氏距离,精准判断查询集样本所属类别。

2.2 故障诊断流程

本文提出的故障诊断方法整体流程如图 4 所示,具体步骤如下:

- 1)收集并整理变压器油色谱数据样本。
- 2)将数据样本划分为训练集和测试集,训练集和测试集中各自包含支持集和查询集,对样本数据进行批归一化。
- 3)构建基于改进原型网络的变压器故障诊断模型,模型主要包括特征提取模块、注意力增强原型构建模块以及距离度量模块。
- 4)在训练阶段,将训练集输入改进原型网络模型,通过负对数似然损失进行优化,并利用反向传播更新参数;同时引入梯度裁剪、L2 正则化的 AdamW 优化器和余弦退火学习率策略以提升训练稳定性。
- 5)在测试阶段,将查询集样本输入训练完成的网络模型,输出预测的故障类型。

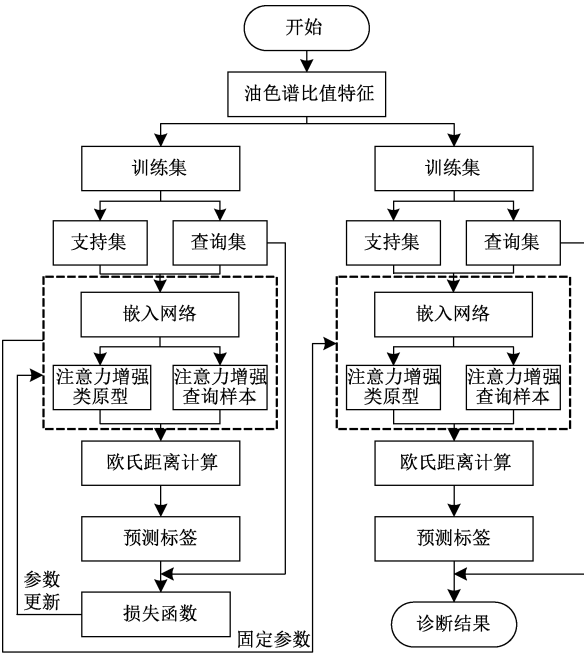


图 4 故障诊断流程

Fig. 4 Fault diagnosis flowchart

3 实验结果与对比分析

3.1 实验数据集

本文采用文献[25-26]收集的电力变压器故障 DGA 数据进行实验分析。根据故障机理,变压器故障可分为热性故障与电性故障两大类,其中热性故障细分为低温过热、中温过热与高温过热,电性故障包括低能放电、高能放

电和局部放电。将上述 6 种故障结合正常运行状态,共形成 7 类运行状态,编号为 0~6,具体数据样本分布情况及编号如表 1 所示。

表 1 运行状态编码及样本分布

Table 1 Operating condition codes and sample distribution

运行状态	状态编码	训练集样本数量	验证集样本数量
正常	0	144	36
低温过热	1	144	36
中温过热	2	144	36
高温过热	3	144	36
局部放电	4	144	36
低能放电	5	144	36
高能放电	6	144	36

7 类运行状态共 1 260 个样本,每一类别样本共 180 个,其中训练集为 144 个,验证集为 36 个。为提高模型训练与评估的可靠性,在保持上述数据划分形式的情况下,采用分层 5 折交叉验证对模型的性能进行验证,以减少单次划分带来的偶然性并反映结果稳定性。

3.2 特征选择

传统人工智能诊断方法通常直接将氢气(H₂)、甲烷(CH₄)、乙炔(C₂H₂)、乙烯(C₂H₄)、乙烷(C₂H₆)5 类特征气体的浓度作为模型输入特征。然而,仅依赖上述气体的浓度信息可能无法全面反映故障信号,易导致模型在训练时陷入局部最优解,进而降低诊断性能。已有研究表明,油中溶解气体的浓度比值与变压器运行状态之间存在更紧密的关联性^[27]。原型网络是一种度量学习方法,其度量机制强调类间相对分布,对故障气体的原始浓度数据较为敏感。因此,在保留 5 种原始气体浓度数据特征的基础上,借鉴气体比值特征构建方法^[28],新增 12 种气体浓度比值特征,得到 17 种故障特征如表 2 所示。

表 2 变压器故障特征集

Table 2 Transformer fault diagnosis feature set

编号	故障特征	编号	故障特征
S1	H ₂	S10	C ₂ H ₂ /C ₂ H ₄
S2	CH ₄	S11	C ₂ H ₂ /C ₂ H ₄ /0. 1
S3	C ₂ H ₂	S12	C ₂ H ₄ /TH
S4	C ₂ H ₄	S13	C ₂ H ₂ /THD/0. 21
S5	C ₂ H ₆	S14	H ₂ /EWTHD
S6	C ₂ H ₄ /C ₂ H ₆	S15	H ₂ /THH
S7	CH ₄ /H ₂	S16	C ₂ H ₄ /EWTH
S8	C ₂ H ₂ /EWTHD	S17	C ₂ H ₄ /EWTHH
S9	C ₂ H ₂ /THD		

表 2 中,S1~S5 为 5 种关键气体原始浓度特征,S6~

SI7 为 12 种浓度比值特征,其中:

THH=C₂H₂+C₂H₄+H₂;

THD=C₂H₂+C₂H₄+CH₄;

TCH=CH₄+C₂H₄+C₂H₆+C₂H₂;

TH=H₂+CH₄+C₂H₄+C₂H₆+C₂H₂;

EWTHD=CH₄+2.64C₂H₄+7.56C₂H₂;

EWTHH=2.76H₂+2.64C₂H₄+7.56C₂H₂;

EWTCCH=CH₄+1.63C₂H₆+2.64C₂H₄+7.56C₂H₂;

EWTH=2.76H₂+CH₄+1.63C₂H₆+2.64C₂H₄+7.56C₂H₂.

3.3 支持集样本量对识别性能影响

在小样本学习任务中,常用 N -way、 K -shot 的方式来描述任务设定,其中 N -way 表示需要分类的 N 个类别数, K -shot 表示每个类别使用 K 个样本来用于模型训练或原型构建。在工业故障诊断任务中将每类训练样本 ≤ 10 视为典型的 few-shot 设置^[29]。因此,为验证本文模型在小样本情况下的诊断性能,根据变压器的 7 种运行状态,将样本类别数设定为 7-way,支持样本采用 1-shot、5-shot、10-shot 和 20-shot 的验证模式,模拟真实场景中极少样本、标准小样本、中等小样本和有限样本的情况。不同 K -shot 条件下本文模型诊断准确率如图 5 所示。

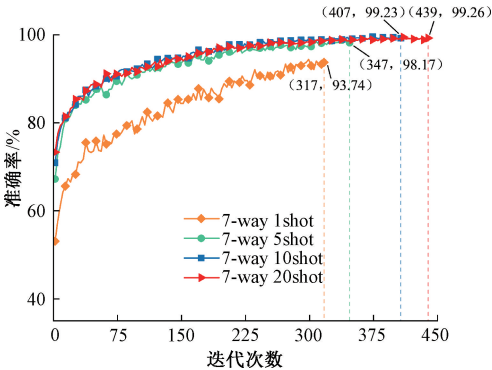


图 5 不同 K -shot 下模型准确率

Fig. 5 Accuracy of the proposed model under different K -shot

由图 5 可知,模型在所有设置下准确率均能逐步趋于收敛;支持样本数量从 1-shot 提升至 5-shot 时候,模型性能提升最为显著,说明在极少样本情况下,少量额外支持样本可以显著改善类原型构建;而从 5-shot 至 20-shot 的增益呈现递减趋势,表明使用每类 5 个支持样本就能获得接近最优的性能,若继续增加样本数量则收益有限。

为评估改进原型网络在不同 K -shot 条件下的诊断能力,将其与原型网络进行了对比,进行分层 5 折交叉验证,每折进行一次完整训练、测试流程,验证结果如图 6 所示。

由图 6 可知,改进原型网络在样本量极少的情况下仍有很好的诊断效果,而原型网络随着支持样本数量的减少,诊断准确率明显下降;与原型网络相比,改进原型网络在 1-shot、5-shot、10-shot 和 20-shot 的验证模式下,准确率

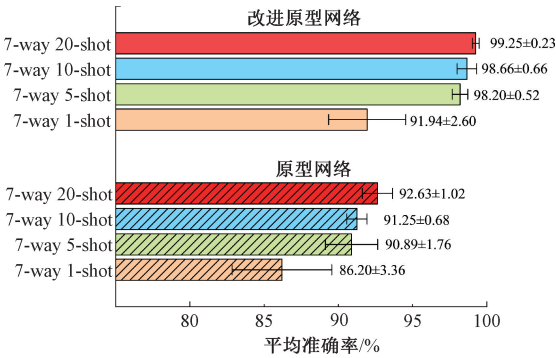


图 6 改进前后平均准确率对比

Fig. 6 Comparison of average accuracy before and after improvement

分别提高了 5.74%、7.31%、7.41% 和 6.62%,说明改进原型网络能够有效提高类原型的判别能力,提升了模型在小样本条件下的鲁棒性与适应性。

3.4 消融实验

为验证模型改进的有效性,在 few-shot 标准小样本 5-shot 的验证模式下,进行消融实验分析,保证较高诊断准确率的同时具备更快的训练收敛速度和更低的计算开销。实验组合设置如下:1)原型网络;2)将原始嵌入网络替换为 ResMLP 进行特征提取;3)在原型网络中加入样本注意力机制;4)在原型网络中加入任务条件多头注意力机制;5)本文方法。分层 5 折交叉验证结果如表 3 所示。

表 3 消融实验结果

Table 3 Ablation study results

组合 序号	模型 Protonet	组件 ResMLP SA TCA			评估指标 平均准确率/%
1	✓				90.89±1.76
2	✓	✓			94.59±1.82
3	✓		✓		92.81±1.37
4	✓			✓	97.02±0.41
5	✓	✓	✓	✓	98.20±0.52

由表 3 可知,使用原型网络进行故障诊断的准确率最低,为 90.89%,添加任一组件均能提升诊断准确率;组合 1 使用 ResMLP 特征提取模块替代传统 MLP,诊断准确率提升了 3.70%,表明引入残差连接有助于减轻深层网络训练时的梯度衰减现象,能够通过更深层的结构有效捕捉更复杂的特征;组合 3 中引入 SA 进行样本加权,使诊断准确率提升了 1.90%,SA 能够有效提高类原型的质量与稳定性,进一步细化匹配过程,进而提高查询样本和类原型的匹配精度;组合 4 引入 TCA,使诊断准确率提升了 6.13%,TCA 能够根据当前任务的类别上下文进行自适应调整注意力分布,使原型网络能够有效区分支持样本和查询集之间的细微差异。

为进一步分析不同组件对诊断性能的影响,通过混淆

矩阵对诊断结果进行可视化展示,如图 7 所示。由图 7(c)、(d)可知,在组合 3、4 中不同类型故障的诊断准确率均有明显提升,但仍存在较多将故障样本误判为正常的情况;由图 7(b)可知,在组合 2 中 ResMLP 的加入有助于在嵌入空间中提升各类别间的可分性,从而有利于形成区分度更

高的类原型并显著降低类别间混淆与错误分类;3 种组件配合使用,将故障诊断准确率提升至 98.20%。综上所述,通过 5 组实验验证了本文方法中各个模块对变压器故障诊断性能提升起到重要作用,验证了模型设计的合理性与有效性。

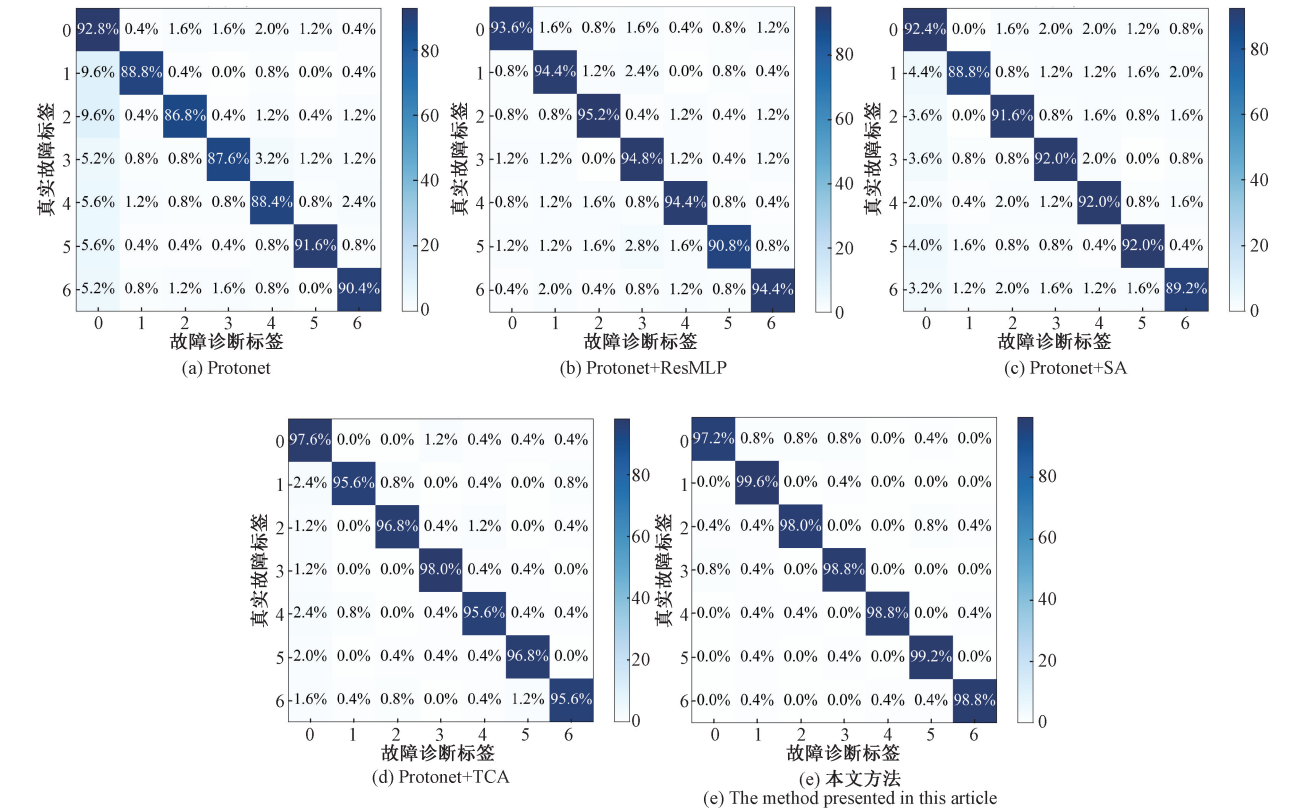


图 7 诊断结果混淆矩阵

Fig. 7 Confusion matrix of diagnostic results

3.5 泛化性实验

为验证本文模型的泛化性能,选取 IEEE Dataport^[30]和 IEC TC 10 Dataport^[31]两种公开数据集中,与变压器油色谱相关的气体特征及其对应运行状态标签进行故障诊断分析。其中,IEEE Dataport 共包含 182 条故障样本,IEC TC 10 Dataport 共包含 89 条故障样本,具体数据分布如表 4 所示。

以本文所用训练集训练模型,采用 IEC 三比值法、ITR 改进三比值法、随机森林(random forest, RF)、反向传播神经网络(backpropagation neural network, BPNN)、PN、匹配网络(matching networks, MN)、关系网络(relation networks, RN)和本文方法对两个公开数据集进行故障诊断验证。其中,RF 的决策树数量为 50,最小叶子数为 1;BPNN 学习率设置为 0.01,迭代次数为 200 轮;MN 和 RN 两种元学习方法采用与本文模型相同的 7way 5shot 设置。诊断结果如表 5 所示。

由表 5 可知,本文提出的故障诊断方法在两个公开数

表 4 公开数据集运行状态编码及样本分布

Table 4 Operating condition codes and sample distribution of the public datasets

运行状态	状态编码	IEEE	IEC TC 10
正常	0	0	0
低温过热	1	18	0
中温过热	2	8	0
高温过热	3	38	16
局部放电	4	20	4
低能放电	5	48	24
高能放电	6	50	45

据集上的故障诊断准确率最高,分别为 82.87% 和 83.40%。与 IEC、ITR、RF、BPNN、PN、MN 和 RN 相比,本文模型在 IEEE 数据集上的诊断准确率分别提升了 31.77%、26.28%、19.74%、23.41%、15.14%、21.49% 和 18.68%;在 IEC TC 10 数据集上的诊断准确率分别提升

表 5 不同模型在公开数据集上的诊断准确率对比

Table 5 Comparison of diagnostic accuracy among different models on public datasets (%)		
模型	IEEE	IEC TC 10
IEC	51.10	67.42
ITR	56.59	69.66
RF	63.13	66.92
BPNN	59.46	74.16
PN	67.73	76.97
MN	61.38	76.50
RN	64.19	74.75
本文方法	82.87	83.40

了 15.98%、13.74%、16.48%、9.24%、6.43%、6.90% 和 8.65%。

不同模型在公开数据集上运行的结果表明,IEC、ITR 等方法依赖固定阈值和经验规则,难以准确反映气体浓度间的复杂关系;RF 虽稳健,但依赖手工特征,易受无关或高度相关特征与噪声干扰;BPNN 表示能力不足且易过拟合;PN、MN 和 RN 等模型在处理高维复杂特征关系时能力有限,对噪声和分布变化敏感,限制了泛化能力,从而导致这些方法在公开数据集上诊断准确率低,而本文方法在相同数据集上取得了更高的准确率,验证了方法的可行性和泛化能力。

3.6 变压器故障案例分析

为进一步验证本文方法的有效性和泛化性,对国家电网两起实际故障案例进行诊断分析。

1)案例 1

采用国家电网某供电公司 110 kV 变压器油色谱异常数据进行故障诊断分析^[32],具体数据如表 6 所示。

表 6 案例 1 变压器油色谱数据

Table 6 Case 1 dissolved gas analysis data (μL/L)					
试验日期	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂
2023-02-09	520.34	1 103.65	152.36	403.42	1.78
2023-03-10	946.67	1 237.61	779.60	2 685.12	2.37

该变压器型号为 SSZ11-50000/110,投运日期为 2018 年,故障发生后,经检查变压器关键部位确定为由于变压器铁心引线过长与硅钢片接触,导致引线绝缘层被硅钢片划破,铁心与引线接触发生多点接地,引起高温过热故障,引线破损烧蚀部位如图 8 所示。

采用本文提出的故障诊断方法对油色谱样本数据进行诊断分析,并计算异常样本与不同类原型间的欧氏距离如图 9 所示。

由图 9 可知,异常样本与各故障类原型之间的欧氏距离存在明显差异,其中红色柱表示欧氏距离最小的故障类原型。状态编码 3 对应的高温过热故障类原型与异常样

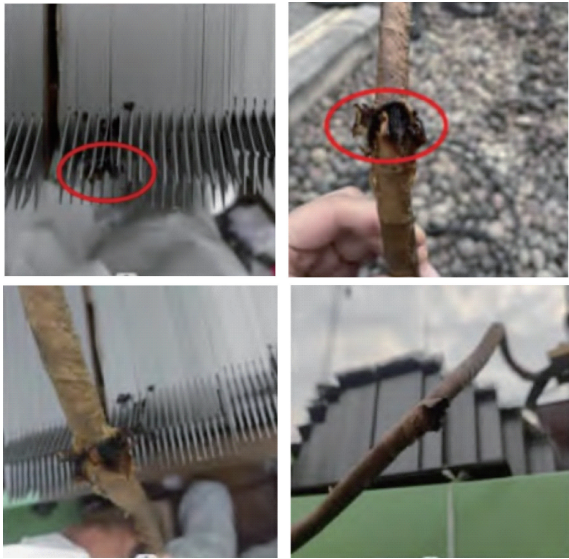


图 8 引线破损烧蚀部位

Fig. 8 Damaged and ablated area of the transformer lead

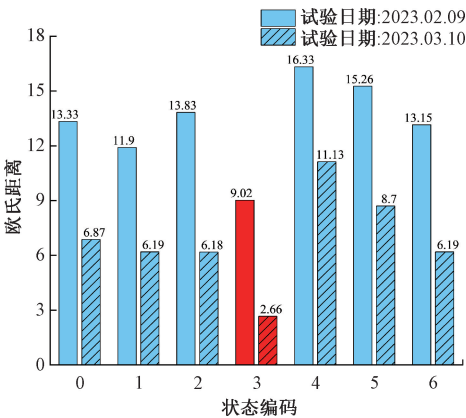


图 9 案例 1 样本与不同故障类原型间的欧氏距离

Fig. 9 Comparison of Euclidean distances from the case 1 sample to different fault class prototypes

本距离最近,因此诊断结果为高温过热故障。在本案例中,本文方法诊断结果与变压器实际故障类型一致。

2)案例 2

采用国家电网某供电公司 110 kV 变压器油色谱异常数据进行故障诊断分析^[33],具体数据如表 7 所示。

表 7 案例 2 变压器油色谱数据

Table 7 Case 2 dissolved gas analysis data (μL/L)					
取油位置	H ₂	CH ₄	C ₂ H ₆	C ₂ H ₄	C ₂ H ₂
上部油门	634	139	11	136	194
下部油门	600	139	10.7	138	196

该变压器型号为 SFZ11-31500/110,出厂日期为 2007 年 8 月,2014 年 6 月变压器重瓦斯保护动作,变压器在运烧毁。检查发现低压 a 相绕组存在明显内陷变形,绝缘受

损,顶部线匝和铁心之间发生了高能量放电导致线匝和铁心硅钢片融化,绕组变形部位如图 10 所示。



图 10 绕组变形部位

Fig. 10 Deformed section of the winding

采用本文提出的故障诊断方法对油色谱样本数据进行诊断分析,并计算异常样本与各类别原型之间的欧氏距离如图 11 所示。其中红色柱表示欧氏距离最小的故障类原型。

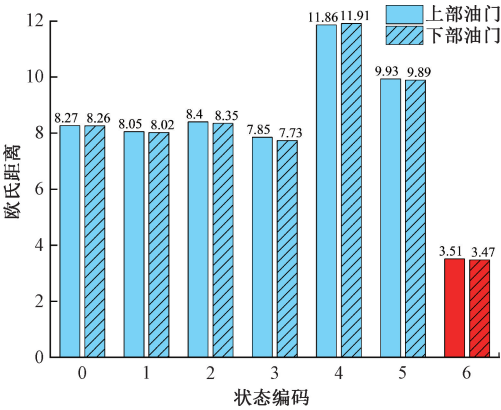


图 11 案例 2 样本与不同故障类原型间的欧氏距离

Fig. 11 Comparison of Euclidean distances from the case 2 sample to different fault class prototypes

由图 11 可知,该变压器上、下部油位油色谱样本均与状态编码 6 高能放电故障类原型的欧氏距离最近,诊断为高能放电故障,与实际情况结果一致。

3)对比分析

使用不同方法对变压器故障案例进行诊断,诊断结果如表 8 所示。

表 8 不同模型对案例的诊断结果

Table 8 Diagnostic results of different models on the cases

诊断方法	案例 1		案例 2	
	2022-12-17	2023-02-09	上部油门	下部油门
MN	1	2	5	5
RN	5	3	4	6
PN	3	2	6	3
本文方法	3	3	6	6
现场分析	3	3	6	6

由表 8 可知,在两个实际案例诊断分析中,除本文方法外,MN、RN 和 PN 均存在误判,本文方法能够准确识别不同类型的故障状态,并且诊断结果与现场检查结论完全一致。这表明本文方法不仅在实验数据集上有效,在真实工程场景中同样具有较强的适应性和可靠性,验证了其有效性和泛化性。

4 结 论

本文针对极少量带标签样本条件下的变压器故障诊断难题,提出一种基于改进原型网络的变压器故障诊断方法,结合实验分析得到以下结论。

在特征提取阶段,本文将残差结构引入 MLP,能够更深层地提取 DGA 数据特征,提高了模型的鲁棒性;在原型构建阶段,提出了任务条件注意力与样本注意力相结合的增强机制,有效解决了原型网络在跨任务适应性和类内判别性方面的不足,构建更具代表性的类原型,实现查询样本的精准匹配。在 IEEE Dataport 与 IEC TC 10 Dataport 两个公开数据集上的实验表明,本文方法较基础原型网络故障诊断准确率提升了 15.14% 和 6.43%,验证了本文方法解决小样本条件下变压器故障诊断问题可行性和泛化能力。

本文提出的改进原型网络在小样本条件下性能显著提升,其主要是针对原型构建进行了优化,生成更加判别性的类原型。然而,在查询样本与类原型的匹配过程中,仅使用欧氏距离来衡量相似度,缺乏可学习的相似度判别器以捕捉更复杂的非线性相似性。未来工作可考虑引入可学习的相似度判别模块,以捕捉更丰富的非线性关系,从而进一步提高鲁棒性与泛化性能。

参考文献

[1] 邢致恺,何怡刚,姚其新. 基于多模态信息融合的变压器在线故障诊断方法[J]. 电子测量与仪器学报, 2024, 38(9): 96.
Xing Zhikai, He Yigang, Yao Qixin. Transformer online fault diagnosis method based on multi-modal information fusion[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 96.
[2] 朱朋,邓蕾,汤宝平,等. 数字孪生驱动小波注意力迁移网络的齿轮箱故障诊断方法[J]. 仪器仪表学报, 2025, 46(10): 86-95.
Zhu Peng, Deng Lei, Yang Baoping, et al. Digital twin driven wavelet attention transfer network for gearbox fault diagnosis [J]. Chinese Journal of Scientific Instrument, 2025, 46(10): 86-95.
[3] 倪炳阳,何青. 基于 CNN-BiLSTM-Attention 的风电机组故障预警[J]. 电子测量技术, 2025, 48(11): 79.
Ni Bingyang, He Qing. Early fault warning of wind turbines based on CNN-BiLSTM-Attention [J].

- Electronic Measurement Technology, 2025, 48(11): 79.
- [4] 冯治国, 金日, 罗冲, 等. 基于Transformer神经网络的变压器状态监测[J]. 国外电子测量技术, 2023, 42(2): 146.
- Feng Zhiguo, Jin Ri, Luo Chong, et al. Power transformer state monitoring based on Transformer deep neural network[J]. Foreign Electronic Measurement Technology, 2023, 42(2): 146.
- [5] 陈仁祥, 张旭, 徐向阳, 等. 噪声标签下注意力特征混合的旋转机械故障诊断[J]. 仪器仪表学报, 2023, 44(9): 257.
- Chen Renxiang, Zhang Xu, Xu Xiangyang, et al. Fault diagnosis of rotating machinery with attentive feature mixup in noisy labels[J]. Chinese Journal of Scientific Instrument, 2023, 44(9): 257.
- [6] 汤茂祥, 王聪, 朱超平, 等. 少标签下油浸式变压器双层故障诊断模型[J]. 电子测量技术, 2023, 46(16): 113.
- Tang Maoxiang, Wang Cong, Zhu Chaoping, et al. Transformer fault diagnosis method based on GBDT and K-means gain clustering[J]. Electronic Measurement Technology, 2023, 46(16): 113.
- [7] Ding Y, Ma L, Ma J, et al. A generative adversarial network-based intelligent fault diagnosis method for rotating machinery under small sample size conditions [J]. IEEE Access, 2019, 7: 149736.
- [8] 王欢, 王昕, 张峰, 等. 基于改进生成对抗网络的变压器声纹故障诊断[J]. 智慧电力, 2024, 52(4): 24.
- Wang Huan, Wang Xin, Zhang Feng, et al. Transformer voiceprint fault diagnosis based on improved generative adversarial network[J]. Smart Power, 2024, 52(4): 24.
- [9] Li W Y, Nie Y X, Yang F. Multi-variable transformer-based meta-learning for few-shot fault diagnosis of large-scale systems[J]. Sensors, 2025, 25(9): 2941.
- [10] 刘俊, 秦晓瑞, 陶剑, 等. 面向小样本文本分类的互学习原型网络[J]. 北京邮电大学学报, 2024, 47(3): 31.
- Liu Jun, Qin Xiaorui, Tao Jian, et al. Mutual learning prototype network for few-shot text classification[J]. Journal of Beijing University of Posts and Telecommunications, 2024, 47(3): 31.
- [11] 李庆旭, 王巧华. 基于原型网络的小样本禽蛋图像特征检测方法[J]. 农业机械学报, 2021, 52(11): 377.
- Li Qingxu, Wang Qiaohua. Feature detection method of small sample poultry egg image based on prototypical network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(11): 377.
- [12] He Q C, Li S B, Li C J, et al. A hybrid matching network for fault diagnosis under different working conditions with limited data [J]. Computational Intelligence and Neuroscience, 2022, 2022: 3024590.
- [13] Li X, Jiang A P, Zheng X Q, et al. Few-shot fault diagnosis based on an attention-weighted relation network[J]. Entropy, 2023, 26(1): 22.
- [14] Tang T, Qiu C H, Yang T Y, et al. A novel lightweight relation network for cross-domain few-shot fault diagnosis[J]. Measurement, 2023, 213: 112697.
- [15] Chen Z Y, Wang Y H, Wu J, et al. Wide residual relation network-based intelligent fault diagnosis of rotating machines with small samples [J]. Sensors, 2022, 22(11): 4161.
- [16] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning [C]//Advances in Neural Information Processing Systems 30 (NeurIPS 2017), 2017: 4078
- [17] Wang Y, Chen L, Liu Y, et al. Wavelet-prototypical network based on fusion of time and frequency domain for fault diagnosis[J]. Sensors, 2021, 21(4): 1483.
- [18] 余浩帅, 汤宝平, 张楷, 等. 小样本下混合自注意力原型网络的风电齿轮箱故障诊断方法[J]. 中国机械工程, 2021, 32(20): 2475.
- Yu Haoshuai, Tang Baoping, Zhang Kai, et al. Fault diagnosis method of wind turbine gearboxes mixed with attention prototype networks under small samples [J]. China Mechanical Engineering, 2021, 32(20): 2475.
- [19] Wang Z L, Shen H H, Xiong W Z, et al. Method for diagnosing bearing faults in electromechanical equipment based on improved prototypical networks[J]. Sensors, 2023, 23(9): 4485.
- [20] Li K, Shang C, Ye H. Reweighted regularized prototypical network for few-shot fault diagnosis[J]. IEEE Trans Neural Netw Learn Syst, 2024, 35(5): 6206.
- [21] 赵志宏, 张然, 刘克俭, 等. 一种改进原型网络的小样本轴承故障诊断方法[J]. 振动与冲击, 2023, 42(20): 215.
- Zhao Zhihong, Zhang Ran, Liu Kejian, et al. An improved prototype network method for small sample bearing fault diagnosis[J]. Journal of Vibration and Shock, 2023, 42(20): 215.
- [22] Oreshkin B, Rodriguezlopez P, Lacoste A. TADAM: Task dependent adaptive metric for improved few-shot learning [C]//Advances in Neural Information Processing Systems 31 (NeurIPS), 2018: 719.
- [23] Wang Y K, Xu C M, Liu C, et al. Instance credibility

- inference for few-shot learning[C]//Advances in Neural Information Processing Systems (NeurIPS), 2020: 22068.
- [24] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[C]//Advances in Neural Information Processing Systems (NeurIPS), 2016: 3630.
- [25] Sami S M, Bhuiyan M I H. Power transformer fault diagnosis with intrinsic time-scale decomposition and XGBoost classifier[J]. Proceedings of the International Conference on Big Data, IoT, and Machine Learning, 2021,95:527-537.
- [26] 何志洋. 油浸式变压器故障特征研究与分析[D]. 乌鲁木齐: 新疆大学, 2023: 30.
He Zhiyang. Analysis and research on fault characteristics of oil-immersed transformer [D]. Urumqi: Xinjiang University, 2023: 30.
- [27] Zhang Y, Chen H C, Du Y P, et al. Power transformer faults diagnosis considering data imbalance and data set fusion[J]. High Voltage, 2021, 6(3): 543-554.
- [28] He Z Y, Kari T, Yilihamu Y, et al. Hybrid feature selection approach for power transformer fault diagnosis based on whale optimization algorithm and extreme learning machine[J]. IEEE Access, 2021, 9: 135757.
- [29] Deng W H, Xiong W, Lu Z Y, et al. Few-shot power transformers fault diagnosis based on gaussian prototype network [J]. International Journal of Electrical Power & Energy Systems, 2024, 160:110146.
- [30] Li E W, Wang L N, Song B. Fault diagnosis of power transformers with membership degree [J]. IEEE Access, 2019, 7:28791-28798.
- [31] Duval M, De Pabla A. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases [J]. IEEE Electrical Insulation Magazine, 2001, 17(2): 31-41.
- [32] 胡露, 舒鸿科, 肖懿, 等. 一起 110 kV 主变铁心多点特殊接地故障诊断消缺的典型案例分析[J]. 变压器, 2025, 62(1): 67-71.
Hu Lu, Shu Hongke, Xiao Yi, et al. Analysis of typical case of diagnosis and elimination of multiple special grounding faults in core of 110 kV main transformer[J]. Transformer, 2025, 62(1): 67-71.
- [33] 张冰倩, 咸日常, 朱锋, 等. 一起 110 kV 变压器绕组变形故障的案例分析[J]. 变压器, 2021, 58(2): 70-73.
Zhang Bingqian, Xian Richang, Zhu Feng, et al. Case analysis of 110 kV transformer winding deformation fault[J]. Transformer, 2021, 58(2): 70-73.

作者简介

韦强宇, 硕士研究生, 主要研究方向为变压器故障诊断。

E-mail: 852515570@qq.com

吐松江·卡日(通信作者), 博士, 教授, 博士生导师, 主要研究方向为电力系统数字化、新能源发电与并网和人工智能与模式识别等。

E-mail: minyun229@163.com