

基于可泛化图知识蒸馏的油浸式电力变压器故障检测

张瑞佳¹, 马慧芳¹, 张映月¹, 彭生江²

(1. 西北师范大学计算机科学与工程学院, 甘肃 兰州 730070; 2. 国网甘肃省电力公司武威供电公司, 甘肃 武威 733000)

摘要: 溶解气体分析(DGA)旨在通过监测绝缘油中的溶解气体来识别潜在的故障类型。然而, 现有 DGA 方法受到有限标记数据制约导致性能不佳。为此, 提出一种新的图知识蒸馏方法(GKDG), 旨在提高 DGA 的准确性和效率。采用双视角图构建策略从样本邻域中获得额外的监督, 通过传播直接从其他样本中聚合信息。进一步地, 将教师图神经网络(GNN)中的知识蒸馏到学生 GNN 模型中, 确保学生模型能够有效地捕捉并解释溶解气体之间的复杂关系。此外, 为了对齐嵌入空间中的学生图和教师图, 引入多种知识, 从而增强学生模型的学习能力, 使其更好地学习教师模型。实验结果验证了 GKDG 在提升 DGA 性能方面的显著效果, 其能为电力设备的维护和故障检测提供有力支持。

关键词: 油浸式电力变压器; 故障检测; 溶解气体分析; 图神经网络; 知识蒸馏

源代码链接: <https://github.com/vv20327/GKDG>

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070580

Fault Detection in Oil-Immersed Power Transformers Based on Generalizable Graph Knowledge Distillation

ZHANG Ruijia¹, MA Huifang¹, ZHANG Yingyue¹, PENG Shengjiang²

(1. College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, Gansu, China;

2. State Grid Gansu Electric Power Company Wuwei Power Supply Company, Wuwei 733000, Gansu, China)

【Abstract】 Dissolved Gas Analysis (DGA) aims to identify potential fault types by monitoring the dissolved gases in insulating oil. However, existing DGA methods exhibit limited performance because of the constraints imposed by the scarcity of labeled data. To address this issue, a novel Graph Knowledge Distillation method (GKDG) is proposed to enhance the accuracy and efficiency of the DGA. It employs a dual-view graph construction strategy to obtain additional supervision from sample neighborhoods, aggregating information directly from other samples through propagation. Furthermore, knowledge from the teacher Graph Neural Network (GNN) is distilled into the student GNN model, ensuring that the student model can effectively capture and interpret the complex relationships among the dissolved gases. Additionally, to align the student and teacher graphs in the embedding space, multiple types of knowledge are introduced, thereby enhancing the learning capability of the student model and enabling it to learn better from the teacher model. The experimental results validate the effectiveness of the GKDG in improving the DGA performance, providing strong support for the maintenance and fault detection of power equipment.

【Key words】 oil-immersed power transformers; fault detection; Dissolved Gas Analysis (DGA); Graph Neural Network (GNN); knowledge distillation

0 引言

油浸式电力变压器是电力系统的核心设备, 其安全运行对电网稳定至关重要^[1-2]。油中溶解气体分析(DGA)是一种广泛应用于油浸式电力变压器状态估计和故障检测的技术^[3-4], 通过对变压器油中的溶解气体进行分析, 可以有效判断变压器是否存在故障, 以确保电网持续可靠供电, 并降低设备故障导致的停电及其经济影响^[5-6]。

传统 DGA 方法^[7]通常基于自身经验分析, 具有很高的不确定性, 亟需稳定可靠的方法来提升故障检测的质量和精度^[8-9]。近年来, 研究人员开始将人工智能应用于 DGA, 在很大程度上提升了变压器故障检测的准确性, 此类研究通常基于以下两点建模: 其一, 溶解气体含量和故障标签之间的非线性关系; 其二, 目标气体样本和其余样本的相似性度量。文献[10]考虑了当前样本和已知故障标签样本之间的相似性。文献[11-13]对不同气

基金项目: 甘肃省重点基础研究项目(24JRRA123); 甘肃省产业支撑项目(2022CYZC11)。

作者简介: 张瑞佳, 女, 硕士研究生, 主研方向为电力变压器故障诊断; 马慧芳(CCF 会员、通信作者), 教授、博士; 张映月, 硕士研究生; 彭生江, 教授级高级工程师、博士。

收稿日期: 2024-11-05

修回日期: 2025-03-03

E-mail: mahuifang@yeah.net

体浓度和故障标签之间复杂的非线性关系进行建模。近期,已有研究人员将图神经网络(GNN)用于 DGA^[14-15]。文献[14]利用对比双通道图滤波器,采用生成少数样本策略以解决数据不平衡问题。文献[15]通过采用双通道图编码和排名对比损失,提高了 DGA 诊断的准确性。因此,将 GNN 用于 DGA 不仅可以有效挖掘溶解气体含量和故障标签之间的非线性关系,还可以使用图的邻接矩阵更好地建模样本之间的关系,从而进一步提高故障检测的效率。

尽管 GNN 在 DGA 故障检测中展现出显著优势,但是 DGA 面临标记数据可用性不足的挑战。当只有少量标记节点进行模型训练时,性能往往不佳。为了解决上述问题,本文提出一种图知识蒸馏方法(GKDG)来进行 DGA 故障检测。首先,为了获得对样本的额外监督,采用双视角建图策略,即构建相似图和不相似图,将相似图作为学生网络,将不相似图作为教师网络;然后,引入多种知识在不同空间中对齐学生网络和教师网络,使学生模型从教师模型获得更多知识,有效捕捉和解释气体之间的复杂关系,从而实现少样本标记下的性能提升。本文的主要贡献总结如下:

1) 强调将图知识蒸馏纳入 DGA 的重要意义。

2) 创建双视角图来同时捕捉气体之间的相关性。此外,通过知识蒸馏技术增强 DGA 的能力,实现对具有多个约束条件的无标签数据的有效利用。

3) 在样本标签有限的场景下,对 2 个不同的 DGA 数据集进行实验,结果表明,GKDG 能够有效应对标签不足情形下的挑战。

1 准备知识

1.1 符号定义

用 $\mathcal{G}=(\mathcal{V}, \mathcal{E})$ 表示节点集为 $\mathcal{V}$ 、边集为 $\mathcal{E}$ 的无向图。 n 表示节点数, m 表示边数,连接节点 i 和 j 的边记作 e_{ij} 。 $\mathcal{C}=\{c_0, c_1, \dots, c_{t-1}\}$ 表示节点标签集合, t 为标签数量。节点的特征矩阵记作 $\mathbf{X} \in \mathbb{R}^{n \times F}$,其中, F 为节点特征数。 $\mathcal{G}$ 的结构用邻接矩阵 $\mathbf{A}$ 描述, $\mathbf{A}'$ 表示添加自环的邻接矩阵。 $\hat{\mathbf{A}}=\mathbf{D}^{-(1/2)} \mathbf{A}' \mathbf{D}^{-(1/2)}$ 表示带自环的对称归一化邻接矩阵, $\mathbf{D}$ 是 $\mathbf{A}'$ 的度矩阵。图拉普拉斯矩阵被定义为 $\mathbf{L}=\mathbf{D}-\mathbf{A}'$,对称拉普拉斯矩阵被定义为 $\hat{\mathbf{L}}=\mathbf{I}-\hat{\mathbf{A}}$ 。图构建过程见下文(第 2.2 节)。

1.2 GNN 表示学习

GNN 旨在聚合邻居节点信息来迭代更新节点表示。大多数 GNN 可以看作低通图滤波器,它在

图结构上平滑特征,使得相邻节点的表征趋于相似。然而,在一些图中,相连节点类别不一致,若平滑节点特征,可能会导致信息混淆。为此,相关研究^[16]引入了高通图滤波器来解决这个问题。

在后文构建的相似图上采用低通图滤波器进行特征提取,如下:

$$\mathbf{h}^{(l)}=f\left(\mathbf{h}^{(l-1)}, \hat{\mathbf{A}}\right) \quad (1)$$

式中: $\mathbf{h}^{(l)} \in \mathbb{R}^{n \times d}$ 为 l 层的节点表示; $\hat{\mathbf{A}}$ 是对称归一化的邻接矩阵; $f(\cdot, \cdot)$ 表示更新函数。

在后文构建的不相似图中,采取高通图滤波器实现节点表示学习,从而增强节点特征区分度,保留高频图信号,形式化定义为:

$$\mathbf{h}^{(l)}=f\left(\mathbf{h}^{(l-1)}, \hat{\mathbf{L}}\right), \hat{\mathbf{L}}=\mathbf{I}-\hat{\mathbf{A}} \quad (2)$$

通过将相似图作为学生 GNN 的输入,将不相似图作为教师 GNN 的输入,使得教师 GNN 为学生 GNN 提供不同视角的信息,辅助学生在更广泛的特征空间中进行学习。此外,在训练过程中,可以对教师 GNN 施加额外的约束,有助于防止其过拟合,从而提高模型的泛化能力。

1.3 问题定义

基于 DGA 的故障检测问题可被形式化为在标记数据有限的情形下的半监督节点分类任务。具体来说,给定标记节点数量有限的训练集 $\mathcal{V}^L$,目标是预测 $\mathcal{V}^U$ 中未标记节点的标签。问题定义如下:

输入 特征矩阵 $\mathbf{X}$,构造 k 近邻(k -NN)图的邻居数 k ,有标签的节点集合 $\mathcal{V}^L$,待预测的节点集合 $\mathcal{V}^U$
输出 $\mathcal{V}^U$ 的预测结果 $\hat{\mathbf{Y}}$

2 方法设计

2.1 概述

GKDG 的框架如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同),该框架包括 2 个关键设计策略:1)为了捕获不同气体样本之间的复杂关联并增加额外的监督信息,采用双视角图构建方式;2)构建一个与学生 GNN 具有相同架构的教师 GNN,以充分利用无标签数据,并引入多种知识以增强 DGA 的能力。

2.2 双视角图构建

2.2.1 相似图构建

为解决 DGA 标记数据有限的问题,本文提出双视角图构建策略 GKDG,分别描述样本之间的相似性与差异性。 k -NN 图在有效捕获样本间局部结构信息的同时,还能保证图的连通性,因此,本文采用 k -NN 方法来构建邻接矩阵,从而为特征提取提供更加精确的监督信息。

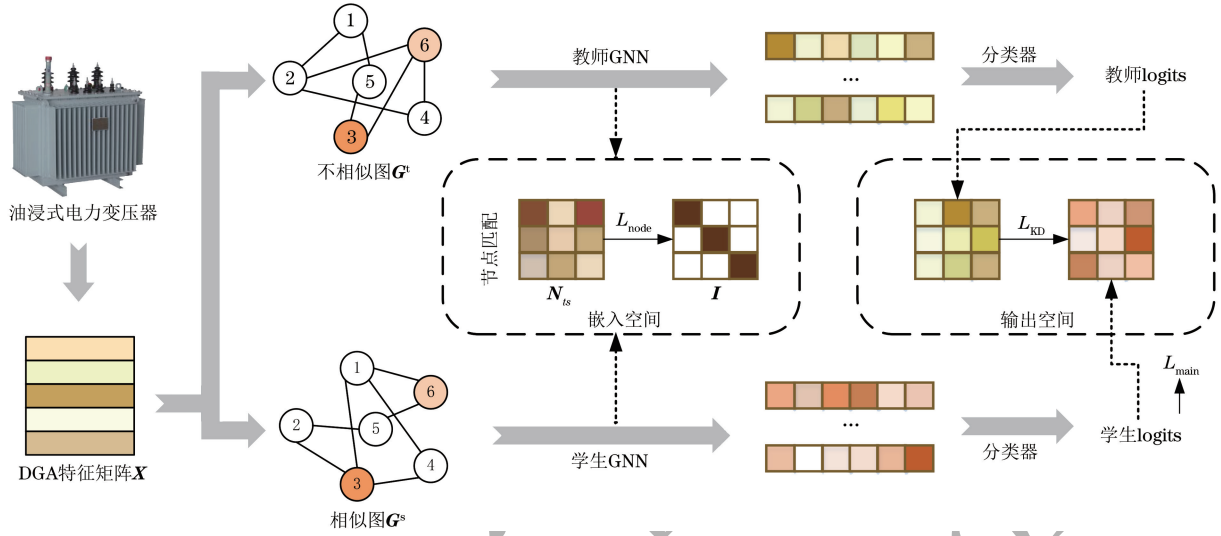


图 1 GKDG 的整体框架

Fig. 1 The overall framework of GKDG

计算任意 2 个样本之间的余弦相似度,得到相似度矩阵,记作 $S \in \mathbb{R}^{n \times n}$ 。选择与每个样本点相似度最高的前 k 个样本点作为其邻居(不包括自身),将节点 i 的相似节点集合表示为 $\mathcal{N}_i^s$ 。接着,构建相似 k -NN 图 $G^s = (\mathcal{V}, \mathcal{E}^s)$,邻接矩阵 A^s 元素计算如下:

$$A_{ij}^s = \begin{cases} s_{ij}, & j \in \mathcal{N}_i^s \\ 0, & j \notin \mathcal{N}_i^s \end{cases} \quad (3)$$

2.2.2 不相似图构建

受文献[16]的启发,本文构建不相似 k -NN 图来为样本提供更多额外的监督。在形式上,将不相似 k -NN 图记作 $G^d = (\mathcal{V}, \mathcal{E}^d)$, A^d 为图 G^d 的邻接矩阵, $\mathcal{N}_i^{ds}$ 表示节点 i 的不相似节点集合。

$$A_{ij}^d = \begin{cases} s_{ij}, & j \in \mathcal{N}_i^{ds} \\ 0, & j \notin \mathcal{N}_i^{ds} \end{cases} \quad (4)$$

2.2.3 设计讨论

上述构图策略可以有效实现双视角建模气体关联。相似图的构建旨在通过 GNN 学习,使得相邻节点在特征空间中更加接近,这有助于同类故障节点之间特征融合;不相似图的构建通过推开不同类别节点的特征表示,使得它们在学习过程中“越学越不像”,从而增强节点间的区分度。该过程优势如下:1)通过 2 种不同视角,丰富信息获取途径,强化带标签数据和无标签数据之间的联系,从而提高其泛化能力;2)双视角建模使得模型能够学习到更加可靠的特征表示,模型的泛化能力得到进一步增强,能够在标签稀缺时保持较高的性能。

2.3 多种知识蒸馏

本节引入多种知识,旨在使学生模型可以更好

地对齐教师模型。

2.3.1 预测级蒸馏

现有方法在标记数据有限时往往未能捕捉到足够丰富的信息,从而陷入性能瓶颈。图知识蒸馏技术^[17]通过在不同 GNN 之间传递知识,可以增强学习模型的能力,并从中获取更丰富的信息。

基于 Logits^[18] 的知识蒸馏通常用来在预测层面引导学生模型向教师模型看齐,该方法得到了广泛应用^[19-21]。近年来,该技术已扩展到图数据处理上^[22-24],具体来说,分别对学生 GNN 和教师 GNN 进行训练,然后对齐教师网络的 logits t_i 与学生网络的 logits z_i ,如式(5)所示:

$$L_{KD} = -\frac{1}{n} \sum_{i \in \mathcal{V}} \hat{\mathbf{y}}_i^T \text{lb}(\hat{\mathbf{y}}_i^S) \quad (5)$$

$$\hat{\mathbf{y}}_i^T = \text{Softmax}(t_i), \hat{\mathbf{y}}_i^S = \text{Softmax}(z_i)$$

式中: $\hat{\mathbf{y}}_i^S$ 和 $\hat{\mathbf{y}}_i^T$ 分别表示学生 GNN 和教师 GNN 预测的概率分布。

2.3.2 嵌入级蒸馏

相同故障节点在嵌入空间中可能表现为相似表示,但预测级蒸馏仅关注输出,未考虑学生与教师模型在嵌入空间中的差异。为提升泛化能力,与文献[25]类似,本文通过图结构在嵌入空间中建模实例及其关系,促进学生模型与教师模型对齐,从而在学习过程中捕捉到更细微的差异性,使知识有效传递。

利用皮尔逊相关系数构建结构嵌入图,指导学生 GNN 与教师 GNN 对齐,其计算公式如下:

$$e_{ij}^p = \frac{\sum_{k=1}^d (h_i^k - \bar{h}_i) (h_j^k - \bar{h}_j)}{\sqrt{\sum_{k=1}^d (h_i^k - \bar{h}_i)^2} \sqrt{\sum_{k=1}^d (h_j^k - \bar{h}_j)^2}} \quad (6)$$

式中: i 和 j 是嵌入空间中的样本点; d 是嵌入空间特征的维度; e_{ij}^d 为嵌入空间中连接节点 i 和节点 j 的边, 通过它们的相关性来量化它们的关系。

给定一批嵌入节点 $\mathcal{H} = \{h_1, h_2, \dots, h_B\}$, B 为同一批节点的数量, 依据式(7)对节点之间的相关性进行编码, 边矩阵可以表示为:

$$\mathbf{E}^p(\mathbf{h}, \mathbf{h}) = (e_{ij}^p) \in \mathbb{R}^{B \times B} \quad (7)$$

构建节点矩阵 $\mathbf{N}_{ts} = \mathbf{E}^p(\mathbf{h}_t, \mathbf{h}_s)$, 设计节点匹配损失函数, 使得同批次教师 GNN 和学生 GNN 同一节点的相关性向 1 靠近, 不同节点的相关性向 0 靠近, 损失函数定义如下:

$$L_{\text{node}} = \|\mathbf{N}_{ts} - \mathbf{I}\|_2 \quad (8)$$

式中: $\mathbf{I} \in \mathbb{R}^{B \times B}$ 是单位矩阵。

2.3.3 设计讨论

上述设计具有以下优势: 1) 标签高效性: 通过构建与学生 GNN 具有相同架构的教师 GNN, 能够有效地从未标记的数据中提取有用信息, 从而在不依赖大量标签的情况下提高学习效率; 2) 多种图知识蒸馏: 通过构建图结构, 可以有效揭示 DGA 数据中不同气体成分之间的复杂关联, 利用多种知识为学生 GNN 提供更为全面和丰富的监督信号。

2.4 模型优化

为了端到端地训练学生模型, 联合优化分类损失和图知识蒸馏损失, 完整的训练目标定义如下:

$$L = L_{\text{main}} + \gamma L_{\text{KD}} + \mu L_{\text{node}} \\ L_{\text{main}} = -\frac{1}{n} \sum_0^{n-1} \sum_0^{t-1} y_i^c \ln \hat{y}_i^c \quad (9)$$

式中: L_{main} 为主训练损失; y^c 和 $\hat{y}^c$ 分别表示标签为 c 时样本的真实值和预测值; γ 、 μ 是平衡参数。通过最小化训练目标, 使丰富的未标记数据和稀缺但有价值的标记数据共同为模型提供额外的监督信号。

3 实验分析

本节设计一系列实验来评估 GKDG 的有效性, 通过 3 个验证实验, 旨在回答以下研究问题:

1) 问题 1. GKDG 相对于基线方法性能怎么样, 特别是在标签有限的情况下?

2) 问题 2. GKDG 中的不同组成部分对其性能有何影响?

3) 问题 3. 超参数的变化对于 GKDG 影响如何?

3.1 数据描述

3.1.1 数据集

本文在 DGA-200 (<https://IEEE-dataport.org/documents/dissolved-gas-data-transformer-oil-fault->

diagnosis-power-transformers-membership-degree) 和 DGA-2000 (<https://github.com/alan-456/transformer-fault-dataset>) 2 个公开的数据集上进行广泛的实验。2 个数据集的特征均为氢气、甲烷、乙烷、乙烯以及乙炔的浓度含量。DGA-200 数据集共有 201 条数据, 标签依次为高温过热、中温过热、低温过热、低中温过热、局部放电、火花放电和电弧放电。DGA-2000 数据集共有 2 321 条数据, 标签依次为正常、低温过热、中温过热、高温过热、局部放电、低能放电和高能放电。利用这 2 个数据集构造 k -NN 图, 图中节点表示 DGA 样本, 边表示不同样本之间的相关性。

3.1.2 数据划分

为了确保模型之间的比较更加公正和全面, 在标记数据有限的场景下对每个数据集进行评估。数据的具体划分方式如下: 随机选取数据集的 50%、20%、10% 和 5% 的样本作为训练样本, 以评估模型在标记数据适中和标记数据有限条件下的表现。

3.2 实验设置

1) 环境设置。本文所使用的环境参数如表 1 所示。

表 1 环境参数

Table 1 Environmental parameters

环境参数	相关信息
处理器	Intel® Core™ i5-10300H CPU@2.50 GHz
内存/GB	16
操作系统	Windows 10
编程语言	Python
编程工具	Python 3.8.10
深度学习框架	PyTorch
图形处理器	NVIDIA Quadro P620

2) 评价指标。为了评估 GKDG 的有效性, 采用准确率(Accuracy)、精确率(Precision)和 ROC 曲线下面积(AUC)对模型性能进行评估, 使用粗体表示最佳性能。进行 10 次重复实验, 计算指标的平均值和标准差。

3) 基线方法。将 GKDG 与 DGA 中的机器学习方法、改进后的机器学习和深度学习方法以及基于图的方法进行比较, 具体包括:

(1) k -NN(2017 年)。 k -NN 利用不同的距离度量方法, 选择与该样本最接近的 k 个样本, 若样本邻域中的大多数属于某一个类别, 则该样本也属于这个类别。

(2) Tree(2022 年)。Tree 是一种树形结构的机

器学习算法,目标是构建决策树模型,用于对新的输入样本进行分类预测。

(3)WSO SVM(2023 年)。WSO SVM 通过无编码比率法提取变压器气体特征,并利用 WSO 算法优化 SVM 的核参数和惩罚因子,从而提高变压器故障诊断的准确率。

(4)IMBP(2023 年)。IMBP 通过残差网络和 SVM,是一种基于改进 BP 神经网络的电力变压器故障诊断方法。

(5)GCDGA(2024 年)。GCDGA 采用双通道图编码和排名对比损失进行电力变压器故障诊断。

4)实现细节。为了有效地聚合邻居节点的信息,在 DGA-200 数据集上设置 $k=2$,而在数据量更大的 DGA-2000 数据集上 k 值增加至 5,以便在更广泛的邻居信息中进行特征提取。同时,将嵌入维度依次设置为 128 和 256 维,GNN 层数依次为 2 和 3 层,增强模型的表达能力,同时避免过度复杂化。选用 Adam 优化器,学习率在 DGA-200 数据集上设置为 0.001,在 DGA-2000 数据集上则为 0.01,这样的设置有助于模型在训练过程中更快地收敛。表 2 所示为参数设置细节。

3.3 性能分析

本节对 GKDG 在标记适中和标记稀缺条件下的性能进行详细评估,并与其他基准方法进行比较,性能对比如图 2 所示,可以得出如下结果:

表 2 GKDG 的架构细节(以 DGA-2000 数据集为例)

Table 2 The architectural details of GKDG
(taking the DGA-2000 dataset as an example)

模块	输入大小	层类型	输出大小
学生 GNN	(2 321, 5)	GCNConv(5,256)	(2 321, 256)
	(2 321, 256)	GCNConv(256,256)	(2 321, 256)
	(2 321, 256)	GCNConv(256,128)	(2 321, 128)
教师 GNN	(2 321, 5)	GCNConv(5,256)	(2 321, 256)
	(2 321, 256)	GCNConv(256,256)	(2 321, 256)
	(2 321, 256)	GCNConv(256,128)	(2 321, 128)
全连接层	(2 321, 128)	Linear(dim 为 7)	(2 321, 7)

1)带标签样本适中的情况下 GKDG 的性能。当训练集占比为 50%和 20%时,GKDG 在 DGA-200 数据集上表现卓越,这表明在标签样本数量适中时,GKDG 具有良好的泛化能力。然而,在 DGA-2000 数据集上,GKDG 的准确率和精确率表现不及 k -NN,可能是由于以下原因:(1) k -NN 算法基于简单的局部性原理,在标签样本适中的情况下, k -NN 能够有效地利用这些样本的局部结构来进行分类,而 GKDG 可能无法高效地捕捉到这种局部关系;(2) k -NN 通过平均邻近样本的标签来减少噪声和异常值的影响,在标签样本适中的情况下,这种方法可能比 GKDG 更有效。值得注意的是,在图 2(f)中,GKDG 和 GCDGA 的 AUC 远高于其他方法,充分说明了双视角构图的有效性。

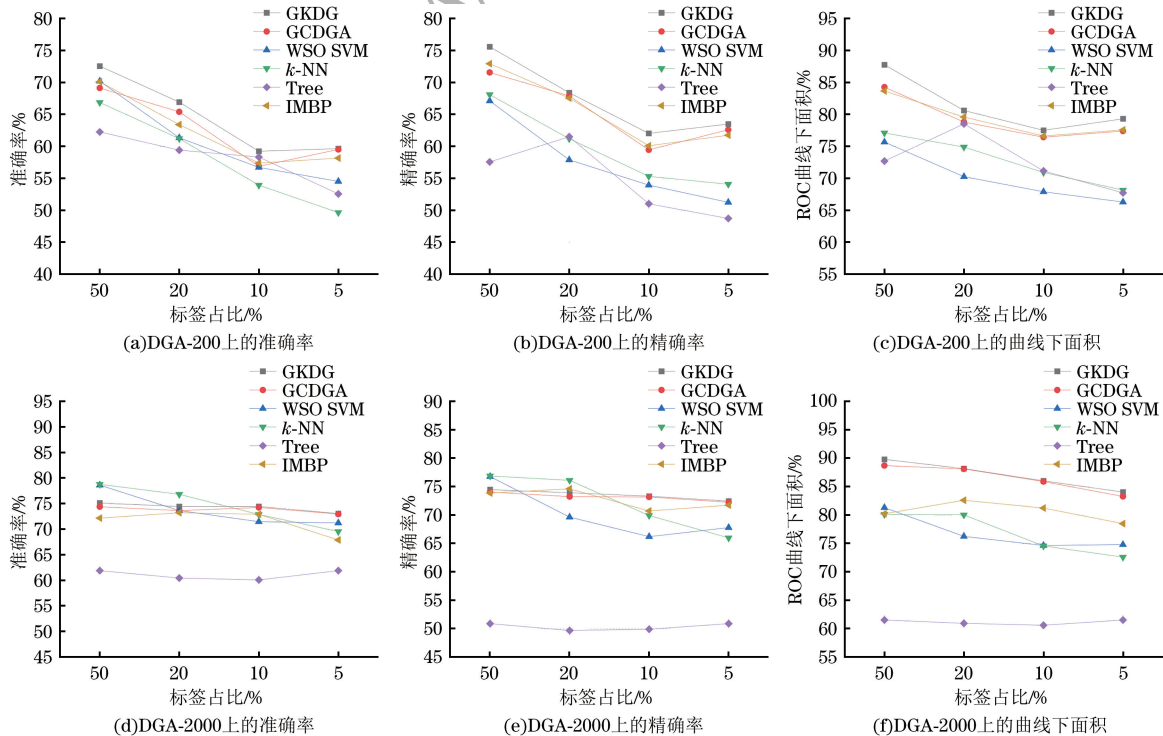


图 2 GKDG 与基线方法在不同数据集上的表现

Fig.2 Performance of GKDG and baseline methods on different datasets

2)带标签样本有限的情况下 GKDG 的性能。在训练样本占比降至 10%和 5%时,DGA-200 数据集中各评价指标普遍有所下降,尽管如此,GKDG 仍然优于其他基线方法。在 DGA-2000 数据集中,尽管训练样本进一步减少,GKDG 在各项指标上的表现却相对稳定,而一些基线方法的性能则出现了明显下降,特别是带标签样本适中时表现较好的 k -NN,在标注数据更少的情况下,其性能显著下降。相比之下,本文方法在标注数据稀缺时展现出了最优性能。值得注意的是,Tree 在 DGA-2000 数据集中性能远低于其他方法,原因可能在于 DGA-2000 数据集的样本数量众多,Tree 构建的决策树结构过于复杂,这可能导致过拟合问题,进而影响了性能表现。

3.4 组件有效性分析

本节对 GKDG 模型的关键组件进行性能评估,进行 k -NN 合理性验证以及蒸馏有效性检验。

1) k -NN 构图合理性。为了验证 k -NN 构图策略的有效性,采用 ϵ -邻域图作为对照组。在实验中, ϵ 取值为 $\{0.3, 0.5, 0.7, 0.9\}$,并与基于 k -NN 的构图方法进行比较。 k -NN 的最近邻数 k 选用第 3.5 节确定的最佳 k 值。实验结果如表 3 所示。GKDG 模型在使用 k -NN 构图时性能明显优于 ϵ -邻域图方法,这充分证明了本文构图策略的合理性。原因可能包括:

(1) k -NN 能够自适应地调整局部邻域结构,而 ϵ -邻域图对密度变化较敏感。

(2) k -NN 通过选择最近的 k 个点来减少噪声的影响,而 ϵ -邻域图可能将噪声点包含在邻域内。

此外,由于 DGA-2000 数据集规模较大,构建 ϵ -邻域图的时间复杂度较高,而 k -NN 方法在效率上更具优势。

表 3 k -NN 合理性分析

Table 3 k -NN rationality analysis

ϵ	DGA-200		DGA-2000	
	Accuracy/%	Precision/%	Accuracy/%	Precision/%
0.3	62.25±4.95	68.25±5.94	59.23±2.27	64.56±3.37
0.5	65.20±4.05	69.51±2.60	66.59±1.25	70.76±1.50
0.7	67.06±2.81	68.40±4.13	72.39±1.02	73.55±2.23
0.9	66.47±2.97	67.81±4.18	71.08±1.08	71.37±2.71
GKDG	72.55±2.24	75.57±3.15	75.13±1.24	74.46±1.41

2)蒸馏有效性。本次实验每次从模型中删除一个组件,w/o PL (Predicted Level) loss 表示去除预测级损失,w/o EL (Embedded Level) loss 表示去除嵌入级损失。实验结果如表 4 所示,从中可以看出:一方面,GKDG 在 2 个数据集上均显著优于

移除了预测级损失的变体,这表明,即便是在数据规模增大的情况下,预测级损失依旧能够有效提升模型的学习成效,由此可见,预测级损失确实在促进 GKDG 从教师 GNN 中学习和整合知识方面发挥了关键作用;另一方面,GKDG 表现得比没有嵌入级损失的 GKDG 更好,这表明嵌入级损失在模型训练过程中具有积极作用,其原因可能在于相似节点在嵌入空间中具有相似的节点表示,通过在嵌入空间中对齐学生图和教师图,有效地减少了节点分类的错误,从而提升了模型的准确性。总体来说,GKDG 的 2 个核心组成部分(嵌入空间和输出空间的对齐机制)对于确保学生 GNN 与教师 GNN 之间的知识传递至关重要,这种将 2 种损失函数相结合的策略,对于提升模型性能起着关键作用。此外,这一方法不仅为 DGA 领域开辟了新的研究视角,也进一步验证了将图知识蒸馏技术引入 DGA 领域的可行性与有效性,为后续研究提供了坚实的理论与实践基础。

表 4 蒸馏有效性分析

Table 4 Analysis of distillation effectiveness

模型	DGA-200		DGA-2000	
	Accuracy	Precision	Accuracy	Precision
GKDG	72.55±2.24	75.57±3.15	75.13±1.24	74.46±1.41
w/o PL loss	71.96±1.26	74.36±2.62	74.86±1.23	73.35±1.44
w/o EL loss	71.47±1.72	73.24±2.82	75.04±1.57	74.42±1.33

3.5 超参数敏感性分析

1) k 值分析。不同 k 值对 GKDG 的影响如图 3 所示。在图 3(a)中,GKDG 在 $k=2$ 时性能达到最佳, $k=1$ 时性能相对较低,而在其他 k 值下整体相对稳定,波动不大。这可能是因为 DGA-200 数据量较少,较小的 k 值无法提供足够的信息量,而过大的 k 值则可能导致将不够相似(或者不相似)的节点错误地连接起来,从而降低构造的图的可靠性。在图 3(b)中, $k=5$ 时模型达到最佳性能,DGA-2000 由于其较大的数据规模,可能包含了更多与特定样本相似的邻居节点。因此,选择较大的 k 值有助于模型在更丰富的邻域信息支持下提升其整体性能。

2)多种知识蒸馏权重分析。 γ 和 μ 分别为平衡预测级和嵌入级损失的超参数。图 4(a)和图 4(b)分别展示了在 DGA-200 和 DGA-2000 数据集下 GKDG 的准确率变化。可以看出,随着 γ 和 μ 的变化,GKDG 的性能波动不大,说明 GKDG 对不同知识蒸馏权重的调整具有较高的鲁棒性。进一步地,可以发现当 γ 和 μ 均为 0.1 时,模型性能略优于其他权重配置,因此,设置 $\gamma=0.1$ 、 $\mu=0.1$ 以优化模型性能。

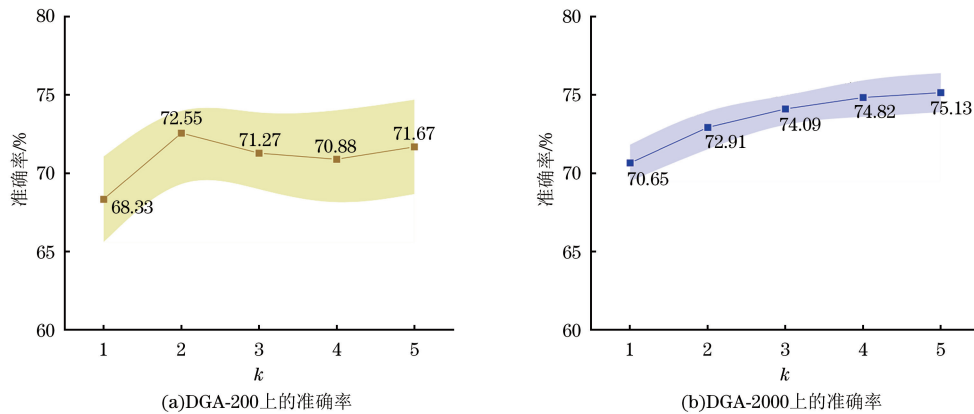
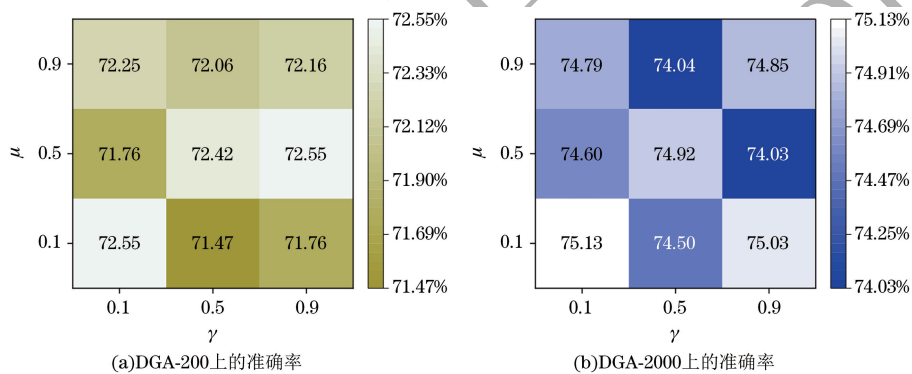
图 3 不同 k 值下 GKDG 的性能Fig.3 Performance of GKDG under different values of k 

图 4 多种知识蒸馏权重下 GKDG 的性能

Fig.4 Performance of GKDG under multiple knowledge distillation weights

4 结束语

实现一个针对数据标签稀缺情形的可泛化模型对于 DGA 至关重要。本文提出了一个用于 DGA 的图知识蒸馏方法 GKDG,用于解决标签数据有限的问题。采用双视角建图策略从样本领域中获得额外监督,并构建一个和学生 GNN 具有相同框架的教师 GNN,更好地捕获气体之间的复杂关系。为了使学生模型可以更好地向教师模型学习,引入多种知识来对齐学生模型和教师模型。实验结果证明了本文方法相对于其他基线方法的优越性。

参考文献

- [1] LIAO W, YANG D, WANG Y, et al. Fault diagnosis of power transformers using graph convolutional network[J]. CSEE Journal of Power and Energy Systems, 2020, 7(2): 241-249.
- [2] 罗文萱. 基于胶囊神经网络的电力变压器故障诊断方法研究[J]. 高压电器, 2024, 60(5): 92-98.
LUO W X. Research on fault diagnosis method of power transformer based on capsule networks[J]. High Voltage Apparatus, 2024, 60(5): 92-98. (in Chinese)
- [3] CHEN H C, ZHANG Y. Rethinking shallow and deep learnings for transformer dissolved gas analysis: a review[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2025, 32(1): 3-10.
- [4] PATIL M, PARAMANE A, DAS S, et al. Hybrid algorithm for dynamic fault prediction of HVDC converter transformer using DGA data [J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2024, 31(4): 2128-2135.
- [5] 毛业栋, 张春辉, 陈杰. 融合特征分析及机器学习的可演进变压器故障诊断模型[J]. 计算机工程, 2024, 50(8): 379-388.
MAO Y D, ZHANG C H, CHEN J. Evolvable transformer fault diagnosis model combining feature analysis and machine learning[J]. Computer Engineering, 2024, 50(8): 379-388. (in Chinese)
- [6] ZHANG X, YANG K Y. Transformer fault diagnosis method based on MTF and GhostNet [J]. Measurement, 2025, 249: 117056.
- [7] GOUDA O E, ELHOSHY S H, TAMALY H H E L. Condition assessment of power transformers based on dissolved gas analysis[J]. IET Generation, Transmission & Distribution, 2019, 13(12): 2299-2310.
- [8] 王宇, 祁琦, 王纯, 等. 储能变流器信号高精度故障诊断方法[J]. 计算机工程, 2024, 50(8): 389-396.
WANG Y, QI Q, WANG C, et al. High-precision fault diagnosis method for energy storage inverter signals [J]. Computer Engineering, 2024, 50(8): 389-396. (in Chinese)
- [9] CHEN H C, ZHANG Y. Dissolved gas analysis using knowledge-filtered oversampling-based diverse stack learning[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 2505211.
- [10] THOTE P B, DAIGAVANE M B, DAIGAVANE P M, et al. An intelligent hybrid approach using KNN-GA to enhance the performance of digital protection transformer scheme[J]. Canadian Journal of Electrical and Computer

- Engineering, 2017, 40(3): 151-161.
- [11] MENEZES A G C, ARAUJO M M, ALMEIDA O M, et al. Induction of decision trees to diagnose incipient faults in power transformers [J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2022, 29(1): 279-286.
- [12] LI K, YANG G Q, WANG K, et al. Fault diagnosis method of transformer based on WSO and SVM[C]//Proceedings of the 7th IEEE Conference on Energy Internet and Energy System Integration (EI2). Washington D. C., USA: IEEE Press, 2024: 3732-3736.
- [13] JIN Y S, WU H, ZHENG J F, et al. Power transformer fault diagnosis based on improved BP neural network [J]. Electronics, 2023, 12(16): 3526.
- [14] SHU K, MA H F, YANG J P, et al. GraphSmin: imbalanced dissolved gas analysis with contrastive dual-channel graph filters[J]. Advanced Engineering Informatics, 2024, 62: 102839.
- [15] ZHANG Y Y, MA H F, ZHANG D, et al. Graph contrastive learning for dissolved gas analysis [EB/OL]. [2024-10-05]. https://link.springer.com/chapter/10.1007/978-981-96-0821-8_12.
- [16] JIN W Y, MA H F, ZHANG Y Y, et al. Multi-view discriminative edge heterophily contrastive learning network for attributed graph anomaly detection[J]. Expert Systems with Applications, 2024, 255: 124460.
- [17] TIAN Y J, PEI S C, ZHANG X L, et al. Knowledge distillation on graphs: a survey [J]. ACM Computing Surveys, 2025, 57(8): 1-16.
- [18] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network [EB/OL]. [2024-10-05]. <https://arxiv.org/abs/1503.02531>.
- [19] LIU J J, KE W J, WANG P, et al. Towards continual knowledge graph embedding via incremental distillation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(8): 8759-8768.
- [20] SUN W J, CHEN D F, LÜ S W, et al. Knowledge distillation with refined logits [EB/OL]. [2024-10-05]. <https://arxiv.org/abs/2408.07703>.
- [21] WANG Q, ZHOU J J. Multi-perspective contrastive logit distillation[EB/OL]. [2024-10-05]. <https://arxiv.org/abs/2411.10693>.
- [22] TIAN Y J, XU S K, LI M Y. Decoupled graph knowledge distillation: a general logits-based method for learning MLPs on graphs[J]. Neural Networks, 2024, 179: 106567.
- [23] GUO Z, WANG D, HE Q, et al. Leveraging logit uncertainty for better knowledge distillation [J]. Scientific Reports, 2024, 14: 31249.
- [24] XU L X, WANG Z W, BAI L, et al. Multi-level knowledge distillation with positional encoding enhancement[J]. Pattern Recognition, 2025, 163: 111458.
- [25] MA Y C, CHEN Y B, AKATA Z. Distilling knowledge from self-supervised teacher by embedding graph alignment [EB/OL]. [2024-10-05]. <https://arxiv.org/abs/2211.13264>.

文字编辑 吴云芳
栏目编辑 宋 圆