自动化学报
ACTA AUTOMATICA SINICA

# 面向文本属性图基础模型的“检索分治”测试时自适应方法

靳毅凡 李懿 宋飞 王瑞 李江梦 刘立祥 孙富春

# Divide-and-conquer Test-time Adaptation With Retrieval-augmentation for Text-attributed Graph Foundation Model

JIN Yi-Fan, LI Yi, SONG Fei, WANG Rui, LI Jiang-Meng, LIU Li-Xiang, SUN Fu-Chun

在线阅读 View online: https://doi.org/10.16383/j.aas.c250517

---

# 面向文本属性图基础模型的“检索−分治”测试时自适应方法

靳毅凡[1,2] 李 懿[1,2] 宋 飞[1,2] 王 瑞[2] 李江梦[2] 刘立祥[1,2] 孙富春[3]

**摘 要** 文本属性图基础模型 (TAG-FM) 旨在通过在大规模图数据集上预训练, 实现稳健的跨域泛化能力. 尽管 TAG-FM 已在多项下游任务中展现出良好的性能, 但是通过深入分析发现, 其在泛化能力方面仍存在关键弱点: 域适应阈值效应, 即当预训练域与测试域之间的分布偏移超过某一临界阈值时, 模型性能将出现显著退化. 由于基础模型复杂度高且测试数据稀缺, 引入测试时自适应方法成为缓解该问题的可行方案. 然而现有测试时自适应方法在实例层和域层均存在局限, 严重制约着 TAG-FM 对分布外数据的适应性能. 针对此问题, 提出一种面向 TAG-FM 的“检索−分治”测试时自适应方法, 该方法通过检索增强的特征共形融合模块生成融合图结构和节点文本语义的统一特征, 并引入一种分而治之的测试时自适应策略, 通过渐进式自适应过程学习泛化性语义. 在 7 个基准文本属性图数据集和 12 个严重域偏移文本属性图数据集上的实验结果表明, 本文方法能够显著增强 TAG-FM 的跨域泛化能力.

**关键词** 文本属性图基础模型; 域泛化; 测试时自适应; 检索增强; 分而治之

**引用格式** 靳毅凡, 李懿, 宋飞, 王瑞, 李江梦, 刘立祥, 孙富春. 面向文本属性图基础模型的“检索−分治”测试时自适应方法. 自动化学报, 2026, **52**(8): 1690−1709

**DOI** 10.16383/j.aas.c250517 **CSTR** 32138.14.j.aas.c250517

## Divide-and-conquer Test-time Adaptation With Retrieval-augmentation for Text-attributed Graph Foundation Model

JIN Yi-Fan[1,2] LI Yi[1,2] SONG Fei[1,2] WANG Rui[2] LI Jiang-Meng[2] LIU Li-Xiang[1,2] SUN Fu-Chun[3]

**Abstract** Text-attributed graph foundation model (TAG-FM) is designed to attain robust cross-domain generalization capability through pretraining on large-scale graph datasets. Although TAG-FM has demonstrated promising performance on various downstream tasks, our in-depth analysis reveals a critical limitation in its generalization capability: The domain adaptation threshold effect, i.e., model performance undergoes catastrophic degradation when distributional shifts between pretraining and testing domains surpass critical thresholds. Due to the excessive foundation model complexity and scarcity of testing data, a feasible solution to mitigate this issue is to introduce test-time adaptation approaches for TAG-FM. Nevertheless, existing test-time adaptation approaches exhibit substantial limitations at both the instance and domain levels, severely constraining the ability of TAG-FM to adapt to out-of-distribution data. To address this issue, we innovatively propose the divide-and-conquer test-time adaptation with retrieval-augmentation for TAG-FM. Specifically, a retrieval-augmented feature conformal integration module is introduced to construct consolidated node features by integrating graph structures with node textual semantics. Furthermore, a divide-and-conquer test-time adaptation strategy is developed to progressively capture generalizable semantics through an iterative adaptation process. Empirically, we verify that the proposed method achieves the state-of-the-art performance on 7 benchmark text-attributed graph datasets and 12 constructed text-attributed graph datasets exhibiting severe domain shifts, which confirms that our method effectively enhances the cross-domain generalization capability of TAG-FM.

**Keywords** text-attributed graph foundation model; domain generalization; test-time adaptation; retrieval-augmentation; divide-and-conquer

**Citation** Jin Yi-Fan, Li Yi, Song Fei, Wang Rui, Li Jiang-Meng, Liu Li-Xiang, Sun Fu-Chun. Divide-and-conquer test-time adaptation with retrieval-augmentation for text-attributed graph foundation model. *Acta Automatica Sinica*, 2026, **52**(8): 1690−1709

收稿日期 2025-09-30 录用日期 2026-03-04
Manuscript received September 30, 2025; accepted March 4, 2026
国家自然科学基金 (62406313) 资助
Supported by National Natural Science Foundation of China (62406313)
本文责任编委 金连文
Recommended by Associate Editor JIN Lian-Wen

图数据广泛存在于众多关键领域, 如社交网络[1−3]、网络安全[4] 等. 图神经网络通过对图结构的高效编

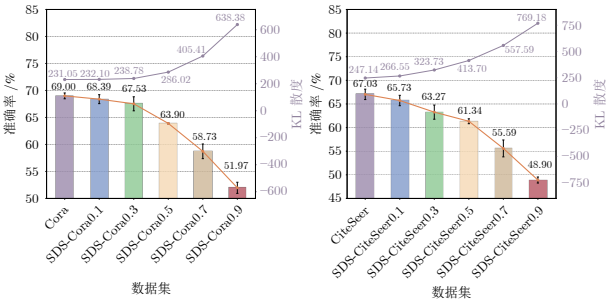
1. 中国科学院大学 北京 101408 2. 中国科学院软件研究所 北京 100190 3. 清华大学 北京 100084
1. University of Chinese Academy of Sciences, Beijing 101408
2. Institute of Software, Chinese Academy of Sciences, Beijing 100190
3. Tsinghua University, Beijing 100084

码, 在社区发现、异常检测等多项任务中取得显著成效[1−5]. 然而, 随着信息规模的急剧扩张, 现实世界中的图节点往往附带丰富的文本属性, 由此催生的文本属性图 (text-attributed graphs, TAGs) 目前已成为图表示学习领域新的研究焦点. 近期研究 [6−8] 表明, 有效融合节点的文本属性信息可显著提升下游任务的性能.

大语言模型 (large language models, LLMs) 凭借其卓越的深层语义理解能力[9−11], 能够理解 TAGs 中丰富的文本语义信息, 进而产生一系列突破性研究, 推动了 TAGs 方法的发展. 例如, OFA[12] 和 ZeroG[13] 摒弃浅层的文本嵌入方法, 如词袋模型, 转而采用基于 LLMs 的高阶特征提取器, 从而获得更具任务判别性的节点表示. 另一条主流研究路线致力于使 LLMs 直接理解图结构信息[14−15]. 具体而言, GraphGPT[14] 与 LLaGA[15] 借助图神经网络将拓扑结构编码为离散的图单元, 并借助附加的投影层将这些单元映射到文本嵌入空间. 尽管上述工作已取得显著进展, 但其对人工标注标签的高度依赖不仅限制了大规模无标注语料的高效利用, 也严重制约了 TAGs 模型的规模化扩展. 依据缩放定律[16−17], 这一局限性从根本上限制了 TAGs 模型的性能上限. 受视觉−语言自监督学习 (如 CLIP[18]) 取得巨大成功的启发, Zhu 等[19] 提出一种图−文对比自监督预训练策略, 该策略降低了对高质量人工标注数据的依赖, 同时支持模型尺寸的规模化扩展, 最终得到具备跨域泛化能力的文本属性图基础模型 (text-attributed graph foundation model, TAG-FM).

然而, 本文通过探索实验发现, TAG-FM 在进行域泛化时仍存在关键缺陷——域适应阈值效应[20], 即当预训练域 $\mathcal{S}$ 与测试域 $\mathcal{T}$ 之间的分布偏移超过某一临界阈值时, 模型性能将出现灾难性骤降. 考虑到 KL 散度可作为域偏移程度的度量指标, 图 1 展示了 TAG-FM 在不同域偏移强度的测试数据集上进行节点分类任务的准确率, 以及预训练域与测试域之间的 KL 散度变化[1]. 图 1 中, 以“SDS-Cora-0.5”为例, 其表示基于 Cora 数据集构建的严重域偏移数据集, 通过随机删除图中 50% 的边实现, 域偏移程度为 0.5. 实验结果显示, 随着预训练域与测试域之间的 KL 散度逐渐增大, TAG-FM 在测试域上的分类准确率整体呈下降趋势, 且在 Cora 和 CiteSeer 数据集上分别于域偏移强度达到 0.5 和 0.3 时

[1] 本文将 TAG-FM 的预训练数据视为预训练域, 将 Cora、CiteSeer 等测试数据集视为测试域, 一个测试数据集可视为一个测试域. 值得注意的是, TAG-FM 的预训练数据仅用于探索实验, 所提方法的实际实现并未使用任何 TAG-FM 预训练数据.




(a) Cora　　(b) CiteSeer

图 1　不同域偏移强度下 TAG-FM 性能及预训练域与测试域之间的 KL 散度

Fig. 1　TAG-FM performance and the KL divergence between the pretraining and testing domains under varying degrees of domain shifts

出现明显骤降. 由于基础模型的高度复杂性和测试数据的稀缺性, 引入测试时自适应 (test-time adaptation, TTA) 方法[21−24] 是缓解该问题的可行方案. 但本文研究发现, 现有 TTA 方法存在两个层面的固有缺陷, 从而严重制约了 TAG-FM 在分布外 (out-of-distribution, OOD) 数据上的适应效果.

1) 在实例层面, 现有针对图数据的 TTA 方法[22−24] 侧重挖掘图中的局部结构信息, 并隐含假设局部结构能够稳定刻画节点在测试阶段下游任务中的判别语义. 然而, 在分布外场景下, 局部结构往往呈现出显著的域间差异性, 致使其极易受噪声干扰而降低可靠性, 进而导致基于局部结构学习得到的节点表示发生语义偏离. 在此背景下, 节点文本属性作为具有语义不变性的重要信息来源, 其在测试阶段对于纠正由局部结构引起的语义偏离的作用亟须被充分挖掘. 为深入探究此问题, 本文采用预训练域语义空间作为刻画节点任务语义的近似参照系, 从预训练域和测试域中, 筛选出 20 对在语义空间高度邻近的代表性实例对, 进而系统比较不同节点编码策略对其语义相似度的影响, 实验结果如图 2 所示. 图 2 中“g”表示仅编码结构特征, “g+t”表示增加节点自身文本属性, “g+t+n”表示进一步融合邻域节点的文本属性. 可以观察到, 在节点编码过程中引入节点文本属性后, 测试实例的语义偏离程度显著减小; 当进一步融合邻域节点的文本属性时, 这种偏离现象得到进一步缓解. 这表明, 尽管分布偏移削弱了局部结构的语义可靠性, 但文本属性中蕴含的丰富语义信息能够有效弥补这一缺陷, 为实例层表示提供关键的语义约束.

2) 在域层面, 现有测试时方法普遍忽视了严重域偏移下的样本复杂度假定, 即“样本诅咒效应”. 根据统计学习理论, 为实现模型从预训练域 $\mathcal{S}$ 到

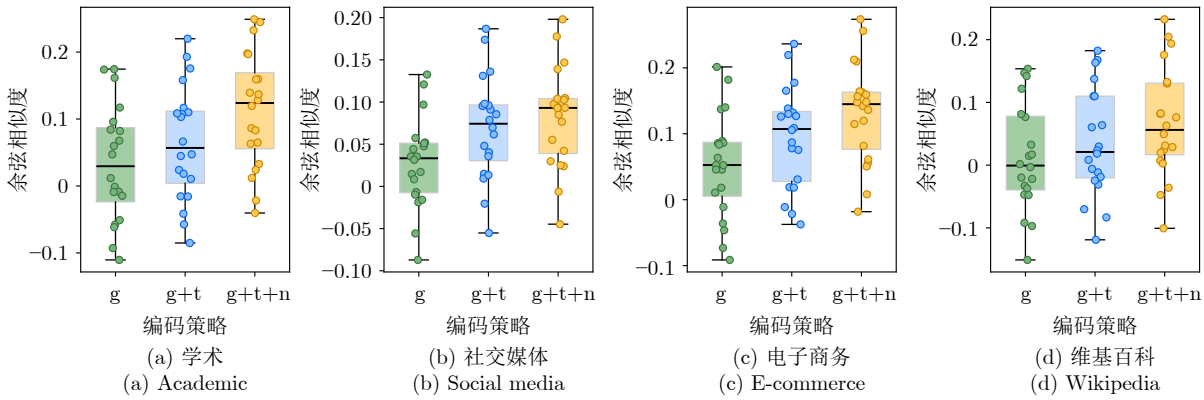



图 2 代表性实例对在不同编码策略下的语义相似度比较

Fig. 2 Comparison of semantic similarity for representative instance pairs across different encoding strategies

测试域 $\mathcal{T}$ 的适应, 所需的最小支撑样本量 $N^*$ 取决于两域之间的 KL 散度[25−26]. 具体而言, $N^*$ 与两域间的 KL 散度呈现指数级增长关系, 即满足 $N^* \propto \exp(\mathrm{KL}(\mathcal{S}||\mathcal{T}))$. 如图 1 所示, 随着测试域偏移程度加剧, 预训练域与测试域之间的 KL 散度也显著增大, 进而导致所需样本量 $N^*$ 急剧上升. 然而, 现有测试时方法仅基于熵最小化准则直接进行从 $\mathcal{S}$ 到 $\mathcal{T}$ 的适应, 在实际可用样本量远小于 $N^*$ 的情况下, 难以实现有效适应.

针对上述问题, 本文提出一种面向文本属性图基础模型的“检索−分治”测试时自适应方法 (divide-and-conquer test-time adaptation with retrieval-augmentation, DORA), 通过“检索–分治”机制同时修正实例层与域层缺陷, 实现在 OOD 数据上的稳健泛化. 具体而言, 为充分利用节点文本属性以缓解分布偏移带来的影响, 提出检索增强的特征共形融合 (retrieval-augmented feature conformal integration, RFCI) 模块, 该模块首先基于检索增强的节点属性总结 (retrieval-augmented node attribute summarization, RAS) 模块获得增强的节点文本总结, 之后双模态特征共形融合 (dual-modal feature conformal integration, DFCI) 模块采用能量分数将结构特征与文本特征融合得到统一节点特征. 此外, 为避免预训练域 $\mathcal{S}$ 与测试域 $\mathcal{T}$ 之间的分布鸿沟带来的样本诅咒效应, 引入分而治之测试时自适应 (divide-and-conquer test-time adaptation, D&C-TTA) 策略, 通过构造多个插值过渡域连接 $\mathcal{S}$ 与 $\mathcal{T}$, 实现渐进式自适应. 在多个基准 TAGs 数据集上的对比实验表明, DORA 显著优于现有最优方法.

总体而言, 本文的贡献可概括为以下三个方面:

- 揭示 TAG-FM 在 OOD 数据上的域适应阈值效应, 并从实例层与域层两个维度阐明现有 TTA 方法难以有效支持 TAG-FM 的根本原因.
- 首次将 TTA 范式引入 TAG-FM, 并创新性地提出 DORA, 通过“检索增强的特征共形融合”与“分而治之测试时自适应”两大核心模块实现在 OOD 数据上的稳健泛化.
- 构建更具挑战性的严重域偏移 TAGs 数据集, 同时在 19 个不同域偏移程度的 TAGs 数据集上取得了当前最优性能.

## 1 相关研究

### 1.1 文本属性图方法

目前, 利用 LLMs 进行文本属性图表示学习的研究已成为前沿热点[27]. OFA[12] 将多源图数据统一转换为文本属性图格式, 借助 LLMs 完成跨域一致编码. ZeroG[13] 利用语言模型同步编码节点属性与类别语义, 实现跨域特征维度统一. GraphGPT[14] 通过图指令微调将结构知识注入 LLMs, 使其具备理解复杂图结构的能力. LLaGA[15] 将文本属性图编码为结构感知的序列表示, 实现 LLMs 对图结构信息的直接处理. 近期研究进一步强化了结构与语义的协同建模能力. BiGTex[28] 通过堆叠式图文融合单元将 GNN 与 LLMs 紧密结合, 实现图结构和文本语义信号的双向交互. GraphiT[29] 则探索将文本属性图转化为文本序列, 并结合 LLMs 提示优化机制完成节点级预测任务. 为降低对高质量人工标注数据的依赖并扩大模型规模, GraphCLIP[19] 在 CLIP[18] 的双编码架构基础上, 通过自监督对比学习将节点文本语义与图结构表示映射至统一嵌入空间, 形成了具有代表性的 TAG-FM 范式. 该类模型强调在通用无标注数据上学习稳健的跨模态对齐表示, 从而支持下游任务的统一适配与迁移. 然而, 本文进一步发现目前的 TAG-FM 在测试阶段存在域适应阈值效应这一关键泛化缺陷, 并提出 DORA,

引入检索增强的特征共形融合和分而治之 TTA 方法显著提升其分布外泛化能力.

### 1.2　测试时自适应

测试时自适应最早应用于计算机视觉领域, 目标是在无法获取测试集标签及预训练数据的情况下, 实现在线更新主干模型. Tent[21] 首次提出通过熵最小化策略更新批归一化层参数. COME[30] 揭示了传统熵最小化策略因过度自信而导致模型崩溃的致命局限, 并创新性地提出通过狄利克雷先验分布显式建模预测不确定性, 以实现更保守、稳定的自适应优化. 进一步地, DeYO[31] 则指出单一熵度量不足以衡量置信度, 进而引入熵和伪标签概率差 (pseudo-label probability difference, PLPD) 的联合度量以筛选样本. 针对伪标签在模型自适应过程中难以被灵活优化的固有问题, PASLE[32] 框架提出通过选择性标签增强为不确定样本分配候选伪标签集, 以提升伪标签的可靠性与模型性能.

近年来, 针对图数据的 TTA 方法亦取得进展. GAPGC[33] 与 GT3[34] 率先将自监督对比学习引入图编码器的推理阶段. GTrans[24] 从数据视角出发, 对节点特征与图结构施加可学习扰动以实现自适应. GraphPatcher[23] 针对度分布偏移, 迭代生成虚拟节点以提高低度节点预测性能. SOGA[35] 通过联合优化结构一致性与信息最大化目标, 将源模型适应至测试图数据. Matcha[22] 则聚焦结构偏移, 动态调整图神经网络中的聚合参数以缓解分布偏移. PGNDS[36] 探索在不更新图神经网络参数的前提下, 通过引入基于双重条件扩散的生成式投影机制, 将测试图分布逆向映射回模型熟悉的训练分布. ASSESS[37] 通过基于子图互信息的自适应样本选择与原型正则化监督, 实现可靠自适应. 然而, 实例层与域层的双重缺陷显著限制了现有 TTA 方法在 TAG-FM 面对 OOD 数据时的泛化能力. 在此背景下, 本文首次探索将 TTA 范式引入文本属性图基础模型领域, 以期突破现有方法的局限.

### 1.3　渐进式域适应

渐进式域适应 (gradual domain adaptation, GDA) 最早在计算机视觉领域受到关注. Gadermayr 等[38] 在 2018 年率先将疾病进展的有序阶段视为天然的中间域序列, 提出一个面向病理全切片分割任务的无监督渐进式域适应框架. 随后, Kumar 等[39] 从理论层面系统分析了渐进式域适应的有效性, 证明了渐进式适应相较于传统的一步式直接域适应具有更显著的优势. 在此基础上, 针对缺乏中间域序列信息的现实需求, Chen 等[40] 提出中间域标记算法, 通过由粗到细的域发现与序列细化机制, 从未索引的中间数据中自动恢复过渡域序列. 近期, 渐进式域适应的研究开始拓展至图学习场景. GGDA[41] 通过引入融合 Gromov-Wasserstein 度量的桥接中间图生成机制, 并结合节点级选择与自适应质量衰减策略, 逐步构造中间域序列, 从而缓解严重分布偏移条件下的图域适应困难. Gadget[42] 则进一步从图渐进式域适应的最优路径角度展开研究, 通过理论分析证明融合 Gromov-Wasserstein 测地线可作为渐进式适应的最优过渡路径, 并据此生成中间图以实现逐步迁移. 尽管上述方法为渐进式域适应在图学习中的应用提供了重要启发, 但现有方法均建立在预训练域数据可访问的前提之上, 本文进一步探索在未知预训练域数据条件下, 通过分治策略构造过渡域, 实现渐进式测试时自适应算法.

## 2　预备知识与分析

### 2.1　问题定义

本文聚焦于 TAGs 的节点分类任务. 与传统图不同, TAGs 中的每个节点都有相应的文本属性. 形式上, 将 TAG 定义为 $\mathcal{G}=\{\mathcal{V},\ \{T_n\}_{n=1}^{N},\ X,\ A,\ Y\}$. 其中 $\mathcal{V}=\{v_1,\ \cdots,\ v_N\}$ 为节点集, $N=|\mathcal{V}|$ 表示节点总数; $T_n$ 为节点 $v_n$ 对应的文本属性; $X$ 为初始节点特征矩阵; $A\in\mathbf{R}^{N\times N}$ 为邻接矩阵; $Y\in\{0,\ 1\}^{N\times C}$ 为标签矩阵, $C$ 表示类别总数. 节点分类任务的目标是预测给定节点的类别标签.

### 2.2　文本属性图基础模型

本文提出的方法以预训练的 TAG-FM 为主干模型, 该模型遵循 CLIP[18] 框架, 包含图编码器和文本编码器两个组件. 其中图编码器通常基于图 Transformer[43] 实现. 对于给定节点 $v_n$, TAG-FM 首先通过随机游走采样[44] 等方法, 构建以 $v_n$ 为中心的自我中心图 $G_n=(X_n,\ A_n,\ P_n)$, 其中 $X_n$ 为初始节点特征矩阵, $A_n$ 为邻接矩阵, $P_n$ 表示节点位置嵌入矩阵. 随后, 图编码器通过以下公式得到图 $G_n$ 的中心节点 $v_n$ 的结构特征:

$$z_n^s=\mathrm{MLP}(\mathrm{POOL}(g(X_n,\ A_n,\ P_n))) \tag{1}$$

其中, $\mathrm{MLP}(\cdot)$ 表示多层感知机; $\mathrm{POOL}(\cdot)$ 表示平均池化函数; $g(\cdot)$ 表示图编码器; $z_n^s$ 的上标 $s$ 特指“结构”, 与第 3.1.2 节的文本特征相对应.

文本编码器采用基于 Sentence-Transformer 架构的语言模型, 输入为与类别相关的提示模板 $\Phi_c$ (例如“This paper has a topic on {class}”), 其嵌入向量 $z_c$ 通过以下公式得到:

$$z_c = \mathrm{POOL}(f([\mathrm{CLS}],\ \Phi_c)) \tag{2}$$

其中, [CLS] 表示 Transformer 架构中用于捕获输入序列全局语义表示的特殊分类标记; $f(\cdot)$ 表示文本编码器.

节点 $v_n$ 的最终预测结果由两步计算得到. 首先将节点的结构特征 $z_n^s$ 与各类别对应的文本提示嵌入向量 $\{z_c\}_{c=1}^{C}$ 进行相似度匹配, 并通过温度系数 $\tau$ 对相似度进行缩放, 从而得到对数几率 (logits):

$$p_c^s = \frac{\mathrm{sim}(z_n^s,\ z_c)}{\tau} \tag{3}$$

其中, $\mathrm{sim}(\cdot,\ \cdot)$ 表示余弦相似度; $p_c^s$ 表示基于结构特征 $z_n^s$ 预测节点 $v_n$ 属于第 $c$ 类的对数几率. 随后, 对上述对数几率施加 softmax 运算, 以获得类别预测概率:

$$\hat{p}_c^s = \mathrm{softmax}(p_c^s) = \frac{\exp(p_c^s)}{\sum\limits_{c=1}^{C}\exp(p_c^s)} \tag{4}$$

其中, $\hat{p}_c^s$ 表示基于结构特征 $z_n^s$ 预测节点 $v_n$ 属于第 $c$ 类的预测概率. 最终, 节点 $v_n$ 的预测类别 $\hat{y}_n$ 由其与文本提示匹配程度最高的类别确定, 即

$$\hat{y}_n = \arg\max_{c\in[1,\ C]}\ \hat{p}_c^s \tag{5}$$

## 2.3 域适应阈值效应

### 2.3.1 形式化定义

随着测试域分布逐步偏离预训练域, TAG-FM 的性能并不总是随分布偏移强度平滑退化. 相反, 当分布偏移达到某一临界水平时, 模型性能会出现显著下降, 并在后续更严重的分布偏移条件下呈现持续退化趋势. 本文将该现象称为域适应阈值效应, 并基于性能下降比率对其进行形式化定义, 以作为后续实验分析与方法比较的统一判定标准.

**定义 1 (域适应阈值效应).** 记模型在不同分布偏移强度 $\epsilon$ 下的测试性能为 $\mathrm{Acc}(\epsilon)$. 设 $\epsilon_0 < \epsilon_1 < \cdots < \epsilon_m < \cdots < \epsilon_M$ 表示预训练域与测试域之间分布偏移逐步增强的一系列测试设置, 其中 $\epsilon_0$ 为原始测试数据集的分布偏移强度, $\epsilon_M$ 表示实验中所考虑的最大分布偏移强度. 若存在某一偏移强度 $\epsilon^* = \epsilon_m$, 使得模型在该设置下的测试性能相对于原始数据集出现显著下降, 即

$$\frac{\mathrm{Acc}(\epsilon_0) - \mathrm{Acc}(\epsilon_m)}{\mathrm{Acc}(\epsilon_0)} > \eta \tag{6}$$

且在随后的更强分布偏移设置下模型性能呈持续下降趋势, 则称在 $\epsilon^*$ 处发生域适应阈值效应, 并将 $\epsilon^*$ 定义为对应的域适应阈值. $\eta$ 为超参数, 本文在所有实验中统一采用 $\eta = 0.05$, 以保证不同数据集和模型配置下阈值判定标准的一致性.

### 2.3.2 实证分析

为系统刻画文本属性图基础模型在跨域测试场景下的域适应阈值效应, 本文在不同域偏移强度条件下, 对 GraphCLIP 与 GraphCLIP+GTrans 两种模型配置在多个数据集上的节点分类性能变化趋势进行系统评估, 如图 3 和图 4 所示. 实验中, 域偏移强度通过随机删除图中边的比例进行控制, 以模拟由结构扰动引起的分布偏移.

图 3 展示了基线方法 GraphCLIP 在不同数据集及不同域偏移强度下的性能表现. 可以观察到, 尽管不同数据集的具体退化模式存在差异, 但是 GraphCLIP 在所有数据集上均出现了明显的域适应阈值效应, 即当域偏移强度超过某一临界水平后, 模型性能发生显著退化. 然而, 不同数据集上临界阈值 $\epsilon^*$ 的出现位置并不一致. 例如, 在 Cora 数据

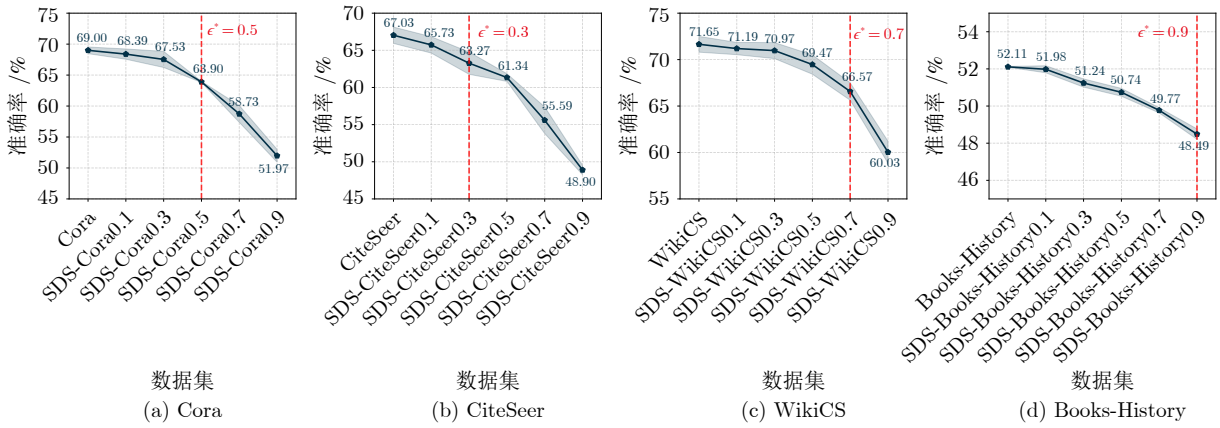



图 3 GraphCLIP 在不同域偏移设置下的性能表现

Fig. 3 Performance of GraphCLIP under different domain shift settings

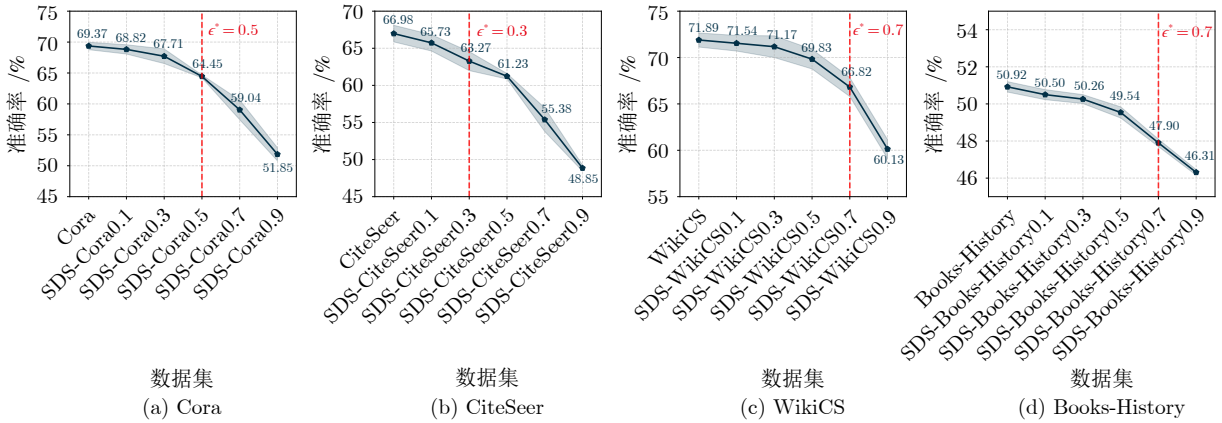



图 4　GraphCLIP+GTrans 在不同域偏移设置下的性能表现

Fig. 4　Performance of GraphCLIP+GTrans under different domain shift settings

集上, 模型性能在偏移强度为 0.5 时开始出现明显骤降; 而在 CiteSeer 数据集上, 该阈值提前至 0.3. 相比之下, WikiCS 与 Books-History 的性能随域偏移增强呈现出更为平缓的下降趋势, 其显著退化通常出现在更高的偏移强度. 这一结果表明, 域适应阈值并非固定常数, 而是与具体数据集的统计特性及其对分布扰动的敏感性密切相关.

为进一步探究现有测试时自适应方法是否能够缓解上述阈值效应, 本文在 GraphCLIP 基础上引入针对图数据的测试时自适应方法 GTrans, 其在不同域偏移强度下的性能表现如图 4 所示. 通过对比 GraphCLIP 与 GraphCLIP+GTrans 可以发现, 引入 GTrans 后, 模型在不同数据集下的整体性能变化趋势与 GraphCLIP 基本一致, 均表现出随分布偏移增强而逐步退化的规律. 在部分数据集 (如 Cora 和 WikiCS) 上, GraphCLIP+GTrans 相较于 GraphCLIP 在部分偏移区间内取得了一定的性能提升, 但该提升并未改变域适应阈值效应的出现; 而在 Books-History 数据集上, 引入 GTrans 不仅未带来性能改善, 反而导致整体性能显著下降, 并使性能显著退化发生于更低的域偏移强度.

从整体趋势来看, 域适应阈值效应在不同数据集和不同模型配置下均呈现出显著的共性特征: 随着分布偏移程度的持续增强, 与原始数据集相比, 模型性能在超过某一临界区间后会出现明显下降, 并在更强偏移条件下持续退化. 同时, 阈值出现的具体位置及退化幅度在不同数据集之间存在显著差异, 反映了数据分布特性对模型跨域泛化能力的影响. 尽管引入现有的测试时自适应方法在部分场景下能够一定程度上提升模型性能, 但其效果具有不稳定性, 且难以从根本上缓解域适应阈值效应, 甚至在部分情况下可能加剧性能退化过程.

### 2.4　测试时自适应

测试时自适应的目标是在不依赖预训练数据的前提下, 利用包含 $N_t$ 个样本的无标签测试集 $\mathcal{D}_{test} = \{x_i\}_{i=1}^{N_t}$ 将主干模型适应至目标分布. 现有方法 [21, 27] 通常将最小化预测熵作为优化目标, 仅更新模型中的少量参数, 以实现主干模型的高效自适应. 常用的目标函数可表示为:

$$\mathrm{Ent}(x_i) = -\sum_{c=1}^{C} p_c \log p_c \tag{7}$$

其中, $p_c$ 是样本 $x_i$ 为第 $c$ 个类别的预测概率; $\mathrm{Ent}(\cdot)$ 表示预测概率分布的信息熵.

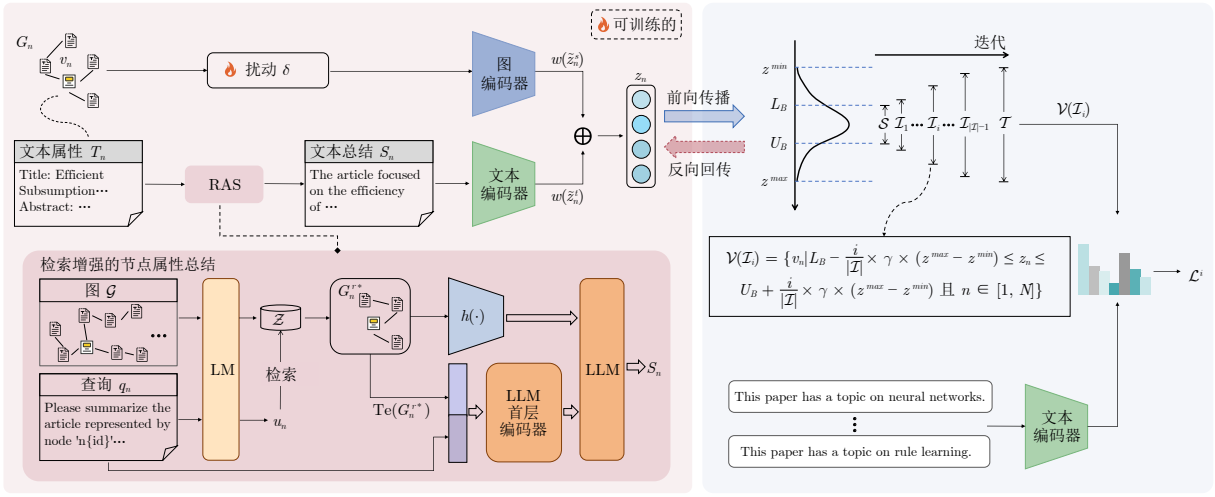
## 3　方法

本节将详细阐述所提出的 DORA, 其总体流程如图 5 所示. 针对 TAG-FM 在分布外场景下面临的实例层语义偏离与域层样本诅咒的双重挑战, 本文提出由实例层与域层协同驱动的统一测试时自适应框架 DORA.

在实例层面, 针对分布外场景下局部结构语义可靠性下降所引发的实例层语义偏离问题, 本文设计检索增强的特征共形融合模块, 该模块首先通过检索增强的节点属性总结组件扩充节点文本描述的信息量; 随后, 在双模态特征共形融合阶段, 将图模态与文本模态融合为对域偏移具有鲁棒性的统一特征. 在域层面, 针对跨域适应过程中普遍存在的样本诅咒效应, 本文进一步引入分而治之测试时自适应模块, 通过分治策略在测试阶段实现渐进式自适应, 进一步提升模型在 OOD 数据上的泛化能力.

### 3.1　检索增强的特征共形融合

当 TAG-FM 面临严重域偏移时, 局部结构的




(a) 检索增强的特征共形融合
(a) Retrieval-augmented feature conformal integration

(b) 分而治之测试时自适应
(b) Divide-and-conquer test-time adaptation

图 5 DORA 的整体框架

Fig. 5 The overall framework of DORA

域间差异性削弱了其语义可靠性, 使得仅依赖自我中心图难以捕捉稳健的判别特征. 然而, 与不稳定的局部结构相比, 节点文本属性凭借其语义不变性, 能够提供关键的语义支撑. 为此, 本文提出检索增强的特征共形融合模块, 以通过引入文本模态来提升实例层表示在分布偏移条件下的语义稳定性. 如图 5(a) 所示, 检索增强的节点属性总结模块首先在全局图结构中检索与目标节点语义高度相关的子图, 并对检索到的文本信息进行语义总结, 以生成更加聚焦的节点文本语义; 在此基础上, 双模态特征共形融合进一步根据不同模态特征在具体实例上的置信度差异, 自适应调节结构模态与文本模态的融合权重, 从而获得对分布偏移更加鲁棒的节点表示.

### 3.1.1 检索增强的节点属性总结

图 2 的实验结果表明, 仅依赖目标节点自身的文本属性时, 其表示在跨域场景下仍存在一定语义偏离; 而在节点编码过程中进一步引入邻域节点的文本信息后, 该偏离现象能够得到显著缓解. 这一现象表明, 邻域节点的文本属性在语义层面为目标节点提供了重要的上下文补充, 有助于稳定其判别语义. 然而, 若直接基于局部拓扑结构聚合邻域文本, 容易引入语义无关或噪声信息. 因此, 本文设计检索增强的节点属性总结模块. 该模块从全局图结构中检索与目标节点语义相似的子图信息, 并对其进行属性总结, 以构建更稳定且更具判别性的节点文本表示. 以引文网络的节点分类任务为例, 网络中的节点表示学术论文, 边则刻画了论文之间的引用关系. RAS 的设计直观模拟了对学术论文进行深入理解的过程: 在阅读论文摘要的基础上, 进一步检索并综合分析其核心参考文献及相关研究. 基于这一思想, 本文将上述过程形式化为由检索和生成组成的两阶段框架.

检索阶段旨在从原图 $\mathcal{G}$ 中抽取与目标节点 $v_n$ 语义高度相关的子图, 通过引入一组语义一致且具有互补信息的参考节点, 为节点表示提供稳定的语义参照, 从而缓解仅依赖局部邻域所导致的语义偏离问题. 具体而言, 首先利用预训练语言模型对 $\mathcal{G}$ 中所有节点的文本属性进行编码, 得到嵌入向量库 $\mathcal{Z}=\{z_1^l, \cdots, z_N^l\}$; 同时, 基于 $v_n$ 的文本属性 $T_n$ 构造查询 $q_n$, 并通过同一预训练语言模型进行编码. 该过程可形式化描述如下:

$$z_n^l = \mathrm{LM}(T_n),\ u_n = \mathrm{LM}(q_n) \tag{8}$$

其中, $\mathrm{LM}(\cdot)$ 表示预训练语言模型; $z_n^l$ 为节点 $v_n$ 的嵌入; $u_n$ 为查询 $q_n$ 的嵌入. 之后基于余弦相似度, 选取前 $k$ 个最相关节点:

$$\mathcal{V}_n^{top} = \underset{z_m^l \in \mathcal{Z}}{\arg\ \mathrm{top}k}\ \mathrm{sim}(u_n, z_m^l) \tag{9}$$

其中, $\arg\ \mathrm{top}k$ 表示从嵌入向量库 $\mathcal{Z}$ 中选取余弦相似度前 $k$ 大的节点集合, $k$ 是超参数. 为保证检索子图的连通性, 本文在节点集 $\mathcal{V}_n^{top}$ 上执行带惩罚的斯坦纳树 (prize-collecting Steiner tree, PCST) 算法[45], 并采用基于排序的节点权重作为奖励, 以固定值 $C_e$ 作为边的代价. 节点奖励可形式化为:

$$\mathrm{prize}(v) = \begin{cases} k-i, & \text{如果 } v \in \mathcal{V}_n^{top} \text{ 且 } v \text{ 是} \\ & \text{第 } i \text{ 大的节点} \\ 0, & \text{否则} \end{cases} \tag{10}$$

由此, 子图检索问题可被形式化为:

$$G_n^{r\,*} = \arg\max_{G_n^r \subseteq \mathcal{G}} \left( \sum_{v \in \mathcal{V}_{G_n^r}} \text{prize}(v) - \text{cost}(G_n^r) \right) \quad (11)$$

其中, $G_n^r$ 是 $\mathcal{G}$ 的一个连通子图; $\mathcal{V}_{G_n^r}$ 表示图 $G_n^r$ 的节点集合; $\text{cost}(G_n^r) = E \times C_e$, $E$ 表示 $G_n^r$ 中的边数.

基于检索子图 $G_n^{r\,*}$ 和查询 $q_n$, 生成阶段利用 LLMs 为节点 $v_n$ 生成信息丰富的文本总结. 为充分利用 $G_n^{r\,*}$ 中的信息, 本文引入一种双分支编码策略: 一方面, 通过编码器 $h(\cdot)$ 对检索子图进行编码, 以捕获其结构信息; 另一方面, 使用文本化函数 $\text{Te}(\cdot)$ 将子图转换为文本形式, 并与查询 $q_n$ 拼接后, 输入至 LLMs 的首层网络 $\text{Embedder}(\cdot)$ 以获取文本嵌入向量. 最终, 将两个分支的输出进行拼接, 输入到后续的 LLMs 中, 生成节点 $v_n$ 的文本总结 $S_n$. 该过程可形式化为:

$$S_n = \text{LLMs}([h(G_n^{r\,*});\ \text{Embedder}([\text{Te}(G_n^{r\,*});\ q_n])]) \quad (12)$$

其中, $h(\cdot)$ 由图编码器与投影头组合而成; $[\cdot\,;\,\cdot]$ 代表拼接操作. 从直观上看, 生成阶段相当于对检索阶段得到的检索子图中的文本信息进行语义聚合与冗余抑制, 使生成的节点文本总结能够更加集中地表征目标节点的核心语义. 需要说明的是, 与现有的主要面向图问答或规划任务的图检索增强生成 (graph retrieval-augmented generation, GraphRAG) 方法不同, 本文关注的是文本属性图基础模型在分布偏移场景下的测试时自适应问题, 其中检索增强生成模块用于构造目标节点的稳健语义表示, 而非直接生成最终任务输出.

#### 3.1.2　双模态特征共形融合

在获得增强的文本语义后, 如何在分布偏移场景下有效融合结构特征与文本特征仍是实例层建模的关键问题. 传统的静态加权融合方式 (如等权求和) 假设各模态具有固定且相同的置信度, 然而在分布偏移场景下, 不同模态及不同节点受分布偏移影响的程度往往存在显著差异, 该假设难以成立[46]. 为此, 本文提出双模态特征共形融合. 其中“共形”指的是特征分布的一致性, 旨在衡量测试样本在特定模态下的特征分布与预训练源域分布的契合程度. 具体而言, DFCI 引入能量分数[47] 作为这种共形程度的量化指标, 对各模态特征的置信度进行自适应估计, 并据此动态调整其融合权重, 使置信度更高的模态在融合过程中占据主导地位, 从而显著提升统一节点特征在分布偏移条件下的鲁棒性.

首先, 为对结构特征进行训练, 本文在初始节点特征中引入可学习扰动 $\delta$, 并依据式 (1) 计算节点 $v_n$ 的结构特征:

$$\tilde{z}_n^s = \text{MLP}(\text{POOL}(g(X_n + \delta,\ A_n,\ P_n))) \quad (13)$$

同时, 将文本总结 $S_n$ 输入 TAG-FM 的文本编码器, 经式 (2) 获得节点 $v_n$ 的文本特征 $\tilde{z}_n^t$.

给定结构特征 $\tilde{z}_n^s$ 与文本特征 $\tilde{z}_n^t$, 本文引入能量分数作为不同模态特征分布偏移程度的估计量, 从而实现融合权重的自适应调整. 具体而言, 能量分数将亥姆霍兹自由能 (Helmholtz free energy) 与样本密度相关联, 其数值越大表示样本越可能处于分布外区域, 从而其对应特征的置信度越低. 基于此, 本文分别对 $\tilde{z}_n^s$ 与 $\tilde{z}_n^t$ 计算能量分数, 具体形式如下:

$$\text{Energy}(\tilde{z}_n^s) = -\tau^s \times \log \sum_{c=1}^{C} \mathrm{e}^{\frac{p_c^s}{\tau^s}} \quad (14)$$

$$\text{Energy}(\tilde{z}_n^t) = -\tau^t \times \log \sum_{c=1}^{C} \mathrm{e}^{\frac{p_c^t}{\tau^t}} \quad (15)$$

其中, $\tau^s$ 和 $\tau^t$ 是不同模态对应的温度系数; $p_c^s$ 和 $p_c^t$ 分别表示结构特征 $\tilde{z}_n^s$ 与文本特征 $\tilde{z}_n^t$ 关于第 $c$ 类标签的对数几率, 其值由式 (3) 计算得到. 之后, 可通过双模态特征共形融合获得节点 $v_n$ 的统一特征:

$$z_n = w(\tilde{z}_n^s) \times \tilde{z}_n^s + w(\tilde{z}_n^t) \times \tilde{z}_n^t \quad (16)$$

其中, $w(\tilde{z}_n^s) = -\text{Energy}(\tilde{z}_n^s)$; $w(\tilde{z}_n^t) = -\text{Energy}(\tilde{z}_n^t)$. 特征 $\tilde{z}_n^s$, $\tilde{z}_n^t$ 的能量分数与它们的置信度呈负相关, 即, $\text{Energy}(\tilde{z}_n^s)$ 越高, 对应 $\tilde{z}_n^s$ 的置信度越低; 反之, $w(\tilde{z}_n^s)$ (或 $w(\tilde{z}_n^t)$) 的值越大, 则表明 $\tilde{z}_n^s$ (或 $\tilde{z}_n^t$) 的置信度越高. 根据式 (16), 置信度更高的特征在融合过程中被赋予更大的权重, 从而显著提升 $z_n$ 对分布偏移的鲁棒性. 相比传统静态加权方法, 基于能量分数的双模态特征共形融合的优势在于: 其能够自适应地反映每个模态特征的置信度, 动态降低分布外模态对融合结果的影响; 同时, 它允许对结构与文本特征分别计算权重, 使双模态信息得到差异化利用.

### 3.2　分而治之测试时自适应

在域层面, 当预训练域与测试域之间存在显著分布偏移时, 直接将主干模型向测试域进行适应, 通常在统计意义上难以成立. 其根本原因在于, 随着域间分布差距的增大, 实现有效跨域适应所需的最小样本量将呈指数级增长, 而实际可用的无标签测试样本规模通常远不足以支撑这一过程. 在此条件下, 现有基于熵最小化的直接适应范式容易出现不稳定优化问题, 表现为明显的域适应阈值效应.

为缓解该问题, 本文构造一个渐进式过渡域序列 $\mathcal{I}=\{\mathcal{I}_0, \cdots, \mathcal{I}_i, \cdots, \mathcal{I}_{|\mathcal{I}|}\}$, 用于桥接预训练域 $\mathcal{S}$ 与测试域 $\mathcal{T}$, 其中 $\mathcal{I}_0=\mathcal{S}$, $\mathcal{I}_{|\mathcal{I}|}=\mathcal{T}$, 且 $i\in[0,\ |\mathcal{I}|]$. 在此基础上, 本文将传统 TTA 过程的 $\mathcal{S}\to\mathcal{T}$ 替换为分而治之的渐进式自适应过程, 即 $\mathcal{S}\to\cdots\to\mathcal{I}_i\to\cdots\to\mathcal{T}$. 由于相邻过渡域之间的分布差异远小于预训练域与测试域之间的差异, 即满足 $\mathrm{KL}(\mathcal{I}_i||\mathcal{I}_{i+1})\ll\mathrm{KL}(\mathcal{S}||\mathcal{T})$, 该方法能够在测试样本有限的情况下有效弥补直接自适应的缺陷. 下文将详细阐述过渡域的构造方法.

鉴于难以直接推导出 $\mathcal{I}_i$ 的解析形式, 本文通过筛选"近似服从 $\mathcal{I}_i$ 分布的样本"来构造过渡域. 若 $\mathcal{S}$ 与 $\mathcal{T}$ 均已知, 则可通过线性插值轻松构造过渡域, 如 $x=(1-i/|\mathcal{I}|)\times x_1+i/|\mathcal{I}|\times x_2$ (其中 $x_1\sim\mathcal{S}$, $x_2\sim\mathcal{T}$) 即可将 $x$ 视作来自 $\mathcal{I}_i$ 的样本. 然而, 在本文的设定中, 预训练域 $\mathcal{S}$ 不可见. 为此, 本文借鉴统计离群点检测[48] 的思想, 依据群体特征的统计信息来筛选服从预训练域分布的样本. 该方法基于以下假设: 正常样本服从某一可估计的概率分布, 而离群点则位于该分布的低概率区域. 通过识别处于低概率区域的节点作为离群样本, 可反向推演出服从预训练域分布的代表性样本. 具体而言, 给定测试域中全部节点特征 $\{z_n\}_{n=1}^N$, 若某节点特征 $z_n$ 满足:

$$L_B\le z_n\le U_B \tag{17}$$

则该节点可被视为服从预训练域分布的样本. 其中, 上/下界定义为:

$$U_B=z^{min}+(1-\gamma)\times(z^{max}-z^{min}) \tag{18}$$

$$L_B=z^{min}+\gamma\times(z^{max}-z^{min}) \tag{19}$$

其中, $z^{min}$ 表示特征的最小值; $z^{max}$ 表示特征的最大值; $\gamma$ 为预定义的超参数. 据此, 可构造出预训练域样本集[2]:

$$\begin{aligned}&\mathcal{V}(\mathcal{I}_0)=\mathcal{V}(\mathcal{S})=\\&\quad\{v_n|L_B\le z_n\le U_B\ \text{且}\ n\in[1,\ N]\}\end{aligned} \tag{20}$$

而整个测试域节点则自然构成测试域样本集, 即

$$\begin{aligned}&\mathcal{V}(\mathcal{I}_{|\mathcal{I}|})=\mathcal{V}(\mathcal{T})=\\&\quad\{v_n|z^{min}\le z_n\le z^{max}\ \text{且}\ n\in[1,\ N]\}\end{aligned} \tag{21}$$

注意到 $|z^{min}-L_B|=|z^{max}-U_B|=\gamma\times(z^{max}-z^{min})$, 通过线性插值可进一步获得服从 $\mathcal{I}_i$ 分布的样本集:

[2] 对任意节点 $v_n$, 均存在对应的统一特征 $z_n$, 故可反向利用 $z_n$ 检索节点 $v_n$.

$$\begin{aligned}\mathcal{V}(\mathcal{I}_i)=\Bigg\{&v_n|L_B-\frac{i}{|\mathcal{I}|}\times\gamma\times(z^{max}-z^{min})\le z_n\le\\&U_B+\frac{i}{|\mathcal{I}|}\times\gamma\times(z^{max}-z^{min})\ \text{且}\ n\in[1,\ N]\Bigg\}\end{aligned} \tag{22}$$

其满足 $\mathcal{V}(\mathcal{I}_0)=\mathcal{V}(\mathcal{S})$ 且 $\mathcal{V}(\mathcal{I}_{|\mathcal{I}|})=\mathcal{V}(\mathcal{T})$. 构造渐进式过渡域序列 $\mathcal{I}=\{\mathcal{I}_0, \cdots, \mathcal{I}_i, \cdots, \mathcal{I}_{|\mathcal{I}|}\}$ 的示意图见图 5(b). 需要注意的是, 这里的 $\mathcal{V}(\mathcal{I}_0)$ 并非预训练域样本, 而是从测试域数据中构造的、在统计意义上近似服从预训练域 $\mathcal{S}$ 分布的样本集, 且该过程未实际引入任何预训练域数据.

此外, 为进一步提升节点 $v_n$ 的结构特征 $\tilde{z}_n^s$ 与文本特征 $\tilde{z}_n^t$ 的对齐程度, 本文引入一项正则化项. 尽管 KL 散度可用于衡量并促进模态间对齐, 但单纯最小化 KL 散度容易因两类特征质量不一致而导致性能退化. 为此, 本文借助统一特征 $z_n$ 协调在模态质量不平衡条件下的结构信息与文本信息, 即

$$\begin{aligned}\mathcal{L}(v_n)=&\mathrm{KL}\left(p^s\|\frac{1}{2}(p^t+p)\right)+\mathrm{KL}\left(p^t\|\frac{1}{2}(p^s+p)\right)=\\&\sum_{c=1}^{C}p_c^s\log\frac{2p_c^s}{p_c^t+p_c}+\sum_{c=1}^{C}p_c^t\log\frac{2p_c^t}{p_c^s+p_c}\end{aligned} \tag{23}$$

其中, $\mathcal{L}(v_n)$ 表示节点 $v_n$ 的损失; $p^s$、$p^t$、$p$ 分别表示节点 $v_n$ 的结构特征 $\tilde{z}_n^s$、文本特征 $\tilde{z}_n^t$ 及统一特征 $z_n$ 所对应的类别预测概率. 实际训练中, $|\mathcal{I}|$ 为一个训练周期内的迭代次数, $i$ 指代第 $i$ 次迭代. 由此, DORA 在第 $i$ 次迭代的最终损失 $\mathcal{L}^i$ 可形式化为:

$$\mathcal{L}^i=\sum_{v_n}\alpha(z_n)\mathbb{I}_{\{v_n\in\mathcal{V}(\mathcal{I}_i)\}}(\mathrm{Ent}(z_n)+\lambda\mathcal{L}(v_n)) \tag{24}$$

其中, 权重系数 $\alpha(z_n)=1/\exp(\mathrm{Ent}(z_n)-\mathrm{Ent}_0)$, $\mathrm{Ent}_0$ 为预设熵阈值[49]; 指示函数 $\mathbb{I}_{\{v_n\in\mathcal{V}(\mathcal{I}_i)\}}$ 表示当节点 $v_n$ 属于 $\mathcal{V}(\mathcal{I}_i)$ 时, 其值为 1, 否则为 0; $\lambda$ 是超参数. DORA 的完整训练伪代码见算法 1.

**算法 1. DORA 算法**

**输入.** 文本属性图 $\mathcal{G}$, 节点 $v_n$ 的自我中心图 $G_n$ 及文本属性 $T_n$, 一个训练周期的迭代次数 $|\mathcal{I}|$.

**输出.** 更新后的参数 $\delta$.

**for** 迭代次数 $i=1$ to $|\mathcal{I}|$ **do**

// 检索增强的特征共形融合模块

1) 根据 $T_n$ 构建查询 $q_n$;

2) 基于 $q_n$ 和 $\mathcal{G}$, 按式 (11) 获取检索子图 $G_n^{r*}$;

3) 基于 $G_n^{r*}$ 和 $q_n$, 按式 (12) 生成文本总结 $S_n$;

4) 按式 (13) 计算带扰动 $\delta$ 的节点 $v_n$ 的结构特征 $\tilde{z}_n^s$;

5) 按式 (2) 对文本总结 $S_n$ 编码, 得到节点 $v_n$ 的文本特征 $z_n^t$;

6) 按式 (16) 融合特征 $z_n^s$ 和 $z_n^t$, 获得节点统一特征 $z_n$;

// 分而治之测试时自适应模块

7) 依照式 (22) 筛选当前过渡域的样本;

8) 基于筛选样本, 按式 (24) 计算损失 $\mathcal{L}^i$;

9) 根据 $\mathcal{L}^i$ 反向传播更新 $\delta$;

**end for**

# 4　实验

## 4.1　实验设置

### 4.1.1　数据集

本文的实验采用如表 1 所示的 7 个经典 TAGs 数据集, 涵盖学术、电子商务、维基百科与社交媒体四大领域, 详细介绍如下.

Cora[50] 是一个引文网络, 包含 2 708 个节点和 5 429 条边. 节点对应学术论文, 边表示论文间的引用关系. 每个节点的文本属性由论文的标题与摘要构成, 节点类别涵盖 7 个研究方向: 案例推理、遗传算法、神经网络、概率方法、强化学习、规则学习及理论.

CiteSeer[51] 也是引文网络, 共 3186 个节点与 4277 条边. 节点对应学术论文, 共划分为 6 大研究方向: 智能体、机器学习、信息检索、数据库、人机交互及人工智能.

Ele-Photo[52] 源自 Amazon Electronics 数据集[53], 节点代表电子产品, 边表示共同购买或共同浏览关系. 节点的文本属性来自点赞数最高的用户评论, 若无点赞记录, 则随机选取一条评论. 该数据集的任务是将产品划分到 12 个类别中.

Ele-Computers[52] 同样来自 Amazon Electronics 数据集[53], 结构和文本属性的构建方式与 Ele-Photo 一致, 但产品被划分为 10 个类别.

Books-History[52] 提取自 Amazon-Books 数据集[53], 聚焦历史类图书. 节点表示书籍, 边表示共同购买或浏览关系, 文本属性包含书名与描述信息, 分类任务涵盖 12 个类别.

表 1　文本属性图数据集统计

Table 1　Statistics of text-attributed graph datasets

| 数据集 | 节点数 | 边数 | 领域 | 类别数 |
|---|---|---|---|---|
| Cora | 2 708 | 5 429 | 学术 | 7 |
| CiteSeer | 3 186 | 4 277 | 学术 | 6 |
| Ele-Photo | 48 362 | 500 928 | 电子商务 | 12 |
| Ele-Computers | 87 229 | 721 081 | 电子商务 | 10 |
| Books-History | 41 551 | 358 574 | 电子商务 | 12 |
| WikiCS | 11 701 | 215 863 | 维基百科 | 10 |
| Instagram | 11 339 | 144 010 | 社交媒体 | 2 |

WikiCS[54] 是一个基于维基百科的基准数据集, 专为图神经网络设计. 节点类别共 10 类, 对应计算机科学的各分支方向.

Instagram[55] 为社交网络数据集, 节点代表用户, 边代表用户之间的关注关系, 预测任务旨在将用户划分为商业账号或普通个人账号.

此外, 为模拟严重域偏移情况, 本文进一步在 Cora、CiteSeer、WikiCS 及 Books-History 四个数据集基础上, 每个数据集分别随机删除 50%、70% 和 90% 的边, 共构造出 12 个严重域偏移的 TAGs 数据集, 将其依次记为 SDS-Cora0.5/0.7/0.9、SDS-CiteSeer0.5/0.7/0.9、SDS-WikiCS0.5/0.7/0.9 与 SDS-Books-History0.5/0.7/0.9. 以 SDS-Cora0.5 为例, 其表示在 Cora 中随机丢弃 50% 的边后得到的数据集.

### 4.1.2　实验细节

DORA 严格遵循标准 TTA 范式的设定: 仅利用无标签测试数据更新主干模型 TAG-FM, 全程不访问任何预训练数据, 且测试域与预训练域相互独立. 具体而言, 本文采用预训练的 GraphCLIP[19] 作为 TAG-FM 的主干框架, 其中图编码器为 Graph-GPS[43], 文本编码器为 all-MiniLM-L6-v2[56]. 在 RAS 阶段, 式 (8) 中的预训练语言模型同样使用 all-MiniLM-L6-v2[56], 式 (12) 中的大语言模型则采用 LLaMA-3.1-8B-Instruct[11], 超参数 $k$ 在 {1, 2, 3, 4} 范围内调优, 边代价 $C_e$ 固定为 0.5. DFCI 模块的温度参数设为 $\tau^s = 10$, $\tau^t = 10$. 根据文献 [48], D&C-TTA 模块中的参数 $\gamma$ 被设置为 0.25, 超参数 $\lambda$ 在 {0.05, 0.5, 5, 50, 500, 5 000} 内选取最优值. 所有实验均使用 Adam 优化器, 批量大小设为 32, 学习率和权重衰减均为 $1 \times 10^5$, 并在 7 个 NVIDIA V100 GPU 上并行训练. 所有实验结果均经 3 次独立运行后取平均值并给出标准差.

表 2 给出了各类别提示, 其中“{class}”为类别名称, “{class_desc}”为类别描述. 表 3 汇总了构造查询 $q_n$ 所用提示模板, 其中“n{id}”表示节点唯一编号, “{desc}”对应其文本属性.

### 4.1.3　基线方法

本文与 21 个基线方法开展对比实验, 这些方法涵盖以下 5 类代表性方案.

1) 预训练语言模型方法: 仅编码文本信息而不利用图结构信息, 包括 BERT[57]、SBERT[58]、DeBERTa[59]、E5[60]、Qwen2-7B-Instruct[61]、Qwen2-72B-Instruct[61]、LLaMA-3.1-8B-Instruct[11] 与 LLaMA-3.1-70B-Instruct[11];

2) 当前最优的 TAGs 方法: 包括 OFA[12]、Zero-

表 2 类别相关提示模板汇总
Table 2 Summary of the class-relevant prompt templates

| 数据集 | 提示模板 |
|---|---|
| Cora | this paper has a topic on {class} {class_desc} |
| CiteSeer | good paper of {class} {class_desc} |
| Ele-Photo | this product belongs to {class} {class_desc} |
| Ele-Computers | is {class} category {class_desc} |
| Books-History | this book belongs to {class} {class_desc} |
| WikiCS | it belongs to {class} research area {class_desc} |
| Instagram | {class} {class_desc} |

G[13]、GraphGPT[14] 与 LLaGA[15];

3) 自监督图表示学习算法: 仅建模图结构信息, 包括 DGI[62]、GRACE[63]、BGRL[64]、GraphMAE[65] 与 G2P2[66];

4) 文本属性图基础模型方法: 在预训练阶段同时编码图结构和节点文本属性信息的 GraphCLIP[19];

5) 测试时自适应方法: 包括最早提出的 TTA 方法 Tent[21] 和当前针对图数据表现最优的 TTA 方法 GTrans[24] 与 Matcha[22].

## 4.2 分布偏移下的性能表现

### 4.2.1 轻度域偏移设置

表 4 展示了 DORA 在 7 个经典 TAGs 数据集上的节点分类准确率. 可以看出, DORA 在所有数据集上均取得当前最优性能, 相比于基线方法 GraphCLIP, 最高提升达 4.69%, 平均提升 2.03%. 进一步分析表明, 仅依赖图结构信息的自监督图表示学习方法因缺乏文本语义支撑, 泛化能力有限; 而 LLaMA 等大规模预训练语言模型凭借海量文本预训练获得了具有竞争力的性能. DORA 通过在测试阶段有效融合图结构与文本语义, 充分挖掘 TAGs 的双模态特性, 从而实现了最优分类表现. 表 5 对比了不同 TTA 方法的性能. 表 5 中的 TTA 方法与 DORA 保持一致, 均以预训练的 GraphCLIP[19] 作为主干模型. Tent[21] 沿用原始实现, 即仅更新批归一化层; GTrans[24] 与 Matcha[22] 分别使用各自的损失函数对可学习扰动 $\delta$ 进行单步更新. 对所有数据集, 每条测试样本仅执行一次适应步骤. 结果表明, DORA 以 1.29% 的平均优势领先目前最优的基线方法 GTrans, 验证了在相同分布偏移 (即预训练域与测试域的 KL 散度保持不变) 且样本受限的条件下, 所提的 D&C-TTA 模块能够显著缓解域差距下的样本诅咒问题, 从而提升 TAG-FM 在 OOD 数据上的泛化能力.

### 4.2.2 严重域偏移设置

为验证 DORA 在严重域偏移场景下的有效性, 本文在构造的 12 个严重域偏移 TAGs 数据集上进行系统实验. 其中, 以 Cora 数据集为例, SDS-Cora0.5、SDS-Cora0.7 与 SDS-Cora0.9 表示域偏移

表 3 构建查询 $q_n$ 的提示模板汇总
Table 3 Summary of prompt templates used to construct the query $q_n$

| 数据集 | 构建查询 $q_n$ 的提示模板 |
|---|---|
| Cora | The textual description of the citation network centered on node 'n{id}' is shown above. Each node in the network represents a scholarly article, and each edge signifies a citation relationship between articles. The textual description of node 'n{id}' is: '{desc}'. Please summarize the core content of the article represented by node 'n{id}' using the information provided above in 200 words: |
| CiteSeer | The textual description of the citation network centered on node 'n{id}' is shown above. Each node in the network represents a scholarly article, and each edge signifies a citation relationship between articles. The textual description of node 'n{id}' is: '{desc}'. Please summarize the core content of the article represented by node 'n{id}' using the information provided above in 200 words: |
| Ele-Photo | The textual description of the Amazon Electronics graph centered on node 'n{id}' is shown above. Each node in the graph represents an electronic product, and each edge signifies frequent co-purchases or co-views. The user review with the most votes of node 'n{id}' is: '{desc}'. Please summarize the information provided above and re-describe the electronic product represented by node 'n{id}' based on the results of the summary in 200 words: |
| Ele-Computers | The textual description of the Amazon Electronics graph centered on node 'n{id}' is shown above. Each node in the graph represents an electronic product, and each edge signifies frequent co-purchases or co-views. The user review with the most votes of node 'n{id}' is: '{desc}'. Please summarize the information provided above and re-describe the electronic product represented by node 'n{id}' based on the results of the summary in 200 words: |
| Books-History | The textual description of the Amazon-Books graph centered on node 'n{id}' is shown above. Each node in the graph represents a book focusing on items labeled as 'History', and each edge signifies frequent co-purchases or co-views between two books. The title and description of the book of node 'n{id}' is: '{desc}'. Please summarize the information provided above and re-describe the history book represented by node 'n{id}' based on the results of the summary in 200 words: |
| WikiCS | The textual description of the Wikipedia-based graph centered on node 'n{id}' is shown above. The textual description of node 'n{id}' is '{desc}'. Please summarize the information provided above and re-describe node 'n{id}' based on the results of the summary in 200 words: |
| Instagram | The textual description of the social network centered on node 'n{id}' is shown above. Each node in the graph signifies a user, and each edge signifies following relationships. The textual description of node 'n{id}' is: '{desc}'. Please summarize the information provided above and re-describe the user represented by node 'n{id}' based on the results of the summary in 200 words: |

表 4　轻度域偏移数据集上的节点分类任务准确率 (%)

Table 4　Accuracy for the node classification task on mild domain shift datasets (%)

| 方法 | Cora | CiteSeer | WikiCS | Instagram | Ele-Photo | Ele-Computers | Books-History | 平均值 |
|---|---|---|---|---|---|---|---|---|
| BERT[57] | 19.56 ± 0.98 | 33.26 ± 2.35 | 29.37 ± 0.00 | 57.02 ± 0.57 | 21.80 ± 0.14 | 13.88 ± 0.29 | 9.95 ± 0.42 | 26.41 |
| SBERT[58] | 54.35 ± 1.26 | 50.47 ± 0.90 | 48.16 ± 0.00 | 48.34 ± 1.23 | 35.96 ± 0.44 | 41.82 ± 0.22 | 30.45 ± 0.19 | 44.22 |
| DeBERTa[59] | 16.42 ± 1.26 | 16.42 ± 1.26 | 15.29 ± 0.00 | 39.81 ± 0.58 | 12.38 ± 0.26 | 10.62 ± 0.15 | 9.70 ± 0.26 | 17.23 |
| E5[60] | 44.65 ± 0.82 | 42.57 ± 0.54 | 31.49 ± 0.00 | 61.28 ± 0.97 | 35.14 ± 0.28 | 16.54 ± 0.14 | 12.92 ± 0.48 | 34.94 |
| Qwen2-7B-Instruct[61] | 61.44 ± 1.29 | 53.57 ± 0.86 | 58.72 ± 0.25 | 39.13 ± 0.78 | 45.55 ± 0.12 | 59.18 ± 0.20 | 23.79 ± 0.34 | 48.77 |
| Qwen2-72B-Instruct[61] | 62.18 ± 0.98 | 60.97 ± 0.87 | 60.91 ± 0.08 | 47.70 ± 0.31 | 52.41 ± 0.39 | 60.88 ± 0.30 | 53.56 ± 0.64 | 56.94 |
| LLaMA-3.1-8B-Instruct[11] | 57.75 ± 1.21 | 53.54 ± 1.71 | 58.32 ± 0.21 | 39.37 ± 1.14 | 34.38 ± 0.25 | 46.98 ± 0.21 | 22.28 ± 0.18 | 44.66 |
| LLaMA-3.1-70B-Instruct[11] | 65.72 ± 1.24 | 62.79 ± 1.24 | 62.82 ± 0.04 | 43.68 ± 0.52 | 51.26 ± 0.53 | 61.62 ± 0.42 | 53.33 ± 0.55 | 57.32 |
| OFA[12] | 37.25 ± 1.38 | 29.64 ± 0.19 | 45.52 ± 1.06 | 32.71 ± 0.16 | 33.03 ± 0.64 | 22.09 ± 0.39 | 16.87 ± 0.93 | 31.02 |
| ZeroG[13] | 62.32 ± 1.91 | 52.55 ± 1.23 | 54.93 ± 0.06 | 48.97 ± 0.78 | 45.12 ± 0.65 | 56.20 ± 0.35 | 40.74 ± 0.65 | 51.55 |
| GraphGPT[14] | 23.25 ± 1.45 | 18.04 ± 1.45 | 6.30 ± 0.26 | 45.12 ± 1.16 | 7.62 ± 0.22 | 29.71 ± 0.83 | 15.92 ± 0.14 | 20.85 |
| LLaGA[15] | 21.44 ± 0.65 | 16.07 ± 1.15 | 2.65 ± 0.72 | 41.12 ± 0.94 | 6.50 ± 0.53 | 23.10 ± 0.33 | 11.17 ± 0.58 | 17.44 |
| DGI[62] | 24.03 ± 1.40 | 18.71 ± 1.22 | 18.86 ± 0.25 | 61.42 ± 1.12 | 13.96 ± 0.17 | 27.12 ± 0.03 | 15.77 ± 0.02 | 25.70 |
| GRACE[63] | 13.69 ± 1.27 | 22.88 ± 1.49 | 16.07 ± 0.32 | 62.23 ± 0.93 | 10.16 ± 0.13 | 10.94 ± 0.12 | 32.39 ± 0.11 | 24.05 |
| BGRL[64] | 31.99 ± 1.06 | 26.50 ± 1.22 | 18.35 ± 0.22 | 61.45 ± 0.82 | 5.21 ± 0.22 | 24.12 ± 0.22 | 16.28 ± 0.35 | 26.27 |
| GraphMAE[65] | 23.25 ± 1.07 | 20.75 ± 0.88 | 12.14 ± 0.20 | 62.39 ± 0.84 | 12.53 ± 0.08 | 8.36 ± 0.06 | 21.76 ± 0.17 | 23.03 |
| G2P2[66] | 41.51 ± 0.78 | 51.02 ± 0.62 | 31.92 ± 0.15 | 52.87 ± 0.78 | 22.21 ± 0.12 | 32.52 ± 0.13 | 26.18 ± 0.25 | 36.89 |
| GraphCLIP[19] | 67.31 ± 1.76 | 63.13 ± 1.13 | 70.19 ± 0.10 | 64.05 ± 0.34 | 53.40 ± 0.64 | 62.04 ± 0.21 | 53.88 ± 0.35 | 62.00 |
| DORA | **70.29** ± 0.54 | **67.82** ± 0.52 | **72.42** ± 0.67 | **64.83** ± 0.25 | **54.93** ± 0.42 | **63.25** ± 0.50 | **54.69** ± 0.43 | **64.03** |

表 5　不同测试时自适应方法在轻度域偏移数据集上的节点分类任务准确率 (%)

Table 5　Accuracy for the node classification task on mild domain shift datasets with different TTA methods (%)

| 方法 | Cora | CiteSeer | WikiCS | Instagram | Ele-Photo | Ele-Computers | Books-History | 平均值 |
|---|---|---|---|---|---|---|---|---|
| Tent[21] | 64.27 ± 1.35 | 63.06 ± 2.49 | 57.27 ± 0.83 | 50.56 ± 0.39 | 44.90 ± 0.14 | 49.35 ± 0.20 | 24.99 ± 0.48 | 50.63 |
| GTrans[24] | 69.37 ± 0.60 | 66.98 ± 1.12 | 71.89 ± 0.77 | 64.02 ± 0.19 | 54.83 ± 0.29 | 61.16 ± 0.37 | 50.92 ± 0.28 | 62.74 |
| Matcha[22] | 69.49 ± 1.13 | 66.78 ± 1.39 | **72.82** ± 0.89 | 64.02 ± 0.50 | 51.34 ± 0.74 | 56.06 ± 0.30 | 45.34 ± 1.35 | 60.84 |
| DORA | **70.29** ± 0.54 | **67.82** ± 0.52 | 72.42 ± 0.67 | **64.83** ± 0.25 | **54.93** ± 0.42 | **63.25** ± 0.50 | **54.69** ± 0.43 | **64.03** |

程度依次加剧. 实验结果如图 6 所示, 随着数据集的域偏移程度增强, 基线方法 GraphCLIP 的性能出现了显著下降; 而 DORA 在不同程度的域偏移设置下均展现出优于 GraphCLIP 的性能. 更为重要的是, 随着域偏移程度的加剧, DORA 相对于 GraphCLIP 的性能优势愈发明显, 如图 6(a) ~ 6(c) 所示. 上述结果充分验证了所提方法在缓解 TAG-FM 的域适应阈值效应方面的有效性, 即便在预训练域与测试域之间存在严重分布偏移时, DORA 仍能保持优异的泛化性能.

### 4.3　消融实验

为评估 DORA 中各组件的贡献, 本节开展消融实验, 结果汇总于表 6, 首行表示基线方法 GraphCLIP 的性能, 最后一行为 DORA 的性能.

1) RFCI 模块的有效性. 表 6 中第二行将 RAS 组件移除, 转而直接融合节点自身文本属性, 与 DORA 相比平均性能下降 0.94%, 表明引入检索增强的节点属性总结对于生成高信息量文本特征至关重要. 表 6 中第三行将 DFCI 组件替换为简单加权平均 $(0.5 \times \tilde{z}_n^s + 0.5 \times \tilde{z}_n^t)$, 性能显著下降, 说明 DFCI 对于准确估计跨模态分布偏移并实现有效特征融合至关重要.

2) D&C-TTA 模块的有效性. 表 6 中第四行将所提出的 D&C-TTA 模块替换为传统的 TTA 策略 (即式 (7)), 这一操作导致性能出现一定程度的下降, 表明传统的 TTA 策略不足以应对 TAG-FM 面临的域偏移问题. 相比之下, 本文提出的 D&C-TTA 能够有效缓解这一问题.

### 4.4　超参数实验

本节对 DORA 的关键超参数 $k$ 与 $\lambda$ 进行敏感性分析, 以评估模型对参数设置的鲁棒性.

1) 相似度选择阈值 $k$. 超参数 $k$ 决定了检索子

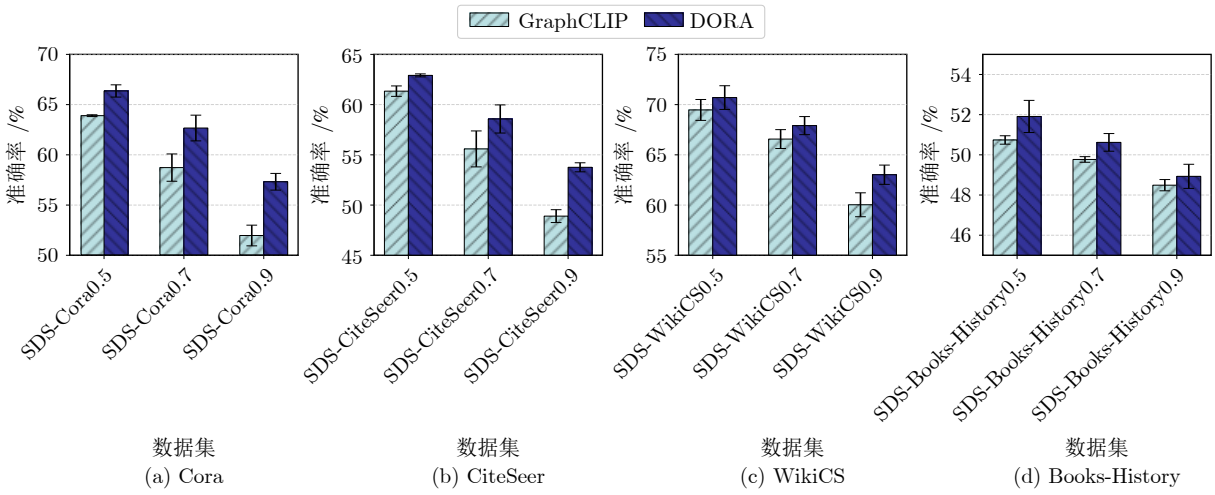



图 6 严重域偏移数据集上的节点分类任务准确率

Fig. 6 Accuracy for the node classification task on severe domain shift datasets

表 6 消融实验结果 (%)

Table 6 The results of the ablation experiment (%)

| RFCI | | D&C-TTA | Cora | WikiCS | Instagram |
|---|---|---|---|---|---|
| RAS | DFCI | | | | |
| × | × | × | 67.31 ± 1.76 | 70.19 ± 0.10 | 64.05 ± 0.34 |
| × | ✓ | ✓ | 69.13 ± 1.23 | 71.39 ± 0.74 | 64.21 ± 0.33 |
| ✓ | × | ✓ | 70.23 ± 0.57 | 72.16 ± 0.68 | 64.60 ± 0.31 |
| ✓ | ✓ | × | 70.05 ± 0.46 | 71.41 ± 1.35 | 64.80 ± 0.28 |
| ✓ | ✓ | ✓ | **70.29** ± 0.54 | **72.42** ± 0.67 | **64.83** ± 0.25 |

图 $G_n^{r*}$ 的大小. 为探究超参数 $k$ 对模型性能的影响, 本文在 Cora 和 CiteSeer 数据集上对 $k$ 进行调优, 实验结果如图 7 所示. 可以看出, 模型性能随 $k$ 值的增大呈现先上升后下降的趋势: 当 $k$ 取值较小时, 适当增加相关节点数量有助于引入更多语义相关信息, 从而提升性能; 然而当 $k$ 过大时, 可能引入不相关或噪声节点, 导致文本特征质量下降, 性能出现衰退. 基于此观察, 本文在主实验中统一设定 $k = 3$, 以在相关性与噪声之间取得平衡.

2) 权重参数 $\lambda$. 参数 $\lambda$ 控制了损失 $\mathcal{L}^i$ 中正则化项的权重. 图 8 展示了在固定 $k = 3$ 的条件下, DORA 在不同 $\lambda$ 取值下的性能表现, 其中横坐标使用对数刻度. 实验结果表明, 尽管 $\lambda$ 取值有所变化, DORA 在 Cora 和 CiteSeer 数据集上的性能均稳定优于基线方法, 且波动范围较小. 这充分说明, DORA 对 $\lambda$ 参数不敏感, 能够在不同的权衡设置下持续带来性能增益, 展现出较强的鲁棒性.

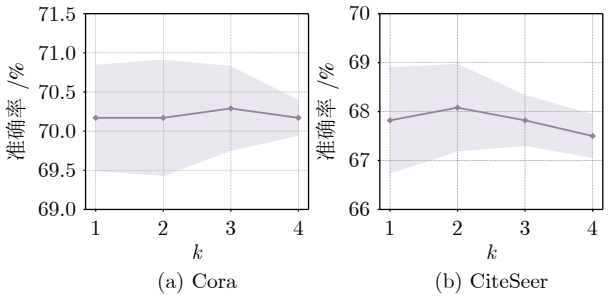



图 7 超参数 $k$ 对模型性能的影响

Fig. 7 The impact of hyperparameter $k$ on model performance

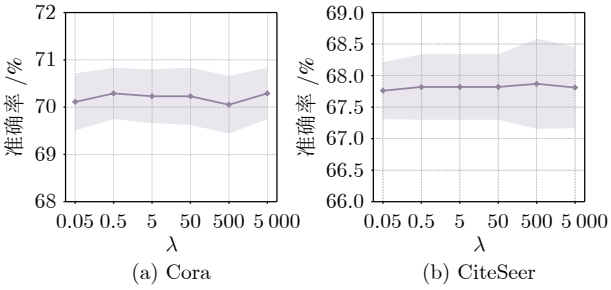


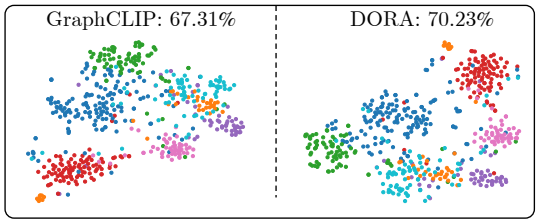

图 8 超参数 $\lambda$ 对模型性能的影响

Fig. 8 The impact of hyperparameter $\lambda$ on model performance

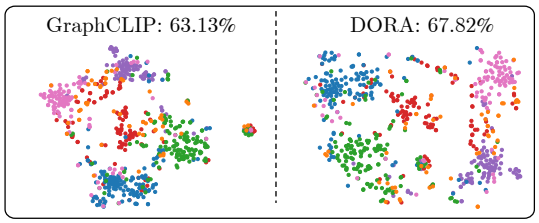
## 4.5 可视化分析

### 4.5.1 轻度域偏移设置

为直观地对比基线方法 GraphCLIP 与本文提出的 DORA, 图 9 给出了它们在 Cora 和 CiteSeer 数据集上的 t-SNE[67] 可视化结果, 其中不同颜色的圆点代表不同类别的样本. 可以看出, GraphCLIP 得到的特征分布呈现以下两个明显缺陷: 1) 不同类别节点间的决策边界模糊; 2) 存在明显离群点. 相比之下, DORA 通过检索增强的特征共形融合模块充分利用 TAGs 中的图结构和文本属性信息, 同时




(a) Cora




(b) CiteSeer

图 9　Cora 和 CiteSeer 数据集上 GraphCLIP 与 DORA 所得特征的 t-SNE 可视化

Fig. 9　t-SNE visualization of features produced by GraphCLIP and DORA on Cora and CiteSeer datasets

通过精心设计的分而治之测试时自适应策略实现渐进式自适应, 从而呈现出更加清晰的类间边界, 各类中心间距显著增大, 且离群点数量大幅减少. 这一可视化结果直观地验证了 DORA 能够有效缓解 TAG-FM 在 OOD 数据上的性能退化.

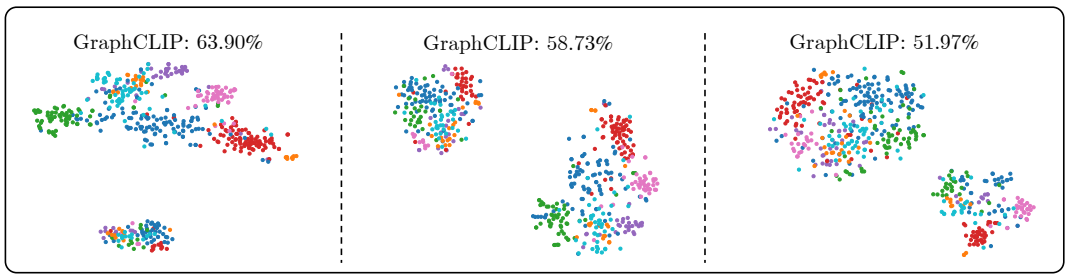
#### 4.5.2　严重域偏移设置

为进一步验证 DORA 在严重域偏移场景下的分布外泛化能力, 本文对基线方法 GraphCLIP 与 DORA 在 SDS-Cora 和 SDS-CiteSeer 系列数据集上得到的特征进行 t-SNE 可视化, 结果见图 10 与图 11. 可以看出, 随着域偏移程度逐渐加剧, GraphCLIP 得到的特征类别边界逐渐模糊, 簇间混叠加剧, 聚类结构逐渐趋于松散. 在相同条件下, DORA 得到的特征聚类形态虽也出现一定程度的松散, 但仍保持相对集中的簇结构, 其簇内分布更为紧凑. 这一优势得益于 DORA 所提出的“检索–分治”机制: 该机制一方面通过检索增强的节点文本属性, 在测试阶段为 TAG-FM 补充有益信息; 另一方面借助分而治之测试时自适应策略, 实现从预训练域到测试域的渐进式自适应, 从而缓解域间严重偏移带来的影响. 实验结果表明, DORA 在不同程度的域偏移下均表现出更优的鲁棒性与泛化能力, 有效减轻了 TAG-FM 在 OOD 数据上所面临的域适应阈值效应.

### 4.6　分而治之测试时自适应的机制分析

为深入剖析本文提出的分而治之测试时自适应方法在提升文本属性图基础模型泛化能力方面的内在机理, 本节从特征分布、梯度更新稳定性以及误差传播行为三个角度, 对其作用机制进行系统分析, 并与非分治策略在稳定性层面进行本质性比较.

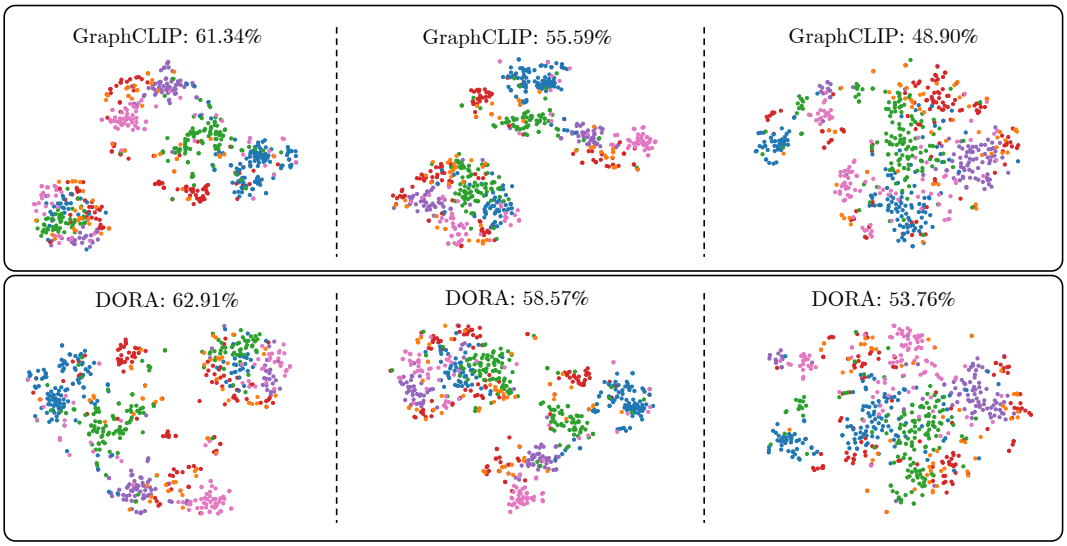
从特征分布的视角看, 分治策略通过构建渐进式过渡域, 重构了模型的自适应优化轨迹. 传统的非分治方法试图在有限样本条件下, 强制模型直接跨越预训练域与测试域之间的整体分布鸿沟, 往往因样本稀缺而难以实现有效的分布适应. 相比之下,




(a) SDS-Cora0.5　(b) SDS-Cora0.7　(c) SDS-Cora0.9

图 10　SDS-Cora 系列严重域偏移数据集上 GraphCLIP 与 DORA 所得特征的 t-SNE 可视化

Fig. 10　t-SNE visualization of features produced by GraphCLIP and DORA on SDS-Cora series datasets with severe domain shifts




(a) SDS-CiteSeer0.5 (b) SDS-CiteSeer0.7 (c) SDS-CiteSeer0.9

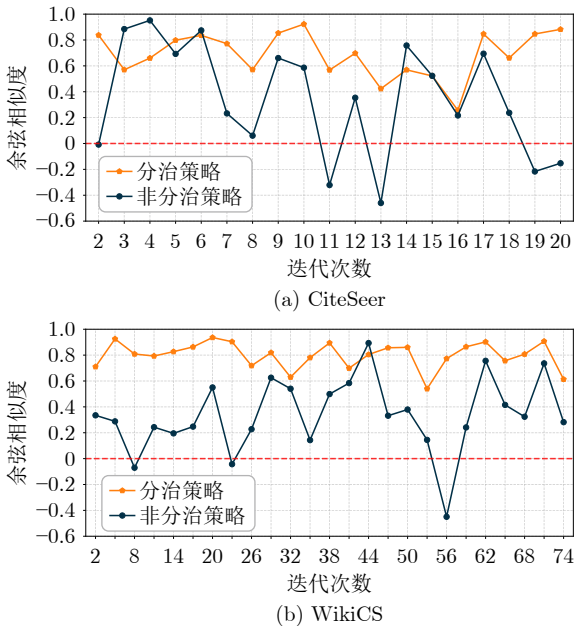
图 11 SDS-CiteSeer 系列严重域偏移数据集上 GraphCLIP 与 DORA 所得特征的 t-SNE 可视化

Fig. 11 t-SNE visualization of features produced by GraphCLIP and DORA on SDS-CiteSeer series datasets with severe domain shifts

分治策略优先在特征空间中选择与预训练域分布相近的样本子集参与自适应, 并随迭代过程动态扩展优化集合, 确保模型在各阶段仅需拟合与当前参数状态邻近的局部特征分布. 该机制有效地将全局域适应任务分解为多步域间差距显著减小的渐进式适应过程, 从而在样本受限场景下显著增强了域泛化的稳健性.

为进一步探究分治策略对优化过程的影响, 本文以 CiteSeer 和 WikiCS 数据集为例, 对分治策略与非分治策略在测试时自适应过程中相邻迭代步间的梯度余弦相似度进行对比分析. 其中, 非分治策略指在保持检索增强的特征共形融合模块不变的前提下, 将本文提出的 D&C-TTA 模块替换为传统的测试时自适应策略 (如式 (7) 所示). 实验结果如图 12 所示, 分治策略在整个测试时自适应过程中保持了较高且相对稳定的梯度相似度, 而非分治策略的梯度相似度则呈现出明显波动, 部分迭代阶段甚至出现负相关现象. 这表明, 非分治策略由于直接进行从预训练域到测试域的跨域适应, 在域偏移较大时易产生方向不一致的梯度更新, 从而导致优化过程失稳; 相比之下, 分治策略通过构建渐进式过渡域, 有效缩小了相邻两个迭代间的域偏移程度, 从而使参数更新在连续迭代之间具有更强的一致性.

此外, 在基于熵最小化的测试时自适应场景下, 高确定性错误样本因其预测分布的低熵特性, 往往被视为高可靠样本, 从而在自适应过程中被持续强化, 成为误差传播的重要来源. 为刻画该现象, 本文将预测类别错误且预测概率大于 0.7 的样本定义为




(a) CiteSeer

(b) WikiCS

图 12 测试时自适应过程中相邻迭代步之间梯度的余弦相似度

Fig. 12 Cosine similarity of gradients between consecutive iterations during test-time adaptation

“高确定性错误样本”, 并统计了测试时自适应过程中该类样本数量的变化情况, 实验结果如图 13 所示. 可以观察到, 非分治策略在适应过程中持续累积了大量高确定性错误样本, 而分治策略则始终将此类样本控制在较低水平. 该结果表明, 分治策略通过降低高确定性错误样本在优化过程中的出现频

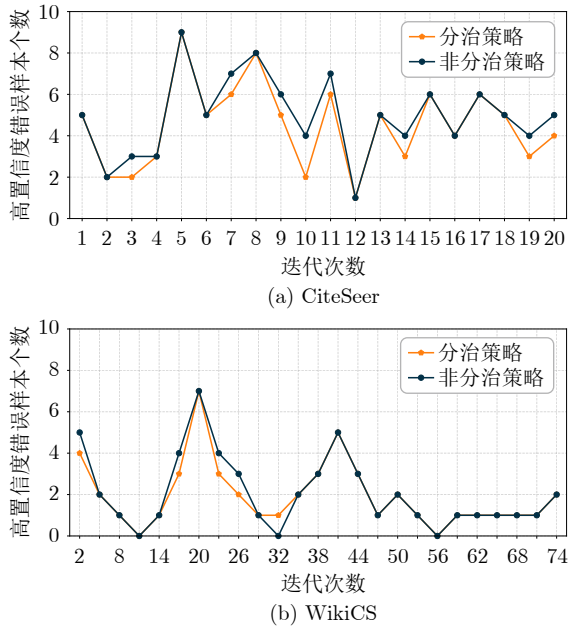



图 13　测试时自适应过程中高确定性错误样本数量

Fig. 13　Number of misclassified samples with high certainty during test-time adaptation

率, 有效抑制了误差在迭代过程中的传播, 从而显著提升了测试时自适应过程的稳定性.

综上所述, 非分治策略默认模型能够在有限测试样本的支持下, 一次性完成从预训练域到测试域的直接自适应. 然而, 当域偏移较为显著时, 这种直接适应方式往往导致梯度更新过程震荡、不稳定, 并引发误差在迭代中的逐步累积. 与之相对, 本文提出的分治策略通过构建渐进式过渡域, 使每一步的梯度更新更为平稳, 并有效抑制了高确定性错误样本带来的误差传播, 从而在面临严重域偏移时, 显著增强了 TAG-FM 在测试时自适应过程的泛化能力和稳定性.

### 4.7　案例分析

为进一步验证 DORA 在缓解实例层语义偏离方面的有效性, 本文从 Cora 数据集中选取两个具有代表性的论文节点进行案例分析, 如图 14 所示. 在这两个案例中, 仅依赖局部结构的基线方法 GraphCLIP 均因局部拓扑噪声而发生误判, 而 DORA 凭借对文本语义的深度挖掘实现了准确预测.

如图 14(a) 所示, 待分类论文 $v_n$ 主要研究“查询委员会方法在计算学习理论框架下的理论性质”, 重点关注“预测误差的收敛行为及其理论下界”, 属于典型的理论类研究工作. 基线方法 GraphCLIP 仅依赖论文节点的自我中心图进行分类决策. 然而, 其自我中心图中包含了大量应用导向的邻居节点, 例如讨论“Active Learning ··· for Text Categorization”以及提及“the C4.5 rule induction program”的文献, 导致局部结构语义呈现出明显偏离. 在缺乏全局语义校正机制的情况下, GraphCLIP 难以有效区分“Rule Learning (规则学习)”与“Theory (理论)”这两个类别之间的细粒度差异, 最终产生错误分类结果. 相比之下, DORA 通过其检索增强机制在全局语义空间中构建检索子图, 成功检索到一组在计算学习理论与查询委员会方法框架下高度相关的论文节点, 为分类决策提供了关键的语义依据. 此外, 基于检索子图生成的节点文本总结中显著包含“positive lower bound”、“prediction error decreases exponentially”等典型理论分析术语, 强化了论文的理论属性表达, 从而有效纠正了基线方法的误判.

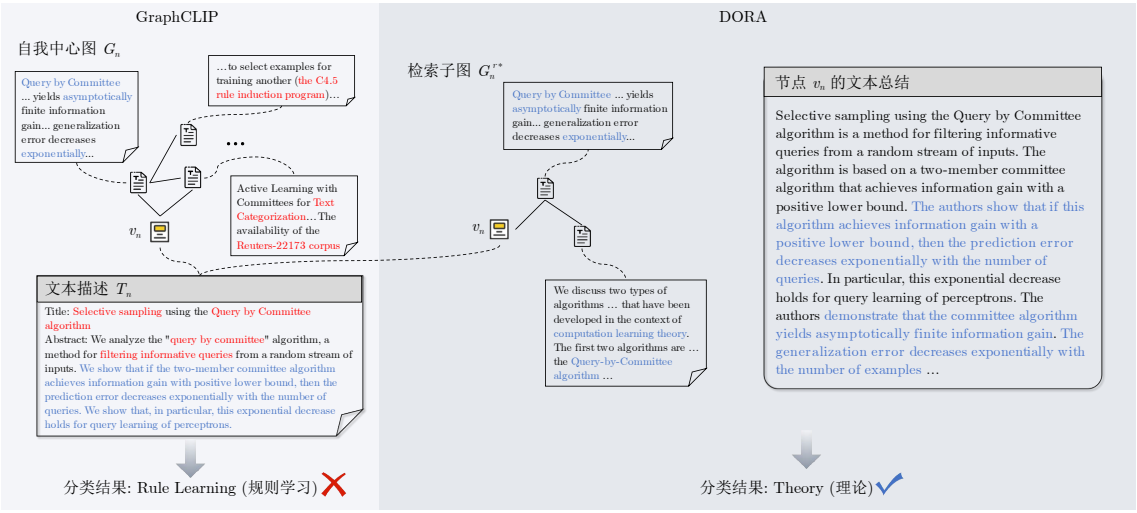
图 14(b) 中的待分类论文聚焦于最大化感知机模型鲁棒性的高效学习算法设计, 并特别针对离散权重引发的组合问题进行优化设计, 属于神经网络领域. 但由于涉及组合优化问题, 该论文的自我中心图中包含了大量关于“Genetic Algorithms”和“evolutionary algorithm”的文献, 这种局部拓扑结构使得 GraphCLIP 得到了错误的预测结果; 而 DORA 构建的检索子图有效过滤了上述噪声邻居, 转而检索到一组与神经网络模型和感知机学习机制高度相关的论文节点. 更重要的是, DORA 生成的文本总结在首句即明确指出了文章的核心目标是最大化感知机的鲁棒性, 并强调了其在神经网络硬件实现中的地位, 从语义层面进一步消除了分类歧义.

上述案例分析表明, 在分布偏移场景下, 仅依赖局部拓扑结构极易引入语义噪声, 导致节点表示发生语义偏离. 而 DORA 通过引入全局检索增强与节点文本总结, 充分挖掘了文本属性在纠正实例层语义偏离中的关键作用.

## 5　适用性与局限性讨论

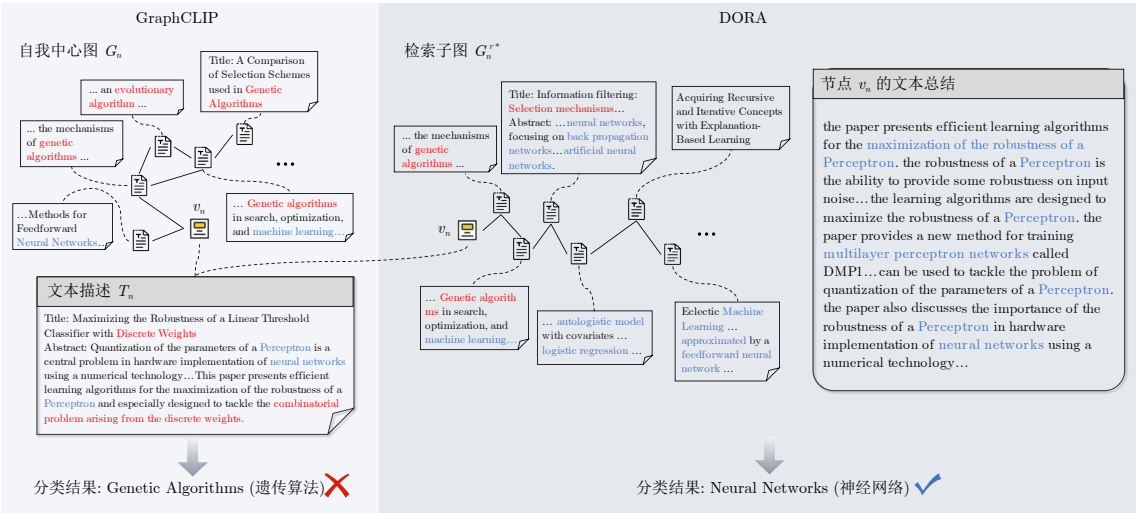
前述实验验证了 DORA 的有效性, 但深入探究其特性与适用边界, 对于指导后续应用具有重要意义. 本节将从方法的架构通用性与潜在的数据依赖局限两个维度对 DORA 进行讨论.

从方法适用性与推广前景来看, DORA 并不受限于单一的模型架构或固定的模态组合, 而是展现出显著的通用性, 这种灵活性源于其模块化的设计思想. 首先, 分而治之测试时自适应模块具有天然的模型无关性, 可独立应用于各类基础模型, 以缓解严重分布偏移场景下的优化不稳定性. 其次, 检




(a) 规则学习节点分类示例

(a) Example of a node classified as rule learning




(b) 遗传算法节点分类示例

(b) Example of a node classified as genetic algorithms

图 14 GraphCLIP 与 DORA 在 Cora 数据集中两个典型节点上的分类结果对比分析

Fig. 14 Comparative analysis of classification results of GraphCLIP and DORA on two representative nodes in the Cora dataset

索增强机制通过引入额外的上下文信息提升测试阶段对样本的表达能力, 其具体实现形式可根据数据模态特征与可用的知识资源进行相应调整. 在此基础上, 针对同时具备图结构与文本语义的场景, DORA 能够引入双模态特征共形融合以强化协同建模能力. 因此, 面向更广泛的推广场景, 该框架具备良好的适应性. 对于仅依赖拓扑结构的传统图基础模型, 虽然跨模态融合机制不再适用, 但这并不妨碍检索增强和分而治之测试时自适应这一核心逻辑的独立运用. 对于 CLIP 等通用的多模态基础模型, 只要明确了适配的检索知识库与检索范式, DORA 所蕴含的“检索–分治”测试时自适应思想同样可以自然扩展. 综上所述, DORA 为不同建模范式下的测试时自适应提供了一套统一而灵活的解决方案.

然而, 尽管 DORA 在多种设置下展现了优异的鲁棒性, 但作为一种利用上下文信息辅助决策的方法, 其性能增益仍受限于数据本身的丰富程度. 具体而言, 检索增强机制的核心在于从邻域中挖掘潜在的结构与语义关联, 若测试域数据的拓扑关系稀疏或其邻域信息包含大量语义噪声, 检索模块所捕获的辅助信息可能缺乏足够的判别力. 在这种极端的数据条件下, 检索增强对样本表示的修正能力会自然衰减, 从而使性能增益受到一定程度的影响.

## 6 结束语

本文聚焦 TAG-FM, 通过探索实验揭示了其在泛化能力方面存在的一个关键局限: 域适应阈值效应, 即当预训练域与测试域之间的分布偏移超过某一临界阈值时, 模型性能将出现显著衰退. 为探究可行的解决路径, 本文系统审视了现有 TTA 方法, 发现其在实例层与域层存在双重缺陷, 因而难以有效支撑 TAG-FM 对 OOD 数据的适应. 针对上述问题, 本文创新性地提出 DORA, 一种面向 TAG-FM 的“检索−分治”测试时自适应方法. 该方法首先通过检索增强的特征共形融合模块获得融合了图结构和文本语义的统一表示, 并采用分而治之测试时自适应策略实现从预训练域到测试域的渐进式自适应. 在 19 个 TAGs 数据集上的实验结果表明, DORA 在不同域偏移程度的数据集上均达到最优性能, 验证了方法的有效性与先进性. 本文的未来工作将探索如何进一步提升 TAG-FM 在动态图或持续域偏移场景下的泛化性, 以期在更复杂的 OOD 任务中实现更优的性能表现.

## 参考文献


1 Rozemberczki B, Allen C, Sarkar R. Multi-scale attributed node embedding. *Journal of Complex Networks*, 2021, **9**(2): Article No. cnab014

2 Hu W H, Fey M, Zitnik M, Dong Y X, Ren H Y, Liu B W, et al. Open graph benchmark: Datasets for machine learning on graphs. arXiv preprint arXiv: 2005.00687, 2021.

3 Pareja A, Domeniconi G, Chen J, Ma T F, Suzumura T, Kanezashi H, et al. EvolveGCN: Evolving graph convolutional networks for dynamic graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020. 5363−5370

4 Wang Zhen-Dong, Xu Zhen-Yu, Li Da-Hai, Wang Jun-Ling. Construction and analysis of meta graph neural network for intrusion detection. *Acta Automatica Sinica*, 2023, **49**(7): 1530−1548 (王振东, 徐振宇, 李大海, 王俊岭. 面向入侵检测的元图神经网络构建与分析. 自动化学报, 2023, **49**(7): 1530−1548)

5 Chen Lu, Shang Jia-Xing, Liu Da-Jiang, Zhang Yu-Fang, Ni Wan-Cheng. An interpretable wargame situation prediction method based on heterogeneous graph neural networks. *Acta Automatica Sinica*, 2025, **51**(6): 1248−1260 (陈露, 尚家兴, 刘大江, 张玉芳, 倪晚成. 基于异构图神经网络的可解释兵棋态势预测方法. 自动化学报, 2025, **51**(6): 1248−1260)

6 Li Y X, Hooi B. Prompt-based zero- and few-shot node classification: A multimodal approach. arXiv preprint arXiv: 2307.11572, 2023.

7 Tang Y R, Qiu R H, Liu Y L, Li X, Huang Z. CaseGNN: Graph neural networks for legal case retrieval with text-attributed graphs. In: Proceedings of the 46th European Conference on Information Retrieval. Glasgow, UK: Springer, 2024. 80−95

8 He X X, Tian Y J, Sun Y F, Chawla N V, Laurent T, LeCun Y, et al. G-retriever: Retrieval-augmented generation for textual graph understanding and question answering. arXiv preprint arXiv: 2402.07630, 2024.

9 Xu Zheng-Fei, Xin Xin. An empirical study of Chinese entity linking based on large language model. *Acta Automatica Sinica*, 2025, **51**(2): 327−342 (徐正斐, 辛欣. 基于大语言模型的中文实体链接实证研究. 自动化学报, 2025, **51**(2): 327−342)

10 Qin Long, Wu Wan-Sen, Liu Dan, Hu Yue, Yin Quan-Jun, Yang Dong-Sheng, et al. Autonomous planning and processing framework for complex tasks based on large language models. *Acta Automatica Sinica*, 2024, **50**(4): 862−872 (秦龙, 武万森, 刘丹, 胡越, 尹全军, 阳东升, 等. 基于大语言模型的复杂任务自主规划处理框架. 自动化学报, 2024, **50**(4): 862−872)

11 Vavekanand R, Sam K. LLaMA 3.1: An in-depth analysis of the next-generation large language model [Online], available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6139407, July 1, 2026

12 Liu H, Feng J R, Kong L C, Liang N Y, Tao D C, Chen Y X, et al. One for all: Towards training one graph model for all classification tasks. arXiv preprint arXiv: 2310.00149, 2024.

13 Li Y H, Wang P S, Li Z X, Yu J X, Li J. ZeroG: Investigating cross-dataset zero-shot transferability in graphs. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain: ACM, 2024. 1725−1735

14 Tang J B, Yang Y H, Wei W, Shi L, Su L X, Cheng S Q, et al. GraphGPT: Graph instruction tuning for large language models. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington, USA: ACM, 2024. 491−500

15 Chen R J, Zhao T, Jaiswal A, Shah N, Wang Z Y. LLaGA: Large language and graph assistant. In: Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria: JMLR, 2024. Article No. 306

16 Rosenfeld J S. *Scaling Laws for Deep Learning*. Cambridge: Massachusetts Institute of Technology, 2021.

17 Hernandez D, Kaplan J, Henighan T, McCandlish S. Scaling laws for transfer. arXiv preprint arXiv: 2102.01293, 2021.

18 Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning. Virtual Event: PMLR, 2021. 8748−8763

19 Zhu Y, Shi H Z, Wang X T, Liu Y C, Wang Y K, Peng B C, et al. GraphCLIP: Enhancing transferability in graph foundation models for text-attributed graphs. In: Proceedings of the ACM on Web Conference 2025. Sydney, Australia: ACM, 2025. 2183−2197

20 Han Z Y, Gui X J, Sun H L, Yin Y L, Li S. Towards accurate and robust domain adaptation under multiple noisy environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, **45**(5): 6460−6479

21 Wang D Q, Shelhamer E, Liu S T, Olshausen B A, Darrell T. Tent: Fully test-time adaptation by entropy minimization. In: Proceedings of the 9th International Conference on Learning Representations. Virtual Event: OpenReview.net, 2021.

22 Bao W X, Zeng Z C, Liu Z N, Tong H H, He J R. Matcha: Mitigating graph structure shifts with test-time adaptation. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025.

23 Ju M X, Zhao T, Yu W H, Shah N, Ye Y F. GRAPHPATCHER: Mitigating degree bias for graph neural networks via test-time augmentation. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2023. Article No. 2434

24 Jin W, Zhao T, Ding J Y, Liu Y Z, Tang J L, Shah N. Empowering graph representation learning with test-time graph transformation. In: Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: OpenReview.net, 2023.

25 Chatterjee S, Diaconis P. The sample size required in importance sampling. *The Annals of Applied Probability*, 2018, **28**(2): 1099−1135

26 McAllester D, Stratos K. Formal limitations on the measurement of mutual information. In: Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics. Palermo, Italy: PMLR, 2020. 875−884

27 Su G X, Wang H C, Wang J W, Zhang W J, Zhang Y, Pei J. Large language models meet text-attributed graphs: A survey of integration frameworks and applications. arXiv preprint arXiv: 2510.21131, 2025.

28 Beiranvand A, Vahidipour S M. Integrating structural and semantic signals in text-attributed graphs with BiGTex. arXiv

preprint arXiv: 2504.12474, 2025.

29 Khoshraftar S, Abedini N, Hajian A. GraphiT: Efficient node classification on text-attributed graphs with prompt optimized LLMs. In: Proceedings of the ACM on Web Conference 2025. Sydney, Australia: ACM, 2025. 1824−1829

30 Zhang Q Y, Bian Y T, Kong X K, Zhao P L, Zhang C Q. COME: Test-time adaption by conservatively minimizing entropy. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025.

31 Lee J, Jung D, Lee S, Park J, Shin J, Hwang U, et al. Entropy is not enough for test-time adaptation: From the perspective of disentangled factors. In: Proceedings of the 12th International Conference on Learning Representations. Vienna, Austria: OpenReview.net, 2024.

32 Hu Y H, Qiao C Y, Geng X, Xu N. Selective label enhancement learning for test-time adaptation. In: Proceedings of the 13th International Conference on Learning Representations. Singapore: OpenReview.net, 2025.

33 Chen G Z, Zhang J Y, Xiao X, Li Y. GraphTTA: Test time adaptation on graph neural networks. arXiv preprint arXiv: 2208.09126, 2022.

34 Wang Y Q, Li C Z, Zhao J N, Li R. Test-time training for graph neural networks. In: Proceedings of the 30th International Conference on Database Systems for Advanced Applications. Singapore: Springer, 2026. 349−364

35 Mao H T, Du L, Zheng Y J, Fu Q, Li Z L, Chen X, et al. Source free graph unsupervised domain adaptation. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. Merida, Mexico: ACM, 2024. 520−528

36 Zheng X, Huang W, Zhou C, Li M, Pan S R. Test-time graph neural dataset search with generative projection. In: Proceedings of the 42nd International Conference on Machine Learning. Vancouver, Canada: PMLR, 2025.

37 Zhao Y S, Zhang Q X, Luo X, Luo J Y, Ju W, Xiao Z P, et al. Test-time adaptation on graphs via adaptive subgraph-based selection and regularized prototypes. In: Proceedings of the 42nd International Conference on Machine Learning. Vancouver, Canada: PMLR, 2025.

38 Gadermayr M, Eschweiler D, Klinkhammer B M, Boor P, Merhof D. Gradual domain adaptation for segmenting whole slide images showing pathological variability. In: Proceedings of the 8th International Conference on Image and Signal Processing. Cherbourg, France: Springer, 2018. 461−469

39 Kumar A, Ma T Y, Liang P. Understanding self-training for gradual domain adaptation. In: Proceedings of the 37th International Conference on Machine Learning. Virtual Event: PMLR, 2020. 5468−5479

40 Chen H Y, Chao W L. Gradual domain adaptation without indexed intermediate domains. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual Event: Curran Associates Inc., 2021. Article No. 627

41 Lei P I, Chen X M, Sheng Y J, Liu Y Y, Gong Z G, Yang Q. Gradual domain adaptation for graph learning. arXiv preprint arXiv: 2501.17443, 2025.

42 Zeng Z C, Qiu R Z, Bao W X, Wei T X, Lin X, Yan Y C, et al. Pave your own path: Graph gradual domain adaptation on fused Gromov-Wasserstein geodesics. arXiv preprint arXiv: 2505.12709, 2026.

43 Rampáček L, Galkin M, Dwivedi V P, Luu A T, Wolf G, Beaini D. Recipe for a general, powerful, scalable graph transformer. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2022. 14501−14515

44 Hu Zheng-Ping, Meng Peng-Quan. Graph presentation random walk salient object detection algorithm based on global isolation and local homogeneity. *Acta Automatica Sinica*, 2011, **37**(10): 1279−1284
(胡正平, 孟鹏权. 全局孤立性和局部同质性图表示的随机游走显著目标检测算法. 自动化学报, 2011, **37**(10): 1279−1284)

45 Bienstock D, Goemans M X, Simchi-Levi D, Williamson D. A note on the prize collecting traveling salesman problem. *Mathematical Programming*, 1993, **59**(1−3): 413−420

46 Han Z B, Zhang C Q, Fu H Z, Zhou J T. Trusted multi-view classification with dynamic evidential fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, **45**(2): 2551−2566

47 Liu W T, Wang X Y, Owens J D, Li Y X. Energy-based out-of-distribution detection. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2020. Article No. 1802

48 Vinutha H P, Poornima B, Sagar B M. Detection of outliers using interquartile range technique from intrusion dataset. In: Proceedings of the 6th International Conference on FICTA. Singapore: Springer, 2018. 511−518

49 Niu S C, Wu J X, Zhang Y F, Chen Y F, Zheng S J, Zhao P L, et al. Efficient test-time model adaptation without forgetting. In: Proceedings of the 39th International Conference on Machine Learning. Baltimore, USA: PMLR, 2022. 16888−16905

50 Sen P, Namata G, Bilgic M, Getoor L, Gallagher B, Eliassi-Rad T. Collective classification in network data. *AI Magazine*, 2008, **29**(3): 93−106

51 Giles C L, Bollacker K D, Lawrence S. CiteSeer: An automatic citation indexing system. In: Proceedings of the 3rd ACM Conference on Digital Libraries. Pittsburgh, USA: ACM, 1998. 89−98

52 Yan H, Li C Z, Long R S, Yan C, Zhao J N, Zhuang W W, et al. A comprehensive study on text-attributed graphs: Benchmarking and rethinking. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc., 2023. Article No. 755

53 Ni J, Li J C, McAuley J. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China: ACL, 2019. 188−197

54 Mernyei P, Cangea C. Wiki-CS: A wikipedia-based benchmark for graph neural networks. arXiv preprint arXiv: 2007.02901, 2020.

55 Huang X W, Han K Q, Yang Y, Bao D Z, Tao Q J, Chai Z W, et al. Can GNN be good adapter for LLMs? In: Proceedings of the ACM Web Conference 2024. Singapore: ACM, 2024. 893−904

56 Reimers N, Gurevych I. Making monolingual sentence embeddings multilingual using knowledge distillation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Virtual Event: ACL, 2020. 4512−4525

57 Devlin J, Chang M W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, USA: ACL, 2019. 4171−4186

58 Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: ACL, 2019. 3982−3992

59 He P C, Liu X D, Gao J F, Chen W Z. DeBERTa: Decoding-enhanced Bert with disentangled attention. In: Proceedings of the 9th International Conference on Learning Representations. Virtual Event: OpenReview.net, 2021.

60 Wang L, Yang N, Huang X L, Jiao B X, Yang L J, Jiang D X, et al. Text embeddings by weakly-supervised contrastive pre-training. arXiv preprint arXiv: 2212.03533, 2022.

61 Yang A, Yang B S, Zhang B C, Hui B Y, Zheng B, Yu B W, et al. Qwen2.5 technical report. arXiv preprint arXiv: 2412.15115, 2024.

62 Velickovic P, Fedus W, Hamilton W L, Liò P, Bengio Y, Hjelm R D. Deep graph infomax. In: Proceedings of the 7th International Conference on Learning Representations. Barcelona, Spain: OpenReview.net, 2019.

63 Zhu Y Q, Xu Y C, Yu F, Liu Q, Wu S, Wang L. Deep graph contrastive representation learning. arXiv preprint arXiv: 2006.04131, 2020.

64 Thakoor S, Tallec C, Azar M G, Munos R, Veličković P, Valko M. Bootstrapped representation learning on graphs. arXiv pre-

print arXiv: 2102.06514, 2021.

65 Hou Z Y, Liu X, Cen Y K, Dong Y X, Yang H X, Wang C J, et al. GraphMAE: Self-supervised masked graph autoencoders. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington, USA: ACM, 2022. 594–604

66 Wen Z H, Fang Y. Augmenting low-resource text classification with graph-grounded pre-training and prompting. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. Taipei, China: ACM, 2023. 506–516

67 van der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 2008, **9**(86): 2579–2605

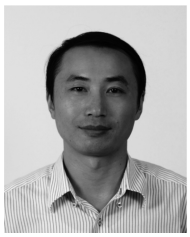
**靳毅凡** 中国科学院软件研究所博士研究生. 2020 年获得中南大学学士学位. 主要研究方向为图表示学习.
E-mail: yifan2020@iscas.ac.cn
(**JIN Yi-Fan** Ph.D. candidate at the Institute of Software, Chinese Academy of Sciences. She received her bachelor degree from Central South University in 2020. Her main research interest is graph representation learning.)

**李 懿** 中国科学院软件研究所博士研究生. 2021 年获得福州大学学士学位. 主要研究方向为多模态表征学习.
E-mail: liyi2022@iscas.ac.cn
(**LI Yi** Ph.D. candidate at the Institute of Software, Chinese Academy of Sciences. He received his bachelor degree from Fuzhou University in 2021. His main research interest is multimodal representation learning.)

**宋 飞** 中国科学院软件研究所博士研究生. 2021 年获得河南大学学士学位. 主要研究方向为多模态提示学习.
E-mail: songfei2022@iscas.ac.cn
(**SONG Fei** Ph.D. candidate at the Institute of Software, Chinese Academy of Sciences. She received her bachelor degree from Henan University in 2021. Her main research interest is multimodal prompt learning.)

**王 瑞** 中国科学院软件研究所高级工程师. 2012 年获得山东大学硕士学位. 主要研究方向为深度强化学习, 多媒体技术.
E-mail: wangrui@iscas.ac.cn
(**WANG Rui** Senior engineer at the Institute of Software, Chinese Academy of Sciences. She received her master degree from Shandong University in 2012. Her research interests include deep reinforcement learning and multimedia technology.)

**李江梦** 中国科学院软件研究所助理研究员. 2023 年获得中国科学院软件研究所博士学位. 主要研究方向为可信多模态信息融合. 本文通信作者.
E-mail: jiangmeng2019@iscas.ac.cn
(**LI Jiang-Meng** Assistant researcher at the Institute of Software, Chinese Academy of Sciences. He received his Ph.D. degree from the Institute of Software, Chinese Academy of Sciences in 2023. His main research interest is trustworthy multimodal information fusion. Corresponding author of this paper.)



**刘立祥** 中国科学院软件研究所研究员. 2002 年获得上海交通大学博士学位. 主要研究方向为天基综合信息系统智能组网, 大规模网络一体化系统地面验证和复杂信息系统体系结构. E-mail: lixiang@iscas.ac.cn
(**LIU Li-Xiang** Researcher at the Institute of Software, Chinese Academy of Sciences. He received his Ph.D. degree from Shanghai Jiao Tong University in 2002. His research interests include intelligent networking of space-based integrated information systems, ground verification of large-scale network integrated system, and complex information system architecture.)

**孙富春** 清华大学计算机科学与技术系教授. 1997 年获得清华大学博士学位. 主要研究方向为人工智能, 智能控制与机器人和认知系统中的信息感知与处理.
E-mail: fcsun@mail.tsinghua.edu.cn
(**SUN Fu-Chun** Professor in the Department of Computer Science and Technology, Tsinghua University. He received his Ph.D. degree from Tsinghua University in 1997. His research interests include artificial intelligence, intelligent control and robotics, and information sensing and processing in cognitive systems.)