

◆特邀栏目◆

基于图多层感知机的节点分类算法*

文竹¹, 袁立宁^{1,2}, 黄伟^{3**}, 黄琬雁¹, 莫嘉颖¹, 冯文刚²

(1. 广西警察学院信息技术学院, 广西南宁 530028; 2. 中国人民公安大学国家安全学院, 北京 100038; 3. 南宁职业技术学院人工智能学院, 广西南宁 530008)

摘要:多数图神经网络(Graph Neural Networks, GNN)通过设计复杂的节点信息传递和聚合方式, 以提升节点分类等图分析任务的实验表现, 而本文提出了一种无需信息传递和聚合的图多层感知机(Multi-Layer Perceptron, MLP)模型 A&T-MLP, 利用属性和拓扑信息引导的对比损失来增强模型表征能力。A&T-MLP 首先使用属性矩阵和邻接矩阵计算节点间的属性和拓扑相似度; 然后使用基于相似度信息引导的对比损失, 增大特征空间中相似节点的一致性和不相似节点的差异性; 最后构建多层感知机模型并引入交叉熵损失进行端到端训练。在节点分类任务中, A&T-MLP 表现优于基线模型, Wikipedia 数据集上的 Micro-F1 和 Macro-F1 相较 GNN 模型图卷积网络(Graph Convolutional Network, GCN)提升了 15.86% 和 13.64%。实验结果表明, A&T-MLP 能够通过对比损失保留丰富原始图的信息, 提升模型性能。此外, A&T-MLP 在处理拓扑信息不准确的图数据时具有较为明显的优势, 即使在缺失 80% 拓扑信息的极端情况下, 其实验表现依然优于基线模型。

关键词:图神经网络; 多层感知机; 节点属性; 对比学习; 节点分类

中图分类号: TP183 文献标识码: A 文章编号: 1005-9164(2023)05-0942-09

DOI: 10.13656/j.cnki.gxkx.20231121.013

图是一种常用的数据形式, 用于表征实体间的关系, 如引文网络中论文之间的互引信息、社交网络账号之间的关注信息等。节点分类^[1]是利用图中已知的拓扑结构、节点属性以及节点标签等信息对无标签节点进行类别预测。因此, 节点分类的关键在于如何

提取原始图中蕴含的特征信息。图神经网络(Graph Neural Networks, GNN)^[2]是一类高效的图特征学习方法, 它通过传递和聚合节点间信息来生成保留拓扑结构和属性信息的低维节点表示, 并将其用于节点分类等下游图分析任务。

收稿日期: 2023-09-08

修回日期: 2023-10-19

* 广西哲学社会科学规划研究课题(21FGL027), 广西法学会法学研究课题(GFKT2023-C3), 广西警察学院校级科研项目(2022KYZ17)和中央高校基本科研业务费专项资金项目(2022JKF02002)资助。

【第一作者简介】

文竹(1982-), 女, 硕士, 副教授, 主要从事机器学习、大数据技术研究。

【**通信作者简介】

黄伟(1979-), 男, 硕士, 副教授, 主要从事人工智能研究, E-mail: hweiwz@foxmail.com。

【引用本文】

文竹, 袁立宁, 黄伟, 等. 基于图多层感知机的节点分类算法[J]. 广西科学, 2023, 30(5): 942-950.

WEN Z, YUAN L N, HUANG W, et al. Node Classification Based on Graph Multi-Layer Perceptron [J]. Guangxi Sciences, 2023, 30(5): 942-950.

近年来,已有许多研究对 GNN 进行改进,这些改进一般通过引入高效的信息传递和转换方式以增强 GNN 的表征能力。图卷积网络(Graph Convolutional Network, GCN)^[3]将经典的卷积神经网络从欧式数据(图像、音频)推广到非欧式数据(图),实现了对拓扑结构和节点属性信息的编码,其在计算过程中仅使用节点的一阶邻域进行聚合,这提高了模型层间信息的传播效率。简化图卷积(Simplified Graph Convolution, SGC)^[4]改进了 GCN 的编码策略,通过直接使用线性网络编码原始图信息以减少模型参数,降低模型复杂度;图注意力网络(Graph Attention Network, GAT)^[5]在 GCN 的基础上引入注意力机制,为一阶邻域中各节点分配不同的权值,使信息聚合过程能够更加关注影响力较大的高权值邻居节点;GATv2^[6]在 GAT 的基础上引入动态注意力机制,修改了 GAT 中权值计算过程,构建近似注意函数,从而提升了 GAT 的表征能力;结构感知路径聚合图神经网络(PathNet)^[7]使用最大熵路径采集器采样节点上下文路径,然后引入结构感知循环单元学习上下文路径中蕴含的语义信息。由于浅层 GNN 架构限制了模型学习高阶邻域信息的能力,而深层 GNN 架构容易出现过平滑^[8],为解决以上问题,Zhang 等^[9]提出的浅层双仿射图卷积网络(Shallow Biaffine Graph Convolutional Networks, BAGCN)不仅能够学习节点对远距离邻居的直接依赖性,还能通过一阶消息传递捕获多跳邻居的特征信息。

与上述通过复杂的节点信息传递和聚合方式提升模型性能的研究思路不同,本文提出了基于多层感知机(Multi-Layer Perceptron, MLP)^[10]的 A&T-MLP。A&T-MLP 无需信息传递和聚合,而是通过属性和拓扑信息引导的对比损失以保留原始图相关特征。具体来说,A&T-MLP 首先使用属性矩阵和邻接矩阵计算节点间的相似度信息;然后使用基于相似度信息引导的对比损失,优化 MLP 隐藏层生成的特征表示;最后引入交叉熵损失对 A&T-MLP 进行端到端训练。

1 相关工作

1.1 MLP

MLP 是一种基于前馈神经网络的深度学习模型,通常由输入层、隐藏层和输出层构成,相邻两层之间所有输入神经元都和输出神经元相连,因此也称为全连接神经网络。其中,输入层用于接收原始数据,

隐藏层利用权重系数和激活函数对数据进行特征提取和非线性变换,输出层利用隐藏层生成的特征向量得到预测结果。MLP 具有较强的表征能力,可通过特定的损失函数和反向传播算法对其进行训练优化。MLP 能够自动地提取数据中蕴含的潜在信息,因此可用于分类、回归等机器学习任务。

在计算机视觉和自然语言处理领域,MLP 被广泛应用于语音识别、图像识别、语义分割等众多任务。Hou 等^[11]提出了基于 MLP 的高效视觉识别架构,使用线性投影分别对高度和宽度特征表示进行编码,捕获长距离依赖关系,进而提升视觉识别网络的性能。Gong 等^[12]提出了基于 Cycle-MLP 的 DriftNet 模型,该模型通过神经元渐进激活以促进空间和光谱信息的融合,进而提升了高光谱图像像素分类的实验表现。Sun 等^[13]提出了多模态情绪分析模型 CubeMLP, CubeMLP 将所有模态特征作为输入,并将输入分配到 3 个不同的 MLP 单元提取特征,最后对混合多模态特征进行展平以实现情绪分析任务。在图分析领域,Hu 等^[14]提出了 Graph-MLP, Graph-MLP 通过拓扑结构中的邻接关系构建保留拓扑信息的对比损失,使仅使用节点属性作为输入的 MLP 能够匹配 GCN 在节点分类任务中的实验性能。

上述研究为 MLP 在其他领域的应用提供了理论基础,其性能也能够匹配多数复杂模型的实验表现。因此,本文在已有研究的基础上,提出了一种基于属性和拓扑信息增强 MLP 模型 A&T-MLP,并应用于节点分类任务。

1.2 对比学习

对比学习是从大量未标记的数据中提取可转移的特征,并将其扩展为特定的自监督信息^[15],使模型在训练过程中能够对相似样本和不相似样本进行区分,因此对比学习的核心思想就是增大特征空间中同类数据表示的相似性和不同类数据表示的差异性。分类任务中常见的对比损失函数的表达式为

$$\text{loss}(x_i) = -\log \frac{\exp(s_{i,i})}{\sum_{k=1}^N \mathbf{1}_{[k \neq i]} \exp(s_{i,k}) + \exp(s_{i,i})}, \quad (1)$$

式中, $s_{i,i}$ 表示正样本, $s_{i,k}$ 表示负样本。通过对比损失调整,使第 i 个数据与 $s_{i,i}$ 之间的相似度尽可能大,与 $s_{i,k}$ 之间的相似度尽可能小,从而提升模型表征原始数据信息的能力。

近年来,基于对比学习增强的图深度学习模型已成为研究的热点方向。除 Graph-MLP 模型,Wang

等^[16]提出了基于集群感知监督对比损失的 Cluster-SCL, ClusterSCL以节点聚类分布的形式保留图信息。Xia等^[17]提出了一种无需数据增强的 GNN 对比学习框架 SimGRACE, SimGRACE 使用 GNN 编码器和扰动 GNN 编码器提取原始图特征, 然后通过对比两个视图来保留关键的语义信息。此外, Wang等^[18]对图对比学习中的结构公平性进行了探究, 其研究表明引入对比学习后学习到的高度节点和低度节点的特征表示均优于 GCN 模型。

本文提出了一种基于属性信息和拓扑信息引导的图多层感知机模型, 通过计算属性相似度和拓扑相似度, 推动特征空间中相似节点的低维表示相互靠近, 不相似节点的低维表示进一步远离, 从而提升模型在节点分类任务上的表现。

2 A&T-MLP

本节主要介绍 A&T-MLP 的算法原理及模型构建。首先提出模型整体架构, 然后介绍用于对比损失的属性相似度和拓扑相似度, 接着讨论用于模型训练的对比损失和交叉熵损失, 最后介绍模型的优化过程。

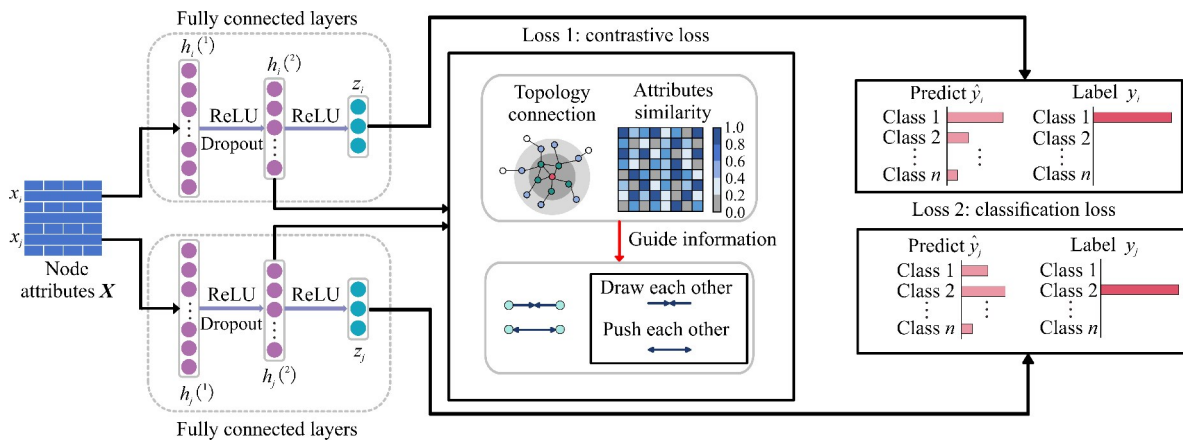


图1 A&T-MLP 的框架

Fig. 1 The framework of A&T-MLP

2.2 属性相似性与拓扑相似性

为了充分保留原始图中蕴含的属性与拓扑关联信息, 分别使用节点属性矩阵 \mathbf{X} 和邻接矩阵 \mathbf{A} 计算用于对比损失函数的节点相似性。

对于拓扑结构, 可以通过图的一阶相似性进行表示, 即直接相连的节点关系更加紧密^[20]。如果节点 i 和 j 之间存在 1 条边, 那么 i 和 j 的一阶相似性为边的权重值 (无权图中, 权重值默认为 1), 如果 i 和 j 之间没有边, 那么两个节点的一阶相似性为 0。由于上

2.1 模型结构

MLP 通过全连接层 (Fully Connected Layer, FC) 提取数据中蕴含的潜在信息, 生成用于下游任务的特征表示。A&T-MLP 在 MLP 的基础上, 通过引入兼顾节点属性相似性与拓扑结构相似性的对比损失以增强模型对原始图信息的表征能力, 其整体结构如图 1 所示。由于过多的神经网络层数易导致模型过拟合^[19], 因此 A&T-MLP 使用 3 层网络结构。首先, 使用 MLP 提取属性矩阵 \mathbf{X} 的特征, 在第 1 层和第 2 层网络之间引入激活函数 ReLU 和 Dropout 机制, 这样不仅能增强模型的非线性变换能力, 还能防止参数过拟合; 然后, 使用属性和拓扑信息引导的对比损失, 对第 2 层网络生成的特征向量进行优化, 使属性和拓扑相似的节点在特征空间中进一步接近, 不相似的节点进一步远离; 最后, 使用第 3 层网络生成的特征向量计算交叉熵损失, 从而实现节点分类任务。综上, A&T-MLP 的网络结构可表示为

$$\mathbf{H}^{(1)} = \text{Dropout}(\text{ReLU}(\text{FC}(\mathbf{X}))), \quad (2)$$

$$\mathbf{H}^{(2)} = \text{ReLU}(\text{FC}(\mathbf{H}^{(1)})), \quad (3)$$

$$\mathbf{Z} = \text{FC}(\mathbf{H}^{(2)}). \quad (4)$$

述边信息通常使用图的邻接矩阵 \mathbf{A} 进行存储, 所以 \mathbf{A} 中元素 A_{ij} 能够直接表示节点 i 和 j 的拓扑相似性。

对于节点属性, 属性矩阵 \mathbf{X} 中元素 X_{ij} 表示节点 i 和属性 j 之间的关联信息, 不能直接表示节点间的相似性。因此, 属性相似性计算的关键在于如何表示节点在属性空间的相似性, 并通过与拓扑相似性相同的形式进行表示。首先, 使用相似性度量算法, 计算两个节点的属性相似性值 S_{ij} , 本文使用余弦相似性

获取相似度矩阵 \mathbf{S} 。

余弦相似度即使用两个向量之间夹角的余弦值度量相似度：

$$\mathbf{S}_{ij} = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{|\mathbf{x}_i| |\mathbf{x}_j|}, \quad (5)$$

式中, \mathbf{x}_i 和 \mathbf{x}_j 表示节点 i 和 j 的属性向量。根据 \mathbf{S} 中数据元素大小为每个节点选取前 k 个最相似的节点并保留相似度值(其余位置为 0), 构建属性相似度矩阵 $\mathbf{S}^{(k)}$, 此时 $\mathbf{S}^{(k)}$ 中元素 $\mathbf{S}_{ij}^{(k)}$ 能够直接表示节点 i 和 j 的属性相似度。

拓扑相似度和属性相似度分别通过邻接矩阵 \mathbf{A} 和属性相似度矩阵 $\mathbf{S}^{(k)}$ 表示, 但是 \mathbf{A} 中元素 \mathbf{A}_{ij} 的取值为 $\{0, 1\}$, $\mathbf{S}^{(k)}$ 中元素 $\mathbf{S}_{ij}^{(k)}$ 的取值为 $[0, 1]$ 。为了消除两类不同信息在量纲和数值上的差异, 对数据进行规范化处理 $\tilde{\mathbf{A}} = \mathbf{D}_1^{-1} \mathbf{A}$, $\tilde{\mathbf{S}}^{(k)} = \mathbf{D}_2^{-1} \mathbf{S}^{(k)}$, 其中 \mathbf{D}_1 表示邻接矩阵 \mathbf{A} 的度矩阵, \mathbf{D}_2 表示属性相似度矩阵 $\mathbf{S}^{(k)}$ 的度矩阵。最终, 用于对比损失计算的拓扑和属性引导信息可表示为

$$\gamma = \tilde{\mathbf{A}} + \tilde{\mathbf{S}}^{(k)}, \quad (6)$$

式中, γ 表示包含原始图中属性与拓扑相似信息的节点关联强度矩阵。

2.3 损失函数

在 MLP 提取数据特征生成节点表示的过程中, 使用属性和拓扑引导信息 γ 计算对比损失可以推动特征空间中相似节点的低维表示彼此接近, 不相似节点的低维表示进一步远离。基于对比损失的一般形式[公式(1)], A&T-MLP 使用的对比损失 loss_{CL} 表达式为

$$\text{loss}_{CL} = -\log \frac{\sum_{j=1}^N \mathbf{1}_{[j \neq i]} \gamma_{ij} \exp(\text{sim}(h_i, h_j))}{\sum_{k=1}^N \mathbf{1}_{[k \neq i]} \exp(\text{sim}(h_i, h_k))}, \quad (7)$$

式中, sim 表示余弦相似度函数, h 表示节点。 loss_{CL} 借助节点的相似度信息以增大特征空间中相似节点的一致性和增大不相似节点的差异性。由于 A&T-MLP 是用于分类任务的模型, 因此在训练过程中引入交叉熵损失 loss_{CE} :

$$\text{loss}_{CE} = -\sum_{i \in \mathbf{Y}} \sum_{m=1}^M Y_{im} \log(Z_{im}), \quad (8)$$

式中, \mathbf{Y} 为节点标签集合, M 为标签的类别数, Y_{im} 为符号函数(样本 i 的真实类别等于 m 则取值 1, 否则取 0), Z_{im} 表示观测样本 i 属于类别 m 的预测概率。最后, 将对比损失 loss_{CL} 和交叉熵损失 loss_{CE} 进行组

合, 构建完整的损失函数 loss :

$$\text{loss} = \text{loss}_{CE} + \alpha \text{loss}_{CL}, \quad (9)$$

式中, α 表示平衡 loss_{CL} 和 loss_{CE} 的加权系数。在确定网络结构和损失函数后, A&T-MLP 优化的具体过程见算法 1。

算法 1 A&T-MLP

输入: 属性矩阵 \mathbf{X} , 邻接矩阵 \mathbf{A} , 节点标签集合 \mathbf{Y} , 隐藏层维度 d 、属性相似节点数 k , 权重系数 α , 可训练参数 θ 。

输出: 模型优化参数 θ 。

1. 计算属性矩阵 \mathbf{X} 的余弦相似度得到相似度矩阵 \mathbf{S}
2. 选取每个节点前 k 个相似度值构建属性相似度矩阵 $\mathbf{S}^{(k)}$
3. 对 \mathbf{A} 和 $\mathbf{S}^{(k)}$ 进行归一化, 计算对比损失引导信息 γ
4. Repeat
5. 使用 \mathbf{X} 、 θ 和公式(3)、(4)生成特征矩阵 $\mathbf{H}^{(2)}$ 和 \mathbf{Z}
6. 使用 $\mathbf{H}^{(2)}$ 、 \mathbf{Z} 、 γ 、 α 和公式(9), 计算 loss
7. 计算损失函数梯度 ∇loss , 利用反向传播更新参数矩阵 θ
8. Until convergence

3 结果与分析

3.1 实验设置

本文使用 3 个基准图数据集 Wikipedia、Cora 和 Citeseer^[21] 评估基线模型和 A&T-MLP 在节点分类任务中的实验性能, 使用的基线模型和构建的基线模型均使用 Python 语言和深度学习框架 Pytorch 实现。数据集相关统计信息见表 1。基线模型分别选择 GNN 模型 GCN^[3]、SGC^[4]、GAT^[5] 和 GATv2^[6], 以及感知机模型 MLP^[10] 和 Graph-MLP^[14]。

节点分类是一种多分类任务, 利用图的拓扑结构和节点属性确定每个节点所属类别, 因此采用常见多分类指标 Micro-F1 和 Macro-F1 进行评估。对于每个数据集, 各模型采用相同数据集划分, 随机抽取 10% 的节点标签作为训练集, 剩余节点标签中随机抽取 50% 作为测试集。

所有基线模型都先按照原始论文中建议的参数进行初始化, 然后对部分模型进一步调整, 以获取最佳性能。对于 A&T-MLP, 参数主要包括学习率 lr 、训练次数 epoch、隐藏层维度 d 、属性相似节点数 k

以及平衡损失函数的权重系数 α 。A&T-MLP 的最终参数设置如表 2 所示。

表 1 数据集信息

Table 1 Dataset information

数据集 Dataset	#节点 # Nodes	#边 # Edges	#属性 # Attributes	#标签 # Labels
Wikipedia	2 405	17 981	4 973	19
Cora	2 708	5 429	1 433	7
Citeseer	3 327	4 732	3 703	6

表 2 A&T-MLP 参数设置

Table 2 Parameters setting of A&T-MLP

数据集 Dataset	学习率 Learning rate (lr)	训练次数 Epoch	隐藏层维度 Dimension of hidden layers (d)	属性相似节点数 Number of nodes with similar attributes (k)	权重系数 Weighting coefficient (α)
Wikipedia	10^{-2}	150	512	30	6
Cora	10^{-3}	500	512	20	10
Citeseer	10^{-3}	400	512	30	10

3.2 实验结果

本节通过节点分类任务评估不同模型的实验性能,实验结果如表 3 所示。从表 3 可以看出,在 3 个数据集上,A&T-MLP 模型的 Micro-F1 和 Macro-F1 分数始终高于基线模型,表明 A&T-MLP 能够有效保留原始图的属性和拓扑特征信息,并将其编码到节点特征向量中,从而提升了节点分类能力。在属性信息丰富的 Wikipedia 数据集上,所有 GNN 模型表现均弱于 MLP 模型,同时引入拓扑对比损失的 Graph-MLP 弱于仅使用属性信息的 MLP。这表明在以属性为主要信息的数据集上,依赖拓扑结构进行特征提取和保留的方法限制了模型的代表能力,而同时使用属性和拓扑信息引导对比损失的 A&T-MLP 表现最好,能够有效平衡特征向量中蕴含的属性和拓扑信息。在拓扑信息丰富的 Cora 和 Citeseer 数据集上,所有 GNN 模型优于仅使用属性信息的 MLP。GNN 模型的高性能得益于信息传递和聚合过程中拓扑信息的重复使用,Graph-MLP 则是通过对比损失保留了拓扑关联信息,而 A&T-MLP 的对比损失在关注拓扑相似性的同时,也对属性相似性这一重要的原始图信息进行关注,能够更为充分地保留分类任务所需的特征信息。

综上,不同基线模型在不同数据集上的表现差异较大,并且复杂的信息传递和聚合方式不一定能够获

得最佳的性能。相较于基线模型,A&T-MLP 在不同数据集上均能获得最佳性能,证明了同时使用属性和拓扑相似度信息引导的对比损失,不仅能够提升模型的预测性能,而且也增强了模型对不同数据的泛化能力。

表 3 节点分类实验结果

Table 3 Results of the node classification experiment

模型 Model	Wikipedia		Cora		Citeseer	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
GCN	56.83	41.11	76.37	<u>75.18</u>	71.42	64.44
SGC	53.13	35.19	74.74	73.15	67.64	60.45
GAT	55.33	39.49	76.07	74.75	70.47	64.34
GATv2	54.71	37.38	73.41	71.45	72.45	65.56
MLP	<u>69.07</u>	<u>49.73</u>	61.60	54.93	64.72	54.79
Graph-MLP	67.84	49.15	<u>78.73</u>	75.03	<u>74.43</u>	<u>67.69</u>
A&T-MLP	72.69	54.75	82.79	81.31	76.70	70.09

Note; bold is the best result, underlined is the second best result.

为了进一步探究 A&T-MLP 在 3 个数据集上的性能,本文通过减少训练集中边缘的数量评估基线模型和 A&T-MLP 在信息缺失条件下的实验表现。在边缘缺失实验中,每次减少训练集中 20% 的边信息,实验结果见图 2。

由图 2 可知,随着训练集中包含的边信息不断减少,各模型的实验性能出现较为明显的波动(由于 MLP 仅使用属性信息,实验结果不受边信息减少影响),但 A&T-MLP 模型的预测结果始终优于其他基线模型。在属性信息丰富的 Wikipedia 数据集上,边信息的减少不一定导致模型性能下降,相反 GNN 模型在边缘缺失 80% 时取得了最佳实验结果。在拓扑信息丰富的 Cora 和 Citeseer 数据集上,除 MLP 外所有模型的实验性能随边信息的减少呈现出下降趋势,特别是 GATv2 在边缘缺失 80% 时性能下降 9.53%。综上,在边信息大量缺失时 A&T-MLP 相较于基线模型能够更好地学习数据特征分布,改善节点分类任务的实验表现。

3.3 消融实验

为了验证拓扑信息和属性信息对 A&T-MLP 模型性能的影响,在 3 个数据集上进行消融实验,记录 Micro-F1 和 Macro-F1。其中,不使用对比损失的变体为 MLP,仅使用拓扑信息引导对比损失的变体为 T-MLP,仅使用属性信息引导对比损失的变体为 A-MLP。为保证实验结果的公平性,保持各模型参数一致。消融实验结果如图 3 所示。相较于不使用原

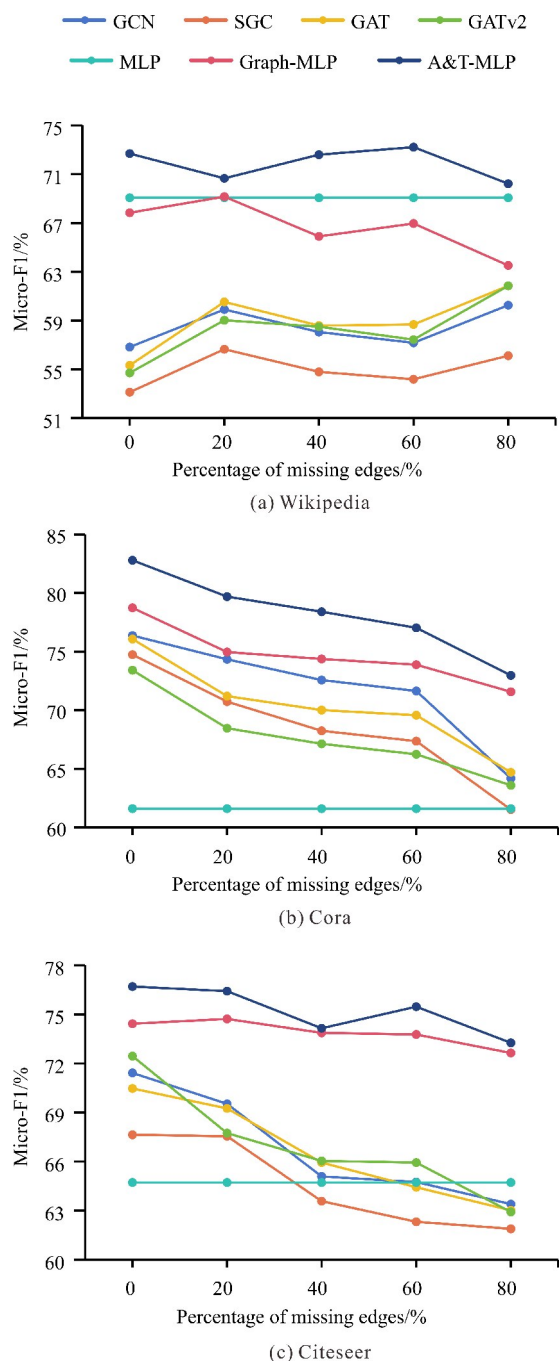


图2 边缘缺失实验结果

Fig. 2 Results of the missing edge experiment

始图信息、单独使用拓扑信息或属性信息的变体,同时使用两类信息的 A&T-MLP 模型获得了最佳性能,证明同时保留属性和拓扑信息能够进一步增强模型的代表能力,提升分类任务中的实验表现。

3.4 参数实验

为了分析 A&T-MLP 性能受参数的影响,使用 Citeseer 数据集进行参数实验,并记录 Micro-F1 和 Macro-F1 分数。此外,为了保证参数实验的公平性,除验证参数外,其余参数按照表 3 进行设置。参数实

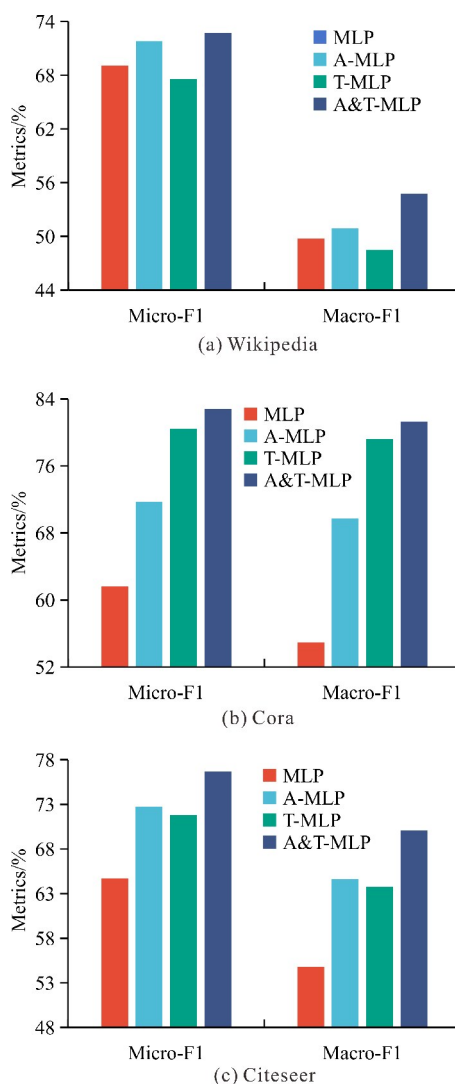


图3 消融实验结果

Fig. 3 Results of ablation experiment

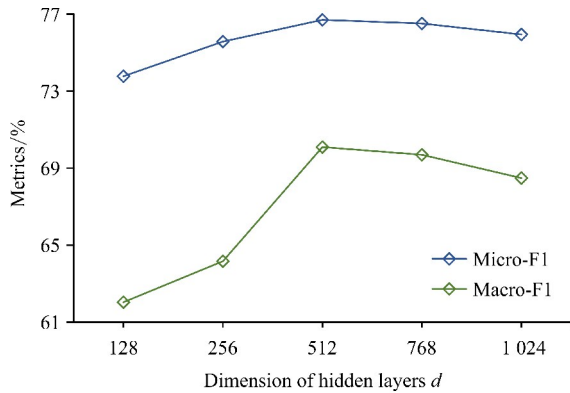
验结果如图 4 所示。

为了验证隐藏层维度 d 对 A&T-MLP 性能的影响,使用不同 d 值进行实验,结果如图 4(a) 所示。最初, Micro-F1 和 Macro-F1 随网络维度的增大而提高,这是因为较高的维度代表了更多的神经元,能够编码更多的有益信息,提升实验表现。但是,随着维度进一步增大, Micro-F1 和 Macro-F1 呈现出下降的趋势,这是因为过大的维度使节点表示中编码了噪声信息,从而影响了 A&T-MLP 的表现。

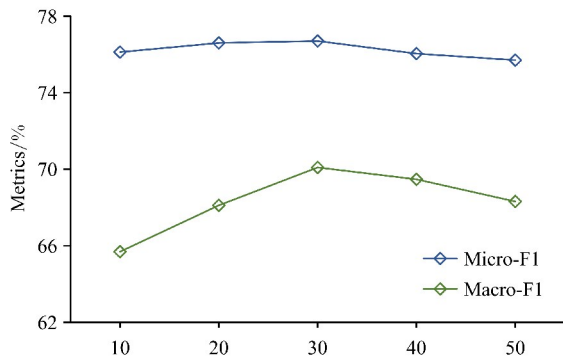
为了验证属性相似节点数 k 对 A&T-MLP 性能的影响,使用不同 k 值进行实验,结果如图 4(b) 所示。从结果看, Micro-F1 和 Macro-F1 呈现先上升再下降的趋势,过大和过小的 k 值下 A&T-MLP 的表现均较为一般,这是因为 k 值过小无法保留充足的属性相似度信息,而 k 值过大会在对比损失计算过

程中引入大量噪声节点,影响 A&T-MLP 的性能。

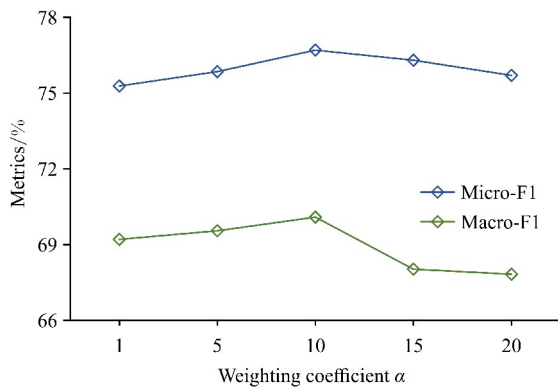
为了验证损失函数中的权重系数 α 对模型性能的影响,使用不同 α 值进行实验,结果如图 4(c) 所示。权重系数 α 调控了对比损失在整个损失函数中所占的比重, α 值越大模型训练过程中越关注对比损失的优化。从实验结果来看,A&T-MLP 的预测性能随着 α 值的增大先上升后下降,这是因为过于关注对比损失的优化会导致对交叉熵损失的忽略,从而影响了 A&T-MLP 的性能。



(a) Experimental results of parameter d



(b) Experimental results of parameter k



(c) Experimental results of parameter α

图 4 参数实验结果

Fig. 4 Results of parametric experiment

3.5 可视化

节点表示蕴含了原始图的相关信息,对其进行可视化能够直观地反映原始图的某些特征。对于可视化任务,首先使用 t -SNE^[22] 将各模型生成的节点特征向量降至 2 维,然后根据 Citeseer 数据集的标签信息将节点分别标记为 6 种颜色,最后在二维平面上进行可视化,实验结果如图 5 所示。节点分类准确的可视化结果通常为相同颜色节点接近,不同颜色节点远离。由图 5 可知,GCN、MLP、Graph-MLP 和 A&T-MLP 均能提取原始图中相关信息形成社区结构,但是使用属性和拓扑信息引导对比损失的 A&T-MLP 具有更高的类内相似性和类间界限。可视化实验反映了模型保留同一社群节点相似特征的能力,证明了使用对比损失能够增强模型表征能力,也更为直观地反映了模型的性能。

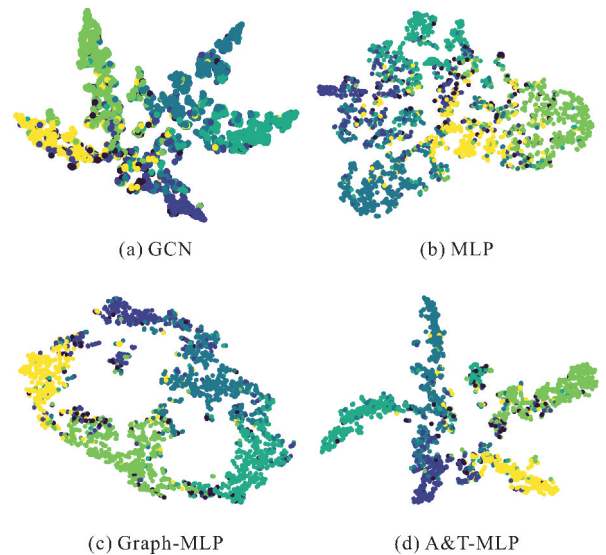


图 5 Citeseer 可视化

Fig. 5 Visualization of Citeseer

3.6 训练时间

为了比较不同模型的训练复杂度,记录 Wikipedia 数据集上迭代 100 次后单次迭代的平均训练时间(包括前向传播、损失函数计算、反向传播过程),实验结果如图 6 所示。对比不同模型的训练时间,直接对原始图信息编码的 GCN 和 MLP 计算速度较快,在此基础引入了更加复杂的特征提取方式的改进模型训练时间进一步增加。虽然基于 MLP 的改进模型网络结构更为简单,但是为了保留原始图相关信息,需要借助对比损失进行优化,导致模型的训练时间增加。此外,对比使用对比损失的 Graph-MLP 和 A&T-MLP,尽管 A&T-MLP 引入了属性信息,但是

只保留了每个节点前 k 个相似度值最大的值, 因此相较于 GAT 和 GATv2 训练时间差异较大, 相较于 Graph-MLP 训练时间差异较小, 并且实验性能提升更为明显。

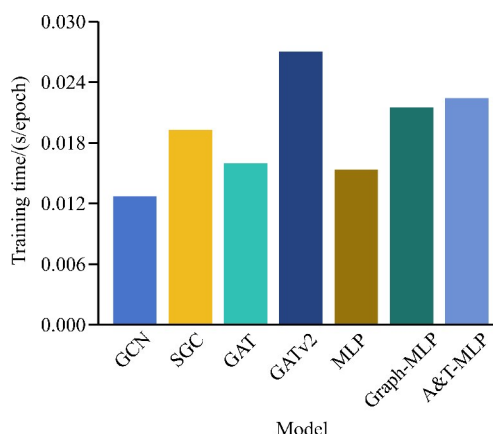


图6 训练时间对比

Fig. 6 Comparison of training time

4 结论

本文设计了一种无需信息传递和聚合的图多层感知机模型 A&T-MLP, 通过属性和拓扑信息引导的对比损失有效保留原始图特征。实验结果表明, 基于属性相似度与拓扑相似度的对比损失能够显著提升 MLP 模型在节点分类任务中的实验性能, 并且优于较为先进的 GNN 基线模型。此外, A&T-MLP 在处理拓扑信息不准确的图数据时具有较明显的优势, 即使在拓扑信息缺失 80% 的极端情况下, 依然具有良好的性能。在当前工作中, 通过简单的 MLP 构建神经网络基本框架, 使得模型训练过程中缺乏对特征中重要信息的关注。因此, 在后续工作中, 将在多层神经网络中引入注意力机制, 增强关键特征信息的保留, 并将模型应用于社交网络恶意账户检测和犯罪组织成员身份发现等任务。

参考文献

- [1] ZOU M H, GAN Z X, CAO R Z, et al. Similarity-navigated graph neural networks for node classification [J]. *Information Sciences*, 2023, 633: 41-69.
- [2] VELICKOVIC P. Everything is connected: graph neural networks [J]. *Current Opinion in Structural Biology*, 2023, 79: 102538.
- [3] BHATTI U A, TANG H, WU G L, et al. Deep learning with graph convolutional networks: an overview and latest applications in computational intelligence [J]. *International Journal of Intelligent Systems*, 2023, 2023: 8342104.
- [4] YANG F, ZHANG H Y, TAO S M. Simplified multilayer graph convolutional networks with dropout [J]. *Applied Intelligence*, 2022, 52: 4776-4791.
- [5] HE L C, BAI L, YANG X, et al. High-order graph attention network [J]. *Information Sciences*, 2023, 630: 222-234.
- [6] BRODY S, ALON U, YAHAV E. How attentive are graph attention networks? [C]// *Proceedings of the 2022 International Conference on Learning Representations (ICLR 2022)*. Virtual; OpenReview, 2022: 1-26.
- [7] SUN Y F, DENG H R, YANG Y, et al. Beyond homophily: structure-aware path aggregation graph neural network [C]// *Proceedings of the 2022 International Joint Conference on Artificial Intelligence (IJCAI-22)*. San Francisco; Morgan Kaufmann, 2022: 2233-2240.
- [8] WU G C, LIN S K, SHAO X X, et al. QPGCN: graph convolutional network with a quadratic polynomial filter for overcoming over-smoothing [J]. *Applied Intelligence*, 2023, 53: 7216-7231.
- [9] ZHANG A C, HUANG J C, LI P, et al. Building shortcuts between distant nodes with biaffine mapping for graph convolutional networks [EB/OL]. (2023-02-17) [2023-09-01]. <https://arxiv.org/pdf/2302.08727.pdf>.
- [10] KRUSE R, MOSTAGHIM S, BORGELT C, et al. Multi-layer perceptrons [M]// *Computational Intelligence: a methodological introduction*. Cham: Springer, 2022: 53-124.
- [11] HOU Q B, JIANG Z H, YUAN L, et al. Vision permutator: a permutable MLP-like architecture for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(1): 1328-1334.
- [12] GONG N, ZHANG C L, ZHOU H, et al. Classification of hyperspectral images via improved cycle-MLP [J]. *IET Computer Vision*, 2022, 16: 468-478.
- [13] SUN H, WANG H Y, LIU J Q, et al. CubeMLP: an MLP-based model for multimodal sentiment analysis and depression estimation [C]// *Proceedings of the 30th ACM International Conference on Multimedia*. Lisbon, Portugal; ACM, 2022: 3722-3729.
- [14] HU Y, YOU H X, WANG Z C, et al. Graph-MLP: node classification without message passing in graph [EB/OL]. (2021-06-08) [2023-09-01]. <https://arxiv.org/pdf/2106.04051.pdf>.
- [15] RANI V, NABI S T, KUMAR M, et al. Self-supervised learning: a succinct review [J]. *Archives of Computational Methods in Engineering*, 2023, 30: 2761-2775.
- [16] WANG Y L, ZHANG J, LI H Y, et al. ClusterSCL: cluster-aware supervised contrastive learning on graphs [C]// *Proceedings of the 2022 ACM Web Conference*.

- Lyon, France; ACM, 2022; 1611-1621.
- [17] XIA J, WU L R, CHEN J T, et al. SimGRACE: a simple framework for graph contrastive learning without data augmentation [C]//Proceedings of the 2022 ACM Web Conference. Lyon, France; ACM, 2022; 1070-1079.
- [18] WANG R J, WANG X, SHI C, et al. Uncovering the structural fairness in graph contrastive learning [C]//36th Conference on Neural Information Processing Systems (NeurIPS 2022). New Orleans, USA: MIT Press, 2022; 32465-32473.
- [19] WANG R Q, ZHU F, ZHANG X Y, et al. Training with scaled logits to alleviate class-level over-fitting in few-shot learning [J]. Neurocomputing, 2023, 522: 142-151.
- [20] 袁立宁, 蒋萍, 莫嘉颖, 等. 基于二阶图卷积自编码器的图表示学习[J/OL]. 计算机工程与应用, 2023; 1-9. [2023-10-10]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20230626.1839.016.html>.
- [21] WAIKHOM L, PATGIRI R. A survey of graph neural networks in various learning paradigms: methods, applications, and challenges [J]. Artificial Intelligence Review, 2023, 56: 6295-6364.
- [22] WANG Y F, HUANG H Y, RUDIN C, et al. Understanding how dimension reduction tools work: an empirical approach to deciphering t -SNE, UMAP, TriMAP, and PaCMAP for data visualization [J]. Journal of Machine Learning Research, 2021, 22: 1-73.

Node Classification Based on Graph Multi-Layer Perceptron

WEN Zhu¹, YUAN Lining^{1,2}, HUANG Wei^{3* * *}, HUANG Wanyan¹, MO Jiaying¹, FENG Wengang²

(1. School of Information Technology, Guangxi Police College, Nanning, Guangxi, 530028, China; 2. School of National Security, People's Public Security University of China, Beijing, 100038, China; 3. School of Artificial Intelligence, Nanning College for Vocational Technology, Nanning, Guangxi, 530008, China)

Abstract: Many Graph Neural Networks (GNN) improve the performance of graph analysis tasks, such as node classification, by designing efficient information propagation and aggregation methods. However, this paper proposes a graph Multi-Layer Perceptron (MLP) model A&T-MLP does not rely on information propagation and aggregation. It enhances its representation ability using through the utilization of contrastive loss guided by attribute and topology information. Firstly, attribute matrix and adjacency matrix were used to calculate the attribute and topology similarity between nodes. Secondly, the contrastive loss guided by similarity information is applied to increase the consistency of similar nodes and the difference of dissimilar nodes in feature space. Thirdly, the multi-layer perceptron is constructed, and the end-to-end training process incorporates the cross-entropy loss function. A&T-MLP outperforms the baselines in the node classification. Micro-F1 and Macro-F1 increased by 15.86% and 13.64% respectively, compared to the GNN model Graph Convolutional Network (GCN) on Wikipedia dataset. The results show that A&T-MLP effectively preserves richer graph information, leading to improved performance. Besides, the proposed method exhibits notable advantages in inaccurate topology. Even in the extreme case where 80% of edges are missing, A&T-MLP can still superior against baselines.

Key words: graph neural network; multi-layer perceptron; node attributes; contrastive learning; node classification

责任编辑: 陆雁, 陈少凡
