基于改进图神经网络的小样本分类研究

张中天[1] 王 涛[1] 谭莹莹[1] 熊 伟[1]
ZHANG Zhongtian WANG Tao TAN Yingying XIONG Wei

摘 要 大多数现有的元学习方法都用来解决欧几里得领域中的图像和文本等小样本学习问题。然而，将元学习应用于非欧几里得领域的工作较少，且图神经网络模型在小样本节点分类问题上表现欠佳。针对此问题，提出一种新型的分类算法 MGAT，首先采用动态注意力机制灵活提取和聚合邻居节点的特征信息，然后通过对类似小样本学习任务进行训练以获取分类器的先验知识，再给出少量标记样本对新类中的节点进行分类。在两个基准数据集上进行的实验表明，所提出的方法在小样本节点分类问题上获得了优于当前各个分类模型的性能。

关键词 图神经网络；元学习；小样本学习；注意力机制；分类算法

doi：10.3969/j.issn.1672-9528.2024.05.012

0 引言

近期，使用各种深度学习方法分析图结构数据引起了研究者的广泛研究兴趣[1]。已经有大量的模型被开发出来，以解决不同的图结构问题，包括图表示、图节点分类[2]等。早期的方法主要侧重于以无监督的方式嵌入节点捕获全局或局部网络结构，而最近的工作将复杂的深度学习模型（已成功用于欧几里得领域）应用于非欧几里得图结构数据，从而产生了许多著名的基于图神经网络（GNN）的方法，如 GCN[3]、GraphSAGE[4]、SGC[5]等。

尽管图神经网络取得了突破，但它在小样本学习问题上依然面临着持续的挑战。现有的 GNN 模型在遇到新类时总是需要重新学习其参数以融入新信息，如果每个新类中的节点数量很少，模型的性能将遭受灾难性的下降。故当前 GNN 的主要挑战之一就是无法处理只存在一个或很少样本的情况。为了解决这个问题，最近的一些研究集中于解决图数据上的小样本学习，其中有的是通过协同训练和自训练进行训练[6]，这些方法通过利用未标记的数据来增强模型的泛化能力。有的是通过堆叠转置图卷积层来扩展重建输入特征[7]，这种方法通过堆叠转置图卷积层来扩展和重建输入特征，从而使得模型在只有少量样本的情况下也能够捕捉到足够的信息来进行有效的学习。有的则通过图数据增强，修改图结构如添加、删除节点或边，可以生成新的图样本，从而增加训练数据的多样性，帮助模型更好地泛化到新类别上。但是当标签极其稀缺时，直观上需要更多轮的聚合，因此需要更深的网络；然而，更深版本的网络通常会导致性能更差。这主要是因为一些关键特征信息可能会在迭代平均的过程中丢失[8]。

元学习（即学会学习）近年来在人工智能界引起了广泛关注，因为它能够快速适应新任务，并通过少量样本学习任务之间的可迁移知识。与许多标准机器学习技术不同，后者涉及对单个任务进行训练并对该任务中保留的样本进行测试。学习大量任务和测试新任务的过程分别被称为元训练和元测试。采用元学习进行图像和文本学习已经取得了重大进展，并且最近提出了几种模型和算法，例如匹配网络[9]、原型网络[10]、关系网络[11]、MAML[12]等。尽管在分析具有类欧几里得属性的数据（例如图像和文本）方面存在大量的研究且有很不错的结果，但将元学习应用于图的工作却很少。其中主要原因是图数据更加不规则、有噪声且节点之间的关系更加复杂。

为了弥补这一差距，提出了一种称为 MGAT 的图元学习模型，重点关注图数据上的小样本节点分类问题。其不是仅仅依赖现有基于图神经网络的模型中节点的辅助信息和来自邻居的聚合信息，而是在大量类似的小样本学习任务上进行训练和优化，以更好地泛化学习新任务。总的来说，本文的主要贡献如下。

（1）为节点分类制定了一种新的小样本学习模型。它与以前的工作不同，目标是对来自新类的节点进行分类，每个类只有很少的样本。

（2）在网络结构中添加动态注意力，提升模型的灵活性以及抵抗噪声的能力。

（3）在两个基准数据集上证明了本文的方法相对于几种先进的 GNN 模型的优越性。

1 图神经网络

本文将无向图表示为 $G=(V, E, A, X)$ 的四元组形式，其中 $V=\{v_1,v_2,\cdots,v_i,\cdots,v_n\}$ 是节点集，$E=\{e_{i,j}=(v_i,v_j)\} \subseteq (V\times V)$ 是边

1. 华北电力大学计算机系 河北保定 071000

集，$\boldsymbol{A}\in R^{n\times n}$ 是一个对称邻接矩阵，其中 a_{ij} 表示节点 v_i 和 v_j 之间的边的权重，$\boldsymbol{X}\in R^{n\times d}$ 是特征矩阵，其中 $x_i\in R^d$ 表示给定节点 v_i 的特征。

对于元学习范式中的小样本节点分类问题，其目标是获得一个能够适应训练过程中未见过的新类的分类器，每个新类中只给出少量样本。形式上，训练集 D_{train} 中的每个节点 v_i 属于 C_1 中的类之一。给定的不相交测试集 D_{test} 中的节点与完全不同的新类 C_2 相关联。在 D_{test} 中，少量节点具有可用的标签。目标是找到一个函数 f_θ，它能够将其余未标记的节点分类到 C_2 中的类之一，并且误分类率较低。如果每个类中标记节点的数量为 K，则该任务称为 $|C_2|$-way K-shot 学习问题，其中 K 是一个很小的数字。

现代图神经网络联合利用图的结构信息和节点特征 X 来学习节点的新表示向量 $\boldsymbol{h}_i$，通常遵循邻域聚合方案。经过 n 次聚合迭代后，节点从其 n 跳邻居中捕获结构信息并更新表示。图 1 展示了一个说明性示例，其中当图经过 GNN 的第一层时，深色节点聚合来自节点 1 到 4 的信息，在第二层之后，它收集来自节点 5 和节点 7 的信息。

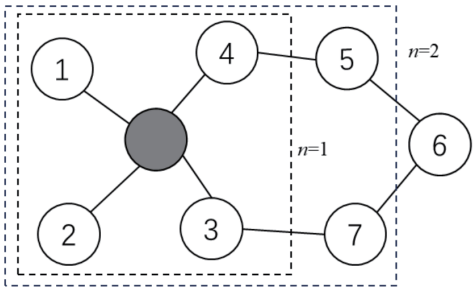

图 1 邻域聚合示例

正式地，GNN 的第 n 层定义为：

$$\boldsymbol{a}_i^n=\boldsymbol{h}_i^{n-1}\cdot aggregate^n(\boldsymbol{h}_j^{n-1}) \tag{1}$$

$$\boldsymbol{h}_i^n=\boldsymbol{h}_i^{n-1}\cdot combine^n(\boldsymbol{a}_i^n) \tag{2}$$

式中：$\boldsymbol{h}_i^n$ 是节点 i 在第 n 层的特征向量，$\boldsymbol{h}_j$ 代表 $\boldsymbol{h}_i$ 的邻居节点特征。最终表示 $\boldsymbol{h}_i^N$ 将用于下游任务，例如链路预测和节点分类。

图卷积网络（GCN）和图注意力网络（GAT）在本质上都致力于通过整合邻近节点的特征来更新目标节点的特征表示。它们的工作方式都是聚合周边顶点的信息以生成目标顶点的新特征。只是 GCN 依赖于拉普拉斯矩阵来进行信息传播，而 GAT 则引入了注意力系数。在模型中融合节点间特征的相关性方面，GAT 表现得更为出色 [13]。

对于图中的某个节点，不同的邻居对其的影响也有不同，对应的边的权重大小由其和各邻居的关系决定。根据与邻居的关系设置边的权重会让网络更加灵活，图注意力网络计算了每个节点和邻居的注意力系数，通过这个系数加权计算各边的权重。具体计算过程为：

$$\alpha_{ij}=\frac{\exp(LeakyReLU(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}\boldsymbol{h}_i\,\|\,\boldsymbol{W}\boldsymbol{h}_j]))}{\sum_{k\in N_i}\exp(LeakyReLU(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}\boldsymbol{h}_i\,\|\,\boldsymbol{W}\boldsymbol{h}_k]))} \tag{3}$$

式中：α_{ij} 即是节点 i 和节点 j 的注意力系数。W 是一个共享的线性映射，它对顶点的特征进行了增维；将顶点 i, j 变换后的特征进行拼接后，再使用 $\boldsymbol{a}^{\mathrm{T}}$ 把这个高维特征映射到一个实数上；最后再进行一个归一化处理得到注意力系数。N_i 代表 i 的邻居节点集合。计算出注意力系数后，新的节点特征即是将邻居节点乘以相对应的权值，然后做相加：

$$\boldsymbol{h}_i'=\sigma\left(\sum_{j\in N_i}\alpha_{ij}\boldsymbol{W}\boldsymbol{h}_j\right) \tag{4}$$

2 元图注意力算法

本文的方法基于元学习方法，通过使用采样的类似任务的训练来解决新的小样本学习问题（即元测试任务 T_{mt}）。子任务中相应的训练集和测试集分别被称为支持集和查询集。子任务结构示意图如图 2。

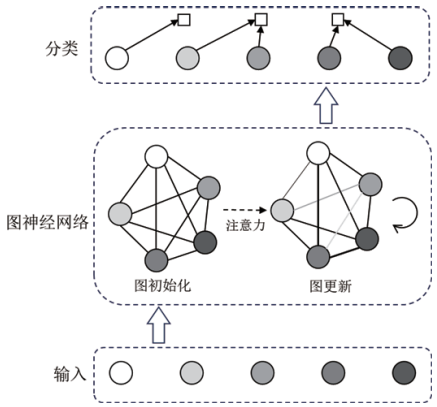

图 2 子任务网络结构示意图

本文将 MGAT 模型表示为 f_θ，参数为 θ，训练集表示为 $D_{\text{train}}=\{(x_1, y_1),\cdots, (x_i, y_i), \cdots, (x_N, y_N)\}$，其中 $y_i\in C_1$ 且 N 是训练集中的节点数。从训练集中提取出 M 个元训练任务：$T=\{T_1, T_2, \cdots, T_M\}$。元训练任务 T_i 的支持集 S_i 和查询集 Q_i 都从 D_{train} 中采样。在支持集中，有 $S_i=\{v_{i1}, v_{i2}, \cdots, v_{is}\}=\{(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \cdots, (x_{is}, y_{is})\}$，其中 $s=|S_i|$，x_{is} 是带有标签 y_{is} 的节点 v_{is} 的输入向量。

2.1 元训练

元训练任务的生成需要从 C_1 中采样 $|C_2|$ 个类别，然后为每个类别随机采样 K 个节点，以此来模拟小样本节点分类。具体来说，首先随机采样 $|C_2|$ 个来自 C_1 的类，表示为 C。然后可以得到 D_C，它是训练集中包含元素 (x_i, y_i) 的子集，其中 y_i 是 C 中的类之一。接下来，随机采样 $K\times|C_2|$ 个来自 D_C 的节点形成支持集 S_i，其中 K 是 S_i 的每个类别中的节点数。最后从 D_C 的剩余节点中随机采样 P 个节点来组成查询集 Q_i，该查询集又用于构造元训练任务 $T_i=S_i+Q_i$。重复上述步骤 M 次，产生 M 个元训练任务。

当学习任务 T_i 时，首先将支持集 S_i 馈送到 MGAT，并计算交叉熵损失：

$$\ell_{T_i}(f_\theta)=-\left(\sum_{x_{is},y_{is}} y_{is}\log f_\theta(x_{is})+(1-y_{is})\log(1-f_\theta(x_{is}))\right) \quad (5)$$

然后使用任务 T_i 中的一个或多个步骤的简单梯度下降来执行参数更新。简洁起见，将在本节的其余部分中仅描述一个梯度更新。

$$\theta_i'=\theta-\alpha_1\frac{\partial\ell_{T_i}(f_\theta)}{\partial\theta} \quad (6)$$

式中：α_1 是任务学习率，模型参数经过训练以优化元训练任务中 $f_{\theta_i'}$ 的性能。更具体地说，元目标为：

$$\theta=\arg\min_\theta\sum_{T_i\sim p(T)}\ell_{T_i}(f_{\theta_i'}) \quad (7)$$

式中：$p(T)$ 是元训练任务的分布。元优化是在模型参数 θ 上执行的，而目标是使用更新的模型参数 θ' 计算的。这是因为对于所有类似的小样本节点分类任务都需要良好的初始化参数 θ，而不是某些仅在特定任务 T_i 上表现良好的更新参数 θ_i'。本质上，MGAT 的目标是优化模型参数，以便在一次或少量梯度下降更新后最大化新任务的节点分类性能。跨任务的元优化通过随机梯度下降法（SGD）进行，模型参数 θ 更新为：

$$\theta\leftarrow\theta-\alpha_2\frac{\partial\sum_{T_i\sim p(T)}\ell_{T_i}(f_{\theta_i'})}{\partial\theta} \quad (8)$$

式中：α_2 是元学习率，即模型引入的额外超参数。对于元测试，只需要将新的小样本学习任务的支持集的节点提供给 MGAT，并使用式（6）通过一个或少量梯度下降步骤更新参数。因此，MGAT 的性能可以在 T_{mt} 的查询集上轻松评估。

MGAT 的整体运行流程如图 3 所示，θ 表示随机初始化的参数，θ_i' 表示经过 i 次元更新后的参数，θ' 则表示经过所有元更新后的参数。其中 M 是元训练任务的数量。

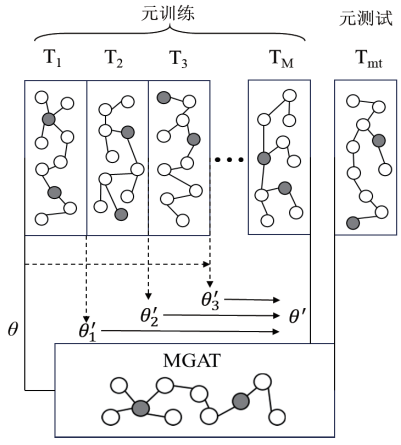

图 3 MGAT 流程

2.2 动态注意力机制

传统 GAT 使用的注意力计算过程存在注意力分数不以查询节点为条件的限制，这会影响 GAT 的表达能力 [14]。因此，本文调整注意力系数的内部计算顺序，以达到动态注意力机制的效果。

从式（3）可以看出，传统 GAT 的问题在于计算系数的函数中 W 和 $\boldsymbol{a}^{\mathrm{T}}$ 是依次运算的，它们可能会退化到单个线性层的效果。因此，将 $\boldsymbol{a}^{\mathrm{T}}$ 移到非线性结果之外，并且先将节点特征与邻居特征进行拼接，再使用 W 进行线性变换。动态注意力系数具体计算过程为：

$$\alpha_{ij}'=\frac{\exp(\boldsymbol{a}^{\mathrm{T}}\cdot LeakyReLU(\boldsymbol{W}\cdot[\boldsymbol{h}_i\,\|\,\boldsymbol{h}_j]))}{\sum_{k\in N_i}\exp(\boldsymbol{a}^{\mathrm{T}}\cdot LeakyReLU(\boldsymbol{W}\cdot[\boldsymbol{h}_i\,\|\,\boldsymbol{h}_k]))} \quad (9)$$

之后再按照式（4）进行特征聚合，以完成整个注意力机制的计算。

3 实验

本文展示了在 Cora 和 Citeseer 两个常用数据集上节点分类性能的实证结果。由于本文工作重点是小样本学习问题，所以对于数据集划分进行了一些修改，以此确保小样本学习的可行性。预处理后的数据集统计结果如表 1 所示。

表 1 数据集描述及划分

	Cora	Citeseer
节点数	2708	3327
特征数	1433	3703
标签数	7	6
$\lvert C_1\rvert$	5	4
$\lvert C_2\rvert$	2	2

对于 Cora 和 Citeseer 数据集，随机划分两个类作为元测试类（即形成 D_{test} 且 $|C_2|$=2 的相应节点），其余节点（即 D_{train}）用于生成元训练。元训练和元测试任务支持集中的每个类对于所有数据集只有一个或三个样本（即 K=1 或 K=3）。因为观察到当支持集极小时，每种方法的性能对节点选择很敏感，所以在支持集中随机选择的 50 个节点上评估所有模型，并报告 Cora 和 Citeseer 上的平均准确度。为了公平、准确地评估 MGAT 和其他基线模型的性能，在 Cora 和 Citeseer 上分别进行了 10 次和 5 次交叉验证。

本文将 MGAT 与 GCN、GraphSAGE 和 SGC 等基线模型在各数据集上进行比较。值得注意的是，仅修改数据集分区以满足元学习范式中的小样本学习设置，每个模型的其他设置与其原始实现相同。

为了加速模型收敛，将 Cora 和 Citeseer 的 batch_size 设置为 5，α_1 和 α_2 在 MGAT 中分别设置为 0.1 和 0.001。

表 2 显示了 MGAT 和基线之间的性能比较结果，从结果中可以明确看出，模型在各个数据集上均取得了最优的表现。一般来说，GNN 模型（包括本文的模型）在少量样

本学习场景中显著优于基于嵌入的模型，例如 DeepWalk 和 Node2Vec。在 GNN 模型中，即使与 GCN 和 SGC 相比，GraphSAGE 也没有表现出有竞争力的结果；即使与 GAT 相比，也存在一定提升。结果表明，以前的图学习模型在应对新增节点时能够展现出良好的性能，但在推广到全新类别的场景中却遇到了困难。

表 2 MGAT 与基线模型准确率

	Cora		Citeseer	
	1-shot/%	3-shot/%	1-shot/%	3-shot/%
DeepWalk	16.06	25.67	14.52	21.18
Node2Vec	15.15	25.66	12.98	20.02
GraphSAGE	50.89	53.12	53.49	55.01
SGC	61.64	75.67	56.91	65.67
GCN	60.33	75.15	58.44	67.99
GAT	63.45	76.03	59.56	69.21
MGAT	**65.27**	**78.41**	**60.35**	**71.78**

MGAT 在 Cora 和 Citeseer 上的优越性随支持集中样本数量的变化而变化，即样本越少，MGAT 相对于基线模型的改进就越大。这也对应了本文的工作重点，即将元学习融入 GNN 模型中以进行小样本图学习。

4 总结

本文提出了一种用于小样本节点分类的新模型，该模型利用元学习和动态注意力机制来优化图神经网络的初始化参数。本文提出的 MGAT 模型可以很好地适应标记样本较少的新学习任务，并显著提高元学习范式下小样本节点分类的性能。其在两个广泛使用的数据集上获得了很好的结果。在未来的工作中，希望扩展本文的模型来解决其他问题，例如小样本文本分类或零样本节点分类。

参考文献：

[1]WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J].IEEE transactions on neural networks and learning systems, 2021,31:4-24.

[2] 张丽英 , 孙海航 , 石兵波 . 基于图卷积神经网络的节点分类方法研究综述 [J/OL]. 计算机科学 :1-19[2024-01-07].http://kns.cnki.net/kcms/detail/50.1075.TP.20230925.1655.162.html.

[3]KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of the 5th International Conference on Learning Representations, Conference Track Proceedings. Toulon:OpenReview.net,2017:1-14.

[4]WILLIAM L H, REX Y, JURE L.Inductive representation learning on large graphs[EB/OL].(2017-06-07)[2023-12-23]. https://arxiv.org/abs/1706.02216.

[5]WU F, ZHANG T, SOUZA A H D, et al.Simplifying graph convolutional networks[EB/OL].(2019-02-19)[2023-12-18]. https://arxiv.org/abs/1902.07153.

[6]LI Q, HAN Z, WU X M.Deeper insights into graph convolutional networks for semi-supervised learning[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence,Volume Five.Palo Alto:AAAI Press,2018:3538-3545.

[7]ZHANG S, ZHOU Z, HUANG Z, et al. Few-shot classification on graphs with structural regularized GCNs[C]//Proceedings of the 7th International Conference on Learning Representations, Conference Track Proceedings. New Orleans:OpenReview.net, 2019:1-11.

[8]XU K, LI C, TIAN Y, et al.Representation learning on graphs with jumping knowledge networks[C]//35th International Conference on Machine Learning,volume 12 of 13.New York: Curran Associates, Inc.,2018:8662-8675.

[9]VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[EB/OL].(2016-06-13)[2023-12-16].https://arxiv.org/abs/1606.04080.

[10]SNELL J, SWERSKY K, ZEMEL R S.Prototypical networks for few-shot learning[EB/OL].(2017-03-15)[2023-12-09]. https://arxiv.org/abs/1703.05175.

[11]SUNG F, YANG Y, ZHANG L, et al. Learning to compare: relation network for few-shot learning[EB/OL].(2017-11-16)[2024-01-03].https://arxiv.org/abs/1711.06025.

[12]FINN C, ABBEEL P, LEVINE S.Model-agnostic meta-learning for fast adaptation of deep networks[C]//34th International Conference on Machine Learning,volume 3 of 8.New York:Curran Associates, Inc.,2018:1856-1868.

[13] 代佳梅 , 孔韦韦 , 王泽 , 等 .BERT 和 LSI 的端到端方面级情感分析模型 [J/OL]. 计算机工程与应用 ,1-13[2024-02-21]. http://kns.cnki.net/kcms/detail/11.2127.TP.20230706.1954.034.html.

[14]BRODY S, ALON U, YAHAV E.How attentive are graph attention networks?[EB/OL].(2021-05-30)[2024-01-16].https://arxiv.org/abs/2105.14491.

【作者简介】

张中天（2000—），通信作者（email：zhangzhongtian@qq.com），男，河南安阳人，硕士，研究方向：计算机视觉、自然语言处理。

王涛（2000—），男，安徽安庆人，硕士，研究方向：计算机视觉。

谭莹莹（1999—），女，湖南株洲人，硕士，研究方向：计算机视觉。

熊伟（1976—），男，辽宁锦州人，博士，高级工程师，研究方向：计算机网络、机器学习、自然语言处理。

（收稿日期：2024-02-22）