

# 基于元路径的图 Transformer 神经网络<sup>①</sup>



梁书晴, 蒋运承

(华南师范大学 计算机学院, 广州 510631)

通信作者: 蒋运承, E-mail: [ycjiang@scnu.edu.cn](mailto:ycjiang@scnu.edu.cn)

**摘要:** 图神经网络作为一种新的深度学习模型, 被广泛运用在图数据中, 并极大地推动了推荐系统、社交网络、知识图谱等应用的发展. 现有的异构图神经网络通常事先定义了多条元路径来学习异构图中的复合关系. 然而, 这些模型通常在特征聚合步骤中只考虑单条元路径, 导致模型只关注了元路径的局部结构, 忽略了元路径之间的全局相关性; 还有一些模型则是忽略掉了元路径的中间节点和边信息, 导致模型无法学习到元路径内部的语义信息. 针对以上问题, 本文提出一种基于元路径的图 Transformer 神经网络 (MaGTNN). 该模型首先将异构图采样为基于元路径的多关系子图, 利用提出的位置编码和边编码的方法来获取元路径中的语义信息. 随后使用改进的图 Transformer 层计算出目标节点与其元邻居的相似度, 并利用该相似度来聚合其所有的元邻居信息. 在 3 个公开数据集的节点分类和节点聚类任务中, MaGTNN 均高于最新的基准模型.

**关键词:** 异构图; 异构图嵌入; 元路径; 图神经网络; Transformer

引用格式: 梁书晴, 蒋运承. 基于元路径的图 Transformer 神经网络. 计算机系统应用, 2022, 31(11): 1-9. <http://www.c-s-a.org.cn/1003-3254/8802.html>

## Metapath-based Graph Transformer Neural Network

LIANG Shu-Qing, JIANG Yun-Cheng

(School of Computer Science, South China Normal University, Guangzhou 510631, China)

**Abstract:** As new deep learning models, graph neural networks are widely used in graph data and promote various applications, such as recommendation systems, social networks, and knowledge graphs. Most existing heterogeneous graph neural models usually predefine multiple metapaths to capture composite relationships in heterogeneous graphs. However, some models usually consider one metapath during the feature aggregation, leading to models only learning neighbor structure but ignoring the global correlation of multiple metapaths. Others omit intermediate nodes and edges along the metapath, which means models cannot learn the semantic information in each metapath. To address those limitations, this study proposes a new model named metapath-based graph Transformer neural network (MaGTNN). Specifically, MaGTNN first samples heterogeneous graph as metapath-based multi-relation graph and then uses the proposed position encoder and edge encoder to capture the semantic information in a metapath. Subsequently, all the metapath-based neighbor information is aggregated to the target node through their similarity, which is calculated by the improved graph Transformer layer. Extensive experiments on three real-world heterogeneous graph datasets for node classification and node clustering show that MaGTNN achieves more accurate prediction results than state-of-the-art baselines.

**Key words:** heterogeneous graph; heterogeneous graph embedding; metapath; graph neural network; Transformer

<sup>①</sup> 基金项目: 国家自然科学基金 (61772210)

收稿时间: 2022-03-03; 修改时间: 2022-04-02; 采用时间: 2022-04-15; csa 在线出版时间: 2022-07-15

许多真实世界的数据集都是用图来表示的,如社交网络<sup>[1,2]</sup>,物理系统<sup>[3,4]</sup>,交通网络<sup>[5,6]</sup>,引用网络<sup>[1,7,8]</sup>,推荐系统<sup>[9,10]</sup>和知识图谱<sup>[11,12]</sup>等.由于图数据的每个节点的邻域大小都是不一致的,而绝大多数深度学习模型,如CNN, RNN等都先假设输入的数据为有序、大小固定的欧几里得数据,这使得我们无法直接使用现有的深度学习模型来处理图数据.因此,研究者们提出了图嵌入模型将图数据转化为欧几里得数据并应用于实际任务中,如社群分割、交通流量预测、蛋白质结构预测等.

图嵌入模型是一种将图数据映射到低维向量空间的算法.早期的图嵌入模型,如LINE<sup>[13]</sup>通过计算节点之间的一阶和二阶邻域的相似度来生成节点嵌入. DeepWalk<sup>[14]</sup>、Node2Vec<sup>[15]</sup>和TADW<sup>[16]</sup>等方法,利用随机游走算法,生成图的节点序列,并利用跳字模型(skip-gram<sup>[17]</sup>)来计算节点的嵌入.随着深度学习的快速发展,图神经网络(GNNs)被提出,它使用专门设计的神经层来学习图的表示.基于谱域的图神经网络,包括ChebNet<sup>[18]</sup>和GCN<sup>[19]</sup>等,先对图数据进行图傅立叶变换,将空域信号转化为谱域信号.在此基础上使用卷积算子来进行特征扩散,最后使用逆傅立叶变换将谱信号转化为空域信号.基于空间的图神经网络,包括GraphSAGE<sup>[1]</sup>, GAT<sup>[20]</sup>和许多其他变体<sup>[6,10,21]</sup>,通过直接在空域中进行信息传播,解决了基于谱的模型的可扩展性低、泛化能力不好和计算量大的问题.

尽管图神经网络在许多任务上都取得了很好的结果,但是他们都在异构图上表现不佳.其主要原因是这些方法没有考虑异构图的异构信息,如:节点/边的属性/类型等、图的时间信息等.现有多数异构图模型都基于元路径的思想.元路径是一条抽象节点与具体边交替而成的序列.比如在包含作者(authors)、论文(papers)、会议(venue)3种节点类型的学者网络中,作者-论文-作者(APA)和作者-论文-会议-论文-作者(APVPA)是两条可以描述作者节点之间不同关系的元路径.元路径APA表示为两个作者为共同作者;APVPA则可以表示为两个作者都在同一个会议里面发表过论文.我们发现相比于传统的图嵌入模型和图神经网络方法,元路径包含了异构图中的异构信息、语义信息和高阶结构信息.目前基于元路径的异构图嵌入方法主要有如下问题:1)模型通常只考虑首尾的节点,忽略了元路径中间节点和关系(如HERec<sup>[22]</sup>和HAN<sup>[23]</sup>).2)模型

只考虑单一元路径(ESim<sup>[24]</sup>、HIN2vec<sup>[25]</sup>)信息,这会导致节点丢失掉来自其他元路径的邻居信息,从而导致次优的性能.

为了解决这些局限性,我们提出了一种新的基于元路径聚合与Transformer模型<sup>[26]</sup>的异构图神经网络(MaGTNN).MaGTNN通过修改过的Transformer模型,来捕获当前节点下所有的元路径邻居和元路径的语义信息.具体来说, MaGTNN首先利用特定的线性变换矩阵来将不同属性的节点映射到相同的向量空间中.接下来, MaGTNN会预先计算当前节点与其元邻居之间的位置编码.随后, MaGTNN利用节点自身的特征、位置编码和边编码来计算当前节点和目标节点之间的注意力,并在特征聚合操作中利用计算出来的注意力将元邻居的特征聚合到当前节点.在元路径注意力计算步骤中, MaGTNN可以从相邻节点和中间的元路径上下文中捕获异构图的结构和语义信息;在元邻居节点聚合的步骤中, MaGTNN可以捕获多条元路径的特征.

总之,本文的主要工作有以下几个方面的贡献:

(1) 本文提出了一种新的基于元路径聚合与Transformer模型的异构图神经网络(MaGTNN).

(2) 本文设计了一个全新元路径注意力计算方法,该方法可以考虑到了元路径中所有的节点特征和边信息,具有更好的特征捕获能力.

(3) 本文在DBLP、ACM和IMDB数据集上进行了节点分类和节点聚类的实验,以评估本文提出的模型的性能.在所有这些数据集中和任务上的实验表明, MaGTNN学习的节点嵌入在分类、聚类两个下游任务上均优于其他基准模型生成的节点嵌入.

## 1 相关工作

现有的异构图嵌入方法主要目的是尽可能地保留图的异构结构信息. PME<sup>[27]</sup>、EOE<sup>[28]</sup>和HeGAN<sup>[29]</sup>等方法只考虑节点一阶信息,即节点的链接.这类方法将每个链接类型视为一个关系,并使用特定的关系矩阵将节点转换到不同的度量空间.这样,通过不同类型的链路连接的节点可以在不同的度量空间中彼此接近,从而捕获图的异构性.由于基于链接的方法只能获得异构图的局部结构,因此研究者们试图引入复杂的高阶结构信息来获得一个更好的图嵌入.比较常用的异

异构图高阶结构信息有路径和子图. Metapath2Vec<sup>[30]</sup>及其衍生 Metapath2Vec++, Spacey<sup>[31]</sup>, JUST<sup>[32]</sup>, BHIN2vec<sup>[33]</sup>, HHNE<sup>[34]</sup> 主要利用元路径引导的随机游走生成语义丰富的异构节点序列; 然后利用跳字模型 (skip-gram) 来保持节点 $v$ 与其上下文节点之间的邻近性, 即随机游走序列中的邻居; Metagraph2Vec<sup>[35]</sup>, DHNE<sup>[36]</sup>, HEBE<sup>[37]</sup> 等方法则是考虑了子图信息. Metagraph2Vec 利用元图引导的随机游走生成异构节点序列. 然后采用跳字模型来学习节点嵌入. 基于此策略, Metagraph2Vec 可以捕获丰富的结构信息和节点间的高阶相似性. DHNE 和 HEBE 则是基于超边的图嵌入方法. DHNE 使用了 3 层神经网络和非线性的元组相似度以学习超边的高阶结构信息. HEBE 则是把与一个事件相关的所有节点都关联到一个超边中, 并且在训练中使节点与其所

属超边之间的近邻度最大化. 在最大化近邻度后, HEBE 可以保持同一超边内的节点相似度, 同时降低属于不同超边的节点相似度.

## 2 预备知识

### 2.1 异构图

异构图 (heterogeneous graph) 可表示为  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , 其中  $\mathcal{V}$  表示节点的集合,  $\mathcal{E}$  表示边的集合. 对于每一个节点存在一种映射关系  $\phi(v): \mathcal{V} \rightarrow \mathcal{A}$ , 对于每一条边存在一种映射关系  $\varphi(e): \mathcal{E} \rightarrow \mathcal{R}$ ,  $\mathcal{A}$  和  $\mathcal{R}$  分别代表节点类型和边类型, 且  $|\mathcal{A}| + |\mathcal{R}| > 2$ . 图 1(a) 即为一个异构图, 该异构图包含 3 个类型的节点, 分别为作者节点、论文节点和会议节点; 一个类型的边, 即代表两个节点之间存在某种关联.

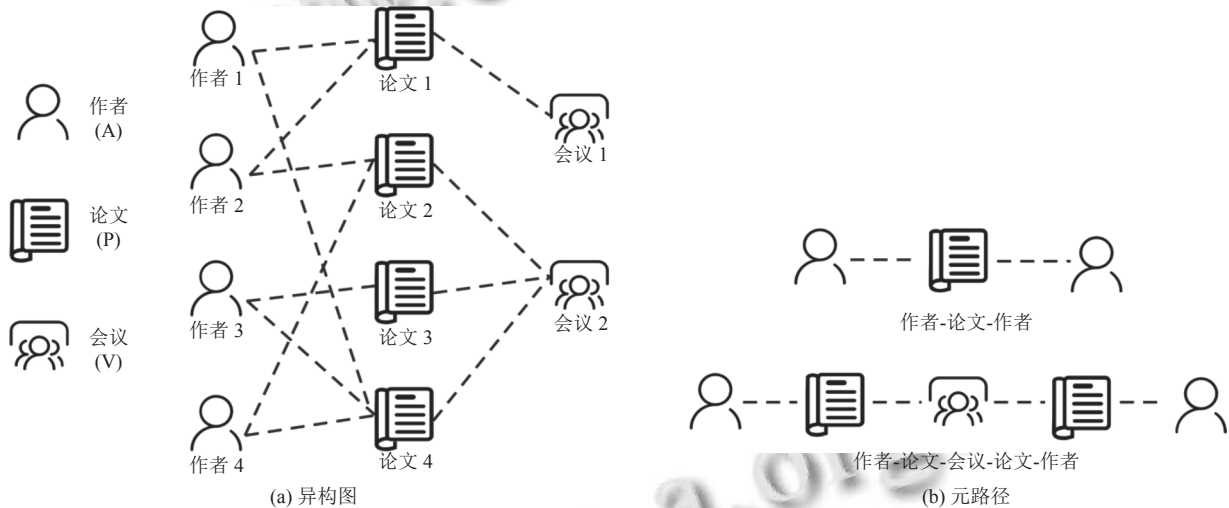


图 1 异构图和元路径图示

### 2.2 异构图嵌入

对于给定的一个异构图  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , 对图节点类型  $A_i \in \mathcal{A}$ , 图节点属性矩阵为  $X_{A_i} \in \mathbb{R}^{\mathcal{V}_{A_i} \times d_{A_i}}$ . 异构图嵌入的任务是学习一个映射函数  $\mathcal{F}: \mathcal{V} \rightarrow \mathbb{R}^d$ , 其中  $d \ll |\mathcal{V}|$ , 使得图中的节点能表示为隐空间  $\mathbb{R}^d$  中的向量, 并且保留图中丰富的结构信息和语义信息.

### 2.3 元路径

对于给定的一个异构图, 元路径  $P$  可定义为一组连接多种类型节点的复合关系, 具体形式为  $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_l} A_{l+1}$ , 其中  $R_i$  表示节点间的关系. 节点类型  $A_1, A_{l+1}$  之间的复合关系记为  $R = R_1 \circ R_2 \circ \dots \circ R_l$ . 图 1(b) 即为预先定义的两条元路径, 其分别代表共同作者关系和共同在某

个会议发表过论文关系.

### 2.4 Transformer 架构

Transformer 架构是由一系列 Transformer 层组成, 每一个 Transformer 层都含有如下两个模块: 自注意力模块和前馈神经网络模块. 令  $H = [h_1^T, h_2^T, \dots, h_n^T]^T \in \mathbb{R}^{n \times d}$  为自注意力模块的输入, 其中,  $d$  为隐藏层维度. 输入  $H$  由 3 个向量  $Q, K, V$  来计算其注意力:

$$Q = HW_Q, K = HW_K, V = HW_V \quad (1)$$

$$E = \frac{QK^T}{\sqrt{d_K}}, \text{Att}(H) = \text{Softmax}(E)V \quad (2)$$

其中,  $W_Q \in \mathbb{R}^{d \times d_K}, W_K \in \mathbb{R}^{d \times d_K}, W_V \in \mathbb{R}^{d \times d_K}$  分别为  $Q, K, V$

的映射矩阵.

### 3 基于元路径的图 Transformer 网络

本文提出了一种基于元路径的图 Transformer 神经网络 (MaGTNN) 来同时捕获元路径自身的语义信息与多条元路径间的全局相关性. 在第 3.2 节和第 3.3 节, 我们介绍了 MaGTNN 几个关键的特征计算方法. 在第 3.4 节, 我们则展示了每一个 MaGTNN 层的具体实现细节. 下面将对 MGTNN 中的各个组件进行详细介绍.

#### 3.1 模型输入

首先, 对于一个输入的异构图  $\mathcal{G}$  及其节点特征  $h_i \in \mathbb{R}^{d_n}$ , 节点  $ij$  之间的边特征  $e_{ij} \in \mathbb{R}^{d_e}$ . 由于节点的类型不同, 他们的自身特征属性也有可能不一样, 比如在 IMDB 中, 电影节点和演员节点的属性特征是不一样的, 电影节点可能包含电影类型、电影时长等特征, 而演员节点可能包含演员的姓名、性别等特征. 因此, 我们首先对节点特征进行一个线性变换:

$$h'_i = M_{A_i} \cdot h_i \quad (3)$$

其中,  $M_{A_i}$  为转换矩阵,  $A_i$  为节点  $i$  所对应类型. 将变换之后的节点特征作为初始输入.

#### 3.2 位置编码

在自然语言处理中, 基于 Transformer 架构的模型通常对每一个单词的词嵌入上加入单词的位置编码. 这一步的目的在于确保句子中每一个单词的编码都是独一无二的. 并且, 由于注意力机制通常会丢失单词之间的顺序信息, 因此加上单词的位置编码也可以保留词之间的顺序关系<sup>[37]</sup>. 然而对于图结构来说, 节点并没有组织成序列. 因此, 早期的图神经网络在单层网络中只考虑了直接连接节点, 并使用堆叠 (stack) 的方法来获得图的高阶结构信息. 这些网络通常只能获得局部的结构信息, 缺少图的全局结构信息. 最近的一些研究 RPGNN<sup>[38]</sup>, CGNNs<sup>[39]</sup>, DEGNN<sup>[40]</sup> 和文献 [41] 都在尝试在训练时同时融入节点的位置信息和图的结构信息. 本文的位置编码参考了文献 [41, 42], 使用了拉普拉斯特征向量作为位置编码. 拉普拉斯向量定义如下:

$$\Delta = I - D^{-1/2} A D^{-1/2} = U^T \Lambda U \quad (4)$$

其中,  $A$  为  $n \times n$  的邻接矩阵,  $D$  为度矩阵,  $\Lambda$ ,  $U$  分别代表特征值和特征向量. 对于节点  $i$ , 我们选择了  $k$  个最小的非平凡特征向量作为其位置编码并记为  $\lambda_i$ . 值得注意的

是, 直接将节点位置编码与节点特征相加会显著提高显存消耗. 因此本文将位置编码使用一个线性变换将其作为注意力系数的偏置量, 这个做法与和节点特征相加是等效的.

#### 3.3 边编码

在异构图中, 边通常也包含结构信息, 如 IMDB 中, 演员-电影和电影-导演之间的边都包含不同的语义信息. 这种语义信息对于整个图的特征表示和节点的特征聚合是非常重要的. 在之前的基于元路径的图神经网络中, 研究者们通常只关注元路径中的节点特征, 忽略掉了元路径中的边的特征. 因此, 为了获得一个更加合理的, 我们在注意力计算步骤中加入了节点编码, 并将其作为一个偏置量, 来区分在不同元路径下节点的注意力.

记  $P_r(i, j) = (R_1, R_2, \dots, R_N)$  为节点  $i, j$  中一条元路径中的边序列, 则元路径中边编码计算为:

$$\delta_{ij} = \frac{1}{N} \sum_{n=1}^N x_{R_n} (w_n^R)^T \quad (5)$$

其中,  $x_{R_n}$  为边序列  $P_r(i, j)$  中的第  $n$  条边的特征,  $w_n^R \in \mathbb{R}^{d_e}$  为对应的权重向量,  $\mathbb{R}^{d_e}$  为边特征的维度.

#### 3.4 基于元路径的图 Transformer 层

基于元路径的图 Transformer 层是在原始 Transformer 架构上的改进. 下面我们来介绍如何在节点特征在层  $\ell$  处是如何进行更新.

对于给定的元路径集  $\mathcal{P}$ , 我们首先利用路径集  $\mathcal{P}$  来构建基于元路径的邻接矩阵  $Meta_{adj}$ , 并利用式 (4) 来计算节点  $i$  与其元路径邻居  $j$  之间的距离编码  $\lambda_{ij}$ .

对于  $Meta_{adj}$  中的一条元路径  $P$ , 记  $P(i, j)$  为一条元路径实例,  $i$  为当前聚合节点,  $j \in \mathcal{N}_i^P$  为元路径邻居. 为了学习到元路径  $P(i, j)$  中的中间节点的特征, 我们首先将其聚合为一个向量. 由于异构图中含有多个关系, 我们在这里参考了 MaGNN<sup>[42]</sup>, RotatE<sup>[12]</sup>, 将关系向量映射到了复空间. 记  $P(i, j) = (t_0, t_1, \dots, t_n)$ ,  $t_i$  为异构图的节点, 且  $t_0 = i$ ,  $t_1 = j$ ,  $R_m$  为节点  $t_{m-1}$  和  $t_m$  之间的关系,  $r_m$  为关系向量, 则元路径聚合步骤为:

$$\begin{cases} o_0 = h'_{t_0} = h'_i \\ o_m = h'_{t_i} + o_{i-1} \odot r_i \end{cases} \quad (6)$$

$$h_{P(i,j)} = \frac{o_n}{n+1} \quad (7)$$

在完成中间节点聚合之后, 我们使用 Transformer

的注意力计算方式来计算元路径实例 $P(i, j)$ 与目标节点 $j$ 之间的注意力系数, 并利用该注意力系数来更新目标节点 $j$ 的特征. 注意力计算公式及节点更新函数为:

$$e_{ij}^{\ell} = \frac{(h_i'^{(\ell)} W_Q^{k\ell})(h_j'^{(\ell)} W_K^{k\ell})}{\sqrt{d}} + b^{\ell} \cdot \lambda_{ij} + \delta_{ij} \quad (8)$$

$$\alpha_{ij}^{\ell} = \text{Softmax}_j(e_{ij}^{\ell}) \quad (9)$$

$$h_i''^{(\ell)} = O_h^{\ell} \parallel \left( \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{k,\ell} V^{k,\ell} h_j^{\ell} \right) \quad (10)$$

其中,  $W_Q^{k\ell}, W_K^{k\ell}$ 为式(1)在第 $\ell$ 层 $k$ 个注意力头中的映射矩阵,  $b^{\ell}$ 为一个可学习矩阵,  $\delta_{ij}$ 为式(4)中边的编码,  $O_h^{\ell} \in \mathbb{R}^{d \times d}$ ,  $k \in [1, T]$ 为多头注意力数量,  $\parallel$ 为向量拼接操作.

整个基于元路径的 Transformer 层的最终输出还需要经过一个正则化层 (layer norm)、残差连接 (residual connection) 和一个前馈神经网络 (FNN):

$$\hat{h}_i^{(\ell+1)} = \text{Norm}(h_i'^{(\ell)} + h_i''^{(\ell+1)}) \quad (11)$$

$$\hat{h}_i^{(\ell+1)} = W_2^{\ell} \text{ReLU}(W_1^{\ell} \hat{h}_i^{\ell}) \quad (12)$$

$$h_i^{(\ell+1)} = \text{Norm}(\hat{h}_i^{\ell} + \hat{h}_i^{\ell+1}) \quad (13)$$

其中,  $W_1^{(\ell)} \in \mathbb{R}^{2d \times d}$ ,  $W_2^{\ell} \in \mathbb{R}^{d \times 2d}$ 为可学习映射矩阵.

## 4 实验

### 4.1 数据集简介

为了评估 MaGTNN 模型, 本文在 3 个异构图数据集: DBLP (<http://web.cs.ucla.edu/~yzsun/data/>), ACM (<https://github.com/Jhy1993/HAN>) 和 IMDB (<https://www.kaggle.com/datasets/karrimba/movie-metadatasv>) 上进行节点分类与节点聚类实验. 表 1 中展示了实验使用的异构图数据集的基本统计信息, 各数据集的划分遵循文献 [23] 的设置.

表 1 数据集统计信息

数据集	节点	边数	特征维度	元路径	训练集	验证集	测试集
DBLP	#作者 (A): 4057	#A-P: 19645	334	APA	400	400	3257
	#论文 (P): 14328	#P-T: 85810		APTPA			
	#主题 (T): 7723	#P-V: 14328		APVPA			
	#会议 (V): 20						
ACM	#文章 (P): 3025	#P-A: 9936	1902	PAP	400	400	2225
	#作者 (A): 5835	#A-S: 3025		PSP			
	#主题 (S): 57						
IMDB	#电影 (M): 4278	#M-D: 4278	1256	MAM	400	400	3478
	#导演 (D): 2081	#M-A: 12828		MDM			
	#演员 (A): 5257			DMD			
				DMAMD			
				AMA			
				AMDMA			

DBLP 是一个计算机科学领域的文献集成网站. 在本文中, 我们使用的是从 DBLP 中抽取的一个子集<sup>[43,44]</sup>, 其中包含 4057 个作者, 14328 篇论文, 7723 个主题和 20 个出版社. 作者被划分为 4 个研究领域 (数据库, 数据挖掘, 人工智能和信息检索). 每一个作者都由其论文中的关键词所构成的词袋来表示. 在本文中, 作者节点被划分为训练集, 验证集和其测试集, 其中各个划分所包含的节点数目分别为: 400 个 (9.86%), 400 个 (9.86%), 和 3257 个 (80.28%).

IMDB 是一个互联网电影数据库. 我们从 IMDB 中抓取了一个子集来构建数据集, 其中包括 4278 部电影, 2081 个导演, 和 5257 个演员. 电影节点被分为

3 个类型 (动作片, 喜剧和戏剧). 每个电影都的特征由其情节的关键词所构成的词袋表示. 在本文中, 电影节点被划分为训练集, 验证集和其测试集, 其中各个划分所包含的节点数目分别为: 400 个 (9.35%), 400 个 (9.35%) 和 3478 个 (81.30%).

ACM 数据集是从发表在 KDD, SIGMOD, SIGCOMM, MobiCOMM 和 VLDB 中论文抽取构建来. 我们从其中抽取了 3025 篇论文, 其中包含 5835 个作者和 56 个主题. 论文被大致划分为 3 类 (数据库, 无线通信和数据挖掘). 论文节点的特征由其关键词词袋所表示, 标签为其发表的会议. 同时, 在本文中, 论文节点被划分为训练集, 验证集和其测试集, 其中各个划分

所包含的节点数目分别为: 400 个 (13.22%), 400 个 (13.22%) 和 2225 个 (73.56%)。

## 4.2 对比方法及参数设置

本文选取了同构图嵌入模型、同构图神经网络和异构图神经网络作为基准模型与 MaGTNN 的实验结果进行比较, 各基准模型介绍如下。

**Node2Vec<sup>[15]</sup>**: 基于随机游走算法的同构图嵌入模型。该模型使用随机游走算法来生成图的节点序列, 并利用跳字模型来学习节点嵌入。

**Metapath2Vec<sup>[25]</sup>**: 基于元路径随机游走算法的异构图嵌入模型。该模型通过元路径随机游走生成节点序列后使用跳字模型来学习节点嵌入。

**HERec<sup>[22]</sup>**: 该模型在元路径随机游走的基础上删除掉游走节点序列中与起始节点类型不同的节点, 将其原始的异构节点序列转化为同构节点序列, 随后在所获得的同构游走序列上使用 Node2Vec 来计算节点嵌入。该模型可以在学习到异构信息的同时降低训练的难度。

**GCN<sup>[8]</sup>**: 基于谱域的卷积图神经网络。该模型通过对图的拉普拉斯矩阵上的卷积操作, 并堆叠了多个图卷积层, 来捕获图的高阶结构信息。

**GAT<sup>[20]</sup>**: 基于自注意力机制的图神经网络。该模型通过自注意力机制来对邻居节点进行聚合, 实现了对不同邻居权值的自适应匹配。GAT 堆叠多个自注意力层来获得图的高阶结构信息。

**HAN<sup>[23]</sup>**: 基于层次注意力的异构图神经网络。该模型包含了节点级注意力和语义级注意力。通过使用节点级注意力来学习元路径的邻居节点权重并将其聚合获得当前语义的节点嵌入, 最后利用语义级别注意力机制评判每个元路径的权重并进行聚合, 获得最终的节点嵌入。

实验中对于基于随机游走算法的模型 (Node2Vec、Metapath2Vec、HERec), 我们将游走窗口设置为 5, 游走长度设置为 100, 每个节点游走次数为 40 次, 负采样设置为 5。对于同构图模型 (Node2Vec、GCN 和 GAT), 本文将原始异构图直接输入, 忽略节点和边的异质性。对于基于注意力的图神经网络模型 (GAT、HAN 以及 MaGTNN), 多头注意力的注意力头数统一设置为 8, 注意力向量维度设置为 128。为了保证公平性, 所有模型生成的节点嵌入维度均设置为 128。由于图神经网络模型 (GCN、GAT、HAN 和 MaGTNN) 在训练的损失函数计算时使用了训练集和验证集的节点标签, 因此,

为了保证标签不被泄漏, 我们仅使用测试集的节点作为下游任务的实验评估。对于下游任务的评价指标, 在节点分类任务上, 我们采用 *micro-F1* 作为评价指标, 其常用来衡量多分类任务下模型的准确度; 在节点聚类任务上, 我们采用聚类任务常用的评价指标: 标准化互信息 (NMI) 和调整兰德系数 (ARI)。

## 4.3 节点分类任务

在节点分类任务中, 本文将各模型训练出来的节点嵌入直接输入到一个线性支持向量机 (support vector machine, SVM) 中, 计算 SVM 分类结果的 *micro-F1* 值作为评估。我们还测试了各模型在不同训练率 (在训练集中随机选取 20%–80% 的节点进行训练) 下的节点分类结果。表 2 记录的是在同样的数据集划分和训练率, 所有模型 10 次运行结果的平均值。

从表 2 中可以看出, MaGTNN 在不同数据集和训练率下都普遍优于其他的基准模型。同时在 3 个数据集中我们可以看到, 忽略掉异构信息的模型 Node2Vec 反而在 ACM 和 IMDB 数据集上都普遍优于只考虑单条元路径异构图模型 Metapath2Vec, HERec。这是由于元路径本来就是一个信息损失的过程, 使用元路径会在一定程度上减少模型可以获得的路径数量, 这很大程度上影响了模型的效果。图神经网络模型, 尤其是异构图神经网络模型在所有数据集和训练率都优于传统图嵌入模型。相比于目前的最优表现模型 HAN, MaGTNN 在 3 个数据集上面都有 1%–4% 的提升。这表示在注意力计算时, 考虑位置和边信息的注意力计算会有更好的表现。

## 4.4 节点聚类任务

我们也在所有数据集上对各模型进行了节点聚类实验。在这一部分中, 我们使用 K-均值 (K-means) 算法对各模型训练出来的节点嵌入进行聚类运算。我们将数据集的节点标签数设置为 K-means 算法的聚类簇数。与节点分类实验相同, 我们对每个数据集进行了 10 次实验, 并取所有结果平均值作为最终结果记录于表 3 中。

从表 3 中可见, 在 ACM 和 DBLP 数据集中, 异构图神经网络模型 HAN、MaGTNN 效果要优于传统图嵌入模型 (Node2Vec、Metapath2Vec、HERec) 和同构图神经网络模型 (GCN、GAT), 这证明了结合元路径信息和图神经网络能更有效学习图的异构信息。同时, 相对于 HAN, MaGTNN 在以上两个数据集上均提升了 2% 左右, 其中一个原因为 MaGTNN 在进行结构学

习时考虑了边信息对注意力的影响;此外, Node2Vec 模型在聚类任务中表现优异,这是由于在随机游走过程中,邻近的节点在有更大的概率在同一条游走序列中,因此在特征聚合时,这些节点在欧式空间中的映射也会分布得更加靠近,在聚类任务中也更容易分为同一簇.相对于 HAN, MaGTNN 在输入的特征中加入了

位置编码,这可能是 MaGTNN 表现优异的另一个原因.此外,值得注意的是在聚类任务中,所有模型在 IMDB 数据集上表现均不佳,其原因可能是由该数据集中电影节点的标签导致:在 IMDB 中一部电影含有多个类别作为其标签,而在聚类问题中,我们只能选择一个类别作为其标签,这极大的影响到了聚类的准确性.

表2 数据集在节点分类任务上的结果 (%)

数据集	训练率	Node2Vec	Metapath2Vec	HERec	GCN	GAT	HAN	MaGTNN
DBLP	20	87.23	89.52	91.26	92.44	93.51	93.37	<b>94.27</b>
	40	88.26	90.31	92.05	92.83	93.65	93.40	<b>94.30</b>
	60	89.08	91.04	92.66	92.78	93.72	93.6	<b>94.99</b>
	80	89.96	91.80	92.78	93.12	93.95	93.76	<b>95.40</b>
ACM	20	76.66	68.61	69.72	89.45	88.67	90.19	<b>92.17</b>
	40	80.67	73.25	71.62	89.48	89.24	90.66	<b>92.54</b>
	60	82.38	75.30	72.89	89.96	89.57	91.21	<b>92.32</b>
	80	83.75	76.40	74.66	90.27	91.29	90.17	<b>92.77</b>
IMDB	20	49.32	47.63	46.23	52.80	53.64	56.31	<b>59.88</b>
	40	51.77	48.98	47.87	53.76	55.56	57.47	<b>60.85</b>
	60	52.79	49.87	48.19	54.23	56.52	58.65	<b>61.58</b>
	80	52.72	50.56	49.11	54.63	57.98	59.70	<b>62.13</b>

表3 数据集在节点聚类任务上的结果 (%)

数据集	评价指标	Node2Vec	Metapath2Vec	HERec	GCN	GAT	HAN	MaGTNN
DBLP	NMI	75.01	74.19	69.65	73.17	75.89	78.63	<b>81.33</b>
	ARI	80.75	78.11	73.73	77.24	82.13	84.41	<b>86.92</b>
ACM	NMI	51.16	19.63	20.50	67.41	67.92	71.88	<b>73.89</b>
	ARI	45.39	9.77	22.97	72.82	73.38	76.43	<b>79.27</b>
IMDB	NMI	0.71	0.89	0.39	7.42	7.83	10.90	<b>14.28</b>
	ARI	0.28	0.22	0.11	7.25	8.87	<b>12.88</b>	10.23

## 5 结论与展望

本文针对异构图提出了一种基于元路径的图 Transformer 神经网络 (MaGTNN) 模型来学习节点嵌入.我们提出的模型首先将异构图使用元路径进行采样,构建出基于元路径的多关系子图,通过引入了元路径上的位置编码和节点编码, MaGTNN 相比其他模型能够获得更多的语义信息.并且, MaGTNN 使用了经过改进的图 Transformer 层来聚合元邻居下的所有节点特征,可以学习到元路径间的全局相关性.本文在 3 个真实数据集上进行了节点聚类和分类实验,实验表明 MaGTNN 在这两个任务上相较其他基准模型一致取得最好的预测效果.

### 参考文献

- Hamilton WL, Ying Z, Leskovec J. Inductive representation learning on large graphs. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 1025–1035.
- Wang DX, Cui P, Zhu WW. Structural deep network embedding. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. California: ACM, 2016. 1225–1234. [doi: 10.1145/2939672.2939753]
- Battaglia P, Pascanu R, Lai M, *et al.* Interaction networks for learning about objects, relations and physics. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 4509–4517.
- Fout A, Byrd J, Shariat B, *et al.* Protein interface prediction using graph convolutional networks. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6533–6542.
- Li YG, Yu R, Shahabi C, *et al.* Diffusion convolutional recurrent neural network: Data-driven traffic forecasting.

- Proceedings of the 6th International Conference on Learning Representations. Vancouver: ICLR, 2018.
- 6 Zhang JN, Shi XJ, Xie JY, *et al.* GaAN: Gated attention networks for learning on large and spatiotemporal graphs. Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence. Monterey: AUAI Press, 2018. 339–349.
  - 7 Atwood J, Towsley D. Diffusion-convolutional neural networks. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 2001–2009.
  - 8 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. Proceedings of the 5th International Conference on Learning Representations. Toulon: ICLR, 2017.
  - 9 van den Berg R, Kipf TN, Welling M. Graph convolutional matrix completion. arXiv: 1706.02263, 2017.
  - 10 Zhang JN, Shi XJ, Zhao SL, *et al.* STAR-GCN: Stacked and reconstructed graph convolutional networks for recommender systems. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: IJCAI, 2019. 4264–4270. [doi: [10.24963/ijcai.2019/592](https://doi.org/10.24963/ijcai.2019/592)]
  - 11 Bordes A, Usunier N, Garcia-Durán A, *et al.* Translating embeddings for modeling multi-relational data. Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2013. 2787–2795.
  - 12 Sun ZQ, Deng ZH, Nie JY, *et al.* RotatE: Knowledge graph embedding by relational rotation in complex space. Proceedings of the 7th International Conference on Learning Representations. New Orleans: ICLR, 2019.
  - 13 Tang J, Qu M, Wang MZ, *et al.* LINE: Large-scale information network embedding. Proceedings of the 24th International Conference on World Wide Web. Florence: International World Wide Web Conferences Steering Committee, 2015. 1067–1077. [doi: [10.1145/2736277.2741093](https://doi.org/10.1145/2736277.2741093)]
  - 14 Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014. 701–710. [doi: [10.1145/2623330.2623732](https://doi.org/10.1145/2623330.2623732)]
  - 15 Grover A, Leskovec J. Node2Vec: Scalable feature learning for networks. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. California: ACM, 2016. 855–864. [doi: [10.1145/2939672.2939754](https://doi.org/10.1145/2939672.2939754)]
  - 16 Yang C, Liu ZY, Zhao DL, *et al.* Network representation learning with rich text information. Proceedings of the 24th International Conference on Artificial Intelligence. Buenos: AAAI Press, 2015. 2111–2117.
  - 17 Mikolov T, Chen K, Corrado G, *et al.* Efficient estimation of word representations in vector space. Proceedings of the 1st International Conference on Learning Representations. Scottsdale: ICLR, 2013.
  - 18 Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 3844–3852.
  - 19 Veličković P, Cucurull G, Casanova A, *et al.* Graph attention networks. Proceedings of the 6th International Conference on Learning Representations. Vancouver: ICLR, 2018.
  - 20 Li YJ, Tarlow D, Brockschmidt M, *et al.* Gated graph sequence neural networks. Proceedings of the 4th International Conference on Learning Representations. San Juan: ICLR, 2016.
  - 21 Shi C, Hu BB, Zhao WX, *et al.* Heterogeneous information network embedding for recommendation. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(2): 357–370. [doi: [10.1109/TKDE.2018.2833443](https://doi.org/10.1109/TKDE.2018.2833443)]
  - 22 Wang X, Ji HY, Shi C, *et al.* Heterogeneous graph attention network. Proceedings of The World Wide Web Conference. San Francisco: ACM, 2019. 2022–2032. [doi: [10.1145/3308558.3313562](https://doi.org/10.1145/3308558.3313562)]
  - 23 Shang JB, Qu M, Liu JL, *et al.* Meta-path guided embedding for similarity search in large-scale heterogeneous information networks. arXiv: 1610.09769, 2016.
  - 24 Fu TY, Lee WC, Lei Z. HIN2vec: Explore meta-paths in heterogeneous information networks for representation learning. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore: ACM, 2017. 1797–1806. [doi: [10.1145/3132847.3132953](https://doi.org/10.1145/3132847.3132953)]
  - 25 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
  - 26 Chen HX, Yin HZ, Wang WQ, *et al.* PME: Projected metric embedding on heterogeneous networks for link prediction. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018. 1177–1186. [doi: [10.1145/3219819.3219832](https://doi.org/10.1145/3219819.3219832)]



- 19986]
- 27 Xu LC, Wei XK, Cao JN, *et al.* Embedding of embedding (EOE): Joint embedding for coupled heterogeneous networks. Proceedings of the 10th ACM International Conference on Web Search and Data Mining. Cambridge: ACM, 2017. 741–749. [doi: [10.1145/3018661.3018723](https://doi.org/10.1145/3018661.3018723)]
- 28 Hu BB, Fang Y, Shi C. Adversarial learning on heterogeneous information networks. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 120–129. [doi: [10.1145/3292500.3330970](https://doi.org/10.1145/3292500.3330970)]
- 29 Dong YX, Chawla NV, Swami A. Metapath2Vec: Scalable representation learning for heterogeneous networks. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax: ACM, 2017. 135–144. [doi: [10.1145/3097983.3098036](https://doi.org/10.1145/3097983.3098036)]
- 30 He Y, Song YQ, Li JX, *et al.* HeteSpaceyWalk: A heterogeneous spacey random walk for heterogeneous information network embedding. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019. 639–648. [doi: [10.1145/3357384.3358061](https://doi.org/10.1145/3357384.3358061)]
- 31 Hussein R, Yang DQ, Cudré-Mauroux P. Are meta-paths necessary? Revisiting heterogeneous graph embeddings. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino: ACM, 2018. 437–446. [doi: [10.1145/3269206.3271777](https://doi.org/10.1145/3269206.3271777)]
- 32 Lee S, Park C, Yu H. BHIN2vec: Balancing the type of relation in heterogeneous information network. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019. 619–628. [doi: [10.1145/3357384.3357893](https://doi.org/10.1145/3357384.3357893)]
- 33 Wang X, Zhang YD, Shi C. Hyperbolic heterogeneous information network embedding. Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019. 5337–5344. [doi: [10.1609/aaai.v33i01.33015337](https://doi.org/10.1609/aaai.v33i01.33015337)]
- 34 Zhang DK, Yin J, Zhu XQ, *et al.* MetaGraph2Vec: Complex semantic path augmented heterogeneous network embedding. Proceedings of the 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining. Melbourne: Springer, 2018. 196–208. [doi: [10.1007/978-3-319-93037-4\\_16](https://doi.org/10.1007/978-3-319-93037-4_16)]
- 35 Tu K, Cui P, Wang X, *et al.* Structural deep embedding for hyper-networks. Proceedings of the 32nd Conference on Artificial Intelligence. New Orleans: AAAI, 2018. 426–433.
- 36 Gui H, Liu JL, Tao FB, *et al.* Large-scale embedding learning in heterogeneous event data. Proceedings of the IEEE 16th International Conference on Data Mining (ICDM). Barcelona: IEEE, 2016. 907–912. [doi: [10.1109/ICDM.2016.0111](https://doi.org/10.1109/ICDM.2016.0111)]
- 37 Wang BY, Shang LF, Lioma C, *et al.* On position embeddings in BERT. Proceedings of the 9th International Conference on Learning Representations. Austria: ICLR, 2021.
- 38 Srinivasan B, Ribeiro B. On the equivalence between positional node embeddings and structural graph representations. Proceedings of the 8th International Conference on Learning Representations. Addis Ababa: ICLR, 2020.
- 39 Li P, Wang YB, Wang HW, *et al.* Distance encoding: Design provably more powerful neural networks for graph representation learning. Proceedings of the 34th Neural Information Processing Systems. NIPS, 2020. 4465–4478.
- 40 Dwivedi VP, Joshi CK, Luu AT, *et al.* Benchmarking graph neural networks. arXiv: 2003.00982, 2020.
- 41 Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation, 2003, 15(6): 1373–1396. [doi: [10.1162/089976603321780317](https://doi.org/10.1162/089976603321780317)]
- 42 Fu XY, Zhang JN, Meng ZQ, *et al.* MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding. Proceedings of the Web Conference. Taipei: ACM, 2020. 2331–2341. [doi: [10.1145/3366423.3380297](https://doi.org/10.1145/3366423.3380297)]
- 43 Gao J, Liang F, Fan W, *et al.* Graph-based consensus maximization among multiple supervised and unsupervised models. Proceedings of the 22nd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2009. 585–593.
- 44 Murphy RL, Srinivasan B, Rao VA, *et al.* Relational pooling for graph representations. Proceedings of the 36th International Conference on Machine Learning. Long Beach: ICML, 2019. 4663–4673.

(校对责编:牛欣悦)