



多重注意力指导下的异构图嵌入方法

孟祥福, 温晶, 李子函, 纪鸿樟

引用本文:

孟祥福, 温晶, 李子函, 纪鸿樟. 多重注意力指导下的异构图嵌入方法[J]. 智能系统学报, 2023, 18(4): 688–698.

MENG Xiangfu, WEN Jing, LI Zihan, JI Hongzhang. Heterogeneous graph embedding method guided by the multi-attention mechanism[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(4): 688–698.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202204006>

您可能感兴趣的其他文章

基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion

智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

层次化双注意力神经网络模型的情感分析研究

Hierarchical double-attention neural networks for sentiment classification

智能系统学报. 2020, 15(3): 460–467 <https://dx.doi.org/10.11992/tis.201812017>

基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism

智能系统学报. 2020, 15(5): 956–963 <https://dx.doi.org/10.11992/tis.201903001>

旅游知识图谱特征学习的景点推荐

Tourism knowledge-graph feature learning for attraction recommendations

智能系统学报. 2019, 14(3): 430–437 <https://dx.doi.org/10.11992/tis.201810032>

引入外部词向量的文本信息网络表示学习

Representation learning using network embedding based on external word vectors

智能系统学报. 2019, 14(5): 1056–1063 <https://dx.doi.org/10.11992/tis.201809037>

数据偏序结构关系中的知识发现可视化方法

A novel knowledge discovery visualization method based on data partial ordered structure

智能系统学报. 2016, 11(4): 475–480 <https://dx.doi.org/10.11992/tis.201606019>

DOI: 10.11992/tis.202204006

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.TP.20230322.0825.002.html>

多重注意力指导下的异构图嵌入方法

孟祥福, 温晶, 李子涵, 纪鸿樟

(辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

摘要: 现有的异构图嵌入学习方法存在两个方面的问题, 一是没有考虑不同节点属性间的深层联系, 二是通过注意力机制聚合邻居节点来生成目标节点的向量表示, 忽略了目标节点的特征在向量表示中起的作用。为解决上述问题, 本文提出了一种多重注意力指导下的异构图神经网络, 从点-线-网 3 个角度学习异构节点嵌入向量。使用双向长短期记忆模型 (bidirectional long short-term memory networks, Bi-LSTM) 挖掘不同节点的属性间的深层关联关系并将其映射到同一向量空间, 利用级联网络对单条元路径实例上的邻居节点和目标节点的特征信息进行融合, 从而增强嵌入向量对目标节点信息的表达能力, 提出一种多重注意力机制来聚合多条元路径实例上的节点信息, 生成最终的节点嵌入向量表示。在 3 个大型异构图上的实验结果表明, 本文提出的模型在异构图嵌入的效果方面优于现有基线模型, 并且对于增强节点属性信息上的表达展现出了良好的性能。

关键词: 异构信息网络; 图表示学习; 异构图嵌入; 元路径; 元路径实例; 图注意力; 异构图; 图神经网络

中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2023)04-0688-11

中文引用格式: 孟祥福, 温晶, 李子涵, 等. 多重注意力指导下的异构图嵌入方法 [J]. 智能系统学报, 2023, 18(4): 688-698.

英文引用格式: MENG Xiangfu, WEN Jing, LI Zihan, et al. Heterogeneous graph embedding method guided by the multi-attention mechanism[J]. CAAI transactions on intelligent systems, 2023, 18(4): 688-698.

Heterogeneous graph embedding method guided by the multi-attention mechanism

MENG Xiangfu, WEN Jing, LI Zihan, JI Hongzhang

(School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China)

Abstract: There are two problems in the existing heterogeneous graph embedding learning methods. One is that the deep relationship between different node attributes is not considered, the other is the problem of ignorance of the role of the features of the target node in the vector representation when generating the vector representation of the target node by aggregating neighboring nodes through attention mechanism. In order to solve above problems, this paper proposes a heterogeneous graph neural network under the guidance of multiple attentions, which learns the embedding vectors of heterogeneous nodes from three perspectives of Point-Line-Net. Bi-LSTM is used to mine the deep relationship between attributes of different nodes and map them to the same vector space. A cascaded network is used to fuse the feature information of neighbor nodes and target nodes on a single meta-path instance, so as to enhance the expression ability of embedded vectors to target node information. A multi-attention mechanism is proposed to aggregate node information on multiple meta-path instances and generate the final node embedding vector representation. Experimental results on three large heterogeneous graphs show that the proposed model is superior to the existing baseline model in the embedding effect of heterogeneous graphs, and shows good performance in enhancing the expression of node attribute information.

Keywords: heterogeneous information network; graph representation learning; heterogeneous graph embedding; metapath; metapath instance; graph attention; heterogeneous graph; graph neural network

收稿日期: 2022-04-04. 网络出版日期: 2023-03-22.

基金项目: 国家自然科学基金项目 (61772249); 辽宁省教育厅项目 (LJKZ0355).

通信作者: 孟祥福. E-mail: marxi@126.com.

现实世界中, 很多数据都是以图结构进行表示的, 如社交网络^[1-3]、蛋白质结构、推荐系统^[4-6]等。这些包含多种不同类型边和节点的图称为异

构图(heterogeneous graph, HG),也叫异构信息网络(heterogeneous information network, HIN)^[7]。如, DBLP(digital bibliography & library project)学术网络可以用HG表示,它由4种类型的节点(作者A、论文P、术语T、地点V)和3种类型的边(作者撰写论文、论文包含术语、论文发表在会议上)组成。基于这些节点和边,可以推导出更复杂的语义。A-P-A代表着同一个论文的合著者, P-A-P代表一个作者发表多篇论文。为了将这些图数据应用到下游任务,如节点分类^[8]、链路预测^[9]、节点聚类^[10]和推荐^[11]等,需要对异构图进行映射和嵌入处理,使得原始网络中的节点和边的结构和语义信息得到有效保持,并且压缩冗余信息。图嵌入作为一种高效的大规模网络表示和管理方法,将网络的拓扑结构映射到低维向量空间中,使得原始网络节点的邻近信息得到很好地保留和利用。

早期的嵌入技术着重研究同构图,如DeepWalk^[12],将图上的一组随机游动序列输入SkipGram^[13]模型来近似这些游动中的节点共现概率,并获得节点嵌入表示。与它类似的还有Node2vec^[14]。随着深度学习的迅速发展,图神经网络(graph neural networks, GNNs)被提出,它利用专门设计的神经层学习图表示。尽管GNN在许多任务中取得了先进的成果,但大多数基于GNN的模型都假设输入的是同构图,如Esim^[15]。为了解决异构图嵌入问题,研究者提出了元路径的概念,通过元路径将异构图转化为同构图,从而进行节点表示,如Metapath2vec^[16]、HAN(heterogeneous graph attention network)^[17]、MAGNN(metapath aggregated graph neural network for heterogeneous graph embedding)^[18]等。也有研究不采用元路径,直接在异构图上进行操作,如HetSANN^[19]、HetGNN^[20]、GATNE^[21]等。

然而,现有的异构图嵌入方法大多是通过简单的线性变换将不同类型节点的不同属性信息投影到同一向量空间,忽视了不同节点的属性信息间的关联。如DBLP数据集有“作者”节点、“会议”节点、“作者”具有“研究方向”、“发表论文”等属性,研究方向为“数据库”的作者发表的文章都是与数据库相关的,并且大多数代表性的论文都是发表在SIGMOD、VLDB、ICDE等会议上。“研究方向”决定着论文发表到什么样的会议或者期刊上。由此可见,同一节点的不同属性之间具有语义关联,不同节点的属性信息间也存在语义关联。同时,异构图嵌入方法大多是基于元路径的

思想,通过注意力机制聚合单条路径上的邻居节点,从而生成目标节点的嵌入,该类方法过于强调邻居节点的重要性而忽视了目标节点本身的信息。

为了解决上述问题,本文提出了一种多重注意力指导下的异构图神经网络嵌入方法(multiple attention for heterogeneous graph embedding method, MAN)。MAN从Point、Line、Net 3个角度进行节点嵌入,生成最终的嵌入向量。首先,把节点的属性信息输送进双向长短期记忆模型(bidirectional long short-term memory networks, Bi-LSTM),通过正向和反向建模捕获节点属性间的双向联系。然后,利用图注意力机制聚合邻居节点。为了解决图注意力只能聚合元路径实例上的首尾节点的问题,通过引入一个邻居节点聚合器学习每个邻居节点在元路径实例中的上下文向量表示;为了解决目标节点最终向量中包含较少节点属性信息的问题,设计了一种级联网络,该网络通过将图注意力生成的向量和目标节点属性向量进行线性操作,生成Line层次的向量表示。最后,通过多重注意力机制混合多条路径进行Net层次聚合。注意力机制按照元路径类型对图中所有路径进行分类和平均转换,聚合平均向量生成最终嵌入向量。然而对同一种元路径下的多条路径未区分重要性,因此,为解决这一问题,模型在将注意力作用到根据不同类型元路径形成的平均向量的同时,引入一个可学习的注意力参数,形成多重注意力机制,减小聚合平均向量为最终嵌入向量带来的偏差。

本文工作的主要创新点如下:

- 1)提出了基于Bi-LSTM的节点属性间关联关系的建模方法。
- 2)设计了级联网络和多重注意力机制来聚合元路径内部以及元路径之间的邻居节点。
- 3)在3个大型异构图上进行了节点分类和链路预测任务,评估所提模型的性能。

1 相关工作

1.1 图神经网络

图神经网络^[22-24]的目标是学习图中每个节点的低维向量表示,从而将这些节点应用于下游任务。图卷积网络(graph convolutional network, GCN)^[25]是一种在图中结合拓扑结构和顶点属性信息学习节点的嵌入表示的方法,然而GCN要求在一个确定的图中学习嵌入表示,无法泛化到在训练过程中没有出现过的节点。为解决这一问

题, Hamilton^[26]提出了一种通过学习一个对邻居顶点进行聚合的函数来产生目标节点的嵌入向量的方法。图注意力(graph attention networks, GAT)^[27]将注意力机制应用到图表示学习上,将原本GCN的标准化函数替换为使用注意力权重的邻居节点特征聚合函数,从而学习节点的嵌入向量。不同于传统的多头注意力机制均衡地消耗所有的注意力头, GaAN^[28]使用了一个卷积子网络来控制每个注意力头的重要性。

1.2 异构图嵌入

异构图嵌入,旨在通过保留网络拓扑架构和节点内容信息,将节点表示到低维向量空间中。现有的方法大多是基于元路径,如 Metapath2vec^[16]利用元路径引导生成随机游动,将其馈送到 Skip-Gram^[13]模型学习到节点的嵌入表示; HERec^[29]借助元路径的邻居将异构图转化为同构图,将节点送入 DeepWalk^[12]模型学习到嵌入表示。利用基于元路径的邻居将异构图转化为同构图,降低了图嵌入的复杂性。对于转化后的同构图, HAN^[17]使用图注意力架构聚合来自邻居节点的信息,并利用注意力机制来组合各种元路径; MAGNN^[18]改进了 HAN 和 HERec 只考虑元路径起始节点和末尾节点的问题,利用 RotatE^[30]编码元路径实例,最终得到包含丰富节点信息的嵌入表示。ie-HGCN^[31]打破了需要预先定义元路径的局限性,设计了一种基于邻接矩阵的异构图卷积网络,能够在粗粒度和细粒度2个层面上发现对于当前任务最优的元路径。HetSANN^[19]通过一个类型感知的注意力层替换传统 GNN^[32-34]中的卷积层,在不使用元路径的前提下,直接编码异构图中的结构信息。HetGNN^[20]采用重启随机游走为每个节点采样固定数目的强关联异质邻居,对于同类邻居和不同类邻居分别采用 Bi-LSTM^[35]和注意力机制进行聚合,生成嵌入向量。GATNE^[21]提出了一种基于属性多元异构网络的表示学习方法,通过节点之间存在的多种邻近类型,产生具有多个视图的网络,进行多重网络嵌入学习。

2 问题定义

本节将介绍本文使用的与异构图相关的一些重要术语的形式化定义。

定义1 异构图。异构图由 $G=(\mathcal{V}, \mathcal{E})$ 表示, 它由节点集合 \mathcal{V} 和边集合 \mathcal{E} 组成; $\phi: \mathcal{V} \rightarrow \mathcal{A}$ 为节点类型映射函数, $\varphi: \mathcal{E} \rightarrow \mathcal{R}$ 为边类型映射函数; \mathcal{A} 和

\mathcal{R} 表示预定义对象类型和链接类型的集合, $|\mathcal{A}|+|\mathcal{R}|>2$, 如图1所示。

定义2 元路径。元路径 p_i 由 $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} A_3 \xrightarrow{R_3} \dots \xrightarrow{R_i} A_{i+1}$ 来表示, 简称 $p_i = \{A_1, A_2, \dots, A_{i+1}\}$ 。其中, $R = R_1 \circ R_2 \circ \dots \circ R_i$ 是节点类型 $A_1 \sim A_{i+1}$ 的组合关系, \circ 表示关系上的组合操作。

定义3 元路径上的 k 跳邻居。对于一条元路径 $p_i = \{A_1, A_2, \dots, A_{i+1}\}$, 目标节点 A_1 的元路径 p_i 上的 k 跳邻居为 A_1 的 k 阶邻居 A_{k+1} 。

定义4 异构图嵌入。给定一个异构图 $G=(\mathcal{V}, \mathcal{E})$, 异构图嵌入是通过设计的模型学习节点的 d 维向量表示, 该向量尽可能包含节点在图 G 中的语义信息和结构信息。

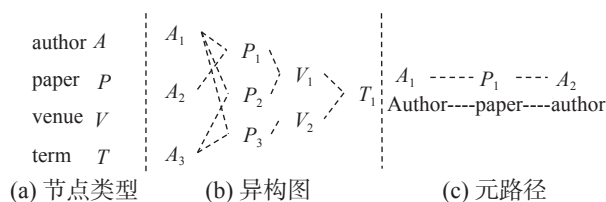


图1 异构图

Fig. 1 Heterogeneous graph

3 多重注意力指导下的异构图神经网络模型

本节描述了一种新的多重注意力指导下的异构图神经网络——MAN, 用于生成异构图嵌入。图2给出了单个节点的嵌入向量生成过程。如图2所示, Bi-LSTM 捕捉节点属性之间的关系, 并将节点投射到相同的向量空间中, 级联网络融合元路径上节点的上下文信息, 多重注意力模型生成节点嵌入表示。

3.1 Point 层次-异质节点属性抽取

通过预训练得到节点不同属性的嵌入表示, 这些表示向量的数据形状由属性内容决定。节点的异质性决定了属性的异质性, 不同类型的节点具有不同的属性, 属性的维度也是不相同的, 无序无规格的属性数据将对后续的操作造成阻碍。以往的解决方法大多是将不同的属性特征进行拼接, 或利用转换矩阵将其映射到同一向量空间。单一的线性转换无法反映出节点属性间存在的联系。例如, 学术论文数据集 (Database systems and logic programming, DBLP) 中有“作者”节点, “作者”具有“研究领域”、“发表论文类型”等属性, 对于研究领域为“推荐系统”的作者 A, 他发表论文的关键字通常是“推荐”、“协同过滤”等。作者的研究领域与发表论文类型也是密切相关的。因

此,设计了一种基于 Bi-LSTM 的模型来捕获节点深层次的特征交互信息,从而增强模型的表达能力。节点 v_i 的内容嵌入表示为

$$h_{vi} = \frac{\sum_{i \in \mathbf{x}_v} [\text{lstmpr}\{f_c \theta x(\mathbf{x}_i)\} \oplus \text{lstmnr}\{f_c \theta x(\mathbf{x}_i)\}]}{|\mathbf{x}_v|} \quad (1)$$

式中: $h_{vi} \in \mathbf{R}^d$ 为节点 v_i 经过 Bi-LSTM 模型得到的内容嵌入; \mathbf{R}^d 为 d 维欧氏空间; $f_c \theta x$ 为节点特征转换器,可以是一个恒等式(即 f_c 的输入与输出相同),也可以是一个参数为 θx 的全连接神经网络,用于调整嵌入维度的大小; \mathbf{x}_v 为节点的内容信息; \oplus 为向量间的串联操作; lstmpr、lstmnr 分别为

正向反向的 LSTM 网络。LSTM 的计算过程为

$$\begin{aligned} z_i &= \sigma(U_z f_c \theta x(\mathbf{x}_i) + W_z h_{i-1} + b_z) \\ f_i &= \sigma(U_f f_c \theta x(\mathbf{x}_i) + W_f h_{i-1} + b_f) \\ o_i &= \sigma(U_o f_c \theta x(\mathbf{x}_i) + W_o h_{i-1} + b_o) \\ c'_i &= \tanh(U_c f_c \theta x(\mathbf{x}_i) + W_c h_{i-1} + b_c) \\ c_i &= f_i \odot c_{i-1} + z_i \odot c'_i \\ h_i &= \tanh(c_i) \odot o_i \end{aligned} \quad (2)$$

Point 层模型首先使用不同的 FC 层来转换节点的内容信息,然后将转换好的内容信息输入 Bi-LSTM 进行编码,从而挖掘节点特征的深层交互,最后利用所有隐藏状态上的池化层输出一个基于内容信息的节点 v_i 的嵌入表示。

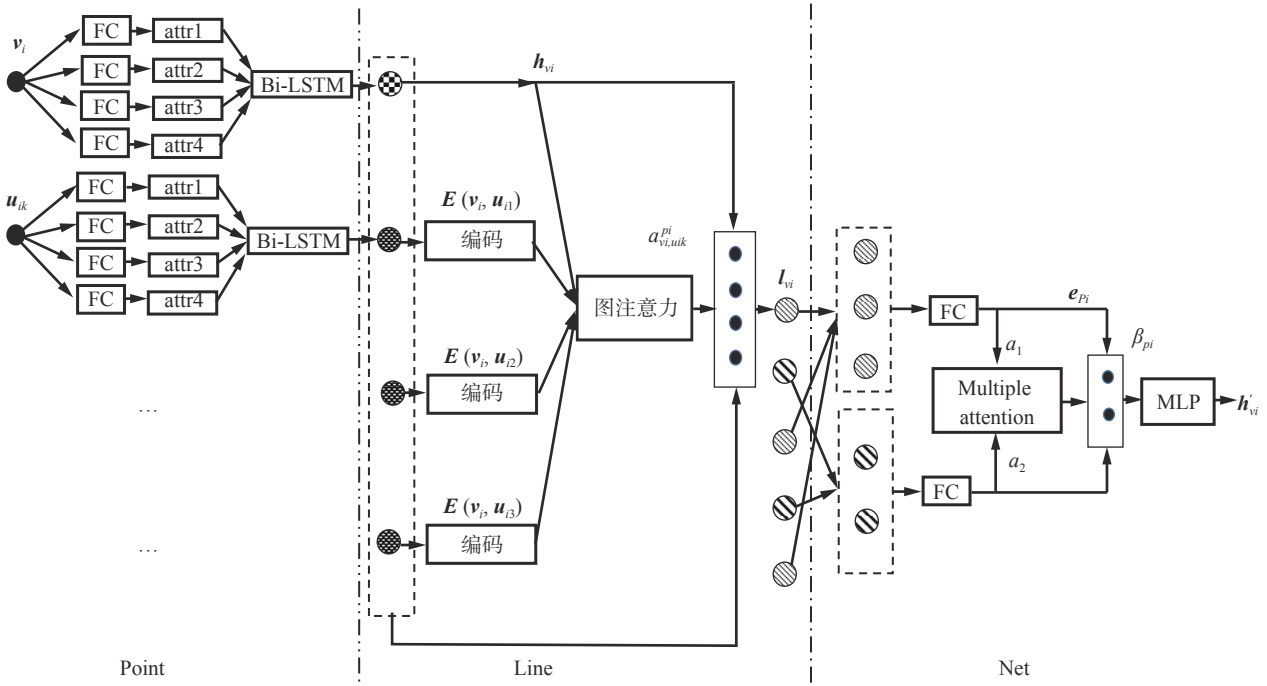


图2 MAN 工作原理

Fig. 2 Working principle of MAN

3.2 Line 层次-节点上下文信息融合

Point 层学习节点属性信息, Line 层学习节点在一条元路径实例上基于上下文的结构和语义信息。给定目标节点 v_i , $p_i = \{v_i, u_{i_1}, u_{i_2}, \dots, u_{i_k}\}$ 为 v_i 所在的一条元路径实例, 设 u_{i_k} 为目标节点在元路径 p_i 上的 k 跳邻居。Line 层使用邻居节点聚合器将 v_i 的邻居节点编码成一个向量。基于复杂空间的关系旋转的邻居节点聚合器定义为

$$\begin{aligned} E_0 &= h_{vi} \\ E_k(v_i, u_{i_k}) &= h_{u_{i_k}} + E_{k-1}(v_i, u_{i_{k-1}}) \odot r_k \\ E(v_i, u_{i_k}) &= \frac{E_{l_{pi}}(v_i, u_{i_k})}{l_{pi} + 1} \end{aligned} \quad (3)$$

式中: $E(v_i, u_{i_k}) \in \mathbf{R}^d$ 为经过聚合器编码后的元路径实例 p_i 下的第 k 个邻居节点 u_{i_k} , $k \in (0, l_{pi}]$; l_{pi} 为路

径 p_i 上的最大节点数; u_{i_0} 代表节点 v_i ; r_k 为节点 $u_{i_{k-1}}$ 与 u_{i_k} 之间的关系; \odot 为元素间乘积操作。邻居节点聚合器以“累加”的思想将 k 跳邻居信息聚合到节点 u_{i_k} 上, 从而实现了邻居节点上下文信息的融合。图3描述了邻居节点聚合器的原理。

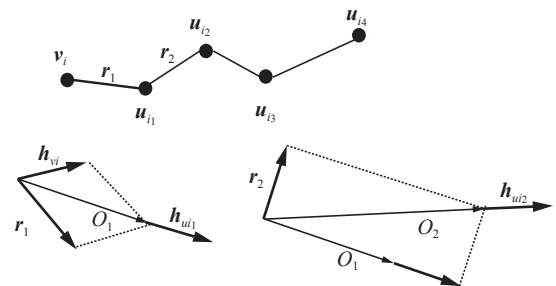


图3 邻居节点聚合器

Fig. 3 Neighbor node aggregator

将邻居节点聚合为向量表示后,采用一个基于图注意力的级联网络来编码元路径实例 p_i 上的目标节点 v_i 的向量。使用注意力机制抽取一条路径上不同类型的节点特征,得到邻居节点对目标节点的重要程度,直接邻居节点的权重通常要高于间接邻居节点。给定节点对 (v_i, u_{ik}) 和连接它们的元路径实例 u_{ik} 对于 v_i 的重要程度的计算方法为

$$e_{vi,u_{ik}}^{p_i} = \sigma(\theta_{p_i}^T [\mathbf{h}_{vi} \| \mathbf{E}(v_i, u_{ik})]) \quad (4)$$

利用注意力向量和激活函数从目标节点与邻居节点的串联向量中提取出邻居节点对于目标节点的重要程度。 $\theta_{p_i} \in \mathbf{R}^{2d'}$ 为元路径实例 p_i 的参数化的注意力向量; $\sigma(\cdot)$ 为激活函数; $e_{vi,u_{ik}}^{p_i}$ 为经过邻居节点聚合器编码后的节点 u_{ik} 对于目标节点 v_i 在 p_i 上的权重系数; \mathbf{h}_{vi} 为目标节点的嵌入向量。经过归一化,将权重映射到 $[0,1]$,计算方法为

$$a_{vi,u_{ik}}^{p_i} = \frac{\exp(e_{vi,u_{ik}}^{p_i})}{\sum_{j \in (0, J_{p_i})} \exp(e_{vi,u_{ij}}^{p_i})} \quad (5)$$

最后将归一化后的注意力系数 $a_{vi,u_{ik}}^{p_i}$ 作用到 $\mathbf{E}(v_i, u_{ik})$ 上,加入目标节点 v_i 的原始向量表示,形成基于注意力的级联网络,经过一个激活函数得到目标节点 v_i 的Line层嵌入,计算方法为

$$\mathbf{l}_{vi} = \sigma \left(\sum_{j \in (0, J_{p_i})} a_{vi,u_{ij}}^{p_i} \mathbf{E}(v_i, u_{ij}) + \mathbf{h}_{vi} \right) \quad (6)$$

Line层首先通过邻居节点聚合器重新定义邻居节点的向量表示,不同于以往只考虑邻居节点的内容信息,MAN考虑到邻居节点的内容信息以及其在元路径实例上的语义和位置信息。不同类型节点对学习目标节点表示的贡献是不同的,采用图注意力机制来捕获贡献度。最后,借鉴残差的思想,将包含 v_i 属性信息的原始向量与图注意力下的邻居节点向量进行线性操作,设计了一个级联网络。由此生成的向量表示既包含邻居节点的语义和结构信息,又包含 v_i 的语义信息,从而实现了对目标节点向量表示的校正操作。

3.3 Net 层次-节点拓扑结构编码

对于类型为A的目标节点 v_i ,以及一组元路径 $p_A = \{p_1, p_2, \dots, p_M\}$,Line层为 v_i 生成 M 个针对特定元路径实例的向量表示,记为 $\{\mathbf{l}_{vi}^1, \mathbf{l}_{vi}^2, \dots, \mathbf{l}_{vi}^M\}$ 其中,每个 \mathbf{l}_{vi}^p 都表示节点 v_i 在A类元路径下隐含的一种语义信息。根据元路径类型对Line层得到的嵌入向量进行划分,即要对同属于一种元路径类型的向量表示进行聚合。计算方法为

$$\mathbf{e}_{p_i} = \left(\frac{1}{s_A} \sum_{i \in I_A} \sigma(\mathbf{M}_A \cdot \mathbf{l}_{vi} + \mathbf{b}_A) \right) \cdot \mathbf{a}_i \quad (7)$$

式中: s_A 为类型为A的元路径的数量; $\sigma(\cdot)$ 为激活函数; \mathbf{e}_{p_i} 为基于元路径类型聚合后的嵌入向量; \mathbf{a}_i 为作用在根据元路径类型划分后的向量上的参数; t_A 为节点类型为A; \mathbf{M}_A 、 \mathbf{b}_A 为可学习的参数矩阵。通过设计一个多重注意力机制混合节点 v_i 的基于元路径类型的向量表示来生成Net层次节点 v_i 的向量表示,计算方法为

$$\begin{aligned} \beta_{p_i} &= \text{softmax}(\beta_A \cdot \mathbf{e}_{p_i}) \\ \mathbf{h}'_{vi} &= \sum_{i \in \text{type}} \beta_{p_i} \cdot \mathbf{l}_{p_i} \end{aligned} \quad (8)$$

式中: β_A 为参数化的注意力向量; β_{p_i} 为基于元路径类型的嵌入向量的注意力系数; \mathbf{h}'_{vi} 为节点 v_i 的最终嵌入表示。

Net层将编码范围扩大到整个网络,首先对Line层得到的每个基于元路径实例的节点表示按照元路径的类型划分并进行向量转换。这里将每种元路径解释为一条特定语义,以DBLP(digital bibliography & library project)数据集中为例,类型为A-P-A的元路径的特定语义为“合著者及论文”。同一类型元路径下不同实例对于目标节点 v_i 的影响也是不同的,因此模型对每一条元路径实例进行线性转换时加入一个可学习的“Multiple”系数,以平衡对于同种元路径实例取平均带来的误差。然后,得到了不同语义下目标节点 v_i 的表示向量。同样,不同语义对于目标节点的表示向量的贡献度也是不同的。因此,采用注意力机制为不同语义分配注意力权重。

4 实验与评估

4.1 实验设计

4.1.1 数据集

实验采用MAGNN^[18]中提供的3个异构图形数据集作为实验数据,分别是DBLP、IMDB(internet movie database)和Last-fm(last favorite music),并与最先进的基线模型进行比较。DBLP和IMDB数据集进行节点分类和聚类,Last-fm进行链路预测。对于没有属性的节点采用one-hot编码作为虚拟属性。

4.1.2 对比方法

通过与最新的基线算法进行对比以验证MAN的有效性。基线模型包括:

LINE^[36]:利用节点间的一阶和二阶接近度对图进行嵌入。通过忽略图结构的异构性和删除所有节点内容特征将其应用于异构图。

Node2vec^[14]:作为deepwalk的升级版应用于异构图。

Esim^[15]: 一种异构图嵌入模型, 可以从采样的元路径实例中学习节点表示。Esim 要求每个元路径具有预定义的权重, 因此, 实验为所有元路径分配相等的权重。

Metapath2vec^[16]: 通过将元路径引导得到的实例馈送到 SkipGram 模型中, 得到节点嵌入。该模型依赖于单个用户指定的元路径, 因此, 实验分别对所有元路径进行测试, 并取具有最佳结果的元路径。

GCN^[25]: 一个半监督图卷积网络, 用于齐次图。

GAT^[13]: 模型结合注意机制, 在图形空间域执行卷积运算。实验在基于元路径的同构图上测试 GAT, 并展示来自最佳元路径的结果。

GATNE^[21]: 从基嵌入和边嵌入生成节点表示。本文展示的是表现最好的 GATNE 变体的结果。

HAN^[17]: 一个异质 GNN。它从不同的基于元路径的同构图中学习元路径特定的节点嵌入, 并利用注意机制将它们组合为每个节点的一个向量表示。

MAGNN^[18]: 在 HAN 的基础上, 利用一个编码器将元路径实例中的邻居信息进行融合, 从而生成包含丰富语义信息的节点嵌入。

4.1.3 参数设置

对于 LINE、Node2vec、Esim、Metapath2vec 和 HERec, 窗口大小设置为 5, 漫游长度设置为 100, 每个节点漫游 40 次, 负采样设置为 5。对于 GNN 模型, 包括 GCN、GAT、HAN 和 MAGNN, dropout 设置为 0.5, 使用相同的训练集、验证集和测试集; 使用 Adam 优化器, 学习率设置为 0.005, 重量衰减设置为 0.001; 对 GNN 进行 100 个 epoch 的培训, 并以 30 min 提前停止。对于节点分类和节点聚类, GNN 以半监督方式进行训练, 将其中一小部分节点标记为引导。对于 GAT、HAN 和 MAGNN, 注意力头的数量设置为 8。对于 HAN 和 MAGNN, 元路径间聚合中的注意向量的维数设置为 128。为了公平比较, 上述所有模型的嵌入维度设置为 64。

4.2 模型应用

4.2.1 节点分类效果测试

为比较不同模型在节点分类任务上的性能, 对 IMDB 和 DBLP 数据集进行实验。将不同模型生成的节点嵌入表示输送进线性支持向量机 (support vector machine, SVM) 分类器中。为保证实验的可靠性, 只将测试集中的数据送入 SVM 中。以每个模型运行 10 次的平均 A_{F_1} 和 I_{F_1} 作为评估标准。

经过 SVM 分类后的节点, T_P 为真实标签为正预测标签为真的真正例, F_P 为真实标签为正预测标签为假的假正例, F_N 为真实标签为假预测结

果为正的假负例, T_N 为真实标签为假预测标签为假的真负例。准确率、精确率、召回率和 F_1 表示为

$$\begin{cases} A_{cc} = \frac{T_P + T_N}{t_{total}} \\ P = \frac{T_P}{T_P + F_P} \\ R = \frac{T_P}{T_P + F_N} \\ F_1 = \frac{2 \times P \times R}{P + R} \end{cases} \quad (9)$$

M_{F_1} 和 I_{F_1} 的计算为

$$\begin{cases} M_{F_1} = \frac{2 \times P \times R}{P + R} \\ I_{F_1} = \frac{1}{n} \sum_{i=1}^n F_i \end{cases} \quad (10)$$

从表 1 可以看出, MAN 模型在不同数据集上的分类效果始终优于其他基线模型。同时, 基于元路径的 Esim、GAT、HAN、MAGNN 和 MAN 的 M_{F_1} 和 I_{F_1} 均高于其他模型, 这也验证了元路径对于提升异构图嵌入质量的有效性的观点。MAN 在最佳基线 MAGNN 上获得的性能增益为 2%~4%。

4.2.2 节点聚类效果测试

为比较不同模型在节点聚类上的表现效果, 选用 DBLP 和 IMDB 2 个数据集进行试验, 并将不同模型得到的节点表示结果嵌入到 K-means 算法中, K-means 中的聚类数设置为每个数据集的类数, 即 DBLP 的聚类数为 4, IMDB 的聚类数为 3。归一化互信息 (N_{MI})^[37] 和调整后的兰德指数 (A_{RI})^[38] 作为评估指标。

$$\begin{aligned} p_i &= \frac{|U_i|}{N} \\ H(U) &= - \sum_{i=1}^R p_i \log p_i \\ M_I(U, V) &= - \sum_{i=1}^R \sum_{j=1}^C p_{i,j} \log \left(\frac{p_{i,j}}{p_i \times p_j} \right) \\ N_{MI}(U, V) &= 2 \frac{M_I(U, V)}{H(U) + H(V)} \end{aligned} \quad (11)$$

$$A_{RI} = \frac{R_I - E(R_I)}{\max(R_I) - E(R_I)} \quad (12)$$

取模型运行 10 次的 N_{MI} 和 A_{RI} 平均值为最终结果, 并在表 2 中给出最终结果。

传统模型中 node2vec 表现较好, 这是因为 node2vec 作为 deepwalk 的变体, 基于随机游走采集邻居节点, 使得图中临近的点在嵌入到低维向量空间中时也保持了邻近性。基于图神经网络的模型中, MAGNN 的效果高于其他模型, 这是由于 MAGNN 引入了一个邻居节点编码器来聚合元

路径实例的上下文信息,目标节点融合了部分邻居节点的信息,因此在聚类任务中目标节点与其邻居节点更容易聚合到同一类中。MAN采用级

联网络突出目标节点的原始向量表示,在生成向量表示时不再过度依赖邻居节点,因此在聚类任务中性能提升程度较小。

表1 节点分类
Table 1 Node classification

数据集	指标	训练/%	LINE	Node2vec	Esim	Metapath2vec	GCN	GAT	HAN	MAGNN	MAN
DBLP	A_{F_1}	20	87.16	86.70	90.68	88.47	88.00	91.05	91.69	93.13	94.11
		40	88.85	88.07	91.61	89.91	89.00	91.24	91.96	93.23	94.61
		60	88.93	88.69	91.84	90.50	89.43	91.42	92.14	93.57	95.19
		80	89.51	88.93	92.27	90.86	89.98	91.73	92.50	94.10	96.03
	I_{F_1}	20	87.68	87.21	91.21	89.02	88.51	91.61	92.33	93.61	94.55
		40	89.25	88.51	92.05	90.36	89.22	91.77	92.57	93.68	94.94
		60	89.34	89.09	92.28	90.94	89.57	91.97	92.72	93.99	95.55
		80	89.96	89.37	92.68	91.31	90.33	92.24	93.23	94.47	96.32
IMDB	A_{F_1}	20	44.04	49.00	48.37	46.05	52.73	53.64	56.19	59.35	61.11
		40	45.45	50.63	50.09	47.57	53.67	55.50	56.15	60.27	61.38
		60	47.09	51.65	51.45	48.17	54.24	56.46	57.29	60.66	62.70
		80	47.49	51.49	51.37	49.99	54.77	57.43	58.51	61.44	64.67
	I_{F_1}	20	45.21	49.94	49.32	47.22	52.80	53.64	56.32	59.60	61.05
		40	46.92	51.77	51.21	48.17	53.76	55.56	57.32	60.50	61.28
		60	48.35	52.79	52.53	49.87	54.23	56.47	58.42	60.88	62.57
		80	48.98	52.72	52.54	50.50	54.63	57.40	59.24	61.53	64.51

表2 节点聚类
Table 2 Node cluster

数据集	指标	LINE	Node2vec	Esim	Metapath2vec	GCN	GAT	HAN	MAGNN	MAN
DBLP	N_{MI}	71.02	77.01	68.33	74.18	73.45	70.73	77.49	80.81	81.14
	A_{MI}	76.52	81.37	72.22	78.11	77.50	76.04	82.95	85.54	86.13
IMDB	N_{MI}	1.13	5.22	1.07	0.89	7.46	7.84	10.79	15.58	15.93
	A_{MI}	1.20	6.02	1.01	0.22	7.69	8.87	11.11	16.74	16.89

4.2.3 链路预测效果测试

Last-fm数据集用于评估链路预测任务中MAN和其他基线的性能。在GNNs中,将连接的用户艺术家对视为正节点对,并将所有未连接的用户艺术家链接视为负节点对。将相同数量的随机抽样负节点对添加到验证集和测试集。采用链路预测常用标准 A_{UC} 和 A_P 衡量模型性能。定义为

$$A_{UC} = \frac{n' + 0.5n''}{n} \quad (13)$$

$$A_P = \frac{m}{L}$$

A_{UC} 为在测试集中随机选择一条边的分数值比随机选择一条不存在的边的分数值大的概率。独立比较 n 次,大于则+1(设有 n' 次),等于则+0.5(设有

n'' 次),小于则不加。 A_P 侧重于衡量前几条边预测是否准确。假设 m 个预测准确,排在前 L 的边中有 m 个在测试集中,精确度定义为 m/L ,对于给定的 L ,精确度越大预测越准确。

从表3中可以看出,MAN的性能大大优于其他基准模型。最强的传统模型是metapath2vec,它从由元路径引导的随机游动生成的节点序列中学习节点表示。MAGNN比metapath2vec获得更好的评价,这表明考虑单个metapath是次优的。在GNN基线中,MAGNN获得了最好的结果,因为它具有异构性意识并结合了多条元路径。与MAGNN相比,MAN效果提升0.7%。因此MAN的效果优于其他模型。

表3 链路预测
Table 3 Link predict

数据集	指标	Line	Node2vec	Esim	Metapath2vec	GCN	GAT	HAN	MAGNN	MAN
Last-fm	A_{UC}	85.76	67.14	82.00	92.20	90.97	92.36	93.40	98.91	99.63
	A_P	88.07	64.11	82.19	90.11	91.65	91.55	92.44	98.93	99.65

4.3 消融与超参数实验

4.3.1 消融实验

为了验证 MAN 模型中的每个组成部分的有效性,对不同的 MAN 变体进行了实验,并将在不同数据集上进行的节点分类、节点聚类 and 链路预测,结果在表4中给出。为直观反映模型的效果,对节点分类任务中得到的不同训练比例的 A_{F1} 和 I_{F1} 取均值进行展示。MANpoint 是在不考虑节点

属性间存在的联系的情况下提出的模型,使用一个简单的线性变换将不同节点的不同属性映射到同一向量空间中。MANline 是不考虑级联网络的情况下进行的实验,用于验证目标节点的原始向量表示对于最终向量表示的有效性。MANnet 是只考虑 Point 和 Line 层模型而不考虑 Net 层中“Multiple”系数对于注意力权重的影响进行的实验,用于验证多重注意力模型的效果。

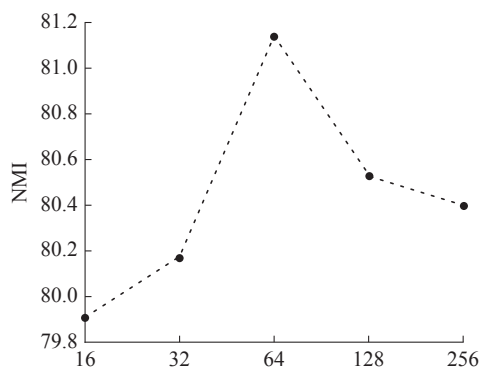
表4 消融实验
Table 4 Ablation experiments

变量	DBLP				IMDB				Last-fm	
	A_{F1}	I_{F1}	N_{MI}	A_{RI}	A_{F1}	I_{F1}	N_{MI}	A_{MI}	A_{UC}	A_P
MANpoint	93.39	93.8628	79.06	84.20	60.66	60.55	15.08	13.16	99.17	99.20
MANline	94.52	94.86	79.60	84.45	61.18	61.06	15.12	14.54	98.94	98.98
MANnet	94.74	95.16	80.41	85.41	62.11	62.03	15.05	14.08	99.02	99.03
MAN	94.98	95.34	81.14	86.13	62.462	62.35	15.93	16.89	99.63	99.65

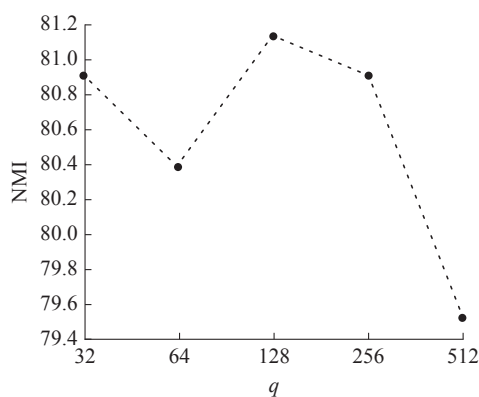
可以看出,MANnet 的效果与 MAN 差距较小,也就是说,Bi-LSTM 模型和级联网络的加入,对于异构图嵌入的影响显著。而多重注意力模型对最终节点嵌入起一个校正作用,这恰恰验证了本文的观点。上述模型在 DBLP 数据集上的效果高于 IMDB 数据集,这是因为 IMDB 数据集“脏”。

4.3.2 超参数实验

超参数在 MAN 中起着重要作用,因为它们决定了如何生成节点嵌入。在 DBLP 数据集上研究了不同参数对于节点分类结果的影响,并将节点聚类的 NMI 结果在图4中给出。



(a) 嵌入向量维度 z 与 NMI 关系



(b) 注意力向量维度 q 与 NMI 关系

图4 超参数改变情况下的效果对比

Fig. 4 Comparison of different hyperparameters

嵌入向量的维度 z 。首先对最终生成的嵌入向量的维度进行测试。随着维度的增加,MAN 的性能逐渐变好,在 $z=64$ 处达到性能最大值,维度继续增加,MAN 性能下降。也就是说过短的编码无法包含节点全部的语义信息,过长的编码将会导致额外的冗余。

注意力向量的维度 q 。对注意力向量的维度进行测试。从图4中可以看到维度的变化是不稳定的,在 $q=128$ 处达到性能最大。维度增大后性能下降,这是由过拟合引起的。

4.3.3 可视化

为了更加直观的理解不同模型的嵌入效果,进行可视化任务。将模型学习到的嵌入向量通过 t-SNE^[39] 投影到二维向量空间中,从而观察节点分布。具体来说,从 DBLP 中取 1000 个作者节点,并根据作者所属领域对其进行着色。

从图 5 中可以发现, GAT 得到的节点集群过于分散,不同领域节点分布散乱。这是由于 GAT 在不考虑元路径的情况下分配注意力权重,忽略了节点上下文间的语义信息。HAN 相比 GAT 有了明显的改善,节点集群的边缘开始明显化,但是仍有部分节点分散在其他领域中。有了元路径的指导, HAN 学习到的嵌入包含了语义信息,具有了一定的类内相似性。MAN 的节点集群边缘明显,同领域节点能够聚集到一起。MAN 在通过聚合邻居节点以捕获上下文信息的同时关注到目标节点自身语义,得到的节点嵌入更加全面准确。

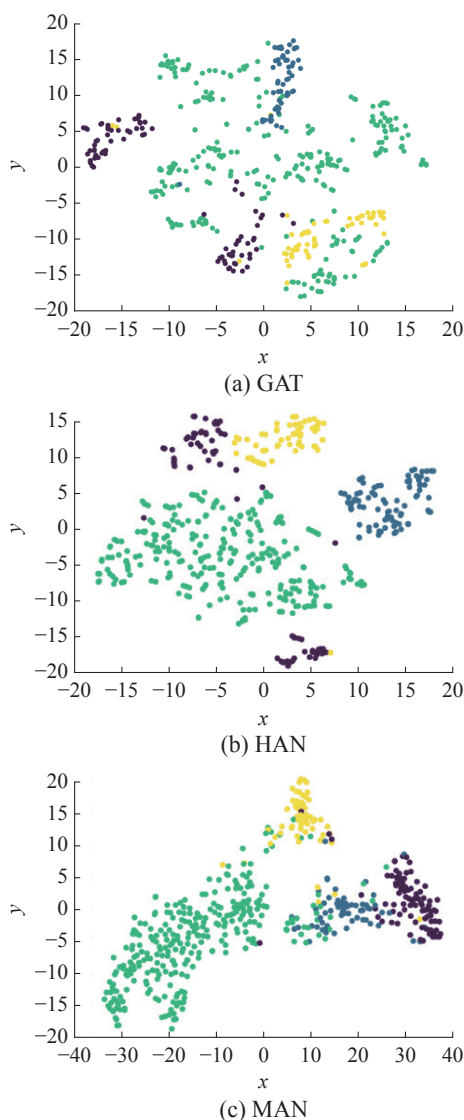


图 5 嵌入效果可视化

Fig. 5 Visualization of embedding

5 结束语

本文针对异构图嵌入中存在的问题,提出了一种新的多重注意力指导下的异构图神经网络——MAN。该模型从 Point-Line-Net 层面对节点进行建模以生成富含丰富语义信息的节点表示: Point 层挖掘节点属性间的联系; Line 层聚合一条元路径实例上的邻居节点; Net 层聚合多条元路径实例生成节点嵌入表示。对于各种图挖掘任务,如节点分类、节点聚类、链路预测, MAN 的性能优于其他方法。消融实验证明了 MAN 3 个层次在提高模型性能方面的功效。未来,我们将探索 MAN 应用于知识图谱进行其他下游任务。

参考文献:

- [1] BILÒ D, FRIEDRICH T, LENZNER P, et al. Selfish creation of social networks[C]//The 35th AAAI Conference on Artificial Intelligence. Online: AAAI, 2021: 5185–5193.
- [2] WANG Daixin, CUI Peng, ZHU Wenwu. Structural deep network embedding[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1225–1234.
- [3] GAO Yuan, LI Yi, SUN Yunchuan, et al. IEEE access special section: privacy preservation for large-scale user data in social networks[J]. *IEEE access*, 2022, 10: 4374–4379.
- [4] SHEN Jiayi, WANG Haotao, GUI Shupeng, et al. UMEC: Unified model and embedding compression for efficient recommendation systems[C]//International Conference on Learning Representations. Addis Ababa: Ithaca, 2020: 1–13.
- [5] HAN Xiaotian, SHI Chuan, WANG Senzhang, et al. Aspect-level deep collaborative filtering via heterogeneous information networks[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. New York: ACM, 2018: 3393–3399.
- [6] HU Binbin, SHI Chuan, ZHAO W X, et al. Leveraging meta-path based context for top-N recommendation with a neural co-attention model[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1531–1540.
- [7] SHI Chuan, LI Yitong, ZHANG Jiawei, et al. A survey of heterogeneous information network analysis[J]. *IEEE transactions on knowledge and data engineering*, 2017,

- 29(1): 17–37.
- [8] ZHANG Yizhou, XIONG Yun, KONG Xiangnan, et al. Deep collective classification in heterogeneous information networks[C]//Proceedings of the 2018 World Wide Web Conference. New York: ACM, 2018: 399–408.
- [9] CHEN Mingren, HUANG Ping, LIN Yu, et al. SSNE: effective node representation for link prediction in sparse networks[J]. *IEEE access*, 2021, 9: 57874–57885.
- [10] TANG Fengqin, WANG Chunling, SU Jinxia, et al. Spectral clustering-based community detection using graph distance and node attributes[J]. *Computational statistics*, 2020, 35(1): 69–94.
- [11] REN Xiang, LIU Jialu, YU Xiao, et al. ClusCite: effective citation recommendation by information network-based clustering[C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2014: 821–830.
- [12] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2014: 701–710.
- [13] MIKOLOV T, CHEN KAI, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. (2013-01-16)[2022-04-04].<https://arxiv.org/abs/1301.3781>.
- [14] GROVER A, LESKOVEC J. Node2vec: scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855–864.
- [15] SHANG Jinbo, QU Meng, LIU Jialu, et al. Meta-path guided embedding for similarity search in large-scale heterogeneous information networks[EB/OL]. (2016–10–31)[2022–04–04]. <https://arxiv.org/abs/1610.09769>.
- [16] DONG Yuxiao, CHAWLA N V, SWAMI A. Metapath2vec: scalable representation learning for heterogeneous networks[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2017: 135–144.
- [17] WANG Xiao, JI Houye, SHI Chuan, et al. Heterogeneous graph attention network[C]//The World Wide Web Conference. New York: ACM, 2019: 2022–2032.
- [18] FU Xinyu, ZHANG Jiani, MENG Ziqiao, et al. MAGNN: metapath aggregated graph neural network for heterogeneous graph embedding[C]//Proceedings of The Web Conference 2020. New York: ACM, 2020: 2331–2341.
- [19] HONG Huiting, GUO Hantao, LIN Yucheng, et al. Attention-based graph neural network for heterogeneous structural learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 4132–4139.
- [20] ZHANG Chuxu, SONG Dongjin, HUANG Chao, et al. Heterogeneous graph neural network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 793–803.
- [21] CEN Yukuo, ZOU Xu, ZHANG Jianwei, et al. Representation learning for attributed multiplex heterogeneous network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 1358–1368.
- [22] LI Yujia, TARLOW D, BROCKSCHMIDT M, et al. Gated graph sequence neural networks[C]//The 4th International Conference on Learning Representations, Caribe Hilton, USA, 2016.
- [23] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. *IEEE transactions on neural networks*, 2009, 20(1): 61–80.
- [24] GORI M, MONFARDINI G, SCARSELLI F. A new model for learning in graph domains[C]//Proceedings of 2005 IEEE International Joint Conference on Neural Networks. Piscataway: IEEE, 2005: 729–734.
- [25] KIPF T N, WELLMING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2016–09–09)[2022–04–04]. <https://arxiv.org/abs/1609.02907>.
- [26] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 1025–1035.
- [27] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2017–10–31)[2022–04–04]. <https://arxiv.org/abs/1710.10903>.
- [28] ZHANG Jiani, SHI Xingjian, XIE Junyuan, et al. Gated attention networks for learning on large and spatiotemporal graphs[C]//The Conference on Uncertainty in Artificial Intelligence. California: PMLR, 2018: 339–349.
- [29] SHI Chuan, HU Binbin, ZHAO W X, et al. Heterogeneous information network embedding for recommendation[J]. *IEEE transactions on knowledge and data engineering*, 2019, 31(2): 357–370.
- [30] ZHOU Xiaohan, YI Yunhui, JIA Geng. Path-RotatE: knowledge graph embedding by relational rotation of path in complex space[C]// IEEE/CIC International Conference on Communications in China. Piscataway: IEEE,

- 2021: 905–910.
- [31] YANG Yaming, GUAN Ziyu, LI Jianxin, et al. Interpretable and efficient heterogeneous graph convolutional network[J]. *IEEE transactions on knowledge and data engineering*, 2023, 35(2): 1637–1650.
- [32] PARK H, LEE S, HWANG D, et al. Learning augmentation for GNNs with consistency regularization[J]. *IEEE access*, 2021, 9: 127961–127972.
- [33] YU Yue, CHEN Jie, GAO Tian, et al. DAG-GNN: DAG structure learning with graph neural networks[EB/OL]. (2019–04–22)[2022–04–04]. <https://arxiv.org/abs/1904.10098>.
- [34] CHEN Zhengdao, VILLAR S, CHEN Lei, et al. On the equivalence between graph isomorphism testing and function approximation with GNNs[EB/OL]. (2019–05–29)[2022–04–04]. <https://arxiv.org/abs/1905.12560>.
- [35] LIMA M, SILVA R, LOPES DE SOUZA MENDES F, et al. Inferring about fraudulent collusion risk on Brazilian public works contracts in official texts using a Bi-LSTM approach[C]//Findings of the Association for Computational Linguistics: EMNLP 2020. Stroudsburg: Association for Computational Linguistics, 2020: 1580–1588.
- [36] TANG Jian, QU Meng, WANG Mingzhe, et al. LINE: large-scale information network embedding[C]//Proceedings of the 24th International Conference on World Wide Web. New York: ACM, 2015: 1067–1077.
- [37] YU Xie, CHENG Chen, GONG Maoguo, et al. Graph embedding via multi-scale graph representations[J]. *Information sciences*, 2021, 578: 102–115.
- [38] LEE Mengyuan, YU Guanding, LI G Y. Graph embedding-based wireless link scheduling with few training samples[J]. *IEEE transactions on wireless communications*, 2021, 20(4): 2282–2294.
- [39] VANDER M L, HINTON G. Visualizing data using t-SNE[J]. *Journal of machine learning research*, 2008, 9(11): 2579–2605.

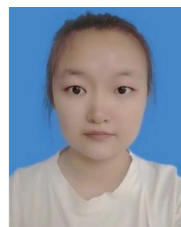
作者简介:



孟祥福, 教授, 博士, 主要研究方向为空间关键字查询、大数据分析、可视化、机器学习和推荐系统。主持国家自然科学基金项目 20 项, 获授权发明专利 1 项。出版专著 1 部, 发表学术论文 50 余篇。



温晶, 硕士研究生, 主要研究方向为图表示学习、图数据挖掘。



李子函, 硕士研究生, 主要研究方向为知识图谱、图表示学习。