

一种基于图注意力网络的异质信息网络表示学习框架

康世泽* 吉立新 张建朋

(战略支援部队信息工程大学 郑州 450001)

摘要: 常用的异质信息网络有知识图谱和具有简单模式层的异质信息网络, 它们的表示学习通常遵循不同的方法。该文总结了知识图谱和具有简单模式层的异质信息网络之间的异同, 提出了一个通用的异质信息网络表示学习框架。该文提出的框架可以分为3个部分: 基础向量模型, 基于图注意力网络的传播模型以及任务模型。基础向量模型用于学习基础的网络向量; 传播模型通过堆叠注意力层学习网络的高阶邻居特征; 可更换的任务模型适用于不同的应用场景。与基准模型相比, 该文所提框架在知识图谱的链接预测任务和异质信息网络的节点分类任务中都取得了相对不错的效果。

关键词: 异质信息网络; 知识图谱; 图注意力网络; 表示学习

中图分类号: TN919.2

文献标识码: A

文章编号: 1009-5896(2021)04-0915-08

DOI: 10.11999/JEIT200034

Heterogeneous Information Network Representation Learning Framework Based on Graph Attention Network

KANG Shize JI Lixin ZHANG Jianpeng

(Information Engineering University, Zhengzhou 450001, China)

Abstract: Commonly used heterogeneous information networks include knowledge graphs and heterogeneous information networks with simple schemas. Their representation learning follows usually different methods. The similarities and differences between knowledge graphs and heterogeneous information networks with simple schemas are summarized, and a general heterogeneous information network representation learning framework is proposed. The proposed framework can be divided into three parts: the basic vector model, the graph attention network based propagation model, and the task model. The basic vector model is used to learn basic network vector; The propagation model learns the high-order neighbor features of the network by stacking attention layers. The replaceable task module is suitable for different application scenarios. Compared with the benchmark model, the proposed framework achieves relatively good results in the link prediction task of the knowledge graph and the node classification task of the heterogeneous information network.

Key words: Heterogeneous information network; Knowledge graph; Graph attention network; Representation learning

1 引言

异质信息网络为具有多种类型(类型数量大于1)节点或边的复杂网络^[1]。学术领域广泛研究的异质信息网络有两种: (1)具有简单模式层的异质信息网络; (2)具有复杂模式层(本体层^[2])的知识图谱。异质信息网络中不同类型的边与节点使其蕴含丰富且复杂的语义信息, 这为对其进行表示学习带来了

挑战。

简单模式层的异质信息网络与知识图谱通常遵循不同的表示学习方法。之前主流的知识表示模型为TransE^[3]及其相关变体。近年来又有基于卷积神经网络^[4]或图神经网络^[5]的知识表示方法。而简单模式层的异质信息网络大多基于元路径捕获网络结构信息以实现表示学习^[6,7]。

上述表示学习方法的不同是由两种网络的差异造成的, 具体表现在: (1)知识图谱的关系信息更加复杂。常用的异质信息网络在两个实体节点之间通常仅存在一种类型的边, 而知识图谱中两个实体节点之间可以存在多种类型的关系。(2)两种网络面向的任务不完全相同。知识图谱面向的任务偏向

收稿日期: 2020-01-09; 改回日期: 2020-12-06; 网络出版: 2020-12-15

*通信作者: 康世泽 xiaozebixia@163.com

基金项目: 国家自然科学基金(61521003)

Foundation Item: The National Natural Science Foundation of China (61521003)

推理,而简单模式层的异质信息网络面向的任务偏向分类。

为了同时表征知识图谱和简单模式层的异质信息网络,本文提出一种基于图注意力网络的异质信息网络表示学习框架。该框架可以分为3部分:基础向量部分,传播模型部分和任务模型部分。其中基础向量用于训练网络的基础向量,传播模型用于学习网络中的高阶信息,而任务模型用于执行不同的任务。本文的贡献如下:

(1) 本文对简单模式层的异质信息网络和知识图谱进行了异同点的总结,并提出了一种通用的异质信息网络表示学习框架,该框架既可以应用于知识图谱也可以应用于简单模式层的异质信息网络。

(2) 本文在多个数据集进行了广泛的实验。实验结果表明,本文所提模型与基准模型相比可以取得相对不错的效果。

2 定义

定义1(3元组) 给定一个异质信息网络 $G = (V, E)$, 本文将每组节点 v_1, v_2 和它们之间直接相连的边 r 定义为一个3元组 (v_1, r, v_2) 。对于知识图谱, v_1 是头实体, v_2 是尾实体, r 为两个实体之间的关系。

3 模型

本节详述基于图注意力网络的异质信息网络表示学习框架(HINs Embedding framework via Graph Attention Network, HE-GAN),该框架包括基础向量(basic vector)部分,传播模型(propagation model)部分,以及任务模型(prediction)部分。本文引入Conv-TransE对知识图谱执行链接预测任务,构成面向链接预测任务的表示学习模型(HE-GAN toward Link Prediction, HE-GAN-LP)。本文通过将任务模型设计成节点分类模型,构建面向节点分类的表示学习模型(HE-GAN toward Node Classification, HE-GAN-NC)。

3.1 基础异质信息网络向量

基础向量用于保持网络的基础结构信息。本文采用在知识图谱领域广泛使用的TransE模型学习知识图谱的节点向量和边向量。由于简单模式层异质信息网络中直接相连的两个节点间仅存在一种边,本文认为这些边没有丰富的语义信息。因此本文没有学习简单模式层的异质信息网络的边向量,而是采用欧氏距离作为度量保留网络中的1阶和2阶相似度。

3.1.1 简单模式层异质信息网络的基础向量

对于图 G 任一3元组 (v_1, r, v_2) ,其对应的向量为 $v_1, v_2(v_1, v_2 \in \mathbb{R}^n)$ 。本文采用欧氏距离为该3元组建模

$$f_{ED}(v_1, r, v_2) = \|v_1 - v_2\|_2^2 \quad (1)$$

本文旨在最小化现有3元组对应分数函数的距离,因此定义基于间隔的损失函数

$$L_{ED} = \sum_{r \in R_u} \sum_{(v_i, v_j) \in D_r} \sum_{(v_i', v_j') \in D_r'} \max[0, \lambda + f_u(v_i, r, v_j) - f_u(v_i', r, v_j')] \quad (2)$$

其中, R_u 是 G 中的边集合, D_r 是边 r 对应的正3元组集合, D_r' 是其对应的负样本集合。本文采用负采样的方式^[8]给每个正3元组生成负样本:对于一个特定的3元组 (u, r, v) ,首先固定 u 和 r ,通过将 v 替换成 v' 生成 K 个负3元组(v 和 v' 属于相同的节点类型,并且 (u, r, v') 在 G 中不存在)。本文全部采用相同的负采样方式。

3.1.2 知识图谱的基础向量

对于知识图谱 G 中的任一3元组 (v_1, r, v_2) ,其对应的向量为 v_1, v_2 和 r 。本文采用TransE为该3元组定义的分函数为

$$f_{TD}(v_1, r, v_2) = \|v_1 + r - v_2\| \quad (3)$$

对于知识图谱中的所有3元组,本文定义基于间隔的损失函数

$$L_{TD} = \sum_{r \in R_u} \sum_{(v_i, v_j) \in D_r} \sum_{(v_i', v_j') \in D_r'} \max[0, \lambda + f_u(v_i, r, v_j) - f_u(v_i', r, v_j')] \quad (4)$$

其中, D_r 是关系 r 对应的正3元组集合, D_r' 是其对应的负3元组集合。

3.2 传播模型

文献^[5,9]中的结果证明通过图神经网络融合高阶的邻居信息可以提高知识图谱链接预测任务的性能;文献^[7]也通过使用图注意力网络融合高阶邻居信息提升了异质信息网络的分类性能。本文借鉴这些方法^[5,7,9]提出了一种既可以融合知识图谱高阶邻居信息,也可以融合简单模式层高阶邻居信息的图神经网络作为传播模型。

本文所提传播模型的示意图如图1所示,该图展示了为节点“中国”生成向量的过程。该图中向量下方的数字表示节点的编号,相同的编号表示同一节点。向量上方的数字为向量编号,相同的数字表示同一向量。虚线表示连接操作。对于传播模型的第1层,其输入是基础向量层的全体节点向量,对于知识图谱来说还包括基础向量层的全体关系向量。基础向量层的所有节点向量可以构成矩阵 $E \in \mathbb{R}^{N \times m}$,其中 N 为网络中的节点总数。知识图谱在基础向量层对应的关系向量构成矩阵 $R \in \mathbb{R}^{M \times m}$,其中 M 为网络中关系的总数。通过一层注意力层后

生成的新向量可以构成矩阵 $\mathbf{E}^{(1)} \in \mathbb{R}^{N \times m_1}$, m_1 为新生成向量的维度；再通过一层注意力层后生成的新关系向量可以构成矩阵 $\mathbf{R}^{(1)} \in \mathbb{R}^{M \times m_2}$, m_2 为新生成向量的维度。

对于知识图谱中任一节点 v_i 对应的3元组 (v_i, r, v_j) , 本文定义注意力系数

$$c_{irj} = \mathbf{a}_i^T [\mathbf{W}_1 \mathbf{v}_i, \mathbf{W}_2 [\mathbf{r} \parallel \mathbf{v}_j]] \quad (5)$$

其中, 注意力系数 c_{irj} 表示节点 v_j 在关系 r 的连接下对节点 v_i 的重要性; $\mathbf{W}_1 \in \mathbb{R}^{k_1 \times m}$ 和 $\mathbf{W}_2 \in \mathbb{R}^{k_2 \times 2m}$ 为线性变换矩阵; $\mathbf{a}_i^T \in \mathbb{R}^{k_1+k_2}$ 为线性变换向量; \mathbf{v}_i , \mathbf{r} 和 \mathbf{v}_j 为 v_i , r 和 v_j 对应的向量; $[\cdot, \cdot]$ 和 \parallel 都表示连接操作 (concatenation)。

对于简单模式层的异质信息网络中的一个节点 v_i 和其任意一阶邻居 v_j , 本文定义注意力系数

$$c_{i,j} = \mathbf{a}_i^T [\mathbf{W}_1 \mathbf{v}_i, \mathbf{W}_2 \mathbf{v}_j] \quad (6)$$

其中, $\mathbf{W}_1 \in \mathbb{R}^{k_1 \times m}$ 和 $\mathbf{W}_2 \in \mathbb{R}^{k_2 \times 2m}$ 为线性变换矩阵, $\mathbf{a}_i^T \in \mathbb{R}^{k_1+k_2}$ 为线性变换向量。与知识图谱不同, 简单模式层的异质信息网络没有关系向量。

接下来, 本文使用 softmax 来归一化与节点 v_i (知识图谱与简单模式层的异质信息网络通用) 相关的注意力系数

$$\alpha_{irj} = \text{softmax}(c_{irj}) = \frac{\exp(\text{LeakyReLU}(c_{irj}))}{\sum_{r' \in R_{ik}} \sum_{v_k \in N_{v_i}} \exp(\text{LeakyReLU}(c_{ir'k}))} \quad (7)$$

其中, N_{v_i} 为节点 v_i 的1阶邻居集合, R_{ik} 表示节点 v_i 和节点 v_k 之间所有的关系集合, LeakyReLU 为神经网络中常用的激活函数。

为了聚合节点 v_i 在该传播层的邻居信息, 本文将不同注意力系数线性结合并采用多头注意力^[10], 为知识图谱定义的聚合公式为

$$\mathbf{v}_i^{(1)} = \parallel_{k=1}^K \sigma \left(\sum_{r_q \in R_{ij}} \sum_{v_j \in N_{v_i}} \alpha_{irj}^k \mathbf{W}_2^k [\mathbf{r}_q, \mathbf{v}_j] \right) \quad (8)$$

其中, K 是多头注意力机制对应头的数量, R_{ij} 为节点 v_i 和 v_j 之间的关系集合, \parallel 为连接操作。如图1所示, 向量1和向量2由不同的注意力头生成, 再采样连接操作将二者融合。

对于简单模式层的异质信息网络, 本文定义的聚合公式为

$$\mathbf{v}_i^{(1)} = \parallel_{k=1}^K \sigma \left(\sum_{v_j \in N_{v_i}} \alpha_{i,j}^k \mathbf{W}_2^k \mathbf{v}_j \right) \quad (9)$$

以上是经过一个注意力层对一个特定节点的操作。所有新生成的节点向量构成矩阵 $\mathbf{E}^{(1)}$ 。对于关系向量, 本文利用线性变换矩阵 \mathbf{W}^R 转化生成本注意力层对应的关系向量

$$\mathbf{R}^{(1)} = \mathbf{W}^R \mathbf{R} \quad (10)$$

之后, $\mathbf{E}^{(1)}$ 和 $\mathbf{R}^{(1)}$ 可以作为下一层的输入, 按照以上聚集邻居信息的方法, 生成 $\mathbf{E}^{(2)}$ 和 $\mathbf{R}^{(2)}$ 。重复这个过程, 最终可以生成 n 层的传播模型。本文将最后一层的节点和关系向量矩阵表示为 $\mathbf{E}^{(f)}$ 和 $\mathbf{R}^{(f)}$ 。最后一层知识图谱的节点聚合公式为

$$\mathbf{v}_i^{(f)} = \sigma \left(\frac{1}{K} \sum_{r_q \in R_{ij}} \sum_{v_j \in N_{v_i}} \alpha_{i,j}^k \mathbf{W}_2^k [\mathbf{r}_q^{(n-1)}, \mathbf{v}_j^{(n-1)}] \right) \quad (11)$$

其中, $\mathbf{r}_q^{(n-1)}$ 和 $\mathbf{v}_j^{(n-1)}$ 为第 $n-1$ 层的传播模型生成的向量。对于最后一层, 本文将不同注意力头生成的向量直接相加, 而不是采样连接操作。如图1所示, 传播模型的输出向量未增加维度。

简单模式层的异质信息网络, 对应的节点聚合公式为

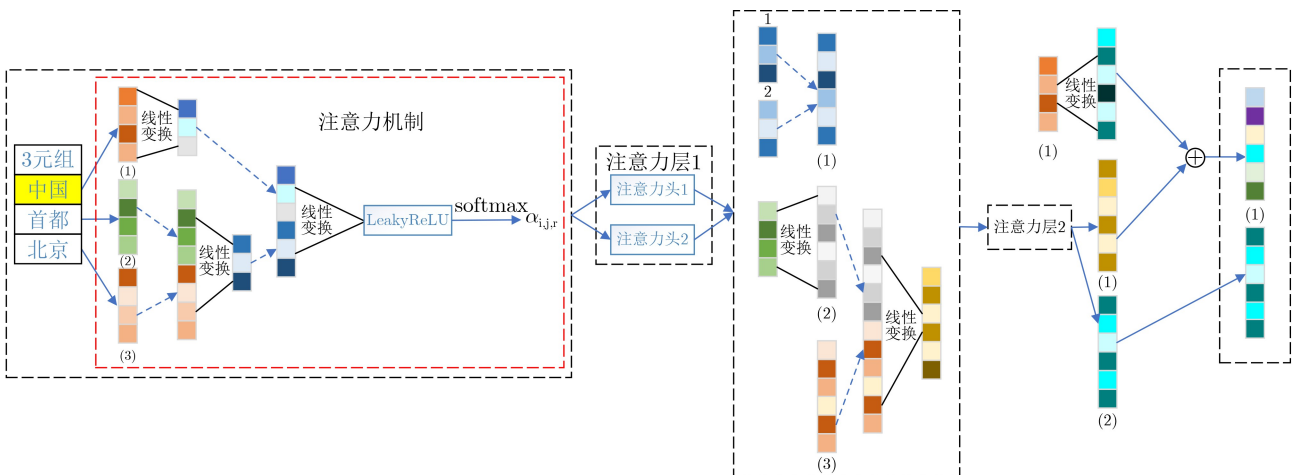


图1 知识图谱传播模型示意图

$$v_i^{(f)} = \sigma \left(\frac{1}{K} \sum_{v_j \in N_{v_i}} \alpha_{i,j}^k \mathbf{W}_2^k v_j^{(n-1)} \right) \quad (12)$$

通过传播模型获得的是网络数据的高阶邻居信息。对于知识图谱，本文将其高阶邻居信息和基础向量信息融合共同送入任务模型。具体的做法是利用线性矩阵 \mathbf{W}^E 转化基础向量 \mathbf{E} 并加上 $\mathbf{E}^{(f)}$ 构成知识图谱传播模型对应的最终向量

$$\mathbf{E}^{(f)} = \mathbf{W}^E \mathbf{E} + \mathbf{E}^{(f)} \quad (13)$$

3.3 任务模型

对于简单模式层的异质信息网络，本文对其执行节点分类任务；对于知识图谱，本文对其执行链接预测任务。

3.3.1 知识图谱的链接预测

本文采用文献[9]提出Conv-TransE来执行知识图谱的链接预测任务，该模型既可以生成比较有效的特征又可以保留TransE模型的翻译特性。该模型对应的示意图如图2所示。

本文要求传播模型每一层的节点向量维度都和关系向量维度相等。链接预测任务旨在给出一个3元组的头实体和关系，预测尾实体。对于任务中的一个3元组 (v_s, r, n) ，首先从 $\mathbf{E}^{(f)}$ 和 $\mathbf{R}^{(f)}$ 中分别取出它们对应的向量 v_s 和 r (其维度都为 m^l)，再将两个向量堆叠在一起。之后，利用 C 个卷积核对堆叠在一起的向量执行卷积操作，其中第 c 个卷积操作为

$$m_c(v_s, r, n) = \sum_{\tau=0}^{F-1} \omega_c(\tau, 0) \hat{v}_s(n + \tau) + \omega_c(\tau, 1) \hat{r}(n + \tau) \quad (14)$$

其中， F 是卷积核的宽度， n 为输出向量各单元的索引并且 $n \in [0, m^l - 1]$ ， \hat{v}_s 和 \hat{r} 分别为 v_s 和 r 的padding版本，而且卷积参数 ω_c 是可训练的。该卷积操作相当于将向量 v 和 r 相加，因此该操作保留了TransE模型的特性。这个卷积操作可以产生一个向量 $\mathbf{M}_c(v_s, r) =$

$[m_c(v_s, r, 0), m_c(v_s, r, 1), \dots, m_c(v_s, r, m^l - 1)]$ 。最后将所有卷积核生成的向量堆叠在一起生成矩阵 $\mathbf{M}(v_s, r) \in \mathbb{R}^{C \times m^l}$ 。

经过卷积操作之后的分数函数为

$$\varphi(v_s, v_o) = f(\text{vec}(\mathbf{M}(v_s, r))\mathbf{W})v_o \quad (15)$$

其中， $\mathbf{W} \in \mathbb{R}^{Cm^l \times m^l}$ 是一个线性变换矩阵，而 f 是一个非线性变换。矩阵 $\mathbf{M}(v_s, r)$ 被转换为一个向量 $\text{vec}(\mathbf{M}) \in \mathbb{R}^{Cm^l}$ 。在训练的过程中，本文对分数函数采用logistic sigmoid函数，如式(16)

$$p(v_s, r, v_o) = \sigma(\varphi(v_s, v_o)) \quad (16)$$

3.3.2 简单模式层的异质信息网络的节点分类

经过 N 层传播模型后，每个节点都会获得一组向量表示 $\{v_i, v_i^{(1)}, \dots, v_i^{(N)}\}$ 。最终，本文采用连接操作来生成任务模型的输入向量

$$v_i = v_i \parallel v_i^{(1)} \parallel \dots \parallel v_i^{(N)} \quad (17)$$

给定任务模型的输入向量，本文执行节点分类任务。为了执行该任务，首先将分类标签编码为独热(one-hot)向量，再给输入向量接入几个全连接层，以使输出向量的维度等于标签对应独热向量的维度。最后，采用交叉熵损失来估计标签节点和预测值之间的差距

$$L_{CE} = - \sum_{v_i \in Q} \mathbf{Y}^l \ln(\theta \cdot v_i^*) \quad (18)$$

其中， Q 为标记节点集合； \mathbf{Y}^l 是对应标签的独热向量； v_i^* 是全连接层的最后一层输出的向量； θ 为分类器的参数。

4 实验

本文主要执行两个任务：节点分类和链接预测。本文基于pytorch编程，所使用服务器的详细参数为：志强64核CPU；256 GB内存；8块TiTAN XP GPU。

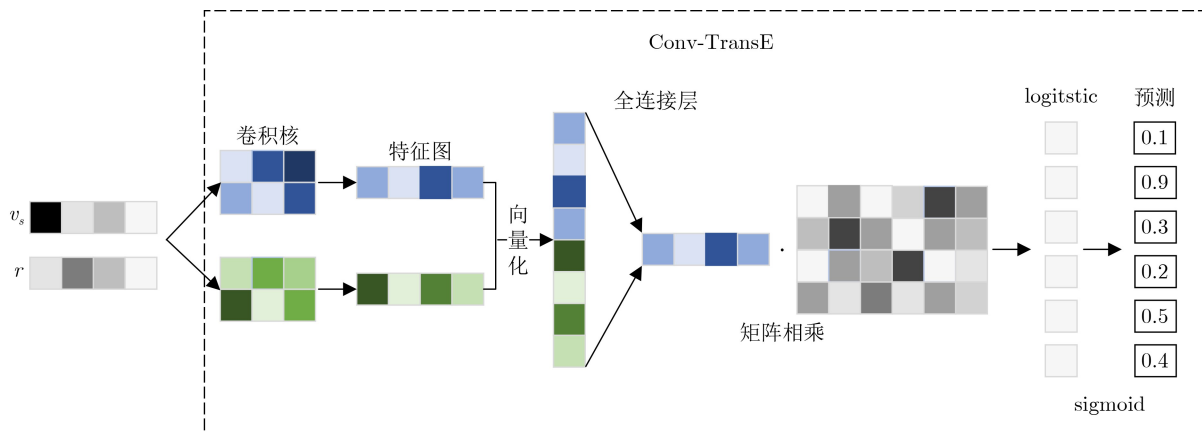


图2 Conv-TransE示意图

4.1 数据集

4.1.1 简单模式层的异质信息网络数据集

DBLP是记录计算机领域学术论文信息的数据集。本文抽样了一个DBLP的子网络构建数据集。所构建的数据集包含4种类型的节点：论文(Paper, P)，作者(Author, A)，会议(Conference, C)和术语(Term, T)。数据集相关的4个研究领域包括机器学习、数据库、数据挖掘和信息检索。对于任一作者，如果他/她的大部分论文来自会议X，便将与会议X相关的研究领域标签分配给该作者。最后，本文选择了可以被明确分类的4000名作者用于节点分类任务。

IMDB是一个关于电影的数据库，包括演员、导演、评论、简介和电影的其他信息。本文从IMDB中抽取信息构建了一个由电影(Movie, M)、演员(Actor, A)和导演(Director, D)组成的数据集。最终，本文选择了3000部可以被明确分类为动作、喜剧或戏剧的节点来执行节点分类。

本文构建数据集的具体统计数据如表1所示。

4.1.2 知识图谱数据集

本文采用了两个知识图谱数据集^[5,11]：WN18RR和FB15k-237。WN18RR和FB15k-237是为了解决相关关系问题^[5]为WN18和FB15k分别创建的子集。

4.2 节点分类

本文采用KNN分类器来执行节点分类并设置KNN的参数 $k=5$ 。本文采用Micro-F1和Macro-F1作为评估实验结果的指标。

4.2.1 基准算法

DeepWalk^[12]将异质信息网络视作同质信息网络为每个节点生成向量。

Metapath2vec^[6]利用随机游走获取每个节点的邻居信息，并利用异质Skip-Gram来学习每个节点的表示。本文对DBLP分别采用集合{APA, APCPA, APTPA}中的元路径生成向量；对IMDB分别采用集合{MAM, MDM}中的元路径生成向量。

Esim^[13]使用预定义的元路径作为向导来学习向

量。与Metapath2vec不同，Esim在学习的过程中可以使用多个元路径，而Metapath2vec在学习过程中只能采用一个元路径。

HAN^[7]也是一种基于图注意力网络的表示学习模型，该模型分别针对节点级别和语义级别(元路径)的邻居信息进行建模。本文为DBLP采用元路径{APA, APCPA, APTPA}；为IMDB采用元路径{MAM, MDM}。

Variant1为了测试简单模式层的异质信息网络是否需要学习关系向量，本文提出了令简单模式层的异质信息网络利用知识图谱的基础向量模型和传播模型生成向量的变体模型。

Variant2为了验证简单模式层的异质信息网络是否需要基础向量层，本文提出了仅包含传播模型和任务模型的变体模型。此外，本文首先训练了基于欧氏距离的节点向量，再用这些向量对传播模型进行了初始化。

训练 本文利用Adam^[14]对模型进行训练。本文对模型中所有的向量和图神经网络所有的线性变换矩阵实施 L_2 正则化来防止训练过拟合。本文设置多头注意力机制对应头的数量为4；学习率 $\lambda = 0.001$ ；图神经网络的层数 $l = 2$ 。与HAN的原论文不同，本文的数据集没有使用任何额外的特征。本文令基础向量模型的向量维度 $d = 50$ ，其他基准模型的向量维度也为50。经过第1层注意力层，单注意力头输出维度为25，总的维度为100；经过第2层注意力层，向量的输出维度为100；通过将不同层的向量连接，送入任务模型的向量维度为250。本文令负样本的大小 $n = 5$ 。对于DeepWalk, Esim和Metapath2vec，本文将每个节点的遍历数设置为50，并将步长设置为100。对于HE-GAN-NC和HAN，本文进行了10次实验并报告了平均结果。对于Metapath2vec，仅在测试所有给定的元路径后报告最佳结果。

4.2.2 实验结果

实验结果如表2所示。从实验结果可以看出，

表1 简单模式层异质信息网络数据集的统计信息

数据集	节点	#节点	关系	#连边	#训练	#验证	#测试
DBLP	Paper(P)	13769					
	Author(A)	13941	AP	30632			
	Conference(C)	20	PC	13769	2400	400	1200
	Term(T)	8623	PT	86739			
IMDB	Movie(M)	5473					
	Actor(A)	6725	MA	15814	1800	300	900
	Director(D)	2761	MD	5473			

表2 简单模式层异质信息网络的节点分类性能

数据集	指标	DeepWalk	Esim	Metapath2vec	HAN	HE-GAN-NC	Variant1	Variant2
DBLP	Macro-F1	83.15	92.47	91.63	92.52	94.31	92.26	94.16
	Micro-F1	85.77	93.60	92.64	93.67	95.17	93.42	94.81
IMDB	Macro-F1	48.34	33.89	45.13	52.29	53.58	50.35	53.11
	Micro-F1	52.48	35.25	49.38	55.86	57.92	53.96	57.32

本文所提HE-GAN-NC优于所有的基准算法,证明本文所提模型对简单模式层的异质信息网络具有比较好的学习能力。此外,图神经网络模型(HAN和HE-GAN)的整体性能优于其他传统网络表示学习模型(DeepWalk, Esim和Metapath2vec),表明图神经网络生成的特征具有更强的表征能力。

本文所提HE-GAN-NC的性能在所有数据集中都优于HAN。可能的原因是HE-GAN-NC不仅使用注意力机制来选择有用信息,而且还采用了欧氏距离来学习可以保留节点1阶和2阶相似度的基础向量。此外,由于HAN的效果是基于所给定的元路径的,相比之下本文所提的模型还具有更强的适用性。

本文所提模型的效果好于Variant1,说明学习简单模式层的异质信息网络的边信息并不能促进实验的效果。本文所提模型的效果好于Variant2,说明基础向量模型的提出可以更进一步提升模型性能。

4.3 知识图谱的链接预测

知识图谱的链接预测任务旨在预测3元组中丢失的头实体或者尾实体,即对于一个3元组 (v_1, r, v_2) 给定 (r, v_2) 预测 v_1 或者给定 (v_1, r) 预测 v_2 。对于测试集中的每一个3元组 (v_1, r, v_2) ,本文通过将 v_1 或 v_2 换成实体集合 E 中的其他实体来构建损坏3元组并限定这些损坏3元组没有在知识图谱的训练、验证和测试集中出现过。本文采用的指标有平均倒数(Mean Reciprocal Rank, MRR)和Hits@ N (正确的3元组在前 N 项中的排序),其中 N 在本文取1, 3和10。更高的MRR值和Hits@ N 值表示更好的实验效果。

4.3.1 基准算法

TransE^[3]是比较简单有效的模型,它将尾实体看作从头实体经过关系的翻译。

ConvE^[4]是一种多层卷积神经网络模型,其参数利用率高,善于学习复杂结构。

ConvKB^[11]通过卷积神经网络捕获实体和关系间的全局关系与翻译特性。

SACN(GCN+Conv-TransE)^[9]是一个端到端的图卷积网络模型,该模型利用加权的图卷积网络学习知识向量,并利用Conv-TransE执行知识图谱的链接预测任务。

relationPrediction(GAT+ConvKB)^[5]使用图

注意力网络并融合关系信息学习知识图谱的结构信息,并使用ConvKB作为解码器。

Variant3为了验证知识图谱的表示学习是否需要基础向量层,本文提出了仅包含传播模型和任务模型的变体模型。此外,本文首先利用TransE模型训练了知识向量,再用这些向量对传播模型进行了初始化。

训练 本文利用Adam对模型进行训练。本文对模型中所有的向量和图神经网络所有的线性变换矩阵实施 L_2 正则化来防止训练过拟合。本文设置多头注意力机制对应头的数量为4;学习率 $\lambda = 0.001$;图神经网络的层数 $l = 2$ 。对于卷积核数 C ,本文在FB15k-237数据集上取100,在WN18RR数据集上取200。本文令基础向量模型的向量维度 $d = 64$ 。经过第1层注意力层,单注意力头输出向量维度为64,总的维度为256;经过第2层注意力网络,最终的输出维度为256。本文令负样本的大小 $n = 5$ 。

4.3.2 实验结果

从表3的实验结果可以看出,本文提出的HE-GAN-LP与基准算法相比取得了比较好的效果。在所有的模型中,TransE的实验效果最不理想。可能的原因是TransE是一种实验参数较少的线性模型,而其他神经网络模型可以生成更有效的特征。SACN的效果整体优于ConvKB,可能的原因是SACN提出的Conv-TransE可以更加有效地保留3元组的翻译特性。本文所提模型比其变体模型Variant3整体效果略好,可能的原因是基础向量层可以使输入图注意力网络的知识向量始终保持翻译特性。本文所提模型的效果整体上稍好于relation-Prediction(GAT+ConvKB)模型。其可能的原因是,本文所提模型具有基础向量层,而且利用了可以生成有效特征的Conv-TransE模型。

此外,在FB15k-237数据集上,基于图注意力网络的模型(relationPrediction和HE-GAN-LP)效果优于基于图卷积网络的模型(SACN);然而,在WN18RR数据集上,基于图注意力网络的模型并未比SACN模型有显著提升。可能的原因是FB15k-237数据集的平均入度比较高,注意力方法可以从各节点的邻居中选择有用的节点提升实验性能,而

表3 知识图谱的链接预测任务性能

数据集	FB15k-237				WN18RR			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TransE	0.279	0.19	0.38	0.44	0.242	0.04	0.44	0.53
ConvE	0.315	0.24	0.35	0.49	0.461	0.42	0.47	0.53
ConvKB	0.285	0.19	0.32	0.47	0.263	0.06	0.45	0.55
SACN	0.352	0.26	0.39	0.54	0.463	0.43	0.48	0.54
relationPrediction	0.518	0.46	0.54	0.63	0.440	0.36	0.48	0.58
HE-GAN-LP	0.523	0.46	0.56	0.66	0.468	0.41	0.50	0.59
Variant3	0.520	0.45	0.55	0.64	0.447	0.37	0.48	0.58

WN18RR数据集的平均入度太小(2.12)，图注意力网络难以发挥优势。

5 总结与展望

本文提出了一种通用的异质信息网络表示学习框架，该框架可以分为3部分：基础向量模型，基于图注意力网络的传播模型以及任务模型。对于每一个模型，本文针对不同的异质信息网络，采用了不完全相同的设计方法，并解决了不同类型网络中存在的一些问题。该框架可以应用于多个任务并取得良好的效果。本框架的问题在于各模块对于不同类型的网络还需要特定的设计，未来计划提出更加具有通用性的表示学习模型。

参考文献

- [1] 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. doi: [10.1109/TKDE.2016.2598561](https://doi.org/10.1109/TKDE.2016.2598561).
- [2] 于洪涛, 丁悦航, 刘树新, 等. 一种基于超节点理论的本体关系消冗算法[J]. *电子与信息学报*, 2019, 41(7): 1633–1640. doi: [10.11999/JEIT180793](https://doi.org/10.11999/JEIT180793).
YU Hongtao, DING Yuehang, LIU Shuxin, *et al.* Eliminating structural redundancy based on super-node theory[J]. *Journal of Electronics & Information Technology*, 2019, 41(7): 1633–1640. doi: [10.11999/JEIT180793](https://doi.org/10.11999/JEIT180793).
- [3] BORDES A, USUNIER N, GARCIA-DURÁN A, *et al.* Translating embeddings for modeling multi-relational data[C]. *The 26th International Conference on Neural Information Processing Systems*, Lake Tahoe, USA, 2013: 2787–2795.
- [4] DETTMERS T, MINERVINI P, STENETORP P, *et al.* Convolutional 2D knowledge graph embeddings[C]. *The 32nd AAAI Conference on Artificial Intelligence (AAAI-18)*, New Orleans, USA, 2018: 1811–1818.
- [5] NATHANI D, CHAUHAN J, SHARMA C, *et al.* Learning attention-based embeddings for relation prediction in knowledge graphs[C]. *The 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019: 4710–4723. doi: [10.18653/v1/P19-1466](https://doi.org/10.18653/v1/P19-1466).
- [6] DONG Yuxiao, CHAWLA N V, SWAMI A, *et al.* Metapath2vec: Scalable representation learning for heterogeneous networks[C]. *The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA, 2017: 135–144. doi: [10.1145/3097983.3098036](https://doi.org/10.1145/3097983.3098036).
- [7] WANG Xiao, JI Houye, SHI Chuan, *et al.* Heterogeneous graph attention network[C]. *The World Wide Web Conference*, San Francisco, USA, 2019: 2022–2032. doi: [10.1145/3308558.3313562](https://doi.org/10.1145/3308558.3313562).
- [8] MIKOLOV T, CHEN Kai, CORRADO G, *et al.* Efficient estimation of word representations in vector space[C]. *The 1st International Conference on Learning Representations*, Scottsdale, Arizona, 2013: 1–12.
- [9] SHANG Chao, TANG Yun, HUANG Jing, *et al.* End-to-end structure-aware convolutional networks for knowledge base completion[C]. *The AAAI Conference on Artificial Intelligence*, Hawaii, USA, 2019: 3060–3067. doi: [10.1609/aaai.v33i01.33013060](https://doi.org/10.1609/aaai.v33i01.33013060).
- [10] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, *et al.* Graph attention networks[C]. *International Conference On Learning Representations*, Vancouver, Canada, 2018: 1–12.
- [11] NGUYEN D Q, NGUYEN T D, NGUYEN D Q, *et al.* A novel embedding model for knowledge base completion based on convolutional neural network[C]. *2018 Conference of the North American Chapter of the Association for Computational Linguistics*, New Orleans, USA, 2018: 327–333. doi: [10.18653/v1/N18-2053](https://doi.org/10.18653/v1/N18-2053).
- [12] PEROZZI B, AL-RFOU R, SKIENA S, *et al.* DeepWalk: Online learning of social representations[C]. *The 20th ACM SIGKDD International Conference on Knowledge Discovery*

and Data Mining, New York, USA, 2014: 701–710. doi: [10.1145/2623330.2623732](https://doi.org/10.1145/2623330.2623732).

- [13] SHANG Jingbo, QU Meng, LIU Jialu, *et al.* Meta-path guided embedding for similarity search in large-scale heterogeneous information networks[J]. arXiv preprint arXiv: 1610.09769v1, 2016.
- [14] KINGMA D P and BA J. Adam: A method for stochastic optimization[C]. The 3rd International Conference for

Learning Representations, San Diego, USA, 2015: 1–15.

康世泽: 男, 1991年生, 博士生, 研究方向为知识图谱.

吉立新: 男, 1969年生, 研究员, 博士生导师, 研究方向为电信网
关信息防护、大数据分析.

张建朋: 男, 1988年生, 助理研究员, 研究方向为大数据分析.

责任编辑: 余 蓉