

基于层级注意力机制的链接预测模型研究

赵晓娟^{1,2}, 贾焰¹, 李爱平¹, 陈恺¹

(1. 国防科技大学计算机学院, 湖南 长沙 410073; 2. 湖南工业大学商学院, 湖南 株洲 412007)

摘要: 为了解决已有图注意力机制在进行链接预测相关任务时, 容易造成注意力分配向某些出现频率高的关系倾斜的问题, 提出了一种基于层级注意力机制的链接预测模型。在链接预测任务中, 通过设计分层注意力机制, 根据预测任务中的关系对知识图谱中与给定实体相连的不同类型的关系给予不同的注意力。在关注多跳邻居实体特征的同时, 更关注关系特征以找到符合目标关系的关系类型。在多个基准数据集上与主流模型进行对比实验, 实验结果表明, 所提模型性能优于主流模型, 并具有较好的稳健性。

关键词: 层级注意力机制; 链接预测; 知识图谱嵌入

中图分类号: TP391

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021057

Research on link prediction model based on hierarchical attention mechanism

ZHAO Xiaojuan^{1,2}, JIA Yan¹, LI Aiping¹, CHEN Kai¹

1. College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China

2. College of Business, Hunan University of Technology, Zhuzhou 412007, China

Abstract: In order to solve the problem that the existing graph attention mechanism tends to cause attention distribution to certain relations with high frequency when performing link prediction related tasks, a new link prediction model based on hierarchical attention mechanism was proposed. In the link prediction task, a hierarchical attention mechanism was designed to give different attention to the relationships of different relationship types connected to a given entity in the knowledge graph according to the relationship in the prediction task. While the characteristics of multi-hop neighbor entities were pay attention to, the relationship characteristics was pay more attention to find the relationship type that matches the target relationship. Through comparison experiments with the mainstream models on multiple benchmark data sets, the results show that the performance of the model is better than the mainstream models and has good robustness.

Keywords: hierarchical attention mechanism, link prediction, knowledge graph embedding

1 引言

在资源描述框架 (RDF, resource description

framework) 下, 知识可以表示为事实三元组, 即 (头实体, 关系, 尾实体) 的形式, 例如 “A 是一名运动员” 可表示为(A, 职业, 运动员), 其中, “A”

收稿日期: 2020-08-20; 修回日期: 2020-11-06

通信作者: 李爱平, liaiping@nudt.edu.cn

基金项目: 国家重点研发计划基金资助项目 (No.2017YFB0802204, No.2016QY03D0603, No.2016QY03D0601, No.2017YFB0803301, No.2019QY1406); 广东省重点研发计划基金资助项目 (No.2019B010136003); 国家自然科学基金资助项目 (No.61732004, No.61732022, No.61672020); 湖南省重点研发计划基金资助项目 (No.2018GK2056); 湖南省教育厅科研基金资助项目 (No.19C0597)

Foundation Items: The National Key Research and Development Program of China (No.2017YFB0802204, No.2016QY03D0603, No.2016QY03D0601, No.2017YFB0803301, No.2019QY1406), The Key Research and Development Program of Guangdong Province (No.2019B010136003), The National Natural Science Foundation of China (No.61732004, No.61732022, No.61672020), The Key Research and Development Project of Hunan Province (No.2018GK2056), Scientific Research Found of Hunan Provincial Education Department (No.19C0597)

“职业”“运动员”分别是三元组的头实体、关系和尾实体。知识也可以表示为有向多关系图，即知识图谱，其中，每个节点对应一个实体，连接节点的每条边对应一种关系。知识图谱中的一个三元组或者事实 h, r, t 表示一对实体及实体之间的关系。图 1 为 RDF 表示与知识图谱表示的示例。

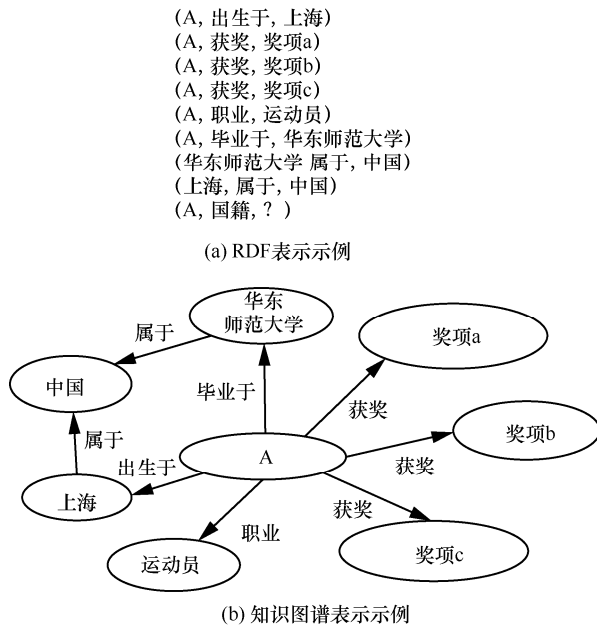


图 1 RDF 表示与知识图谱表示的示例

知识图谱嵌入旨在学习知识图谱中的每个元素（实体和关系）在连续低维度向量空间中的潜在表示，使知识图谱具有可计算性，并且更容易与深度学习模型集成。它支撑着许多实际应用，包括知识问答^[1-3]、推荐系统^[4-6]和其他自然语言理解任务^[7-9]。然而，即使 YAGO^[10]、DBpedia^[11]、Freebase^[12]这样包含了数十亿个事实的大型知识图谱，也不能避免关系或者节点缺失的问题^[2,13-14]，从而引出了知识图谱补全任务的研究。这类研究任务一方面利用知识图谱中已经存在的信息来自动推断缺失的事实，另一方面预测三元组是否有效。

在知识图谱中，实体存在的意义很大程度上取决于其连边的关系类型，同一个实体在不同的关系下，其表示的重点也应有所不同。例如，对于图 1 中的实体“A”作为(A, 职业, 运动员)的头实体，在嵌入表示时更多地体现作为一个运动员的属性；作为(A, 毕业于, 华东师范大学)的头实体，在嵌入表示时则更多地体现作为学生的一些属性。考虑(A, 国籍, ?)这样一个链接预测问题，根据问题中的关系“国籍”可知，“职业”和“获奖”这 2 种关系对推

理“国籍”的贡献不会比“出生于”和“毕业于”这样的关系大，因为与“出生于”和“毕业于”相连的尾实体才是与国籍有关的地点。

事实上，当使用知识图谱嵌入技术将实体和关系投影到低维连续向量空间时，根据已知事实(A, 出生于, 上海)和(上海, 属于, 中国)，本文可以把“A”和“中国”都映射到“上海”附近的向量空间，所以它们在向量空间中的距离比较近，而且，在关系嵌入的向量空间中，相对于“职业”和“获奖”，“出生于”与“国籍”的嵌入向量所表达的语义更接近。因此，通过“出生于”推导出“国籍”的尾实体比通过“职业”或“获奖”推导出的结果更可信。

深度学习的注意力机制模拟人类的选择性视觉注意力，从众多信息中选择与当前预测目标最相关的信息，并根据这些信息做出预测，近年来被各个领域的不同任务广泛采用。文献[15]提出一种在知识图谱推理中关注关系的方法，对不同的关系赋予不同的注意力，但是该方法与其他图神经网络一样，在训练过程中，随着网络层数的增加和迭代次数的增加，每个节点的隐层表征会趋向于收敛到同一个值，即通常所说的过度平滑问题。

基于前述问题，本文提出一种新的层级注意力机制的链接预测模型。模型的主要思想是通过分层聚合来避免过度平滑问题，同时，在关系子图之间的信息聚合时，设计一种更简洁的关系注意力机制，可以根据预测问题中的目标关系对知识图谱中给定实体不同类型的关系给予不同的注意力。具体来说，在低维向量空间中，学习给定实体及其多跳邻域的实体和关系的特征，根据与目标关系的距离分配注意力，通过将更多的注意力分配给语义更接近的关系来获得更准确的尾实体预测结果。

本文的主要贡献是设计了一种基于分层注意力机制的嵌入模型，并将模型应用在知识图谱链接预测任务中。分层注意力机制除了关注多跳邻居实体特征外，能更加关注关系特征以找到符合目标关系的关系类型。

2 相关研究工作

以图神经网络(GNN, graph neural network)^[16]为基础的图卷积网络(GCN, graph convolutional network)^[17]能同时对节点特征信息与结构信息进行端对端学习，是目前对图数据学习任务的最佳选

择。从 GCN 开始, 研究者越来越多地关注将卷积运算引入图领域, 这里的图是指图论中用顶点和边建立相应关系的拓扑图。但是, 传统的离散卷积在知识图谱这种非欧几里得数据上无法保持平移不变性, 如何定义能够处理大小可变的邻居集和共享参数的操作是一个具有挑战性的问题。

为了解决这个问题, 文献[18]提出 GraphSAGE (graph sample and aggregate) 模型, 从每个节点的邻居节点中提取固定数量的节点, 然后使用聚合函数融合这些邻居节点的信息。当然, 这也意味着模型不能采样到所有的邻居。与文献[18]类似, 图注意力 (GAT, graph attention) [19]也是该研究领域一个典型的模型, 这个模型是基于所有邻居节点, 而且对邻居节点的顺序没有要求。虽然该模型在图网络结构中取得了成功, 但不能直接用于知识图谱, 因为这个模型仅考虑了节点, 而忽略了知识图谱中非常重要的一部分信息, 即知识图谱中实体之间的关系。本质上, GCN [17]和 GAT [19]都是将邻居节点的特征聚合到中心节点, 利用图上的局部平衡状态学习新的节点特征表达式; 不同之处在于 GCN 使用了拉普拉斯矩阵, 而 GAT 使用的是注意力机制。CompGCN [20]是一个考虑多种关系信息的图神经网络框架, 它通过共同学习多关系知识图谱中关系和节点的向量表示, 解决了传统 GCN 的难点。为了避免随着关系数量增加, 参数数量显著增加的问题, CompGCN 使用一组基作为可学习的基向量, 而不是为每个关系定义嵌入。

实际上, 对于每个节点, 模型选取它的一些邻居, 每个邻居对节点都有一定的影响, 但每个邻居的影响力都可能不同, 文献[17, 20]没考虑每个邻居对节点的影响力的差异。因此, 文献[21]提出了一种端到端的模型, 该模型在聚合每一层节点特征时考虑了知识图谱中的关系, 但是, 该模型在计算注意力时只考虑了关系和实体表示的简单拼接, 并没有反映不同关系对特定推理任务的重要程度不同。文献[22]的 Minerva 模型将每一种关系设置为一种任务类型, 在预训练中针对特定任务进行模型训练, 这实际上是一种考虑知识图谱中不同类型关系的实践, 该模型根据关系对三元组进行分类, 并没有分别考虑邻居节点和关系对目标节点嵌入的影响。

3 任务描述

在知识图谱链接预测任务中, 目标是当 u 或 v

丢失时, 推断一个三元组 (u, r, v) 是否是有效三元组, 即给定 (r, v) 推导 u 或者给定 (u, r) 推导 v , 本文研究的任务包括 $(u, r, ?)$ 和 $(?, r, v)$ 。对于这 2 种情形, 本文并没有单独为每一种情况训练一个模型, 而是训练了同一个模型用于这 2 种情况的评估。对于每个测试三元组 (u, r, v) , 本文用知识图谱中存在的所有实体替换每个头实体, 构造 $(n-1)$ 个被损坏的三元组, 然后对每一个这样的三元组评分。最后将这些分数按升序排序, 得到正确的三元组的排列。本文也可以用同样的方法替换三元组的尾实体。

知识图谱用 $G=(E, R, V)$ 表示, 这里的 E, R, V 分别表示实体、关系、三元组的集合。 $V = \{(u, r, v) \in E \times R \times E\}$, 其中, $u, v \in E$ 是实体, $r \in R$ 是实体之间的关系。嵌入模型尝试学习一个有效的函数 $f(g)$, 对于给定的输入三元组 $T=(u, r, v)$, $f(T)$ 给出 T 是一个有效三元组的可能性。

4 模型设计

4.1 模型整体框架

本节将详细描述本文提出的模型。异构图注意力 [23] 在异构图神经网络中使用分层注意力机制, 包括节点级别的注意力和语义级别的注意力。受该思想的启发, 本文提出了一种新的基于层级注意力机制的链接预测模型。整个链接预测模型主要由两部分组成: 编码部分和解码部分, RAKGR (relation attention based knowledge graph reasoning) 作为编码器, ConvKB [24] 作为解码器。RAKGR 由多层组成, 整体框架如图 2 所示, 其每一层由 GAT 和关系注意力 (RAT, relation attention) 两部分组成。每一个 GAT 和 RAT 都由多个注意力头组成。本文假设 RAT 和 GAT 的注意力个数相同。图 2 中的多头注意力拼接展示了将每个注意力头的输出结果 concat 之后再降维, 也可以直接求各个注意力头输出的平均值。ConvKB 由一个二维卷积 Conv2D 表示, 其相关内容详见文献[24]。

图 3 给出了某一层中的某个注意力头 “GAT+RAT” 示例, 来说明图 2 中所示的 “GAT+RAT” 的工作原理。这里仅表示单一注意力头, 且仅说明其中一层的情况。中心节点 0 有 9 个一阶直接邻居, 将这些邻居分成 3 个邻居子图 {1, 2}、{3, 4, 5, 6}、{7, 8, 9}, 这 3 个子图与中心节点之间的关系分别为 r_1 、 r_2 和 r_3 。图 3 中右侧黑色圆圈表示每个中心节点 0 为每个邻居子图设置的对应

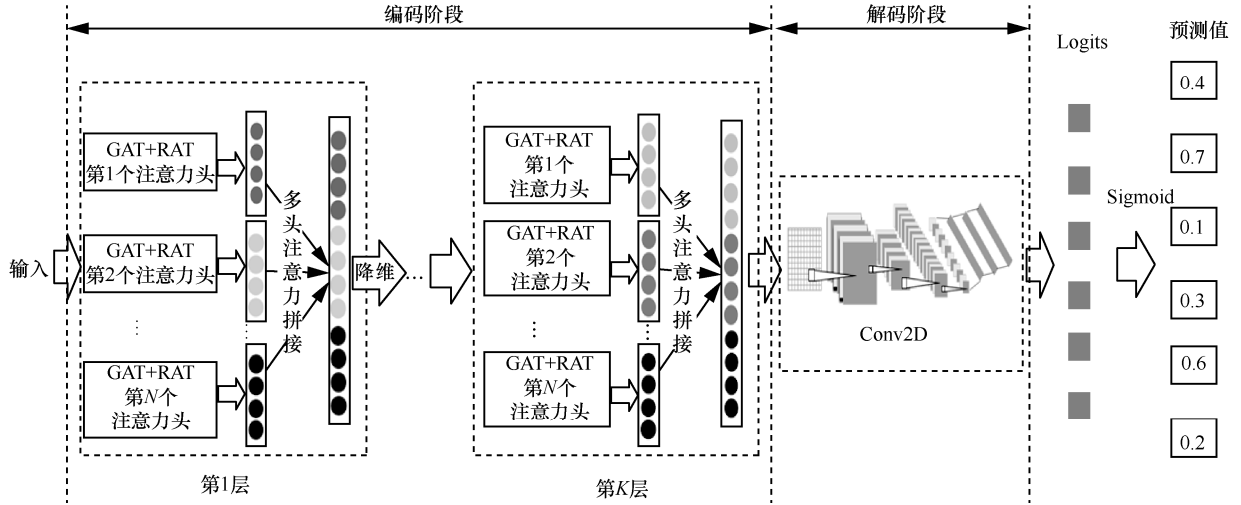


图2 RAKGR 整体框架

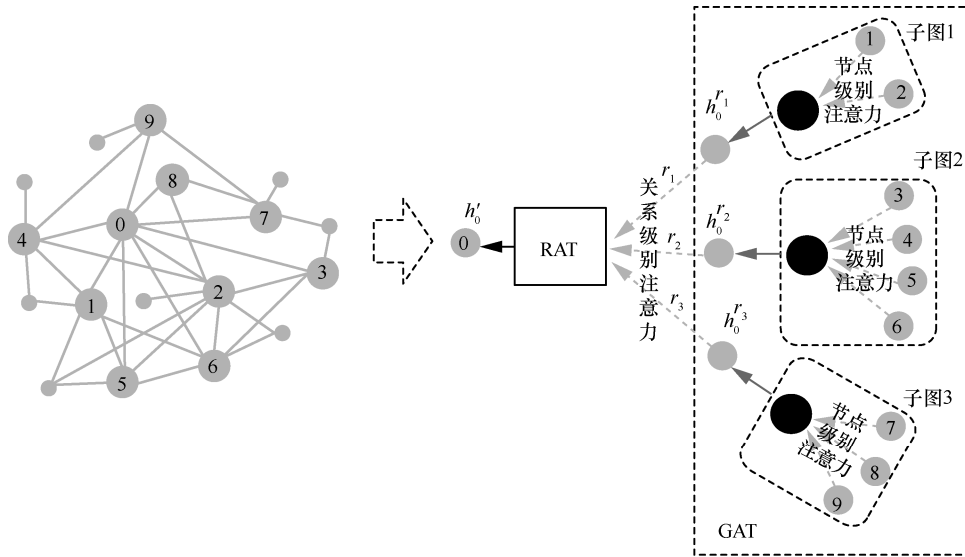


图3 某一层中的某个注意力头“GAT+RAT”示例

的虚拟节点。相应地， $h_0^{r_1}$ 、 $h_0^{r_2}$ 和 $h_0^{r_3}$ 分别对应基于关系 r_1 、 r_2 和 r_3 的子图在聚合邻居信息后的中心节点0的嵌入表示。首先，用GAT聚合子图内部各节点的信息；然后，用关系注意力机制聚合这3个虚拟节点的信息，以获得更新后的中心节点0的嵌入表示。

4.2 构建邻居子图

给定三元组 $(u, r, v) \in G$ ， h_u^0 、 h_r^0 、 h_v^0 分别表示 u 、 r 、 v 对应的初始嵌入。本文用TransE^[25]获得模型的初始嵌入，然后分别用2种类型的转换矩阵 W_v 和 W_r 将节点和关系投射到相同的特征空间。本文将某个中心节点的所有一阶邻居根据其与中心节点之间的关系类型分成不同的子图，也就是说，中心节点与它的一阶邻居之间有多少种类型的关系，就会生成多少个邻居子图。每个子图中所有节

点与中心节点之间的关系是一样的。与文献[23]类似，本文将聚合的过程分为2个级别：子图内部的信息聚合和关系子图之间的信息聚合。

4.3 子图内部的信息聚合

为了获得节点 v 的新的嵌入表示，本文将与该节点相连的每一个实体表示为

$$a_{uv} = \text{LeakyReLU}\{W_1[h'_v \| h'_u]\} \quad (1)$$

其中， W_1 为一个线性变换矩阵。

在考虑邻居节点对目标节点的重要性时，本文用 a_{uv} 表示注意力的绝对值。然后，类似于GAT^[19]，将上述注意力的绝对值进行归一化，如式(2)所示。

$$\alpha_{uv}^r = \text{soft max}(a_{uv}) = \frac{\exp(a_{uv})}{\sum_{n \in N(v)} \exp(a_{un})} \quad (2)$$

其中, α_{uv}^r 为节点 u 与节点 v 在关系 r 下的相关权重系数, $N(v)$ 为与节点 v 直接相连的邻居节点的集合。根据式(2)中得到的相关性权重系数, 用式(3)将子图中每个邻居的信息进行聚合。

$$h'_{N(v_r)} = \sigma \left(\sum_{uv} \alpha_{uv}^r h'_u \right) \quad (3)$$

其中, $h'_{N(v_r)}$ 为关系 r_t 下节点 v 聚合邻居节点信息后的嵌入表示; $N(v_r)$ 为与节点 v 之间存在关系 r_t 的邻居节点的集合。

使用多头注意力机制可以获取更多关于邻居节点的信息, 因此, 对应多头注意力机制的情况, 可将式(3)转换为

$$h'_{N(v_r)} = \parallel \sigma \left(\sum_{uv} \alpha_{uv}^r h'_u \right) \quad (4)$$

其中, \parallel 表示拼接操作, 是一种集成多个注意力头输出结果的聚合方式。多头和单头的区别在于每个注意力头的权重系数不一样, 但模型是一样的。

假设用一个虚拟节点来表示某个邻居子图的所有信息, 那么中心节点与邻居子图之间的关系就简化成了节点与节点之间的关系。对应于关系 r_t 的虚拟节点可表示为

$$h_{N(v_r)} = f(h'_v, h'_{N(v_r)}) \quad (5)$$

其中, $f(\cdot)$ 是转换函数, h'_v 是中心节点 v 在高维空间中的嵌入表示。由此可以得到 T 组特定关系的节点嵌入为 $\{h_{N(v_r)}, h_{N(v_r)}, \dots, h_{N(v_r)}\}$ 。

4.4 关系子图之间的信息聚合

根据目标关系给每个邻居子图分配不同的注意力值, 也称为关系子图级别的聚合。本文要给每个邻居子图计算一个注意力分数 a_r 。区别于 A2N (attending to neighbor) [15] 的注意力机制, 本文将子图之间信息聚合环节的注意力分数表示为

$$a_r = f(v, r, N(v_r)) = \mathbf{W} [h'_v \parallel h'_r \parallel h_{N(v_r)}] \quad (6)$$

对上述注意力分数 a_r 进行标准化处理, 得到每个邻居子图对中心节点 v 的相关性权重值为

$$\alpha'_r = \text{soft max}(a_r) = \frac{\exp(a_r)}{\sum_{i=1}^T \exp(a_r)} \quad (7)$$

将学习到的权重 α'_r 作为系数, 聚合这些特定关系的嵌入, 得到节点 v 更新后的嵌入表示为

$$h_v = \sigma \left(\sum_t \alpha'_t h_{N(v_r)} \right) \quad (8)$$

其中, α'_r 代表与中心节点之间存在关系 r_t 的邻居集对中心节点 v 的嵌入表示的影响, 也就是 $h_{N(v_r)}$ 对 h_v 的影响; h_v 是中心节点 v 更新后的嵌入表示。

上述过程仅描述了 RAKGR 模型的单层单头注意力聚合过程。同理, 可以采用多头注意力机制获得更丰富的邻居信息。类似地, M 个注意力的情况表示为

$$h_v = \parallel \sigma \left(\sum_t \alpha'_t h_{N(v_r)} \right) \quad (9)$$

上述聚合过程可以扩展到多层, 使模型具有高阶传播的特点, 即

$$h_v^{k+1} = f(h_v^k, h_{N(v)}^k) \quad (10)$$

对于一个节点与另一个节点之间存在多个关系的情况, 本文将这个节点复制 n 次得到 n 个不同的节点, 因此, 本文在后续知识图谱推理的应用中根据知识图谱中边的数量来构建邻居子图。

4.5 模型优化目标

本文的模型第一阶段训练目标借鉴了 TransE^[25] 的平移平分函数的思想。对于一个给定三元组 $t_{ij}^k = (u, r, v)$, 学习一种嵌入表示, 其能量函数定义为

$$d(u, r, v) = \|u + r - v\|_2^2 \quad (11)$$

本文使用基于边际的评分函数作为训练目标, 定义为

$$L(u, r, v) = \sum_{(u, r, v) \in T} \sum_{(u', r, v) \in T^-} [\gamma + d(u + r, v) - d(u' + r, v)]_+ \quad (12)$$

其中, $[x]_+$ 表示取 x 正的部分, $\gamma > 0$ 表示一个边际超参数, $(u, r, v) \in T$, T 表示有效三元组集合, 也称为正样本, 而 T^- 则是负样本, 表示为

$$T^- = \{(u', r, v) | u' \in E \setminus u\} \cup \{(u, r, v') | v' \in E \setminus v\} \quad (13)$$

本文通过用其他实体替换三元组的头实体或者尾实体来得到负样本。

第二阶段的 ConvKB 采用与文献[21]类似的方法, 定义评分函数为

$$f(t_{ij}^k) = \left(\parallel_{m=1}^K \text{ReLU}(\text{concat}(e_i, e_r, e_j) * \omega^m) \right) \mathbf{W} \quad (14)$$

其中, e_i 、 e_r 和 e_j 分别表示头实体的最终嵌入、关系的最终嵌入和尾实体的最终嵌入; concat 表示将

表 1

数据集的统计情况

数据集	实体数量/个	关系数量/个	训练集三元组/个	验证集三元组/个	测试集三元组/个	三元组总数/个
FB15k-237	14 541	237	272 115	17 535	20 466	310 116
WN18RR	40 943	11	86 835	3 034	3 134	93 003

这些嵌入表示拼接起来；*表示卷积操作； κ 表示卷积核的个数； ω^m 表示第 m 个卷积核； W 表示一个线性变换矩阵，用于计算三元组的最终得分。本文使用 soft-margin 损失对模型进行训练，表示为

$$L = \sum_{t_{ij}^k \in (T \cup T^-)} \log \{1 + \exp(L_{t_{ij}^k} f(t_{ij}^k))\} + \frac{\lambda}{2} \|W\|_2^2 \quad (15)$$

$$\text{其中, } L_{t_{ij}^k} = \begin{cases} 1, & t_{ij}^k \in T \\ -1, & t_{ij}^k \in T^- \end{cases}.$$

5 实验

5.1 实验数据集

本文提出的层级注意力机制中，子图之间的注意力分配重点考虑了关系在预测任务中的意义，所以，本文选取 FB15k-237^[26]、WN18RR^[27]这 2 个数据集评估前述模型。一方面，FB15k-237 中有 237 种关系，是一种典型的多关系数据集，可以用来验证本文的模型在关系类型较多的情况下其优势更加明显，WN18RR 的节点数比较多，但关系只有 11 种；另一方面，由于 WN18^[28]和 FB15K^[25]中存在很多逆关系，这些逆关系会影响本文预测任务的结果，因此，本文使用的是 WN18 和 FB15K 删除了逆关系之后的子集 WN18RR 和 FB15k-237。数据集的统计情况如表 1 所示。

5.2 实验设置

本文的实验中用到的初始实体嵌入和关系嵌入用 TransE^[25]获得。模型的训练分成 2 个步骤。首先，使用 RAKGR 模型对知识图谱中的实体和关系进行编码；然后，使用 ConvKB 对得到的节点和关系表示进行解码，以得到符合链接预测任务的知识表示方式。

在这类任务中，通常的做法是将正确答案的排列顺序记录在有序列表中，以便确定是否可以将正确答案排列在错误答案之前。常用的 3 个评估指标是平均排序 (MR, mean rank)、平均倒数排序 (MRR, mean reciprocal rank) 和 N 个正确排序所占的比例 (Hits@ N)。MR 是所有排序的均值；MRR 与 MR 类似，但是 MRR 是一种比 MR 更稳健的衡量方法，因

为某一的特别糟糕的排名可以在很大程度上影响平均排名；Hits@ N 中的 N 可以取 1、3 和 10。对于本文模型，MRR 和 Hits@ N 的值越大意味着模型性能越好，MR 的值越小意味着模型性能越好。本文选择 MRR 和 Hits@1、Hits@3、Hits@10 作为评估指标。

为了评估 RAKGR 模型性能，本文选择了几类目前较先进的知识图谱嵌入模型进行对比，包括 A2N^[15]、Minerva^[22]、ConvKB^[24]、TransE^[29]、DistMult^[29]、Complex^[30]、ConvE^[28]。

5.3 实验结果与分析

实验结果如表 2~表 5 所示。A2N、DistMult、Complex、ConvE 的实验结果来自文献[15]，该文献下载了公开的源代码来复现所有的结果；ConvKB、TransE 在 2 个数据集上的实验结果是本文下载开源代码并复现的结果；Minerva 在 2 个数据集上的结果来自文献[22]。文献[21]指出，对于一个 n 层模型，其输入信息是从 n 跳的邻域上累积的。所以，本文实验设置为 2 层模型，相当于聚合了 2 跳邻域的信息。当然，本文提出的模型是可以扩展到任意层数的。

表 2 数据集 FB15k-237 预测尾实体的实验结果

模型	Hits@1	Hits@3	Hits@10	MRR
Complex	0.303	0.434	0.572	0.394
DistMult	0.275	0.417	0.568	0.37
ConvE	0.313	0.457	0.6	0.41
Minerva	0.217	0.329	0.456	0.293
A2N	0.328	0.464	0.608	0.422
RAKGR	0.331	0.476	0.621	0.457

表 3 数据集 FB15k-237 预测尾实体或者头实体的实验结果

模型	Hits@1	Hits@3	Hits@10	MRR
TransE	0.198	0.376	0.441	0.279
Complex	0.158	0.275	0.428	0.247
DistMult	0.196	0.304	0.444	0.278
ConvE	0.237	0.356	0.501	0.325
ConvKB	0.198	0.324	0.471	0.289
A2N	0.232	0.348	0.486	0.317
RAKGR	0.242	0.371	0.52	0.332

表 4 数据集 WN18RR 预测尾实体的实验结果

模型	Hits@1	Hits@3	Hits@10	MRR
Complex	0.382	0.433	0.48	0.415
DistMult	0.41	0.441	0.475	0.433
ConvE	0.403	0.452	0.519	0.438
Minerva	0.413	0.456	0.513	0.448
A2N	0.45	0.5	0.55	0.49
RAKGR	0.445	0.49	0.535	0.481

表 5 数据集 WN18RR 预测尾实体或者头实体的实验结果

模型	Hits@1	Hits@3	Hits@10	MRR
Complex	0.382	0.433	0.48	0.415
DistMult	0.41	0.441	0.475	0.433
ConvE	0.403	0.452	0.519	0.438
Minerva	0.413	0.456	0.513	0.448
A2N	0.45	0.5	0.55	0.49
RAKGR	0.445	0.49	0.535	0.481

表 2 和表 4 中提到的预测尾实体是指单独针对 $(h, r, ?)$ 这样的任务模型得到的结果, ConvKB、TransE 的源代码中并没有提供单独预测尾实体的代码, 所以本文没有展示其结果。表 3 和表 5 中的结果是指测试集中同时存在头实体缺失 $(?, r, v)$ 或者尾实体缺失 $(u, r, ?)$ 这 2 种情况。文献[21]没有提供 Minerva 在同时预测头实体或者尾实体缺失情况下的实验结果, 所以在本文的表 3 和表 5 中没有相应的结果展示。表 2 展示了 FB15k-237 在已知头实体和关系, 预测尾实体的结果。可以看出, 本文模型的所有 4 个评价指标都具有较显著的优越性。表 3 展示了在测试集中同时存在头实体缺失或者尾实体缺失的情况, 尽管这种情况下模型性能比表 2 中展示的结果稍逊色, 但本文模型性能明显优于其他模型。

表 4 和表 5 展示了 WN18RR 数据集上的实验结果。可以发现, 在 WN18RR 数据集上, 本文模型性能与 A2N 接近, 并没有绝对优势。因为 WN18RR 有 40 943 个实体, 但是仅有 11 种类型的关系。本文模型在较多关系的数据集上具有优势, 对于这一类关系较少而节点数又较多的数据集并不能很好地体现模型的优越性。

从上述结果可以看到, 所有的模型在已知头实体和链接预测尾实体情况下的实验结果普遍比同时预测尾实体或者头实体的情况好。这是因为

FB15k-237 与 WN18RR 中删除了逆关系, 这种逆关系对于已知尾实体和关系预测头实体的情况会有影响, 但对已知头实体和关系预测尾实体的情况基本没有影响。另外, 在 2 个数据集上的实验结果表明, 本文模型具有较好的稳健性, 在不同的数据集上性能稳定。

5.4 模型的输出结果测试实例

为了对层级注意力机制有更加清晰的支持和解释, 本文从 UMLS (unified medical language system) 数据集^[31]的测试集中随机选取了一个实体 n_{79} 和一个关系 r_{40} 来进行链接, 然后预测 $(n_{79}, r_{40}, ?)$, 这个测试用例在训练集中是没有出现过的。模型的输出结果测试实例如图 4 所示, 中心节点 n_{79} 的一阶邻居节点共 56 个, 关系类型共 8 种。其中, 17 个节点与 n_{79} 之间的关系是 r_{17} , 9 个节点与 n_{79} 之间的关系是 r_{10} , 21 个节点与 n_{79} 之间的关系是 r_0 。为了简化, 图 4 只表示了主要关系和节点, 邻居节点集合分别用 $N(v_{r_7})$ 、 $N(v_{r_{10}})$ 、 $N(v_{r_0})$ 表示, 且这 3 个邻居节点集合中所有节点在训练集中都没有作为 n_{90} 的头实体。将式(6)进行修改, 直接用类似 RAT 的机制计算关系注意力分数, 如式(16)所示。

$$a_{r_i} = f(v, r, u) = W[h'_v \parallel h'_r \parallel h'_u] \quad (16)$$

观察图 4 可以得到以下 2 个重要的结果。

1) 图 4(a)是用修改后的 RAT 得到的结果。注意力分数排名前三的关系分别为 r_0 (0.321 4)、 r_{10} (0.140 7)、 r_{17} (0.230 5)。 n_{79} 有 56 个一阶邻居, 该方法中, 注意力分数在 56 个三元组中进行分配, 前述注意力分数对应所有与 n_{79} 之间存在该类关系的三元组注意力分数之和; 三元组 (n_{79}, r_{43}, n_{41}) 、 (n_{79}, r_{43}, n_{39}) 、 (n_{79}, r_{43}, n_{29}) 得到的注意力分数分别是 0.020 3、0.0180 4、0.016 5。图 4(b)是利用本文的层级注意力机制得到的结果, 排名前三的关系分别为 r_{43} (0.136 6)、 r_0 (0.131 7)、 r_{38} (0.111 4)。该方法中, 注意力分数在 8 种不同的关系中进行分配。

2) 注意力模型可以加强链接预测结果的可解释性。从关系的语义可以看到, 本测试实例的关系 r_{40} (adjacent_to, 与...相邻)属于空间上相关, r_{43} (surrounds, 包围)也属于空间上相关, 两者在语义上比较接近; 排名第二的关系 r_0 (location_of,

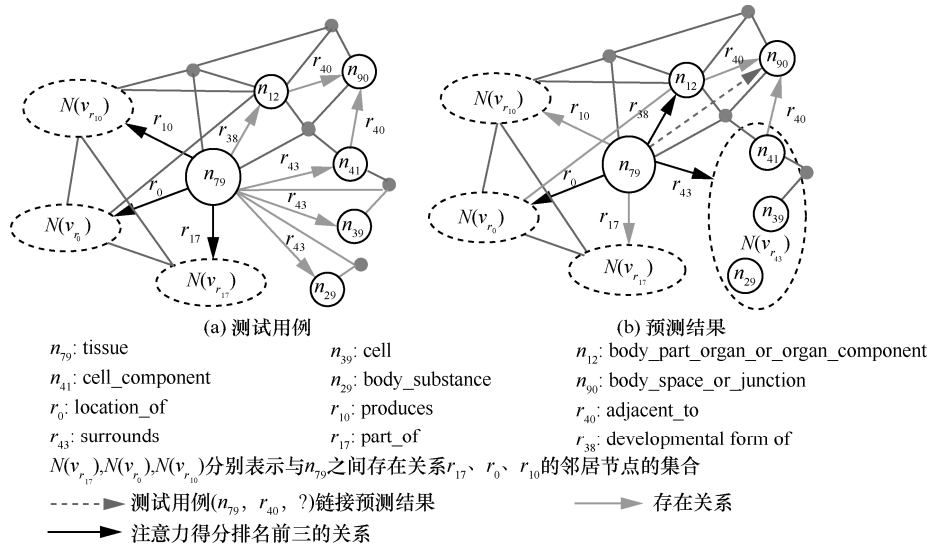


图4 模型的输出结果测试实例

位于)属于空间上相关; 排名第三的关系 r_{38} (developmental form of, ...的发育阶段)属于概念上相关。

6 结束语

本文介绍了一种基于层级注意力机制的链接预测模型, 并将其应用于知识图谱链接预测任务。所提模型取得了比目前较先进的模型更好的结果。在未来的工作中, 可以扩展所提模型, 针对不同类型的节点设计不同的线性变换矩阵, 在考虑知识图谱结构的同时, 考虑节点的语义信息。另外, 可以考虑在嵌入表示时融合知识图谱中实体和关系的相关描述信息和文本信息, 从而进一步提高知识图谱推理的性能, 而不是仅考虑知识图谱内在的信息。本文提出的模型具有高阶传播特性, 可以捕获给定实体周围多跳的关系信息, 在未来的工作中可以考虑将该模型应用在基于知识图谱的多跳推理以及基于知识图谱的复杂关系问答等应用中。

参考文献:

[1] BAO J, DUAN N, YAN Z, et al. Constraint-based question answering with knowledge graph[C]//International Conference on Computational Linguistics. Saarland: DBLP, 2016: 2503-2514.
 [2] WEST R, GABRILOVICH E, MURPHY K, et al. Knowledge base completion via search-based question answering[C]//The 23rd International Conference on World Wide Web. New York: ACM Press, 2014: 515-526.

[3] HUANG X, ZHANG J, LI D, et al. Knowledge graph embedding based question answering[C]//The Twelfth ACM International Conference. New York: ACM Press, 2019: 105-113.
 [4] ZHANG F, YUAN N J, LIAN D, et al. Collaborative knowledge base embedding for recommender systems[C]// The 22nd ACM SIGKDD International Conference. New York: ACM Press, 2016: 353-362.
 [5] 全拥, 贾焰, 张良, 等. 在线社交网络个体影响力算法测试与性能评估[J]. 通信学报, 2018, 39(10): 1-10.
 QUAN Y, JIA Y, ZHANG L, et al. Performance analysis and testing of personal influence algorithm in online social networks[J]. Journal on Communications, 2018, 39(10): 1-10.
 [6] 吴玺煜, 陈启买, 刘海, 等. 基于知识图谱表示学习的协同过滤推荐算法[J]. 计算机工程, 2018, 44(2): 226-232, 263.
 WU X Y, CHEN Q M, LIU H, et al. Collaborative filtering recommendation algorithm based on representation learning of knowledge graph[J]. Computer Engineering, 2018, 44(2): 226-232, 263.
 [7] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//The 26th International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2013: 926-934.
 [8] 黄杨琛, 贾焰, 甘亮, 等. 基于远程监督的多因子人物关系抽取模型[J]. 通信学报, 2018, 39(7): 103-112.
 HUANG Y C, JIA Y, GAN L, et al. Multi-factor person entity relation extraction model based on distant supervision[J]. Journal on Communications, 2018, 39(7): 103-112.
 [9] NIE B, SUN S. Knowledge graph embedding via reasoning over entities, relations, and text[J]. Future Generation Computer Systems, 2018, 91: 426-433.
 [10] SUCHANEK F M, KASNECI G, WEIKUM G, et al. Yago: a core of semantic knowledge[C]//The 6th International Conference on World

- Wide Web. New York: ACM Press, 2007: 697-706.
- [11] LEHMANN J, ISELE R, JAKOB M, et al. DBPedia—a large-scale, multilingual knowledge base extracted from Wikipedia[J]. Social Work, 2015, 6(2): 167-195.
- [12] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//The 2008 ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2008: 1247-1250.
- [13] 赵晓娟, 贾焰, 李爱平, 等. 多源知识融合技术研究综述[J]. 云南大学学报(自然科学版), 2020, 42(3): 459-473.
ZHAO X J, JIA Y, LI A P, et al. A survey of the research on multi-source knowledge fusion technology[J]. Journal of Yunnan University: Natural Sciences Edition, 2020, 42(3): 459-473.
- [14] KROMPAS D, BAIER S, TRESP V, et al. Type-constrained representation learning in knowledge graphs[C]//14th International Semantic Web Conference. Berlin: Springer, 2015: 640-655.
- [15] BANSAL T, JUAN D, RAVI S, et al. A2N: attending to neighbors for knowledge graph inference[C]//The 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL Press, 2019: 4387-4392.
- [16] 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.
- [17] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv Preprint, arXiv:1609.02907, 2016.
- [18] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//The 31st International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2017: 1024-1034.
- [19] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv Preprint, arXiv:1710.10903, 2017.
- [20] VASHISHTH S, SANYAL S, NITIN V, et al. Composition-based multi-relational graph convolutional networks[J]. arXiv Preprint, arXiv:1911.03082, 2019.
- [21] 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. Stroudsburg: ACL Press, 2019: 4710-4723.
- [22] DAS R, DHULIAWALA S, ZAHEER M, et al. Go for a walk and arrive at the answer: reasoning over paths in knowledge bases using reinforcement learning[J]. arXiv Preprint, arXiv:1711.05851, 2017.
- [23] WANG X, JI H, SHI C, et al. Heterogeneous graph attention network[C]//The World Wide Web Conference. New York: ACM Press, 2019: 2022-2032.
- [24] DAI Q N, TU D N, NGUYEN D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network[J]. arXiv Preprint, arXiv:1712.02121, 2017.
- [25] BORDES A, USUNIER N, GARCIADURAN A, et al. Translating embeddings for modeling multi-relational data[C]//The 26th International Conference on Neural Information Processing System. New York: ACM Press, 2013: 2787-2795.
- [26] TOUTANOVA K, CHEN D, PANTEL P, et al. Representing text for joint embedding of text and knowledge bases[C]//Empirical Methods in Natural Language Processing. [S.n.s.l.], 2015: 1499-1509.
- [27] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2D knowledge graph embeddings[J]. arXiv Preprint, arXiv:1707.01476, 2017.
- [28] BORDES A, GLOROT X, WESTON J, et al. A semantic matching energy function for learning with multi-relational data[J]. Machine Learning, 2014, 94(2): 233-259.
- [29] YANG B, YIH W, HE X, et al. Embedding entities and relations for learning and inference in knowledge bases[J]. arXiv Preprint, arXiv:1412.6575, 2014.
- [30] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[J]. arXiv Preprint, arXiv:1606.06357, 2016.
- [31] KOK S, DOMINGOS P. Statistical predicate invention[C]//The 24th International Conference on Machine Learning. New York: ACM Press, 2007: 433-440.

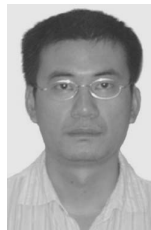
[作者简介]



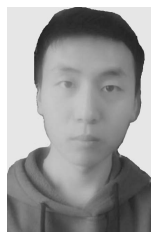
赵晓娟(1975-),女,湖南娄底人,国防科技大学博士生,主要研究方向为知识图谱表示、知识图谱推理、网络空间安全。



贾焰(1960-),女,四川成都人,博士,国防科技大学教授、博士生导师,主要研究方向为数据挖掘、大数据分析、信息安全等。



李爱平(1974-),男,山东诸城人,博士,国防科技大学研究员,主要研究方向为大数据分析、数据挖掘和网络信息安全。



陈恺(1994-),男,湖北十堰人,国防科技大学硕士生,主要研究方向为知识图谱推理、网络空间安全。