

基于多跳路径的知识库补全

王引苗 韩志敏[†]

杭州电子科技大学自动化学院人工智能研究所, 杭州 310018; [†] 通信作者, E-mail: hanzm@hdu.edu.cn

摘要 为了对知识库(KBs)进行补全, 提出一种新的基于路径的推理方法, 使用注意力机制, 将实体与其类型相结合, 共同对路径中的实体进行表示, 并使用注意力机制对每条路径预测的关系向量与给定关系的表示向量之差的绝对值进行汇总来计算模型的置信度。在基准数据集 WN18RR 和 FB15k-237 上的实验结果表明, 与现有的基于路径的关系推理方法相比, 所提方法具有更好的性能。

关键词 知识库(KBs)补全; 基于路径的推理; 注意力机制

Knowledge Bases Completion Based on Multi-hop Paths

WANG Yinmiao, HAN Zhimin[†]

Artificial Intelligence Institute, School of Automation, Hangzhou Dianzi University, Hangzhou 310018;
[†] Corresponding author, E-mail: hanzm@hdu.edu.cn

Abstract In order to complement knowledge bases (KBs), the authors propose a new path-based reasoning method, which uses the attention mechanism to combine entities and their types to represent the entities in the path and use the attention mechanism to summarize the absolute value of the difference between the relationship vector predicted by each path and the representation vector of the given relationship to calculate the confidence of the model. The results of experiment on benchmark data sets WN18RR and FB15k-237 show that the proposed model has better performance than the existing path-based relational reasoning methods.

Key words knowledge bases (KBs) completion; path-based reasoning; attention mechanism

知识库(knowledge bases, KBs)是由实体和有向的实体关系组成的网络, 以(头实体, 实体关系, 尾实体)的形式存储事实三元组。目前, 大规模的KBs(如 WordNet^[1]、Freebase^[2]和 YAGO^[3])已广泛用于不同的领域(如推荐系统^[4]和问答系统^[5])。KBs将实体作为节点, 实体关系作为有向边, 组成大规模的事实三元组。但是, 现实世界的KBs中往往存在缺失, 会对其下游任务造成影响。

基于路径的推理方法通过组合实体对之间的路径来推断实体对之间未知的关系。事实上, 通过利用KBs中两个或多个关系组成的多跳路径进行更复杂的推理, 对KBs的补全更有帮助。如图1所示, 三元组事实(奥巴马, 国籍, 美国)可通过观察KBs中“奥巴马-出生地-檀香山-位于-夏威夷-属于-美

国”这条路径来推断。

在KBs中, 任意两个实体之间都存在大量的多跳路径。因此, 在基于路径的关系推理中, 准确的推理往往来自多条路径信息的整合^[6]。图2中, 4条路径都不直接包含“史蒂夫·乔布斯的国籍是美国”这个信息, 如果综合考虑这些路径, 将从这4条路径中获得许多支持这一事实的证据。

Lao 等^[7]提出基于路径的推理方法 PRA(path ranking algorithm), 该方法将每条路径都视为一个

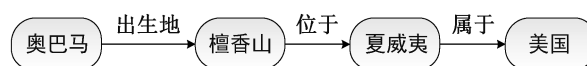


图1 多跳路径
Fig. 1 Multi-hop path

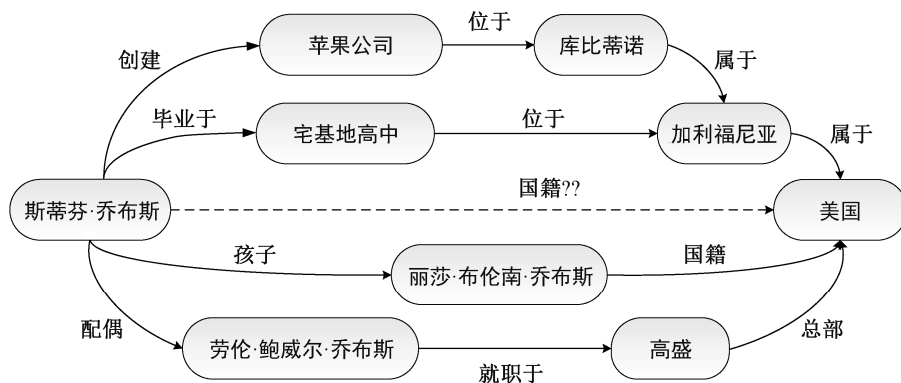


图 2 多跳路径集合^[6]
Fig. 2 Multi-hop path collection^[6]

原子特征,使得它很难适用于拥有数亿计路径的大规模KBs。为了解决这个问题,Neelakantan等^[8]提出一种基于神经网络的Path-RNN方法,但未将路径中的实体作为输入,且对每一个给定的关系单独训练一个模型,没有充分地共享模型参数。为了提高Path-RNN算法的性能,Das等^[9]进行扩展,纳入实体信息,且只训练一个RNN模型来预测所有的关系类别,不仅共享模型参数,还提高了模型精度。上述工作证明了基于路径的推理方法的有效性^[7-10],但存在以下两方面的问题。

1) 忽略路径中实体与实体类型的依赖关系。在KBs中,实体出现的频率是重尾分布。由于大多数KBs中都标注了实体的类型,对于出现频率较低的实体,很难学习到良好的实体表示。例如图1中,奥巴马有总统、政治家、律师和美国公民等实体类型。为了解决重尾分布的问题,Jiang等^[6]和Das等^[9]将实体类型映射到向量空间,并将实体对应的不同实体类型的表示向量平均值作为该实体的表示向量,Zhu等^[11]则将实体对应的不同实体类型的表示向量之和作为该实体的表示向量。在上述两项研究中,仅使用实体类型,均未利用实体类型与实体之间的相互依赖关系。

2) 计算置信度的方法没有充分地利用路径信息。在计算路径集合与实体对之间给定关系的置信度阶段,Das等^[9]采取对不同路径预测的关系向量与给定关系表示向量的内积取平均值、最大值或LSE(LogSumExp)的方法,不能有效地利用路径集合的信息。这是因为,采用最大值的方法只有一条路径用于推理,其他路径的信息被忽略;连接实体对的路径集合通常非常大,仅有几条路径可能有助于推

断。所以,取平均值的方法会使模型受到噪声影响;LSE是最大值的平滑近似值,不能有效地整合来自多条路径的信息。

针对上述问题,本文提出一个新的基于路径推理的模型——基于路径推理的注意力机制模型,将基于路径的推理转化为依据实体之间的一组路径来判断该实体对之间是否存在给定的关系。为了将KBs中缺失的关系补全,本文采用不需要逻辑规则且具有可解释性的基于路径的推理方法。

1 任务的定义

KBs被定义为一个三元组,也称关系事实,

$$X = \{(e_i, r_j, e_m) \mid e_i, e_m \in E \cap r_j \in R\},$$

其中, E 表示实体集合, R 表示关系集合。KBs也可以表示为多关系网络 G ,其中节点代表实体,边代表关系。每个三元组 (e_i, r_j, e_m) 都包含一个从 e_i 到 e_m 有向边 r_j 。 e_i 到 e_m 之间的路径用 $p = \{e_i, r_i, e_{i+1}, r_{i+1}, \dots, r_{m-1}, e_m\}$ 表示,路径的长度定义为路径中的关系数量,在路径 p 中为 $m-i$ 。对于 G 中的任一对实体 e_i 和 e_m ,可以有 N 条不同的路径,从而组成路径集合, $P_{im} = \{p_1, p_2, \dots, p_N\}$ 。本文的目标是从残缺的KBs中提取路径集合 P_{im} 来预测实体对 e_i 与 e_m 之间是否存在缺失的关系 r 。

本文模型需要基于训练数据,学习一个关系分类器,然后使用它来预测测试集中实体对之间的关系,即需要为实体对 e_i 和 e_m 给定一个关系 r ,通过计算得到实体对 e_i 和 e_m 的路径集合 P_{im} 与给定关系 r 的置信度得分,据此判断KBs中 e_i 与 e_m 之间是否存在缺失的关系 r 。

2 基于路径推理的注意力机制模型

本文提出的模型采用随机初始化的方式,将每条路径中涉及的实体、实体对应的不同实体类型和路径中包含的关系映射到向量空间,将向量作为模型的输入。在基于实体类型的注意力层中,利用实体表示向量和实体类型表示向量计算不同实体类型的权重,对不同的实体类型赋予不同的权重,以便突出同一实体的不同实体类型在不同路径或关系中的差异以及个别实体类型对关系推理的重要性。在计算模型置信度的阶段使用注意力机制,通过对每条路径预测的关系向量与给定关系的表示向量之差的绝对值进行汇总来计算模型的置信度。使用注意力机制,将每条路径上产生的误差绝对值进行加权组合,从而能够充分地利用多条路径的聚合信息。并且,注意机制可以为每条路径自动地分配权重。使用向量差的绝对值体现了预测关系向量与给定关系的表示向量在向量空间中每个维度的差异,便于后续对差异进行减小和优化。

本文提出的基于路径推理的注意力机制模型将实体 e_i 与 e_m 之间的路径集合 $P_{im} = \{p_1, p_2, \dots, p_N\}$ 作为输入,并输出给定关系 r 连接 e_i 和 e_m 的置信度 $P(r|e_i, e_m)$ 。如图3所示,本文模型由基于实体类型的注意力层、实体编码层、关系编码层和基于关系路径的注意力层4层组成。

2.1 基于实体类型的注意力层

给定实体对 e_i 与 e_m 之间的一条路径 $p_1 = \{e_i, r_i, e_{i+1}, r_{i+1}, \dots, r_{m-1}, e_m\} \in P_{im}$ 和路径中实体对应的实体类型集合 $\{\tau_i, \dots, \tau_{m-1}, \tau_m\}$, 其中每个实体对应 n 个实体类型,即 $\tau_k = \{t_{k,1}, \dots, t_{k,n-1}, t_{k,n}\}$ 。基于实体类型的

注意力机制将实体 $\{e_i, \dots, e_{m-1}, e_m\}$ 和对应的实体类型 $\{\tau_i, \dots, \tau_{m-1}, \tau_m\}$ 作为输入,对输入进行编码,得到实体的最终表示 $[E_i, \dots, E_{m-1}, E_m]$ 。为突出部分实体类型对实体间关系的主导作用,本文用注意力机制给不同实体类型分配不同的权重。实体 e_i 的第 o 个实体类型 $t_{i,o}$ 在该层所占权重 $\alpha_{i,o}$ 的计算方法如下:

$$\alpha_{i,o} = \frac{\exp(z_{i,o})}{\sum_{k=1}^n \exp(z_{i,k})}, \quad (1)$$

其中,

$$z_{i,o} = f(V(t_{i,o}) + V(e_i)), \quad (2)$$

其中, $z_{i,o}$ 和 $z_{i,k}$ 是中间变量, $z_{i,o}$ 将 $V(t_{i,o})$ 与 $V(e_i)$ 关联, $z_{i,k}$ 将 $V(t_{i,k})$ 与 $V(e_i)$ 关联; $f(\cdot)$ 表示前馈网络; $V(t_{i,o})$ 为实体 e_i 的第 o 个类型 $t_{i,o}$ 的表示向量; $V(e_i)$ 为实体 e_i 的表示向量。利用式(1)和(2)得到不同实体类型的权重后,为了规避实体的重尾分布问题,本文利用实体类型表示向量来获得实体 e_i 的最终表示 E_i :

$$E_i = \sum_{o=1}^n \alpha_{i,o} * V(t_{i,o}), \quad (3)$$

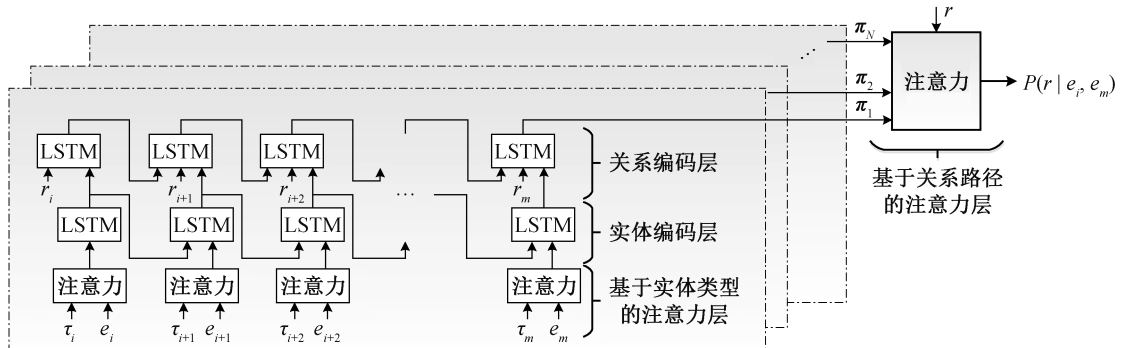
其中, * 表示 Hadamard 积。

2.2 实体编码层

为了充分利用实体的上下文信息,实体编码层将实体的最终表示 $[E_i, \dots, E_{m-1}, E_m]$ 作为本层的输入,使用能缓解 RNN 长期依赖问题的 LSTM^[12] 对 $[E_i, \dots, E_{m-1}, E_m]$ 进行编码,将 LSTM 每个时间步的隐藏状态 $[H_i, \dots, H_{m-1}, H_m]$ 作为本层的输出。

2.3 关系编码层

为了提高算法的性能,将实体编码层的输出 $[H_i, \dots, H_{m-1}, H_m]$ 与关系表示向量 $[V(r_i), \dots, V(r_{m-1})]$,



e_i 代表路径中的实体, τ_i 代表实体 e_i 对应的实体类型, r_i 代表路径中的关系, r 代表给定关系, π_1 代表路径 p_1 预测实体 e_i 与 e_m 之间的关系向量, $P(r|e_i, e_m)$ 代表路径集合 P_{im} 与给定关系 r 的置信度

图3 基于路径推理的注意力机制模型

Fig. 3 Attention mechanism model based on path reasoning

$V(r_m)$ 按照时间步进行拼接,得到本层的输入 $[[H_i; V(r_i)], \dots, [H_{m-1}; V(r_{m-1})], [H_m; V(r_m)]]$ 。为了使路径中的实体数量与关系数量相匹配,为每条路径增加相同的路径结束标志向量 $V(r_m)$ 。对拼接之后的向量,通过 LSTM 进行编码,将 LSTM 最后一个时间步的隐藏状态作为路径 p_1 提供给实体对 e_i 与 e_m 之间关系的预测向量 π_1 。

2.4 基于关系路径的注意力层

使用上述方法对路径集合 P_{im} 中剩余的 $N-1$ 条路径进行编码,最终得到预测关系向量集合 $\Pi=[\pi_1, \dots, \pi_{N-1}, \pi_N]$ 。同时,使用一个可训练的向量 u 来表示 e_i 与 e_m 之间给定的关系 r ,使用向量 u 与每条路径的预测关系向量 π_i 之差的绝对值 $\Delta_i=|u-\pi_i|$ 来表示模型在每条路径上产生的误差。使用类似实体类型的注意力层,对每条路径上产生的误差进行汇总,并将汇总后的误差转化为对应的置信度 $P(r|e_i, e_m)$,计算方式如下:

$$P(r|e_i, e_m) = \delta(f(\boldsymbol{\varepsilon}(P_{im}, r))), \quad (4)$$

$$\boldsymbol{\varepsilon}(P_{im}, r) = \sum_{o=1}^N \beta_o * \Delta_o, \quad (5)$$

$$\beta_i = \frac{\exp(h_i)}{\sum_{k=1}^N \exp(h_k)}, \quad (6)$$

$$h_i = f(\pi_i + u), \quad (7)$$

其中, $\delta(\cdot)$ 表示 sigmoid 函数, β_i 代表注意力机制在每条路径上分配的权重, $\boldsymbol{\varepsilon}(P_{im}, r)$ 代表路径集合 P_{im} 与给定关系 r 之间的误差总和。

基于路径推理的注意力机制模型未使用外部工具产生的特征,因此可以进行端到端的训练。为了预测每个关系,本文使用训练集中的真假三元组作为正例($(e_i, r, e_m) \in D_r^{\text{Train}+}$)和负例($(e_i, r, e_m) \in D_r^{\text{Train}-}$)来训练模型。使用交叉熵损失函数对模型进行优化,模型的损失计算方式如下:

$$L = -\sum_{(e_i, r, e_m) \in D_r^{\text{Train}+}} \log(P(r|e_i, e_m)) + \sum_{(e_i, r, e_m) \in D_r^{\text{Train}-}} \log(P(r|e_i, e_m)) \quad (8)$$

3 实验

3.1 数据集及预处理

在 FB15k-237 和 WN18RR 两个基准数据集上评估本文方法与基线方法。首者是常识 KBs Freebase^[2]的子集,后者则是英语词汇数据库 WordNet^[11]的子集。表 1 列出两个数据集的统计数据。

参照 Liu 等^[10]的方式处理数据。从每个数据集的所有真三元组 X 中建立一个完整的实验数据集:

$$D = \{(e_i, r_j, e_m) | e_i, e_m \in E \cap r_j \in R \cap y \in \{0, 1\}\},$$

y 表示三元组是真还是假。该数据集包含额外的假三元组,这些假三元组是基于个性化 PageRank 算法,从 X 中采样得到。负采样是必要的,因为两个基准数据集中只包含真三元组,负例的数量对算法的性能有显著的影响^[13]。本文为每个正例抽取 10 个负例,并将数据集分为 80% 的训练集和 20% 的测试集。因为基于路径的方法通常分别对每个关系进行建模,所以会根据关系对数据做进一步的划分。

为了提取实体之间的路径,本文用数据集中的真三元组构建 KBs。按照 Lao 等^[7]的方法,用反向关系扩充 KBs。使用双向广度优先算法(breadth first search, BFS)提取路径。实验中,最大路径长度的确定与数据集中三元组的连接紧密程度相关。对多跳路径来说,最大路径长度越大,在给定一种关系的情况下实体对之间的路径数量越多,但无效路径被抽取的概率也越大,不仅给实验带来噪声,而且占用的空间和消耗的时间更多。本文根据 Liu 等^[10]对最大路径的选择,在 WN18RR 数据集中,最大路径长度设置为 6;在连接更密集的 FB15k-237 数据集中,最大路径长度设置为 4。这样,不仅能提供足够的路径用于训练,而且消耗的资源较少。为每对实体随机抽取 200 条路径(即 $N=200$)。在 WN18RR 数据集中,使用 WordNet^[11]中的实体上位词作为实体类型;在 FB15k-237 数据集中,使用 Xie 等^[14]的实体类型数据。然后,遵循 Das 等^[9]的方法,为

表 1 数据集 FB15k-237 和 WN18RR 的统计数据
Table 1 Statistics for the data set of FB15k-237 and WN18RR

数据集	数量								路径长度		
	关系	实体	三元组	被测试的关系	实体类型	每个被测试关系的平均实体对		每个实体拥有的实体类型		最大值	平均值
						训练集	测试集	最大值	平均值		
WN18RR	11	40943	134720	9	8092	38039	11888	14	4.6	6	5.3
FB15k-237	235	13545	254290	10	1029	16696	5219	7	6.4	4	3.5

FB15k-237 数据集中每个实体选择 7 种最常见的类型。对于 WN18RR 数据集, 对每个实体使用全部的上位词, 还使用预先训练的 Google News word2vec 模型, 将实体类型映射到向量中。

3.2 实验参数及细节

本文使用反向传播算法来更新可训练的模型参数。本文模型中有如下超参数: 1) 实体类型表示向量 $V(t_{i,o})$, 其维度在 FB15k-237 中为 50, WN18RR 中为 300; 2) 实体表示向量 $V(e_i)$, 其维度为 50; 3) 关系表示向量 $V(r_i)$, 其维度为 50; 4) 实体对之间给定关系的表示向量 u , 其维度为 100; 5) 关系编码层和实体编码层中两个 LSTMs 的维度均为 150。在前馈网络 $f(\cdot)$ 中使用的激活函数是 Relu 函数。对所有可训练的参数采用随机初始化的方式赋初值, 用 Adam 优化器^[15]对模型进行优化。将 MAP 作为模型的评价标准(AP 是针对一种关系中所有实体对的平均精度, MAP 是针对所有关系的平均 AP)。实验中还将早期停止(early stopping)作为正则化手段, 防止模型过拟合。3.3 节和 3.4 节的实验都进行 5 次, 取 5 次实验结果的平均值作为最终结果。

3.3 验证基于实体类型的注意力层的有效性

本文使用实体类型表示向量的平均值与实体类型表示向量之和来获取实体表示, 得到表 2 中的结果。可以看出使用均值或者和的方式取得的 MAP 值都低于使用注意力机制, 说明使用注意力机制可以对实体进行更好的表示, 不仅可以增强实体与其类型的依赖关系, 还能提高实验结果的精度。

3.4 验证基于关系路径的注意力层的有效性

本文采取对每条路径预测的关系向量与给定关系向量的内积取平均值、最大值或 LSE 的方法获取置信度, 结果见表 3。可以看出, 基于关系路径的注意力层取得更好的实验结果。这是因为在汇总多条路径信息的过程中, 注意力机制不只依赖权重最

表 2 实体向量的不同获取方式对实验结果的影响
Table 2 Effect of different acquisition methods of entity vectors on experimental results

实体向量的获取方式	MAP/%	
	WN18RR	FB15k-237
Avg(type)	78.2	55.9
Sum(type)	79.0	57.1
Attention(type, entity)	79.6	57.4

说明: Avg(type)代表实体类型表示向量的平均值, Sum(type)代表实体类型表示向量之和, Attention(type, entity)表示基于实体类型的注意力层; 粗体数字表示本文方法的结果更好, 下同。

大的路径, 或对所有路径分配相同的权重, 而是综合全部路径的信息, 并使用每条路径预测的关系向量与给定关系的表示向量之差的绝对值来计算每条路径的权重, 会便于后续对差异进行减小和优化。

3.5 与基线模型对比

将本文提出的基于路径推理的注意力机制模型与以下基线模型进行比较, 在相同的数据集上获得的结果如表 4 所示。

1) PRA^[7]将路径作为特征, 输入二进制对数线性分类器中, 判断实体对间是否存在给定的关系。

2) SFE^[16]首先执行广度优先算法, 抽取每个实体周围的子图, 再对子图进行特征抽取, 得到特征向量, 进而实现 KBs 补全。

3) Path-RNN₁^[8]通过 PRA 模型获取需要的路径, 使用 RNN 对路径进行编码, 进而实现 KBs 补全。

4) Path-RNN₂^[9]和 Path-RNN₃^[9]使用 RNN 对路径进行编码, 使用注意力机制对路径集合进行汇总, 完成 KBs 的补全。其中, Path-RNN₂ 和 Path-RNN₃ 使用的数据类型不同。

从表 4 中 Path-RNN₂ 与 Path-RNN₃ 的对比实验结果可以看出, 添加实体信息有助于更准确地预测

表 3 不同置信度的计算方式对实验结果的影响
Table 3 Influence of different calculation methods of confidence on experimental results

置信度计算方式	MAP/%	
	WN18RR	FB15k-237
Avg($P_{im,r}$)	71.0	56.6
Max($P_{im,r}$)	69.4	53.6
LSE($P_{im,r}$)	68.2	51.4
Attention($P_{im,r}$)	79.6	57.4

说明: Avg($P_{im,r}$), Max($P_{im,r}$)和 LSE($P_{im,r}$)分别代表对每条路径预测的关系向量和给定关系向量的内积取平均值、最大值和 LSE, Attention($P_{im,r}$)代表基于关系路径的注意力层。

表 4 与基线模型的比较
Table 4 Comparison with baseline model

模型	所用数据	MAP/%	
		WN18RR	FB15k-237
PRA	关系	38.8	34.3
SFE	关系	30.7	36.7
Path-RNN1	关系	67.2	45.6
Path-RNN2	关系, 实体类型	50.8	51.2
Path-RNN3	关系, 实体, 实体类型	51.0	52.1
本文模型	关系, 实体	71.3	54.4
	关系, 实体, 实体类型	79.6	57.4

实体关系。本文模型同样添加了实体信息,并在路径集合上使用注意力机制,效果大大好于 Path-RNN₃,证明在实体类型上使用注意力机制能够增强实体与其类型的依赖关系,进而增强路径的表示,提高了实验结果的精度。在本文提出的模型中,使用实体类型与不使用实体类型取得的结果有所差距,证明实体类型与实体的依赖关系对模型性能的提升有益。与以上基线模型相比,本文模型取得最好的结果,尤其在 WN18RR 数据集上,MAP 提升 18%,说明本文提出的基于实体类型的注意力层,使用注意力机制对每条路径预测的关系向量与给定关系的表示向量之差的绝对值进行汇总来计算模型置信度的方法是行之有效的。

4 结论

本文实现基于路径的关系推理,为 KBs 补全提供一种有效的新方法。本文引入基于实体类型的注意力层,利用实体类型与实体的相互依赖关系,对实体间关系的推理起到不可或缺的作用。本文提出计算模型置信度的新方法,不仅使模型的整体效果大大提升,还便于误差的减小和优化。在 WN18RR 和 FB15k-237 两个基准数据集上的实验结果表明,本文方法比基于路径排序的方法有显著的改进。

参考文献

- [1] Miller G A. WordNet: a lexical database for English. *Communications of the ACM*, 1995, 38(11): 39–41
- [2] Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge // *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. Vancouver, 2008: 1247–1250
- [3] Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge // *Proceedings of the 16th International Conference on World Wide Web*. Banff, 2007: 697–706
- [4] Wang Xiang, Wang Dingxian, Xu Canran, et al. Explainable reasoning over knowledge graphs for recommendation // *Proceedings of the AAAI Conference on Artificial Intelligence*. Hawaii, 2019: 5329–5336
- [5] 胡月, 周光有. 基于 graph transformer 的知识库问题生成 // 第 19 届中国计算语言学大会. 海口, 2020: 324–335
- [6] Jiang Xiaotian, Wang Quan, Qi Baoyuan, et al. Attentive path combination for knowledge graph completion // *Proceedings of the 9th Asian Conference on Machine Learning*. Seoul, 2017: 590–605
- [7] Lao N, Mitchell T, Cohen W W. Random walk inference and learning in a large scale knowledge base // *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Edinburgh, 2011: 529–539
- [8] Neelakantan A, Roth B, Mccallum A. Compositional vector space models for knowledge base completion // *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*. Beijing, 2015: 156–166
- [9] Das R, Neelakantan A, Belanger D, et al. Chains of reasoning over entities, relations, and text using recurrent neural networks // *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia, 2017: 132–141
- [10] Liu Weiyu, Daruna A A, Kira Z, et al. Path ranking with attention to type hierarchies // *Proceedings of the AAAI Conference on Artificial Intelligence*. New York, 2020: 2893–2900
- [11] Zhu Qiannan, Zhou Xiaofei, Tan Jianlong, et al. Knowledge base reasoning with convolutional-based recurrent neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(5): 2015–2028
- [12] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780
- [13] Kadlec R, Bajgar O, Kleindienst J. Knowledge base completion: baselines strike back // *Proceedings of the 2nd Workshop on Representation Learning for NLP*. Vancouver, 2017: 69–74
- [14] Xie Ruobing, Liu Zhiyuan, Sun Maosong. Representation learning of knowledge graphs with hierarchical types // *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. New York, 2016: 2965–2971
- [15] Kingma D, Ba J. Adam: a method for stochastic optimization [EB/OL]. (2014–12–01) [2021–05–30]. <https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6980K>
- [16] Gardner M, Mitchell T. Efficient and expressive knowledge base completion using subgraph feature extraction // *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, 2015: 1488–1498