

融合HousE和注意力机制的知识推理模型

朱玉亮, 刘俊涛, 饶子昀, 张毅, 曹万华

引用本文

朱玉亮, 刘俊涛, 饶子昀, 张毅, 曹万华. [融合HousE和注意力机制的知识推理模型](#) [J]. 计算机科学, 2024, 51(6A): 230600209-8.

ZHU Yuliang, LIU Juntao, RAO Ziyun, ZHANG Yi, CAO Wanhua. [Knowledge Reasoning Model Combining HousE with Attention Mechanism](#) [J]. Computer Science, 2024, 51(6A): 230600209-8.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于SAMNV3的滚动轴承智能故障诊断方法](#)

Intelligent Fault Diagnosis Method for Rolling Bearing Based on SAMNV3

计算机科学, 2024, 51(6A): 230700167-6. <https://doi.org/10.11896/jsjx.230700167>

[基于LSTM和注意力机制的远程会诊需求预测](#)

Forecasting Teleconsultation Demand Based on LSTM and Attention Mechanism

计算机科学, 2024, 51(6A): 230800119-7. <https://doi.org/10.11896/jsjx.230800119>

[融入类别标签和主题信息的用户兴趣识别方法](#)

User Interest Recognition Method Incorporating Category Labels and Topic Information

计算机科学, 2024, 51(6A): 230500169-8. <https://doi.org/10.11896/jsjx.230500169>

[基于注意力的多尺度蒸馏异常检测](#)

Attention-based Multi-scale Distillation Anomaly Detection

计算机科学, 2024, 51(6A): 230300223-11. <https://doi.org/10.11896/jsjx.230300223>

[基于改进Deeplabv3+算法的滚珠丝杠驱动表面点蚀缺陷检测](#)

Detection of Pitting Defects on the Surface of Ball Screw Drive Based on Improved Deeplabv3+ Algorithm

计算机科学, 2024, 51(6A): 240200058-6. <https://doi.org/10.11896/jsjx.240200058>

融合 HousE 和注意力机制的知识推理模型

朱玉亮 刘俊涛 饶子昀 张毅 曹万华

武汉数字工程研究所 武汉 430205

(jayyl709@163.com)

摘要 知识推理技术是解决知识图谱缺失问题所提出的方法,并在近年来不断发展。为了解决推理中准确度低、可解释性差、适用性不强等问题,提出了一种融合注意力机制和 HousE 的知识推理模型 Att-HousE。该模型由一个带注意力机制的规则生成器和一个带 HousE 嵌入的规则预测器组成,规则生成器生成推理需要的规则并传入预测器,预测器更新并得到不同规则的得分,然后通过 EM 算法不断训练优化生成器与预测器。具体而言,该模型是建立在 RNNLogic 的基础上并作出改进,注意力机制可以选取更值得关注的关系作为规则,提高了模型准确度,HousE 嵌入则在处理复杂关系上更具有灵活性,并适用于建立多边关系。在公开实验数据集上的结果表明,Att-HousE 在 FB15K-237 上做推理任务时,MRR 指标整体比 RNNLogic 高出 6.3%;在稀疏数据集 WN18RR 上,Hits@10 指标整体比 RNNLogic 高出 2.7%,证明了引入 HousE 和注意力机制后可以更全面地抓取和形成多边关系,提升知识推理的精度。

关键词: 知识图谱补全;知识推理;注意力机制;知识表示;EM 算法

中图分类号 TP391

Knowledge Reasoning Model Combining HousE with Attention Mechanism

ZHU Yuliang, LIU Juntao, RAO Ziyun, ZHANG Yi and CAO Wanhua

Wuhan Digital Engineering Institute, Wuhan 430205, China

Abstract Knowledge reasoning technology is a method proposed to solve the problem of missing knowledge graphs and has been continuously developed in recent years. In order to solve the problems of low accuracy, poor interpretability, and weak applicability in knowledge reasoning, a knowledge reasoning model called Att-HousE, which combines HousE with Attention Mechanism, is proposed. It consists of a rule generator with attention mechanism and a rule predictor with HousE. The rule generator generates the rules required for reasoning and passes them into the predictor, which updates and then obtains scores for different rules. After that, the generator and predictor are continuously trained and optimized by the EM algorithm. Specifically, the model is based on RNNLogic and has been improved. The attention mechanism can select more noteworthy relationships as rules, improving the accuracy of the model. HousE has more flexibility in handling complex relationships and is suitable for establishing multilateral relationships. According to experimental results on public datasets, it indicates that Att-HousE's MRR is 6.3% higher than that of RNNLogic when doing reasoning tasks on FB15K-237. For the sparse dataset WN18RR, the Hits@10 of Att-HousE is 2.7% higher than that of RNNLogic. It is demonstrated that the introduction of HousE and attention mechanism can more comprehensively grasp and form multilateral relationships, which can improve the accuracy of knowledge reasoning.

Keywords Knowledge graph completion, Knowledge reasoning, Attention mechanism, Knowledge representation, EM algorithm

1 引言

随着知识图谱技术的发展,越来越多的知识图谱技术在各行各业中得到应用。知识图谱的本质^[1]是形为(实体,关系,实体)的三元组集合,在此基础上,构建了许多不同种类的知识图谱,如 Freebase^[2], WordNet^[3], NLL^[4]等。但是无论是人工创建还是自然生成的知识图谱,由于不可能收集所有的事实,因此都存在数据不完善的问题^[5]。所以现在知识图谱技术主要解决的问题是知识图谱的推理补全的问题^[6],即发现知识图谱中缺失的实体以及关系。

事实上,大部分知识推理的补全问题通过路径推理和知识表示学习其中的逻辑规则然后进行补全。现有大多数传统方法,如路径排序^[7]和马尔可夫逻辑网络^[8],都是将知识图谱上的关系路径当作候选逻辑规则,然后学习每个规则的权重,作为规则质量的评估。最近还有一些基于神经逻辑编程^[9]的方法和神经定理证明器^[10],它们能够以可微的方式同时学习逻辑规则及其权重。尽管这些方法在实验中对预测有效,但其指数级的搜索复杂度导致开销过大,因此识别高质量的逻辑规则需要进行大量训练。此外,还有通过知识表示将实体和关系映射到向量空间进行推理的方法。例如,TransE^[11]

基金项目:十四五装备预先研究项目(50902010503)

This work was supported by the 14th Five Year Equipment Pre-research Project(50902010503).

通信作者:刘俊涛(570710065@qq.com)

试图将关系向量表示为空间上的平移,但却无法表示关系的对称性。RotatE^[12]则将关系表示为复平面中的旋转以建模4个关系模式,但由于旋转距离的限制,它无法处理一对多及多对一的关系,因此该方法在复杂关系的推理中表现不尽人意。

为此,本文提出了一种融合知识表示与注意力机制的创新方法 Att-HousE。首先,将知识图谱中的三元组信息通过知识表示方法映射到高维向量空间,充分保留其语义信息,同时引入EM算法充分挖掘知识图谱中的有效规则。具体来说,它包括一个基于RNN和注意力机制的规则生成器和一个推理规则的规则预测器,它们通过EM算法同时被训练以相互增强。规则生成器为规则预测器提供推理所需的逻辑规则,其中规则的产生引入注意力机制以抓取更有效的规则,而规则预测器为规则生成器的训练提供有效的奖励。此方法因为对知识图谱逻辑规则产生的路径进行了调节,所以有助于减小搜索空间。

同时,在知识图谱的嵌入阶段,即规则预测器进行推理时,受Li等^[33]的启发,采用HousE学习实体和关系的高维表示。该嵌入可以建模任意关系模式,并且在多边关系的处理上也取得了不错的效果。实验结果显示,本文方法在FB15K-237、WN18RR数据集上都优于现有的部分知识表示和传统深度学习的算法,提高了模型可解释性。

本文的主要贡献为:(1)提出了一种基于EM算法的RNN网络处理模块,同时融合注意力机制,抓取更值得注意的规则,提高了模型训练的准确度;(2)在知识推理的融合学习方面提供了新的方法,引入新的知识表示模块HousE,解决了知识推理中建模多边关系以及可解释性的问题。

本文第2章调查了相关工作;第3章提出Att-HousE并解释其具体细节;第4章介绍了实验部分并进行对比与分析;最后总结全文并展望未来。

2 相关工作

知识图谱的实体和关系缺失问题普遍存在,而知识推理是解决缺失问题的主要手段。一般说来,推理补全方法分为3类,分别是:基于规则学习的推理算法、基于知识表示的推理算法和融合知识表示和规则学习的算法。本章将依次介绍这3种方法。

2.1 基于规则学习的推理算法

规则学习是一种从样本数据中自动学习规则的方法,广泛应用于数据挖掘、机器学习、自然语言处理等领域。近年来,将规则学习与知识推理相结合的研究成为处理知识缺失问题的热点,在此方面取得了许多进展。

此类方法都是基于知识图谱的结构特性进行推理,因为知识图谱中的每个实体都可以通过其他实体依靠关系路径相连接,所以关系的推理就是通过枚举查询实体和答案实体之间的关系路径作为候选逻辑规则,并进一步学习每个规则的标量权重来评估质量,再通过规则得到缺失的实体和关系。例如,Tushar等^[14]使用马尔可夫逻辑网络,该方法在函数梯度提升的基础上同时学习规则的权重和结构,使学习规则问题被转化为一组关系函数近似问题,从而生成更加精确和可解释的规则。Christian等^[15]提出了一种自下而上的基于

强化学习的规则学习方法,用于从大型的知识图谱中学习逻辑规则。Lao等^[7]使用路径排名算法来学习路径中不同约束下规则的加权组合,该方法能够发现和使用关系型数据的复杂路径特征。最近还有一些方法通过以可微的方式同时学习逻辑规则和权重来扩展这个领域,如Minervini^[16]等提出了一种可以自动完成的端到端可微证明器CTP,该证明器对子符号表示进行操作。所构建的神经网络能够计算关于符号的矢量表示,从而对知识图谱进行端到端的子符号链接训练。然而,现有的这些方法试图同时学习逻辑规则及其权重,因为权重的优化与规则的更新存在滞后性,所以单纯的神经网络在两者的优化方面难以同时进行。

2.2 基于知识表示的推理算法

表示学习指从数据中学习高维特征表示的一类方法,通过学习高维空间中已知不同实体嵌入的距离来优化嵌入模型,再根据嵌入模型寻找最近的实体嵌入解决知识图谱的推理问题。自word embedding^[17]这个概念被提出后,在知识三元组的表示学习上出现了许多新的方法。Trouillon等^[18]在处理复杂嵌入的关系时使用埃尔米特点积,解决了标准点积无法反应对称与反对称的问题。Bordes等^[11]提出TransE,将关系解释为对实体的低维嵌入,并以此来对关系进行建模并学习一些简单的组合规则。Sun等^[12]提出RotatE,它通过将每个关系定义为复杂向量空间中从源实体到目标实体的旋转,能够建模和推断对称/反对称、反转和合成在内的各种关系模式。Zhang等^[19]则在RotatE基础上提出超复数的定义,可以恰当地捕捉到潜在的相互依赖关系,从而鼓励更高效、更具表现力的表示学习过程。Li等^[13]则提出了一个更强大的KGE框架,名为HousE,它引入了一种基于Householder变换的新参数,以实现关系模式全面建模的能力。但上述方法都存在以下问题:(1)将知识推理转化为单一的向量计算,带来了可解释性差的问题;(2)未能充分利用关系路径等重要信息,推理能力受限,推理准确率尚有较大的提升空间。

2.3 融合知识表示和规则学习的算法

以上两种方法在知识推理方面的短板:一方面只能推理简单路径中的规则;另一方面无法兼顾参数与规则的同时学习。如何平衡这两方面是当下知识推理研究的重点之一。Xiong等^[20]设计了DeepPath,将知识推理过程转化为马尔可夫序列决策过程,以实体集合作为状态空间,以关系集合作为动作空间,智能体通过选择最优动作以拓展其路径来实现知识库中的推理。Lin等^[21]提出了PTransE,该模型使用语义连接算法表示路径关系,将关系路径视为实体表示学习时的一种参数,设计了一种路径约束资源分配算法来衡量关系路径的可靠性,并通过关系嵌入的语义组合来表示关系路径,从而显著提高推理能力。Li等^[22]提出一种基于教育词典的微调双向编码器表示模型,称为EduBERT,它可以通过注意力机制自适应地捕捉教育领域的有效信息,从而提供更好的教育服务。Qu等^[23]提出RNNLogic,该模型将逻辑规则视为潜在变量,并同时使用规则生成器和预测器来训练模型。但上述融合模型由于侧重于提高推理准确率,引入的参数过多,模型的复杂度较高,推理效率较低。

2.4 RNNLogic 框架

本节将介绍改进前的模型RNNLogic,它在之前的工作

中被用于学习知识图谱推理的逻辑规则并进行推理。该模型分别由规则生成器、规则预测器和 EM 算法组成,因此本节将分 3 部分来详细介绍。

2.4.1 问题概率化

假设 $p(G, q, a)$ 是我们用到的训练数据,其中 G 是由一组三元组 (h, r, t) 所表示的知识图谱,我们也可以将其写为 $r(h, t)$,其中 $q = (h, r, ?)$ 是一个查询, $a = t$ 是答案。给定 G 和查询 q ,目标是预测正确答案 a 。更直观地说,我们的目标是对概率分布 $P(a|G, q)$ 进行建模。

首先以概率的方式形式化知识推理,其中逻辑规则集 z 被视为潜在变量,规则集 z 中的规则由知识图谱中的部分关系组成,如其中一条规则 $play \leftarrow friend \wedge play$,意味着由 friend 和 play 可以继续延伸推导出 play 这个关系,这里的 friend 和 play 都是知识图谱中已知的关系。目标概率分布 $P(a|G, q)$ 将由规则生成器和规则预测器联合建模。规则生成器 $P(\theta(z|q))$ 定义了以查询 q 为条件,生成规则集 z 的先验概率分布,而规则预测器 $P_w(a|G, q, z)$ 则给出了以逻辑规则集 z 、查询 q 和知识图 G 为条件的答案 a 的后验分布。 θ 和 w 分别是生成器和预测器的参数,因此 $P_{w,\theta}(a|G, q)$ 的计算式如式(1)所示:

$$\begin{aligned} p_{w,\theta}(a|G, q) &= \sum_z p_w(a|G, q, z) p_\theta(z|q) \\ &= \mathbb{E}_{p_\theta(z|q)} [p_w(a|G, q, z)] \end{aligned} \quad (1)$$

本文目标是联合训练规则生成器和规则预测器,以最大限度地提高训练数据的后验估计。形式上,训练的目标函数呈现如下:

$$\begin{aligned} \max_{\theta, w} \ell(\theta, w) &= \mathbb{E}_{(G, q, a) \sim p_{\text{data}}} [\log p_{w,\theta}(a|G, q)] \\ &= \mathbb{E}_{(G, q, a) \sim p_{\text{data}}} [\log E_{p_\theta(z|q)} [p_w(a|G, q, z)]] \end{aligned} \quad (2)$$

2.4.2 规则生成器

对于查询 q ,规则生成器旨在生成用于回答查询的潜在逻辑规则集 z 。形式上,给定查询 $q = (h, r, ?)$,通过仅考虑关系 r 而不考虑头实体 h 来生成组合逻辑规则,这允许所生成的规则在多个实体之间进行推广。因此我们可以通过递归神经网络(RNN)有效地建模组成关系 r 的规则,并以此来参数化规则生成器。给定一个查询关系 r ,RNN 顺序地生成规则体中的每个关系,直到它到达规则的结尾。在该过程中,同时计算所生成的规则的先验概率。利用这样的先验概率,我们将规则集 z 的分布定义为如下多项式分布:

$$p_\theta(z|q) = \text{Mu}(z|N, \text{RNN}_\theta(\cdot|r)) \quad (3)$$

其中, Mu 代表多项式分布, $\text{RNN}_\theta(\cdot|r)$ 是生成规则的神经网络, θ 是生成器的超参数, N 为规则集中规则的个数, RNN 生成的规则 S_i 以关系 r 为规则头, $i \in N, S_i \in z$ 。规则集 z 的生成过程非常直观,只需用 RNN 即可生成 N 条规则 S_i 来形成规则集 z 。

2.4.3 规则预测器

接下来引入规则预测器,其作用是计算出使用规则后预测出的候选答案的后验概率。首先规则预测器定义答案 a 的分布为 $P_w(a|G, q, z)$,对于查询 q ,预测器使用生成器生成的规则 z 在知识图 G 上推理并预测答案 a , w 为预测器的超参数。

假设对于每个查询 $q = (h, r, ?)$,同样的规则能够在图 G

上找到导致不同候选答案的不同路径。例如,给定查询 $q = (\text{Wang}, \text{play}, ?)$,那么根据规则集 z 中的一条规则 $play \leftarrow \text{friend} \wedge play$,可以从知识图谱中推导出两个不同答案: $play(\text{Wang}, \text{Basketball}) \leftarrow play(\text{Wang}, \text{Li}) \wedge play(\text{Li}, \text{Basketball})$ 和 $play(\text{Wang}, \text{Football}) \leftarrow play(\text{Wang}, \text{Xu}) \wedge play(\text{Xu}, \text{Football})$,因此产生两个候选答案 Basketball 和 Football。

设集合 A 是由集合 z 中的逻辑规则推导出的候选答案的集合。对于每个候选答案 $e \in A$,则计算不同候选答案 e 的得分为 $score_w(e)$:

$$\begin{aligned} score_w(e) &= \sum_{rule \in z} score_w(e|rule) \\ &= \sum_{rule \in z} \sum_{path \in \mathcal{P}(h, rule, e)} \psi_w(rule) \cdot \varphi_w(path) \end{aligned} \quad (4)$$

其中, $\mathcal{P}(h, rule, e)$ 是以 h 开始且在 e 处结束的不同路径的集合, $\psi_w(rule)$ 和 $\varphi_w(path)$ 是每个规则和路径的权重。直观地说,每条候选答案 e 的得分是不同规则贡献的得分之和,即 $score_w(e|rule)$ 。为了求得 $score_w(e|rule)$,需要对知识图 G 中找到的每条候选路径进行加权求和。

一旦有了每个候选答案的得分,我们就可以使用如下 softmax 函数进一步计算查询 q 的答案 a 是尾实体 e 的概率分布,如式(5)所示:

$$p_w(a=e|g, q, z) = \frac{\exp(score_w(e))}{\sum_{e' \in A} \exp(score_w(e'))} \quad (5)$$

2.4.4 EM 算法训练流程

接下来将介绍如何用 EM 算法训练规则预测器和规则生成器。因为在每次训练迭代中,我们的目标是最大化等式(2),但同时训练等式(2)两个未知参数 θ 和 w 又十分困难,所以对于每次查询 q ,可以先固定 q 的先验,即已知规则生成器参数 θ 的情况下,近似等式(2)的目标函数为:

$$\max_{\theta, w} \ell(\theta, w) = \max_w \ell_{(G, q, a)}(w) = \max_w (\log p_w(a|g, q, \hat{z})) \quad (6)$$

因此我们先从生成器中采样一组规则 \hat{z} ,并将 \hat{z} 送到规则预测器中。然后更新规则预测器的参数 w 以最大化答案 a 的对数似然函数。

优化完 w 后,接下来用 EM 算法对参数 θ 进行优化。在算法 1 的步骤 E 中,我们的目标是从生成的 \hat{z} 中识别 K 个高质量规则,并将其纳入一个集合 z' ,即 $z' \in \hat{z}$, z' 集合大小为 K 。假设其中一个查询 $q = (h, r, ?)$,答案 $a = t$ 。对于由规则生成器生成的一组规则 \hat{z} ,我们可以为每条规则 $rule$ 计算以下得分 $G, rule \in \hat{z}$:

$$\begin{aligned} G(rule) &= eq\{score_w(t|rule) - \frac{1}{|A|} \sum_{e \in A} score_w(e|rule)\} + \\ &\quad \log \text{RNN}_\theta(rule|r) \end{aligned} \quad (7)$$

其中, A 是由所有 \hat{z} 中的规则发现的候选答案组成的集合。 $score_w(e|rule)$ 是每条 \hat{z} 中的不同规则对发现相同候选答案的得分函数, $score_w(e|rule)$ 的计算方式由等式(4)给出, $\text{RNN}_\theta(rule|r)$ 是生成器对当前规则已知的先验得分函数, eq 是平衡后验与先验的超参数。

因此我们可以将每条规则的 $G(rule)$ 视为规则质量的评估,它考虑了两个因素。(1)基于规则预测器 P_w 的参数 w ,在等式中被用来计算某条规则对正确答案 t 的得分函数减去该

规则对其他候选答案的得分函数。(2)基于规则生成器 P_θ 的参数 θ , 因为每条规则生成时都伴随着先验, 所以在函数中将先验进行对数化后传入。

在 M 步骤中, 因为 E 步骤中获得一组高质量的规则 z' , 所以进一步利用这些规则来更新 M 步骤中规则生成器的参数 θ 。具体来说, 我们将得到的规则集 z' 视为现在训练数据 (G, q, a) 的一部分, 并通过最大化 z' 的对数似然来更新规则生成器参数 θ :

$$\begin{aligned} \mathcal{O}_{(G, q, a)}(\theta) &= \log p_\theta(z' | q) \\ &= \sum_{rule \in z'} \log RNN_\theta(rule | r) + const \end{aligned} \quad (8)$$

利用上述目标函数, 可以有效地将提取自规则预测器的规则 z' 反馈到规则生成器中。通过这种方式, 将大大减小搜索空间并生成更准确的结果。完整的算法流程 1 如算法 1 所示。

算法 1 RNNLogic 知识推理算法

初始化: 对于每个实例 (G, q, a) , 使用规则生成器随机生成一组规则

\hat{z} ; 将生成的规则 \hat{z} 送入规则预测器, 基于 \hat{z} 和式 (6) 更新规则预测器

For i in EM 迭代轮次:

E 步: 根据式 (7) 中的 $G(rule)$ 从 \hat{z} 中识别 K 个高质量规则 z'

M 步: 根据所识别的规则 z' 以及式 (8) 更新规则生成器

结束阶段: 对于每个查询 $q = (h, r, ?)$, 使用规则生成器生成 N 个规则并输入规则预测器, 在进行推理之后选择得分最高的候选答案即为推理答案 a。

3 融合 HousE 和注意力机制的知识推理模型

现有方法中, 基于规则学习的方法对规则和规则权重的更新无法同时进行, 基于知识表示的学习则是将问题完全转化为向量距离的计算, 缺乏可解释性。受 Qu 等^[23] 提出的 RNNLogic 的启发, 我们在模型中选择分离规则的生成和规则权重的学习, 同时本文创新方法 Att-HousE 分别引入一个融合注意力的规则生成器和融合 HousE 嵌入的规则预测器, 二者可以相互增强对方, 解决了规则和规则权重更新不同步的问题, 并通过产生的规则预测缺失的实体, 预测所使用的规则即为可解释路径。完整流程图如图 1 所示, 在 E 步骤中, 规则生成器生成少量高质量的逻辑规则, 规则预测器只关注学习这些高质量规则的权重, 这显著减小了搜索空间, 并计算出更好的推理结果。在 M 步骤中, 规则预测器反过来可以帮助识别一些有用的逻辑规则, 以改善规则生成器。同时, 在规则生成时引入注意力机制以减小规则搜索空间, 在规则预测器的推理时, 选用 HousE 嵌入, 用于学习低维表示的实体和关系, 将其作为预测的有效工具。

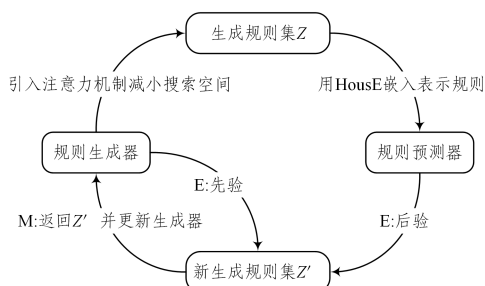


图 1 Att-HousE 算法流程图

Fig. 1 Att-HousE algorithm flowchart

本节将介绍所提出的 Att-HousE 模型, 它是对规则生成器和规则预测器的改进, 分别引入了注意力优化以及嵌入优化, 因此本节将分成两部分来详细介绍。

3.1 对规则生成器进行注意力优化

RNNLogic 的规则生成器生成的规则是由全部训练数据训练后随机输出产生, 因此用于之后的规则预测器时增加了许多冗余的训练量。本文的优化在于, 输出生成规则序列时, 希望可以将注意力集中在那些对当前任务更重要的规则上。也就是给每个经过 RNN 输出的规则中的关系都分配一个权重, 将所有的输出规则中的关系加权平均后再输出。假设初始输出规则为 h_i , 规则中一条关系为 h_{it} , 分配的权重为 a_{it} , 其中, i 表示规则集中规则的条数, t 表示单个规则中关系的条数, h_{it} 表示第 i 个规则中第 t 个关系。则合并后的输出 S_i 即为要得到的最终规则集 z 中的一条规则。

规则 S_i 的生成可以分为以下几个步骤:

(1) 首先将训练数据中以 r 为起始关系的一组关系向量 R_i 作为 RNN 的输入, 在经过一个时间步的 RNN 的计算后, 得到一组新的向量 h_i , 它包含着输入词向量的语义表示, 即该关系衍生出的初始规则。

(2) 接下来引入注意力机制, 目的是找出对输入关系最重要、对规则贡献最大的关系。将 h_{it} 输入一个单层的感知机中, 得到的结果 u_{it} 作为隐含表示。计算式如下:

$$u_{it} = \tanh(W h_{it} + b) \quad (9)$$

其中, W 和 b 为初始化的感知机参数。

(3) 然后为了衡量不同关系的重要性, 我们用 u_{it} 和一个随机初始化的向量 u_m 来做自注意力操作, 经过 softmax 操作获得了一个归一化的 attention 权重矩阵, 其中 a_{it} 代表第 i 个规则中第 t 个关系向量的权重。

$$a_{it} = \frac{\exp(u_{it}^T u_m)}{\sum_t \exp(u_{it}^T u_m)} \quad (10)$$

(4) 有了 attention 权重矩阵 a_{it} 以后, 可以将需要的规则向量 S_i 看作组成该规则的关系向量的加权求和, 之后使用 multinomial 操作将规则向量 S_i 映射为关系的组合。具体操作为对 S_i 的每一行做 n 次随机取值, 输出的值是每一次取值时 S_i 向量对应的行标, n 为取样次数的超参数。又因为随机取值导致返回的值为出现次数最多的向量的下标, 所以可以认为返回的下标即为需要的关系的下标, 对 S_i 全部进行 multinomial 操作后即规则集 z 中的一条规则。这里的随机初始化向量 u_m 是在迭代中不断学习更新的。因此最终的规则 S_i 表示为:

$$S_i = \sum_t a_{it} h_{it} \quad (11)$$

使用注意力机制后能更好的表征关系信息, 可以使训练出的模型具有更好的表现。

3.2 对规则预测器进行嵌入优化

知识图嵌入的有效性在很大程度上取决于对内在复杂关系的建模能力。然而, 先前使用的大部分嵌入方法只能捕捉其中的一些复杂关系, 导致建模能力不足。在本文工作中, 我们通过把 HousE 融合到规则预测器中, 将实体和关系的语义信息映射到高维向量空间。

因为 HousE 引入 Householder 反射变换作为基本操作变换, 并由多个 Householder 反射组合而成 Householder 旋

转和 Householder 投影,所以实现了强大而且全面的建模能力。

在规则预测器中,对于每个规则的标量权重 $\phi_w(rule)$,这是一个可学习优化的超参数;而对于不同路径的得分 $\varphi_w(path)$,RNNLogic 的方法是设置 $\varphi_w(path)=1$ 。但是,此方法无法区分不同的路径。为了解决该问题,本文使用嵌入算法 HousE 为每个实体引入嵌入表示,并将组成规则的每个关系建模为不同实体嵌入上的旋转算子。比如对于根据相同规则 $rule$ 推导出头实体 h 和不同尾实体 e ,如果我们将规则中的关系定义为旋转算子,并且用于旋转变换实体 h 的嵌入表示,然后计算与不同实体 e 嵌入表示之间的距离,距离即为 $\varphi_w(path)$,它可以被看作是每条路径的可靠性和一致性的度量。例如,给定一条路径 $Wang \xrightarrow{\text{friend}} Li \xrightarrow{\text{play}} Football$,我们用 friend 和 play 定义的旋转算子旋转 $Wang$ 的嵌入表示,然后计算 $Wang$ 的旋转变换嵌入与 $Football$ 的实体嵌入的距离 $d(x)$ 作为 $\varphi_w(path)$ 。这种方法允许我们为每条路径计算一个特定的 $\varphi_w(path)$,可以帮助我们对不同候选答案得到更精确的得分。具体而言, $\varphi_w(path)$ 使用的嵌入表示方法就是根据复空间中每个实体和关系的嵌入进行初始化。形式上,实体 e 的嵌入表示为 x_e ,关系 r 的嵌入表示为 x_r ,对于一个存在的路径 $path=e_0 \xrightarrow{r_1} e_1 \xrightarrow{r_2} e_2 \cdots e_{l-1} \xrightarrow{r_l} e_l$,可以定义 $\varphi_w(path)$ 为:

$$\varphi_w(path)=\sigma(\delta-d(x_{e_0} \circ x_{r_1} \circ x_{r_2} \circ \cdots \circ x_{r_l}, x_{e_l})) \quad (12)$$

其中, $\sigma(x)$ 是激活函数, $d(x)$ 是两个复向量之间的距离函数, δ 是超参数, \circ 是复空间中的 Hadward 乘积。

4 实验与分析

4.1 数据集及实验设置

在本文实验中,我们采用知识推理领域通用的两个基准数据集 FB15K-237^[24] 和 WN18RR^[25] 作为实验的数据集,两者都是较大数据集的子集,其中 FB15K-237 是从 FB15K 中去除了冗余关系后得到的三元组的数据集。WN18RR 是 WordNet 的子集,其中更多地保留了原数据集中的对称关系、非对称关系和组合关系,而去除了反转关系。以上两个数据集对应的实体、关系、三元组的划分如表 1 所列。

表 1 实验使用的数据集

Table 1 Datasets used in experiment

数据集	关系数	实体数	训练集	验证集	测试集
FB15K-237	237	14541	272115	17535	20466
WN18RR	11	40943	86835	3034	3134

其中实验使用的操作系统为 Ubuntu20.04,GPU 使用显存为 32GB 的 GTX3090,开发环境为 Python3.8,Pytorch1.7.1。本文对 Att-HousE 框架中各个部分的参数设置如表 2 所列。

表 2 实验使用的参数

Table 2 Parameters used in experiment

数据集	超参数 δ	学习率	超参数 u_m	超参数 eq
FB15K-237	9	1×10^{-2}	1.5	0.9
WN18RR	6	1×10^{-2}	1.5	0.9

对于规则生成器,对于 FB15k-237,生成规则的最大长度设置为 4,对于 WN18RR 设置为 5。最大长度是根据验证

数据的正态分布选择的。RNN $_{\theta}$ 中的输入层和隐藏层的大小分别被设置为 512 和 256,学习率被设置为 1×10^{-2} 。同时使用注意力机制生成高置信度的逻辑规则,防止规则生成器在训练开始时探索无意义的逻辑规则,注意力超参数 u_m 被设置为单个元素为 1.5 的列向量。

规则预测器:在使用 HousE 嵌入加快预测的情况下,FB15k-237 的三元组嵌入维数被设置为 500,WN18RR 的三元组嵌入维数被设置为 200。对于 $\varphi_w(path)$ 的第二种初始化方法,我们将 FB15k-237 的超参数 δ 设置为 9,将 WN18RR 的超参数 δ 设置为 6。

EM 优化过程:在平衡 EM 学习过程中的先验与后验时,将超参数 eq 设置为 0.9。同时将 EM 循环次数定为 10 次,在每次循环时,每条查询从规则生成器中生成 1000 条规则,在 E 步骤中,我们识别 300 条规则作为高质量的逻辑规则,并在 M 步骤中传回生成器。

4.2 评估指标

对于知识推理模型的评估指标,通常使用事实预测判断生成的三元组是否正确,设计的指标有 MR(预测排名的平均值)、MRR(倒数排名的平均值)、Hits@ n (排名位置不大于 n)。在评估时,假设输入被分为 m 批三元组,每一批有 n 个三元组,对于每一批正确的三元组,需要将该批三元组的尾实体换成任意一种其他实体,共替换 $n-1$ 个(因为正确的三元组占有一个名额),然后将这一批 n 个三元组传入模型得出分数,将这 n 个三元组按照得分从高到低排列,然后找出正确的三元组在这 n 个三元组中的排名,记为 $rank1$;再根据上面的方法计算出另外 $(m-1)$ 批三元组中每批正确的实体在其批次中的各自的排名 $rank2, rank3, \dots, rankm$,则 MRR, MR, Hits@ n 的值的计算式如下:

$$MRR = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{rank_i} \right) \quad (13)$$

$$MR = \frac{1}{m} \sum_{i=1}^m (rank_i) \quad (14)$$

$$Hits@n = \frac{1}{m} \sum_{i=1}^m \text{if}(rank_i \leq n) \quad (15)$$

4.3 实验结果及分析

为了验证 Att-HousE 模型的有效性,将本文方法 Att-HousE 与马尔可夫逻辑网络^[14]、路径排名算法^[7]这些传统算法做比较,我们还考虑神经逻辑编程方法,如 CTP^[16],一种基于神经定理证明器的可微分方法。我们还比较了一些嵌入方法,包括 TransE^[11],Complex^[18],RotatE^[12]。同时,我们还将本文方法与其他融合算法相比较,如 RNNLogic^[23],DeepPath^[20],PTransE^[21]。以上对比实验都在 FB15k-237 和 WN18RR 这两个公开数据集上进行,实验结果如表 3 和表 4 所列。为了对本文模型做显著性检验,我们将每个模型运行 5 次,每次使用不同的种子,并使用配对 t 检验比较 ATT-HousE 是否整体上明显优于 RNNLogic。我们选定显著性水平 $p=0.05$,计算统计结果后得出各个指标的单尾数值均低于 0.05,证实了 ATT-HousE 整体上明显优于 RNNLogic。为方便起见,表格中数据选择每次实验的平均数与标准差,其中“—”指结果与别的算法差异过大,不具有统计学意义,因此忽略不计。

表 3 各方法在 FB15k-237 上的事实预测结果

Table 3 Fact prediction results of different methods on FB15k-237

种类	算法	FB15k-237				
		MR	MRR	Hits@1	Hits@3	Hits@10
基于规则学习的算法	路径排名	—	0.088±0.004	8.5±0.3	9.8±0.3	12.3±0.5
	马尔可夫网络	—	0.135±0.002	12.5±0.5	17.6±0.5	33.4±0.4
	CTP	—	0.079±0.009	7.2±0.2	8.5±0.2	10.3±0.3
基于知识表示的算法	TransE	356±2	0.295±0.004	10.3±0.6	58.4±0.4	46.6±0.6
	Complex	337±5	0.289±0.006	15.6±0.4	27.3±0.1	41.5±0.3
	RotatE	177±3	0.338±0.003	24.1±0.2	40.5±0.4	55.3±0.2
	RNNLogic	538±5	0.288±0.002	20.8±0.6	31.5±0.2	44.5±0.4
融合算法	DeepPath	168±6	0.327±0.006	23.8±0.5	35.6±0.8	50.8±0.6
	PTransE	172±8	0.331±0.012	23.9±1.1	36.4±1.5	51.2±1.8
本文算法	Att-HousE	181±4	0.351±0.003	25.9±0.4	36.1±0.5	54.5±0.6

在 FB15k-237 数据集上, Att-HousE 性能提升显著。Att-HousE 取得了比传统基于规则学习和知识表示的算法更好的结果, 比其中效果最好的 RotatE 在 MRR 指标上整体高出 1.3%。这是因为 Att-HousE 的规则生成器和推理预测器可以相互协作, 以减小搜索空间并学习更好的规则。Att-HousE 也优于 RNNLogic, DeepPath 和 PTransE 等融合算法, 在 MRR 指标上整体比 RNNLogic 高出 6.3%。原因是 Att-HousE 使用 HousE 嵌入进行了优化, 其中推理预测器为规则生成器提供了更有用的反馈, 从而提高了准确度。

但在 WN18RR 数据集上, Att-HousE 整体上却不如 RotatE 和 RNNLogic, 不过在 Hits@3 指标上依然具有优势, 整体上分别高出 0.2% 和 1.9%, 在 Hits@10 指标上高出 4.7% 和 2.7%。且相比其他融合算法, Att-HousE 依然有较大提升, 在 MRR 指标上比 DeepPath 整体高出 6.7%。经过思考, 原因是 HousE 嵌入的维度过高, 对简单路径的 WN18RR 数据集空间冗余消耗过大, 而计算单一的简单

嵌入在此数据集上适用性更高。

经过以上对比分析可知, 本文方法在关系较多的 FB15K-237 数据集上的效果优于 WN18RR 数据集, 主要原因是: 其他算法在应用于关系更复杂的 FB15K-237 数据集中时, 因为不同实体之间路径长度比 WN18RR 更长, 选择路径时更容易出错, 所以模型性能降低。而本文方法的优势在于降低错误路径的选择率, 减小复杂路径的搜索空间, 因此在 FB15K-237 数据集上效果提升更明显。

4.4 消融实验

为了验证 Att-HousE 模型中的 HousE 嵌入模型优于其他嵌入, 以及验证注意力机制的有效性, 我们进一步进行消融实验, 以度量注意力机制与 HousE 嵌入方法在提升实体预测准确度中的作用。消融实验在 Att-HousE 的基础上去除注意力机制, 得到 NoAtt-HousE; 接着替换 Att-HousE 中的嵌入为 TransE, 得到 Att-TransE。对这两个模型进行实验时, 使用和 Att-HousE 相同的数据集以及初始化参数进行训练与验证, 实验结果如表 4 所列。

表 4 各方法在 WN18RR 上的事实预测结果

Table 4 Fact prediction results of different methods on WN18RR

种类	算法	WN18RR				
		MR	MRR	Hits@1	Hits@3	Hits@10
基于规则学习的算法	路径排名	—	0.188±0.003	16.9±0.2	19.5±0.3	21.2±0.5
	马尔可夫网络	—	0.195±0.005	17.8±0.4	21.5±0.1	23.8±0.3
	CTP	—	0.165±0.007	13.2±0.5	16.8±0.2	18.3±0.7
基于知识表示的算法	TransE	3385±12	0.225±0.006	21.5±0.4	32.2±0.3	49.5±0.2
	Complex	5266±18	0.424±0.003	40.2±0.2	45.4±0.2	50.8±0.4
	RotatE	3340±6	0.476±0.003	42.8±0.1	49.2±0.1	50.1±0.3
	RNN-Logic	7527±19	0.455±0.012	41.4±0.7	47.5±0.3	53.1±0.1
融合算法	DeepPath	4246±15	0.351±0.008	38.4±0.4	45.2±0.9	49.5±0.5
	PTransE	4581±8	0.393±0.007	42.5±0.3	46.1±0.7	51.2±0.2
本文算法	Att-HousE	4656±13	0.418±0.006	41.3±0.2	49.4±0.3	55.8±0.2

如表 4 所列, NoAtt-HousE 和 Att-TransE 算法在各指标的实验结果都不及 Att-HousE 算法, 主要原因是 NoAtt-HousE 没有对生成器进行注意力优化, 导致冗余规则过多, 知识推理难度增大。Att-TransE 算法各指标在两个数据集上低于 Att-HousE 算法的原因则是嵌入的选择, TransE 对于对称及反对称等复杂关系难以进行优化, 未能充分利用关系路径等重要信息, 因此推理能力受限。实验结果表明, 算法引入注意力机制与 HousE 嵌入对推理任务的完成具有明显的提升效果。

4.5 可解释性分析

为了证明 ATT-HousE 在实验中对知识推理任务可解释

性的影响, 对规则的推理仍然具有可解释的意义, 我们列举了模型在实验中学习到的部分规则, 展示了规则体选择的关系序列, 结果如表 5 所列。

在表 6 中, 本文模型对于不同的关系查询产生多条不同的推理规则, 例如以“play”为关系查询, 产生的规则有“friend ∧ play”和“teammate ∧ play”, 意味着朋友玩过的游戏或队员玩过的游戏, 这种游戏自己也玩过。剩余的推理规则也可进行类似分析, 此处不一一列举。因此推理规则的结果表明 ATT-HousE 对于知识推理的规则依然保持其可解释性。

表 5 消融实验对比结果

Table 5 Comparison of ablation experiment results

数据集	算法	MR	MRR	Hits@1	Hits@3	Hits@10
FB15k-237	NoAtt-HousE	477±5	0.295±0.006	22.4±0.5	35.6±0.7	49.7±0.8
	Att-TransE	275±2	0.298±0.003	12.2±0.4	60.5±0.2	47.4±0.4
	Att-HousE	181±4	0.351±0.003	25.9±0.4	36.1±0.5	54.5±0.6
WN18RR	NoAtt-HousE	5425±15	0.405±0.008	40.8±0.2	48.2±0.1	50.9±0.2
	Att-TransE	3568±8	0.258±0.003	23.4±0.1	35.8±0.3	51.6±0.6
	Att-HousE	4656±13	0.418±0.006	41.3±0.2	49.4±0.3	55.8±0.2

表 6 ATT-HousE 学习到的规则

Table 6 Rules learned by ATT-HousE

关系	推理规则
wife	wife←mother ∧ inv_father
	wife←mother ∧ daughter
	wife←mother ∧ son
play	play←friend ∧ play
	play←teammate ∧ play
	play←family ∧ play
affects	affects←processof ∧ cause
	affects←prevent ∧ disrupts
	affects←occurs in ∧ causes
关系	推理规则
brother	brother←sister ∧ brother
	brother←son ∧ father
	brother←brother ∧ brother

结束语 本文对知识推理任务进行研究,提出了一种新的知识推理算法 Att-HousE 方法。算法包括一个生成规则的规则生成器和一个推理规则的规则预测器,它们通过 EM 算法同时被训练以相互增强,同时在规则生成器中引入注意力机制,大大缩小搜索空间,在规则预测器中引入 HousE 嵌入以提高模型对复杂关系的推理能力以及可解释性。在两个公开数据集上进行比较可知,本文算法在推理准确度上有部分提升,同时消融实验进一步证明 HousE 的选择优于现有的其他嵌入。

在未来的工作中,可从两个方面进行优化:一是对嵌入的优化,实验中发现不同的数据集对不同的嵌入所反馈的结果不一样,原因可能是不同数据集的关系系数不同造成的复杂度不同,因此考虑对不同数据集选择不同嵌入能否进一步提高推理准确度;二是在 EM 算法流程上的优化,当前 EM 算法计算起来比较简单,只涉及目标函数的最大似然,但对初始值很敏感并且容易陷入局部最优解,因此如何进一步提高算法的全局搜索度有待研究。

参考文献

[1] HUANG Q H, YU J, LIAO X, et al. Review of Knowledge Graph Research[J]. Computer Systems & Applications, 2019, 28(6):1-12.

[2] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: A collaboratively created graph database for structuring human knowledge[C]// Proceedings of 2008 ACM SIGMOD International Conference on Management of Data. 2008:1247-1250.

[3] MILLER G A. WordNet: A lexical database for English [J]. Communications of the ACM, 1995, 38(11):39-41.

[4] CARLSON A, BETTERIDGE J, KISIEL B, et al. Toward an architecture for never-ending language learning[C]// Proceedings

of the 24th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2010:1306-1313.

[5] XU Z L, SHENG Y P, HE L R, et al. Overview of Knowledge Graph Technology [J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4):589-606.

[6] TIAN L, ZHANG J C, ZHANG J H, et al. Overview of Knowledge graph: Representation, Construction, Reasoning and Knowledge Hypergraph Theory[J]. Computer Systems & Applications, 2021, 41(8):2161-2186.

[7] LAO N, COHEN W W. Relational retrieval using a combination of path-constrained random walks[J]. Machine Learning, 2010, 81(1):53-67.

[8] SUN S, CHEN J, LIU D, et al. A Posterior-Based Method for Markov Logic Networks Parameters Learning[C]// 2006 5th IEEE International Conference on Cognitive Informatics, Beijing, China, 2006:529-534.

[9] YANG F, YANG Z L, COHEN W W. Differentiable Learning of Logical Rules for Knowledge Base Reasoning[J]. arXiv:1702.08367, 2017.

[10] ASCHEL T R, RIEDEL S. End-to-end differentiable proving [J]. Advances in Neural Information Processing Systems, 2017, 2017(December):3789-3801.

[11] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]// Advances in Neural Information Processing Systems. 2013:2787-2795.

[12] SUN Z, DENG Z H, NIE J Y, et al. Rotate: Knowledge graph embedding by relational rotation in complex space[C]// 7th International Conference on Learning Representations, 2019.

[13] LI R et al. HousE: Knowledge Graph Embedding with Householder Parameterization [J]. International Conference on Machine Learning, 2022.

[14] KHOT T, NATARAJAN S, KERSTING K, et al. Learning markov logic networks via functional gradient boosting[C]// IC-DM. 2011.

[15] MEILICKE C, CHEKOL M W, RUFFINELLI D, et al. Anytime bottom-up rule learning for knowledge graph completion[C]// IJCAI, 2019.

[16] MINERVINI P, BOŠNJAK M, ROCKTÄSCHELT, et al. Differentiable Reasoning on Large Knowledge Bases and Natural Language[J]. arXiv:1912.0824, 2019.

[17] RONAN C, WESTON J. A unified architecture for natural language processing: deep neural networks with multitask learning [C]// International Conference on Machine Learning. 2008.

[18] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]// ICML. 2016.

[19] ZHANG S, TAY Y, YAO L N, et al. Quaternion knowledge graph embeddings[C]// Proceedings of the 33rd Conference on

Neural Information Processing Systems. 2019;2735.

- [20] XIONG W, HOANG T, WANG W Y. DeepPath: A reinforcement learning method for knowledge graph reasoning[J]. arXiv: 1707.06690, 2017.
- [21] LIN Y, LIU Z, LUAN H, et al. Modeling relation paths for representation learning of knowledge bases[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015;705-714.
- [22] LI N, SHEN Q, SONG R, et al. MEduKG: A Deep-Learning-Based Approach for Multi-Modal Educational Knowledge Graph Construction[J]. Information, 2022, 13(2):91.
- [23] QU M, CHEN J, XHONNEUX L, et al. RNNLogic: Learning Logic Rules for Reasoning on Knowledge Graphs[J]. arXiv: 2010.04029, 2020.
- [24] TOUTANOVA K, CHEN D Q. Observed versus latent features for knowledge base and text inference[C]//Workshop on Con-

tinuous Vector Space Models and their Compositionality. 2015.

- [25] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2d knowledge graph embeddings[C]//AAAI. 2018.



ZHU Yuliang, born in 1999, postgraduate. His main research interests include knowledge reasoning and recommender system.



LIU Juntao, born in 1979 Ph.D, professor. His main research interests include recommender system, knowledge computing, and decision support.