



计算机科学

COMPUTER SCIENCE

基于表示学习的知识图谱推理研究综述

李志飞, 赵月, 张龔

引用本文

李志飞, 赵月, 张龔. [基于表示学习的知识图谱推理研究综述](#) [J]. 计算机科学, 2023, 50(3): 94-113.

LI Zhifei, ZHAO Yue, ZHANG Yan. [Survey of Knowledge Graph Reasoning Based on Representation Learning](#) [J]. Computer Science, 2023, 50(3): 94-113.

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)

基于表示学习的知识图谱推理研究综述

李志飞 赵月 张龔

湖北大学计算机与信息工程学院 武汉 430062

摘要 知识图谱以结构化形式描述了现实世界中的客观知识,但面临着构建不完整或者无法处理新增知识等挑战。知识图谱推理方法成为了知识图谱补全和更新的重要手段,该方法旨在基于图谱中已有的事实推断出未知的事实。近年来,基于表示学习的知识图谱推理研究受到了广泛关注,其主要研究思路是将实体和关系嵌入到低维连续向量空间从而进行推理,具有计算效率高、推理性能高等优势。文中以基于表示学习的知识图谱推理方法为研究对象,首先对相关的符号表示、数据集、评价指标、训练方法以及评测任务进行了简要概述;其次介绍了基于平移距离和语义匹配的两种典型知识图谱推理方法;然后对融合多源信息的推理方法进行了分类和梳理,以及详细分析了近期流行的基于神经网络的推理研究进展;最后总结全文,同时对知识图谱推理的未来研究方向进行展望。

关键词: 表示学习;知识图谱推理;平移距离;语义匹配;多源信息;神经网络

中图法分类号 TP391

Survey of Knowledge Graph Reasoning Based on Representation Learning

LI Zhifei, ZHAO Yue and ZHANG Yan

School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China

Abstract Knowledge graphs describe objective knowledge in the real world in a structured form, and are confronted with issues of completeness and newly-added knowledge. As an important means of complementing and updating knowledge graphs, knowledge graph reasoning aims to infer new knowledge based on existing knowledge. In recent years, the research on knowledge graph reasoning based on representation learning has received extensive attention. The main idea is to convert the traditional reasoning process into semantic vector calculation based on the distributed representation of entities and relations. It has the advantages of fast calculation efficiency and high reasoning performance. In this paper, we review the knowledge graph reasoning based on representation learning. Firstly, this paper summarizes the symbolic representation, data set, evaluation metric, training method, and evaluation task of knowledge graph reasoning. Secondly, it introduces the typical methods of knowledge graph reasoning, including translational distance and semantic matching methods. Thirdly, multi-source information fusion-based knowledge graph reasoning methods are classified. Then, neural network-based reasoning methods are introduced including convolutional neural network, graph neural network, recurrent neural network, and capsule network. Finally, this paper summarizes and forecasts the future research direction of knowledge graph reasoning.

Keywords Representation learning, Knowledge graph reasoning, Translational distance, Semantic matching, Multi-source information, Neural network

1 引言

知识是人类智慧的象征,是对现实世界中相关信息的抽象化概括与总结。基于已有知识,人类才能针对现实问题进行分析、对比、演绎,从而找出解决方案。基于此逻辑,愈来愈多研究人员开始关注如何利用已有知识来辅助具体任务。但是随着信息技术的快速发展,互联网上产生了海量的数据并

蕴藏着丰富的知识信息,如何存储和表示这些知识成为了一个难题。在此背景下,知识图谱(Knowledge Graph)作为知识的一种结构化表达,近年来引起了广泛关注。2012年5月,谷歌公司正式发布基于知识图谱的搜索产品——谷歌知识图谱(Google Knowledge Graph),并提出了“Things, Not Strings”¹⁾的理念,旨在链接相关内容和主题,将搜索结果进行结构化展示,使用户更容易发现新知识。如图1所示,当

¹⁾ <https://www.blog.google/products/search/introducing?knowledge?graph-things?not/>

到稿日期:2022-09-15 返修日期:2022-12-01

基金项目:国家自然科学基金(62207011,61977021,62101179);湖北省技术创新专项重大项目(2019ACA144);湖北大学校级项目(202111903000001,202011903000002)

This work was supported by the National Natural Science Foundation of China(62207011,61977021,62101179), Key Project of Technology Innovation in Hubei Province(2019ACA144) and School Project of Hubei University(202111903000001,202011903000002).

通信作者:李志飞(zhifei1993@hubu.edu.cn)

用户搜索“C罗”时,搜索结果不仅返回相关网址,还会以卡片形式展示“C罗”的出生年月、身高、球队等信息,显著地提升了搜索引擎的用户体验。

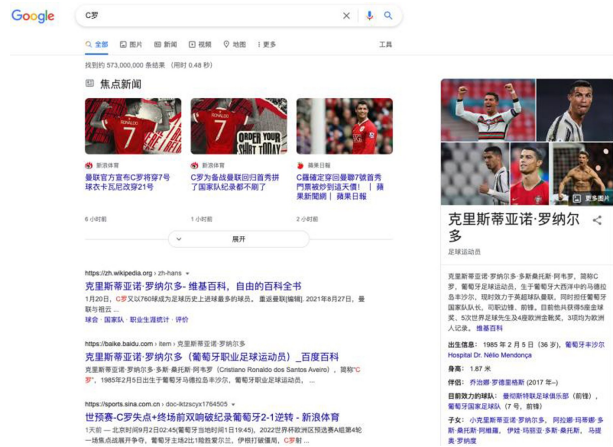


图1 谷歌知识图谱搜索应用

Fig. 1 Search application based on Google knowledge graph

构建知识图谱是从海量互联网数据中抽取知识并进行结构化存储的过程,主要包括信息抽取、知识融合、知识加工以及知识更新等步骤^[1]。近年来,人工智能的快速发展为知识图谱的构建工作提供了新的机遇,突破了传统方法在知识获取方面的瓶颈。

如表1所列,国内外公司及科研机构建立了一大批规模庞大、开放共享的知识图谱。其中国外具有代表性的有语言知识库 WordNet^[2]、世界知识库 Freebase^[3]、多语言知识库

YAGO^[4]、语义网 DBpedia^[5]等。同时国内也涌现了一批知名的知识图谱,包括复旦大学的 CN-DBpedia^[6]、上海交通大学的 zhishi.me^[7]等,这些知识图谱普遍遵循资源描述框架技术标准,包含千万级或者亿级规模的实体,并且这些实体被组织在数十亿甚至百亿的知识三元组中。这些大规模知识图谱被广泛应用于信息检索、问答系统、智能对话等知识驱动型应用^[8-11],成为推动互联网和人工智能发展的核心驱动力之一。

知识图谱为知识存储、组织与管理提供了一种更为有效的方式,通常包含实体(Entity)、关系(Relation)以及相关属性信息。知识图谱以结构化形式描述了现实世界中的客观概念及其相互关系,其基本组成单位是由(头实体,关系,尾实体)组成的知识三元组(Triplet)。其中,节点代表实体,包括具体事物或抽象概念(如人名、地名、机构等),相邻节点的连边代表实体间的关系。由于知识获取不全面,现有知识图谱存在构建不完整或存在错误知识等问题。以世界知识库 Freebase 为例¹⁾,虽然其包含5亿多个实体以及19亿多条知识三元组,但可以发现78.5%的人没有国籍信息以及93.8%的人没有出生地信息,并且有69%的实体缺少属性信息。此外,由于抽取算法的误差或数据来源的缺失等因素,导致知识图谱内部存在知识矛盾或不准确等问题。在知识图谱 NELL^[12]中,知识采集的正确率随着时间的推移而不断下降,其主要原因就是知识抽取算法的不准确导致抽取知识的错误。解决以上问题的办法之一就是知识图谱推理^[13-14](Knowledge Graph Reasoning)方法根据已有知识去推理新的知识。

表1 代表性大规模知识图谱简介

Table 1 Introduction of representative large-scale knowledge graphs

名称	机构	特点	应用
Freebase	MetaWeb (谷歌收购)	实体、语义、属性、关系 自动+人工;维基百科抽取+人工编辑	Google Search
DBpedia	莱比锡大学、 柏林自由大学、 OpenLink Software	实体、语义、属性、关系 维基百科抽取	DBpedia
YAGO	马克思·普朗克 研究所	维基百科、WordNet 和 GeoNames 抽取	YAGO
Wikidata	Wikimedia Foundation	实体、语义、属性、关系 人工(协同编辑)	维基百科
Zhishi.me	上海交通大学	互动百科、百度百科、中文维基百科抽取	Zhishi.me
CN-DBpedia	复旦大学	大规模通用领域结构化百科 百度百科、互动百科、中文维基百科抽取	知识工场
知立方	搜狗	搜索中的结构化数据	搜狗搜索
百度知心	百度	搜索中的结构化数据	百度搜索

进行知识图谱推理研究首先需解决知识表示的问题。在互联网数据呈爆炸性增长的时代,知识图谱常面临海量知识的补充与更新,从而导致知识图谱的规模愈加庞大。针对大规模知识图谱的计算与应用,需重点关注以下两个问题^[15]:1)计算效率问题,传统基于图的算法计算复杂度高、可扩展性差;2)数据稀疏问题,在知识图谱中存在一些罕见的实体或关系,从而导致计算准确率降低。为解决以上两个难题,以深度学习^[16](Deep Learning)为代表的表示学习^[17](Representation Learning)方法得到了广泛关注,

并在自然语言处理、社交网络分析、语音识别等领域获得了研究与发展。表示学习旨在将研究对象的语义信息表示成低维连续向量,两个对象的空间距离越近,说明其语义相似度越高。顾名思义,如图2所示,面向知识图谱的表示学习旨在将实体和关系嵌入到低维连续向量空间,即学习它们的分布式表示^[18-19](Distributed Representation)。这种方法不仅能够有效表示和度量实体、关系间的语义关联,还有助于提升计算效率,缓解数据稀疏,实现异质信息融合。

¹⁾ <https://developers.google.com/freebase/>

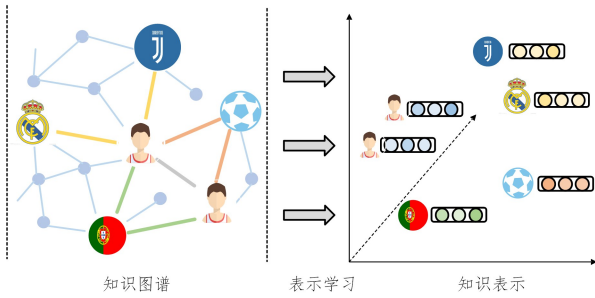


图2 面向知识图谱的表示学习

Fig. 2 Representation learning for knowledge graph

近年来,已有不少关于知识图谱表示与推理的综述性文献陆续发表。譬如,文献[15]介绍了知识表示学习的最新进展以及面临的挑战与解决方案;文献[18]对面向知识图谱的统计关系学习进行了综述;文献[19]对现有的知识图谱表示方法进行了系统的综述,重点关注基于知识图谱本体结构及多源信息融合的方法;文献[20]对常见的大规模知识图谱进行了介绍,并对知识图谱的推理补全及知识去噪相关方法进行了综述;文献[21]从面向知识图谱的知识表示、知识获取与知识应用等方面进行了全方位的综述;文献[22]从神经网络、符号主义与神经网络-符号主义联合这3个层面对知识图谱补全和知识图谱问答过程中的推理方法进行了综述;文献[23]从实验角度对18种知识图谱表示方法进行了详细的对比分析。

相比上述综述文献,本文聚焦于基于表示学习的知识图谱推理方法,主要贡献在于:1)对已有及新增的典型推理方法与信息融合推理方法进行了回顾总结,并详尽梳理了其中的关键创新和解决思路;2)对最新基于神经网络推理方法的研究进展进行了对比分析,重点归纳了相关方法在特征提取与特征汇聚方面的差异区别。此外,本文整理了本领域的常用数据集、评价指标、训练方法与评测任务,总结了本领域的发展现状与未来发展方向,以期为未来的研究工作奠定基础。

本文第2节主要概述了知识图谱推理中涉及的符号表示、评价指标、数据集、训练方法以及评测任务等;第3节主要介绍了知识图谱推理的典型方法,包括平移距离、语义匹配以及其他改进方法;第4节总结融合了多源信息知识图谱推理方法,依据不同信息进行分类描述;第5节对近期流行的基于神经网络的知识图谱推理方法进行了分类介绍;第6节总结全文并展望未来。

2 知识图谱推理相关介绍

2.1 符号表示

表2列出了知识图谱推理中的常用符号。知识图谱通常以实体(Entity)、关系(Relation)以及知识三元组(Triplet)进行表示,可简化表示为:

$$\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T}) \quad (1)$$

其中, \mathcal{G} 代表知识图谱, \mathcal{E} 代表实体集合, \mathcal{R} 代表关系集合, \mathcal{T} 代表知识三元组集合。对于代表某一知识的三元组,可表示为 (h, r, t) 。其中 h 代表头实体, t 代表尾实体, r 代表头尾实体之间的关系,根据定义可以得知 $h, t \in \mathcal{E}, r \in \mathcal{R}$ 以及 $(h, r, t) \in \mathcal{T}$ 。

以三元组(姚明,职业,篮球运动员)为例,其代表了头实体“姚明”和尾实体“篮球运动员”之间存在“职业”的关系。本文使用粗体符号表示其对应的向量表示,例如 $\mathbf{h}, \mathbf{r}, \mathbf{t}$ 即表示头尾实体与关系的向量表示。此外,通常每个知识图谱推理模型会定义一个评分函数 $\phi(h, r, t)$,用于判断三元组的置信度。

表2 知识图谱推理符号表示及说明

Table 2 Symbolre presentation and explanation of knowledge graph

reasoning			
符号	说明	符号	说明
\mathcal{G}	知识图谱	\mathcal{T}	三元组集合
\mathcal{E}	实体集合	\mathcal{R}	关系集合
\mathbf{E}	实体向量矩阵	\mathbf{R}	关系向量矩阵
\mathbf{h}/\mathbf{t}	头/尾实体	r	关系
\mathbf{h}/\mathbf{t}	头/尾实体向量	\mathbf{r}	关系向量
$ \cdot $	集合大小	$\phi(\cdot)$	评分函数

2.2 评价指标

知识图谱推理相关任务有3个常用评测指标,分别为平均排名MR(Mean Rank)、平均倒数排名MRR(Mean Reciprocal Rank)和前 k 命中率(Hit@ k , k 通常取1,3,10),具体计算式如下:

$$\begin{cases} MR: \frac{1}{2|\mathcal{T}_{\text{test}}|} \sum_{i \in \mathcal{T}_{\text{test}}} (\text{rank}_i^h + \text{rank}_i^t) \\ MRR: \frac{1}{2|\mathcal{T}_{\text{test}}|} \sum_{i \in \mathcal{T}_{\text{test}}} \left(\frac{1}{\text{rank}_i^h} + \frac{1}{\text{rank}_i^t} \right) \\ \text{Hit}@k: \frac{1}{2|\mathcal{T}_{\text{test}}|} \sum_{i \in \mathcal{T}_{\text{test}}} [I[\text{rank}_i^h \leq k] + I[\text{rank}_i^t \leq k]] \end{cases} \quad (2)$$

其中, rank_i^h 与 rank_i^t 分别表示头实体和尾实体的排名, $|\mathcal{T}_{\text{test}}|$ 与 $|\mathcal{T}_{\text{test}}|$ 分别表示测试集中的三元组及数量, $I[P]$ 为指示函数,条件 P 成立则返回1,否则返回0。MR与MRR分别表示正确实体(或关系)在候选评分列表中的平均排名与倒数排名,MR值越低或MRR值越高则表示模型整体预测性能越好。以 $k=10$ 为例,Hit@10表示正确的实体(或关系)出现在候选评分列表前10的占比,因此Hit@ k 值越高表明模型推理越精确。

2.3 常用数据集

在知识图谱推理领域,研究人员通常选择从大规模知识图谱中抽取部分信息生成数据集。常用的知识图谱数据集包括FB15k^[24],WN18^[24],FB15k-237^[25],WN18RR^[26],NELL-995^[27]以及YAGO3-10^[28],其中实体数、关系数以及数据集划分等统计信息如表3所列。

表3 知识图谱推理数据集统计信息

Table 3 Dataset statistics of knowledge graph reasoning

数据集	实体数	关系数	数据集划分		
			训练集	验证集	测试集
FB15k	14951	1345	483142	50000	59071
WN18	40943	18	141442	5000	5000
FB15k-237	14541	237	272115	17535	20466
WN18RR	40943	11	86835	3034	3134
NELL-995	74536	200	149678	543	2818
YAGO3-10	123182	37	1079040	5000	5000

(1)FB15k。此数据集抽取于世界知识库Freebase^[3],包含14951个实体和1345种关系,其中描述了有关电影、演员、奖项、体育以及运动队伍等知识。

(2) WN18. 此数据集抽取于语言知识库 WordNet^[2], 包含 40943 个实体和 18 种关系。它的实体代表词义, 而关系定义了实体之间的词汇关系。

(3) FB15k-237 与 WN18RR. 两者分别为 FB15k 与 WN18 的子集, 为检验模型面对复杂关系时的推理性能, 两者剔除了数据集中的所有可逆关系。

(4) NELL-995. 此数据集抽取于卡内基梅隆大学开发的知识图谱 NELL^[12] (Never-Ending Language Learner), 它包含 74 536 个实体以及 200 种关系。

(5) YAGO3-10. YAGO3-10 数据集抽取于多语言知识库 YAGO^[4], 它包含 123 182 个实体以及 37 种关系。其中, 每个实体至少有 10 种关系, 描述了个人的相关属性, 如职业、性别和公民身份。

2.4 模型训练

关于知识图谱推理方法的模型训练通常包含 3 个步骤:

1) 随机初始化所有实体及关系的向量表示; 2) 依据评分函数定义模型的损失函数; 3) 进行实验训练得到实体及关系最终表示并进行推理。具体地, Xavier 方法^[29]通常被用于初始化所有模型参数。常见的损失函数包括基于最大间隔损失 (Margin Loss) 函数

$$\mathcal{L} = \sum_{(h,r,t) \cup (h',r',t')} \max(0, \phi(h,r,t) - \phi(h',r',t') + \gamma) \quad (3)$$

和

$$\mathcal{L} = -\log \sigma(\gamma - \phi(h,r,t)) - \sum_{(h',r',t')} \log \sigma(\phi(h',r',t') - \gamma) \quad (4)$$

以及基于交叉熵损失 (Cross-Entropy Loss) 函数

$$\mathcal{L} = \sum_{(h,r,t) \cup (h',r',t')} \log(1 + \exp(-y\phi(h,r,t))) \quad (5)$$

和

$$\mathcal{L} = - \sum_{(h,r,t) \cup (h',r',t')} (y \log \phi(h,r,t) + (1-y) \log(1 - \phi(h',r',t'))) \quad (6)$$

其中:

$$y = \begin{cases} 1, & \text{for } (h,r,t) \in \mathcal{T} \\ 0, & \text{for } (h',r',t') \notin \mathcal{T} \end{cases} \quad (7)$$

其中, (h',r',t') 表示错误三元组, 即将正确三元组中的头实体或尾实体通过随机选取的方式替换, 生成不存在于知识图谱中的错误三元组。常用的错误三元组生成方式有 1-1 模式 (针对每个正确三元组, 替换头实体或尾实体生成唯一错误三元组) 和 1-N 模式 (针对每个正确三元组, 将头实体或尾实体依次用数据集中的 N 个实体进行替换, 生成 N 个错误三元组)。

算法 1 介绍了知识图谱推理模型的常见训练过程。具体地, 在一些推理模型中, 会通过分阶段使用随机丢失 (Drop-out Rate)^[30] 技术来对模型进行正则化, 并采用批处理归一化 (Batch Normalization)^[31] 方法来加快模型的收敛速度; 此外, 还利用标签平滑 (Label Smoothing)^[32] 技术来减少模型过拟合并提高泛化能力。最后, 通常使用 Adam 优化器 (Adam Optimizer)^[33] 对损失函数进行优化。

算法 1 知识图谱推理模型常见训练过程

输入: 知识图谱 $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$; 向量维度 d ; 批处理大小 b ; 学习率 α

输出: 所有实体及关系的最终向量表示 \mathbf{E} 和 \mathbf{R}

1. 初始化所有实体 \mathcal{E} 及关系 \mathcal{R} 为 d 维向量矩阵 \mathbf{E} 和 \mathbf{R}
2. loop
3. 随机采样批处理大小 b 的三元组集合 $\mathcal{T}_{\text{batch}} \in \mathcal{T}$
4. for $(h, r, t) \in \mathcal{T}_{\text{batch}}$ do
5. 生成错误三元组集合 $\mathcal{T}'_{\text{batch}}$
其中 $\mathcal{T}'_{\text{batch}} = \{(h_1', r, t_1'), (h_2', r, t_2'), \dots\}$
且满足 $(h_1', r, t_1') \notin \mathcal{T}$
6. end for
7. 基于学习率 α 优化式 (3)–(6) 之一更新 \mathbf{E} 和 \mathbf{R}
8. end loop

2.5 评测任务

知识图谱推理模型的优劣需通过不同评测任务进行评价分析, 常见的评测任务包括链接/实体预测、关系预测以及三元组分类, 具体内容如下。

(1) 链接/实体预测。链接预测 (Link Prediction) 或实体预测 (Entity Prediction) 任务是知识图谱推理研究中的经典评测任务, 旨在预测知识图谱中缺失的实体。具体地, 链接预测任务要求模型能够基于三元组 (h, r, t) 中已知的两个对象正确预测出另一对象, 例如根据头实体和关系 (h, r) 能够正确预测出尾实体 t 。

(2) 关系预测。关系预测 (Relation Prediction) 任务与实体预测任务相似, 旨在预测知识图谱中缺失的关系。要求模型能够根据三元组中的头尾实体对 (h, t) 正确推理出它们的关系 r 。

(3) 三元组分类。三元组分类 (Triplet Classification) 旨在判断指定三元组 (h, r, t) 的对错, 其本质上属于一个二分类任务。例如需要判断三元组 (姚明, 职业, 篮球运动员) 是否正确, 也相当于回答了“姚明的职业是否为篮球运动员?” 这个问题。三元组分类能够对实体和关系向量的质量进行评估。

3 知识图谱推理典型方法

3.1 平移距离方法

平移距离 (Translational Distance) 模型将关系看作头实体到尾实体在向量空间的平移。TransE 是最具代表性的基于平移距离的推理模型, 其具有计算效率高特点, 从而被广泛应用于各项基于知识图谱的任务。TransE 模型由 Bordes 等^[24] 于 2013 年提出, 如图 3 所示, Bordes 等受到词表示学习模型 Word2vec^[34] 中语义平移现象的启发, 将知识图谱中的实体和关系的语义信息映射至多维向量空间, 并将关系向量表示为实体向量间的平移操作。在词表示学习中, 词向量之间存在着语义平移的现象, 例如 (姚明, 叶莉) 与 (丈夫, 妻子) 的空间距离往往较为接近。在知识图谱推理任务中, TransE 模型将这种潜在的语义关系直接用关系向量进行表示, 即将关系“职业”表示为头实体“姚明”到尾实体“篮球运动员”的空间向量平移操作。

基于以上平移距离假设, TransE 模型假设每个三元组 (h, r, t) 中的实体与关系向量之间满足 $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ 的条件。形式化地, 模型对三元组 (h, r, t) 定义了如下评分函数:

$$\phi(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{L_1 L_2} \quad (8)$$

该评分函数即 $\mathbf{h} + \mathbf{r}$ 向量和 \mathbf{t} 向量的 L_1 或 L_2 距离。在实际训练中, 平移距离模型使用最大间隔损失 (Margin Loss)

函数进行优化,具体如下:

$$\mathcal{L} = \sum_{(h,r,t) \cup (h',r',t')} \max(0, \| \mathbf{h} + \mathbf{r} - \mathbf{t} \|_{L_1 L_2} - \| \mathbf{h}' + \mathbf{r} - \mathbf{t}' \|_{L_1 L_2} + \gamma) \quad (9)$$

其中, (h, r, t) 为正确三元组, (h', r', t') 为错误三元组, 即将正确三元组中的实体通过随机选取的方式替换, 生成不存在于知识图谱中的错误三元组; 超参数 γ 是正错三元组评分的间隔距离; $\max(0, x)$ 为最大间隔函数。平移距离模型通过随机梯度下降算法进行优化训练, 通过最大化正错三元组之间的距离来优化实体与关系的分布式表示, 从而进行任务推理。

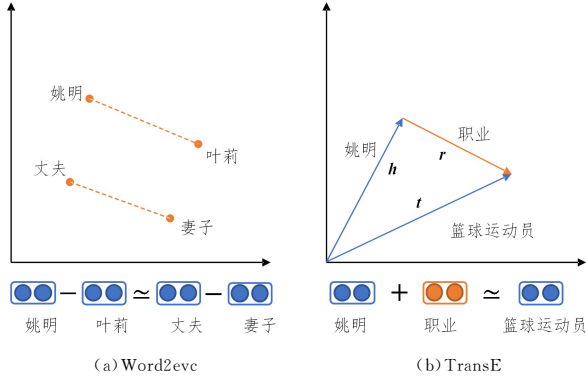


图3 Word2vec 和 TransE 模型示意图

Fig. 3 Schematic diagram of Word2vec and TransE

虽然 TransE 模型简单高效并在知识图谱推理上已经表现出很好的效果, 但是这种方法并不能很好地对复杂关系 (ComplexRelation) 进行建模。根据 TransE 模型的定义, 知识图谱中的关系类别可以分为一对一 (1-to-1)、一对多 (1-to-N)、多对一 (N-to-1) 和多对多 (N-to-N)。针对每个关系 r , 需要计算每个尾实体对应头实体的平均个数 hpt_r 和每个头实体对应尾实体的平均个数 tph_r , 于是 4 种关系可以分类为:

$$\begin{cases} hpt_r < \eta \text{ and } tph_r < \eta \Rightarrow 1\text{-to-}1 \\ hpt_r < \eta \text{ and } tph_r \geq \eta \Rightarrow 1\text{-to-}N \\ hpt_r \geq \eta \text{ and } tph_r < \eta \Rightarrow N\text{-to-}1 \\ hpt_r \geq \eta \text{ and } tph_r \geq \eta \Rightarrow N\text{-to-}N \end{cases} \quad (10)$$

其中, η 一般设置为 1.5。以一对多关系为例, 在该类型关系中的单个头实体会存在多个尾实体与其组成正确的三元组, 例如 (姚明, 队友, 易建联)、(姚明, 队友, 王治郅), 如果按 TransE 模型的平移距离假设, 会出现实体“易建联”与实体“王治郅”的语义向量在空间中相等的情况, 这显然是不符合逻辑的。为解决对复杂关系建模的问题, 一系列基于平移模型的改进方法相继被提出, 以求改进平移模型对复杂关系的建模效果。这些模型关注于如何从知识图谱结构中学习更好的语义信息, 并通过增加实体及关系语义表示的多样性来改进 TransE 模型。本文选取其中具有代表性的模型进行详细阐述, 具体如下。

TransH^[35] 模型认为对于同一实体, 在不同的关系下会有不一样的语义表示, 例如“姚明”在面对“父母”关系和面对“女儿”关系时所表达的语义特征应当是不同的。基于此动机, TransH 模型提出将头实体和尾实体的语义向量投影到一个关于关系的超平面 (Hyperplane), 然后再进行平移操作。具体地, 如图 4 所示, TransH 首先将头实体语义向量 \mathbf{h} 和尾实体

语义向量 \mathbf{t} 投影到关系 r 所在的超平面上, 其过程如下:

$$\begin{cases} \mathbf{h}_r = \mathbf{h} - \mathbf{w}_r^T \mathbf{h} \mathbf{w}_r \\ \mathbf{t}_r = \mathbf{t} - \mathbf{w}_r^T \mathbf{t} \mathbf{w}_r \end{cases} \quad (11)$$

其中, \mathbf{w}_r 为垂直于关系 r 所在超平面的法向量, \mathbf{h}_r 与 \mathbf{t}_r 分别为头实体向量 \mathbf{h} 与尾实体向量 \mathbf{t} 在超平面上的投影向量。因此, TransH 模型的评分函数可表示为:

$$\phi(h, r, t) = \| \mathbf{h}_r + \mathbf{r} - \mathbf{t}_r \|_{L_2} \quad (12)$$

通过语义超平面的投影映射, TransH 在训练时对于头尾实体向量均有不同表示, 这在一定程度上缓解了复杂关系建模困难问题。

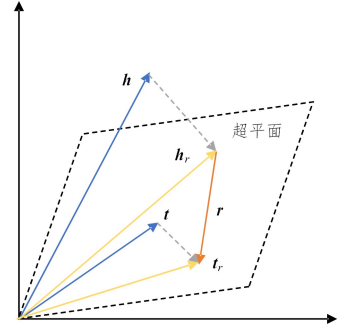


图4 TransH 模型示意图

Fig. 4 Schematic diagram of TransH

TransR^[36] 模型则认为实体和关系应属于不同的语义空间, 并通过投影矩阵将实体向量投影到关系向量所在的语义空间中进行学习表示。具体地, 如图 5 所示, TransR 首先将实体的语义向量从实体空间投影到关系空间中, 其过程如下:

$$\begin{cases} \mathbf{h}_r = \mathbf{M}_r \mathbf{h} \\ \mathbf{t}_r = \mathbf{M}_r \mathbf{t} \end{cases} \quad (13)$$

其中, \mathbf{M}_r 为关系 r 所在的语义空间的投影矩阵。因此, TransR 模型的评分函数可表示为:

$$\phi(h, r, t) = \| \mathbf{M}_r \mathbf{h} + \mathbf{r} - \mathbf{M}_r \mathbf{t} \|_{L_2} \quad (14)$$

通过将实体与关系分配到不同空间, TransR 推理性能在 TransH 的基础上得到了进一步的提升, 但由于引入了空间投影矩阵 \mathbf{M}_r , TransR 模型的参数大幅增加, 计算复杂度也大大提高。此外, 研究人员从实体空间到关系空间的投影矩阵仅仅与关系有关是不合理的, 还应与实体有关。

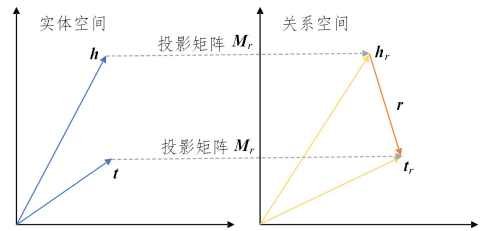


图5 TransR 模型示意图

Fig. 5 Schematic diagram of TransR

如图 6 所示, TransD^[37] 模型为解决上述 TransR 模型存在的不足, 对于给定三元组 (h, r, t) , 设置了两个投影矩阵 $\mathbf{M}_{r,h}$ 和 $\mathbf{M}_{r,t}$, 用于将头实体和尾实体投影到关系空间, 具体定义如下:

$$\begin{cases} \mathbf{M}_{r,h} = \mathbf{h}_p \mathbf{r}_p + \mathbf{I} \\ \mathbf{M}_{r,t} = \mathbf{t}_p \mathbf{r}_p + \mathbf{I} \end{cases} \quad (15)$$

其中, $\mathbf{h}_p, \mathbf{r}_p, \mathbf{t}_p$ 分别代表与头实体 h 、关系 r 、尾实体 t 有关的投影向量, \mathbf{I} 为单位向量。最后, TransD 模型定义了如下评分函数:

$$\phi(h, r, t) = \|\mathbf{M}_{r,h}\mathbf{h} + \mathbf{r} - \mathbf{M}_{r,t}\mathbf{t}\|_{L_2} \quad (16)$$

此外, TransD 模型为减少模型的参数, 将投影矩阵设置成对角矩阵, 解决了 TransR 模型参数数量过多的问题, 并且在计算复杂度上与 TransH 模型相当, 在实验结果上相比之前的模型也得到了提升。

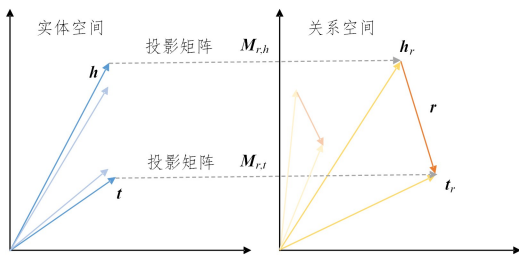


图6 TransD模型示意图

Fig. 6 Schematic diagram of TransD

RotatE^[38]模型是近年来提出的极具新意的知识图谱推理方法, 如图7所示, 该模型将每个关系定义为在复数空间(Complex Space)中从头实体到尾实体的旋转(Rotate), 在相关推理任务上明显优于之前的模型, 在WN18RR数据集上首次将Hit@10指标提升到57%以上。RotatE模型独辟蹊径, 将知识图谱中的常见的关系分类为对称关系(Symmetric, 如“夫妻”)、互逆关系(Inversion, 如“父亲”与“孩子”)以及组合关系(Composition, 如“父亲”的“父亲”可组合为“祖父”)。为对上述关系类型建模, RotatE模型将实体和关系映射到复数空间, 并将每个关系 r 定义为从头实体 h 到尾实体 t 的空间旋转, 定义了如下评分函数:

$$\phi(h, r, t) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|_{L_2} \quad \text{where } \mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^d \quad (17)$$

其中, \circ 代表哈达玛乘积(Hadamard Product), \mathbb{C} 代表复数向量空间。结果表明, RotatE模型能够很好地建模上述3种关系类型, 并在计算复杂度上呈线性增长, 可扩展到大型知识图谱推理任务。

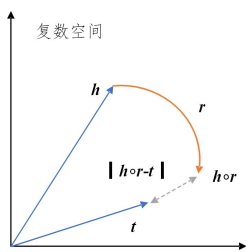


图7 RotatE模型示意图

Fig. 7 Schematic diagram of RotatE

除上述代表性模型以外, 研究人员基于平移距离思想, 还提出了许多新颖的知识图谱推理模型。TransAH^[39]在利用TransH超平面的思想上, 引入加权欧氏距离从而更加灵活地区分正错三元组。TransSparse^[40]为解决知识图谱中存在的异质性和不平衡性问题, 提出使用稀疏矩阵替代TransR中的稠密矩阵。TransG^[41]与KG2E^[42]则认为知识图谱中的实体与关系会存在多重语义, 提出使用高斯分布来表示实体与关系。IWM^[43]通过对实体语义向量进行适当的加权处理, 来实现对实体语义的区分。ManifoldE^[44]以及TorusE^[45-46]

分别将流形(Manifold)和李群(LieGroup)等数学方法扩展到平移距离方法, 从而缓解TransE中 $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ 的约束。KBGAN^[47]针对错误三元组生成质量不高的问题, 提出使用生成对抗网络生成负样本。

3.2 语义匹配方法

语义匹配(Semantic Matching)模型利用了基于相似性的评分函数, 通过匹配实体和关系的潜在语义来衡量知识三元组的准确性, 也被称为矩阵分解(Matrix Factorization)或双线性(Bilinear)模型。作为典型的语义匹配模型, 如图8所示, RESCAL^[48]使用向量表示实体的潜在语义, 并用基于关系的矩阵对这些潜在语义进行交互建模。针对每个知识三元组 (h, r, t) , RESCAL定义了如下评分函数:

$$\phi(h, r, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t} = \sum_{i=1}^d \sum_{j=1}^d [\mathbf{M}_r]_{ij} \cdot [\mathbf{h}]_i \cdot [\mathbf{t}]_j \quad (18)$$

其中, $\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$ 代表 d 维语义向量, $\mathbf{M}_r \in \mathbb{R}^{d \times d}$ 是与关系 r 相关的矩阵。上述评分函数计算了头实体 h 与尾实体 t 在关系 r 情况下的语义匹配程度, 该值越大表示 (h, r, t) 组成正确知识三元组的置信度越高。

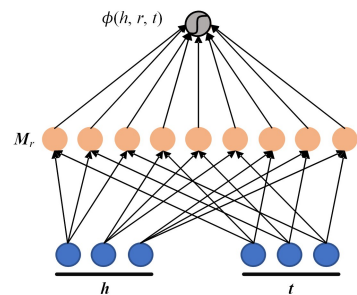


图8 RESCAL模型示意图

Fig. 8 Schematic diagram of RESCAL

可以发现, RESCAL模型的思路简单粗暴, 通过优化一个多阶的三阶张量来学习实体及关系的语义表示。但在面对大规模知识图谱时, 随着实体及关系数量的急剧增加, 需要优化的张量维数也在不断加大, 从而导致RESCAL模型的计算效率大幅下降。为解决上述问题, 一系列基于RESCAL的扩展模型相继被提出, 以求解优化模型的计算复杂度, 其中具有代表性的模型如下。

DistMult^[49]模型通过将矩阵 \mathbf{M}_r 约束为对角矩阵(Diagonal Matrix)来简化RESCAL模型。如图9所示, 对于每个关系 r , 该模型同时生成一个代表关系语义的向量 \mathbf{r} 和一个用于匹配语义的对角矩阵 $\text{diag}(\mathbf{r})$ 。因此, DistMult模型的评分函数被定义为:

$$\phi(h, r, t) = \mathbf{h}^T \text{diag}(\mathbf{r}) \mathbf{t} = \sum_{i=1}^d [\mathbf{h}]_i \cdot [\mathbf{r}]_i \cdot [\mathbf{t}]_i \quad (19)$$

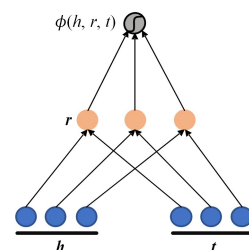


图9 DistMult模型示意图

Fig. 9 Schematic diagram of DistMult

此评分函数只计算了头实体 h 与尾实体 t 在关系 r 情况下同一维度的语义匹配程度,相比 RESCAL 模型,DistMult 模型在参数量大幅减少的情况下仍能提高计算效率。但是,该模型仍存在明显不足,例如三元组(姚明,职业,篮球运动员)的评分结果 $\mathbf{h}^T \text{diag}(\mathbf{r}) \mathbf{t}$ 与三元组(篮球运动员,职业,姚明)的评分结果 $\mathbf{t}^T \text{diag}(\mathbf{r}) \mathbf{h}$ 在模型的假设下完全相等,但后一个知识三元组显然是错误的。

如图 10 所示,HolE^[50] 模型结合 RESCAL 模型的方法思路和 DistMult 的计算效率优势,提出使用全息向量(Holographic Embedding)来学习实体及关系的语义表示,其评分函数的定义如下:

$$\phi(h, r, t) = \mathbf{r}^T (\mathbf{h} * \mathbf{t}) \quad (20)$$

其中,

$$[\mathbf{h} * \mathbf{t}]_i = \sum_{k=1}^d [\mathbf{h}]_k \cdot [\mathbf{r}]_{(k+i) \bmod d} \quad (21)$$

式(20)、式(21)中,符号 $*$ 代表循环相关操作,可看作矩阵乘法的特殊形式,mod 指求余函数(即两个数值作除法运算后的余数)。其中,循环相关操作存在不可逆性,即 $\mathbf{h} * \mathbf{t} \neq \mathbf{t} * \mathbf{h}$,因此解决了 DistMult 模型在非对称关系建模中存在的不足。HolE 在模型可解释性方面仍有待加强,但是也为基于语义匹配的推理模型提供了全新的解决思路,值得关注。

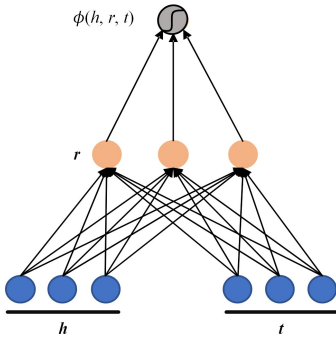


图 10 HolE 模型示意图

Fig. 10 Schematic diagram of HolE

Complex^[51] 模型最先将复数空间引入知识图谱推理任务。考虑到 DistMult 模型中实数向量之间的点积计算具有交换性(使用实数不好处理非对称关系),该模型采用复数空间来建模不对称关系,同时保留点积的效率优势,即空间和和时间复杂度均为线性。相应地,定义了如下评分函数:

$$\phi(h, r, t) = \text{Re}(\mathbf{h}^T \text{diag}(\mathbf{r}) \bar{\mathbf{t}}) \text{ where } \mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^d \quad (22)$$

具体如下:

$$\begin{aligned} \phi(h, r, t) = & \sum_{i=1}^d (\text{Re}(\mathbf{h}_i) \cdot \text{Re}(\mathbf{r}_i) \cdot \text{Re}(\mathbf{t}_i) + \text{Re}(\mathbf{h}_i) \cdot \\ & \text{Im}(\mathbf{r}_i) \cdot \text{Im}(\mathbf{t}_i) + \text{Im}(\mathbf{h}_i) \cdot \text{Re}(\mathbf{r}_i) \cdot \text{Im}(\mathbf{t}_i) - \\ & \text{Im}(\mathbf{h}_i) \cdot \text{Im}(\mathbf{r}_i) \cdot \text{Re}(\mathbf{t}_i)) \end{aligned} \quad (23)$$

其中, $\text{Re}(x)$ 代表复数的实部, $\text{Im}(x)$ 代表复数的虚部。在实验方面,模型简洁的 ComplEx 在多个推理数据集上都取得了不错的效果,明显优于当时表现优异的 HolE 模型。

Tucker^[52] 模型提出了一种基于塔克分解(Tucker Decomposition)的知识图谱推理方法,并证明了其在推理性能提升的同时能够大幅减少参数量。该模型认为知识图谱中许多实体及关系之间的语义特征是相同的(例如实体“姚明”与实体“易建联”之间的性别、职业、国籍等语义特征是相同的),

因此可以用一个核心张量来代替这些共享语义特征,并定义了如下评分函数:

$$\phi(h, r, t) = \mathbf{W} \times_1 \mathbf{h} \times_2 \mathbf{r} \times_3 \mathbf{t} \quad (24)$$

其中, $\mathbf{W} \in \mathbb{R}^{d \times k \times d}$ 代表共享的语义张量, $\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$ 代表每个实体特有的 d 维语义向量, $\mathbf{r} \in \mathbb{R}^k$ 代表每个关系特有的 k 维语义向量,最后 Tucker 模型使用一个逻辑回归(Logistic)函数 $\sigma(\phi(h, r, t))$ 来得到知识三元组 (h, r, t) 的置信度,结果为 1 代表正确,0 代表错误。

通过证明,Tucker 模型具有充分表达性,一些主流语义匹配模型(如 RESCAL, DistMult, ComplEx)均可视为 Tucker 在特定情况下的案例。此外,该模型的参数呈线性增长,其核心张量 \mathbf{W} 的参数规模只会随设置的语义向量维度呈线性变化,与实体及关系的数量无关。最后在实验中,即使给定的语义向量维度较小,Tucker 模型也能取得很好的推理效果。当向量维度为 20 时,在 FB15k-237 数据集上相比 ComplEx 模型 MRR 评价指标提升了 4.2%。

除上述代表性模型以外,研究人员基于语义匹配思想,还提出了许多新颖的知识图谱推理模型。ANALOGY^[53] 在 RESCAL 模型的基础上,引入类比结构(Analogical Structure),从而限制推理结果的范围进而提高推理准确性。Simple^[54] 针对在学习头尾实体语义向量时存在互相独立的问题,引入逆关系(Inverse Relation)使得学习的头尾实体语义向量相关依赖。ComplEx-N3^[55] 首次引入三阶正则化项,用于防止基于语义匹配方法普遍存在的过拟合问题。Dihedral^[56] 认为现有的语义匹配方法忽视了对关系组合的建模,从而导致知识图谱推理缺乏可解释性,首次尝试使用有限非阿贝尔群(Finite Non-Abelian Group)来建模关系组合。MuRP^[57] 观察到知识图谱经常具有复杂的层次结构,因此提出在双曲空间(Hyperbolic Space)学习实体及关系的语义向量。QuatE^[58] 则进一步拓展了 ComplEx 模型,提出在四元数空间(Quaternion Space)为实体和关系建模,并采用哈密顿算子(Hamilton Operator)进行语义计算。

3.3 知识图谱推理典型方法总结

本节介绍了知识图谱推理的典型方法,包括以 TransE 模型为代表的平移距离方法和以 RESCAL 模型为代表的语义匹配方法。如表 4 所列,本文对上述方法的评分函数和实体及关系特征维度进行了比较。其中,基于平移距离的推理模型把关系看作头实体和尾实体在语义空间的平移操作,具有计算效率高及应用范围广等特点,后续研究从复杂关系建模(TransH, TransR, TransD, TransAH)、语义差异(TransG, TransSparse, KG2E, IWM)、数学表达优化(ManifoldE, RotatE, TorusE)、负样本生成(KBGM)等角度进行改进,均取得了显著的性能提升;基于语义匹配的推理模型通过计算实体和关系的潜在语义来判断三元组的准确性,在模型拟合速度和特征学习效率方面具有明显优势。后续研究一方面从计算复杂度角度进行优化(DistMult, HolE, Tucker),另一方面通过引入附加约束(ANALOGY, Simple, ComplEx-N3, Dihedral)来提高推理性能。此外,研究人员尝试引入更高语义维度空间(ComplEx, MuRP, QuatE 等)对实体及关系建模,但在参数规模方面有待优化。

表 4 典型知识图谱推理方法比较

Table 4 Comparison of representative knowledge graph reasoning methods

分类	模型	评分函数 $\phi(h, r, t)$	实体特征	关系特征
平移距离方法	TransE	$\ h+r-t\ _{L_1/L_2}$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^d$
	TransH	$\ h-w_r^T h w_r + r - t - w_r^T t w_r\ _{L_2}$	$h, t \in \mathbb{R}^d$	$r, w_r \in \mathbb{R}^d$
	TransR	$\ M_r h + r - M_r t\ _{L_2}$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^k, M_r \in \mathbb{R}^{k \times d}$
	TransD	$\ M_{r,h} h + r - M_{r,t} t\ _{L_2}$	$h, t \in \mathbb{R}^d$	$r, M_{r,h}, M_{r,t} \in \mathbb{R}^d$
	RotatE	$\ h \cdot r - t\ _{L_2}$	$h, t \in \mathbb{C}^d$	$r \in \mathbb{C}^d$
语义匹配方法	RESCAL	$h^T M_r t$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^d, M_r \in \mathbb{R}^{d \times d}$
	DistMult	$h^T \text{diag}(r) t$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^d$
	HolE	$r^T (h \times t)$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^d$
	ComplEx	$\text{Re}(h^T \text{diag}(r) \bar{t})$	$h, t \in \mathbb{C}^d$	$r \in \mathbb{C}^d$
	TuckER	$W \times_1 h \times_2 r \times_3 t$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^k, W \in \mathbb{R}^{d \times k \times d}$

除上述研究以外,基于表示学习的知识图谱推理仍面临许多挑战:1)如何融合多源信息辅助构建更好的知识表示,从而提高知识图谱推理相关任务的性能;2)如何利用当前流行的基于深度学习的神经网络方法来提高模型的语义特征提取及学习能力。因此,接下来本文将重点介绍融合多源信息的知识图谱推理方法和基于神经网络的知识图谱推理方法。

4 融合多源信息的知识图谱推理方法

融合多源信息可以辅助构建更好的知识表示,从而提高知识图谱推理相关任务的性能。传统的知识图谱推理模型往往重点关注知识图谱自身的高度结构化信息,忽略了丰富的跨模态多源信息。随着互联网的快速发展,海量的文本、音频、图片和视频数据被上传并共享,如何有效和高效地利用从文本到视频等多源信息正成为知识图谱推理中一个关键且具有挑战性的问题。在融合多源信息的知识图谱推理方面,已经有一些研究工作,本节将重点介绍以下代表性方法。

4.1 文本描述

DKRL^[59]模型提出了一种基于实体描述的知识图谱推理模型,旨在利用 Freebase 知识图谱中丰富的实体描述信息。如图 11 所示,DKRL 模型根据头尾实体的描述信息生成头尾实体的向量表示,然后基于 TransE 模型的损失函数进行模型训练。具体地,DKRL 模型定义了如下评分函数:

$$\phi(h, r, t) = \|h_s + r - t_s\|_{L_1} + \|h_d + r - t_d\|_{L_1} + \|h_s + r - t_d\|_{L_1} + \|h_d + r - t_s\|_{L_1} \quad (25)$$

其中, h_s/h_d 和 t_s/t_d 分别为基于结构的 (Structure-based) 和基于描述的 (Description-based) 头尾实体特征向量。前者用于建模知识图谱中客观存在的结构信息,后者用于融合实体描述的文本特征。在文本特征表示方面,DKRL 模型采用了两种不同的方法,分别为连续词袋方法和卷积神经网络方法。

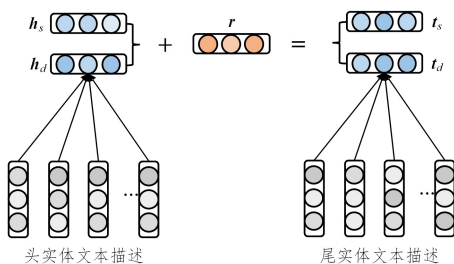


图 11 DKRL 模型示意图

Fig. 11 Schematic diagram of DKRL

进行训练优化,实验结果表明该模型相比 TransE 模型在推理任务上具有显著的性能提升。此外,DKRL 模型有助于实现对新增实体的表示。面对知识图谱中未曾出现的实体时,借助于实体的简短文本描述,该模型能够生成实体的特征表示,这对于知识图谱的补全具有重要意义。

TEKE^[60]模型针对知识图谱推理中的复杂关系和数据稀疏等问题,提出将丰富的文本信息融合到实体的特征表示中,并且每一个关系对于不同的头实体和尾实体都有不同的表示,从而更好地解决一对多、多对多关系。TEKE 模型主要分为以下 3 个步骤:首先,对文本描述中的实体进行标注,构建一个由实体和描述词组成的共现网络 (Co-occurrence Network);然后,对于每个实体 h 和 t ,在共现网络中定义其上下文集合 $n(h)$ 和 $n(t)$ 并构建其上下文嵌入 (Context Embedding) $n(h)$ 和 $n(t)$;最后,通过融入上下文嵌入来学习文本增强的实体及关系表示。其计算过程如下:

$$\begin{cases} \hat{h} = \mathbf{A}n(h) + h \\ \hat{r} = \mathbf{B}n(h, t) + r \\ \hat{t} = \mathbf{A}n(t) + t \end{cases} \quad (26)$$

其中, \mathbf{A} 和 \mathbf{B} 代表权重矩阵, h, r, t 分别表示头实体、关系和尾实体的偏置向量。基于平移距离假设,可以得到 TEKE 模型的评分函数:

$$\phi(h, r, t) = \|\hat{h} + \hat{r} - \hat{t}\|_{L_2} \quad (27)$$

通过将文本描述的上下文嵌入融入到知识图谱语义空间中,TEKE 模型能够很好地解决知识图谱中实体多、关系少的数据稀疏问题。此外,对于给定的关系,关系特征的表示也会根据不同的头实体和尾实体发生相应变化,使得该模型在处理复杂关系推理时表现依旧出色。

此外,ATE^[61]模型利用实体描述和特定关系来增强语义表示,提出关系和实体描述间的相互注意机制来学习更准确的文本表示。KEC^[62]模型通过引入微软概念图谱 (Microsoft Concept Graph) 中的概念信息,将实体和概念联合嵌入到语义空间来增强实体的特征向量学习。RLKB^[63]模型计算每个三元组的概率以及相应的实体描述,同时学习描述中实体、关系和文本的上下文嵌入。SSP^[64]模型提出在关于文本的语义空间中学习特征向量,以建模三元组与文本描述之间的强相关性。上述模型均提出将文本描述信息融合到实体及关系的特征表示中,说明文本描述信息能够有效提升知识图谱推理相关任务的性能,特别是可以有效处理新实体的表示问题。

在模型训练方面,DKRL 模型基于最大间隔损失函数

但是,也可以发现,上述模型在对文本特征进行表征时所使用的方方法仍较为传统,学习到的文本语义表示质量不高。近年来,随着 Transformer^[65]模型在自然语言处理与计算机视觉领域的成功应用,研究人员尝试将该模型与知识图谱推理任务相结合。一方面,CoKE^[66]模型和 KG-BERT^[67]模型均采用 Transformer 编码器学习实体及关系包含上下文语义信息的特征表示,另一方面,可利用 Transformer 模型生成更为精准的文本表示结果,有助于文本描述信息在知识图谱推理任务上的高效利用。

4.2 路径组合

为解决 TransE 模型对每个三元组进行孤立评分的局限,PtransE^[68]模型提出了考虑关系路径的知识图谱推理方法。在知识图谱中,头实体到尾实体间往往存在多跳(Multi-hop)关系路径,可表示为 $h \rightarrow r_1 \rightarrow \dots \rightarrow r_n \rightarrow t$,这些关系路径可为推理提供丰富的语义信息和可解释性。如图 12 所示,PtransE 模型考虑建立多步关系路径的向量表示 p 来建模从头实体 h 到尾实体 t 的平移距离。具体地,PtransE 模型通过对路径上的所有关系向量进行语义组合形成路径向量,并采用 3 种代表性的语义组合操作,其计算过程如下:

$$\begin{cases} \text{Addition: } p = r_1 + \dots + r_n \\ \text{Multiplication: } p = r_1 \circ \dots \circ r_n \\ \text{RNN: } c_i = f(W[c_{i-1}; r_i]) \end{cases} \quad (28)$$

其中,Addition, Multiplication, RNN 分别代表语义相加、语义相乘和循环网络(Recurrent Neural Network),符号 \circ 为哈达玛乘积(Hadamard Product), c_i 为第 i 个关系的累积路径向量, W 为共享参数矩阵,函数 $[\cdot]$ 和 $f(\cdot)$ 分别代表向量拼接和非线性映射。PtransE 同样基于平移距离的模型假设,确定了如下评分函数:

$$\phi(h, r, t) = \|h + p - t\|_{L_2} \quad (29)$$

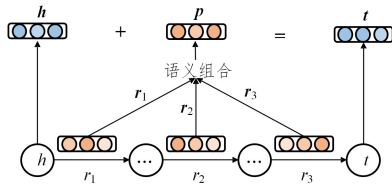


图 12 PTransE 模型示意图

Fig. 12 Schematic diagram of PTransE

在知识图谱数据集上的实验结果表明,PtransE 模型能够极大提升知识图谱推理相关任务的性能。其中关系路径向量的语义相加在 3 种语义组合操作中的效果最好。此外,考虑到两步关系路径就可以达到最佳结果,随着关系路径增加,实验性能反而降低,该结果表明在推理过程中没有必要考虑太长的关系路径。

此外,PaSKoGE^[69]模型考虑了不同关系路径之间的差异,对给定头尾实体间的关系和多步关系路径的相关性进行编码,自适应地确定其基于关系路径的损失函数。TKGE^[70]模型提出了一种增强的 LSTM 模型来建模任意长度的关系路径特征,并结合路径中的实体和关系,利用基于注意力(Attention)的多实例学习方法对实体语义向量进行建模。

PBKG^[71]模型将双向长短期记忆(Bi-LSTM)、卷积神经网络与注意机制相结合,建立头尾实体之间的关系路径的语义相关性。上述模型均提出利用头尾实体间的关系路径来提高知识图谱推理的准确性,均取得了显著的效果,进一步说明关系路径中蕴藏着丰富的信息。但是,关于关系路径建模的工作还处于比较初级的阶段,相关工作较少,在关系路径的语义组合、关系路径的权重选择等方面还有待深入研究。

4.3 时序特征

面对动态知识图谱^[72]推理研究需要重点关注如何对时序特征进行建模与表示。现有的知识图谱推理模型在对知识三元组进行建模时,往往忽略了知识图谱中包含的时序信息。动态知识图谱的基本组成单位可表示为 (h, r, t, τ) ,其中 τ 代表时间戳(Timestamps),例如(C 罗, 俱乐部, 皇家马德里, 2009-2018)。基于 TransE 模型, TTransE 模型^[73]将时序特征也融合到头实体到尾实体在语义空间的平移,并定义了如下评分函数:

$$\phi(h, r, t, \tau) = \|h + r + \tau - t\|_{L_1/L_2} \quad (30)$$

其中, τ 表示时序的特征向量。同样地, TTransE 模型基于最大间隔损失(Margin Loss)函数进行训练优化,从而得到所有实体、关系及时序信息的特征向量。由于 TTransE 模型具有计算效率高等特点,从而被广泛应用于各项基于动态知识图谱的表示与应用。

HyTE^[74]模型借鉴 TransH 构建超平面的思想,如图 13 所示,提出将实体及关系首先投影到基于时序 τ 的超平面(Hyperplane),从而进行空间的平移距离操作,其计算过程如下:

$$\begin{cases} h_\tau = h - w_\tau^T h w_\tau \\ r_\tau = r - w_\tau^T r w_\tau \\ t_\tau = t - w_\tau^T t w_\tau \end{cases} \quad (31)$$

其中, w_τ 表示基于 τ 的超平面的法向量,则 h_τ, r_τ, t_τ 分别表示头实体、关系及尾实体在超平面上的投影向量,因此基于平移距离的模型假设可以得到 HyTE 模型的评分函数:

$$\phi(h, r, t, \tau) = \|h_\tau + r_\tau - t_\tau\|_{L_1/L_2} \quad (32)$$

实验结果表明,该模型与传统模型或者同类模型相比都有着突出的表现。此外,HyTE 模型不仅能够利用时序特征进行知识图谱推理,还能够为知识图谱中缺失时序特征的三元组进行时间范围预测。

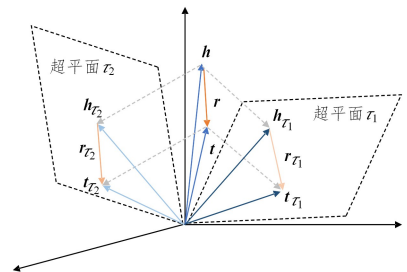


图 13 HyTE 模型示意图

Fig. 13 Schematic diagram of HyTE

TA^[75]模型为融合时序特征,利用循环神经网络来学习带有时序感知的关系特征表示。具体地,如图 14 所示,该模型

基于关系向量 r 与时序特征向量 τ , 通过传入神经网络生成带有时序感知的关系向量 r_τ 。为减小时序特征向量的规模和更好地区分时序差异, TA 模型将时序特征向量 τ 分解为年 τ_y (year)、月 τ_m (month)、日 τ_d (day) 3 个标记(Token)。于是, TA 模型定义了如下两种评分函数:

$$\begin{cases} \text{TA-TransE: } \phi(h, r, t, \tau) = \|h + r_\tau - t\|_{L_2} \\ \text{TA-DistMult: } \phi(h, r, t, \tau) = h^T \text{diag}(r_\tau) t \end{cases} \quad (33)$$

TA-TransE 模型与 TA-DistMult 模型分别是基于平移距离和语义匹配的模型假设, 将带有时序感知的关系向量 r_τ 看作头实体到尾实体在语义空间的平移距离或语义匹配操作。实验结果表明, 上述模型相比其原始模型均有显著的性能提升, 验证了这些方法的有效性。

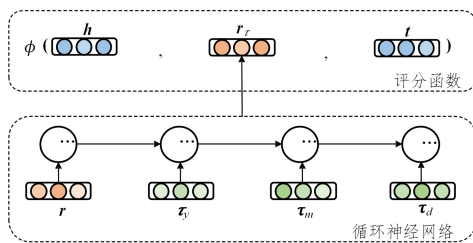


图 14 TA 模型示意图

Fig. 14 Schematic diagram of TA

DE^[76] 模型提出为每个实体学习一个历时的向量(Diachronic Embedding), 用于建模实体特征随时间变化而改变, 提供实体在任何时间点的特征表示, 例如个人的性格会随时间的推移而改变。基于此动机, DE 模型提出了历时实体向量函数, 以头实体 h 为例, 其在时间戳 τ 时的特征向量 h^τ 计算过程如下:

$$h_n^\tau = \begin{cases} e_n f(w_n \tau_n + b_n), & \text{if } 1 \leq n \leq \gamma d \\ e_n, & \text{if } \gamma d < n \leq d \end{cases} \quad (34)$$

其中, $e \in \mathbb{R}^d$ 代表基于实体的语义向量; $w_n, b_n \in \mathbb{R}^{\gamma d}$ 分别表示基于时序的权重向量和偏执向量; $f(\cdot)$ 表示非线性激活函数 ReLU; $\gamma \in [0, 1]$ 是一个超参数, 用于确定随时序变化的特征比例, 例如当 $\gamma = 0.2, d = 100$ 时, 表示实体向量有 20 维特征会随时序改变而变化。

基于上述得到的实体向量, DE 模型同样基于平移距离和语义匹配的模型假设, 提出 DE-TransE, DE-DistMult 与 DE-Simple 这 3 种扩展模型。实验结果表明, DE-Simple 模型在多个时序动态知识图谱数据集上取得了 SOTA(State-of-the-art)性能。由于 DE 模型的假设非常高效新颖, 因此可以扩展到任何基于平移距离和语义匹配方法上。

此外, Know-Evolve^[77] 模型提出了一种深度进化知识网络, 它可以随时间学习非线性演化的实体向量表示。DyRep^[78] 模型构建了双时间尺度(Two-time scale)的深度学习模型, 可以从知识图谱中不断获取实体随时间改变的演化特征, 从而更有效地学习丰富的实体表示。RE-Net^[79] 将知识图谱的推理建模成基于以往知识并以时间序列为条件的概率分布, 使用循环神经网络编码器对过去的知识进行编码, 并使用邻域聚合器对同一时间戳的事实之间的连接进行建模。TIMEPLEX^[80] 利用三元组的周期性和成对关系之间的时间

交互, 学习和利用所有关系的软排序和跨度约束。TComplex^[81] 模型受语义匹配 Complex 模型的启发, 提出了一种基于四阶张量分解的时序动态知识图谱推理模型。CyGNet^[82] 基于时间感知复制生成机制(Time-aware Copy-generation Mechanism), 可以根据已知的事实预测未来的事实。上述模型均针对带有时序特征的知识图谱推理展开了研究, 其主要思路是将时间信息融合到知识图谱实体及关系的特征表示中。未来的主要研究方向在于如何更好地对时序特征进行表示学习, 以及生成更高质量的融合时序特征的实体及关系特征向量。

4.4 其他信息

除了上述 3 种附加信息外, 也有一些研究试图将其他信息纳入到知识图谱推理中, 用于增加模型的先验知识, 均在知识图谱推理上取得了不错的表现。

(1) 属性类别。在知识图谱中, 实体往往具有多样的类别信息, 例如“姚明”和“C 罗”虽同属“运动员”, 但通过细分可以知道前者属于“篮球运动员”而后者属于“足球运动员”。因此, 实体在特定三元组中所突出的特征是不同的, 需要在不同情境下突出实体不同方面的属性。为解决以上问题, TKRL^[83] 模型基于平移模型的模型假设, 考虑了知识图谱的结构信息, 同时引入实体类别信息, 以构建更加精准的实体特征表示。具体地, TKRL 模型定义了如下评分函数:

$$\phi(h, r, t) = \|M_{r,h,1} h + r - M_{r,t,2} t\|_{L_2} \quad (35)$$

其中, $M_{r,h,1}$ 和 $M_{r,t,2}$ 分别表示基于头实体和尾实体的类型特化映射矩阵, 例如 $M_{r,h,1}$ 代表了由 h 在关系 r 上作为头实体时所突出的类型。此外, TKRL 模型提出了两种层次类型编码器, 用于构建实体的类型特化映射矩阵。SSE^[84] 模型有类似的思路, 它要求相同类型的实体在语义向量空间的距离相近, 并采用了两种流形学习算法来模拟平滑假设, 该模型可扩展到其他知识推理方法中。KR-EAR^[85] 模型对知识图谱中的复杂属性进行了建模, 以提高在面对复杂关系时的推理能力。HAKE^[86] 模型提出将实体特征映射到极坐标系统中来建模实体的层次类别信息。实验结果表明, 上述模型在知识图谱推理相关任务上均超过了基线模型, 证明了实体层次类型中蕴含了丰富的信息, 能够帮助建立更精确的实体向量表示。

(2) 图像信息。同文本描述信息一样, 有关实体的图像信息同样可用于增强实体的特征向量表示。基于此动机, IKRL^[87-88] 模型尝试融合实体的图像信息, 建立跨模态的实体表示方法, 以提高知识图谱推理有关任务的性能。具体地, 如图 15 所示, IKRL 模型使用图像编码器建立实体图像的表示, 然后引入注意力机制自动选择高质量的实体图像构建实体基于图像的向量表示, 最后基于平移距离的模型假设确定了如下评分函数:

$$\phi(h, r, t) = \alpha_1 \|h_s + r - t_s\|_{L_1} + \alpha_2 \|h_i + r - t_i\|_{L_1} + \alpha_3 \|h_s + r - t_i\|_{L_1} + \alpha_4 \|h_i + r - t_s\|_{L_1} \quad (36)$$

其中, h_s/h_i 和 t_s/t_i 分别为基于结构的(structure-based)和基于图像的(image-based)头尾实体特征向量, $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 是控制各项评分的超参数。同样地, IKRL 模型使用最大间隔函数进行模型训练, 实验结果验证了有关实体的图像蕴含丰富

视觉信息,有助于实体特征向量的表示。

(3)逻辑规则。最后,本节将介绍基于逻辑规则的知识图谱推理模型。现实世界中,许多知识之间存在着一定的逻辑规则,这些逻辑规则包含着丰富的背景信息,在知识获取和推理中得到了广泛的研究。例如,实体“姚明”和“叶莉”既可以用关系“妻子”也可以用关系“丈夫”进行链接。传统基于逻辑规则的推理方法根据知识图谱结构特点运用简单规则进行推理,有助于增加推理过程的可解释性。因此,研究人员通过总结和表示这些逻辑规则,来进一步提升知识图谱推理的性能。如图16所示,KALE^[89-90]模型通过联合建模三元组评分计算和逻辑规则组合来学习更有效的实体及关系向量表示。其中,在三元组评分部分定义了如下方法:

$$\phi(h, r, t) = 1 - \frac{1}{3\sqrt{d}} \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{L_1} \quad (37)$$

其中, d 表示向量的维度,因此 $\phi(h, r, t) \in [0, 1]$ 用于表示三元组的置信度。逻辑规则组合部分采用 t 范数模糊逻辑(t -norm fuzzy logics)进行建模。实验结果表明,KALE模型增强了对新增知识三元组的预测能力。未来的主要研究方向为融合其他类型的逻辑规则来提供更精准的约束信息,例如一阶逻辑(First-order Logic, FOL)规则。

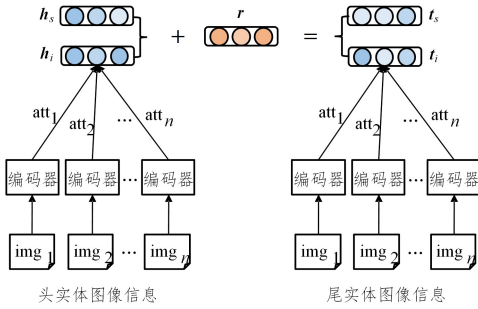


图15 IKRL模型示意图

Fig. 15 Schematic diagram of IKRL

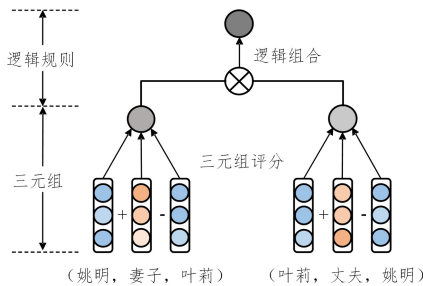


图16 KALE模型示意图

Fig. 16 Schematic diagram of KALE

4.5 融合多源信息知识图谱推理方法总结

本节主要对融合多源信息知识图谱推理模型进行介绍,通过结合文本描述、路径组合、时序特征以及其他附加信息,有助于改善数据稀疏问题,提高知识表示能力,从而进行有效推理。具体而言,融合文本描述模型主要从文本表示方法(DKRL, ATE, KEC)、上下文联合嵌入(TEKE, RLKB, SSP)两方面出发,将文本描述信息融合到实体及关系的特征表示中。以PTransE为代表的融合路径信息的推理模型关注对关系路径进行建模,并从关系的语义差异(PaSKoGE)及

权重选择(TKGE, PBKGC)等方面进行研究。对时序特征进行融合是面向动态知识图谱推理的关键问题,相关研究一方面关注时序特征的表示学习(TTransE, TA, TComplex),即采用平移距离、语义匹配或循环神经网络等方法来学习时序特征,并将其融合到实体及关系语义中;另一方面从实体特征会随时间变化而动态改变这一逻辑出发,学习带有时序感知的实体表示(HyTE, DE, DyRep, TIMEPLEX, Know-Evolve)。此外,最新工作(RE-Net, CyGNet)采用由历史知识推理未来知识的研究思路,为动态知识图谱推理提供了全新视角,值得关注。除上述多源信息外,研究人员也尝试将属性类别(TKRL, SSE, HAKE, KR-EAR, HAKE)、图像信息(IKRL)、逻辑规则(KALE)等信息融合到基于表示学习的知识图谱推理任务中,用于增加模型的先验知识。

5 基于神经网络的知识图谱推理方法

在大规模知识图谱中,一些浅层的推理模型常被用于做链接预测任务,但浅层的推理模型很难提取到深层次的语义特征,链接预测任务的效果无法得到进一步提升。为增加模型学习语义特征的能力,可以采用增加特征向量维度的方法,但不适用于大规模知识图谱,因为参数量与实体以及关系的个数成正比,并且这样容易造成过拟合现象。早期研究人员考虑将神经网络引入知识图谱推理任务,其中NTN^[91]使用神经张量网络进行知识图谱推理,主要思路是使用张量计算取代传统神经网络的线性变换。NAM^[92]利用传统深度神经网络来建立头尾实体间的语义关联。但上述模型在面对海量实体及关系时计算复杂度显著提高,不适用于大规模知识图谱的推理任务。为处理上述参数规模与计算效率的矛盾,研究人员开始研究设计参数高效、计算快速且具有较强泛化能力的知识图谱推理模型,本节将重点介绍以下代表性工作。

5.1 卷积神经网络

近来,以卷积神经网络(CNN)^[32, 93-94]为代表的深度学习方法得到了广泛关注,并在自然语言处理、计算机视觉等领域获得了突破性进展。受此启发,Detters等首次将卷积神经网络运用于知识图谱推理,提出了ConvE模型,如图17所示。ConvE^[26]模型主要分为特征组合、特征提取以及特征匹配3个部分,其中特征组合旨在将头实体及关系的语义特征向量拼接成一个二维矩阵,然后通过卷积神经网络进行特征提取,最后通过全连接及映射操作再与尾实体的语义特征向量进行特征匹配。基于上述操作,ConvE模型定义了如下评分函数:

$$\phi(h, r, t) = f(\text{vec}(f([\bar{\mathbf{h}}; \bar{\mathbf{r}}] * \mathbf{\Omega})\mathbf{W})\mathbf{t}) \quad (38)$$

其中, $f(x)$ 表示非线性函数ReLU^[95], $\bar{\mathbf{h}}$ 和 $\bar{\mathbf{r}}$ 分别表示头实体及关系语义向量的二维变形(2D Reshape),符号 $[x]$ 代表拼接操作,*表示卷积操作, $\mathbf{\Omega}$ 则表示卷积核, \mathbf{W} 表示映射矩阵, $\text{vec}(x)$ 表示全连接操作。此外,不同于其他模型对每个知识三元组 (h, r, t) 进行评分(1-1评分),ConvE模型首次提出了1-N评分方法,即对头实体及关系 (h, r) 进行特征提取后同时与所有实体的特征矩阵 \mathbf{E} 进行特征匹配,该方法适用于大规模知识图谱的推理计算。

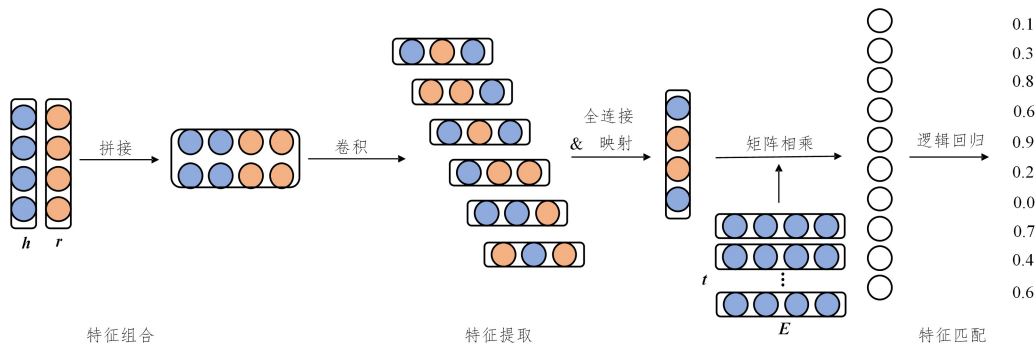


图 17 ConvE 模型示意图

Fig. 17 Schematic diagram of ConvE

ConvE 模型首次将卷积神经网络运用于知识图谱推理,在学习及提取实体及关系的语义特征方面具有很强的优势。此外,该模型提出了 1-N 评分方法,使得模型的评测速度相比 1-1 评分方法提升了 300 倍。相比 DistMult 模型,ConvE 模型不仅在推理准确性方面得到了较大提升,在参数量仅为 46 万的情况下,在 FB15k-237 数据集上取得了 Hit@10 为 42.5% 的优异性能,相比 DistMult 模型,其在同样性能下参数量减少为原来的 1/8。但是该模型相比计算机视觉中的卷积网络架构仍相对简单,且对特征向量的二维变形缺乏可解释性。在 ConvE 模型的整体流程的基础上,相关研究主要在特征组合和特征提取方面进行了优化改进,其中具有代表性的模型如下。

ConvKB^[96]模型将每个知识三元组 (h, r, t) 的语义特征向量组合成一个三阶矩阵,如图 18 所示,并将此矩阵当作卷积神经网络的输入,在卷积层中对矩阵使用多个卷积核以生成不同的特征图,然后将这些特征图全连接到代表输入三元组的单个特征向量中,通过点积操作将特征向量与权重向量相乘以返回知识三元组 (h, r, t) 的置信度评分,最后使用该评分来判断三元组是否正确。

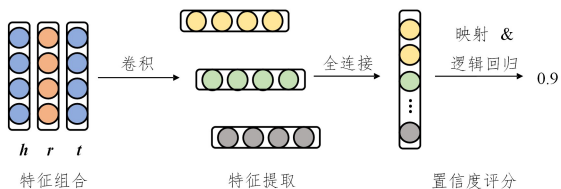


图 18 ConvKB 模型示意图

Fig. 18 Schematic diagram of ConvKB

具体地,ConvKB 模型定义了如下评分函数:

$$\phi(h, r, t) = \text{vec}(f(\llbracket h; r; t \rrbracket * \Omega)) \mathbf{W} \quad (39)$$

其中, $f(x)$ 表示非线性激活函数 ReLU, 符号 $\llbracket x \rrbracket$ 代表拼接操作, $*$ 表示卷积操作, Ω 则表示卷积核, $\text{vec}(x)$ 表示全连接操作, \mathbf{W} 表示映射矩阵。与 ConvE 模型端对端(End-to-End)的训练方式不同,ConvKB 模型首先需要通过 TransE 模型训练学习得到所有实体及关系的语义向量,然后用于知识图谱推理。

ConvR^[97]模型旨在进一步减少 ConvE 模型的参数,提出使用关系的语义特征向量替代卷积核,并定义了如下的评分函数:

$$\phi(h, r, t) = f(\text{vec}(f(\bar{h} * \bar{r})) \mathbf{W}) t \quad (40)$$

其中, \bar{h} 和 \bar{r} 分别表示头实体及关系语义特征向量的二维变形(2D Reshape), $*$ 表示卷积操作, $\text{vec}(x)$ 表示全连接操作, \mathbf{W} 表示映射矩阵, $f(x)$ 表示非线性激活函数 ReLU。ConvR 模型通过卷积神经网络提取到关于头实体与关系的语义特征,最后与尾实体进行语义匹配,从而得到知识三元组的置信度评分。

ConvR 模型根据关系的语义表示自适应地构造卷积核,并将这些卷积核应用于实体语义表示以生成特征图。通过此方式,该模型可以在实体和关系语义表示的不同区域之间进行丰富的交互,并且生成的所有特征图都能够捕获到这些交互。最后通过实验证明,ConvR 模型在参数减少的同时,提高了模型的推理能力。

IntactE 模型^[98]认为,ConvE 模型捕获的实体及关系间的语义交互是有限的,因此该模型通过改变特征组合的方式来捕获更多特征交互,如图 19 所示,并使用了循环卷积操作。

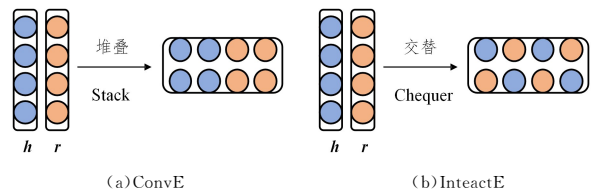


图 19 ConvE 与 IntactE 模型比较

Fig. 19 Comparison between ConvE and IntactE

具体地,IntactE 模型定义了如下的评分函数:

$$\phi(h, r, t) = f(\text{vec}(f(\llbracket h; r \rrbracket \odot \Omega)) \mathbf{W}) t \quad (41)$$

其中, $\llbracket h; r \rrbracket$ 代表头实体及关系的语义向量以交替(Chequer)的方式进行特征组合, \odot 表示深度循环卷积运算(Circular Convolution)操作, Ω 则表示卷积核, $\text{vec}(x)$ 表示全连接操作, \mathbf{W} 表示映射矩阵, $f(x)$ 表示非线性激活函数 ReLU。

IntactE 模型对于 ConvE 模型的特征组合及特征提取部分都进行了重新改造,提出了一种新颖的交替组合方式,并且在特征提取部分使用了循环卷积运算,以尽量捕获实体及关系间的语义交互。通过实验证明,IntactE 模型在知识图谱推理性能方面得到了进一步提升。

此外,Hyper^[99]使用超网络(Hypernetwork)架构来生成特定的每个关系的卷积核,并将这些卷积核应用到头实体向量卷积操作中。ReInceptionE^[100]与 RCN^[101]则针对卷积结构进行改进,从而增加了卷积操作的效率和特征提取能力。

上述模型基于现有方法进行了优化改进,或者在特征提取部分采用不同的深度学习方法,均在知识图谱推理上取得了不错的表现。

5.2 图神经网络

近年来,图神经网络^[102-103]的流行为知识图谱推理提供了新的解决思路。一方面,作为图结构数据的表示学习工具,图神经网络可以有效地聚合每个节点的邻居信息^[104-108];另一方面,通过施加相同的汇聚函数可以显著提高图神经网络的计算效率^[109-112]。通常,令 $\mathbf{H}^{(l+1)}$ 表示第 $(l+1)$ 层的节点特征矩阵,则基于图神经网络的特征汇聚通常可表示为:

$$\mathbf{H}^{(l+1)} = f(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \quad (42)$$

其中, $f(x)$ 表示非线性激活函数; $\tilde{\mathbf{A}}$ 表示图的邻接矩阵(Adjacency Matrix); $\tilde{\mathbf{D}}$ 是关于所有节点的度矩阵(Degree Matrix),并有 $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$; $\mathbf{W}^{(l)}$ 表示第 l 层的权重矩阵。对于含不对称邻接矩阵的有向图,特征汇聚可表示为:

$$\mathbf{H}^{(l+1)} = f(\tilde{\mathbf{D}}^{-1} \tilde{\mathbf{A}} \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \quad (43)$$

但是,上述特征汇聚过程主要针对同质图(Homogeneous Graph),而知识图谱通常带有多种类型的实体和关系,被广

泛称为异质信息网络^[113-114](Heterogeneous Information Network)。如图20所示,针对知识图谱设计有效的图神经网络框架时,需要考虑以下问题:1)知识图谱的异质性,异质性是知识图谱的固有属性,即各种类型的实体具有不同属性,并且它们的特征可能属于不同的向量空间;2)不同关系的重要性,知识图谱的异质性通常由关系路径^[68]反映,并且在不同知识三元组下表现出复杂的语义特征。如表5所列,针对上述问题,研究人员从实体及关系特征汇聚、注意力机制以及异质性等方面设计了不同的图神经网络框架来对知识图谱进行建模。

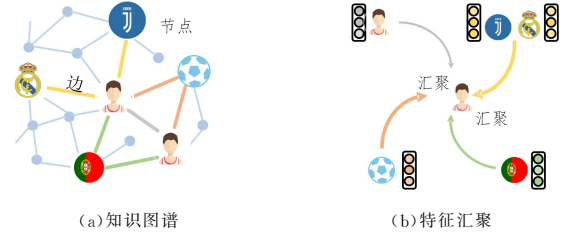


图20 知识图谱与特征汇聚示意图

Fig. 20 Schematic diagram of knowledge graph and feature aggregation

表5 基于图神经网络的知识图谱推理方法比较

Table 5 Comparison of knowledge graph reasoning models based on graph neural networks

模型	特征汇聚	实体向量	关系向量	注意力	异质性
R-GCN	$\mathbf{h}_i^{(l+1)} = f\left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_r^i} \frac{1}{ \mathcal{N}_r^i } \mathbf{h}_j^{(l)} \mathbf{W}_r^{(l)} + \mathbf{h}_i^{(l)} \mathbf{W}^{(l)}\right)$	✓	×	×	✓
KBGAT	$\mathbf{h}_i^{(l+1)} = f\left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{A}_i^k} \alpha_{ij}^k \mathbf{h}_j^{(l)} \mathbf{W}^{(l)}\right)$	✓	×	✓	×
W-GCN	$\mathbf{h}_i^{(l+1)} = f\left(\sum_{j \in \mathcal{A}_i} \alpha_r^{(l)} \mathbf{h}_j^{(l)} \mathbf{W}^{(l)} + \mathbf{h}_i^{(l)} \mathbf{W}^{(l)}\right)$	✓	×	✓	×
CompGCN	$\mathbf{h}_i^{(l+1)} = f\left(\sum_{j \in \mathcal{A}_i} \phi(\mathbf{h}_j^{(l)}, \mathbf{h}_r^{(l)}) \mathbf{W}_r^{(l)}\right)$	✓	✓	×	×

为将知识图谱中的结构信息融合到实体特征表示,R-GCN^[115]模型首次将图神经网络应用于知识图谱推理任务中,并定义了如下特征汇聚函数:

$$\mathbf{h}_i^{(l+1)} = f\left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_r^i} \frac{1}{|\mathcal{N}_r^i|} \mathbf{h}_j^{(l)} \mathbf{W}_r^{(l)} + \mathbf{h}_i^{(l)} \mathbf{W}^{(l)}\right) \quad (44)$$

其中, $\mathbf{h}_i^{(l)}$ 表示第 l 隐藏层的第 i 个节点特征, \mathcal{N}_r^i 表示在关系 r 下与第 i 个实体的相邻实体集合, $\mathbf{W}_r^{(l)}$ 和 $\mathbf{W}^{(l)}$ 为可学习的共享权重矩阵, $\mathbf{h}_i^{(l+1)}$ 则表示经过 l 层更新后的节点特征。

R-GCN模型的每层节点特征 $\mathbf{h}_i^{(l+1)}$ 都是由上一层节点特征 $\mathbf{h}_i^{(l)}$ 得到,且节点的邻居节点特征和自身特征分别通过 $\mathbf{W}_r^{(l)}$ 和 $\mathbf{W}^{(l)}$ 进行加权求和得到新的特征。为避免模型过拟合,R-GCN模型提出了两种正则化方法,分别为基函数分解(Basis Decomposition)和块对角分解(Block Diagonal Decomposition)。基函数分解可以看作是不同关系类型之间的权重共享,而块对角分解可以看作是对每个关系类型的权值矩阵的稀疏约束,皆用于减少模型的参数。通过在链接预测和实体分类任务上的实验,证明了融合知识图谱的图结构信息有助于推理任务。

KBGAT^[116]模型基于GAT^[106]图注意力网络,采用多头(Multi-head)注意力机制有选择性地汇聚邻接节点的特征,并定义了如下特征汇聚函数:

$$\mathbf{h}_i^{(l+1)} = f\left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{A}_i^k} \alpha_{ij}^k \mathbf{h}_j^{(l)} \mathbf{W}^{(l)}\right) \quad (45)$$

其中, \mathcal{A}_i^k 表示第 i 个实体的相邻实体集合; α_{ij}^k 表示第 k 个由注意力机制计算出的归一化后的权重值,其计算过程如图21所示; \mathbf{h}_i 和 \mathbf{h}_j 分别代表中心节点和邻接节点的特征; \mathbf{r} 表示它们之间的关系特征,然后通过线性变换和Softmax激活函数得到邻接节点的权重值 α_{ij} 。基于学习得到的节点特征,KBGAT采用ConvKB模型进行知识图谱推理。

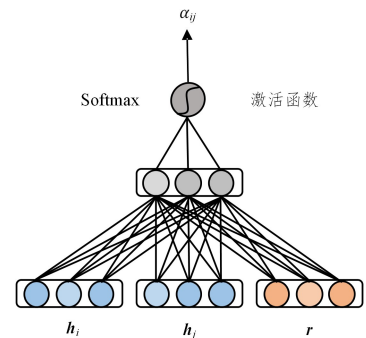


图21 KBGAT模型的注意力机制示意图

Fig. 21 Schematic diagram of attention mechanism of KBGAT

KBGAT模型在特征汇聚过程中,只提供实体特征的

表示方法,忽略了关系(边)的特征。该模型忽略了知识图谱的异质性,无法对不同关系路径下的语义特征进行区分。此外,KBGAT模型采用TransE模型对所有实体及关系向量进行预训练,只需对神经网络中的参数进行训练更新。

W-GCN^[117]模型针对ConvE模型没有考虑图结构信息的不足,提出用加权的(Weighted)图卷积网络来对知识图谱进行建模,并定义了如下特征汇聚函数:

$$\mathbf{h}_i^{(l+1)} = f\left(\sum_{j \in \mathcal{N}_i} \alpha_r^{(l)} \mathbf{h}_j^{(l)} \mathbf{W}^{(l)} + \mathbf{h}_i^{(l)} \mathbf{W}^{(l)}\right) \quad (46)$$

其中, $\alpha_r^{(l)}$ 表示第 l 层关于关系 r 的权重。与R-GCN模型一样,W-GCN模型在汇聚邻接节点特征时同时保留了自己的信息。

在知识图谱推理部分,该模型提出了Conv-TransE方法。具体地,Conv-TransE首先取消了ConvE的二维变形(2D Reshape)以保留平移距离的模型假设;然后用多个卷积核进行卷积,得到特征图并拼接成一个长向量,经过一个线性变换再与实体向量矩阵相乘;最后得到三元组的置信度评分。可以发现,W-GCN模型同样没有考虑知识图谱的异质性和关系特征(边)的汇聚。

针对上述3种图神经网络结构忽略了知识图谱中关系特征汇聚的不足,CompGCN^[118]模型联合学习知识图谱中实体(节点)和关系(边)的向量表示,并定义了如下特征汇聚函数:

$$\mathbf{h}_i^{(l+1)} = f\left(\sum_{j \in \mathcal{N}_i} \psi(\mathbf{h}_j^{(l)}, \mathbf{h}_r^{(l)}) \mathbf{W}_r^{(l)}\right) \quad (47)$$

其中, $\mathbf{h}_j, \mathbf{h}_r$ 分别为邻接实体和关系的特征, $\psi(x)$ 表示汇聚函数。具体地,CompGCN模型定义了如下3种非参数的汇聚函数:

$$\begin{cases} \text{Subtraction: } \psi(\mathbf{h}_j, \mathbf{h}_r) = \mathbf{h}_j - \mathbf{h}_r \\ \text{Multiplication: } \psi(\mathbf{h}_j, \mathbf{h}_r) = \mathbf{h}_j \circ \mathbf{h}_r \\ \text{Circular-correlation: } \psi(\mathbf{h}_j, \mathbf{h}_r) = \mathbf{h}_j * \mathbf{h}_r \end{cases} \quad (48)$$

其中, $\mathbf{h}_j - \mathbf{h}_r$ 是基于TransE平移距离的相减(Subtraction)操作, $\mathbf{h}_j \circ \mathbf{h}_r$ 是基于DistMult语义匹配的乘法(Multiplication)操作, $\mathbf{h}_j * \mathbf{h}_r$ 是基于HolE语义匹配的循环相关(Circular-correlation)操作。最后,模型分别采用了TransE,DistMult和ConvE的知识图谱推理方法,其中ConvE性能最佳。

此外,DPMPN^[119]模型提出了一种基于意识先验(Consciousness Prior)的动态剪枝消息传递网络,具体包括基于全图随机抽样和基于批输入多子图的图神经网络。RGHAT^[120]与HRAN^[121]提出了基于异质信息网络的图神经网络模型,通过实体及关系层级分步汇聚邻接节点特征,并且基于关系路径汇聚了不同类型的语义特征信息。总体来讲,通过图神经网络能够更好地学习实体的特征表示,可进一步提升知识图谱推理的性能,但在针对大规模知识图谱的计算效率方面有待进一步提升。

5.3 循环神经网络

循环神经网络由于在时序依赖建模方面的优势,因此被广泛应用于自然语言及音频处理等任务上。针对知识图谱推理任务,研究人员尝试利用循环神经网络捕获实体的多跳(Multi-hop)关系依赖,从而更好地利用知识图谱的结构信息,其中文献[122]和文献[123]均直接利用循环神经网络来建模关系路径,但前者忽略了实体特征的学习。SSME^[124]模型在

使用循环神经网络学习实体及关系特征的同时,将逆关系镜像引入推理过程,使得模型能够有效处理知识图谱中的对称性/非对称性关系推理。但是上述模型重点关注三元组层级,导致在特征学习过程中难以建模实体的长期关系依赖。针对上述问题,RSN^[125]模型提出了一种循环跳跃网络(Recurrent Skipping Networks),如图22所示。

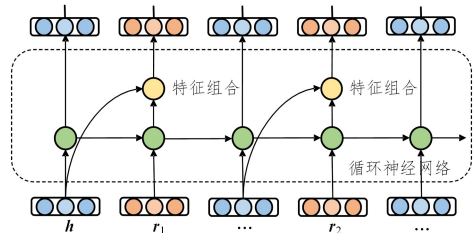


图22 RSN模型的示意图

Fig. 22 Schematic diagram of RSN

该模型将循环神经网络与残差学习相结合,采用跳跃机制来弥合实体之间的空缺信息,从而有效地建模实体特征学习过程中的长期关系依赖。具体而言,给定由实体及关系交替组合的输入序列 (x_1, x_2, \dots, x_N) ,奇数项为实体,偶数项为关系,标准的循环神经网络的计算式如下:

$$\mathbf{h}_n = \tanh(\mathbf{h}_n \mathbf{h}_{n-1} + \mathbf{W}_x \mathbf{x}_n + \mathbf{b}) \quad (49)$$

其中, \mathbf{h}_n 表示输出的隐藏状态特征, \mathbf{W}_h 与 \mathbf{W}_x 皆为权重矩阵, \mathbf{b} 为偏置向量。可以发现,上述计算过程没有区分实体与关系特征作为输入的顺序差异,为此RSN模型采用下述判断进行进一步的优化:

$$\mathbf{h}_n' = \begin{cases} \mathbf{h}_n, & x_n \in \mathcal{E} \\ \mathbf{S}_1 \mathbf{h}_n + \mathbf{S}_2 \mathbf{x}_{n-1}, & x_n \in \mathcal{R} \end{cases} \quad (50)$$

其中, \mathbf{h}_n' 为循环跳跃网络输出的隐藏状态特征, \mathbf{S}_1 与 \mathbf{S}_2 皆为权重矩阵。即当输入 x_n 为实体时, \mathbf{h}_n' 与 \mathbf{h}_n 相等,当输入 x_n 为关系时,则 \mathbf{h}_n' 将与特征 \mathbf{x}_{n-1} 做加权平均。

5.4 胶囊网络

胶囊网络^[126-127](Capsule Network)主要用于解决CNN在特征提取过程中会忽略特征方位的不足。在此背景下,CapsE^[128]模型首次尝试将胶囊网络用于知识图谱推理和个性化搜索任务,充分利用胶囊网络可以捕捉到不同特征映射在同一维度上的深层特征的功能,针对同一维度信息进行深度结构建模。具体地,CapsE模型定义了如下评分函数:

$$\phi(h, r, t) = \| \text{capsnet}(f([\mathbf{h}; \mathbf{r}; \mathbf{t}] * \mathbf{\Omega})) \| \quad (51)$$

其中, $f(x)$ 表示非线性激活函数,符号 $[x]$ 代表拼接操作, $*$ 表示卷积操作, $\mathbf{\Omega}$ 则表示卷积核,capsnet(x)表示胶囊网络计算。该模型与ConvKB模型类似,均采用预训练的实体及关系特征表示进行推理,通过链接预测和个性化搜索实验验证了模型的有效性。CapsE为知识图谱推理任务提供了全新视角,值得后续研究。

5.5 基于神经网络的知识图谱推理方法总结

本节对基于卷积神经网络、图神经网络、循环神经网络及胶囊网络的知识图谱推理模型进行了介绍。其中,基于卷积神经网络的推理模型一方面从实体及关系语义特征组合形式进行设计,以增加卷积过程中的交互次数;另一方面通过结合最新卷积结构来提高模型的特征提取能力。基于图神经网络

的推理模型从实体向量、关系向量、注意力机制、图的异质性 4 个方面设计了不同的特征汇聚方法,以学习实体及关系的特征向量,然后用于知识图谱推理任务;基于循环神经网络的推理模型主要关注实体前后关系路径的建模,使得学习到的特征表示能够捕获上下文结构信息。基于胶囊网络的推理模型仍处于初步探索阶段,有待进一步的研究。总体来讲,相比基于平移距离或语义匹配的推理模型,基于神经网络的推理模型在语义特征学习与提取方面具有较大优势,是当前知识图谱推理的主流研究方向。

6 总结与展望

6.1 总结

如何对知识图谱中存储的知识进行更新或补全,成为了当下热门的研究课题。在此背景下,基于表示学习的知识图谱推理技术应运而生,其主要思路是将传统推理过程转换成基于分布式表示的实体及关系语义向量计算,然后用于知识三元组的推理。此外,基于表示学习的知识图谱推理技术能够显著提升计算效率、有效缓解数据稀疏以及实现异质信息融合,显著地提高了知识图谱推理相关任务的性能。

本文首先对知识图谱推理的相关背景知识进行了介绍,然后对基于表示学习的知识图谱推理方法进行了分类总结,包括传统的平移距离和语义匹配典型方法、融合多源信息的推理方法,以及基于神经网络的推理方法,并介绍了不同的模型假设以及评分函数,总结分析了各类方法的特点与创新。

6.2 展望

近年来,基于表示学习的知识图谱推理在相关任务上展现出了巨大的潜力,但也面临着诸多挑战,本节将对未来的研究方向进行展望。

(1)知识图谱推理可解释性研究。知识图谱推理过程的可解释性(Interpretability)是一个值得关注的研究点,可解释性可以使知识图谱推理过程更符合人的认知过程。尽管近期的基于神经网络的推理模型取得了令人印象深刻的性能提升,但其在推理的过程及结果可解释性方面存在很大局限。相关研究通过结合关系路径搜索^[129]或知识置信度^[130]来增加模型的可解释性,此外 DeepPath^[27]与 MINERVA^[131]模型采用强化学习来实现推理过程中更合理的路径选择,DI-VA^[132]模型在知识图谱推理中引入了变分推理框架,将路径搜索和路径推理紧密结合,从而进行联合推理。但上述模型性能与当前最流行的神经网络方法相差较大,因此,如何平衡推理模型的可解释性与实验性能是一个值得关注的问题。未来可以探索在理论上解决基于神经网络推理模型的可解释性,此外可将传统基于逻辑规则的知识图谱推理方法融合到现有的技术方案中。

(2)更加多源的信息融合与建模。融合多源信息可以辅助构建更好的知识表示,从而提高知识图谱推理相关任务的性能。当前,融合多源信息知识图谱推理包括文本描述、时序特征、图像信息、关系路径等。但这些模型的信息来源比较有限,且对多源信息的使用还处于初级阶段,仍需研究如何设计更合理、有效的知识图谱推理模型,以便更好地利用有关知识图谱的多源信息。未来一方面应考虑建立更为高效的表示

学习方法对多源信息进行特征提取,另一方面可探索知识与跨模态信息(如文本、图像等)的融合表示。此外,可通过实体对齐方法^[133]进行多源知识图谱的融合,有助于建立多语义空间融合的准确知识表示结果。

(3)面向规模庞大的知识图谱推理。互联网中的知识图谱往往规模庞大,现有的知识图谱推理模型往往无法高效地应用于现实场景的具体任务中。例如,截止到目前,谷歌公司建立的世界知识库 Freebase 包含了超过 5 亿个实体,以及超过 19 亿条知识三元组。由于这些知识图谱的规模宏大,现有的很多知识图谱推理模型往往只是在大规模知识图谱的一个子集上验证了模型的有效性,而这些模型由于计算效率的问题往往无法直接应用于现实场景。目前大规模预训练模型^[134](Pre-trained Models)已成功应用于自然语言处理的各个领域,未来可借鉴相关算法实现对大规模知识图谱的预训练并用于推理任务。此外可以参考课程学习^[135](Curriculum Learning)思想,优先学习核心的知识表示,然后学习外围的知识表示,也许能解决大规模的知识图谱推理问题。

(4)基于低资源知识图谱的推理研究。现有的推理模型往往需要高质量的三元组样本进行训练,在面对低资源(Low Resource)实体及关系时推理性能表现不佳。目前已有相关工作对其进行了初步探索,例如文献^[136]提出了一种元学习框架,旨在通过文字描述来处理小样本实体及关系学习。文献^[137-138]采用生成对抗网络(Generative Adversarial Networks)框架,利用从高资源关系中学习到的特征表示辅助相关的低资源知识表示学习。因此,面对低资源知识图谱的推理,一方面可参考前文通过融合附加信息增强少样本知识的表示能力,另一方面可将已建立的高质量知识图谱作为先验,借助高质量样本辅助低资源知识图谱推理。

(5)涉及不同关系类型的推理研究。现有的关系分类方式主要有两种,包括 TransE 模型提出的一对一(1-to-1)、一对多(1-to-N)、多对一(N-to-1)及多对多(N-to-N)划分方法和 RotatE 模型提出的对称(Symmetry)、反对称(Anti-symmetry)、互逆(Inversion)及组合(Composition)划分方法。针对前者的关系类型,已有研究利用语义映射^[35-37]和流形空间^[44-46]的解决方案,在一定程度上缓解了复杂关系的推理难题;针对后者的关系类型,相关工作采取语义空间旋转^[38,86]对 4 种关系进行建模。上述工作表明,涉及不同关系类型需建立专门的知识推理模型,未来有必要针对具体知识图谱推理任务设计相应的关系分类标准。此外,可尝试结合认知科学建立新的关系分类,开展新关系分类体系下的推理研究。

除了上述主要研究方向之外,还有大量关于知识图谱推理的研究工作亟待开展。例如,当前的知识图谱往往基于同一语言进行构建,而多语言的知识图谱对于全球化下的知识共享非常重要,建立跨语言的知识图谱推理模型必将成为未来的发展趋势。另外,知识图谱推理已在推荐系统、知识搜索、智能问答以及机器翻译等领域发挥重要作用,期待知识图谱推理能够带来更多的创新智能应用。

参考文献

[1] LIU Q, LI Y, DUAN H, et al. Knowledge graph construction

- techniques[J]. *Journal of Computer Research and Development*, 2016, 53(3):582-600.
- [2] MILLER G A. WordNet;a lexical database for English [J]. *Communications of the Acm*, 1995, 38(11):39-41.
- [3] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//*Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. 2008:1247-1250.
- [4] SUCHANEK F M, KASNECI G, WEIKUM G. Yago: a core of semantic knowledge[C]//*Proceedings of the World Wide Web*. 2007:697-706.
- [5] AUER R, BIZER C, KOBILAROV G, et al. DBpedia: a nucleus for a web of open data[C]//*Proceedings of the International Semantic Web Conference*. 2007:722-735.
- [6] XU B, XU Y, LIANG J, et al. CN-DBpedia: A never-ending Chinese knowledge extraction system[C]//*Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. 2017:428-438.
- [7] NIU X, SUN X, WANG H, et al. Zhishi. me-weaving chinese linking open data[C]//*Proceedings of the International Semantic Web Conference*. 2011:205-220.
- [8] WANG X, HE X, CAO Y, et al. KGAT: Knowledge Graph Attention Network for Recommendation[C]//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019:950-958.
- [9] HU S, ZOU L, YU J X, et al. Answering Natural Language Questions by Subgraph Matching over Knowledge Graphs[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(5):824-837.
- [10] ZHOU H, YOUNG T, HUANG M, et al. Commonsense Knowledge Aware Conversation Generation with Graph Attention [C]//*Proceedings of the International Joint Conference on Artificial Intelligence*. 2018:4623-4629.
- [11] LIU K, ZHANG Y, JI G, et al. Representation learning for question answering over knowledge base: An overview[J]. *Acta Automatica Sinica*, 2016, 42(6):807-818.
- [12] CARLSON A, BETTERIDGE J, KISIEL B, et al. Toward an architecture for never-ending language learning[C]//*Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*. 2010:1306-1313.
- [13] GUAN S, JIN X, JIA Y, et al. Knowledge graph-oriented knowledge inference methods: A survey[J]. *Journal of Software*, 2018, 29(10):2966-2994.
- [14] CHEN X, JIA S, XIANG Y. A review: Knowledge reasoning over knowledge graph: A Survey [J]. *Expert Systems with Applications*, 2020, 141:1-21.
- [15] LIU Z, SUN M, LIN Y, et al. Knowledge representation learning: A review[J]. *Journal of Computer Research and Development*, 2016, 53(2):247-261.
- [16] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553):436-444.
- [17] BENGIO Y, COURVILLE A C, VINCENT P. Representation Learning: A Review and New Perspectives[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8):1798-1828.
- [18] NICKEL M, MURPHY K, TRESP V, et al. A Review of Relational Machine Learning for Knowledge Graphs[C]//*Proceedings of the IEEE*. 2016:11-33.
- [19] WANG Q, MAO Z, WANG B, et al. Knowledge Graph Embedding: A Survey of Approaches and Applications [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(12):2724-2743.
- [20] PAULHEIM H. Knowledge graph refinement: A survey of approaches and evaluation methods[J]. *Semantic web*, 2017, 8(3):489-508.
- [21] JI S, PAN S, CAMBRIA E, et al. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(2):494-514.
- [22] ZHANG J, CHEN B, ZHANG L, et al. Neural, symbolic and neural-symbolic reasoning on knowledge graphs[J]. *AI Open*, 2021, 2:14-35.
- [23] ROSSI A, BARBOSA D, FIRMANI D, et al. Knowledge Graph Embedding for Link Prediction: A Comparative Analysis[J]. *ACM Transactions on Knowledge Discovery from Data*, 2021, 15(2):1-49.
- [24] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating Embeddings for Modeling Multi-relational Data[C]//*Proceedings of the Advances in Neural Information Processing Systems*. 2013:2787-2795.
- [25] TOUTANOVA K, CHEN D. Observed versus latent features for knowledge base and text inference[C]//*Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*. 2015:57-66.
- [26] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2D Knowledge Graph Embeddings[C]//*Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. 2018:1811-1818.
- [27] XIONG W, HOANG T, WANG W Y. DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning[C]//*Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2017:564-573.
- [28] MAHDISOLTANI F, BIEGA J, SUCHANEK F M, YAGO3: A Knowledge Base from Multilingual Wikipedias[C]//*Proceedings of the Seventh Biennial Conference on Innovative Data Systems Research*. 2015:1-11.
- [29] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]//*Proceedings of the International Conference on Artificial Intelligence and Statistics*. 2010:249-256.
- [30] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. *Journal of Machine Learning Research*, 2014, 15(1):1929-1958.
- [31] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//*Pro-*

- ceedings of the 32nd International Conference on Machine Learning, 2015:448-456.
- [32] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:2818-2826.
- [33] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization[C]// Proceedings of the 3rd International Conference on Learning Representations. 2015:1-15.
- [34] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]// Proceedings of the Advances in Neural Information Processing Systems. 2013:3111-3119.
- [35] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes[C]// Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. 2014:1112-1119.
- [36] LIN Y, LIU Z, ZHU X, et al. Learning entity and relation embeddings for knowledge graph completion[C]// Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015:2181-2187.
- [37] JI G, HE S, XU L, et al. Knowledge Graph Embedding via Dynamic Mapping Matrix[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015:687-696.
- [38] SUN Z, DENG Z H, NIE J Y, et al. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space[C]// Proceedings of the 7th International Conference on Learning Representations. 2019:1-18.
- [39] FANG Y, ZHAO X, TAN Z, et al. A Revised Translation-Based Method for Knowledge Graph Representation[J]. Journal of Computer Research and Development, 2018, 55(1): 139-150.
- [40] JI G, LIU K, HE S, et al. Knowledge Graph Completion with Adaptive Sparse Transfer Matrix[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016:985-991.
- [41] XIAO H, HUANG M, ZHU X. TransG: A generative model for knowledge graph embedding[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016:2316-2325.
- [42] HE S, LIU K, JI G, et al. Learning to Represent Knowledge Graphs with Gaussian Embedding[C]// Proceedings of the 24th ACM International Conference on Information and Knowledge Management. 2015:623-632.
- [43] LIU Q, HAN M, YANG X, et al. Representation Learning Based Relational Inference Algorithm with Semantical Aspect Awareness[J]. Journal of Computer Research and Development, 2017, 54(8): 1682-1692.
- [44] XIAO H, HUANG M, ZHU X. From one point to a manifold: knowledge graph embedding for precise link prediction[C]// Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. 2016:1315-1321.
- [45] EBISU T, ICHISE R. TorusE: Knowledge Graph Embedding on a Lie Group[C]// Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. 2018:1819-1826.
- [46] EBISU T, ICHISE R. Generalized Translation-based Embedding of Knowledge Graph[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(5): 941-951.
- [47] CAI L, WANG W Y, KBGAN: Adversarial Learning for Knowledge Graph Embeddings[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. 2018:1470-1480.
- [48] NICKEL M, TRESP V, KRIEGEL H P. A three-way model for collective learning on multi-relational data[C]// Proceedings of the International Conference on Machine Learning. 2011: 809-816.
- [49] YANG B, YI W, HE X, et al. Embedding Entities and Relations for Learning and Inference in Knowledge Bases[C]// Proceedings of the 3rd International Conference on Learning Representations. 2014:1-12.
- [50] NICKEL M, ROSASCO L, POGGIO T A. Holographic Embeddings of Knowledge Graphs[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016:1955-1961.
- [51] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]// Proceedings of the 33rd International Conference on Machine Learning. 2016:2071-2080.
- [52] BALAŽEVIĆ I, ALLEN C, HOSPEDALES T M. TUCKER: Tensor Factorization for Knowledge Graph Completion[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. 2019:5188-5197.
- [53] LIU H, WU Y, YANG Y. Analogical inference for multi-relational embeddings[C]// Proceedings of the 34th International Conference on Machine Learning. 2017:2168-2178.
- [54] KAZEMI S M, POOLE D. Simple embedding for link prediction in knowledge graphs[C]// Proceedings of the Advances in Neural Information Processing Systems. 2018:4289-4300.
- [55] LACROIX T, USUNIER N, OBOZINSKI G. Canonical Tensor Decomposition for Knowledge Base Completion[C]// Proceedings of the 35th International Conference on Machine Learning. 2018:2869-2878.
- [56] XU C, LI R. Relation Embedding with Dihedral Group in Knowledge Graph[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019:263-272.
- [57] BALAŽEVIĆ I, ALLEN C, HOSPEDALES T. Multi-relational Poincare Graph Embeddings[C]// Proceedings of the Advances in Neural Information Processing Systems. 2019:4465-4475.
- [58] ZHANG S, TAY Y, YAO L, et al. Quaternion Knowledge Graph Embedding[C]// Proceedings of the Advances in Neural Information Processing Systems. 2019:2731-2741.
- [59] XIE R, LIU Z, JIA J, et al. Representation Learning of Knowledge Graphs with Entity Descriptions[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016: 2659-2665.
- [60] WANG Z, LI J. Text-enhanced representation learning for

- knowledge graph[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. 2016;1293-1299.
- [61] AN B, CHEN B, HAN X, et al. Accurate Text-Enhanced Knowledge Graph Representation Learning[C]//Proceedings of the North American Chapter of the Association for Computational Linguistics. 2018;745-755.
- [62] GUAN N, SONG D, LIAO L. Knowledge graph embedding with concepts[J]. Knowledge-Based Systems, 2019, 164:38-44.
- [63] FAN M, ZHOU Q, ZHENG T F, et al. Distributed representation learning for knowledge graphs with entity descriptions[J]. Pattern Recognition Letters, 2017, 93:31-37.
- [64] XIAO H, HUANG M, MENG L, et al. SSP: semantic space projection for knowledge graph embedding with text descriptions [C]//Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. 2017;3104-3110.
- [65] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the Advances in Neural Information Processing Systems. 2017;5998-6008.
- [66] WANG Q, HUANG P, WANG H, et al. CoKE: Contextualized knowledge graph embedding[J]. arXiv:1911.02168, 2019.
- [67] YAO L, MAO C, LUO Y. KG-BERT: BERT for knowledge graph completion[J]. arXiv:1909.03193, 2019.
- [68] LIN Y, LIU Z, LUAN H, et al. Modeling Relation Paths for Representation Learning of Knowledge Bases[C]//Proceedings of the Empirical Methods in Natural Language Processing. 2015;705-714.
- [69] JIA Y, WANG Y, JIN X, et al. Path-specific knowledge graph embedding[J]. Knowledge-Based Systems, 2018, 151:37-44.
- [70] NIE B, SUN S. Knowledge graph embedding via reasoning over entities, relations, and text [J]. Future Generation Computer Systems, 2019, 91:426-433.
- [71] JAGVARAL B, LEE W K, ROH J S, et al. Path-based reasoning approach for knowledge graph completion using CNN-BiLSTM with attention mechanism[J]. Expert Systems with Applications, 2020, 142:1-10.
- [72] KAZEMI S M, GOEL R, JAIN K, et al. Representation Learning for Dynamic Graphs: A Survey. [J] Journal of Machine Learning Research, 2020, 21(70):1-73.
- [73] LEBLAY J, CHEKOL M W. Deriving Validity Time in Knowledge Graph[C]//Proceedings of the 2018 World Wide Web Conference. 2018;1771-1776.
- [74] DASGUPTA S S, RAY S N, TALUKDAR P. HyTE: Hyperplane-based Temporally aware Knowledge Graph Embedding [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018;2001-2011.
- [75] GARCÍA-DURÁN A, DUMANCIC S, NIEPERT M. Learning Sequence Encoders for Temporal Knowledge Graph Completion [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018;4816-4821.
- [76] GOEL R, KAZEMI S M, BRUBAKER M, et al. Diachronic Embedding for Temporal Knowledge Graph Completion[C]//Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. 2020;3988-3995.
- [77] TRIVEDI R, DAI H, WANG Y, et al. Know-Evolve: Deep Temporal Reasoning for Dynamic Knowledge Graphs[C]//Proceedings of the 34th International Conference on Machine Learning. 2017;3462-3471.
- [78] TRIVEDI R, FARAJTABAR M, BISWAL P, et al. DyRep: Learning Representations over Dynamic Graphs[C]//Proceedings of the 7th International Conference on Learning Representations. 2019;1-25.
- [79] JIN W, QU M, JIN X, et al. Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020;6669-6683.
- [80] JAIN P, RATHI S, CHAKRABARTI S. Temporal Knowledge Base Completion: New Algorithms and Evaluation Protocols [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020;3733-3747.
- [81] LACROIX T, OBOZINSKI G, USUNIER N. Tensor Decompositions for Temporal Knowledge Base Completion[C]//Proceedings of the 8th International Conference on Learning Representations. 2020;1-12.
- [82] ZHU C, CHEN M, FAN C, et al. Learning from History: Modeling Temporal Knowledge Graphs with Sequential Copy-Generation Networks [C] // Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence. 2021;4732-4740.
- [83] XIE R, LIU Z, SUN M. Representation learning of knowledge graphs with hierarchical types[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. 2016;2965-2971.
- [84] GUO S, WANG Q, WANG B, et al. SSE: Semantically Smooth Embedding for Knowledge Graphs[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(4):884-897.
- [85] LIN Y, LIU Z, SUN M. Knowledge Representation Learning with Entities, Attributes and Relations[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. 2016;2866-2872.
- [86] ZHANG Z, CAI J, ZHANG Y, et al. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction[C]//Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. 2020;3065-3072.
- [87] XIE R, HEINRICH S, LIU Z, et al. Integrating Image-based and Knowledge-based Representation Learning[J]. IEEE Transactions on Cognitive and Developmental Systems, 2020, 12(2):169-178.
- [88] XIE R, LIU Z, LUAN H, et al. Image-embodied knowledge representation learning[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. 2017;3140-3146.
- [89] GUO S, WANG Q, WANG L, et al. Jointly Embedding Knowledge Graphs and Logical Rules[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016;192-202.
- [90] WANG Q, WANG B, GUO L. Knowledge Base Completion Using Embeddings and Rules[C]//Proceedings of the Twenty-

- Fourth International Joint Conference on Artificial Intelligence. 2015;1859-1866.
- [91] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C] // Proceedings of the Advances in Neural Information Processing Systems. 2013;926-934.
- [92] LIU Q, JIANG H, EVDOKIMOV A, et al. Probabilistic reasoning via deep learning: Neural association models[J]. arXiv: 1603.07704, 2016.
- [93] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015;1-9.
- [94] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8):2011-2023.
- [95] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C] // Proceedings of the 26th Annual Conference on Neural Information Processing Systems. 2012;1097-1105.
- [96] 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] // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. 2018;327-333.
- [97] JIANG X, WANG Q, WANG B. Adaptive Convolution for Multi-Relational Learning[C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. 2019;978-987.
- [98] VASHISHTH S, SANYAL S, NITIN V, et al. InteractE: Improving Convolution-Based Knowledge Graph Embeddings by Increasing Feature Interactions[C] // Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. 2020;3009-3016.
- [99] BALAZEVIC I, ALLEN C, HOSPEDALES T M. Hypernetwork Knowledge Graph Embeddings[C] // Proceedings of the International Conference on Artificial Neural Networks. 2019;553-565.
- [100] XIE Z, ZHOU G, LIU J, et al. ReInceptionE: Relation-Aware Inception Network with Joint Local-Global Structural Information for Knowledge Graph Embedding[C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020;5929-5939.
- [101] LI Z, LIU H, ZHANG Z, et al. Recalibration Convolutional Networks for Learning Interaction Knowledge Graph Embedding[J]. Neurocomputing, 2021, 427;118-130.
- [102] XU B, CEN K, HUANG J, et al. A survey on graph convolutional neural network[J]. Chinese Journal of Computers, 2020, 43(5):755-780.
- [103] WU Z, PAN S, CHEN F, et al. A Comprehensive Survey on Graph Neural Networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1):4-24.
- [104] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C] // Proceedings of the Advances in Neural Information Processing Systems. 2017;1024-1034.
- [105] KIPF T N, WELING M. Semi-Supervised Classification with Graph Convolutional Networks[C] // Proceedings of the 5th International Conference on Learning Representations. 2017;1-14.
- [106] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[C] // Proceedings of the 6th International Conference on Learning Representations. 2018;1-12.
- [107] WANG X, JI H, SHI C, et al. Heterogeneous Graph Attention Network[C] // Proceedings of the World Wide Web Conference. 2019;2022-2032.
- [108] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral Networks and Locally Connected Networks on Graphs[C] // Proceedings of the 2nd International Conference on Learning Representations. 2014;1-14.
- [109] CHEN J, MA T, XIAO C. FastGCN: Fast Learning with Graph Convolutional Networks via Importance Sampling[C] // Proceedings of the 2nd International Conference on Learning Representations. 2018;1-15.
- [110] WU F, SOUZA A H, ZHANG T, et al. Simplifying Graph Convolutional Networks[C] // Proceedings of the 36th International Conference on Machine Learning. 2019;6861-6871.
- [111] YUN S, JEONG M, KIM R, et al. Graph Transformer Networks[C] // Proceedings of the Advances in Neural Information Processing Systems. 2019;11960-11970.
- [112] HASSANI K, AHMADI A H K. Contrastive Multi-View Representation Learning on Graphs[C] // Proceedings of the International Conference on Machine Learning. 2020;3451-3461.
- [113] SHI C, LI Y, ZHANG J, et al. A Survey of Heterogeneous Information Network Analysis[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(1):17-37.
- [114] ZHANG C, SONG D, HUANG C, et al. Heterogeneous Graph Neural Network[C] // Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019;793-803.
- [115] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C] // Proceedings of the 15th European Semantic Web Conference. 2018;593-607.
- [116] NATHANI D, CHAUHAN J, SHARMA C, et al. Learning Attention-based Embeddings for Relation Prediction in Knowledge Graphs[C] // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;4710-4723.
- [117] SHANG C, TANG Y, HUANG J, et al. End-to-End Structure-Aware Convolutional Networks for Knowledge Base Completion[C] // Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence. 2019;3060-3067.
- [118] VASHISHTH S, SANYAL S, NITIN V, et al. Composition-based Multi-Relational Graph Convolutional Networks[C] // Proceedings of the 8th International Conference on Learning Representations. 2020;1-16.
- [119] XU X, FENG W, JIANG Y, et al. Dynamically Pruned Message Passing Networks for Large-scale Knowledge Graph Reasoning[C] // Proceedings of the 8th International Conference on Learning Representations. 2020;1-16.

- ning Representations. 2020;1-30.
- [120]ZHANG Z,ZHUANG F,ZHU H, et al. Relational Graph Neural Network with Hierarchical Attention for Knowledge Graph Completion[C]//Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. 2020;9612-9619.
- [121]LI Z,LIU H,ZHANG Z, et al. Learning Knowledge Graph Embedding with Heterogeneous Relation Attention Networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022,33(8):3961-3973.
- [122]GARDNER M,TALUKDAR P P,KRISHNAMURTHY J, et al. Incorporating Vector Space Similarity in Random Walk Inference over Knowledge Bases[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014;397-406.
- [123]NEELAKANTAN A,ROTH B,MCCALLUM A. Compositional Vector Space Models for Knowledge Base Completion[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. 2015;156-166.
- [124]YANG X,WAN R,ZHANG H, et al. Semantical Symbol Mapping Embedding Learning Algorithm for Knowledge Graph[J]. Journal of Computer Research and Development, 2018,55(8):1773-1784.
- [125]GUO L,SUN Z,HU W. Learning to Exploit Long-term Relational Dependencies in Knowledge Graphs[C]//Proceedings of the 36th International Conference on Machine Learning. 2019:2505-2514.
- [126]SABOUR S,FROSST N,HINTON G E. Dynamic Routing Between Capsules[C]//Proceedings of the Advances in Neural Information Processing Systems. 2017;3856-3866.
- [127]HINTON G E,SABOUR S,FROSST N. Matrix capsules with EM routing[C]//Proceedings of the 6th International Conference on Learning Representations. 2018;1-15.
- [128]NGUYEN D Q,VU T,NGUYEN T D, et al. A Capsule Network-based Embedding Model for Knowledge Graph Completion and Search Personalization[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. 2018;2180-2189.
- [129]ZHANG W,PAUDEL B,ZHANG W, et al. Interaction Embeddings for Prediction and Explanation in Knowledge Graphs [C]//Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. 2019;96-104.
- [130]XIE R,LIU Z,LIN F, et al. Does William Shakespeare really write Hamlet? knowledge representation learning with confidence[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. 2018;4954-4961.
- [131]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[C]//Proceedings of the 6th International Conference on Learning Representations. 2018;1-18.
- [132]CHEN W,XIONG W,YAN X, et al. Variational Knowledge Graph Reasoning[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. 2018;1823-1832.
- [133]ZHAO X,ZENG W,TANG J, et al. An Experimental Study of State-of-the-Art Entity Alignment Approaches[J]. IEEE Transactions on Knowledge and Data Engineering, 2022,34(6):2610-2625.
- [134]QIU X,SUN T,XU Y, et al. Pre-trained models for natural language processing: A survey [J]. Science China Technological Sciences,2020,63(10):1872-1897.
- [135]WANG X,CHEN Y,ZHU W. A Survey on Curriculum Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022,44(9):4555-4576.
- [136]WANG Z,LAI K P,LI P, et al. Tackling Long-Tailed Relations and Uncommon Entities in Knowledge Graph Completion[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. 2019;250-260.
- [137]ZHANG N,DENG S,SUN Z, et al. Relation Adversarial Network for Low Resource Knowledge Graph Completion[C]//Proceedings of the 2020 World Wide Web Conference. 2020:1-12.
- [138]QIN P,WANG X,CHEN W, et al. Generative Adversarial Zero-Shot Relational Learning for Knowledge Graphs[C]//Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. 2020;8673-868.



LI Zhifei, born in 1993, Ph.D, lecture, is a member of China Computer Federation. His main research interests include knowledge graphs and graph neural networks.

(责任编辑:喻藜)