

知识图谱的最新进展、关键技术和挑战

马忠贵[✉], 倪润宇, 余开航

北京科技大学计算机与通信工程学院, 北京 100083

✉通信作者, E-mail: zhongguima@ustb.edu.cn

摘 要 围绕知识图谱的全生命周期技术, 从知识抽取、知识融合、知识推理、知识应用几个层面展开综述, 重点介绍了知识融合技术和知识推理技术. 通过知识抽取技术, 可从已有的结构化、半结构化、非结构化样本源以及一些开源的百科类网站抽取实体、关系、属性等知识要素. 通过知识融合, 可消除实体、关系、属性等指称项与实体对象之间的歧义, 得到一系列基本的事实表达. 通过本体抽取、知识推理和质量评估形成最终的知识图谱库. 按照知识抽取、知识融合、知识推理 3 个步骤对知识图谱迭代更新, 实现碎片化的互联网知识的自动抽取、自动关联和融合、自动加工, 从而拥有词条自动化链接、词条编辑辅助功能, 最终达成全流程自动化知识获取的目标. 最后, 讨论知识图谱未来的发展方向与可能存在的挑战.

关键词 知识工程; 知识图谱; 知识融合; 知识推理; 知识抽取与表示

分类号 TP391.1

Recent advances, key techniques and future challenges of knowledge graph

MA Zhong-gui[✉], NI Run-yu, YU Kai-hang

School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

✉ Corresponding author, E-mail: zhongguima@ustb.edu.cn

ABSTRACT The Google knowledge graph is a knowledge base used by Google and its services to enhance the search engine's results with information gathered from a variety of sources. Since its inception by Google to improve users' quality of experience of the search engine, the knowledge graph has become a term that is recently ubiquitously used in medical, education, finance, e-commerce and other industries to promote artificial intelligence (AI), which evolves from perceptual intelligence to cognitive intelligence. As a branch of knowledge engineering, a knowledge graph is based on the semantic network of knowledge engineering, and it combines the latest advancements achieved in machine learning, natural language processing, knowledge representation, and inference. Both academia and industries are showing keen interest in AI, and several studies are in progress under promotion of big data. With its powerful semantic processing and open interconnection capabilities, the knowledge graph can break the data isolation in different scenarios, and can generate application value in intelligent information services such as intelligent search and recommendation, intelligent question answering, and content distribution networks, thereby making information services more intelligent. The state of the art of knowledge graph technologies is outlined by introducing a process of building a knowledge graph. A knowledge graph is a structured representation of facts, consisting of entities, relations and semantic descriptions. A comprehensive summary of the overall lifecycle technologies of the knowledge graph is provided, including knowledge extraction, knowledge fusion, knowledge reasoning, and knowledge application. But the focus is on knowledge fusion and knowledge reasoning. Entities, relations, attributes, and other knowledge elements can be extracted from existing structured, semi-structured, unstructured data sources, and websites given in encyclopedia using knowledge extraction. With knowledge fusion, the ambiguity between referential items such as entities, relations, and attributes can be eliminated, and a series of basic facts can be obtained. The final knowledge base is formed through ontology extraction, knowledge reasoning and quality

收稿日期: 2020-02-28

基金项目: 中央高校基本科研业务费专项资金资助项目 (FRF-DF-20-12, FRF-GF-18-017B)

evaluation. Following the three steps of knowledge extraction, knowledge fusion, and knowledge reasoning, it can iteratively update the knowledge graph and realize full process automation knowledge acquisition, such as realizing the automatic extraction, automatic association and fusion, automatic processing of fragmented Internet knowledge, and realizing automatic linking of entries and auxiliary functions of entry editing. Finally, the future directions and possible challenges of the knowledge graph are discussed.

KEY WORDS knowledge engineering; knowledge graph; knowledge fusion; knowledge reasoning; knowledge extraction and representation

随着知识的不断积累和科学的飞速发展,人类社会进行了多次改变社会结构的重大生产力革命。最近的生产力革命正是由 Web 技术发展引发的信息革命。伴随着 Web 技术不断地演进与发展,人类即将迈向基于知识互联的崭新“Web3.0”时代^[1]。受语义网络(Semantic network)和语义网(Semantic web)的启发,Google 公司提出了知识图谱(Knowledge graph)^[2],目的是为了提升搜索引擎的智能能力,增强用户的搜索质量和体验。随后,这一概念被传播开来,并广泛应用于医疗、教育、金融、电商等行业中,推动人工智能从感知智能向认知智能跨越。目前,已经涌现出一大批知识图谱,其中国外具有代表性的有 YAGO^[3]、DBpedia^[4]、Freebase^[5]、NELL^[6]、Probase^[7]等;国内出现了开放知识图谱项目 OpenKG^[8],中文知识图谱 CN-DBpedia^[9]、zhishi.me^[10]等。知识图谱的本质是连接实体间关系的图,即揭示实体之间关系的语义网络^[11],普遍采用资源描述框架(Resource description framework, RDF)来描述知识。知识图谱全生命周期主要包括 3 种关键技术:(1)从样本源中获取数据,并将其表示为结构化知识的信息抽取与表示技术;(2)融合异源知识的信息融合技术;(3)根据知识图谱中已有的知识进行知识推理和质量评估。

近几年,越来越多的学者将目光聚焦在了认知智能上,知识图谱受到越来越广泛的关注。除了知识图谱的技术文章爆发式增长之外,综述文章也越来越多。文献[11]针对知识图谱的相关技术进行了全面解析,文献[12-13]综述了知识图谱核心技术的研究进展以及典型应用,文献[14]总结了面向知识图谱的推理方法并展望了未来的研究方向,文献[15]定义知识图谱与本体的关系并简述了已开发的国内外知识图谱。2019 年年末和 2020 年年初,国内有 3 本知识图谱的专著问世^[16-18],我们有了写作本论文的动机。与已有的综述文献相比,本文的主要贡献如下:梳理了知识图谱全生命周期技术,从知识抽取与表示、知识融合、知识推理、知识应用 4 个层面展开综述,建立方法论思

维。限于篇幅,针对知识图谱的 4 个关键技术进行了取舍,重点介绍了知识融合与知识推理技术的最新进展。同时,简要介绍了知识图谱目前的挑战并展望了未来的发展方向。

1 知识抽取与表示

对于知识图谱而言,首要的问题是:如何从海量的数据提取有用信息并将得到的信息有效表示并储存,就是所谓的信息抽取与表示技术。知识抽取与表示,也可以称为信息抽取,其目标主要是从样本源中抽取特定种类的信息,例如实体、关系和属性,并将这些信息通过一定形式表达并储存。对于知识图谱,一般而言采用 RDF 描述知识,形式上将有效信息表示为(主语,谓语,宾语)三元组的结构,某些文献中也表示为(头实体,关系,尾实体)的结构。针对信息抽取种类的不同,知识抽取又可分为实体抽取、关系抽取以及属性抽取。图 1 展示了知识图谱的技术架构。

实体抽取也称为命名实体识别,主要目标是从样本源中识别出命名实体。实体是知识图谱最基本的元素,实体抽取的完整性、准确率、召回率将直接影响知识图谱的质量^[12]。文献^[19]将实体抽取的方法归纳为 3 种:(1)基于规则与词典的方法。通常需要为目标实体编写相应的规则,然后在原始语料中进行匹配,Quimbaya 等^[20]提出了一个基于词典的实体抽取方法,并应用于电子健康记录。(2)基于统计机器学习的方法。主要利用数据来对模型进行训练,然后再利用训练好的模型去识别实体,Liu 等^[21]将 K 近邻(K-nearest neighbors, KNN)算法和线性条件随机场(Conditional random fields, CRF)模型结合来识别实体。(3)面向开放域的抽取方法。主要是针对海量网络数据,Jain 与 Pennacchiotti^[22]提出通过已知实体的语义特征来识别命名实体,并提出实体聚类的无监督开放域聚类算法。Zhang 与 Elhadad^[23]提出一个无监督的实体抽取方法,利用术语、语料库统计信息以及浅层语法知识从生物医学中抽取实体。

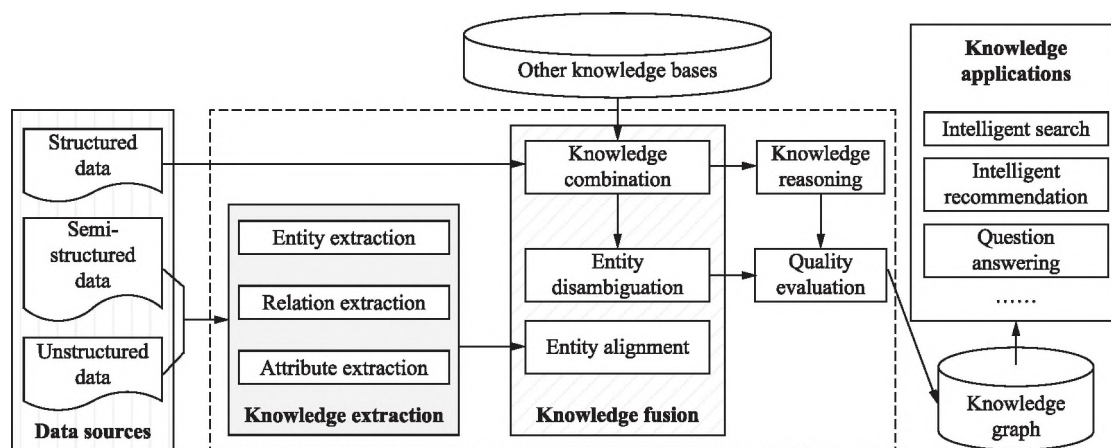


图 1 知识图谱的技术架构

Fig.1 Architecture of the Knowledge Graph

通过实体抽取获取的实体之间往往是离散且无关联的。通过关系抽取,可以建立起实体间的语义链接。关系抽取技术主要分为 3 种:(1)基于模板的关系抽取。使用模板通过人工或者机器学习的方法抽取实体关系,虽然准确率高且针对性强,但是其也具有不适用于大规模数据集、低召回率、难以维护等缺点。(2)基于监督学习的关系抽取。将大量人工标注的数据送入模型中训练,刘克彬等^[24]根据本体知识库训练模型,在开放数据集中对关系进行抽取,取得了极高的准确率。Sun 与 Han^[25]提出了名为 FTK(Feature-enriched tree kernel)的模型,利用设计好的有效特征训练,计算关系实例相似度并通过支持向量机对关系进行分类。(3)基于半监督或无监督学习的关系抽取。基于少量人工标注数据或者无标注数据,使用最大期望(Expectation maximization)等算法的半监督关系抽取方法进行关系抽取。Sun 与 Grishman^[26]提出名为 LGCo-Testing 的主动学习系统,Fu 与 Grishman^[27]则进一步优化了这个系统。Ji 等^[28]提出基于句子级注意力和实体描述的神经网络关系抽取模型 APCNNS。该模型实际采用了多示例学习的策略,将同一关系的样例句子组成样例包,关系分类是基于样例包的特征进行的。实验结果表明,该模型可以有效地提高远程监督关系抽取的准确率。在采用多示例学习策略时,有可能出现整个样例包都包含大量噪声的情况。针对这一问题,Feng 等^[29]提出了基于强化学习的关系分类模型 CNN-RL(Convolutional neural networks and reinforcement learning),该模型包括 2 个重要模块:样例选择器和关系分类器。实验结果表明:该模型获得了比句子级卷积神经网络和样例包级关系分类模型更好

的结果。最近的工作通过强化学习来处理句子级的去噪,这种学习将来自远程监督的标签视为事实。然而,很少有工作专注于直接校正噪声标签的标签级降噪。Sun 等^[30]提出了一种基于强化学习的标签去噪方法,用于远程监督关系提取。该模型由两个模块组成:抽取网络和策略网络。标签去噪的核心是在策略网络中设计一个策略来获取潜在标签,可以在其中选择使用远距离监督标签或从抽取网络预测标签的操作。实验结果表明,强化学习对于噪声标签的校正是有效的,并且所提出的方法可以胜过最新的关系抽取系统。

属性抽取的目标是补全实体信息,通过从样本源中获取实体属性信息或属性值。实体属性可以看作是属性值与实体间的一种关系,因而可以通过关系抽取的解决思路来获得。Wu 与 Weld^[31]利用百科类网站的半结构化数据,训练抽取模型,之后将抽取模型应用在非结构化数据中抽取属性。Chang 等^[32]提出了基于张量分解的关系抽取方法,这一方法也可以应用在属性抽取中,通过利用关于实体种类相应的领域知识来更好地获得实体所缺少的属性值。

2 知识融合

通过知识抽取与表示,初步获得了数量可观的形式化知识。由于知识来源的不同,导致知识的质量参差不齐,知识之间存在着冲突或者重叠。此时初步建立的知识图谱,知识的数量和质量都有待提高。应用知识融合技术对多源知识进行处理,一方面提升知识图谱的质量,另一方面丰富知识的存量。Zhao 等^[33]对最新的知识融合进行了综述。早期的知识融合是通过传统的数据融合方法

完成, Dong 等^[34]比较了传统的数据融合方法,选择了几种方法改良,并应用到知识融合中。随着知识图谱的飞速发展,目前也出现了专门的知识融合方法。下面从实体消歧、实体对齐和知识合并3个方面进行综述。

2.1 实体消歧

对于知识图谱中的每一个实体都应有清晰的指向,即明确对应某个现实世界中存在的事物。初步构建的知识图谱中,因数据来源复杂,存在着同义异名的实体。例如,名称为“乔丹”的实体既可以指美国著名篮球运动员,也可以指葡萄牙足球运动员,还可以指某个运动品牌。为了确保每一个实体有明确的含义,采用实体消歧技术来使得同名实体得以区分。

利用已有的知识库和知识图谱中隐含的信息来帮助进行语义消歧, Han 与 Zhao^[35]提出使用维基百科(Wikipedia)作为背景知识,通过利用 Wikipedia 的语义知识,例如社会关系来更精确地衡量实体间的相似性,从而提升实体消歧的效果。Sen^[36]提出了主题模型,利用知识库中存在的文本信息,学习共有实体组来实现实体集体消歧。Guo 与 Barbosa^[37]基于语义相似性的自然概念提出了两个针对集体消歧的方法。通过在知识库上知识子图中随机游走得到的概率分布来表示实体和文档的语义,之后基于迭代的贪婪逼近算法和学习排序的方法来进行实体消歧任务。Zhu 与 Iglesias^[38]提出了基于语义上下文相似度的命名实体消歧方法,基于上下文和知识图谱中实体的信息词之间的语义相似度来进行实体消歧。另外还提出了 Category2Vec 模型,将目录也用嵌入向量的形式表示出来。主要思想是候选实体和上下文单词间应存在语义联系,利用该联系来帮助选出正确的实体。

在线百科全书由专家和网络用户编写,有着高覆盖率和结构信息丰富的特点。Shen 等^[39]提出 LINDEN(A framework for Linking named entities with knowledge base via semantic knowledge)模型,同时利用 Wikipedia 和 WordNet,基于文本相似性和主题一致性进行实体消歧。Ratinov 等^[40]提出名为 GLOW(Global and local approaches of Wikipedia)的系统, GLOW 组合捕捉实体指称与 Wikipedia 题目间的相关性的本地模型和选择准确歧义语境的方法。统计 Wikipedia 中实体的频率作为候选实体的排序依据。Alokaili 与 Menai^[41]提出了基于支持向量机的集成学习来解决实体消歧问题,使用不

同的支持向量机的核函数来学习不同的集成学习算法,例如 bagging、boosting、voting 等。具体流程是将命名实体作为输入,根据 Wikipedia 中的知识生成候选实体,构造特征向量,最后送入集成学习模块里完成实体消歧。

值得一提的是, Agarwal 等^[42]提出了利用时间的实体消歧思路,通过计算实体的时序特征来和输入的命名实体上下文的时序比较,即使命名实体的上下文提供的信息不充分也可以完成实体消歧任务。Dong^[43]将基于相似度特征的随机森林模型和基于 XGBoost、基于逻辑回归以及基于神经网络的方法进行比较,随机森林模型不仅拥有极高的准确率和召回率,且不像 XGBoost 和神经网络那样容易受到超参数的影响,在实体消歧任务中表现突出。

2.2 实体对齐

在现实生活中,一个事物对应着不止一个称呼,例如,“中华人民共和国”和“中国”都对应于同一个实体。在知识图谱中也同样存在着同义异名的实体,通过实体对齐,将这些实体指向同一客观事物。苏佳林等^[44]提出基于决策树的自适应属性选择的实体对齐方法。通过联合学习将实体嵌入表示在一个向量空间后,由信息增益选出最优约束属性,训练实体对齐模型,计算最优约束属性相似度和实体语义相似度完成实体对齐。

Cheng 等^[45]提出了一个全自动的实体对齐框架,包括候选实体生成器、选择器和清理器,利用搜索引擎使用者的查询信息和查询后的点击记录,计算出实体间的相似度,完成实体对齐任务。Pantel 等^[46]提出了一个大规模相似性模型,在 MapReduce 框架下实施并且部署了超过 2000 亿从互联网上爬取得到的单词。通过计算 5 亿 terms 得到的相似度矩阵来进行实体对齐任务。Chakrabarti 等^[47]通过一个同义发现框架将实体相似性作为输入生成一个满足简单自然属性的同义词,提出了两种新的相似性度量法,并通过在 bing 系统上实际应用,发现可以有效识别同义词。Mudgal 等^[48]综述了基于深度学习的实体对齐方法,通过将这些方法分类,分别组合设计空间中属性嵌入、属性相似度表示、分类的各个方法,得到最具代表性的平滑倒词频(Smooth inverse frequency, SIF)、循环神经网络(Recurrent neural network, RNN)、Attention 和 Hybrid 共 4 种解决方案。

针对基于嵌入表示的实体对齐, Sun 等^[49]提出自举的方法解决标记训练数据不足的问题。根据

全局最优目标来标记可能的对齐,并在迭代中将其加入到训练数据中,不断训练嵌入表示模型。Guan 等^[50]发现基于监督学习的实体对齐方法,普遍在取得标签数据上需要花费大量时间,无监督学习方法的表现则很大程度地依赖于验证集上复杂的相似度衡量方式。Zhang 等^[51]从实体的多视角出发,利用实体的名称、实体间的关系、实体的属性的组合策略来学习实体的嵌入,并根据实体的表示来完成实体对齐任务。

2.3 知识合并

实体消歧和实体对齐更多的是关注知识图谱中的实体,从实体层面上通过各种方法来提升知识图谱的知识质量。知识合并则是从知识图谱整体层面上进行知识的融合,基于现存的知识库和知识图谱来扩大知识图谱的规模,丰富其中蕴含的知识。然而现存的知识库或者知识图谱都是各种机构或者组织根据自己的需求设计创建,其中的知识也存在着多样性和异构性,并且存在很多知识上的重复和错误,因而需要使用知识合并技术来解决这些问题^[52]。知识图谱的合并需要解决 2 个层面的问题:数据层的合并和模式层的合并^[53]。知识合并过程中可能出现的来自两个数据源的同一实体的属性值却不不同的现象,我们称这种知识合并过程中出现的现象为知识冲突。针对知识冲突问题,可以采用冲突检测与消解以及真值发现等技术进行消除,再将各个来源的知识关联合并为一个知识图谱。

冲突消解目前的研究方向是利用图谱自身存在的特征,Trisedya 等^[54]利用属性元组生成属性特征嵌入向量。使用成分函数来表示属性。将多个属性值都转化为单一向量,并将相似的属性映射为相似的向量表示。利用这些属性特征嵌入向量将两个图谱中的实体嵌入转化到同一个空间中,计算实体的相似性。Chen 等^[55]针对多语言知识图谱的合并,提出了利用实体描述的基于嵌入的半监督跨语言学习方法,在一个大规模数据集上通过迭代的方式联合训练一个多语言知识图谱嵌入模型和一个文字描述嵌入模型,训练模型完成图谱的合并。Cao 等^[56]提出多通道图神经网络模型,通过多个通道将两个知识图谱进行鲁棒编码。在每个通道中通过不同的关系加权方案来编码知识图谱,使用知识图谱补全和跨知识图谱注意力策略来分别修剪每个图谱中的独有实体,通过池化技术组合这些通道。

3 知识推理与质量评估

知识推理技术可以提升知识图谱的完整性和准确性。传统的知识推理方法拥有极高的准确率,但无法适配大规模知识图谱。针对知识图谱数据量大、关系复杂的特点,提出了面向大规模知识图谱的知识推理方法,并归纳为以下 4 类^[14, 57]: (1) 基于图结构和统计规则挖掘的推理; (2) 基于知识图谱表示学习的推理; (3) 基于神经网络的推理; (4) 混合推理。

3.1 基于图结构和统计规则挖掘的推理

受传统推理地启发,基于知识图谱的图结构以及挖掘蕴藏在知识图谱中的规则进行推理的方法得以提出,并在知识推理任务上取得一定效果。Lao 与 Cohen^[58]提出了路径排序算法(Path ranking algorithm, PRA),将实体间的路径作为特征,通过随机行走算法来计算实体间是否存在潜在的关系。Wang 等^[59]设计了耦合路径排序算法(Coupled path ranking algorithm, CPRA),并提出一种全新的逐次聚合的策略,通过这一策略使得具有强相关度的关系聚合在一起。使用多任务学习策略预测聚合后的关系。Xiong 等^[60]针对多跳关系路径的学习提出使用强化学习的框架,设计了一个具有连续基于知识图谱嵌入状态的策略 Agent,通过 Agent 在知识图谱的向量空间中寻找最有潜力的关系加入路径完成推理。

Cohen^[61]针对如何将知识整合到梯度学习的系统的问题,描述了一个概率演绎的数据库 Tensorlog,通过可微分的过程来进行推理。Yang 等^[62]研究了基于学习一阶概率逻辑规则进行知识库推理的问题。受到 Tensorlog 的启发,提出了名为神经逻辑规划的框架,将一阶逻辑规则的参数和结构整合到一个端到端的可微分模型中。设计了一个带 Attention 机制和存储功能的神经控制系统来学习组合那些用于完成推理的规则。Kampffmeyer 等^[63]提出深度图传播模型,在利用图结构的便利的同时解决知识过于稀疏的问题。

3.2 基于知识图谱表示学习的推理

表示模型将知识图谱中相应的实体和关系用向量、矩阵或者张量的形式表示,表示后进行运算完成知识推理任务。因其简单高效且适应于大规模知识图谱推理的特点而不断发展。

3.2.1 基于距离的推理模型

Bordes 等^[64]提出了 TransE 模型,将所有的实体和关系表示为同一个空间下的向量,假设事实元

组中头实体向量和关系向量之和应该约等于尾实体的向量。通过随机替换事实元组中的某一项来构建负例。计算元组中头向量和关系向量的和向量与尾向量的距离作为候选实体的得分。尽管 TransE 模型简单且有效,但其仍然具有许多缺陷,因而衍生出很多基于该模型的方法。Wang 等^[65]提出 TransH 模型,每一个关系都有一个特定的超平面,头实体向量和尾实体向量投影至特定的关系超平面计算事实元组得分。Lin 等^[66]提出 TransR 模型,针对特定关系引入了空间。Xiao 等^[67]提出了 ManifoldE 模型,引入了特定关系参数。尾实体向量有效范围是以头实体向量和关系实体的和向量为中心,以特定关系参数为半径的一个超球面。Feng 等^[68]提出的 TransF 模型和 ManifoldE 模型有着相似的思路,放宽了 TransE 中对实体关系向量的要求,仅需要头实体向量位于尾实体向量和负的关系向量的和向量的方向上,同时尾实体向量也位于头实体向量和关系向量的和向量的方向上。

Kzaemi 与 Poole^[69]提出 Simple 模型,允许实体拥有两个独立学习的向量表示,而关系由一个向量表示。Ebisu 与 Ichise^[70]提出了 TorusE 嵌入模型,将 TransE 的思想应用在李群(Lie group)理论中的圆环面上,即在圆环面上计算表示向量间的距离来取得元组得分。Xu 与 Li^[71]提出 DihEdral 模型,针对性地增强了知识推理的可解释性,通过离散值将关系建模成组的元素,显著地缩小了解空间。Sun 等^[72]提出 RotatE 模型,将关系看作是从头实体向量向尾实体向量的旋转,元组得分通过计算旋转后的头实体向量和尾实体向量的距离得到。Zhang 等^[73]引入超复数的概念,提出了 QuatE 模型,通过一个拥有三个虚部的超复数来表示知识图谱中的实体和关系。与 RotatE 想法类似,QuatE 模型将关系看作超复数平面下头实体到尾实体的旋转。

3.2.2 基于语义匹配的推理模型

Nickel 等^[74]提出的 RESCAL 模型将实体和向量联系起来,从而捕捉其中隐含的语义,潜在因子间的相互作用建模后得到关系表示矩阵,计算实体向量与关系矩阵的乘积来得到元组得分。Yang 等^[75]提出 DistMult 模型,每一个关系都表示为向量,再将向量转化为对角矩阵,通过计算头尾实体向量与关系对角矩阵的乘积得到元组的得分。Trouillon 等^[76]提出 Complex 模型,引入复数嵌入针对不对称关系建模。在 Complex 模型中,实体

和关系都由复平面中的向量表示,计算头实体向量和根据关系向量建立的对角矩阵以及尾实体向量的共轭这三者的乘积,结果的实部作为元组的得分。Liu 等^[77]提出 ANALOGY 模型,利用实体和关系的类比性质来建模,实体由嵌入空间中的向量表示,将关系矩阵处理得到一系列稀疏的对角矩阵,减少了关系矩阵的参数。将头尾实体向量与关系矩阵的积作为元组得分。

Balazevic 等^[78]提出了基于 KKT(Karush Kuhn Tucker)分解的 tuckER 模型,将所有实体和关系分别表示为行向量嵌入矩阵,从这两个矩阵中取出头尾实体向量和关系向量,将这些向量和一个核心张量相乘得到元组的得分。针对大部分现存的基于知识图谱嵌入的模型,Kristiadi 等^[79]研究了如何将文字信息整合到现存的表示模型中去,提出了 LiteralE 模型,在实体的嵌入表示上加入文字信息,用实体表示和文字信息的联合表示取代原本模型的单独的实体表示。Zhang 等^[80]提出了 CrossE 模型,基于向量表示实体和关系,生成多个元组的特定嵌入即交互嵌入。由交互表示和尾实体的嵌入表示的匹配程度给出元组得分。

基于表示学习的知识推理模型的比较如表 1 所示。

3.3 基于神经网络的推理

基于神经网络的推理方法将知识图谱中事实元组表示为向量形式送入神经网络中,通过训练神经网络不断提高事实元组的得分,最终通过输出得分选择候选实体完成推理。Socher 等^[81]提出适应于实体间关系推理的神经张量网络(Neural tensor networks, NTN)模型,用双线性张量层取代神经网络层,实体通过连续的词向量平均表示进而提升模型的表现。Neelakantan 等^[82]使用循环神经网络来建模知识图谱中的分布式语义的多跳路径。Das 等^[83]主要是将符号逻辑推理中丰富的多步推理与神经网络的泛化能力相结合。通过学习实体、关系和实体的种类来联合推理,并使用神经网络注意力建模来整合多跳路径。在单层 RNN 中分享参数来表示所有关系的逻辑组成。Graves 等^[84]建立了可微神经计算机模型,将神经网络和记忆系统结合起来,将通过样本学习到的知识储存起来并进行快速知识推理。

Dettmers 等^[85]针对知识图谱中大规模与过拟合的问题,设计了参数简洁且计算高效的二维卷积神经网络(Convolutional 2D, ConvE)模型。Vashishth 等^[86]基于特征排列、新的特征变形以及

表 1 部分基于表示学习的知识推理模型

Table 1 Some knowledge reasoning models based on representation learning

Method	Scoring function	The entity representations	The relation representation
TransE	$-\ h+t-r\ _{1/2}$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^d$
ManifoldE	$-\left(\ h+t-r\ _2^2 - \theta_r^2\right)^2$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^d$
SimplE	$\frac{1}{2} \left(\langle h_{e_i}, v_r, t_{e_j} \rangle + \langle h_{e_j}, v_{r^{-1}}, t_{e_i} \rangle \right)$	$h_e, t_e \in \mathbb{R}^d$	$v_r \in \mathbb{R}^d$
RotatE	$\ h \circ r - t\ $	$h, t \in \mathbb{C}^d$	$r \in \mathbb{C}^d$
QuatE	$h \otimes \frac{r}{ r } \cdot t$	$h, t \in \mathbb{H}^d$	$r \in \mathbb{H}^d$
RESCAL	$h^T M_r t$	$h, t \in \mathbb{R}^d$	$M_r \in \mathbb{R}^{d \times d}$
DistMult	$h^T \text{diag}(r) t$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^d$
ComplEx	$\text{Re}(h^T \text{diag}(r) \bar{t})$	$h, t \in \mathbb{C}^d$	$r \in \mathbb{C}^d$
ANALOGY	$h^T M_r t$	$h, t \in \mathbb{R}^d$	$M_r \in \mathbb{R}^{d \times d}$
CrossE	$\sigma(\tanh(c_t \circ h + c_t \circ h \circ r + b) t^T)$	$h, t \in \mathbb{R}^d$	$r \in \mathbb{R}^d$

循环卷积提出 InteractE 模型。InteractE 模型通过使用多种排列输入,更简单的特征变形方法以及循环卷积来取得比 ConvE 更显著的效果。

3.4 混合推理

对于上面的几类知识推理的方法,各有其优势与缺点,于是考虑结合多种方法的优势来提升推理效果,进而提出了混合推理方法。Guo 等^[87]提出学习规则增强关系来补全知识图谱的方法,使用规则来进一步改善传统关系学习得到的推理结果,提升知识推理的准确性。Lu 等^[88]提出了基于强化学习建模的逻辑概率的知识表示和推理模型,同时在已知的知识和由强化学习整合的经验上进行推理来训练强化学习的 Agent。Xie 等^[89]提出一种利用实体描述的知识表示学习的方法,使用了连续词袋模型和深度卷积模型来编码实体的描述语义。之后进一步学习通过三元组和三元组中实体的描述来学习表示知识。并利用学习到的知识来完成知识推理任务。Wang^[90]提出规则嵌入神经网络(The rule-embedded neural network, ReNN)。ReNN 基于局部的推理检测局部模式,由局部模式领域知识的规则来生成规则调制映射。针对规则引起的优化问题,采用两阶段优化策略。引入规则解决了传统神经网络必须受限于数据集的问题,从而提升了推理的准确率。

Zhang 等^[91]提出了一个名为 IterE 的迭代学习嵌入和规则的框架,目标是同时学习实体嵌入表示和规则,并利用它们各自的优势来弥补对方的不足。Nie 与 Sun^[92]组合了隐形特征和图特征的优势提出了一个名为文本强化型知识图谱嵌入

(Text-enhanced knowledge graph embedding, TKGE) 的组合模型,通过实体、关系和文本来提升推理的表现。Guan 等^[93]基于一个常识图的常识概念信息提出了一个常识伴随的知识图谱嵌入(Knowledge graph embedding with concepts, KEC)模型,将来自于知识图谱的事实元组通过常识概念信息修正,从而使得模型不仅仅关注实体间的关联性还有实体存在的常识概念。因此这个模型具有明确的语义性。

4 类知识推理方法对比如表 2 所示。

3.5 质量评估

通过质量评估技术来对新知识进行筛选,是构建知识图谱中必不可少的环节。Mendes 等^[94]提出了 Sieve,用于简化生成高质量数据的任务,并整合进了链接数据整合框架(Linked data integration framework, LDIF)中,包括一个质量评估模型和一个数据融合模型。质量评估主要利用用户选择的质量因子,通过用户配置的得分函数生成质量得分。数据融合使用质量得分来处理用户设置的冲突消解任务。Fader 等^[95]基于来自网络或 Wikipedia 的 1000 个句子中人工标注的实例来训练 ReVerb 系统的置信函数,通过一个逻辑回归分类器来评估每一个通过 ReVerb 系统抽取得到的实例的置信度。Google 的 Knowledge vault 项目^[96],通过统计全球网络中抽取数据的频率作为评估信息可信度的依据,并通过已有知识库中的知识来修正可信度,这一方法有效降低了评估数据结果的不确定性,从而提升了知识的质量水平。Tan 等^[97]提出了一个名为 CQUAL(Contribution quality predictor)的

表 2 4 类知识推理方法对比

Table 2 Comparisons of 4 kinds of knowledge reasoning methods

Reasoning methods	Advantage	Disadvantage	Typical model
Knowledge reasoning based on graph structure and statistical rule mining	The advantages of graph structure and rules can significantly improve the accuracy of knowledge reasoning	Large-scale knowledge graphs have complex graph structures and rules are not easy to obtain; noise rules can mislead knowledge reasoning	PRA AMIE TensoLog
Knowledge reasoning based on representation learning	Simple and efficient, suitable for large-scale knowledge graph	Does not consider the deeper information in the knowledge graph, which limits its accuracy of reasoning	RESCAL TransE
Knowledge reasoning based on the neural network	Outstanding learning ability and reasoning ability	High complexity, huge number of parameters, and poor interpretability	NTN
Knowledge reasoning based on hybrid methods	Combines the advantages of several inference methods, so its performance is excellent	Most methods are just shallow fusion, not taking full advantage of their respective methods	TKGE

方法来自动预测用户提交至知识库的知识的质
量,主要依据提交用户的领域、提交历史、以及历史
准确率等数据。实验表明这一方法拥有很高的
准确率和召回率。

4 知识图谱应用

知识图谱技术提出之后,因其具有的语义处
理和开放互联的能力,以及其简洁灵活的表达方
式等优势,受到了广泛关注。知识图谱技术的发展
得益于自然语言处理、互联网等技术的发展,而不
断完善的知识图谱技术也可以应用到自然语言处
理、智能问答系统、智能推荐系统等技术中,进一
步促进这些技术的发展,而这些技术以及知识图
谱技术又可以进一步应用在诸如医疗、金融、电
商等垂直行业或领域内,帮助促进行业发展^[16-17]。

构建完备的知识图谱可以帮助自然语言理解
技术发展。针对文本分类问题,Wang 等^[98]首先利
用知识库中的知识将短文本概念化,获得短文本
的嵌入表示后送入卷积神经网络中进行分类。
Lagon 等^[99]提出了知识图谱语言模型,一种拥有
从知识图谱中选择和复制知识的神经语言模型。

智能问答系统可以依靠知识图谱中的知识
来回答查询。Bauer 等^[100]利用关系路径从常识网
络中获取背景常识知识,之后利用多注意力机制
完成多跳推理并通过一个指针生成译码器来合成
问题的答案。朱宗奎等^[101]针对中文知识图谱问答
系统,将 BERT(Bidirectional encoder representations
from transformers)模型和双向长短期记忆网络结
合,之后通过条件随机场模型来预测字符标签,从
而识别出问题中的实体并链接到知识网络中,最
后完成答案的搜索。

知识图谱可作为外部信息整合至推荐系统
中,使得推荐系统获得推理能力。通过利用知识图

谱中诸如实体、关系的信息,许多研究进一步基
于嵌入正则化来提升推荐效果。Wang 等^[102]将图注
意网络应用于实体-关系和用户-物品图的协作知
识图谱上,提出了名为知识图谱注意力网络的模
型,在端到端的模式下通过嵌入传播和基于注意
的聚合对建模知识图谱中的高阶连通性建模。

在垂直行业或领域内,知识图谱已开始应用。
在医疗领域,通过提供更加精确规范的行业数据
以及更加丰富的表达,帮助非行业相关人员获取
医疗知识的同时也帮助行业人员更直观快捷获取
所需医疗知识。在金融领域,借助知识图谱检测数
据的不一致性,来识别潜在的欺诈风险。同时,利
用知识图谱技术分析招股书、年报、公司公告等
金融报告,建立公司和人物的关系,在此基础上做
更进一步的研究和更优的决策。在电商领域,阿里
巴巴已经通过应用知识图谱,建立商品间的关联
信息,为用户提供更全面的商品信息和更智能
化的推荐,从而提升用户的购物服务与体验。同
时,知识图谱也在教育、科研、军事等领域中广泛
应用。

5 知识图谱在知识融合、推理与应用中的
挑战与展望

自谷歌提出知识图谱概念至今,这项技术一
直受到广泛的关注。随着深度学习、自然语言处
理等相关领域的发展,知识图谱的研究热度不断
增加。不可忽略的是,知识图谱发展至今,知识融
合、知识推理等知识图谱关键技术以及知识图谱
的应用仍面临许多挑战。

知识融合技术是知识图谱的关键技术之一。
知识融合主要任务是将新获得的知识融入知识图
谱中。保证知识图谱知识准确率的前提下高效地
引入新知识,是知识融合的关键。存在的挑战如

下: (1) 为了保证融合后知识图谱的质量, 首先要提升知识评估的能力. 现存的知识评估方法大都是针对静态知识进行评估, 缺少动态知识评估手段是目前知识评估面临的一大挑战. (2) 要解决由自然语言的特殊性引发的知识冗余和缺失问题. 当知识图谱不能准确将具有同义异名的实体对齐或将同名异义的实体消歧就会导致知识图谱中出现知识冗余或缺失. (3) 目前, 因自然语言的复杂性, 在单一语言的背景下实体对齐和实体消歧的准确率仍然有待提高, 针对多语言实体对齐或消歧更是一大挑战.

知识推理技术也是知识图谱的关键技术之一, 通过已知的知识推理获得新知识来完善知识图谱. 存在的挑战如下: (1) 知识推理的主要对象多是二元关系, 通常处理多元关系的方法是将其拆分为二元关系进行推理, 然而将多元关系拆分会损失结构信息, 如何尽可能完整地利用多元关系中复杂的隐含信息推理是知识推理的一大挑战. (2) 现有的知识推理往往都是基于大量高质量的数据集训练推理模型, 在相应的测试集中测试优化模型来完成推理. 除了数据集获取成本高的问题, 通过数据集训练的模型的泛化能力也极为有限, 而现实世界中人类通过少量样本学习即可完成推理. 如何模仿人脑机制实现小样本或零样本学习知识推理也是一大挑战. (3) 知识图谱中知识的有效性往往受到时间空间等动态因素约束, 如何合理利用知识的动态约束信息完成动态推理也是知识推理的一大挑战.

知识的表达、存储与查询将是贯穿知识图谱应用始终的问题. 存在的挑战如下: (1) 目前, 应用行业领域的知识图谱因为很大程度上依赖人工的参与构建, 成本高昂. 大多数研究工作主要针对知识图谱的半自动构建^[103], 如何自动构建高质量知识图谱是知识图谱应用所面临的一大挑战. (2) 知识拥有指导功能, 利用知识图谱中的知识引导机器学习中的数据学习, 从而降低数据依赖打破数据红利损耗殆尽后的僵局, 是知识图谱应用面临的一大挑战. (3) 利用人类易懂的符号化知识图谱, 解释各类机器学习特别是深度学习的过程, 补足其在可解释性方面的短板, 也是知识图谱应用面临的一大挑战. (4) 未来, 能否应用知识图谱中的知识, 作为已知的经验, 通过训练构建人工智能层面上的心智模型, 同样是知识图谱应用的一大挑战.

知识图谱意在模仿人类的认知方式, 构建属

于机器的知识库, 是实现机器认知智能的关键技术, 也是网络大数据时代中利用大数据的关键技术. 本文从知识图谱构建过程中的关键技术出发, 简略研究了知识的抽取与表示, 重点分析了知识融合和知识推理技术的研究成果. 然而众多研究成果实用性不强, 知识图谱虽然已经出现了诸如 Magi^[104] 这样的理论实践者, 但距离知识图谱成为机器大脑知识库、实现机器认知智能的终极目标还有不小的距离. 未来的研究中, 基于网络数据自动构建的知识图谱将成为主流. 因而需要进一步提高知识抽取、知识融合和知识推理技术的准确性, 确保获取知识的质量; 同时提高这些技术的效率, 从而保证面对大规模数据量级时的实用性. 同时, 知识图谱虽然已经在公安情报分析、反金融欺诈等实际问题中开始应用, 但是其具有的巨大潜力仍有待挖掘, 如何将知识图谱技术应用在生活中的各个方面, 也将是未来的主要研究方向. 除此之外, 目前存在着的大量知识图谱, 大多有着结构或者语言上的差异, 这种差异增大了知识图谱应用的难度, 制定行业规范、整合各个知识图谱、构建通用知识图谱, 也是未来知识图谱研究的方向之一.

参 考 文 献

- [1] Sheth A, Thirunarayan K. *Semantics Empowered Web 3.0: Managing Enterprise, Social, Sensor, and Cloud-Based Data and Services for Advanced Applications*. Vermont: Morgan & Claypool, 2012
- [2] Singhal A. Official Google Blog: Introducing the Knowledge Graph: things, not strings[J/OL]. *Google Inc* (2012-05-02)[2020-02-28]. <http://googleblog.blogspot.pt/2012/05/introducing-knowledge-graph-things-not.html>
- [3] Hoffart J, Suchanek F M, Berberich K, et al. YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia. *Artif Intelligence*, 2013, 194: 28
- [4] Auer S, Bizer C, Kobilarov G, et al. DBpedia: a nucleus for a web of open data // *Proceedings of the 6th International The Semantic Web and 2nd Asian Conference on Asian Semantic Web Conference*. Busan, 2007: 722
- [5] Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge // *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. Vancouver, 2008: 1247
- [6] Carlson A, Betteridge J, Kisiel B, et al. Toward an architecture for never-ending language learning // *Twenty-Fourth AAAI Conference on Artificial Intelligence*. Atlanta, 2010: 1306

- [7] Wu W T, Li H S, Wang H X, et al. Probase: A probabilistic taxonomy for text understanding // *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*. Scottsdale, 2012: 481
- [8] Peak Labs. About OpenKG. CN[EB/OL]. *Professional Committee of Language and Knowledge Computing, Chinese Information Processing Society of China*(2018-06-01)[2020-02-28]. http://wp.openkg.cn/?page_id=77PeakLabs
(关于OpenKG. CN[EB/OL]. 中国中文信息学会语言与知识计算专业委员会(2018-06-01)[2020-02-28]. http://wp.openkg.cn/?page_id=77PeakLabs)
- [9] Xu B, Xu Y, Liang J Q, et al. CN-DBpedia: a never-ending Chinese knowledge extraction system // *Proceedings of the 30th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Arras, 2017: 428
- [10] Niu X, Sun X R, Wang H F, et al. Zhishi. me-weaving Chinese linking open data // *Proceedings of the 10th International Semantic Web Conference*. Bonn, 2011: 205
- [11] Liu Q, Li Y, Duan H, et al. Knowledge graph construction techniques. *J Comput Res Dev*, 2016, 53(3): 582
(刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述. 计算机研究与发展, 2016, 53(3): 582)
- [12] Xu Z L, Sheng Y P, He L R, et al. Review on knowledge graph techniques. *J Univ Electron Sci Technol China*, 2016, 45(4): 589
(徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述. 电子科技大学学报, 2016, 45(4): 589)
- [13] Qi G L, Gao H, Wu T X. The research advances of knowledge graph. *Technol Intelligence Eng*, 2017, 3(1): 4
(漆桂林, 高桓, 吴天星. 知识图谱研究进展. 情报工程, 2017, 3(1): 4)
- [14] Guan S P, Jin X L, Jia Y T, et al. Knowledge reasoning over knowledge graph: A survey. *J Software*, 2018, 29(10): 2966
(官赛萍, 靳小龙, 贾岩涛, 等. 面向知识图谱的知识推理研究进展. 软件学报, 2018, 29(10): 2966)
- [15] Huang H Q, Yu J, Liao X, et al. Review on knowledge graphs. *Comput Syst Appl*, 2019, 28(6): 1
(黄恒琪, 于娟, 廖晓, 等. 知识图谱研究综述. 计算机系统应用, 2019, 28(6): 1)
- [16] Xiao Y H. *Knowledge Graph: Concepts and Techniques*. Beijing: Publishing House of Electronics Industry, 2020
(肖仰华. 知识图谱: 概念与技术. 北京: 电子工业出版社, 2020)
- [17] Wang H F, Qi G L, Chen H J. *Knowledge Graph: Methodology, Practice and Application*. Beijing: Publishing House of Electronics Industry, 2019
(王昊奋, 漆桂林, 陈华钧. 知识图谱: 方法、实践与应用. 北京: 电子工业出版社, 2019)
- [18] Yan S, Wei K, Hong W F. *Knowledge Graph: Technology and Application*. Beijing: Posts & Telecom Press, 2019
(闫树, 魏凯, 洪万福. 知识图谱技术与应用. 北京: 人民邮电出版社, 2019)
- [19] Zhao J, Liu K, Zhou G Y, et al. Open information extraction. *J Chin Inform Process*, 2011, 25(6): 98
(赵军, 刘康, 周光有, 等. 开放式文本信息抽取. 中文信息学报, 2011, 25(6): 98)
- [20] Quimbaya A P, Munera A S, Rivera R A G, et al. Named entity recognition over electronic health records through a combined dictionary-based approach. *Procedia Comput Sci*, 2016, 100: 55
- [21] Liu X H, Zhang S D, Wei F R, et al. Recognizing named entities in tweets // *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics. Portland, 2011: 359
- [22] Jain A, Pennacchiotti M. Open entity extraction from web search query logs // *Proceedings of the 23rd International Conference on Computational Linguistics*. Beijing, 2010: 510
- [23] Zhang S D, Elhadad N. Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts. *J Biomed Informatics*, 2013, 46(6): 1088
- [24] Liu K B, Li F, Liu L, et al. Implementation of a kernel-based Chinese relation extraction system. *J Comput Res Dev*, 2007, 44(8): 1406
(刘克彬, 李芳, 刘磊, 等. 基于核函数中文关系自动抽取系统的实现. 计算机研究与发展, 2007, 44(8): 1406)
- [25] Sun L, Han X P. A feature-enriched tree kernel for relation extraction // *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Baltimore, 2014: 61
- [26] Sun A, Grishman R. Active learning for relation type extension with local and global data views // *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, New York, 2012: 1105
- [27] Fu L S, Grishman R. An efficient active learning framework for new relation types // *Proceedings of the Sixth International Joint Conference on Natural Language Processing (IJCNLP)*. Nagoya, 2013: 692
- [28] Ji G L, Liu K, He S Z, et al. Distant supervision for relation extraction with sentence-level attention and entity descriptions // *31st AAAI Conference on Artificial Intelligence*. San Francisco, 2017: 3060
- [29] Feng J, Huang M L, Zhao L, et al. Reinforcement learning for relation classification from noisy data // *32nd AAAI Conference on Artificial Intelligence*. New Orleans, 2018: 5779
- [30] Sun T T, Zhang C H, Ji Y, et al. Reinforcement learning for distantly supervised relation extraction. *IEEE Access*, 2019, 7: 98023
- [31] Wu F, Weld D S. Autonomously semantifying Wikipedia // *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*. Lisbon, 2007: 41
- [32] Chang K W, Yih W T, Yang B S, et al. Typed tensor decomposition of knowledge bases for relation extraction // *2014 Conference on Empirical Methods in Natural Language Processing*. Doha, 2014: 1568

- [33] Zhao X J, Jia Y, Li A P, et al. Multi-source knowledge fusion: A survey // *2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC)*. Hangzhou, 2019: 119
- [34] Dong X L, Gabrilovich E, Heitz G, et al. From data fusion to knowledge fusion. *Proc VLDB Endowment*, 2014, 7(10): 881
- [35] Han X P, Zhao J. Named entity disambiguation by leveraging wikipedia semantic knowledge // *Proceedings of the 18th ACM Conference on Information and Knowledge Management*. Hong Kong, 2009: 215
- [36] Sen P. Collective context-aware topic models for entity disambiguation // *Proceedings of the 21st International Conference on World Wide Web*. New York, 2012: 729
- [37] Guo Z C, Barbosa D. Robust named entity disambiguation with random walks. *Semantic Web*, 2018, 9(4): 459
- [38] Zhu G G, Iglesias C A. Exploiting semantic similarity for named entity disambiguation in knowledge graphs. *Expert Syst Appl*, 2018, 101: 8
- [39] Shen W, Wang J Y, Luo P, et al. Linden: linking named entities with knowledge base via semantic knowledge // *Proceedings of the 21st International Conference on World Wide Web*. New York, 2012: 449
- [40] Ratinov L A, Roth D, Downey D, et al. Local and global algorithms for disambiguation to Wikipedia // *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Portland, 2011: 1375
- [41] Alokaili A, Menai M E B. SVM ensembles for named entity disambiguation. *Computing*, 2020, 102(4): 1051
- [42] Agarwal P, Strötgen J, Del Corro L, et al. DiANED: time-aware named entity disambiguation for diachronic corpora // *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne, 2018: 686
- [43] Dong X L. Challenges and innovations in building a product knowledge graph // *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, 2018: 2869
- [44] Su J L, Wang Y Z, Jin X L, et al. Entity alignment method based on adaptive attribute selection. *J Shandong Univ Eng Sci*, 2020, 50(1): 14
(苏佳林, 王元卓, 靳小龙, 等. 自适应属性选择的实体对齐方法. 山东大学学报: 工学版, 2020, 50(1): 14)
- [45] Cheng T, Lauw H W, Paparizos S. Entity synonyms for structured web search. *IEEE Trans Knowl Data Eng*, 2012, 24(10): 1862
- [46] Pantel P, Crestan E, Borkovsky A, et al. Web-scale distributional similarity and entity set expansion // *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*. Stroudsburg, 2009: 938
- [47] Chakrabarti K, Chaudhuri S, Cheng T, et al. A framework for robust discovery of entity synonyms // *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Beijing, 2012: 1384
- [48] Mudgal S, Li H, Rekatsinas T, et al. Deep learning for entity matching: a design space exploration // *Proceedings of the 2018 International Conference on Management of Data*. Houston, 2018: 19
- [49] Sun Z Q, Hu W, Zhang Q H, et al. Bootstrapping entity alignment with knowledge graph embedding // *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*. Stockholm, 2018: 4396
- [50] Guan S P, Jin X L, Wang Y Z, et al. Self-learning and embedding based entity alignment. *Knowl Inform Syst*, 2019, 59(2): 361
- [51] Zhang Q H, Sun Z Q, Hu W, et al. Multi-view knowledge graph embedding for entity alignment // *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*. Macao, 2019: 5429
- [52] Zhuang Y, Li G L, Feng J H. A survey on entity alignment of knowledge base. *J Comput Res Dev*, 2016, 53(1): 165
(庄严, 李国良, 冯建华. 知识库实体对齐技术综述. 计算机研究与发展, 2016, 53(1): 165)
- [53] Deshpande O, Lamba D S, Tourn M, et al. Building, maintaining, and using knowledge bases: a report from the trenches // *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. New York, 2013: 1209
- [54] Trisedya B D, Qi J Z, Zhang R. Entity alignment between knowledge graphs using attribute embeddings // *Proceedings of the AAAI Conference on Artificial Intelligence*. Honolulu, 2019: 297
- [55] Chen M H, Tian Y T, Chang K W, et al. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment // *Twenty-Seventh International Joint Conference on Artificial Intelligence {IJCAI-18}*. Stockholm, 2018: 3998
- [56] Cao Y X, Liu Z Y, Li C J, et al. Multi-channel graph neural network for entity alignment // *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, 2019: 1452
- [57] Lu R, Cai Z P, Zhao S. A survey of knowledge reasoning based on KG. *IOP Conf Ser Mater Sci Eng*, 2019, 569(5): 052058
- [58] Lao N, Cohen W W. Relational retrieval using a combination of path-constrained random walks. *Mach Learn*, 2010, 81(1): 53
- [59] Wang Q, Liu J, Luo Y F, et al. Knowledge base completion via coupled path ranking // *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, 2016: 1308
- [60] Xiong W H, Hoang T, Wang W Y. DeepPath: a reinforcement learning method for knowledge graph reasoning // *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Denmark, 2017: 575
- [61] Cohen W W. Tensorlog: A differentiable deductive database [J/OL]. *arXivpreprint*(2016-07-19)[2020-02-28]. <https://arxiv.org/>

- abs/1605.06523
- [62] Yang F, Yang Z, Cohen W W. Differentiable learning of logical rules for knowledge base reasoning // *31st Conference on Neural Information Processing Systems*. Long Beach, 2017: 2319
- [63] Kampffmeyer M, Chen Y B, Liang X D, et al. Rethinking knowledge graph propagation for zero-shot learning // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, 2019: 11487
- [64] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data // *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*. Lake Tahoe, 2013: 2787
- [65] Wang Z, Zhang J W, Feng J L, et al. Knowledge graph embedding by translating on hyperplanes // *Twenty-Eighth AAAI Conference on Artificial Intelligence*. Québec, 2014: 1112
- [66] Lin Y K, Liu Z Y, Sun M S, et al. Learning entity and relation embeddings for knowledge graph completion // *Twenty-Ninth AAAI Conference on Artificial Intelligence*. Austin, 2015: 2181
- [67] Xiao H, Huang M L, Yu H, et al. From one point to a manifold: orbit models for knowledge graph embedding[J/OL]. *arXiv preprint* (2017-06-17)[2020-02-28]. <https://arxiv.org/abs/1512.04792v5>
- [68] Feng J, Huang M L, Wang M D, et al. Knowledge graph embedding by flexible translation // *Fifteenth International Conference on Principles of Knowledge Representation and Reasoning*. Palo Alto, 2016: 557
- [69] Kazemi S M, Poole D. Simple embedding for link prediction in knowledge graphs // *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Montréal, 2018: 4289
- [70] Ebisu T, Ichise R. Toruse: Knowledge graph embedding on a lie group // *The Thirty-Second AAAI Conference on Artificial Intelligence*. Louisiana, 2018: 1819
- [71] Xu C R, Li R J. Relation embedding with dihedral group in knowledge graph // *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, 2019: 263
- [72] Sun Z Q, Deng Z H, Nie J Y, et al. Rotate: Knowledge graph embedding by relational rotation in complex space[J/OL]. *arXiv preprint* (2019-02-26)[2020-02-28]. <https://arxiv.org/pdf/1902.10197.pdf>
- [73] Zhang S, Tay Y, Yao L N, et al. Quaternion knowledge graph embeddings // *Thirty-third Conference on Neural Information Processing Systems*. Vancouver, 2019: 2735
- [74] Nickel M, Tresp V, Krieger H P. A three-way model for collective learning on multi-relational data // *The 28th International Conference on Machine Learning The International Conference on Machine Learning*. Bellevue, Washington, 2011: 809
- [75] Yang B S, Yih W T, He X D, et al. Embedding entities and relations for learning and inference in knowledge bases[J/OL]. *arXiv preprint* (2015-08-29)[2020-02-28]. <http://arxiv.org/abs/1412.6575>
- [76] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction // *Proceedings of the 33rd International Conference on Machine Learning (ICML)*. New York City, NY, USA, 2016: 2071
- [77] Liu H X, Wu Y X, Yang Y M. Analogical inference for multi-relational embeddings // *Proceedings of the 34th International Conference on Machine Learning*. Sydney, 2017: 2168
- [78] Balazevic I, Allen C, Hospedales T M. TuckER: Tensor factorization for knowledge graph completion // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Hong Kong, 2019: 5184
- [79] Kristiadi A, Khan M A, Lukovnikov D, et al. Incorporating literals into knowledge graph embeddings[J/OL]. *arXiv preprint* (2019-07-18)[2020-02-28]. <https://arxiv.org/abs/1802.00934>
- [80] Zhang W, Paudel B, Zhang W, et al. Interaction embeddings for prediction and explanation in knowledge graphs // *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM)*. Melbourne, 2019: 96
- [81] Socher R, Chen D Q, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion // *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*. Nevada, 2013: 926
- [82] Neelakantan A, Roth B, McCallum A. Compositional vector space models for knowledge base inference[J/OL]. *arXiv preprint* (2015-05-27)[2020-02-28]. <https://arxiv.org/abs/1504.06662>
- [83] Das R, Neelakantan A, Belanger D, et al. Chains of reasoning over entities, relations, and text using recurrent neural networks // *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1*. Valencia, 2017: 132
- [84] Graves A, Wayne G, Reynolds M, et al. Hybrid computing using a neural network with dynamic external memory. *Nature*, 2016, 538(7626): 471
- [85] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings // *Thirty-Second AAAI Conference on Artificial Intelligence*. New Orleans, 2018: 1811
- [86] Vashishth S, Sanyal S, Nitin V, et al. InteractE: improving convolution-based knowledge graph embeddings by increasing feature interactions// *Proceedings of the AAAI Conference on Artificial Intelligence*, New York, 2020: 3009
- [87] Guo S, Ding B Y, Wang Q, et al. Knowledge base completion via rule-enhanced relational learning // *China Conference on Knowledge Graph and Semantic Computing*. Beijing, 2016: 219
- [88] Lu K, Zhang S, Stone P, et al. Robot representation and reasoning with knowledge from reinforcement learning[J/OL]. *arXiv preprint* (2018-11-22)[2020-02-28]. <http://arxiv.org/pdf/1809.11074.pdf>
- [89] Xie R B, Liu Z Y, Jia J, et al. Representation learning of

- knowledge graphs with entity descriptions // *Thirtieth AAAI Conference on Artificial Intelligence*. Arizona, 2016: 2659
- [90] Wang H. ReNN: Rule-embedded neural networks // *24th International Conference on Pattern Recognition*. Beijing, 2018: 824
- [91] Zhang W, Paudel B, Wang L, et al. Iteratively learning embeddings and rules for knowledge graph reasoning // *The World Wide Web Conference*. San Francisco, 2019: 2366
- [92] Nie B L, Sun S Q. Knowledge graph embedding via reasoning over entities, relations, and text. *Future Generat Comput Syst*, 2019, 91: 426
- [93] Guan N N, Song D D, Liao L J. Knowledge graph embedding with concepts. *Knowl Based Syst*, 2019, 164: 38
- [94] Mendes P N, Mühleisen H, Bizer C. Sieve: linked data quality assessment and fusion // *Proceedings of the 2012 Joint EDBT/ICDT Workshops*. Berlin, 2012: 116
- [95] Fader A, Soderland S, Etzioni O. Identifying relations for open information extraction // *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Edinburgh, 2011: 1535
- [96] Dong X, Gabrilovich E, Heitz G, et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion // *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, 2014: 601
- [97] Tan C H, Agichtein E, Ipeirotis P, et al. Trust, but verify: predicting contribution quality for knowledge base construction and curation // *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*. New York, 2014: 553
- [98] Wang J, Wang Z Y, Zhang D W, et al. Combining knowledge with deep convolutional neural networks for short text classification // *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne, 2017: 2915
- [99] Logan I R L, Liu N F, Peters M E, et al. Barack's wife hillary: Using knowledge-graphs for fact-aware language modeling // *The 57th Annual Meeting of the Association for Computational Linguistics*. Florence, 2019: 5962
- [100] Bauer L, Wang Y C, Bansal M. Commonsense for generative multi-hop question answering tasks // *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, 2018: 4220
- [101] Zhu Z K, Zhang P J, Jia Y H, et al. Chinese knowledge base question answering based on multi-label strategy. *Comput Eng*, <https://doi.org/10.19678/j.issn.1000-3428.0056763>
(朱宗奎, 张鹏举, 贾永辉, 等. 基于多标签策略的中文知识图谱问答研究. 计算机工程, <https://doi.org/10.19678/j.issn.1000-3428.0056763>)
- [102] Wang X, He X N, Cao Y X, et al. KGAT: Knowledge graph attention network for recommendation // *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage, 2019: 950
- [103] Ji S X, Pan S R, Cambria E, et al. A survey on knowledge graphs: representation, acquisition and applications[J/OL]. *arXiv preprint* (2020-02-02)[2020-02-28]. <http://arxiv.org/abs/2002.00388>
- [104] Ji Y C. Welcome to Magi[EB/OL]. *Peak Labs* (2019-08-20)[2020-02-28]. <https://www.peak-labs.com/docs/en/magi/intro>