软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2022,33(12):4644−4667 [doi: 10.13328/j.cnki.jos.006522]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

知识图谱可解释推理研究综述*

侯中妮[1,2], 靳小龙[1,2], 陈剑赟[3], 官赛萍[1], 王元卓[1], 程学旗[1,2]

[1](中国科学院网络数据科学与技术重点实验室 (中国科学院 计算技术研究所), 北京 100190)

[2](中国科学院大学 计算机科学与技术学院, 北京 100049)

[3](北京市信息技术研究所, 北京 100094)

通信作者: 靳小龙, E-mail: jinxiaolong@ict.ac.cn

摘　要: 面向知识图谱的知识推理旨在通过已有的知识图谱事实, 去推断新的事实, 进而实现知识库的补全. 近年来, 尽管基于分布式表示学习的方法在推理任务上取得了巨大的成功, 但是他们的黑盒属性使得模型无法为预测出的事实做出解释. 所以, 如何设计用户可理解、可信赖的推理模型成为了人们关注的问题. 从可解释性的基本概念出发, 系统梳理了面向知识图谱的可解释知识推理的相关工作, 具体介绍了事前可解释推理模型和事后可解释推理模型的研究进展; 根据可解释范围的大小, 将事前可解释推理模型进一步细分为全局可解释的推理和局部可解释的推理; 在事后解释模型中, 回顾了推理模型的代表方法, 并详细介绍提供事后解释的两类解释方法. 此外, 还总结了可解释知识推理在医疗、金融领域的应用. 随后, 对可解释知识推理的现状进行概述, 最后展望了可解释知识推理的未来发展方向, 以期进一步推动可解释推理的发展和应用.

关键词: 可解释性; 知识推理; 知识图谱; 事后可解释; 事前可解释

中图法分类号: TP18

中文引用格式: 侯中妮, 靳小龙, 陈剑赟, 官赛萍, 王元卓, 程学旗. 知识图谱可解释推理研究综述. 软件学报, 2022, 33(12): 4644–4667. http://www.jos.org.cn/1000-9825/6522.htm

英文引用格式: Hou ZN, Jin XL, Chen JY, Guan CP, Wang YZ, Cheng XQ. Survey of Interpretable Reasoning on Knowledge Graphs. Ruan Jian Xue Bao/Journal of Software, 2022, 33(12): 4644–4667 (in Chinese). http://www.jos.org.cn/1000-9825/6522.htm

Survey of Interpretable Reasoning on Knowledge Graphs

HOU Zhong-Ni[1,2], JIN Xiao-Long[1,2], CHEN Jian-Yun[3], GUAN Sai-Ping[1], WANG Yuan-Zhuo[1], CHENG Xue-Qi[1,2]

[1](CAS Key Laboratory of Network Data Science & Technology (Institute of Computing Technology, Chinese Academy of Sciences), Beijing 100190, China)

[2](School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China)

[3](Beijing Institute of Information Technology, Beijing 100094, China)

Abstract: Reasoning over knowledge graphs aims to infer new facts based on known ones, so as to make the graphs as complete as possible. In recent years, distributed embedding-based reasoning methods have made great success on this task. However, due to their black-box nature, these methods cannot provide interpretability for a specific prediction. Therefore, there has been a growing interest in how to design user-understandable and user-trustworthy reasoning models. Starting from the basic concept of interpretability, this work systematically studies the recently developed methods for interpretable reasoning on knowledge graphs. Specifically, it introduces the research progress of ante-hoc and post-hoc interpretable reasoning models. According to the scope of interpretability, ante-hoc interpretable models can be further divided into local-interpretable and global-interpretable models. In post-hoc interpretable reasoning models, this study reviews representative reasoning methods and introduces two post-hoc interpretation methods in detail. Next, it also summarizes the application of explainable knowledge reasoning in such fields as finance and healthcare. Then, this study summarizes the current situation

* 基金项目: 国家自然科学基金 (61772501, 62002341, U1911401, U1836206); 国家重点研发计划 (2018YFC0825205)
收稿时间: 2021-03-08; 修改时间: 2021-08-05, 2021-09-22; 采用时间: 2021-10-20; jos 在线出版时间: 2021-12-24

in explainable knowledge learning. Finally, the future technological development of interpretable reasoning models is prospected.
Key words: interpretability; knowledge reasoning; knowledge graph; post-hoc interpretability; ante-hoc interpretability

知识图谱 (knowledge graph) 本质是一种语义网络, 通常用 (头实体, 关系, 尾实体)/ (h,r,t) 这样的三元组来表达事物属性以及事物之间的语义关系. 自谷歌提出知识图谱概念以来, 知识图谱已经为智能问答、对话生成、个性化推荐等多个 NLP 任务领域提供了有力支撑. 虽然目前的知识图谱中存在大量的实体和事实数据, 但是这样大规模的数据仍然不完整, 大量缺失的三元组严重限制了这些下游任务的性能. 知识推理, 这一旨在根据一定的推理机制去预测图谱中缺失三元组的任务, 也吸引了学术界越来越多的目光.

早在 2013 年, Li 等人[1]提出利用表示学习的方法去做知识推理, 通过将实体和关系映射到低维连续的向量空间, 将推理预测任务转化为实体与关系所关联的简单的向量/矩阵操作. 鉴于该方法的自由度高、可计算性好、推理效率高等优点, 该类方法在近几年得到了广泛关注和发展, 并且广泛地应用在推荐系统、对话生成等互联网场景. 在这些场景下, 研究者们更多的关注如何提高知识推理的性能, 忽略知识推理发生错误时的风险问题. 即便推理模型在这些场景下产生错误推理时, 通常来说, 并不会招致非常严重的后果. 然而, 在当今人工智能技术应用的大趋势下, 知识推理不仅可以应用在上述互联网场景, 而且越来越多地被应用在和人类的生产生活息息相关的一些领域 (例如, 智能医疗[2−4]、军事[5]、金融[6,7]、交通运输[8,9]), 这些领域往往对模型的安全性能要求较高, 风险高度敏感. 例如, 在医疗领域, 推理的可靠性会关系到人的生命安全. 通常来说, 在这些领域, 仅仅获得预测结果是不够的, 模型还必须解释是怎么获得这个预测的, 来建立用户和推理模型之间的信任.

随着深度学习的发展, 知识推理方法的模型结构越来越复杂, 仅仅一个网络就可能包含几百个神经元、百万个参数. 尽管这些推理模型在速度、稳定性、可移植性、准确性等诸多方面优于人类, 但由于用户无法对这类模型里的参数、结构、特征产生直观理解, 对于模型的决策过程和模型的推理依据知之甚少, 对于模型的决策过程知之甚少, 不知道它何时会出现错误, 在风险敏感的领域中, 用户仍然无法信任模型的预测结果. 因此, 为了建立用户和推理模型之间的信任, 平衡模型准确率和可解释性之间的矛盾, 可解释性知识推理在近几年的科研会议上成为关注热点.

尽管有很多学者对知识推理领域进行了深入的研究, 并从不同的角度 (如分布式表示角度[10]、图神经网络角度[11]、神经-符号角度[12]等) 对推理模型进行梳理和总结. 然而, 在推理模型的可解释性方面却缺少深入的对比和总结. 为了促进可解释知识推理的研究与发展, 本文对现有的可解释推理模型进行了系统梳理、总结和展望. 本文首先阐述可解释性的定义和可解释性在推理任务中的必要性, 并介绍常见的可解释模型划分标准; 然后, 根据解释产生的方式, 对现有的可解释知识推理模型进行总结和归类, 并讨论相关方法的局限性; 接着, 简单介绍可解释知识推理在金融领域和医疗领域的应用. 最后, 本文讨论可解释知识推理面临的挑战以及可能的研究方向.

1 可解释的知识推理

在详细介绍现有的可解释知识推理模型之前, 首先介绍知识推理的基本概念, 接着对什么是可解释性 (interpretability), 以及为什么要在推理任务中注重可解释性进行介绍, 最后对本文的划分标准做简要说明.

1.1 知识推理的基本概念

2012 年, 谷歌正式提出知识图谱的概念, 用于改善自身的搜索质量. 知识图谱通常用 (h,r,t) 这样的三元组表达实体及其实体之间的语义关系, 其中, h 代表头实体, r 代表实体之间的关系, t 代表尾实体. 例如 (詹姆斯·卡梅隆, 执导, 泰坦尼克号) 即是一个三元组, 其中头实体和尾实体分别为“詹姆斯·卡梅隆”和“泰坦尼克号”, “执导”是两个实体之间的关系. 代表性的知识图谱, 如 DBpedia[13]、Freebase[14]、Wikidata[15]、YAGO[16]等, 虽然包含数以亿计的三元组, 但是却面临非常严重的数据缺失问题. 据 2014 年的统计, 在 Freebase 知识库中, 有 75% 的人没有国籍信息, DBpedia 中 60% 的人缺少没有出生地信息[17]. 知识图谱的不完整性严重制约了知识图谱在下游任务中的效能发挥. 因此, 如何让机器自动基于知识图谱中的已有知识进行推理, 从而补全和完善知识图谱, 成为了工业界和学术界都亟待解决的问题.

总的来说, 面向知识图谱的知识推理实质上是指利用机器学习或深度学习的方法, 根据知识图谱中已有的三元组去推理出缺失的三元组, 从而对知识图谱进行补充和完善. 例如, 已知 (詹姆斯·卡梅隆, 执导, 泰坦尼克号) 和 (莱昂纳多·迪卡普里奥, 出演, 泰坦尼克号), 可以得到 (詹姆斯·卡梅隆, 合作, 莱昂纳多·迪卡普里奥). 知识推理主要包含知识图谱去噪[18]和知识图谱补全 (又称之为链接预测)[19−21]两个任务[22], 其中, 知识图谱去噪任务专注于知识图谱内部已有三元组正确性的判断; 而知识图谱补全专注于扩充现有的图谱. 根据要推理元素的不同, 知识图谱补全任务可以进一步细分为实体预测和关系预测. 其中, 实体预测是指给定查询 $(h,r,?)$, 利用已有事实的关系, 推理出另一个实体并由此构成完整三元组, 同理, 关系预测则是指给定查询 $(h,?,t)$, 推理给定的头尾实体之间的关系. 由于知识图谱中大多数三元组都是正确的, 知识图谱去噪任务通常采用对已有三元组进行联合建模并进一步判断特定三元组是否成立的方法. 在这种情况下, 知识图谱补全任务可以转化为知识图谱去噪任务[23,24]. 为此, 在下面的内容里, 本文以知识图谱补全任务为中心, 对相关的可解释性方法进行梳理和总结.

1.2 可解释性及其在知识推理中的必要性

目前学术界和工业界对于可解释性没有明确的数学定义[25], 不同的研究者解决问题的角度不同, 为可解释性赋予的涵义也不同, 所提出的可解释性方法也各有侧重. Miller 等人[26]和 Buhrmester 等人[27]提出可解释性是人们能够理解决策原因的程度. 如果一个模型比另一个模型的决策过程更简单、明了、易于理解, 那么它就比另一个模型具有更高的可解释性. 目前这种定义被广泛接受.

在某些情况下, 我们不必关心模型为什么做出这样的预测, 因为它们是在低风险的环境中使用的, 这意味着错误不会造成严重后果 (例如, 电影推荐系统), 但是对于某些问题或任务, 仅仅获得预测结果是不够的. 该模型还必须解释是怎么获得这个预测的, 因为正确的预测只部分地解决了原始问题. 通常来说, 以下 3 点原因推动了对可解释性的需求.

1) 高可靠性要求. 尽管可解释性对于一些系统来说并不是不可或缺的, 但是, 对于某些需要高度可靠的预测系统来说很重要, 因为错误可能会导致灾难性的结果 (例如, 人的生命、重大的经济损失). 可解释性可以使潜在的错误更容易被检测到, 避免严重的后果. 此外, 它可以帮助工程师查明根本原因并相应地提供修复. 可解释性不会使模型更可靠或其性能更好, 但它是构建高度可靠系统的重要组成部分.

2) 道德和法律要求. 第一个要求是检测算法歧视. 由于机器学习技术的性质, 经过训练的深度神经网络可能会继承训练集中的偏差, 这有时很难被注意到. 在我们的日常生活中使用 DNN 时存在公平性问题, 例如抵押资格、信用和保险风险评估. 人们要求算法能够解释作出特定预测或判断的原因, 希望模型的解释能够使“算法歧视”的受害者诉诸人权. 此外, 推理模型目前也被用于新药的发现和设计[23]. 在药物设计领域, 除了临床测试结果以外, 新药还需要通常还需要支持结果的生物学机制, 需要具备可解释性才能获得监管机构的批准, 例如国家药品监督管理局 (NMPA).

3) 科学发现的要求. 推理模型本身应该成为知识的来源, 可解释性使提取模型捕获的这些额外知识成为可能. 当深度网络达到比旧模型更好的性能时, 它们一定发现了一些未知的“知识”. 可解释性是揭示这些知识的一种方式.

1.3 本文的划分标准

根据不同的划分标准, 知识推理模型可以被划分成不同的类别. 其中, 根据解释产生的方法, 可以将推理模型划分为两大类: 事前可解释和事后可解释[28−33]. 其中, 事前可解释模型主要指不需要额外的解释方法, 解释蕴含在自身架构之中的模型. 事后可解释性是指模型训练后运用解释方法进行推理过程和推理结果的解释, 解释方法自身是不包含在模型里面的. 一种方法被看作能够对黑盒模型进行解释, 是指该方法可以: (1) 通过可解释和透明的模型 (例如, 浅决策树、规则列表或者稀疏线性模型) 对模型的行为进行近似, 可以为模型提供全局的可解释; (2) 能够解释模型在特定输入样例上进行预测的原因; (3) 可以对模型进行内部检查, 了解模型的某些特定属性, 譬如模型敏感性或深度学习中神经元在某一特定决策中起到的作用[29]. 值得注意的是, 可以将事后解释方法应用于事前可解释的模型上, 例如, 可以从敏感性分析的角度对事前模型进行剖析. 此外, 根据可解释的范围大小——是否解释单个实例预测或整个模型行为, 可以将模型划分为局部可解释和全局可解释两大类[31,32]; 根据解释方法是否特

定于模型, 可以将模型划分为特定于模型和模型无关两种类别[32]. 在接下来的内容里, 本文按照解释产生的方式, 对知识推理模型进行总结和归类.

2 事前可解释的推理

事前解释模型是指模型本身内置可解释性, 或者将可解释的模块整合到自身架构中的模型[33,34]. 对于一个训练好的学习模型, 无需额外的信息就可以理解模型的决策过程或决策依据[32,33]. 在知识推理领域, 事前可解释模型主要围绕规则、本体以及路径等易于理解的特征展开. 下面, 本文根据可解释的范围大小, 将事前可解释的推理划分为全局可解释和局部可解释两大类.

2.1 全局可解释的推理

全局可解释试图将解释推广到尽可能广泛的输入范围, 如果一种解释可以解释整个模型, 而不仅仅是单个的实例, 那么这种解释是全局可解释的[25]. 在知识推理领域, 通常将本体和规则作为一种全局的解释. 学术界和工业界围绕本体发现和规则发现进行了一系列的研究, 具体包括基于本体的推理、基于规则的推理两大类.

2.1.1 基于本体的推理

本体定义了知识图谱的骨架, 是共享概念模型的、明确的、形式化规范说明[35], 并且本体中概念和关系是被共同认可和接受的. 实际上, 在描述逻辑的上下文中[36], 本体通常是根据一元和二元谓词来定义的. 一元谓词通常被称为概念或类别, 并且定义了某些类别, 例如具有特定特征的个人的类别. 与此相反, 二元谓词定义了一对个体之间可能存在的关系, 通常称为关系或角色. 本体真正吸引人的地方在于, 它们通常不仅定义那些谓词, 而且还提供根据它们做出预测的规则. 它包含了简单的推论, 像医生这个类别, 它的每个个体都属于人类这个类别; 同时它也包括考虑了多个概念和关系的更为详尽的推理, 例如 $\text{Carnivore} \subseteq \exists \text{eat.Meat} \cap \text{Animal}$. 基于本体的推理是演绎推理的一种, 其主要思想是利用抽象的本体层面的频繁模式、类型约束进行推理. 由于本体中的概念及层次关系被广泛认可, 作为一种一般到特殊的推理, 基于本体的推理方法可以被人们直观理解.

为了方便用户进行知识推理, 工业界和学术界纷纷推出基于本体推理的高效可扩展的推理机, 如 Racer[37]、KAON2[38]、FaCT++[39]、HermiT[40]、RDFox[41]、Pellet[42]等, 为知识库补全提供有力支撑. 由于已有的推理机大多是基于描述逻辑和基于规则的推理, 不能满足于现实世界中的不确定知识的推理, Ding 等人[43]提出将贝叶斯概率方法应用到本体推理中. 首先, 对 OWL 本体中的各个概念和属性进行概率标记; 其次, 将所有的类转换成贝叶斯网络 (BN) 中的节点, 当本体中存在父类到子类的谓词时, 在对应的节点中添加有向边, 将 OWL 本体转换为 BN 的有向无环图 (DAG); 最后, 通过为 DAG 中的每个节点构造条件概率表 (CPT) 来完成贝叶斯网络的构建, 使用通用贝叶斯网络推断程序 (例如, 信噪比传播或结点树) 来计算概念 C 与描述 e 所表示的概念之间的重叠或包含程度. Jiang 等人[18]提出利用马尔可夫逻辑网 (MLN) 对所有候选事实进行联合概率推理, 将基于描述逻辑的硬性约束转化为软性约束, 进而实现对知识库的清理、提取. 其主要思想是将来自原始信息提取系统的本体约束视为 MLN 中的硬约束, 将单个事实的置信度值视为软约束, 通过人工标记正负例事实来学习软约束的权重, 从而对信息系统提供的置信度值进行校准. 在本体约束中, 主要考虑本体约束包括包含关系、互斥关系以及关系的头尾实体类型约束. 在推理过程中, 模型将查询原子视为网络的中心, 把它们的近邻添加到一起以启用联合推理来实现加速推理. Pujara 等人[44]进一步提出基于概率软逻辑 (PSL) 框架[45], 通过提取置信度和本体信息来共同推理知识图谱. PSL 框架与 MLN 使用 0 或 1 的二值逻辑不同, 它表示的逻辑关系是用概率的形式在区间 [0, 1] 中使用软真值, 可以捕获真实世界知识中固有的不确定性和不完备性. 考虑到在知识库构建阶段, 需要手工输入和编辑大量的数据信息, 费时费力, 并且出错率较高, 难以实现大规模的本体构建. 因此, 需要对现有的本体语言进行扩展. 为此, Bühmann 等人[46]提出半自动的模式构建方法来提高本体的构建效率, 并基于模式对知识库进行补全. 首先, 对知识库进行统计分析, 发现频繁的原子模式; 然后在具体的知识图谱上, 查询原子模式和相关数据, 得到与该原子相关的候选实例; 最后, 计算候选的置信度得分, 将置信度大于特定阈值的候选原子模式作为规则进行知识图谱补全.

总的来说, 基于本体推理的方法, 作为演绎推理的一种, 能够模拟人类的逻辑推理能力, 具有非常高的可解释性, 但是它仅支持预定义的本体公理上的推理, 且高度依赖于本体频繁模式、本体约束的准确性. 对规模庞大的知识库来说, 一次性构建完善、准确的本体模式是不可能的, 这会给推理结果造成很大的影响. 此外, 由于本体约束大多为一种抽象层面的约束, 在推理的过程中, 需要对概念进行实例化, 对于实例数量很大的知识图谱来说, 实例化的过程将会产生昂贵的计算代价[22]. 因此, 本体推理难以扩展到实例数量很大的知识图谱.

2.1.2 基于规则的推理

决策规则遵循 IF-THEN 的一般结构, 由条件和预测组成, 如果条件满足, 那么可以做出一个特定的预测. 这种 IF-THEN 的结构在语义上类似于自然语言和人类的思维方式, 例如: “IF A 是 B 的父亲 AND A 是 C 的配偶 (条件), THEN C 是 B 的母亲 (预测)”. 在知识图谱中, 将单个的条件称之为原子, 可以利用单个或有限数量的原子组合进行预测. 由于知识图谱存在丰富的语义, 基于规则进行知识推理所用到的规则一般也比较复杂, 如传递性规则, 自反性规则、复合性规则等, 人为获取规则的代价比较高, 一般采用机器学习或深度学习通过算法自动挖掘决策规则. 按照使用方法的不同, 该类方法可以划分为基于搜索的传统规则推理、基于分布式表示的神经规则推理和基于深度网络的神经规则推理 3 大类.

(1) 基于搜索的传统规则推理

基于搜索的传统规则推理主要围绕规则搜索和规则修剪两个过程展开[12]. 本节首先详细介绍基于搜索的经典算法 AMIE[47], 然后简要介绍基于 AMIE 方法的一些改进.

AMIE 算法支持从不完备的知识库中, 对闭式规则进行挖掘. AMIE 维护一个规则队列, 初始化为空, 迭代地从队列中取一个规则, 如果取出来的是封闭规则且没有被丢弃, 通过添加悬挂边、添加实例边、添加闭合边 3 种操作对该规则进行扩展. 其中, 悬挂边是指边的一端是一个未出现过的变量, 而另一端 (变量或常量) 是在规则中出现过的; 实例边的一端是规则中出现过的常量或变量, 但是另一端则是未出现过的常量; 而闭合边的两端均在规则中出现过. 将扩展后的规则集合进行评估, 若其置信度大于阈值, 则将扩展后的符合要求规则加入规则队列. 如此循环, 直到没有新的规则被加入. 在规则剪枝方面, AMIE 利用支持度、置信度、针对部分完整性假设的 PCA 置信度对规则质量进行评估, 丢弃小于阈值的规则. 在此基础上, Galárraga 等人[48]提出了 AMIE+, 对 AMIE 进行优化, 通过修改规则扩展过程和规则修剪过程中定义的指标, 可以进行更有效的搜索. 在规则扩展过程中, AMIE+不会在最后一步添加悬挂边, 因为这将引入一个新的变量, 从而会产生非封闭规则. 相反, 实例边和封闭边将在最后一步添加以封闭规则. AMIE 算法在规则的修剪过程中, 计算规则的头覆盖率和置信度需要统计规则得出的事实数, 如果规则体包含具有很多变量的原子, 那推导过程将十分昂贵, 为此, AMIE+提出基于置信度近似的方法来进行规则评估及剪枝, 使得规则可以在分钟级别挖掘完毕. 此外, 为了改进 AMIE 规则置信度的计算过程, Meilicke 等人[49]提出了新的规则挖掘算法 RuleN, 能够基于随机选择的样本集合对置信度进行估算. 具体来说, 对于给定的目标关系 r, RuleN 仅采样 k (=样本大小) 个三元组 (a,r,b), 使用深度优先搜索来确定 a 和 b 之间的长度为 n 的所有可能路径, 并将这些路径作为可能的规则体. 为了挖掘包含常量且长度为 1 的规则, RuleN 为知识图谱中的每一个三元组, 创建两种规则 $r(x,y) \leftarrow r(x,y)$ 和 $r(x,b) \leftarrow r(x,y)$, 为了避免规则实例化所带来的昂贵计算问题, RuleN 对满足规则体的路径进行随机采样, 基于采样到的路径, 进一步判断该路径中的头尾实体是否存在目标关系, 从而完成对置信度的估算. 这种方式简化了置信度的计算过程, 从而可以挖掘到更长的规则.

总体上来看, 基于规则推理的方法通常采用规则挖掘算法进行自动挖掘, 虽然避开了人工获取规则的高代价, 但是, 由于知识图谱本身的不完备性, 置信度得分等统计指标可能会对规则质量进行误判, 不可避免的引入噪音规则, 对推理结果产生影响, 如何设计科学合理的规则评估指标对规则质量进行量化评价, 进而引导规则的剪枝和规律将是未来的一个研究方向. 其次, 目前规则挖掘模型大多只能学习链状规则, 表达能力也十分有限, 有待挖掘更多样、更有效的复杂规则.

(2) 基于分布式表示的神经规则推理

一方面, 基于分布式表示模型 (该类方法本文将在第 3 节进行详细讨论) 通常具有很高的可扩展性, 更能抵抗数据中的干扰. 但是, 另一方面, 它们的预测仅以一定的概率是正确的, 可解释性很差. 与此相反, 基于规则的推理

准确性高, 可解释性强, 但是却面临严重的可扩展性的问题. 基于规则和分布式的混合推理希望可以结合规则和表示学习各自的优点, 充分利用规则的准确性高、可解释性强以及表示学习的训练速度快、可扩展性好的特点来改善推理. 基于规则与分布式的混合推理整体可以分为利用分布式表示辅助规则发现以及利用规则辅助分布式表示两个类别, 但是由于利用规则辅助分布式表示的模型, 如 KALE[50], RUGE[51], RPJE[52] 等, 其本质是为了学习更好的分布式表示, 其模型整体依然不具备可解释性, 为此, 本文不做过多介绍.

考虑到规则学习算法普遍面临计算效率低下的问题, Omran 等人[53] 进一步提出基于嵌入表示的高效可扩展的规则挖掘模型 RLvLR, RLvLR 只针对与查询谓词相关的子图进行采样, 在保存必要信息的前提下减少了算法的搜索空间和计算量. 首先, RLvLR 以目标谓词为中心, 将周围路径长度小于 N 的实体及关系组织成子图. 接着, 基于提取出来的子图, 通过搜索与谓词相关的合理路径进行规则搜索, 利用嵌入模型 RESCAL 为规则中实体、谓词、论元生成嵌入表示, 并基于这些嵌入表示, 利用语义相似和谓词论元共现的思想设计评分函数来指导修剪搜索规则, 得到候选规则集合; 最后, 通过有效的邻接矩阵乘法计算头覆盖率等指标以进行最终评估. 由于知识图谱本身的不完备性, 置信度得分等统计指标可能会误判规则质量, Ho 等人[54] 针对候选规则的排名和修剪问题, 提出一种在外部资源的指导下进行规则学习的方法 RuLES, 该方法允许从不完整的知识图谱中学习高质量的规则. 具体做法是, 利用在原始图谱中得到的规则生成三元组; 同时利用实体的外部文本信息以获得实体关系的嵌入, 利用嵌入技术对规则生成的三元组进行打分, 将事实 $r(h,t)$ 的置信度得分计算为 h 和 t 之间的点积. 然后, RuLES 将学习到的规则的外部质量定义为该规则得出的所有事实的平均置信度得分. 通过在不完备图谱中的规则质量和嵌入反馈的规则质量进行综合评估, 以更精确地判断学习到的规则的质量. 近期, Zhang 等人[55] 提出了一种基于 OWL2 的联合学习框架 IterE, 可以充分建模自反、对称、传递、等价、包含、互逆以及复合 9 种关系. 如图 1 所示, 该模型分为 3 个部分: 嵌入学习、公理池生成以及公理预测及注入. 嵌入学习部分主要从正样例、负样例、规则生成的三元组数据进行学习, 通过分布式模型对实体和关系进行学习. 在得到关系的嵌入表示之后, 进入公理池生成阶段, 该阶段负责生成可能存在的规则. 为了获得规则的初始库, IterE 提出了一种与 AMIE 类似的修剪策略, 但是将遍历和随机选择的操作结合起来以平衡潜在规则的搜索过程和高度可能的规则的收敛性. 之后, 基于实体和关系嵌入从规则中导出新的三元组, 然后基于扩展三元组集. 模型 3 个部分迭代进行, 在学习嵌入表示的同时, 可以学习到更好的规则.

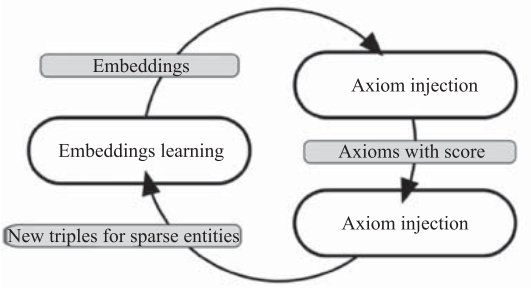

图 1 IterE 模型图[55]

(3) 基于深度网络的神经规则推理

由于传统的规则挖掘方法依靠硬匹配和离散逻辑进行规则搜索, 这对于模棱两可和具有噪音的数据不兼容, 如何利用机器学习、神经网络辅助规则发现的方法在最近几年引起了极大的关注. 提取出来的规则可以作为具体的推理结果的依据. 基于深度网络辅助规则发现的推理模型可以进一步细分为基于矩阵连乘和基于递归匹配的神经-规则挖掘两个大类.

(a) 基于矩阵连乘的神经-规则推理

基于矩阵连乘的规则提取将为每一个关系定义一个矩阵, 将多跳推理形式化为矩阵相乘, 通过 LSTM/RNN 等深度网络来模拟规则推理. 这种可微的矩阵运算, 允许模型在学习规则参数的同时, 对规则结构进行学习. 一方面,

基于神经网络的基本框架可以处理复杂的推理任务, 另一方面, 由于每一层节点都实现了推理的一部分功能, 学习后的权重具有明确的意义, 模型可以从中提取大量推理规则为预测结果作出解释.

Yang 等人[56]基于 TensorLog[57]的思想提出完全可微的系统 Neural LP, 可以同时对一阶逻辑规则的参数和结构进行学习. 首先为每个实体关联一个 one-hot 向量, 为每个关系定义一个{0, 1}操作矩阵, 如果第 i 个实体和第 j 个实体存在对应的关系, (i, j) 位置上的值为 1, 否则为 0, 通过关系矩阵相乘来模拟知识推理和规则提取. 由于传统 TensorLog 的得分函数中, 每一个置信度关联特定的规则, 规则本身的离散特性导致模型很难利用连续可微的方法去学习规则的结构. 为此, Neural LP 对得分函数进行修改, 交换求和和求积的顺序. 虽然修改之后的得分函数可以建模规则置信度和规则结构, 但却将所有的规则长度限制为固定值. 为此, Yang 等人引入辅助记忆向量、操作注意向量、存储注意向量, 以一种循环的形式对规则进行学习, 并利用 RNN 对模型参数进行求解, 并最终通过存储注意向量和操作注意向量进行规则复原. 考虑到 Neural LP 在年龄、体重或者科学测量等数字特征的处理上存在限制, Wang 等人[58]对 Neural LP 进行扩展, 提出 Neural-Num-LP 模型. 与传统一阶规则不同, 数值规则中包含变量间或者变量和数值常量之间的数值比较. Neural-Num-LP 将这些与实体相连的数值视为实体特征, 将数值关系视为依赖于这些特征的函数, 定义比较运算符和否定运算符, 并通过使用动态规划和累积和等方式, 允许模型在保证扩展效率的同时, 可以学习到更为丰富的规则. 同样, Sadeghian 等人[59]提出一种新的端到端的规则学习方法 DRUM. 为了允许模型拥有学习变长规则的能力, DRUM 引入邻接矩阵为单位阵的特殊的关系 B0, 允许该关系出现在规则的任何位置, 并允许其出现任意次; 通过巧妙的数学公式变换, 降低了参数的数量; 同时, 引入双向 RNN 来建模规则头和规则体中关系的关联. 考虑到这些模型大多只能学习一阶逻辑谓词的线性组合范式, 表达能力也十分有限, 而且生成的关系路径是基于特定的查询, 这意味着学习到的规则仅对当前查询有效, 这使得学习知识库中全局一致的规则变得困难, 泛化性能也并不理想. 为此, Yang 等人[60]对上述多跳推理框架进行扩展, 提出高效的基于神经网络的归纳逻辑学习模型 NLIL, 它首先将逻辑谓词转换为一种谓词操作, 进而将所有的中间变量转化为首尾实体的谓词操作表示, 而这样的首尾变量在具体实现时可用随机初始化的向量表示, 这样就摆脱了数据依赖; 随后这样一个个谓词操作组成了逻辑范式的原始表达单位, 这样就极大地拓展了逻辑谓词的表达能力, 从只能表达链式的逻辑规则拓展到树形, 以及规则之间的合取模式.

(b) 基于递归匹配的神经-规则推理

基于递归匹配的神经-规则挖掘主要借鉴 Prolog[61]的思想, 是一种面向演绎推理的逻辑型程序设计语言, 它主要包括匹配和回溯两种操作. 首先寻找能与目标谓词匹配合一的事实或规则头部, 若成功匹配, 则原目标的求解就转化成对规则体的求解. 在程序运行期间, 当一个子目标 (规则体中的原子) 不能满足时, 程序就返回到前一个已经满足的子目标 (如果存在), 并撤销有关变量的值约束, 然后再使其重新满足. 成功后, 再继续满足原来的子目标. 如果失败的子目标前再无子目标, 则程序返回到该子目标的上一级目标 (即该原子所在规则的头部) 使它重新匹配. 与 Prolog 不同之处在于, 它使用嵌入代替 Prolog 提供的严格统一. 因此, 对于任何给定的预测, 他们还可以以证明路径的形式提供解释.

Rocktäschel 等人[62]提出基于 Prolog 编程语言的反向链式算法的端到端的推理模型 NTPs. NTPs 依赖于 3 个模块: 合一模块 (unification module) 将 Prolog 中原子间离散匹配操作替换成了计算其嵌入相似度的可微运算符; 互相递归的 or 和 and 模块, 递归地遮盖掉 KB 中的事实并试图使用其他的事实变量和规则证明出它们, 联合枚举了所有可能的证明路径, 在最后的聚合之前, 选择得分最高的一个来证明给定知识库上的一个查询. NTPs 通过将预测错误反向传播到规则表示, 可以从数据中学习到可解释的规则. 此外, NTPs 中的证明过程是可解释的——得分最大的证明路径表示在推理过程中使用了哪些规则和事实. 然而, 由于计算复杂度的原因, NTPs 只能成功应用于涉及的数据集较小的任务, 可扩展性较差. 而且, 大多数人类知识在 KB 中是不可得的, 但是在自然语言文本上直接进行自动的推理任务又是很困难的. 基于上述原因, Minervini 等人[63]基于 NTPs 模型进行扩展, 提出了 Greedy NTPs 模型, 有效解决了原始 NTP 算法的计算复杂性高和可扩展性差的问题, 使得模型可以在大规模图谱上进行有效推理. Greedy NTPs 采用动态构建 NTPs 的计算图的方式降低, 降低模型的时间和空间复杂度, 在计算图的生成过程中, 只包含推理过程中最有希望的证明路径, 从而得到更有效的模型. 此外, 通过将逻辑事实和自然

语言句子嵌入到共享的向量空间, 实现在 KBs 和文本上的联合推理.

基于分布式表示辅助规则发现和基于深度网络辅助规则发现的方法, 将内置可解释的模型 (机制) 和黑盒模型集成到一起, 从而充分利用两者的优势, 保留或改善黑盒模型的预测性能, 同时对数据提供可解释的预测[34]. 但是由于黑盒模型的引入, 模型依然具有一定的不可解释性.

2.2 局部可解释的推理

与全局可解释不同, 局部可解释更加关注单条样本或一组样本, 通常利用目标输入处的信息 (例如, 与输入相关的案例信息、路径信息、周围子图信息等) 为模型预测作出解释[30]. 在知识推理领域, 局部可解释的推理主要包含基于随机游走的推理、基于案例的推理、基于注意力机制的推理和基于强化学习的推理四大类, 分别将路径权重、相关案例、注意力权重和推理路径作为样本预测结果的解释.

2.2.1 基于随机游走的推理

基于随机游走的方法主要借鉴 PRA (path ranking algorithm) 的思想, 将路径作为特征进行预测. 路径特征是可解释的, 这意味着人们可以很容易理解他们的含义. 如果一个路径特征比另一个具有更高的分值, 则意味着它对模型的预测具有更高的影响. 基于随机游走的方法首先确定目标关系, 判断两个实体之间是否存在该关系, 如果存在, 将其加入到正例集合, 由于知识图谱中只有正例没有负例, 通过对头尾实体进行随机替换构造负例集合. 将两个实体之间的一条路径作为一个特征, 任意枚举两个实体之间的长度不超过特定阈值的所有可能路径, 将这些可能路径放入特征集合, 根据随机游走的思想计算路径的特征值, 进而构成每个样本的特征向量; 通过利用这些正负例样本的特征向量训练 logistic 回归分类器.

基本的 PRA 算法主要通过随机选取路径来建模知识推理任务, 考虑到知识库的巨大模型和数以亿计的路径数量, 这类方法存在一定的限制. 为此, Lao 等人[64]进一步提出基于受限和加权的随机游走模型, 以数据驱动的方式来寻找对推理有帮助的路径, 来缓解枚举对于大规模知识图谱的不适应性. 该方法修改了传统 PRA 中的路径生成方式, 认为路径中的节点应该至少出现在一定比例的训练数据中, 短路径相比长路径通常更有助于推理, 来减少路径的数量. 由于知识图谱存在大量的缺失三元组, 当实体之间没有路径关联时, PRA 失效. 针对低连通图无法抽取实体间有效路径的问题, Lao 等人[65]将 PRA 扩展为使用来自文本解析的句法信息对知识库进行推理. 由于文本语料库中谓词数量极大, 直接向知识图谱中添加这类谓词代价过高, 为了克服路径特征和数据稀疏性的爆炸式增长, Gardner 等人[66]对 PRA 进一步扩展, 除了在图谱中使用传统的词法标签之外, 提出通过将词法标签映射到潜在的嵌入空间, 通过使用潜在嵌入对知识图谱进行增广, 在此基础上去执行 PRA 算法. 在特征值的计算上, 考虑到知识库的巨大模型和数以亿计的路径数量, 通过随机游走的方法计算概率的方式代价昂贵, Gardner 等人[67]提出一种更简单的子图特征提取技术 SEF, 利用路径出现/不出现的二值特征来重建 PRA 所使用的特征空间. 同时, Gardner 等人还利用路径二元特征 (path bigram features)、单边特征 (one-sided features)、向量空间相似性特征 (vector space similarity features) 和任意关系特征 (any-relation features) 等特征进行尝试. 然而, 这些方法为每一个关系预测任务单独设置分类模型, 忽略了关系之间丰富的语义联系. 对于非频繁关系而言, 训练数据的缺失会严重影响模型性能. 为此, Wang 等人[68]提出多任务学习框架 CPRA, 基于公共路径的相似度度量方法将高度相关的关系聚集在一起, 形成不同的簇, 同簇的关系共同学习; 利用共享参数和私有参数分别建模相似关系之间的共同性和关系本身的特性, 实现联合学习. 由于在知识图谱中无目的的纯随机游走来挖掘有价值的推理路径的方式效率较低, 甚至会带来一定的噪音. Wei 等人[69]进一步提出目标引导的规则挖掘算法, 具体地, 为了达到目标引导的机制, 在每一步随机游走的过程中, 算法根据最终目标动态的估计各个邻居的潜在可能性, 为具有更高可能性的邻居分配更高的概率, 使得算法更倾向于访问这些高概率邻居.

基于路径特征的方法, 有很好的可解释性, 而且可以从数据中自动发现关联规则, 准确性往往也可以满足一定的要求, 但是该方法很难处理关系稀疏的数据, 当知识图谱稀疏、低联通时, 对路径特征的提取效率低下且耗时.

2.2.2 基于案例的推理

基于案例的推理 (case-based reasoning) 源于人类的认知心理活动, 通过寻找与当前问题相似的历史案例来进

行推理. 一个典型的案例推理问题求解过程的基本步骤可以归纳为 4 个主要步骤[70]: 案例检索 (retrieve)、案例重用 (reuse)、案例修正 (revise) 以及案例保存 (retain). 在案例推理中, 通常把待解决的问题或工况称为目标案例 (target case), 把历史案例称为源案例 (base case), 源案例的集合称为案例库. 图 2 给出了案例推理解决的问题的基本流程[71]: 一个待解决的新问题出现, 这个就是目标案例; 利用目标案例的描述信息查询过去相似的案例, 即对案例库进行检索, 得到与目标案例相类似的源案例, 由此获得对新问题的一些解决方案; 如果这个解答方案失败将对其进行调整, 最后, 如果修改后的解决方案可用于解决给定的问题, 则将他们保留在内存中, 以便将来使用. 案例推理模型简单, 不需要训练, 符合人类的认知心理活动, 具有良好的可解释性和可扩展性.

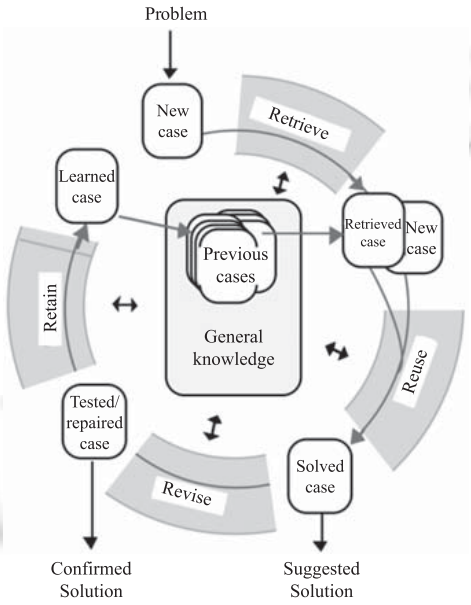

图 2 案例推理流程图[71]

Das 等人[72]基于案例推理的思想, 提出一种简单的非参数的推理方法 CBR, 来进行知识库补全. 在给定一个实体 e_q 和查询关系 r_q 的情况下, CBR 首先在整个知识图谱中检索与查询实体类似且存在 r_q 的实体, 这里, 相似度通过预先计算好的相似度矩阵来计算; 接下来, 对于检索到的实体, 找到与其通过关系 r_q 连接的尾实体, 在这些实体对中挖掘路径并加入路径列表; 将列表中的路径所包含的实体剔除, 并按照出现频率降序排列; 最后的查询目标实体 e_q 周围是否存在类似的推理路径, 如果存在相似的路径并可以到达正确的尾实体, 将该路径保存到记忆单元中. 考虑到即使是出现频率最高的路径, 也不一定适用于所有的查询, 譬如对于出生在美国的科学家, “born_in”可以给出正确的“place_of_death”; 但是对于移民到美国的科学家来说, 该路径却会失效. Das 等人[73]进一步提出基于 KNN 概率的案例推理方法. 主要方法是为每个实体关联一个 m-hot 向量, 如果实体 e_i 有 m 个不同的输出边类型, 则将与这些类型相对应的维设置为 1, 通过计算实体向量之间的 cosine 距离来计算实体之间的相似性, 得到与查询实体类似且存在 r_q 的 k 个实体; 接下来, 对于检索到的实体, 找到与其通过关系 r_q 连接的尾实体, 在这些实体对中挖掘路径并加入路径列表; 通过在给定当前查询的情况下得出正确答案的可能性来权衡每条路径; 在可能性的估计上, 通过将相似的实体聚集成簇, 并估计通过简单的计数统计信息来测量每个簇中路径的先验和精度. 此外, 该模型还扩展了对于未见实体的推理, 使之可以很好地处理未见实体的推理任务.

基于案例的推理本质是一种类比推理方式, 它可以很好的利用案例中隐藏的难以规则化的知识, 来辅助规则推理的不足. 相较于规则获取, 案例的获取通常更为容易, 可以很好地解决规则获取的瓶颈问题. 但是, 案例推理的质量高度依赖于知识库自身的案例数目和质量, 当知识库中案例越多, 覆盖面越广, 越有利于推理质量的提高. 其次, 在相似案例的检索上, 如何找到一个好的相似度度量也会对模型效果产生影响.

2.2.3 基于注意力机制的推理

注意力机制在深度学习的各个领域被广泛使用. 它借鉴了人类的注意力机制, 允许从大量输入信息中选择小部分有用信息来重点处理, 同时忽略其他的可见信息, 可以通过展示特征重要性为模型的预测结果进行解释, 是一种内置的可解释机制[33,61]. 通过注意力机制和神经网络模型进行结合, 可以有效改善神经网络模型自身可解释性差的问题.

在现实生活中, 一个概念可以被多个关系共享, 一个关系可能是一组概念的组合集合, 譬如, 关系“(某人) 因 (某项工作) 获奖”和“(某人) 因 (某项工作) 获得提名”都分别描述了一个人获得奖项或提名的高质量作品这一概念, 而关系 has_part_of, 可能是指地理上的包含概念, 也可能是物品之间的包含概念. 基于嵌入表示的方法, 如 TransR[21] , TransD[19] 等, 由于对投影矩阵没有限制, 概念上相关的关系可能有完全不同的投影空间, 从而阻碍统计规律的发现、共享和推广. 同时, 嵌入表示模型在稀疏关系上存在一定的限制. 针对这两个问题, Xie 等人[74] 提出可解释的知识转移模型 ITransF, 解决数据稀疏问题并鼓励在投影矩阵中共享统计规律. 该模型通过稀疏注意力机制, 学习将共享的概念矩阵组合成特定关系的投影矩阵, 从而获得更好的泛化特性.

近年来, 基于注意力机制的神经网络推理已成为知识推理的一大热点. 这类模型的通常做法是, 首先基于目前所处的位置对不同的实体/关系设置不同的权重, 使得模型能够忽略查询无关信息而关注重点信息, 最后通过可视化注意力矩阵, 来提供对预测结果的解释, 以增强预测自身的可解释性.

Feng 等人[75] 提出了图感知的知识表示方法 GAKE, 基于知识图谱的结构信息来学习实体和关系的向量表示. GAKE 引入了 3 类图上下文信息: 邻居上下文、路径上下文和边上下文, 从不同角度反映知识属性, 同时设计注意力机制, 即实体和关系的权重学习, 学习有代表能力的实体或关系. 其中, 邻居上下文反映三元组关系, 实体的邻居上下文为以该实体为头实体的所有三元组中的关系和尾实体对, 关系的邻居上下文为该关系对应的所有三元组中的头实体和尾实体对. 路径上下文为多步路径上的实体和关系, 实体的边上下文为与该实体相连的所有关系, 关系的边上下文为该关系连接的所有实体. 每类上下文的目标函数为给定上下文, 实体/关系的概率函数和. GAKE 最大化 3 类目标函数的加权和. Nathani 等人[76] 通过关注任意给定实体邻域内的实体和关系特征, 将注意力机制与图神经网络进行结合, 以此作为编码器学习更好的实体和关系向量表示. 具体的, 通过多头注意力机制 (multi-head attention architecture) 捕获不同的一跳邻居信息, 经由类似的多个图注意力层, 来捕获不同实体在具体查询任务中所扮演的角色多样性. Wang 等人[77]认为有效的邻居聚合器需要满足无序性、有冗余意识以及关注重点关系 3 个特性. 具体地, 排列无关指在聚合时, 邻居节点的排列顺序不应对最终编码造成影响; 冗余意识是指各类信息之间存在信息的冗余, 例如“play_for”包含“work_as”的信息, 聚合器应当具有识别这种冗余的机制; 关注重点关系是指聚合器应当着重关注与当前查询相关的关系. 在此基础上, Wang 等人[77] 进一步改进了传播模型, 提出利用规则与注意力机制共同计算邻居权重的逻辑注意力网络 LAN. LAN 考虑邻居的冗余性和查询关系, 以排列不变的方式将不同的权重赋予实体的邻居, 根据在粗关系级别具有逻辑关系的数据和神经注意网络在精细的邻居级别对邻居权重进行估计. Bansal 等人[78] 提出了 A2N 的方法, 这是一种带有近邻注意力的知识图谱嵌入技术. 作者在评估中证明, 从近邻中获取信息可以更好地表示多重关系 (multi-hop relation). 近期, Teru 等人[79] 提出基于图神经网络的关系预测框架 GraIL, 来解决知识图谱上进行归纳式关系预测的问题. GraIL 模型分为 3 个步骤, 首先进行子图挖掘, 对于两个目标节点, 采样出在两个节点之间路径长度最大为 K+1 的所有路径构成子图; 接着, 通过度量子图中每个点和目标节点的距离对节点进行标签初始化, 例如对于目标节点为 u 、 v 的子图, 其中的一点 i , 用一个元组 ($d(i,u)$, $d(i,v)$) 表示, 其中 $d(\cdot,\cdot)$ 表示两点最短距离. 特别的, u 、 v 两点分别以 (0, 1)、(1, 0) 进行表示. 最后, 借鉴 R-GCN[80] 的方法来建模对多关系图的消息传递, 与 R-GCN 的不同之处在于, GraIL 增加了一个新的注意力机制, 该注意力机制不仅仅建模两个相邻节点以及它们之间的关系, 同时也对被预测的目标关系进行建模. 最终利用基于整个图的表示, 计算在给定实体对之间某种关系成立的概率, 得分最高的关系被视为被预测关系. Zhang 等人[81] 进一步提出了一种基于层次注意力的关系图神经网络 RGHAT, 第 1 层是关系级别的注意力, 其灵感来自不同关系对某一实体的指示权重不同的直觉; 第 2 层是实体级别关注, 使得模型能够突出同一关系下不同相邻实体的重

要性. 这种分层的注意力机制, 为模型提供了细粒度的学习过程, 从而提高了模型的可解释性. 为了减小传统 GCN 的复杂性, Xu 等人[82]提出用于顺序推理的神经网络架构 DPMPN, DPMPN 网络包含两个图神经网络, 其中一个执行输入不变的全局信息传递; 另一个在依赖于输入的子图上进行信息传递, 其中, 依赖于输入的子图并非一次构建, 而是采用流式注意力机制, 动态且选择性的扩展.

不可否认, 在利用图神经网络解决知识推理任务时, 注意力机制发挥了巨大的作用, 并可以在一定程度上对 GNN 进行解释, 但是节点的 1-hop 邻居和 2-hop 邻居可能存在重叠, 注意力机制会对同一节点学习到不同的权重, 如何去做出正确的抉择是一个需要解决的问题. 其次, 注意力机制的计算通常会利用简单的神经网络进行计算, 譬如前向神经网络, 这使得注意力模型会增加模型整体的计算复杂度.

2.2.4 基于强化学习的推理

强化学习是一种有效解决序列决策问题的方法, 基于强化学习的推理旨在通过有限步的探索来寻找与当前查询相关的可靠推理路径[83]. 这类方法将路径寻找问题形式化为马尔可夫决策过程, 首先从主题实体出发, 根据问题选择一个关系, 从而跳转到一个新的实体; 基于跳转到的新实体, 继续选择关系, 迭代若干步, 直到达到最大步数或到达正确答案实体. 在此过程中, 基于策略的智能体不断根据当前状态选择最有希望的关系进行状态转换, 一旦找到关系路径, 则通过奖励函数更新策略网络. 通过反复的试验、探索, 智能体可以被用来查找与问题相关的可解释的推理路径, 为当前预测结果做出解释. 虽然强化学习的方法可以显式地给出推理路径, 但是其智能体的策略网络通常是基于 LSTM/CNN 等深度网络来实现的, 模型依然具有一定程度的不可解释性.

Xiong 等人[83]于 2017 年提出 DeepPath 模型, 首次将强化学习方法引入到知识推理任务中, 将知识推理视为部分 Markov 过程. 与传统 PRA 模型不同, DeepPath 利用嵌入表示模型来编码 RL 智能体的状态, 并通过全连接神经网络对关系进行采样来实现推理路径的顺序扩展. 此外, 为了鼓励代理探索更多的有效路径, DeepPath 设计了综合准确性、路径多样性以及路径有效性的打分函数. 鉴于 DeepPath 没有对路径信息编码, 造成历史探索的遗忘, 同时其智能体状态包含答案实体, 无法直接应用于查询问答任务, Das 等人[84]进一步提出新的推理模型 MINERVA, 其模型图如图 3 所示. MINERVA 将 LSTM 作为路径记忆组件, 对路径信息进行编码, 将路径的隐层表示、查询问题与当前节点表示进行拼接, 通过线性层和 ReLU 得到隐层表示; 同时对当前状态的候选动作进行编码, 与当前状态的隐层表示一起进行候选动作决策.

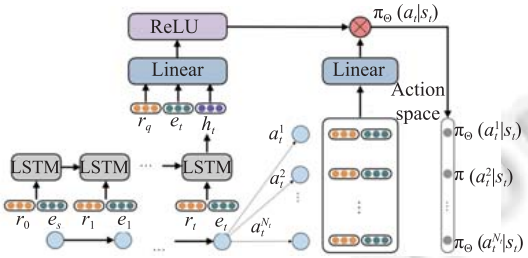

图 3 MINERVA 模型图[84]

考虑到知识图谱搜索空间随路径长度呈指数增长以及参数随机初始化, 基于强化学习的方法普遍面临奖励稀疏性的问题, 同时, 智能体可能会被偶然到达正确答案的虚假搜索路径所欺骗, 如果当前策略错误地为正确的路径分配了低概率, 则智能体很可能在探索期间无法发现这条路径, 因此将无法增加正确路径的概率. 这在每个梯度步骤上都会重复, 从而使正确路径永远保持较低的可能性. 相同的反馈回路也可能导致本已很高概率的伪路径获得更大的概率. 由此可见, 这会造成一种富者越富, 穷者越穷的状态. 为此, Lin 等人[70]在 MINERVA 的基础上, 进一步提出 MultiHopKG 模型以解决上述两个问题. 其中, 通过利用预训练的表示学习模型, 如 ConvE[85]等, 为未观察到的三元组提供软奖励, 从而减少稀疏奖励的影响; 利用动作随机丢弃机制去随机掩盖部分关系, 解决伪路径的问题. Wang 等人[86]提出基于注意力机制的强化推理模型 AttnPath, 将 LSTM 和图注意力机制联合起来共同作为记忆组件, 使得状态的表示更加丰富. 同时, 为了避免智能体在同一实体上陷入停滞, AttnPath 设计了一种新的学习

机制, 使得智能体每一步都能够前进. 近期, 针对强化学习模型依赖于对手工设计的奖励函数以及大量试验 (trial-and-error) 导致的收敛性差两个问题, Li 等人[87] 引入模仿学习的概念, 提出基于生成的对抗模仿学习的插件 DIVINE, 用于增强现有的基于强化学习的方法. DIVINE 通过使用生成式对抗训练, 从专家事例中训练由生成器和判别器组成的推理机, 其中生成器可以是现有基于强化学习的方法中的任何基于策略的代理器, 而判别器可以是作为一种自适应奖励函数, 生成器要尽量生成和专家事例相符的路径, 判别器进行路径质量的判别, 并返回给生成器对应的奖励. 对于不同的数据集, 判别器可以自动调整奖励函数以实现最佳性能, 从而消除传统奖励函数的人工干预. Hildebrandt 等人[88] 提出基于动态辩论的知识图谱推理模型 R2D2, 主要思想是将三元组分类的任务转化成两个强化学习智能体之间的辩论过程. 其中两个智能体被视为稀疏的对抗特征生成器, 分别从知识图谱中提取论点 (知识图谱中的路径), 促进事实为真命题或者否命题. 基于这些论点, 法官 (前向神经网络) 对当前三元组进行判断. 与其他黑盒方法相反, 这些提取的论点可以使用户了解法官的决定, 从而可以被用户很好地理解.

基于强化学习的推理可以有效避免 PRA 路径空间过大的问题, 但是存在以下问题: 首先, 基于强化学习的推理通常将是否到达正确尾实体作为依据来设计奖励函数, 这种二值奖励信号比较稀疏, 导致智能体学习缓慢甚至无法学习到最优策略. 如何设计合理的奖励函数去指导智能体的学习过程, 是一个非常大的挑战. 其次, 大多数强化推理模型采用从头开始学习的方法, 这种方式会使模型面临非常严重地冷启动问题, 如何提供强力且有效的干预来解决冷启动的问题, 加速模型的训练过程将会是未来的一个可能的研究方向.

本文对以上事前可解释的方法做了简单总结, 并且比较了它们的优点和缺点, 具体如表 1 所示.

表 1　事前可解释知识推理模型对比

类别			解释形式	方法优缺点
全局可解释的推理	基于本体的推理		事先定义好的本体库, 如 $Mother(x), Mother \subseteq Women \rightarrow Women(x)$	优点: 可以利用更为抽象化的本体层面的频繁模式、约束或路径进行推理 缺点: 仅支持预定义的本体公理上的推理, 且高度依赖于本体约束的准确性
	基于规则的推理	基于搜索的传统规则推理		优点: 模型准确性高、透明度高、可解释性强 缺点: 对于具有噪音的数据不兼容, 可扩展性比较差.
		基于分布式表示的神经规则推理	基于规则解释整个模型, 如 $bornin \wedge cityof \rightarrow nationalistyof$	优点: 充分利用规则的准确性高、可解释性强以及表示学习的训练速度快、可扩展性好的特点 缺点: 引入黑盒模型, 存在一定的不可解释性
		基于深度网络的神经规则推理		优点: 可以解决传统规则对噪音数据不兼容的问题, 可扩展性强 缺点: 引入黑盒模型, 存在一定的不可解释性
局部可解释的推理	基于随机游走的推理		给定查询, 通过计算路径特征的权重作为预测结果的解释, 权重高的路径对模型的预测结果有决定性作用	优点: 可以从数据中发现隐式的关联规则; 准确性往往也可以满足一定的要求 缺点: 搜索空间较大, 当知识图谱稀疏、低联通时, 对路径特征的提取效率低下且耗时
	基于案例的推理		给定查询, 通过寻找与之相似的案例, 把它重新应用到新问题的环境中来, 相似的案例便是模型预测结果的解释	优点: 可以很好地利用案例中隐藏的难以规则化的知识, 来辅助规则推理的不足 缺点: 案例推理的质量高度依赖于知识库自身的案例数目和质量; 其次, 相似度度量也会影响模型效果
	基于强化学习的推理		给定查询, 借助RL的推理路径来解释智能体的行为, 进而为模型预测结果提供解释	优点: 可以解决PRA面临的搜索空间过大问题 缺点: 奖励信号稀疏, 模型训练依赖于奖励函数设计; 黑盒模型的引入, 导致模型透明度及可解释性的下降
	基于注意力机制的推理		给定查询 $nationalityof$, 基于attention权重的高低来解释模型预测, 高权重的输入单元对模型的预测结果有决定性作用, 如 $bornin, 0.36, friendof, 0.08$	优点: 解释关系推理任务注意力机制拥有良好的可视化操作, 可以为单条样本或一组样本提供解释 缺点: 注意力值的计算依赖于新的神经网络, 增加了计算复杂度, 降低了模型的透明度

可以看出, 在推理模型中, 无论是全局可解释的知识推理还是局部可解释的推理, 都需要针对感兴趣的推理结果学习出正确的关系语义依赖. 其中, 全局可解释的推理主要围绕本体和一阶逻辑规则展开. 针对传统规则挖掘方法面临的可扩展性差、无法兼容噪音数据等问题, 学术界逐渐开始尝试将黑盒模型和规则挖掘进行融合, 利用分布式表示和深度网络在辅助规则挖掘, 从而可以更好地保留两者各自的优势, 在提供可解释性的同时提高模型的性能. 而局部可解释的推理, 关注单条样本或一组样本, 并通过寻找与查询三元组相关的案例、展示推理路径等方式来为模型预测作出解释.

3 事后可解释的推理

事后可解释代表了一种从学习的模型中提取信息的独特方法. 虽然无法准确阐明模型的工作原理, 但是对于一个给定的训练好的分布式推理模型, 通过利用解释方法或构建解释模型, 可以对推理模型的工作机制、决策行为和据测依据进行一定的解释 [32,33]. 基于分布式表示推理方法在最近几年得到了迅速发展, 尽管它们产生了良好的效果, 但是却无法为模型的预测提供解释. 因此, 研究者们通常采用规则提取和敏感性分析等事后解释方法为这类黑盒推理模型提供解释. 其中, 规则提取是指从受训模型中提取解释规则的方式, 提供对黑盒模型整体决策逻辑的理解 [33]; 而敏感性分析基于稳健统计的思想, 通过改变自变量的值来解释因变量守自变量变化影响大小的规律. 下面, 本文简单介绍事后可解释推理模型的代表性工作, 并对可以提供事后解释的解释方法做详细介绍 [31].

3.1 事后可解释推理模型

在面向知识图谱的推理领域, 事后可解释推理模型主要是指基于分布式表示的推理. 此类模型将三元组的实体和关系映射到低维连续的向量空间, 通过向量之间的运算来进行知识推理. 按照使用方法的不同, 基于分布式表示的推理可以大体分为基于距离、基于张量分解和基于神经网络的推理 3 类方法.

(1) 基于距离的推理

基于距离的推理模型基本思想是将三元组的实体和关系映射到低维连续的向量空间, 将关系看作实体间的转移; 根据转移假设设计得分函数来衡量三元组的有效性. 得分越高, 三元组成立的可能性要高.

Borders 等人于 2013 年提出了第一个基于转移的表示模型 TransE[1] , 其主要思想是: 如果三元组 (头实体, 关系, 尾实体) 成立, 头实体的向量与关系向量的和应近似等于尾实体的向量, 否则远离. 由于 TransE 严格要求有效三元组满足头实体加关系在向量空间中与尾实体足够靠近, 因此无法很好的处理 1-*N*, *N*-1 和 *N*-*N* 的关系. 因此, Wang 等人针对上述问题, 提出了 TransH[89]. TransH 依然将实体表示为向量, 但是把关系表示为关系表示向量和关系映射向量, 将实体映射到关系相关的超平面; 然后在该超平面, 将关系表示向量看作映射之后的实体向量之间的转移. TransE 模型及其扩展模型往往只考虑了实体之间的直接关系, 忽略了知识图谱中实体之间多步关系所蕴含的丰富的语义信息, 为突破这类模型孤立学习每个三元组的局限性, Lin 等人基于 TransE 模型进行扩展, 提出 PTransE 模型 [90] , 将知识图谱中的关系路径融入到知识表示模型中. 通过关系的相加、相乘等操作的建模关系路径的复合语义, 并利用资源分配算法为实体间的路径进行加权. 传统的基于转移的方法大多将概念和实例作为同等重要的实体进行编码, 忽略了实例和概念之间的不同. Lv 等人 [91] 对实例和概念进行区分, 将图谱中的概念当作球体, 实例当作向量进行编码, 如果某一实例属于某个概念的范畴, 那么该实例的嵌入应该在概念 C 的球体内部, 并最终通过定义实例三元组之间、实例与所属概念之间、概念与概念之间的损失函数进行模型优化. 此外, Zhang 等人 [92] 进一步对层级实体进行建模, 通过将实体、概念、关系映射到极坐标中, 利用节点距离原点的半径作为节点在层级树中的层级, 将角度作为同一层中不同节点, 从而可以有效区分不同层级上的实体以及同一层级的不同实体, 通过将关系视为节点半径的比例转换来进行知识推理.

(2) 基于张量分解的推理

基于张量分解的推理方法通常将整个知识图谱表示为张量, 然后通过张量分解来学习实体和关系的隐层表示. 分解得到张量通过矩阵相乘得到重构张量来近似整个知识图谱, 重构张量中的元素值的大小作为知识图谱中未知三元组成立的可能性.

Nickel 等人[93]提出 RESCAL 模型, 拉开了利用张量分解建模知识推理的帷幕. 其核心思想如图 4 所示, 首先将整个知识图谱编码为一个三维张量, 如果三元组存在, 则张量中对应位置的元素值为 1, 否则为 0. 由这个张量分解出一个关系张量和一个实体矩阵, 关系张量中每一个二维矩阵切片代表一种关系, 实体矩阵中每一行代表一个实体. 通过最小化重构误差来进行实体和关系的表示. 最后, 关系张量和实体矩阵还原的结果被看作对应三元组成立的概率, 如果概率大于某个阈值, 则对应三元组成立; 否则不成立. 虽然 RESCAL 模型推理准确率高, 但是其张量分解过程复杂, 计算速度慢. 为此, Chang 等人[94]引入实体类型信息对 RESCAL 模型进行加速. 具体地在损失函数的计算中排除不满足关系特定的实体类型约束的三元组. 例如, 关系 spouse_of 的头尾实体类型都必须是人. Yang 等人[95] 提出 DistMult 模型, 通过限制关系矩阵为对称阵的方式, 对 RESCAL 模型进行简化, 来解决 RESCAL 所面临的参数众多的问题, 但是这种过度地简化导致只能处理对称关系. 为了增强模型的表达性, Trouillon 等人[96]对 DistMult 进行扩展, 提出基于复值向量表示的矩阵分解方法. 其中, 每一个实体和关系嵌入不再存在于实空间中, 而是存在于复空间中, 三元组 (s,r,o) 的得分表示为关系 r 的向量表示、头实体 s 的向量表示以及尾实体 o 的向量表示的共轭向量的乘积, 并保留最后结果的实部. 由于复空间的埃尔米特乘积 (Hermitian dot product) 不具有交换性, 因此可以很好的建模对称关系和反对称关系. 在如何引入更简洁的张量分解方法, 同时使模型具有充分的表达性方面, Kazemi 等人[97]提出针对实体出现的位置, 为每个实体关联两个向量, 利用 CP 分解进行三元组的可靠性打分. Balažević等人[98]进一步提出具有完全表达能力的 TuckER 模型. 具体地, 利用 Tucker 分解的思想对图谱张量进行分解, 将图谱张量分解成一个核心张量和关于头尾实体以及关系的 3 个矩阵.

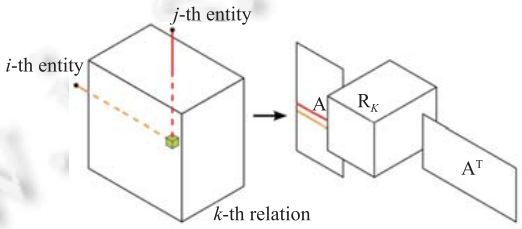

图 4 RESCAL 模型图[93]

(3) 基于神经网络的推理

考虑到神经网络所具备的强大的学习能力, 越来越多的研究者提出将神经网络用于知识推理. 这类方法主要利用 CNN、RNN 和 GNN 等深度学习模型对知识图谱进行建模, 在获得实体和关系的向量表示之后, 进行下一步的推理.

Dettmers 等人[85]提出基于 CNN 的知识推理模型 ConvE, 首先将一对头实体和关系嵌入进行拼接, 利用二维卷积将头实体和关系重塑为二维矩阵, 然后利用二维卷积进行尾实体的预测. Vashishth 等人[99]指出增加实体和关系交互的数量有助于知识推理任务, 并表明 ConvE 可以捕获的交互数量是有限的. 为此 Vashishth 等人基于 ConvE 进行扩展, 通过特征排列、特征重塑和循环卷积来进一步增加实体和关系之间的交互. 其中, 特征排列是指将实体嵌入和关系嵌入随机打乱, 生成关于实体和关系的 t 个排列; 随后, 通过堆叠、交替出现等方式将实体和关系重塑为特征位数的二维矩阵; 最后, 在每一个二维矩阵上应用循环卷积, 将循环卷积的输出进行拼接, 来进行链接预测. Schlichtkrull 等人[80]提出 R-GCN 模型, 开创了使用 GCN 框架去建模关系网络的先河. 与传统图卷积神经网络不同之处在于, R-GCN 提出了关系特定的转换, 对关系的类型和方向进行建模, 每层节点特征都是由上一层节点特征和节点的关系 (边) 得到; 通过对节点的邻居节点特征和自身特征进行加权求和得到新的特征. 将 R-GCN 学习到的实体关系表示输入到 DistMult 模型中进行解码进行链接预测. Shang 等人[100]提出加权图卷积网络 (WGCN), 给不同类型的关系设置不同的权重, 将知识图谱看成多个带有不同强弱的单关系的子图, 每个子图共享相同的卷积操作, 但在信息聚合时, 引入可学习的关系特定的权重, 对各个子图卷积的结果进行加权.

在 RNN 与知识推理结合方面, Das 等人[101]针对关系预测任务, 利用 RNN 对关系、实体、实体类型进行联合建模, 来预测实体之间的可能关系. 针对实体预测任务, Guo 等人[102]提出 RSN 模型, 将 RNN 与残差学习集成

在一起, 来捕获知识图谱中的长距离的依存关系. 具体地, 首先在知识图谱中进行随机游走, 获取 RNN 的训练路径集, 之后利用 RNN 对路径进行建模. 如图 5 与普通 RNN 不同之处在于, RSN 明确区分实体和关系, 对实体使用跳跃连接, 允许实体直接参与预测对象实体.

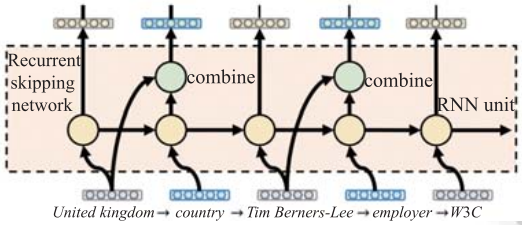

图 5 RSNs 处理 2 跳关系路径的示例[102]

3.2 事后可解释推理模型的解释方法

3.2.1 规则/贝叶斯网络提取

规则/贝叶斯网络提取的基本思想是通过从复杂的黑盒模型中提取解释规则/贝叶斯网络, 在保持保真度的同时, 提供对人类可解释的描述性表示[33]. 现有工作中主要有 3 种技术可以从复杂模型中提取知识: 教学法, 分解法和折中法. 在教学法中, 训好的神经网络模型被看作黑盒, 不利用其模型结构和参数信息, 只利用模型的输入和输出, 提取将输入直接映射到输出的规则[103]; 分解法是指剥离复杂模型并分析每个部分, 例如在深度学习中, 模型组件及包括神经元, 也包括权重; 折衷方法是教学法和分解法的结合.

Sánchez 等人[104] 基于教学法的思想, 针对张量分解模型分别对基于一阶规则提取和贝叶斯网络 (BN) 进行探索. 首先使用矩阵分解模型来预测矩阵所有事实的可能性, 并将相关与之设置为 0.5, 以产生一组要学习的真实三元组和错误三元组, 从而构成训练集 DMF. 接着从 DMF 中挖掘形式为 $b(x,y)\Rightarrow h(x,y)$ 的一阶规则, 其中规则头的变量和规则体的变量要一致. 通过计算规则头和规则体之间的 PMI (点互信息), 进行规则筛选. 此外, Sánchez 等人还进行从张量分解模型中提取贝叶斯网络的尝试, 将 BN 的结构约束为一棵树. 在这种情况下, 学习过程简化为针对由 MF 模型生成的训练集 DMF 中的变量 (即关系) 之间的互信息找到最大生成树, Prim 算法在 $O(N^2)$ 中来进行求解解决此问题, 其中 N 是变量数. 同时, Gusmao 等人[105] 基于教学法, 提出 XKE-PRED 方法从训练好的嵌入表示模型中进行规则提取. 由于嵌入表示模型的输入都是由三元组组成, 这些三元组没有固有的可解释的特征, 为此, Gusmao 等人通过使用从子图结构提取的特征和原始分类器预测的标签, 重新训练逻辑回归模型, 进而得出加权的 Horn 子句的解释. 此外, Yang 等人[95] 提出利用规则提取的方法对 TransE 和张量分解模型进行解释. 具体地, 针对真实三元组 (e_1,r_1,e_2)、(e_2,r_2,e_3)、(e_1,r_3,e_3), 可以通过 $r_1+r_2\approx r_3$ 提取规则 $r_1(e_1,e_2)\wedge r_2(e_2,e_3)\Rightarrow r_3(e_1,e_3)$. 针对张量分解模型, 可以通过 $r_1r_2\approx r_3$ 提取规则 $r_1(e_1,e_2)\wedge r_2(e_2,e_3)\Rightarrow r_3(e_1,e_3)$. Carmona 等人[106] 针对张量分解模型, 分别进行提取决策规则、提取决策树和提取贝叶斯网络 3 组实验, 利用保真度和可解释性对 3 种方法进行评估, 并指出基于逻辑规则的方法和基于决策树的方法由于其自身的确定性, 只存在匹配和不匹配两种情况, 因而产生的三元组的排名和嵌入表示相比有一定差距; 而基于贝叶斯网络由于自身可以建模不确定知识, 可以在提供预测的多步解释的同时, 拥有很高的保真度. 这些规则提取的方法不需要额外的数据源, 并且避免了通过将可解释性纳入目标函数而引入的潜在偏差. 但是, 大多数的规则提取方法所提取到的规则往往不够精确, 因而只能提供近似解释, 不一定能反映推理模型的真实行为. 为此, Nandwani 等人[107] 提出 OXKBC 事后解释方法, 可以基于张量分解的 KBC 模型的预测结果进行解释. 首先根据 KBC 模型计算出的实体张量, 计算实体之间的相似性, 在相似实体之间引入加权边来对知识图谱进行增广; 接着根据路径中的关系与查询关系的相似度以及路径中边的权重, OXKBC 定义解释有效性的打分函数. 考虑到不同的路径中涵盖的边类型不同, 路径得分会有较大差异, OXKBC 进一步为相似路径定义二阶模版, 通过 MLP 选择最有可能性的模版来对预测结果进行解释.

3.2.2 敏感度分析

基于敏感性分析的方法, 其本质是基于稳健统计的, 主要思想是在给定一个查询的情况下, 识别有影响力的三

元组, 可以更好地解释模型的行为和预测. 当某一个三元组从现有的图谱中删除后, 会大大影响模型的参数或者预测结果, 那么该三元组被视为是“有影响力的”[31].

Pezeshkpour 等人[108]提出了对一种对知识图谱进行对抗性修改的新颖模型 CRIAGE, 用来分析链接预测模型的鲁棒性和可解释性. 如图 6 所示, 在模型训练之后, CRIAGE 通过删除目标尾实体的一条邻接边或添加一条虚假边对知识图谱进行扰动. 这些添加和删除的事实, 会更改目标三元组的预测. 通过使用这些单个修改, 可以确定预测链接最有影响力的事实, 并评估模型对于添加假事实的敏感性. 由于不能在每次添加删除操作之后都对所有的三元组进行重新训练, 尤其是候选三元组数量极大的情况下, 模型利用泰勒展开, 引入一阶近似的思想来估计这种修改的效果, 通过梯度计算识别最有影响力的实例并评估模型对于添加假事实的敏感性; 通过总结每个关系中最有影响力的事实, 来对预测结果进行解释. 值得一提的是, CRIAGE 是针对图神经网络 GNNExplainer[109]引入的. GNNExplainer 提供了解图网络的预测方法的第一项工作, 而且是一种与模型无关的方法, 可为任何基于图的机器学习任务上基于 GNN 的模型的预测提供可解释的解释. 给定一个实例, GNNExplainer 会确定紧凑的子图结构和节点特征的一小部分, 这些特征对 GNN 的预测至关重要. 此外, GNNExplainer 可以为整个实例类生成一致而简洁的解释. 它可用于识别影响特定实例的预测的图神经网络的最重要部分和特征 (例如, 新链接、新节点标签).

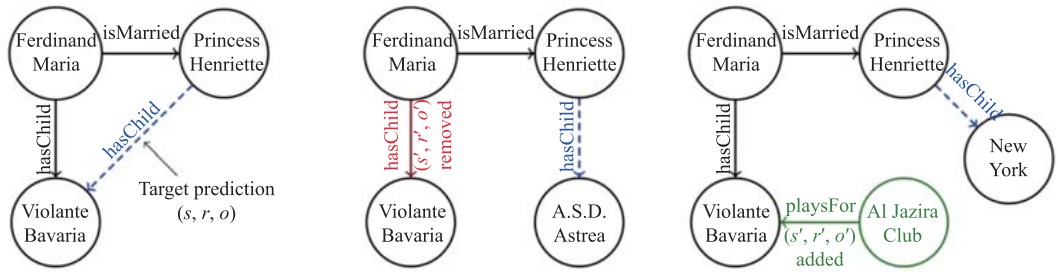

图 6 CRIAGE 对知识图谱进行扰动的两种策略[108]

基于有影响力实例的方法, 通过在特征和预测结果中建立一个因果关系为用户提供解释. 其删除添加操作是模型无关的, 这意味着该方法可以应用于任何模型. 可以通过使用该方法来比较不同的机器学习模型并更好地理解他们的不同行为. 但是删除和添加的计算非常昂贵, 而且这种影响力度量仅考虑单个实体的添加和删除, 而不是针对多个三元组. 三元组之间可能会具有一些交互, 这些交互强烈影响模型的训练和预测.

4 可解释知识推理的应用

目前, 可解释的知识推理已经在医疗领域、金融领域等诸多不同领域展示出良好的应用前景.

当应用于医疗知识图谱时, 可解释的知识推理方法可以帮助医生收集健康数据、诊断疾病和控制错误. 知识推理在医疗领域最常见的一个应用是构建临床决策支持系统. 在手动诊断中, 有时医生无法收集完整的患者病历, 可能影响诊断准确性. 在这种情况下, 临床决策支持系统可以指导医生做出更好的决策. 同样, 在得出结论之前, 医生必须仔细分析系统提供的建议. 在有效的临床决策支持系统的支持下进行的诊断可以大大提高诊断的准确性. Martínez-Romero 等人[110]构建基于本体的急性心脏病危重病人智能监护治疗系统, 专家知识由 OWL 本体和一组 SWRL 规则表示. 基于患者在特定时间的生命体征和领域知识, 该系统执行推理过程, 并向医生提供有关应该进行何种治疗以实现最快恢复的建议. García-Crespo 等人[111]基于设计了一个基于逻辑推理和概率细化的本体驱动的鉴别诊断系统 (ODDIN), 可以根据多个不同参数确定最可能的诊断. Sherimon 等人[112]针对糖尿病患者, 提出了一种基于本体推理的方法来构建临床决策支持系统, 来帮助经验不足的医疗从业人员, 进而提高对患者的护理质量. Bao 等人[3]提出应用本体技术来描述和揭示蒙医药学基础理论、疾病、症状、症候、方剂、药材等资源之间的语义关系, 进而构建知识库, 并基于领域本体进行知识发现. 柳彦平等人[2]结合 RDF 和专家系统, 设计了一种基

于 RDF 的诊疗专家系统, 可以对疾病、流感、痢疾、麻疹等进行例行诊断, 并给出一般性治疗药方. 杨丽[4]针对中医临床诊断决策问题, 结合归纳逻辑程序设计和马尔可夫逻辑网, 来实现诊断决策支持, 并基于案例推理实现处方治疗决策支持.

在金融领域和证券投资领域, 存在大量的欺诈行为. 公司为了获得更多投资, 会对财务报表进行造假 (财务报表欺诈现象). 常见的财务报表舞弊类型包括财务记录遗漏, 伪造或篡改收入、资产, 支出和其他财务变量, 以及对管理讨论和陈述的虚假陈述. 财务报表欺诈严重影响投资者和监管机构, 会对经济和股市造成巨大损失, 并破坏了公众对商业环境的信心. 为此, Tang 等人[7]提出一种基于本体, SWRL 和决策树算法的基于知识的财务报表欺诈检测系统. 首先建立了财务报表的本体, 并使用决策树算法对财务报表欺诈模式进行查找, 并将这些模式转换为可在基于知识的系统中使用的 SWRL 规则. 之后, 利用 Pellet 作为推理器进行欺诈识别. 在防控经济犯罪方面, 熊建英等人[6]结合本体论的思想, 抽取资金流中的相关实体、属性, 来构建经济犯罪侦查大数据图谱, 根据专家知识, 制定风险预警命中规则, 如交易金额、交易对象和账户数量等, 从而可以为办案人员提供参考. 强韶华等人[113]基于本体和案例推理的思想, 构建金融事件本体, 设计基于本体的 SWRL 推理规则, 综合企业财务、非财务和舆情等因素, 成功地预测金融事件对企业股价的影响.

5 挑战与展望

近年来, 可解释推理引发了广泛关注, 为此, 我们对以上几种方法做了简单总结, 并且比较了它们的优点和缺点, 具体如表 2 所示.

表 2 可解释性知识推理模型对比

类别		代表性推理模型	优点	缺点
事前可解释的推理	全局可解释	基于本体的推理基于规则的推理	这类模型内置可解释性, 具有很好的透明度, 用户可以完全理解模型的决策过程	在计算能力上, 和其他方法相比存在一定的差距
	局部可解释	基于随机游走的推理基于案例的推理基于注意力机制的推理基于强化学习的推理	将内置可解释的模型(机制)和黑盒模型集成到一起, 在保留或改善黑盒模型的预测性能的同时, 对数据提供可解释的预测	黑盒模型的引入, 导致模型不透明度的上升, 进而无法在一些对安全要求比较高的领域得到大规模应用
事后可解释的推理		基于距离的推理基于张量分解的推理基于神经网络的推理	计算能力强, 可扩展性好	需要重新设计解释方法, 但是解释方法只能提供近似解释, 会导致解释结果和模型真实行为之间存在一定程度的不一致性. 在运用解释方法的基础, 模型依然存在一定的不可解释性

总体来说, 尽管可解释知识推理的研究已经取得了一系列研究成果, 但是其研究整体还处于起步阶段, 依然面临诸多挑战. 其中, 可解释推理主要面临事前可解释模型推理性能低下, 与应用领域强绑定; 事后解释方法无法反映模型真实行为、适用范围受限; 缺乏统一的评测体系 3 大挑战. 下面, 本文针对上述挑战进行简单介绍并对未来的发展方向进行展望.

(1) 事前解释模型的推理性能问题

从应用角度来看, 目前事前可解释的推理模型由于其自身的高安全性和高可控性, 已经在对安全性要求较高的领域中得到了相对广泛的应用. 但是, 对比经典的嵌入表示模型, 这些方法无法兼容噪音数据, 在针对命中率等指标上的表现并不是很优异, 且存在推理速度慢等诸多问题. 因此, 如何设计高效的事前可解释模型来消除这类方法和经典嵌入表示模型之间的性能差异是事前可解释模型所面临的一大挑战. 此外, 从行业应用的角度看, 在医疗等领域得到广泛使用的方法大多是基于本体和基于决策规则的推理方法, 这类方法本身和领域强绑定, 通用性非常受限, 修改难度也比较大. 考虑到事前可解释模型自身的高安全性、高可控性和高自我解释性, 本文认为进一步发展事前可解释模型将在未来的发展过程中将占有非常重要的一席.

(2) 事后解释方法的设计问题

在本身不具备解释机制的推理模型中, 事后解释方法应当在提供用户可理解的解释的同时, 可以精确的反映模型的内部工作逻辑. 目前的解释方法大多采用基于规则或贝叶斯网络提取的方式, 但是这类解释方法只能提供近似的解释, 这种近似会导致解释结果和模型真实行为之间存在一定程度的不一致性[33], 进而影响用户对模型的信赖度. 也正是由于这种不一致性, 限制了这类模型在运输、军事、金融、智能医疗等领域的应用. 解释方法首先应该忠实于推理模型自身, 要具备精确解释模型内部决策的能力, 其次, 应当尽可能普适和轻量级. 目前, 现有的解释方法往往针对单一模型, 可以广泛适用于分布式表示推理框架的解释方法并不是很多, 模型无关的解释方法效果仍有待进一步提升. 如果存在一个模型无关的、精确的可解释方法, 就可以避免针对每个推理模型分别设计解释方案, 从而减少大量不必要的资源耗费[31]. 此外, 事后解释方法目前主要面向从事不同行业的从业人员, 这类人群对深度学习的了解程度相对不高, 因此, 解释方法主要扮演建立人机共同语言的角色. 除了这类人群, 解释方法还应服务于从事模型开发设计的研究人员, 对模型的各个层进行解释, 对嵌入表示的每一维度进行解释, 以此来帮助设计鲁棒性、扩展性很好的推理模型.

(3) 缺乏统一的评测体系

对于同一个任务/场景, 可以应用不同的推理方法, 除了推理任务本身的精准度之外, 如何去衡量这些方法在可解释性上的优劣是亟需解决的一个问题. 当前, 解释方法的评估没有一个明确的指标, 只能对解释方法进行定性分析, 无法对同类型的研究工作进行严格的、确定性的分析与比较[34]. 针对事前可解释的推理模型而言, 其评估挑战在于如何量化模型内在的解释能力; 对于事后可解释的推理模型而言, 首先需要明确要评估的指标, 进而从指标入手, 建立评估方法.

考虑到事前可解释推理模型的可解释性受到应用场景和终端用户等多种因素的制约, 未来可以从终端用户特征、模型算法本身以及效用评估 (用户在执行具体任务时的表现) 等多个角度来设计事前可解释推理模型的评估指标. 对事后可解释推理模型而言, 如何评估一种事后解释方法的效用, 以及如何去评估不同模型之间可解释性的强弱, 是一直是困扰可解释性研究员的一个难题. 当前, 事后解释方法的评估大多通过可视化方法或者人工评估的方法, 这很大程度上依赖于人类的认知, 因而只能定性分析, 无法对可解释模型的性能进行量化. Jacovi 等人[71]指出可解释的评估应当包含似然性和忠实性的评估, 而人的参与只会把评估变为似然性评估, 对证明模型具有真正的可解释能力——即忠实性方面毫无用处. 值得注意的是, 在忠实性的评估方面, 不应引入人对于解释质量的判断, 因为人的判断会倾向于似然性, 导致忠实性的评估产生偏差. 为此, 对于事后可解释模型来说, 首先需要明确评估指标 (忠实性还是似然性), 并针对不同的评估对象使用适当的方法, 进而建立多因素综合的评估体系[34], 综合评估解释方法的忠实性和似然性.

6 总 结

随着深度神经网络的发展, 关于模型可解释性问题吸引了越来越多的目光. 本文从可解释性的基本概念出发, 系统梳理了当前可解释知识推理的相关工作. 根据解释产生方式的不同, 本文将可解释知识推理分为事前可解释推理模型和事后解释推理模型两大类. 其中, 根据可解释范围的大小, 本文将事前可解释推理模型进一步细分为全局可解释的推理和局部可解释的推理; 在事后解释模型中, 介绍了事后解释推理的 3 类代表性推理模型并详细介绍提供解释的事后解释方法. 接着, 对可解释知识推理在其他领域的应用做了简单介绍. 最后, 本文对可解释知识推理的现状进行总结和归纳, 并展望了可解释知识推理的未来发展方向.

References:

[1] Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: Advances in Neural Information Processing Systems. South Lake Tahoe: HAL, 2013. 2787–2795.

[2] Liu YP, Wang WJ, Rong J. Medical diagnosis expert system based on RDF. Microcomputer & Its Application, 2005, 24(5): 36–39 [doi: 10.3969/j.issn.1674-7720.2005.05.012]

[3] Bao YL. Research on construction of knowledge base and knowledge discovery of traditional Mongolian medicine based on domain

ontology [Ph.D. Thesis]. Changchun: Jilin University, 2018 (in Chinese with English abstract).

[4] Yang L. Research on clinical decision support methods based on knowledge-based reasoning for traditional Chinese medicine diagnosis and treatment [MS. Thesis]. Beijing: Beijing Jiaotong University, 2014 (in Chinese with English abstract).

[5] Chen XJ, Zhang XP, Zhao XZ. Research on event ontology model and fusion method of naval battlefield situation. Military Operations Research and Systems Engineering, 2019, 33(1): 69–74 (in Chinese with English abstract). [doi: 10.3969/j.issn.1672-8211.2019.01.013]

[6] Xiong JY. Research on the application of knowledge graph in the analysis of abnormal capital flow. Financial Technology Time, 2021, 29(1): 28–33 (in Chinese with English abstract). [doi: 10.3969/j.issn.2095-0799.2021.01.005]

[7] Tang XB, Liu GC, Yang J, Wei W. Knowledge-based financial statement fraud detection system: Based on an ontology and a decision tree. Knowledge Organization, 2018, 45(3): 205–219. [doi: 10.5771/0943-7444-2018-3-205]

[8] Huang KP, Jiang CJ. Analyzing and reasoning knowledge of urban transportation: Based on ontology. Computer Science, 2007, 34(3): 192–196 (in Chinese with English abstract). [doi: 10.3969/j.issn.1002-137X.2007.03.051]

[9] Liang YD, Zhai J, Yuan CF. Establishment of logistics distribution system based on ontological reasoning. Logistics Technology, 2015, 34(9): 255–258 (in Chinese with English abstract).

[10] Wang Q, Mao ZD, Wang B, Guo L. Knowledge graph embedding: A survey of approaches and applications. IEEE Trans. on Knowledge and Data Engineering, 2017, 29(12): 2724–2743. [doi: 10.1109/TKDE.2017.2754499]

[11] Arora S. A survey on graph neural networks for knowledge graph completion. arXiv:2007.12374v1, 2020.

[12] Zhang J, Chen B, Zhang LX, Ke XR, Ding HP. Neural, symbolic and neural-symbolic reasoning on knowledge graphs. arXiv:2010.05446, 2020.

[13] Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, Hellmann S, Morsey M, van Kleef Patrick, Auer S, Bizer C. DBpedia–A large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web, 2015, 6(2): 167–195. [doi: 10.3233/SW-140134]

[14] Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J. Freebase: A collaboratively created graph database for structuring human knowledge. In Proc. of the 2008 ACM SIGMOD Int'l Conf. on Management of Data. Vancouver: ACM, 2008. 1247–1250. [doi: 10.1145/1376616.1376746]

[15] Vrandečić D, Krötzsch M. Wikidata: A free collaborative knowledgebase. Communications of the ACM, 2014, 57(10): 78–85. [doi: 10.1145/2629489]

[16] Fabian MS, Gjergji K, Gerhard W. Yago: A core of semantic knowledge unifying wordnet and wikipedia. In: Proc. of the 16th Int'l World Wide Web Conf. Banff: HAL, 2007. 697–706. [doi: 10.1145/1242572.1242667]

[17] Dong X, Gabrilovich E, Heitz G, Horn W, Lao N, Murphy K, Strohmann T, Sun SH, Zhang W. Knowledge vault: A Web-scale approach to probabilistic knowledge fusion. In: Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM, 2014. 601–610. [doi: 10.1145/2623330.2623623]

[18] Jiang SP, Lowd D, Dou DJ. Learning to refine an automatically extracted knowledge base using Markov logic. In: Proc. of the 12th IEEE Int'l Conf. on Data Mining. Brussels: IEEE, 2012. 912–917. [doi: 10.1109/ICDM.2012.156]

[19] Ji GL, He SZ, Xu LH, Liu K, Zhao J. Knowledge graph embedding via dynamic mapping matrix. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing (Vol. 1: Long Papers). Beijing: Association for Computational Linguistics, 2015. 687–696. [doi: 10.3115/v1/P15-1067]

[20] Lin XV, Socher R, Xiong CM. Multi-hop knowledge graph reasoning with reward shaping. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 3243–3253. [doi: 10.18653/v1/D18-1362]

[21] Lin YK, Liu ZY, Sun MS, Liu Y, Zhu X. Learning entity and relation embeddings for knowledge graph completion. In: Proc. of the 29th AAAI Conf. on Artificial Intelligence. Austin: AAAI, 2015. 2181–2187. [doi: 10.5555/2886521.2886624]

[22] Guan SP, Jin XL, Jia YT, Wang YZ, Cheng XQ. Knowledge reasoning over knowledge graph: A Survey on Neural Network Interpretability. Ruan Jian Xue Bao/Journal of Software, 2018, 29(10): 2966–2994 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5551.htm [doi: 10.13328/j.cnki.jos.005551]

[23] Mohamed SK, Nováček V, Nounu A. Discovering protein drug targets using knowledge graph embeddings. Bioinformatics, 2020, 36(2): 603–610. [doi: 10.1093/bioinformatics/btz600]

[24] Huynh VP, Papotti P. Buckle: Evaluating fact checking algorithms built on knowledge bases. Proc. of the VLDB Endowment, 2019, 12(12): 1798–1801. [doi: 10.14778/3352063.3352069]

[25] Lipton ZC. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.

Queue, 2018, 16(3): 31–57. [doi: 10.1145/3236386.3241340]

[26] Biran O, Cotton C. Explanation and justification in machine learning: A survey. Proc. of the IJCAI-17 Workshop on Explainable AI (XAI). 2017, 8(1): 8–13.

[27] Miller T. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 2019, 267: 1–38. [doi: 10.1016/j.artint.2018.07.007]

[28] Buhrmester V, Münch D, Arens M. Analysis of explainers of black box deep neural networks for computer vision: A survey. Machine Learning and Knowledge Extraction, 2019, 3(4): 966–989. [doi: 10.3390/make3040048]

[29] Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM Computing Surveys, 2019, 51(5): 93. [doi: 10.1145/3236009]

[30] Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable ai: A review of machine learning interpretability methods. Entropy, 2020, 23(1): 18.

[31] Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Lulu Press, 2020.

[32] Camburu OM. Explaining deep neural networks. arXiv:2010.01496, 2020.

[33] Ji SL, Li XF, Du TY, Li B. Survey on techniques, applications and security of machine learning interpretability. Journal of Computer Research and Development, 2019, 56(10): 2071–2096 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2019.20190540]

[34] Došilović FK, Brčić M, Hlupić N. Explainable artificial intelligence: A survey. In: Proc. of the 41st Int'l Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). Opatija: IEEE, 2018. 210–215. [doi: 10.23919/MIPRO.2018.8400040]

[35] Studer R, Benjamins VR, Fensel D. Knowledge engineering: Principles and methods. Data & Knowledge Engineering, 1998, 25(1–2): 161–197. [doi: 10.1016/S0169-023X(97)00056-6]

[36] Baader F, Calvanese D, McGuinness D, Nardi D, Peter-Schneider P. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge: Cambridge University Press, 2003.

[37] Haarslev V, Möller R. RACER system description. In: Goré R, Leitsch A, Nipkow T, eds. Proc. of the 1st Int'l Joint Conf. on Automated Reasoning. Siena: Springer, 2001. 701–705. [doi: 10.1007/3-540-45744-5_59]

[38] Motik B, Studer R. KAON2–a scalable reasoning tool for the semantic Web. In: Proc. of the 2nd European Semantic Web Conf. (ESWC2005). Heraklion. 2005, 17.

[39] Tsarkov D, Horrocks I. FaCT++ description logic reasoner: System description. In: Proc. of the 3rd Int'l Joint Conf. on Automated Reasoning. Seattle: Springer, 2006. 292–297. [doi: 10.1007%2F11814771_26]

[40] Glimm B, Horrocks I, Motik B, Stoilos G, Wang Z. HermiT: An OWL 2 reasoner. Journal of Automated Reasoning, 2014, 53(3): 245–269. [doi: 10.1007/s10817-014-9305-1]

[41] Nenov Y, Piro R, Motik B, Horrocks I, Wu Z, Banerjee J. RDFox: A highly-scalable RDF store. In: Proc. of the 14th Int'l Semantic Web Conf. Bethlehem: Springer, 2015. 3–20. [doi: 10.1007%2F978-3-319-25010-6_1]

[42] Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y. Pellet: A practical OWL-DL reasoner. Journal of Web Semantics, 2007, 5(2): 51–53. [doi: 10.1016/j.websem.2007.03.004]

[43] Ding ZL, Peng Y. A probabilistic extension to ontology language OWL. In: Proc. of the 37th Annual Hawaii Int'l Conf. on System Sciences. Big Island: IEEE, 2004. 10. [doi: 10.1109/HICSS.2004.1265290]

[44] Pujara J, Miao H, Getoor L, Cohen W. Large-scale knowledge graph identification using PSL. In: Proc. of the AAAI Symp. on Semantics for Big Data. Virginia: AAAI, 2013.

[45] Brocheler M, Mihalkova L, Getoor L. Probabilistic similarity logic. arXiv:1203.3469, 2012.

[46] Bühmann L, Lehmann J. Pattern based knowledge base enrichment. In: Proc. of the 12th Int'l Semantic Web Conf. Sydney: Springer, 2013. 33–48. [doi: 10.1007/978-3-642-41335-3_3]

[47] Galárraga LA, Teflioudi C, Hose K, Suchanek F. AMIE: Association rule mining under incomplete evidence in ontological knowledge bases. In: Proc. of the 22nd Int'l Conf. on World Wide Web. Riode Janeiro: ACM, 2013. 413–422. [doi: 10.1145/2488388.2488425]

[48] Galárraga L, Teflioudi C, Hose K, Suchanek FM. Fast rule mining in ontological knowledge bases with AMIE+. The VLDB Journal, 2015, 24(6): 707–730. [doi: 10.1007/s00778-015-0394-1]

[49] Meilicke C, Fink M, Wang YJ, Ruffinelli D, Gemulla R, Stuckenschmidt H. Fine-grained evaluation of rule- and embedding-based systems for knowledge graph completion. In: Proc. of the 17th Int'l Semantic Web Conf. Monterey: Springer, 2018. 3–20. [doi: 10.1007/978-3-030-00671-6_1]

[50] Guo S, Wang Q, Wang LH, Guo L. Jointly embedding knowledge graphs and logical rules. In: Proc. of the 2016 Conf. on Empirical

Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 192–202. [doi: 10.18653/v1/D16-1019]

[51] Guo S, Wang Q, Wang LH, Wang B, Guo L. Knowledge graph embedding with iterative guidance from soft rules. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. New Orleans: AAAI, 2018. 4816–4823.

[52] Niu GL, Zhang YF, Li B, Cui P, Liu S, Li JY, Zhang XW. Rule-guided compositional representation learning on knowledge graphs. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 2950–2958. [doi: 10.1609/aaai.v34i03.5687]

[53] Omran PG, Wang KW, Wang Z. Scalable rule learning via learning representation. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence. Stockholm: IJCAI, 2018. 2149–2155. [doi: 10.24963/ijcai.2018/297]

[54] Ho VT, Stepanova D, Gad-Elrab MH, Kharlamov E, Weikum G. Rule learning from knowledge graphs guided by embedding models. In: Proc. of the 17th Int'l Semantic Web Conf. Monterey: Springer, 2018. 72–90. [doi: 10.1007%2F978-3-030-00671-6_5]

[55] Zhang W, Paudel B, Wang L, Chen J Y, Zhu H, Zhang W, Bernstein A, Chen H J. Iteratively learning embeddings and rules for knowledge graph reasoning. In: Proc. of the 2019 World Wide Web Conf. San Francisco: ACM, 2019. 2366–2377. [doi: 10.1145/3308558.3313612]

[56] Yang F, Yang ZL, Cohen WW. Differentiable learning of logical rules for knowledge base reasoning. In: Proc. of the 31st Conf. on Neural Information Processing Systems. Long Beach, 2017. 2319–2328.

[57] Cohen WW. TensorLog: A differentiable deductive database. arXiv:1605.06523, 2016.

[58] Wang PW, Stepanova D, Domokos C, Kolter JZ. Differentiable learning of numerical rules in knowledge graphs. In: Proc. of the 2019 Int'l Conf. on Learning Representations. 2019.

[59] Sadeghian A, Armandpour M, Ding P, Wang DZ. DRUM: End-to-end differentiable rule mining on knowledge graphs. In: Proc. of the 33rd Conf. on Neural Information Processing Systems. Vancouver, 2019. 15347–15357.

[60] Yang Y, Song L. Learn to explain efficiently via neural logic inductive learning. In: Proc. of the 2019 Int'l Conf. on Learning Representations. 2019.

[61] Gallaire H, Minker J, Nicolas JM. Logic and databases: A deductive approach. In: Artificial Intelligence and Databases. Morgan Kaufmann: Elsevier, 1989. 231–247. [doi: 10.1016/B978-0-934613-53-8.50020-0]

[62] Rocktäschel T, Riedel S. End-to-end differentiable proving. In: Proc. of the 31st Conf. on Neural Information Processing Systems. Long Beach, 2017. 3788–3800.

[63] Minervini P, Bošnjak M, Rocktäschel T, Riedel S, Grefenstette E. Differentiable reasoning on large knowledge bases and natural language. arXiv:1912.10824v1, 2020.

[64] Lao N, Mitchell T, Cohen WW. Random walk inference and learning in a large scale knowledge base. In: Proc. of the 2011 Conf. on Empirical Methods In Natural Language Processing. Edinburgh: ACM, 2011. 529–539. [doi: 10.5555/2145432.2145494]

[65] Lao N, Subramanya A, Pereira F, Cohen WW. Reading the web with learned syntactic-semantic inference rules. In: Proc. of the 2012 Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island: Association for Computational Linguistics, 2012. 1017–1026.

[66] Gardner M, Talukdar PP, Kisiel B, Mitchell T. Improving learning and inference in a large knowledge-base using latent syntactic cues. In: Proc. of the 2013 Conf. on Empirical Methods in Natural Language Processing. Seattle: Association for Computational Linguistics, 2013. 833–838.

[67] Gardner M, Mitchell T. Efficient and expressive knowledge base completion using subgraph feature extraction. In: Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing. Lisbon: Association for Computational Linguistics, 2015. 1488–1498. [doi: 10.18653/v1/D15-1173]

[68] Wang Q, Liu J, Luo YF, Wang B, Lin CY. Knowledge base completion via coupled path ranking. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Berlin: Association for Computational Linguistics, 2016. 1308–1318. [doi: 10.18653/v1/P16-1124]

[69] Wei ZY, Zhao J, Liu K. Mining inference formulas by goal-directed random walks. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Austin: Association for Computational Linguistics, 2016. 1379–1388. [doi: 10.18653/v1/D16-1145]

[70] Aamodt A, Plaza E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications, 1994, 7(1): 39–59. [doi: 10.3233/AIC-1994-7104]

[71] Jacovi A, Goldberg Y. Towards faithfully interpretable NLP Systems: How should we define and evaluate faithfulness? In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 4198–4205. [doi: 10.18653/v1/2020.acl-main.386]

[72] Das R, Godbole A, Dhuliawala S, Zaheer M, McCallum A. A simple approach to case-based reasoning in knowledge bases.

arXiv:2006.14198, 2020.

[73] Das R, Godbole A, Monath N, Zaheer M, McCallum A. Probabilistic case-based reasoning for open-world knowledge graph completion. In: Proc. of the 2020 Findings of the Association for Computational Linguistics (EMNLP 2020). Association for Computational Linguistics, 2020. 4752–4765. [doi: 10.18653/v1/2020.findings-emnlp.427]

[74] Xie QZ, Ma XZ, Dai ZH, Hovy E. An interpretable knowledge transfer model for knowledge base completion. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Vancouver: Association for Computational Linguistics, 2017. 950–962. [doi: 10.18653/v1/P17-1088]

[75] Feng J, Huang ML, Yang Y, Zhou XY. GAKE: Graph aware knowledge embedding. In: Proc. of the 26th Int'l Conf. on Computational Linguistics (COLING 2016): Technical Papers. Osaka: The COLING 2016 Organizing Committee, 2016. 641–651.

[76] Nathani D, Chauhan J, Sharma C, Kaul M. Learning attention-based embeddings for relation prediction in knowledge graphs. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 4710–4723.

[77] Wang PF, Han JL, Li CL, Pan R. Logic attention based neighborhood aggregation for inductive knowledge graph embedding. In: Proc. of the 33rd AAAI Conf. on Artificial Intelligence. Honolulu: AAAI, 2019. 7152–7159. [doi: 10.1609/aaai.v33i01.33017152]

[78] Bansal T, Juan DC, Ravi S, McCallum A. A2N: Attending to neighbors for knowledge graph inference. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 4387–4392. [doi: 10.18653/v1/P19-1431]

[79] Teru K, Denis E, Hamilton W. Inductive relation prediction by subgraph reasoning. In: Proc. of the 37th Int'l Conf. on Machine Learning. PMLR, 2020. 9448–9457.

[80] Schlichtkrull M, Kipf TN, Bloem P, van den Berg R, Titov I, Welling M. Modeling relational data with graph convolutional networks. In: Proc. of the 15th Int'l Conf. on the Semantic Web. Heraklion: Springer, 2018. 593–607. [doi: 10.1007/978-3-319-93417-4_38]

[81] Zhang Z, Zhuang FZ, Zhu HS, Shi ZP, Xiong H, He Q. Relational graph neural network with hierarchical attention for knowledge graph completion. In: Proc. of the 2020 AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 9612–9619. [doi: 10.1609/aaai.v34i05.6508]

[82] Xu XR, Feng W, Jiang YS, Xie XH, Sun ZQ, Deng ZH. Dynamically pruned message passing networks for large-scale knowledge graph reasoning. In: Proc. of the 8th Int'l Conf. on Learning Representations. Addis Ababa: OpenReview.net, 2019.

[83] Xiong WH, Hoang T, Wang WY. DeepPath: A reinforcement learning method for knowledge graph reasoning. In: Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 564–573. [doi: 10.18653/v1/D17-1060]

[84] Das R, Dhuliawala S, Zaheer M, Vilnis L, Durugkar I, Krishnamurthy A, Smola A. Andrew McCallum. Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases with reinforcement learning. arXiv:1711.05851, 2018.

[85] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D knowledge graph embeddings. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. New Orleans: AAAI, 2018. 1811–1818.

[86] Wang H, Li SY, Pan R, Mao MZ. Incorporating graph attention mechanism into knowledge graph reasoning based on deep reinforcement learning. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 2623–2631. [doi: 10.18653/v1/D19-1264]

[87] Li RP, Cheng X. DIVINE: A generative adversarial imitation learning framework for knowledge graph reasoning. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 2642–2651. [doi: 10.18653/v1/D19-1266]

[88] Hildebrandt M, Serna JAQ, Ma YP, Ringsquandl M, Joblin M, Tresp V. Reasoning on knowledge graphs with debate dynamics. Proc. of the 2020 AAAI Conf. on Artificial Intelligence, 2020, 34(4): 4123–4131. [doi: 10.1609/aaai.v34i04.6600]

[89] Wang Z, Zhang JW, Feng JL, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: Proc. of the 28th AAAI Conf. on Artificial Intelligence. Québec: AAAI, 2014. 1112–1119.

[90] Lin YK, Liu ZY, Luan HB, Sun MS, Rao SW, Liu S. Modeling relation paths for representation learning of knowledge bases. In: Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing. Lisbon: Association for Computational Linguistics, 2015. 705–714. [doi: 10.18653/v1/D15-1082]

[91] Lv X, Hou L, Li JZ, Liu ZY. Differentiating concepts and instances for knowledge graph embedding. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 1971–1979. [doi: 10.18653/v1/D18-1222]

[92] Zhang ZQ, Cai JY, Zhang YD, Wang J. Learning hierarchy-aware knowledge graph embeddings for link prediction. Proc. of the 2020 AAAI Conf. on Artificial Intelligence, 2020, 34(3): 3065–3072. [doi: 10.1609/aaai.v34i03.5701]

[93] Nickel M, Tresp V, Kriegel HP. A three-way model for collective learning on multi-relational data. In: Proc. of the 28th Int'l Conf. on Int'l Conf. on Machine Learning. Bellevue: Omnipress, 2011. 809–816. [doi: 10.5555/3104482.3104584]

[94] Chang KW, Yih W, Yang B, *et al*. Typed tensor decomposition of knowledge bases for relation extraction. Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP). 2014. 1568–1579.

[95] Yang BS, Yih WT, He XD, Gao JF, Deng L. Embedding entities and relations for learning and inference in knowledge bases. arXiv:1412.6575, 2014.

[96] Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G. Complex embeddings for simple link prediction. In: Proc. of the 33rd Int'l Conf. on Machine Learning. New York: PMLR, 2016. 2071–2080. [doi: 10.5555/3045390.3045609]

[97] Kazemi SM, Poole D. Simple embedding for link prediction in knowledge graphs. In: Proc. of the 32nd Int'l Conf. on Neural Information Processing Systems. Montréal, 2018. 4284–4295.

[98] Balazevic I, Allen C, Hospedales T. TuckER: Tensor factorization for knowledge graph completion. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 5188–5197. [doi: 10.18653/v1/D19-1522]

[99] Vashishth S, Sanyal S, Nitin V, Agrawal N, Talukdar P. InteractE: Improving convolution-based knowledge graph embeddings by increasing feature interactions. Proc. of the AAAI Conf. on Artificial Intelligence, 2020, 34(3): 3009–3016. [doi: 10.1609/aaai.v34i03.5694]

[100] Shang C, Tang Y, Huang J, Bi JB, He XD, Zhou BW. End-to-end structure-aware convolutional networks for knowledge base completion. Proc. of the AAAI Conf. on Artificial Intelligence, 2019, 33(1): 3060–3067. [doi: 10.1609/aaai.v33i01.33013060]

[101] Das R, Neelakantan A, Belanger D, McCallum A. Chains of reasoning over entities, relations, and text using recurrent neural networks. In: Proc. of the 15th Conf. of the European Chapter of the Association for Computational Linguistics (Vol. 1, Long Papers). Valencia: Association for Computational Linguistics, 2017. 132–141. [doi: 10.18653/v1/E17-1013]

[102] Guo LB, Sun ZQ, Hu W. Learning to exploit long-term relational dependencies in knowledge graphs. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 2505–2514.

[103] Puri N, Gupta P, Agarwal P, Verma S, Krishnamurthy B. MAGIX: Model agnostic globally interpretable explanations. arXiv:1706.07160, 2017.

[104] Sánchez I, Rocktäschel T, Riedel S, Singh S. Towards extracting faithful and descriptive representations of latent variable models. In: Proc. of the 2015 AAAI Spring Symp. on Knowledge Representation and Reasoning (KRR): Integrating Symbolic and Neural Approaches. Palo Alto: AAAI, 2015. 35–38.

[105] Gusmão AC, Correia AHC, De Bona G, Cozman FG. Interpreting embedding models of knowledge bases: A pedagogical approach. arXiv:1806.09504, 2018.

[106] Carmona IS, Riedel S. Extracting interpretable models from matrix factorization models. In: Proc. of the 2015 Int'l Conf. on Cognitive Computation: Integrating Neural and Symbolic Approaches. 2015. 78–84.

[107] Nandwani Y, Gupta A, Agrawal A, Chauhan MS, Singla P, Mausam. OxKBC: Outcome explanation for factorization based knowledge base completion. In: Proc. of the 2020 Conf. on Automated Knowledge Base Construction. 2020.

[108] Pezeshkpour P, Tian YF, Singh S. Investigating robustness and interpretability of link prediction via adversarial modifications. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics (NAACL). Minneapolis: Association for Computational Linguistics, 2019. 3336–3347. [doi: 10.18653/v1/N19-1337]

[109] Ying Z, Bourgeois D, You JX, Zitnik M, Leskovec J. GNNExplainer: Generating explanations for graph neural networks. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver, 2019. 829.

[110] Martínez-Romero M, Vázquez-Naya JM, Pereira J, Pereira M, Pazos A, Baños G. The iOSC3 system: Using ontologies and SWRL rules for intelligent supervision and care of patients with acute cardiac disorders. Computational and Mathematical Methods in Medicine, 2013, 2013: 650671. [doi: 10.1155/2013/650671]

[111] García-Crespo Á, Rodríguez A, Mencke M, Gómez-Berbís JM, Colomo-Palacios R. ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements. Expert Systems with Applications, 2010, 37(3): 2621–2628. [doi: 10.1016/j.eswa.2009.08.016]

[112] Sherimon PC, Krishnan R. OntoDiabetic: An ontology-based clinical decision support system for diabetic patients. Arabian Journal for Science and Engineering, 2016, 41(3): 1145–1160. [doi: 10.1007/s13369-015-1959-4]

[113] Qiang SH, Luo YL, Li YP, Wu P. Ontology reasoning for financial affairs with RBR and CBR. Data Analysis and Knowledge

Discovery, 2019, 3(8): 94–104 (in Chinese with English abstract). [doi: 10.11925/infotech.2096-3467.2018.1137]

附中文参考文献:

[2] 柳彦平, 王文杰, 荣江. 基于RDF的医疗诊断专家系统. 微型机与应用, 2005, 24(5): 36–39. [doi: 10.3969/j.issn.1674-7720.2005.05.012]

[3] 鲍玉来. 基于领域本体的蒙医药学知识库构建与知识发现研究 [博士学位论文]. 长春: 吉林大学, 2018.

[4] 杨丽. 基于知识推理的中医临床诊疗决策支持方法研究 [硕士学位论文]. 北京: 北京交通大学, 2014.

[5] 陈行军, 张晓盼, 赵晓哲. 海战场情况事件本体模型及融合方法研究. 军事运筹与系统工程, 2019, 33(1): 69–74. [doi: 10.3969/j.issn.1672-8211.2019.01.013]

[6] 熊建英. 知识图谱在异常资金流分析中的应用研究. 金融科技时代, 2021, 29(1): 28–33. [doi: 10.3969/j.issn.2095-0799.2021.01.005]

[8] 黄珂萍, 蒋昌俊. 基于本体的城市交通的知识分析和推理. 计算机科学, 2007, 34(3): 192–196. [doi: 10.3969/j.issn.1002-137X.2007.03.051]

[9] 梁艺多, 翟军, 袁长峰. 基于本体推理的物流配送系统的构建. 物流技术, 2015, 34(9): 255–258.

[22] 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): 2966–2994. http://www.jos.org.cn/1000-9825/5551.htm [doi: 10.13328/j.cnki.jos.005551]

[33] 纪守领, 李进锋, 杜天宇, 李博. 机器学习模型可解释性方法、应用与安全研究综述. 计算机研究与发展, 2019, 56(10): 2071–2096. [doi: 10.7544/issn1000-1239.2019.20190540]

[113] 强韶华, 罗云鹿, 李玉鹏, 吴鹏. 基于RBR和CBR的金融事件本体推理研究. 数据分析与知识发现, 2019, 3(8): 94–104. [doi: 10.11925/infotech.2096-3467.2018.1137]

侯中妮(1996－), 女, 博士生, 主要研究领域为知识图谱, 事理图谱.

官赛萍(1991－), 女, 博士, 助理研究员, 主要研究领域为知识图谱, 事理图谱, 多元关系推理.

靳小龙(1976－), 男, 博士, 研究员, 博士生导师, CCF 高级会员, 主要研究领域为大数据知识工程, 知识图谱.

王元卓(1978－), 男, 博士, 研究员, 博士生导师, CCF 杰出会员, 主要研究领域为大数据分析, 开放知识网络, 社交演化分析.

陈剑赟(1977－), 女, 博士, 主要研究领域为智能信息处理, 系统工程.

程学旗(1971－), 男, 博士, 教授, 博士生导师, CCF 会士, 主要研究领域为网络科学与社会计算, 互联网搜索与挖掘.