

QA-KGNet:一种语言模型驱动的知识图谱问答模型^{*}



乔少杰¹, 杨国平¹, 于泳^{1,2}, 韩楠³, 覃晓⁴, 屈露露¹, 冉黎琼¹, 李贺⁵

¹(成都信息工程大学 软件工程学院, 四川 成都 610225)

²(数字媒体艺术四川省重点实验室, 四川音乐学院, 四川 成都 610021)

³(成都信息工程大学 管理学院, 四川 成都 610103)

⁴(广西人机交互与智能决策重点实验室, 南宁师范大学, 广西 南宁 530100)

⁵(西安电子科技大学 计算机科学与技术学院, 陕西 西安 710071)

通讯作者: 韩楠, E-mail: hannan@cuit.edu.cn

摘要: 基于知识图谱的问答系统可以解析用户问题,已经成为一种检索知识、自动回答所询问题的有效途径.知识图谱问答系统通常是利用神经网络归纳模型,将自然语言问题转化为逻辑形式,在知识图谱上执行该逻辑形式能够得到答案.然而,使用预训练语言模型和知识图谱的知识问答系统包含两个挑战:1)给定问答(QA, Question-answering)上下文,需要从大型知识图谱(KG, Knowledge Graph)中识别相关知识;2)对 QA 上下文和 KG 进行联合推理.基于此,提出一种语言模型驱动的知识图谱问答推理模型 QA-KGNet,将 QA 上下文和 KG 连接起来形成一个工作图,使用语言模型计算给定 QA 上下文节点与 KG 节点的关联度,并使用多头图注意力网络更新节点表示.在 CommonsenseQA, OpenBookQA 和 MedQA-USMLE 真实数据集上进行实验来评估 QA-KGNet 的性能,实验结果表明,QA-KGNet 优于现有的基准模型,表现出优越的结构化推理能力.

关键词: 知识图谱; 预训练语言模型; QA 上下文; 多头图注意力网络; 联合推理

中图法分类号: TP311

中文引用格式: 乔少杰,杨国平,于泳,韩楠,覃晓,屈露露,冉黎琼,李贺. QA-KGNet:一种语言模型驱动的知识图谱问答模型. <http://www.jos.org.cn/1000-9825/6882.htm>

英文引用格式: Qiao SJ, Yang GP, Yu Y, Han N, Qin X, Qu LL, Ran LQ, Li H. QA-KGNet: a language model-driven knowledge graph question-answering model. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/6882.htm>

QA-KGNet: A Language Model-Driven Knowledge Graph Question-Answering Model

QIAO Shao-Jie¹, YANG Guo-Ping¹, YU Yong^{1,2}, HAN Nan³, QIN Xiao⁴, QU Lu-Lu¹, RAN Li-Qiong¹, LI He⁵

¹(School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

²(Digital Media Art, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music, Chengdu 610021, China)

³(School of Management, Chengdu University of Information Technology, Chengdu 610225, China)

⁴(Guangxi Key Lab of Human-machine Interaction and Intelligent Decision, Nanning Normal University, Nanning 530100, China)

⁵(School of Computer Science and Technology, XiDian University, Xi'an 710071, China)

Corresponding author: QIAO Shao-Jie, E-mail: sjqiao@cuit.edu.cn

* 基金项目: 国家自然科学基金项目(61962006); 四川省科技计划项目(2021JDJQ0021, 2022YFG0186); 四川音乐学院数字媒体艺术四川省重点实验室资助项目(21DMAKL02); 成都市技术创新研发项目(2021-YF05-00491-SN); 成都市重大科技创新项目(2021-YF08-00156-GX); 成都市“揭榜挂帅”科技项目(2021-JB00-00025-GX); 成都市软科学研究项目(2021-RK00-00065-ZF, 2021-RK00-00066-ZF); 广西重大创新驱动项目(桂科 AA22068057); 四川省社会科学高水平团队项目(2015Z177)

收稿时间: 2022-07-02; 修改时间: 2022-08-18, 2022-12-14; 采用时间: 2022-12-28; jos 在线出版时间: 2023-01-13

Abstract: The question-answering system based on knowledge graphs can analyze user questions, and has become an effective way to retrieve relevant knowledge and automatically answer the given questions. The knowledge graph-based question-answering system usually uses a neural program induction model to convert natural language question into a logical form, and the answer can be obtained by executing the logical form on the knowledge graph. However, the knowledge question-answering system by using pre-trained language models and knowledge graphs involves two challenges: 1) given the QA (Question-answering) context, relevant knowledge needs to be identified from a large KG (Knowledge Graph); 2) we need to perform the joint reasoning on QA context and KG. Based on these challenges, a language model-driven knowledge graph question-answering model is proposed, which connects the QA context and KG to form a joint graph, and uses a language model to calculate the relevance of the given QA context nodes and KG nodes, and a multi-head graph attention network is employed to update the node representation. Extensive experiments on the CommonsenseQA, OpenBookQA and MedQA-USMLE real datasets are conducted to evaluate the performance of QA-KGNet and the experimental results show that QA-KGNet outperforms existing benchmark models and exhibits excellent structured reasoning capability.

Key words: knowledge graph; pre-trained language models; QA context; multi-head graph attention network; joint reasoning

知识图谱^[1]是由 Google 公司于正式提出.从狭义角度来说,知识图谱指的是在语义网研究领域中,经过演化而产生的具体知识库,例如 DBpedia^[2]和 Freebase^[3].从广义角度来说,知识图谱指的是基于图结构构建的知识集合^[4],比如资源描述框架(RDF,Resource Description Framework)数据集或者本体(Ontology).知识图谱以图结构描述真实世界的客观实体与概念,以及这些实体、概念之间的关系^[5].进一步讲,知识图谱将 Web 应用的异源异构数据链接融合为统一的形式化知识,使得纷繁庞杂的信息更易于机器的理解与计算,并将 Web 应用上的知识统一组织.最终将互联网海量的信息表示成符合人类认知与理解的形式,使相关知识的关联更紧密,也便于人们获取、溯源所需要的知识^[6,7,8].

传统的搜索方式是搜索引擎使用关键字匹配(如模糊查询),返回给用户包含关键字的一系列粗粒度网页,然后由用户手动从这些网页中筛选,并从网页中识别自己所需要的信息.利用知识图谱的语义搜索具有一定的联想能力,能够根据知识图谱的结构、查询情景与查询意图,将相关联的实体、概念等知识以卡片的形式返回给用户,提供一个相对细粒度的答案.

随着人工智能技术快速发展,基于知识图谱的问答系统应运而生.可以使用户快捷地获取知识,在学界和工业界均引起了广泛关注,成为近年来研究的热点问题^[9,10,11].目前,知识图谱问答已经成为推动人机交互与互联网语义搜索的重要驱动力之一,也在众多实际智能应用中发挥巨大的作用.

知识图谱问答系统必须能够访问相关知识并对其进行推理.通常,对非结构化文本进行预训练的语言模型可以对知识进行隐式编码^[12,13],然后在结构化知识图谱中显式表示^[14,15].其中,实体表示为节点,关系表示为边.最近,预训练语言模型在许多问答任务中都取得了显著的成功^[16,17].然而,虽然语言模型涵盖了广泛的知识,但在结构化推理方面的表现欠佳.而知识图谱更适合结构化推理,并且可以通过提供推理路径实现可解释的预测,但可能缺乏覆盖范围或带有噪音.如何有效地利用语言模型(LM,Language Model)和知识图谱进行推理仍然是一个重要的开放性问题.

研究动机.现有的工作^[18,19,20]通过获取主题实体(即给定 QA 上下文中提到的 KG 实体)及其少跳(few-hop)邻居节点,从 KG 中检索子图.但是,这会引入许多在语义上与 QA 上下文无关的实体节点,尤其是当主题实体或跳点(hops)的数量增加时.此外,现有的 LM+KG 推理方法^[21,22]将 QA 上下文和 KG 视为两种独立的模态.将 LM 简单应用于 QA 上下文,将图神经网络应用于 KG,并不会相互更新或统一它们的表示.这种相对独立的方案可能会限制结构化推理的能力.基于此,Yasunaga 等人^[23]提出了 QA-GNN 模型,将语言模型与知识图谱相结合,使用图注意力网络提取知识图谱节点信息,并进行消息传递.然而,QA-GNN 只考虑了一组随机初始化权重,无法从多个角度对知识图谱的信息进行提取,导致泛化性较低.

方法先进性.图 1 展示了同时考虑 QA 上下文和知识图谱节点(即使用 LM+KG 的方法)的案例,可以看到 QA 上下文被视为知识图谱的特殊节点.QA 上下文节点中的问题词与答案词分别被视为问题实体(Question

entity)与选项实体(Choice entity).给定 QA 上下文节点,从大型 KG 中识别信息化知识,然后获取 QA 上下文的细微差别和 KG 的结构,并结合二者以进行联合推理.图 1 所示案例兼顾了 QA 上下文与知识图谱的联合信息表示,大多数工作^[18,19,20,21,22]不具备联合推理能力,而文献[23]虽然具备联合推理能力,但模型的泛化性较低.

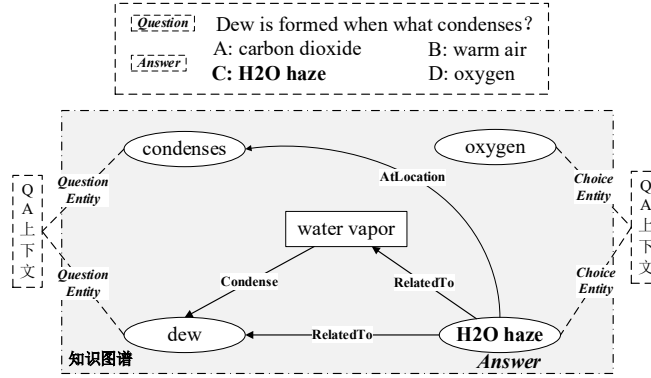


Fig.1 Case of the joint graph of QA and KG

图 1 QA 和 KG 的工作图案例

基于上述分析,提出一种语言模型驱动的知识图谱问答推理模型 QA-KGNet.使用语言模型对 QA 上下文进行编码,并检索 KG 子图.主题实体的所有少跳邻居节点组成 KG 子图,某些实体节点与给定 QA 上下文的关联度较高.本文通过将实体与 QA 上下文连接起来,并使用预训练语言模型计算 QA 上下文实体与 KG 子图实体的关联度.如图 1 所示,引入了 QA 上下文和 KG 的工作图表示,将 QA 上下文视为一个附加节点,并将其连接到 KG 子图中的主题实体.将两种模式统一为工作图,然后用关联度来增加每个节点的特征,并使用多头图注意力网络从多角度对多轮消息进行传递,动态更新节点表示.使用图注意力池化对工作图进行降采样,最后将语言模型表示、上下文节点表示与降采样结果进行整合,由多层感知机输出概率值.

本文的主要贡献包括:

- (1) 提出一种语言模型驱动的知识图谱问答推理模型 QA-KGNet,使用语言模型对 QA 上下文进行编码,并结合 KG 子图的编码.将 QA 上下文与 KG 子图统一为一个工作图,用关联度来增加每个节点的特征,兼顾了 QA 上下文和知识图谱的联合信息表示.
- (2) 提出一种基于多头注意力的消息表示和注意力系数,将该注意力系数应用于多头图注意力网络中,使模型能够在工作图上从多角度对多轮消息进行传递,动态更新节点表示.使用图注意力池化对工作图进行降采样,最后将语言模型表示、上下文节点表示与降采样结果进行整合,由多层感知机输出概率值.
- (3) 本文在 CommonsenseQA, OpenBookQA 和 MedQA-USMLE 数据集上进行实验.实验结果表明,所提 QA-KGNet 的推理准确性超过了其他基准方法.此外,QA-KGNet 具有较强的可解释性与结构化推理能力.

1. 相关工作

由于常识知识并不能完全融入实体与关系的匹配模式中,传统方法仅捕获那些在文本中明确提及的实体或关系,限制了在获取常识知识方面的适用性.而预训练语言模型可以学习潜在的常识知识.Petroni 等人^[12]对预训练语言模型中存在的关系知识进行了深入分析,证明了这些模型无需任何微调即可得出相关知识.Bosselut 等人^[13]通过实验证明了预训练语言模型的隐式知识可以被迁移到常识知识图谱中,进而生成显式知识.Pan 等人^[24]提出利用外部知识来源进行学科领域问答的方法.为了提供更明确和可解释的知识,一些工作将结构化知识集成到语言模型中.Mihaylov 等人^[25]提出神经阅读理解模型,该模型关注相关的外部知识,并在推断答案之前将这些知识与上下文表示相结合.Bosselut 等人^[26]提出一种神经常识知识模型,可以按需生成与上下文相关的符号知识结构.上述工作充分利用预训练语言模型中的隐式知识,将隐式知识迁移到知识图谱中以执行下游任务.一些研究尝试将语言模型与知识图谱相结合(LM+KG)来执行问答任务.例如,Lin 等人^[18]提出用于回答常识

性问题的文本推理框架,该框架有效地利用外部的结构化知识图谱来执行可解释的推理.Feng 等人^[22]提出一种新颖的知识感知方法,该方法为预训练语言模型分配了多跳关系推理模块,对从外部知识图谱中提取的子图执行多跳、多关系推理.此外,发现对 KG 节点/路径进行评分或修剪依赖于基于图的指标.Paul 等人^[27]提出一种从常识知识资源中提取、排序、过滤和选择多跳关系路径的方法,根据人类潜在的需求来解释情感表达.然而,现有基于 LM+KG 的方法没有将上下文节点与知识图谱节点有机结合,二者相对独立.通常的方法是将 LM 简单应用于 QA 上下文,将图神经网络应用于 KG.但这类方法中上下文节点与知识图谱节点并不会相互更新或统一它们的表示.这种相对独立的方案可能会限制结构化推理的能力.

不同形式的问答任务引起了研究者的关注,例如基于段落的问答^[28]和外部知识基问答^[29].Yang 等人^[28]采用注意力机制从外部知识中自适应地选择所需的知识,然后将选择的知识与 BERT(Bidirectional Encoder Representation from Transformers)^[30]融合来实现段落和知识感知的预测.这一方法扩展了简单的问答形式,充分利用预训练语言模型 BERT 中的隐式知识.Yih 等人^[29]证明为知识库问答收集语义分析标签的重要性,使用适当的用户接口以低成本获得高精度的语义解析,创建并共享了语义解析标记数据集,以推进问答研究.这种方法虽然能取得高精度语义,但需要为知识库问答收集语义分析标签,收集过程的代价较大.

外部文本知识和结构化知识的联合表示可以给问答任务带来极大增益.Sun 等人^[31]提出 PullNet 集成框架,可以学习检索内容以及使用异构信息进行推理以找到最佳答案.PullNet 使用迭代过程来构建包含问题相关信息的子图.Xiong 等人^[32]提出一种端到端问答模型,该模型首先从与问题相关的知识库子图中积累实体知识,然后重新构造潜在空间中的问题,并使用积累的实体知识理解文本.这种联合表示的方法是合理的,因为可以同时将外部文本知识和结构化知识视为一个整体,便于图神经网络学习.

此外,现实中越来越多的问题(如知识问答)可以抽象为非欧几里德数据结构(即数据结构中的节点无固定的排列规则和顺序),传统深度学习模型很难学习非欧几里德数据.而图神经网络(GNN, Graph Neural Network)已被证明可有效地对非欧几里德数据进行建模.图神经网络(GNN, Graph Neural Network)已被证明可有效地对基于图形的数据进行建模.Zhang 等人^[33]提出一种扩展图卷积网络,可以并行地在任意依赖结构上汇集信息.为了最大程度地删除不相关内容并融合相关信息,对输入树使用修剪策略.Yasunaga 等人^[34]提出一个程序反馈图,将源代码中与程序修复相关的符号和诊断反馈联系起来,在程序顶部应用一个图神经网络来对推理过程进行建模.文献[34]还提出了一种用于程序修复的自我监督学习范式,它利用在线可用的未标记程序来创建大量额外的程序修复样本,可以利用这些样本训练模型.上述工作依赖于图神经网络对非欧几里德数据的强大建模能力,对输入节点的顺序没有要求,在各个顶点的边中存储着相邻节点的关系,通过对中心节点的邻居节点加权来更新中心节点的隐藏状态.

2. 问题描述

2.1 背景知识

目前的高性能知识图谱问答方法均在问答系统中通常设计了一个语义分析框架.语义分析框架旨在将自然语言问题转化为能在知识图谱上直接执行而得到相应答案的逻辑形式.本文将问题对应的逻辑形式称为标注(annotation),将逻辑形式的执行结果称为预测答案(denotation),而将问题对应的正确答案称为真实答案(ground-truth answer)或黄金答案(golden answer).

在本文中,将知识图谱用 G 表示,问题以 q 表示,其由 n 个单词组成,单词记为 k ,其表达形式为 $q = (k_1, k_2, \dots, k_n)$.所有可能的候选答案集合表示为 C .知识图谱 G 是一组事实(fact)的集合.每一个事实均表示为一个三元组 (s, p, o) ,其中 p 表示谓词(比如属于,相关于), s 和 o 表示实体或值,可以作为三元组的主语和宾语.复合值类型是一种特殊的实体类别,用来表示一个事件,而不是一个真实世界中的实体.

知识图谱问答任务指的是从候选答案集合 C 构成的概率分布中找到概率值最大的答案,并将此答案表示为预测答案 a_d ,形式化表示如公式(1)所示.

$$a_d = \arg \max_{a \in C} P(a | q) \quad (1)$$

本文将问题的真实答案表示为 a_p , 而知识图谱问答的任务目标是使 $P(a_d = a_p | q)$ 的概率最大化.

2.2 问题描述

本文研究目标是使用预训练语言模型和结构化知识图谱的知识来回答自然语言问题. 使用 $f_{task}(f_{enc}(x))$ 来表示所要执行的任务模型, 其中, 编码器 f_{enc} 将文本输入 x 映射为上下文向量 $\mathbf{h}^{LM}_{f_{task}}$ 使用 $\mathbf{h}^{LM}$ 来执行相关性计算任务(参见 3.2 节). 本文使用具体的掩码语言模型作为编码器 f_{enc} . 将知识图谱定义为多关系图 $G=(V, E)$, 其中, V 表示知识图谱中实体节点的集合; $E \subseteq V \times R \times V$ 是连接 V 中节点的边集合, R 表示关系类型的集合.

给定一个问题 q 和一个答案选项 $a \in C$, 将问题和答案选项中提到的实体连接到给定的知识图谱 G . 令 $V_q \subseteq V$ 和 $V_a \subseteq V$ 分别为问题中提到的实体集和答案选项中提到的实体集, 并使用 $V_{q,a} := V_q \cup V_a$ 表示所有出现在问题或答案选项中的实体, 称之为主题实体集. 然后, 从 G 中提取一个问题-答案选项对 (Question-Answer Option Pair) 的子图 $G'_{q,a} = (V'_{q,a}, E'_{q,a})$, 它包括 $V_{q,a}$ 中节点之间 g 跳路径上的所有节点. 少跳节点和多跳节点统称为 g 跳节点, 通常 $g=3$ 为界限^[22]. 这样的划分标准是根据推理精确度, 如果跳数 g 过少, 可能会丢失部分重要的节点, 导致推理准确度的损失. 如果跳数 g 过多, 虽然可以考虑更多的重要节点, 但会增加许多噪音节点, 从而影响准确率.

表 1 总结了本文关于图结构的符号及其含义, 便于查阅.

Table 1 Symbols in graph structures and their meanings
表 1 图结构符号及其含义

符号	含义
V	实体节点集合
E	边集合
R	关系类型集合
G	知识图谱
V_q	问题实体集
V_a	答案选项实体集
$V_{q,a}$	问题-答案选项实体集
$V'_{q,a}$	问题-答案选项子实体集
$E_{q,a}$	问题-答案选项边集
$E'_{q,a}$	问题-答案选项子边集
$G_{q,a}$	由 $V_{q,a}$ 和 $E_{q,a}$ 构成的知识图谱
$G'_{q,a}$	由 $V'_{q,a}$ 和 $E'_{q,a}$ 构成的子知识图谱
T	QA 上下文节点
E_J	包含 T 节点的工作图边集
V_J	包含 T 节点的工作图实体集
G_J	包含 T 节点的工作图
$R_{T,q}$	T 节点与问题实体的关系
$R_{T,a}$	T 节点与答案选项实体的关系

3 构建 QA-KGNet

QA-KGNet 的工作原理如图 2 所示, 给定一个问题 q 和一个答案选项 a , 将它们拼接以获得 QA 上下文 $[q; a]$. 为了使用语言模型和知识图谱中知识对给定的 QA 上下文进行推理, 首先利用语言模型获得 QA 上下文的表示, 并从知识图谱中检索子图. 然后, 引入表示 QA 上下文的节点 T , 将节点 T 链接到主题实体集 $V_{q,a}$, 以便在两个知识源上(即语言模型与知识图谱)拥有同一工作图, 记为 G_J . 为了自适应地捕获 QA 上下文节点与 G_J 中其他节点之间的关系, 本文计算节点 T 和 G_J 中其他节点之间的关联度. 并将该关联度作为每个节点的附加特征. 此外, 提出了一个新颖的多头图注意力网络, 该网络在 G_J 上进行多轮消息传递. 使用注意力池化^[35]对工作图进行降采

样,最后将语言模型表示、QA 上下文节点表示与降采样结果进行整合,由多层感知机输出概率值..

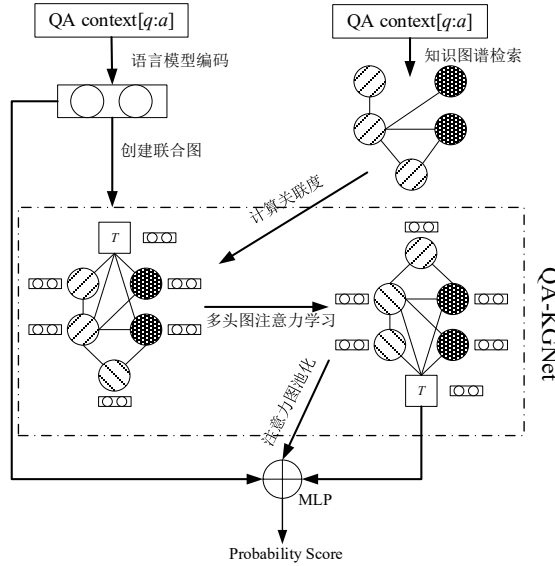


Fig.2 Working mechanism diagram of QA-KGNet

图 2 QA-KGNet 工作原理图

3.1 工作图表示

本文遵循文献[23]的联合图表示法,将语言模型和知识图谱链接在同一图结构中.引入一个表示 QA 上下文的节点 T ,并使用两个关系 $R_{T,q}$ 和 $R_{T,a}$ 分别将节点 T 与主题实体集 $V_{q,a}$ 中的每一个主题实体相连.其中, $R_{T,q}$ 表示节点 T 与某一问题实体之间的关系, $R_{T,a}$ 表示节点 T 与某一答案选项实体之间的关系. $R_{T,q}$ 和 $R_{T,a}$ 关系捕获 QA 上下文与知识图谱中相关实体之间的联系,具体取决于实体是在 QA 上下文的问题还是答案选项中找到.该工作图直观地提供了 QA 上下文和知识图谱的推理空间,将其称为工作图 $G_J = (V_J, E_J)$, 其中 $V_J = V'_{q,a} \cup \{T\}$, $E_J = E'_{q,a} \cup \{(T, R_{T,q}, v) | v \in V_q\} \cup \{(T, R_{T,a}, v) | v \in V_a\}$.

G_J 中的每个节点属于以下四种类型中的一种,即 QA 上下文节点 T 、 V_q 集合中的节点、 V_a 集合中的节点以及其他节点.将 QA 上下文节点 T 和知识图谱节点 $v \in V'_{q,a}$ 的文本分别表示为 $text(T)$ 和 $text(v)$.使用 QA 上下文的语言模型表示作为节点 T 的嵌入,即 $T^{LM} = (f_{enc}(text(T)))$,并通过实体嵌入来初始化 $G'_{q,a}$ 上的每个节点.

3.2 节点关联度

子图 $G'_{q,a}$ 的一些节点在当前 QA 上下文中可能是不相关的.如图 3 所示,本文用节点的深浅来表示关联度的强弱.检索到的子图 $G'_{q,a}$ 可能包括对推理过程没有作用的节点,例如节点“食物(food)”和“脚(foot)”是脱离主题的.这些不相关的节点可能会导致过度拟合或引入不必要的计算开销,尤其是当 $V_{q,a}$ 较大时.例如,当使用 ConceptNet 知识图谱^[14]时,如果仅考虑 3 跳邻居节点,将会检索平均节点数 $|V'_{q,a}| > 400$ 的知识图谱.

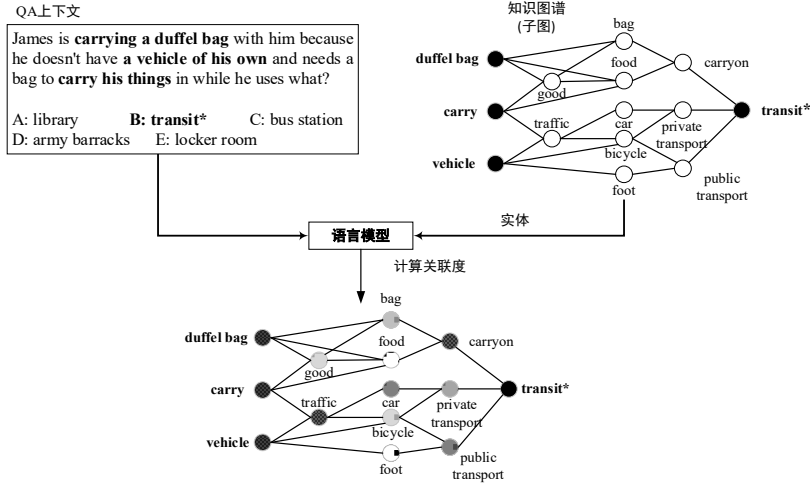


Fig.3 Diagram for calculating node's relevance

图3 节点关联度计算图

使用预训练的语言模型对每个知识图谱节点 $v \in V'_{q,a}$ 的关联度进行计算.在给定 QA 上下文的条件下,对于每个节点 v ,将实体表示 $text(v)$ 与 QA 上下文表示 $text(T)$ 拼接,并计算节点关联度,如公式(2)所示.

$$S = f_{task}(f_{enc}([text(T):text(v)])) \quad (2)$$

其中 f_{enc} 表示 RoBERTa^[17]词嵌入模型, f_{task} 由一组简单的变换组成.选择 RobertaLMHead¹作为 f_{task} .该关联度 S 描述了每个知识图谱节点相对于给定 QA 上下文的重要性,用于推理或修剪工作图 G_J .

3.3 多头图注意力网络推理

为了对工作图 G_J 进行推理,引入一种图注意力网络对图上邻居之间的迭代消息进行传递,然后更新节点表示在 L 层的 QA-KGNet 中,对于每一层 l ,通过公式(3)更新每个节点 $t \in V_J$ 的表示 $h_t^{(l)}$.

$$h_t^{(l+1)} = \rho\left(\sum_{s \in N_t \cup \{t\}} (\alpha_{st} \bullet m_{st})\right) + h_t^{(l)} \quad (3)$$

其中, N_t 表示节点 t 的一阶邻域, s 表示节点 t 的一阶邻域 $N_t \cup \{t\}$ 中的某一节点(包括 t 本身).将注意力系数 α_{st} 与消息特征 m_{st} 相乘,将相乘后的结果进行累加,最后使用激活函数 ρ 进行变换,得到更新后节点 t 的表示.激活函数 ρ 由两个 MLP(Multilayer Perceptron)组成,进行批量标准化(Batch Normalization).对于每个节点 $t \in V_J$,使用矩阵线性变换将初始化节点嵌入映射到同一欧几里得空间,将初始化所得的节点嵌入记为 $h_t^{(0)}$.当节点消息在工作图上传递时,图注意力网络将同时利用并更新 QA 上下文和知识图谱的表示.

为了使图注意力网络的学习过程更加准确,本文受 Vaswani 等人^[36]与 Velickovic 等人^[37]的工作启发,在公式(3)的基础上为图注意力网络提供了不同角度(多头)的考虑,提出了多头图注意力网络.将公式(3)优化为公式(4),如下所示.

$$h_t^{(l+1)} = \parallel_{k=1}^K \left(\rho\left(\sum_{s \in N_t \cup \{t\}} (\alpha_{st}^k \bullet m_{st}^k)\right) \right) + h_t^{(l)k} \quad (4)$$

其中, K 表示不同角度的数量(头的数量),“ $\parallel$ ”表示向量拼接,上标 k 表示第 k 个角度.每个角度的学习过程独立执行,将学到的向量进行拼接.

图4直观地展示了多头图注意力网络案例.图4中不同线型的箭头表示不同的头,将头的数量 K 设为3.不同线型的箭头分别代表一组随机初始化的参数,不同组的参数独立训练.虚线所表示的头(记为 head₁),将节点 $s_1 \sim$ 节点 s_5 分别与节点 t 的注意力关系向量进行整合,得到向量 V_1 .实线所表示的头(记为 head₂),将节点 $s_1 \sim$ 节点

¹ https://github.com/huggingface/transformers/blob/main/src/transformers/models/roberta/modeling_roberta.py

s_5 分别与节点 t 的注意力关系向量进行整合,得到向量 V_2 .同理,对于点划线所表示的头(记为 head_3),整合后得到向量 V_3 .最终将不同头的整合向量 V_1 、 V_2 和 V_3 进行拼接,得到最终向量 V .

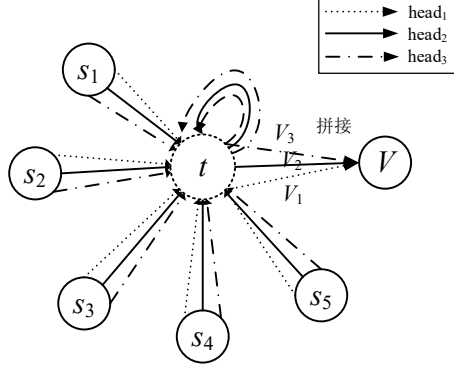


Fig.4 Case of a multi-head graph attention network

图 4 多头图注意力网络案例

进一步对消息表示和注意力系数进行讨论.

消息表示.由于 G_j 是一个多关系图,从源节点传递到目标节点的信息应该包含它们的关系,即边的关系类型和源/目标节点类型.首先,获得每个节点 t 的类型嵌入 u_t ,以及从节点 s 到节点 t 的关系嵌入 R_{st} ,如公式(5)-公式(6)所示.

$$u_t = f_u(u_t^o) \quad (5)$$

$$R_{st} = f_r(e_{st}^o, u_s^o, u_t^o) \quad (6)$$

其中, u_s^o 和 u_t^o 表示 s 和 t 的节点类型, e_{st}^o 表示边 (s, t) 的关系类型,它们均为 one-hot 向量, f_u 是线性变换函数, f_r 由两个 MLP 组成.消息表示 m_{st} 的计算方式如公式(7)所示.

$$m_{st} = f_m(h_s^{(l)}, u_s, R_{st}) \quad (7)$$

其中 f_m 表示线性变换函数.

注意力系数.注意力机制能够捕捉两个节点之间的相关性,取决于它们的节点类型、关系和节点关联度.将每个节点 t 的相关性分数 S 进行嵌入,得到嵌入向量 S_t ,如公式(8)所示.

$$S_t = f_e(S) \quad (8)$$

其中 f_e 表示一个 MLP.为了计算从节点 s 到节点 t 的注意力权重 α_{st} ,需要先计算 query 向量和 key 向量,分别记为 q_s 和 k_t .计算方法如公式(9)-公式(10)所示.

$$q_s = f_q(h_s^{(l)}, u_s, S_s) \quad (9)$$

$$k_t = f_k(h_t^{(l)}, u_t, S_t, R_{st}) \quad (10)$$

其中 f_q 和 f_k 是线性变换函数.注意力权重的计算方式如公式(11)-公式(12)所示, T 表示转置操作, D 表示 q_s 和 k_t 的向量长度.

$$\alpha_{st} = \text{softmax}(R_{st}) = \frac{\exp(R_{st})}{\sum_{t^* \in N_s \cup \{s\}} \exp(R_{st^*})} \quad (11)$$

$$R_{st} = \frac{q_s^T k_t}{\sqrt{D}} \quad (12)$$

给定一个问题 q 和一个答案选项 a ,使用 QA 上下文和知识图谱来计算 a 是正确答案的概率 P ,如公式(13)所示.

$$P(a | q) = \text{softmax}(\text{MLP}(T^{LM}, T^{GNN}, g_p)) \quad (13)$$

$$g_p = \text{AttPool}(T^{LM}, h_v^{(L)}) \quad (14)$$

其中, T^{LM} 表示 QA 上下文节点 T 的语言模型嵌入向量, T^{GNN} 等价于 $h_T^{(L)}$, $h_T^{(L)}$ 为 QA 上下文节点 T 的多头图注意力嵌入向量, 可利用公式(4)计算得到. g_p 表示向量 $h_v^{(L)}$ 经过注意力图池化(AttPool)^[35]后的结果, $v \in V'_{q,a}$. 在训练数据中, 每个问题 q 都有一组答案选项和一个正确选项, 然后使用交叉熵损失优化模型进行梯度下降.

4 实验结果与分析

4.1 数据集

本文根据文献[23]的实验设置, 分别在 CommonsenseQA^[38]、OpenBookQA^[39]和 MedQA-USMLE 真实数据集^[40]上进行实验. CommonsenseQA 包含 12,102 个问题, 每个问题包含 5 个选项, 需要使用常识知识进行推理. 本文根据文献[18]的拆分方式将数据集拆分方式对 CommonsenseQA 进行拆分². OpenBookQA 包含 5,957 个问题, 每个问题包含 4 个选项, 需要使用基础科学知识进行推理. 本文使用文献[25]对 OpenBookQA 官方数据集进行拆分³. MedQA-USMLE 包含 12,723 个问题, 每个问题包含 4 个选项, 需要使用生物医学和临床知识进行推理, 使用文献[39]的原始数据拆分方法.

对于 CommonsenseQA 和 OpenBookQA 数据集, 本文选择通用领域知识图谱 ConceptNet⁵作为结构化知识源 G , 共包含 799,273 个节点和 2,487,810 条边. 使用文献[22]的实体嵌入进行节点嵌入初始化⁶, 它将预训练语言模型应用于 ConceptNet 中的所有三元组, 然后为每个实体获得一个聚合表示.

对于 MedQA-USMLE, 使用了一个自构建的知识图谱, 它集成了统一医学语言系统^[41]和 DrugBank 疾病数据库^[42], 包含 9,958 个节点和 44,561 条边. 使用 SapBERT^[43]的实体名称表示对节点嵌入进行初始化.

为了从领域知识图谱 ConceptNet 中检索包含丰富信息的子图 $G'_{q,a}$, 需要识别 QA 上下文中的实体并将它们链接到 ConceptNet 的实体中, 然后使用这些实体来初始化节点集 $V'_{q,a}$. 然后, 将实体对之间的任何两跳路径中的所有实体添加到 $V'_{q,a}$ 中, 不执行任何剪枝操作, 只需保留 $V'_{q,a}$ 中节点之间所有的边, 最终得到子图 $G'_{q,a}$.

4.2 基准模型

首先, 考虑使用预训练语言模型的基准模型来执行问答任务. 将 RoBERTa-large^[44]用于 CommonsenseQA 数据集, 将 RoBERTa-large^[44]和 AristoRoBERTa^[45]用于 OpenBookQA 数据集. 对于 MedQA-USMLE 数据集, 使用生物医学语言模型 SapBERT^[43]. 其中 RoBERTa-large 和 AristoRoBERTa 均基于 BERT 模型, 属于大型预训练语言模型, 适用于常识知识推理. 而 SapBERT 专注于生物医学领域的语言预训练.

然后, 考虑使用预训练语言模型与知识图谱结合的基准模型, 包括 KagNet^[19]、MHGRN(Multi-hop graph relational network)^[22]、QA-GNN^[23]、关系网络(Relation network, RN)^[46]、RGCN(Relational graph convolutional network)^[47]和 GconAttn^[48]. 其中, RN、RGCN 和 GconAttn 是基于关系感知的图神经网络, KagNet 和 MHGRN 进一步对知识图谱的路径进行建模. 为了保证实验结果的公平可对比性, 所有基准方法和所提 QA-KGNet 使用相同的语言模型.

4.3 实现细节

本文将多头图注意力模块的向量维度 D 设置为 200, 层数 L 为 5, 每层 l 的 dropout 率为 0.25. 选择 Adam^[49]作为参数优化器, 批大小设置为 64, 语言模型的学习率设为 1×10^{-5} , 多头图注意力模块的学习率设为 2×10^{-4} . 实验硬件环境为 Intel(R) Xeon(R) CPU i7-6700k, 32GB 内存, 256G SSD, Ubuntu 20.04 操作系统以及 GPU(GeForce RTX 3090 Ti). 使用 PyTorch 框架⁷实现所提模型.

² <https://github.com/INK-USC/KagNet>

³ <https://github.com/allenai/OpenBookQA>

⁵ <https://github.com/commonsense/conceptnet5>

⁶ <https://github.com/INK-USC/MHGRN>

⁷ <https://github.com/pytorch/pytorch>

4.4 准确率测试

图 5 和图 6 分别展示了不同基准模型在 CommonSenseQA 和 OpenBookQA 数据集上的实验结果.在图 5 中,选用 RoBERTa-large(简称为 RoBERTa)^[44]作为语言模型,RoBERTa(w/o KG)表示仅使用该语言模型执行问答任务.对于选择 RoBERTa-large 作为语言模型的其他基准模型,本文所提 QA-KGNet 在开发集与测试集上的寻找到正确答案的准确率均为最高.在 CommonSenseQA 测试集上,QA-KGNet 比 RoBERTa(w/o KG)的准确率高 5.32%,比 MHGRN^[23]的准确率高 2.41%.在图 6 中,选用两种语言模型(即 RoBERTa-large 和 AristoRoBERTa^[45])分别结合不同的基准模型进行准确率测试.结果表明在 OpenBookQA 数据集上,所提 QA-KGNet 结合不同的语言模型(Language model,LM)的准确率均为最高.上述实验结果说明了基于多头图注意力的关联度和工作图更新是有效的,且 QA-KGNet 与不同的语言模型相结合均能获得大的性能提升,而其他基准模型在设计的时候并未考虑语言模型能够提供的上下文信息,不能有效地整合知识图谱和语言模型的信息.

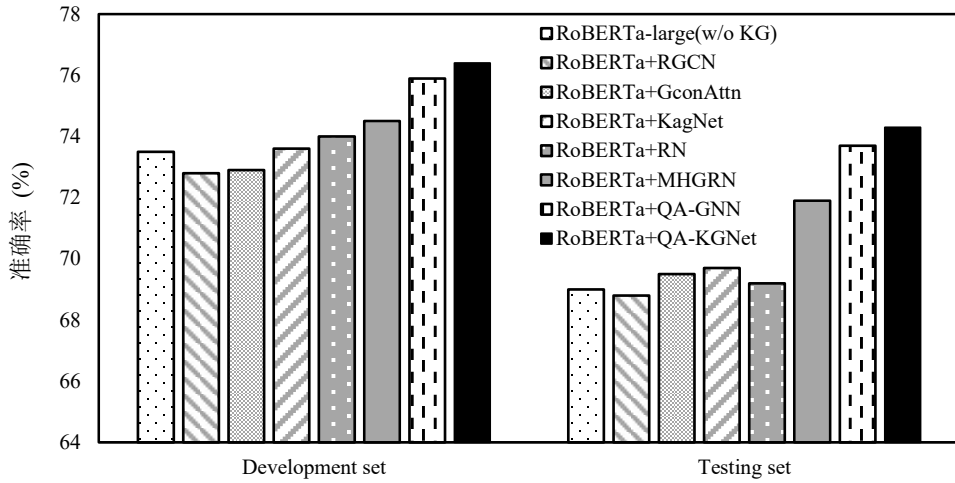


Fig.5 Accuracy comparison of different baseline models on the CommonSenseQA dataset
图 5 不同基准模型在 CommonSenseQA 数据集上的准确率对比

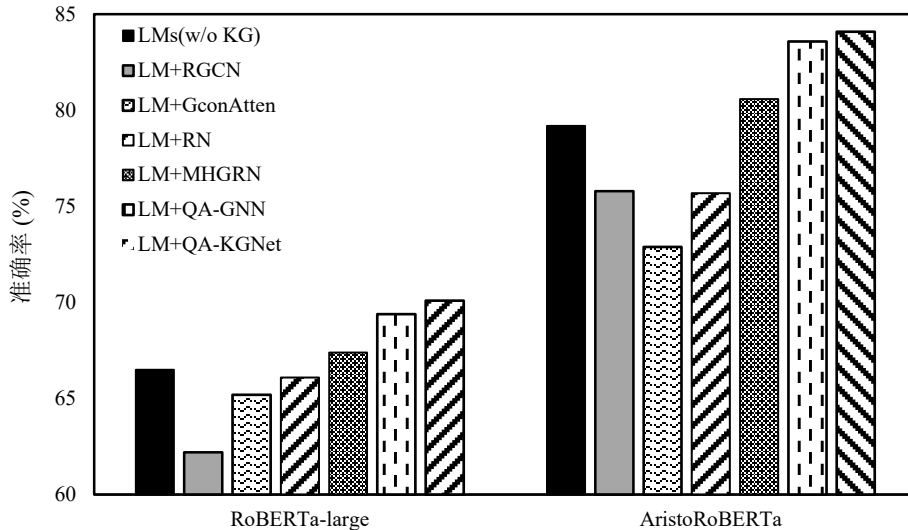


Fig.6 Accuracy comparison of baseline models by using different language models on the OpenBookQA dataset
图 6 使用不同语言模型的基准模型在 OpenBookQA 数据集上的准确率对比

图 7 展示了不同模型在 MedQA-USMLE 医学数据集上的测试集精度. 此处仅使用了语言模型 (BERT^[50], RoBERTa^[44], BioBERT^[51] 和 SapBERT^[43]) 作为基准模型, 目的在于探索所提 QA-KGNet 给语言模型带来的准确率增益. 实验结果表明, SapBERT 的预测准确率高于 BERT、RoBERTa 和 BioBERT, 因为 SapBERT 使用自我监督学习策略, 可以自对齐生物医学实体的表示空间, 进而准确捕捉生物医学领域中的细粒度语义关系, 而其他对比模型对于特定领域所包含的实体及其关系无法准确捕捉. 将所提 QA-KGNet 与 SapBERT 结合, 预测准确率优于其他语言模型基准, 原因在于 QA-KGNet 同时考虑了上下文与知识图谱的联合关系, 并使用 SapBERT 计算节点关联度, 过滤了与主题无关的实体节点, 通过多头图注意力网络进行消息传递与更新. 与主流 LM+KG 方法 QA-GNN 相比, QA-KGNet 的预测准确率略高一些.

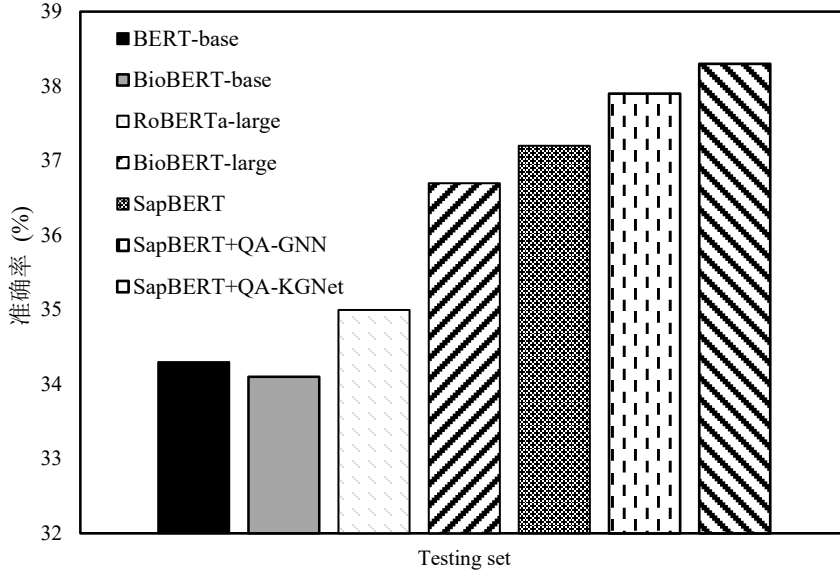


Fig.7 Accuracy comparison of different models on the MedQA-USMLE dataset

图 7 不同模型在 MedQA-USMLE 数据集上的准确率对比

此外,在常识数据集 CommonSenseQA 和 OpenBookQA 上,图 5 和图 6 展示出 QA-KGNet 可以达到 70% 以上的预测准确率.而在专用数据集 MedQA-USMLE 上,图 7 展示出 QA-KGNet 的准确率大约为 38%.造成 QA-KGNet 在不同数据集之间的性能差异主要受预训练语言模型的预测准确率影响.例如图 6 中, RoBERTa-large(LMs(w/o KG)) 的预测准确率远低于 AristoRoBERTa(LMs(w/o KG)) 的预测准确率,使得 QA-KGNet 与 RoBERTa-large 结合的效果弱于与 AristoRoBERTa 结合的效果.如果语言模型精确率较低,将导致计算 QA-KGNet 联合图中上下文节点与各个实体节点之间的相关性分数不准确,进而影响多头图注意力网络中的消息传递,最终影响 QA-KGNet 的预测准确率.在图 7 中, SapBERT 是目前最好的医学语言模型,专用数据集 MedQA-USMLE 中包含大量复杂的医学词汇,即便是 SapBERT 在该数据集上的预测准确率也不高.总之,语言模型影响着 QA-KGNet 的性能,更好的语言模型可以使相关性分数的计算更准确,使得多头图注意力网络能更准确地从不同角度注意节点与节点之间的关系,从而提升整体的问答预测准确率.

4.5 消融测试

为了实验的可靠性,在每一组消融实验中,设置了不同的随机数种子,每组实验重复执行五次,最后取平均准确率.表 2、表 3、图 8 以及图 9 展示了模型中每个组件的消融测试结果.表 2 展示了不同的图连接方式在 CommonsenseQA 常识数据集上的测试集精度. QA-KGNet 应用的一项关键技术是将 T 节点(QA 上下文)与知识图谱中的 QA 主题实体集 $V_{q,a}$ 相连,得到工作图(工作图 G_j).如果不在 QA 上下文和知识图谱的主题节点间建立关联关系,则双方无法相互更新其表示,从而造成性能的下降(76.91%→74.85%).如果将 T 连接到知识图谱中的

所有节点(不仅仅是 QA 实体),则性能略有下降(76.91%→76.44%),因为并不是所有的节点包含有效的信息.

Table 2 Ablation test of different graph connection methods on the CommonsenseQA dataset
表 2 不同的图连接方式在 CommonsenseQA 数据集上的消融测试

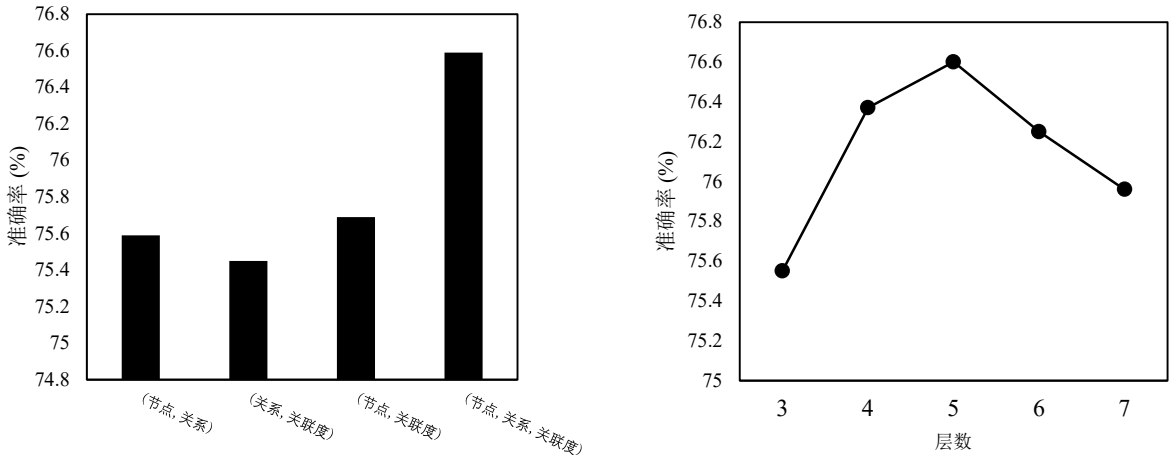
图连接方式	准确率(%)
QA 上下文与知识图谱节点无连接	74.85
QA 上下文与知识图谱节点全连接	76.44
QA 上下文仅与 QA 实体连接	76.91

表 3 展示了知识图谱节点关联度在 CommonsenseQA 数据集上的消融测试.通过计算 QA 上下文节点 T 和子图 $G'_{q,a}$ 中其他节点之间的关联度,使模型性能提升(75.28%→76.69%).本文尝试为每个节点 $v \in V'_{q,a}$ 获取其上下文嵌入 $w_v = fenc([text(T):text(v)])$,然后将 w_v 也作为节点特征.然而,实验结果展示其准确率低于仅使用关联度的模型(76.98%→76.27%).同时使用关联度和上下文嵌入的准确率略低于(或接近于)单独使用关联度的准确率,表明关联度可以获得充足的信息以进行预测.因此,本文所提 QA-KGNet 模型仅使用关联度.

Table 3 Ablation test on the relevance of knowledge graph nodes on the CommonsenseQA dataset
表 3 知识图谱节点关联度在 CommonsenseQA 数据集上的消融测试

关联度	准确率(%)
-	75.28
+上下文嵌入	76.27
+关联度	76.98
+上下文嵌入和关联度	76.65

图 8 展示了在 CommonsenseQA 数据集上不同 GNN 组件的消融测试.图 8(a)通过消融注意力和消息中的节点类型(图中简称节点)、关系和关联度来分别验证模型准确率.结果表明,这些特征均对模型性能的提高有益.此外,图 8(b)展示了不同的 GNN 层数对模型性能的影响, $L=5$ 时的效果最好.图注意力网络的层数过多或过少都会带来一定程度的性能损失.如果层数过少,图注意力网络无法充分学习节点及其关系,导致节点与边无法获得精确的表示,即欠拟合.当层数过多时,图注意力网络会过拟合特定训练集,无法泛化到测试集或其他数据集.



(a) 节点类型、关系与关联度对模型准确率的影响

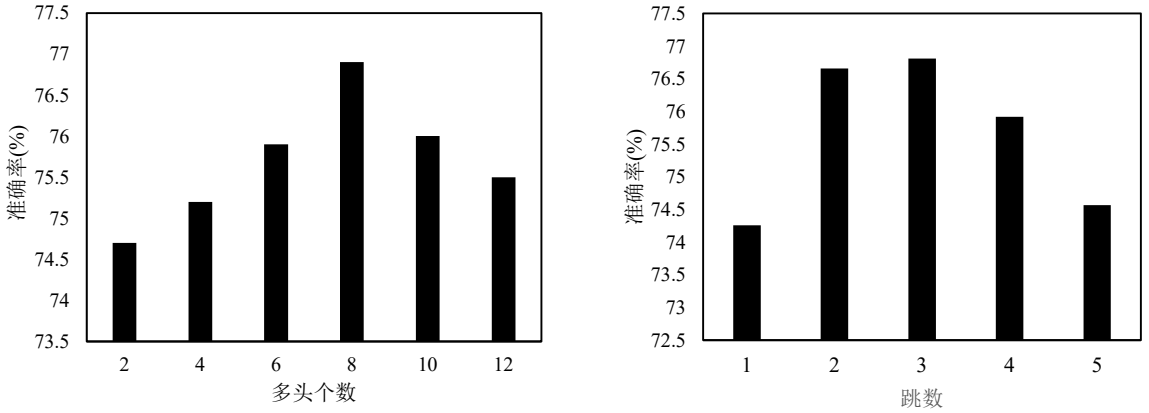
(b) 图注意力网络层数对模型准确率的影响

Fig.8 Ablation test of different GNN components on the CommonsenseQA dataset

图 8 不同的 GNN 组件在 CommonsenseQA 数据集上的消融测试

图 9 给出在 CommonsenseQA 数据集上不同的多头个数 k 与跳数 g 的消融测试结果.图 9(a)设置不同的多头个数 k ,即 $\{2,4,6,8,10,12\}$.结果表明,当 $k=8$ 时,准确率最高.这是因为当 k 过少时,导致图注意力网络无法从多个角度充分获取节点和边的特征,进而使准确率降低.当 k 过多时,图注意力网络会过度关注某些重要特征,从而产生偏差.此外,图 9(b)展示了不同的跳数 g 对模型性能的影响,当 $g=3$ 时的效果最好.如果跳数 g 过少,可能会丢失部分重要的节点,导致推理准确率的损失.如果跳数 g 过多,虽然可以考虑更多的重要节点,但会增加许多噪音节

点,从而影响准确率.值得注意的是,当 $g=2$ 时,模型准确率接近于最优值($g=3$),这是因为所提方法兼顾了上下文节点与知识图谱节点,通过多头图注意力网络进行消息传递,获得精确的更新信息,故而使用更少的跳数也能达到较优的准确率.



(a) 多头个数 k 对模型准确率的影响

(b) 跳数 g 对模型准确率的影响

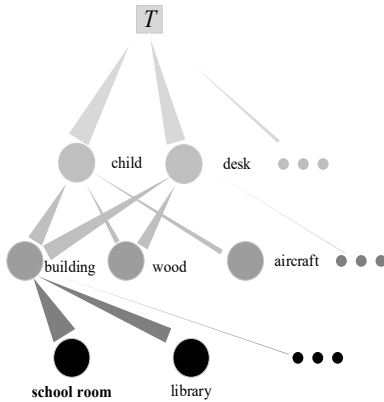
Fig.9 Impact of the number k of multi-heads and the number g of hops on the accuracy of the model

图 9 多头个数 k 与跳数 g 对模型准确率的影响

4.6 模型可解释性研究

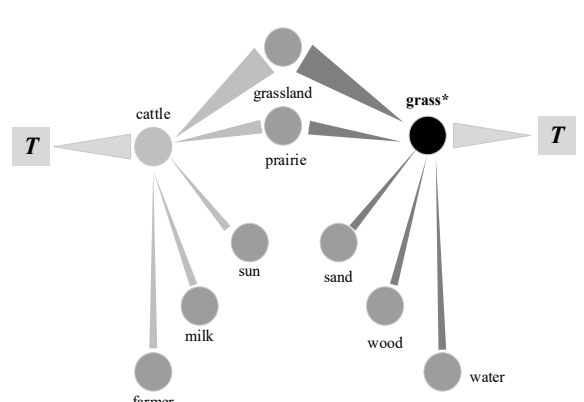
本节通过分析多头图注意力网络的节点间注意力权重来解释 QA-KGNet 的推理过程.图 10 展示了两个案例,图中使用三角形的宽度来表示注意力权重的高低(注意力权重高被称为强关注,反之称为弱关注).在图 10(a)中,工作图中使用最佳优先搜索(BFS, Best First Search)来搜索从 QA 上下文节点 T 到问题实体节点(例如“孩子(child)”和“课桌(desk)”),再到其他实体节点(例如“建筑物(building)”、“木头(wood)”和“飞机(aircraft)”)或答案实体节点(例如“办公楼(school room)”和“教堂(library)”)的注意力权重.可视化结果表明 QA 上下文关注知识图谱中的“child”和“desk”节点,而“child”和“desk”节点强关注于“building”节点,“building”节点与“school room”节点是“isA”关系,则“school room”是答案.在图 10(b)中,从两个方向使用 BFS 来追踪注意力权重.可视化结果展示了知识图谱中的草原概念,即“grassland”和“prairie”这两个概念在 QA 上下文中不一定会被提及,但在问题实体“牛(cattle)”和答案实体“草(grass)”之间产生关联,以便于找到合适的答案.

Q: Where does a *child* likely sit at a *desk*?
A: A. office building B. library C. school room*



(a) 案例 1

Q: *Cattles* live in what sort of environment?
A: A. grass* B. zoo C. city



(b) 案例 2

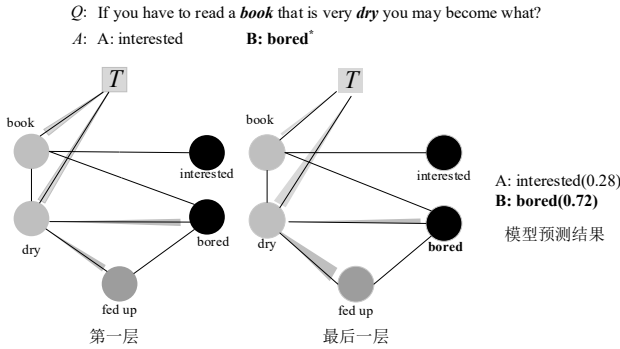
Fig.10 Visualization interpretation of the reasoning process in QA-KGNet

图 10 QA-KGNet 推理过程的可视化解释

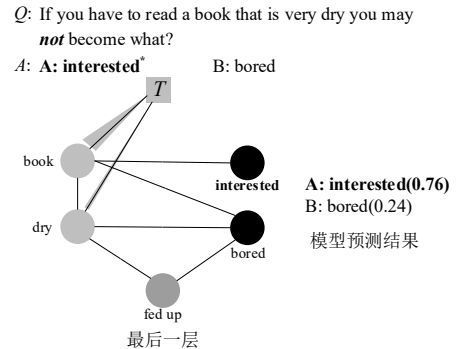
4.8 结构化推理

结构化推理对于做出合理的预测至关重要.图 10 给出了四个案例,用于分析 QA-KGNet 的结构化推理行为,括号内的小数表示预测答案选项的概率,实线表示节点之间的关系,三角形的宽度用于表示注意力权重的高低(注意力权重高被称为强关注,反之称为弱关注).规定推理路径必须遵从 $T \rightarrow$ 问题实体(即“book”或“dry”) $\rightarrow$ 答案选项实体(即“interested”或“bored”)或者遵从 $T \rightarrow$ 问题实体 $\rightarrow$ 其他实体(即“fed up”) $\rightarrow$ 答案选项实体.图 11(a)的问题中不包含任何否定词,可以观察到,在 QA-KGNet 的第一层中,QA 上下文节点 T 对问题实体“book”和“dry”的注意力权重几乎一致.在经过工作图的多轮消息传递后,在最后一层,QA 上下文节点 T 强关注于实体“dry”,而弱关注于另一个实体“book”,此时推理路径为 $T \rightarrow$ “dry”.再者,观察到与“dry”节点相连节点的注意力权重高的为“fed up”节点,推理路径变为 $T \rightarrow$ “dry” $\rightarrow$ “fed up”.最后,通过关系找到正确预测答案“B: bored”.

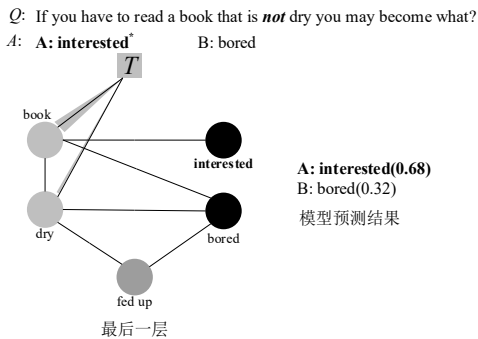
在原始问题(图 11(a))的基础上,分别否定掉其中一个动词(图 11(b)与图 11(c))或双重否定(图 11(d)).在图 11(b)和图 11(c)中,QA 上下文节点 T 强关注于实体“book”,同时弱关注于(否定)实体“dry”,根据上述推理路径的规定,预测了正确答案“A: interested”.在图 11(d)中,可以观察到在双重否定掉所有的动词后 QA 上下文节点 T 对两个问题实体(即“book”和“dry”)均不再强关注,此时模型仍旧预测结果为答案“A: interested”,但此时的正确答案应该是“B: bored”.原因在于模型在处理双重否定时,无法判断哪些实体是更应该被关注,而错失了部分特征信息,导致预测出错.



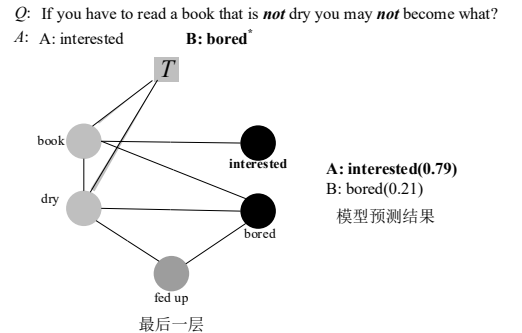
(a) 案例 1



(b) 案例 2



(c) 案例 3



(d) 案例 4

Fig.11 Analysis of the structured reasoning process in QA-KGNet

图 11 QA-KGNet 结构化推理过程分析

5 结论

本文提出一种新型语言模型驱动的知识图谱问答模型 QA-KGNet,其根据给定的 QA 上下文计算与知识图谱节点的相关性,能够对 QA 上下文和知识图谱进行联合推理.通过工作图连接 QA 上下文和知识图谱,并通过多头图注意力网络进行消息传递,同时更新节点表示.实验结果展示了所提 QA-KGNet 在问答任务上的性能提升,以及执行结构化推理的能力.未来的工作将关注于语言模型与知识图谱驱动的双重否定语义的问答任务.

致谢

感谢所有参与本课题研究和本文工作没有在文中署名的课题组成员及对本项目提供技术支持的专家和学者;感谢所有评阅本文的匿名评审人及对本文提出的宝贵修改意见.

References:

- [1] Xiong F, Wang HZ. Mining simple path traversal patterns in knowledge graph. *Journal of Web Engineering*, 2022,21(2): 258-271. [doi: 10.13052/jwe1540-9589.2128]
- [2] Jamehshourani M, Fatemi A, Nematbakhsh MA. Implicit relation-based question answering to answer simple questions over DBpedia. *Turkish Journal of Electrical Engineering and Computer Sciences*, 2020,28(3):1474-1490. [doi: 10.3906/elk-1908-102]
- [3] Jiang K, Wu DK, Jiang H. FreebaseQA: a new factoid QA data set matching trivia-style question-answer pairs with Freebase. In: *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. New York: ACM, 2019. 318-323. [doi: 10.18653/v1/n19-1028]
- [4] Lu XY, Wang LF, Jiang ZJ, He SC, Liu SZ. MMKRL: A robust embedding approach for multi-modal knowledge graph representation learning. *Applied Intelligence*, 2022,52(7): 7480-7497. [doi: 10.1007/s10489-021-02693-9]
- [5] Menin A, Torchelsen RP, Nedel LP. The effects of VR in training simulators: exploring perception and knowledge gain. *Computers & Graphics*, 2022,102(5): 402-412. [doi: 10.1016/j.cag.2021.09.015]
- [6] Li JZ, Hou L. Reviews on knowledge graph research. *Journal of Shanxi University (Nat. Sci. Ed.)*, 2017,40(03): 454-459. [doi: 10.13451/j.cnki.shanxi.univ(nat.sci.).2017.03.008]
- [7] Nayyeri M, Cil GM, Vahdati S, Osborne F, Rahman M, Angioni S, Salatino AA, Recupero DR, Vassilyeva R, Motta E, Lehmann J. Trans4E: Link prediction on scholarly knowledge graphs. *Neurocomputing*, 2021,461(3): 530-542. [doi: 10.1016/j.neucom.2021.02.100]
- [8] Dessì D, Osborne F, Recupero DR, Buscaldi D, Motta E. Generating knowledge graphs by employing Natural Language Processing and Machine Learning techniques within the scholarly domain. *Future Generation Computer Systems*, 2021,116(7): 253-264.[doi: 10.1016/j.future.2020.10.026]
- [9] Lin QK, Zhu YF, Lu H, Shi K, Niu ZD. Improving university faculty evaluations via multi-view knowledge graph. *Future Generation Computer Systems*, 2021,117(5): 181-192. [doi: 10.1016/j.future.2020.11.021]
- [10] Yue Q, Li X, Li D. Chinese relation extraction on forestry knowledge graph construction. *Computer Systems: Science & Engineering*, 2021,37(3): 423-442[doi: 10.32604/csse.2021.014448]
- [11] Yu CM, Wang F, Liu YH, An L. Research on knowledge graph alignment model based on deep learning. *Expert Systems with Applications*, 2021,186(4): 115-129. [doi: 10.1016/j.eswa.2021.115768]
- [12] Petroni F, Rocktaschel T, Riedel S, Lewis PSH, Bakhtin A, Wu YX, Miller AH. Language models as knowledge bases?. In: *Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing*. Washington: ACL, 2019. 2463-2473. [doi: 10.18653/v1/D19-1250]
- [13] Bosselut A, Rashkin H, Sap M, Malaviya C, Celikyilmaz A, Choi Y. COMET: commonsense transformers for automatic knowledge graph construction. In: *Proc. of the 57th Conference of the Association for Computational Linguistics*. Washington: ACL, 2019. 4762-4779. [doi: 10.18653/v1/p19-1470]

- [14] Speer R, Chin J, Havasi C. ConceptNet 5.5: an open multilingual graph of general knowledge. In: Proc. of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2017. 4444-4451.
- [15] Huang SS, Zhu KQ, Liao QZ, Shen LB, Zhao YG. Enhanced story representation by ConceptNet for predicting story endings. In: Proc. of the 29th ACM International Conference on Information and Knowledge Management. New York: ACM, 2020. 3277-3280. [doi: 10.1145/3340531.3417466]
- [16] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou YQ, Li W, Liu PJ. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 2020,21(2):1-14.
- [17] Liao WX, Zeng B, Yin XW, Wei PF. An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. *Applied Intelligence*, 2021,51(6):3522-3533.
- [18] Lin YB, Chen XY, Chen JM, Ren X. KagNet: knowledge-aware graph networks for commonsense reasoning. In: Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing. New York: ACM, 2019. 2829-2839. [doi: 10.18653/v1/D19-1282]
- [19] Bao JW, Duan N, Yan Z, Zhou M, Zhao TJ. Constraint-based question answering with knowledge graph. In: Proc. of the 26th International Conference on Computational Linguistics. Washington: ACL, 2016. 2503-2514.
- [20] Sun HT, Dhingra B, Zaheer M, Mazaitis K, Salakhutdinov R, Cohen WW. Open domain question answering using early fusion of knowledge bases and text. In: Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing. New York: ACM, 2018. 4231-4242. [doi: 10.18653/v1/d18-1455]
- [21] Wang XY, Kapanipathi P, Musa R, Yu M, Talamadupula K, Abdelaziz I, Chang M, Fokoue A, Makni B, Mattei N, Witbrock M. Improving natural language inference using external knowledge in the science questions domain. In: Proc. of the Thirty-Third AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2019. 7208-7215. [doi: 10.1609/aaai.v33i01.33017208]
- [22] Feng YL, Chen XY, Lin YB, Wang PF, Yan J, Ren X. Scalable multi-hop relational reasoning for knowledge-aware question answering. In: Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing. New York: ACM, 2020. 1295-1306. [doi: 10.18653/v1/2020.emnlp-main.99]
- [23] Yasunaga M, Ren HY, Bosselut A, Liang P, Leskovec J. QA-GNN: reasoning with language models and knowledge graphs for question answering. In: Proc. of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Washington: ACL, 2021. 535-546. [doi: 10.18653/v1/2021.naacl-main.45]
- [24] Pan XM, Sun K, Yu D, Chen JS, Ji H, Cardie C, Yu D. Improving question answering with external knowledge. In: Proc. of the 2nd Workshop on Machine Reading for Question Answering. Washington: ACL, 2019. 27-37. [doi: 10.18653/v1/D19-5804]
- [25] Mihaylov T, Frank A. Knowledgeable reader: enhancing cloze-style reading comprehension with external commonsense knowledge. In: Proc. of the 56th Annual Meeting of the Association for Computational Linguistics. Washington: ACL, 2018. 821-832. [doi: 10.18653/v1/P18-1076]
- [26] Bosselut A, Bras RL, Choi YJ. Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering. In: Proc. of Thirty-Fifth AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2021. 4923-4931.
- [27] Paul D, Frank A. Ranking and selecting multi-hop knowledge paths to better predict human needs. In: Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Washington: ACL, 2019. 3671-3681. [doi: 10.18653/v1/n19-1368]
- [28] Yang A, Wang Q, Liu J, Liu K, Lyu YJ, Wu H, She QQ, Li SJ. Enhancing pre-trained language representations with rich knowledge for machine reading comprehension. In: Proc. of the 57th Conference of the Association for Computational Linguistics. Washington: ACL, 2019. 2346-2357. [doi: 10.18653/v1/p19-1226]
- [29] Yih WT, Richardson M, Meek C, Chang MW, Suh J. The value of semantic parse labeling for knowledge base question answering. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics. Washington: ACL, 2019. 546-555. [doi: 10.18653/v1/p16-2033]

- [30] Bevilacqua M, Navigli R. Quasi bidirectional encoder representations from transformers for word sense disambiguation. In: Proc. of the International Conference on Recent Advances in Natural Language Processing. Varna: INCOMA Ltd, 2019. 122-131. [doi: 10.26615/978-954-452-056-4_015]
- [31] Sun HT, Weiss TB, Cohen WW. PullNet: open domain question answering with iterative retrieval on knowledge bases and text. In: Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing. Washington: ACL, 2019. 2380-2390. [doi: 10.18653/v1/D19-1242]
- [32] Xiong WH, Yu M, Chang SY, Guo XX, Wang YW. Improving question answering over incomplete kbs with knowledge-aware reader. In: Proc. of the 57th Conference of the Association for Computational Linguistics. Washington: ACL, 2019. 4258-4264. [doi: 10.18653/v1/p19-1417]
- [33] Zhang YH, Qi P, Manning CD. Graph convolution over pruned dependency trees improves relation extraction. In: Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing. Washington: ACL, 2018. 2205-2215. [doi: 10.18653/v1/d18-1244]
- [34] Yasunaga M, Liang P. Graph-based, self-supervised program repair from diagnostic feedback. In: Proc. of the 37th International Conference on Machine Learning. Berlin: PMLR, 2020. 10799-10808.
- [35] Lee J, Lee I, Kang J. Self-Attention graph pooling. In: Proc. of the 36th International Conference on Machine Learning. California: PMLR, 2019. 3734-3743.
- [36] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. In: Proc. of Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017. New York: ACM, 2017. 5998-6008.
- [37] Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. In: Proc. of the 6th International Conference on Learning Representations. New York: ACM, 2018. 1-12.
- [38] Talmor A, Herzig J, Lourie N, Berant J. CommonsenseQA: a question answering challenge targeting commonsense knowledge. In: Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Washington: ACL, 2019. 4149-4158. [doi: 10.18653/v1/n19-1421]
- [39] Mihaylov T, Clark P, Khot T, Sabharwal A. Can a suit of armor conduct electricity? a new dataset for open book question answering. In: Proc. of the 2018 Conference on Empirical Methods in Natural Language Processing. Washington: ACL, 2018. 2381-2391. [doi: 10.18653/v1/d18-1260]
- [40] Jin D, Pan B, Oufattole N, Weng WH, Fang HY, Szolovits P. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *Applied Sciences*, 2021,11(4):1-17. [doi: 10.3390/app11146421]
- [41] Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 2004,32(4):267-270. [doi: 10.1093/nar/gkh061]
- [42] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu YF, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 2018,46(8):1074-1082. [doi: 10.1093/nar/gkx1037]
- [43] Liu FY, Shareghi E, Meng ZQ, Basaldella M, Collier N. Self-alignment pretraining for biomedical entity representations. In: Proc. of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Washington: ACL, 2021. 4228-4238. [doi: 10.18653/v1/2021.naacl-main.334]
- [44] Liu YH, Ott M, Goyal M, Du JF, Joshi M, Chen DQ, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. RoBERTa: a robustly optimized BERT pretraining approach. *arXiv Preprint arXiv: 1907.11692*, 2019.
- [45] Clark P, Etzioni O, Khashabi D, Khot T, Mishra BD, Richardson K, Sabharwal A, Schoenick C, Tafjord O, Tandon N, Bhakthavatsalam S, Groeneveld D, Guerin M, Schmitz M. From 'f' to 'a' on the NY regent science exams: an overview of the aristo project. *arXiv Preprint arXiv: 1909.01958*, 2019.

- [46] Santoro A, Raposo D, Barrett DGT, Malinowski M, Pascanu R, Battaglia PW, Lillicrap T. A simple neural network module for relational reasoning. In: Proc. of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems. Washington: ACL, 2017. 4967-4976.
- [47] Schlichtkrull MS, Kipf TN, Bloem P, Berg RVD, Titov I, Welling M. Modeling relational data with graph convolutional networks. In: Proc. of the Semantic Web 15th International Conference. Berlin: Springer, 2018. 593-607. [doi: 10.1007/978-3-319-93417-4_38]
- [48] Wang XY, Kapanipathi P, Musa R, Yu M, Talamadupula K, Abdelaziz I, Chang M, Fokoue A, Makni B, Mattei N, Witbrock M. Improving natural language inference using external knowledge in the science questions domain. In: Proc. of the Thirty-Third AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2019. 7208-7215. [doi: 10.1609/aaai.v33i01.33017208]
- [49] Kingma DP, Ba J. Adam: a method for stochastic optimization. In: Proc. of the 3rd International Conference on Learning Representations. Washington: ACL, 2015. 1-15.
- [50] Devlin J, Chang MW, Lee KT, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Washington: ACL, 2019. 4171-4186. [doi: 10.18653/v1/n19-1423]
- [51] Lee J, Yoon W, Kim SD, Kim D, Kim SY, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 2020,36(4):1234-1240. [doi: 10.1093/bioinformatics/btz682]

附中文参考文献:

- [6] 李涓子, 侯磊. 知识图谱研究综述. *山西大学学报(自然科学版)*, 2017,40(3):454-459. [doi:10.13451/j.cnki.shanxi.univ(nat.sci.).2017.03.008]