

基于图文知识融合的常识问答模型

蔡瑞祥, 赵书良, 何家瑶

引用本文

蔡瑞祥, 赵书良, 何家瑶. [基于图文知识融合的常识问答模型](#)[J]. 计算机科学, 2025, 52(11): 237-244.
CAI Ruixiang, ZHAO Shuliang, HE Jiayao. [Commonsense Question Answering Model Based on Graph-Text Integrating](#) [J]. Computer Science, 2025, 52(11): 237-244.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

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

[基于变体注意力的关系与属性感知实体对齐](#)

Relationship and Attribute Aware Entity Alignment Based on Variant-attention
计算机科学, 2025, 52(11): 230-236. <https://doi.org/10.11896/jsjcx.240800140>

[基于审判逻辑的裁判文书生成方法](#)

Method for Generating Judgment Documents Based on Trial Logic
计算机科学, 2025, 52(11): 223-229. <https://doi.org/10.11896/jsjcx.250500054>

[Instruct-Malware:基于控制流图的大型语言模型恶意软件分析](#)

Instruct-Malware:Control Flow Graph Based Large Language Model Analysis of Malware
计算机科学, 2025, 52(11): 40-48. <https://doi.org/10.11896/jsjcx.241100118>

[DF-RAG:基于查询重写和知识选择的检索增强生成方法](#)

DF-RAG:A Retrieval-augmented Generation Method Based on Query Rewriting and Knowledge Selection
计算机科学, 2025, 52(11): 30-39. <https://doi.org/10.11896/jsjcx.241000117>

[基于大语言模型增强的零样本知识抽取方法](#)

Zero-shot Knowledge Extraction Method Based on Large Language Model Enhanced
计算机科学, 2025, 52(11): 22-29. <https://doi.org/10.11896/jsjcx.241000049>

基于图文知识融合的常识问答模型

蔡瑞祥 赵书良 何家瑶

河北师范大学计算机与网络空间安全学院 石家庄 050024

供应链大数据分析 with 数据安全河北省工程研究中心 石家庄 050024

河北省网络与信息安全重点实验室 石家庄 050024

(ssznbdcx@163.com)

摘要 知识图谱在常识性问答中展现出了显著的成效。现有的方法通常利用问题中的实体来从知识图谱(KG)中检索局部子图,然后对其使用图神经网络(GNN)进行编码,随后将 GNN 编码的结果与语言模型(LMs)相结合,以进行答案的推理和问题的回答。然而,使用 GNN 和 LMs 的常识问答系统存在两个挑战:1)如何高效地提取知识图谱中的子图,对其知识与结构信息进行有效的表示和利用;2)如何实现问题上下文与子图知识的深度融合与联合推理。为此,提出了一个基于图文知识融合的常识问答模型(Graph-Text Integrating for Commonsense Question Answering,GTICQA)。该模型首先通过外部词典过滤精炼出关键实体,实现对知识子图的剪枝。然后,使用 LM 对问题上下文进行编码,用 GNN 编码器对精炼后的知识子图进行编码。同时,在子图编码的过程中,引入了一个新的 k 稀疏注意力机制增强对子图全局特征的提取并抑制噪声。最后,使用一种包括细粒度双模态交互融合层和均值交互融合层的知识融合方法,对两种知识表示进行深度融合与动态更新。在 Commonsense-QA, OpenBookQA 和 MedQA-USMLE 这 3 个数据集上对 GTICQA 进行了评估,其分别以 79.12%,72.20%和 39.40%的准确率超越了现有最佳方法,表明了模型在处理常识性问答上的优势。

关键词: 常识性问答;多项选择问答;知识融合;知识图谱;语言模型

中图分类号 TP391

Commonsense Question Answering Model Based on Graph-Text Integrating

CAI Ruixiang, ZHAO Shuliang and HE Jiayao

College of Computer and Cyber Security, Hebei Normal University, Shijiazhuang 050024, China

Hebei Provincial Engineering Research Center for Supply Chain Big Data Analytics & Data Security, Shijiazhuang 050024, China

Hebei Provincial Key Laboratory of Cyber and Information Security, Shijiazhuang 050024, China

Abstract Knowledge graphs have demonstrated significant effectiveness in commonsense question answering. Existing methods typically utilize entities from the question to retrieve local subgraphs from the knowledge graph(KG), which are then encoded using graph neural networks(GNN). Subsequently, the GNN-encoded results are combined with language models(LMs) to infer answers and answer the questions. However, commonsense question answering systems using GNNs and LMs face two challenges: 1) how to efficiently extract subgraphs from the knowledge graph, effectively represent and utilize their knowledge and structural information; 2) how to achieve deep integration and joint reasoning of the question context and subgraph knowledge. This paper proposes a graph-text integrating model for commonsense question answering(Graph-Text Integrating for Commonsense Question Answering, GTICQA). The model initially refines key entities by filtering through an external dictionary, achieving pruning of the knowledge subgraph, and then separately encodes the question context using an LM and the refined knowledge subgraph using a GNN encoder. Additionally, during the subgraph encoding process, a novel k-sparse attention mechanism is introduced to enhance the extraction of global features from the subgraph and suppress noise. Finally, a knowledge fusion method that includes fine-grained bimodal interaction fusion layers and mean interaction fusion layers is used to deeply integrate and dy-

到稿日期:2024-09-12 返修日期:2024-12-09

基金项目:国家社会科学基金重大项目(18ZDA200);河北省省级科技计划(20370301D, 22567606H);河北省引进留学人员资助项目(C20230339);河北师范大学专项科技基金(L2023T03)

This work was supported by the National Social Science Foundation of China(18ZDA200), S&T Program of Hebei(20370301D, 22567606H), Introducing Talents of Studying Overseas Fund of Hebei (C20230339) and Special Science and Technology Fund of Hebei Normal University (L2023T03).

通信作者:赵书良(zhaoshuliang@sina.com)

natically update the two knowledge representations. The GTICQA model is evaluated on three datasets: CommonsenseQA, OpenBookQA, and MedQA-USMLE, achieving accuracy rates of 79.12%, 72.20%, and 39.40%, respectively, surpassing the current best methods, demonstrating the model's advantage in handling commonsense question answering.

Keywords Commonsense QA, Multiple choice QA, Knowledge integration, Knowledge graph, Language model

1 引言

经过大规模文本语料库的训练,预训练模型无需依赖人工标注的标签,仅凭自监督学习机制,便能够学习到丰富且通用的文本表示。这一特性使得预训练模型在自然语言处理(NLP)领域取得了显著的成就。它们不仅能够深入理解自然语言文本,还能生成高质量的文本内容,因此在各种NLP任务中均表现出色。然而由于过于依赖可能已经过时或不够准确的训练数据,当处理实时任务或特定领域知识的问题时,可能会暴露出局限性。这导致大语言模型(LLM)在知识感知问答任务上^[1],即使精心设计了提示策略,也不如经过微调的小型模型^[2]。另外,LLM在知识密集型问答领域展现出的幻觉问题,成为了其性能提升的一大障碍。尤其在需要高精度和可解释性的知识密集型问答任务中,LLM会生成看似合理但实际上并不准确或完全虚构的回答^[3]。

为解决这些问题,研究者探索使用知识图谱(KG)来提升语言模型^[4-6]。知识图谱采用图结构的形式,精确地描绘了现实世界中的实体、概念及其之间的联系。这种结构不仅包含了丰富的知识内容,而且因其高度结构化的特性,极大地促进了基于逻辑的结构化推理过程^[7]。知识图谱的优势在于,它能够清晰地展示知识之间的层次和关联,使得复杂的推理任务变得更加直观和高效。现有研究采用的大多数方法是:首先提取与关键实体相关的子图,再使用图神经网络(GNN)对这些子图所蕴含的结构化知识进行学习。其中关键实体的选择直接决定着子图的质量,进而影响最终的答案预测。此外,这些方法仅仅在最终预测阶段中将其与LM集成,这种浅层和非交互的融合方式使两种模态之间的信息交互十分有限。如何以真正统一的方式有效地融合KG和LM两种模态信息,进而进行推理仍是一个值得研究的问题。

为应对这些挑战,本文提出了一个基于图文知识融合的常识问答模型——GTICQA(Graph-Text Integrating for Commonsense Question Answering),它在其架构中的多个层进行知识的融合,提高了两种模态信息之间的交流,实现了更好的信息融合。本模型首先从问题上下文中提取实体,并结合外部词典进行过滤,得到关键实体。这一过程有助于减少在构建知识子图时出现的噪声实体,实现子图的有效剪枝,提高子图的质量;使用GNN编码器对知识子图进行编码,并设计了一个轻量化注意力机制模块以增强对重要特征的提取并抑制了噪声;使用LM编码器对问题上下文进行编码,为两种模态添加交互信息,再将两种模态的表示使用一个细粒度注意力信息融合模块建立不同模态之间的联系,实现两种模态信息的有效融合。

本文的主要贡献总结为以下4点:

1)通过一种融合外部词典的知识子图精炼剪枝技术,有效剔除了噪声实体,确保了子图的质量。

2)提出了一种kFA的注意力机制模块,在保留GNN局部偏差的基础上,通过提升全局特征提取能力和噪声抑制能力,显著优化了性能。这种设计使GNN不仅能够深入挖掘图结构的局部特征,还能有效整合全局信息,降低了噪声的影响。

3)提出了一种支持KG和LM两种模态的信息融合方案,通过细粒度双模态交互融合层和均值交互融合层在文本和知识子图之间进行双向信息交换和融合,减小了两种模态信息之间的差距。

4)在3个基准数据集CommonsenseQA,OpenbookQA和MedQA-USMLE上的实验结果表明,本文提出的GTICQA模型优于现有方法,通过一系列消融实验和案例研究,验证了GTICQA模型中各模块的有效性。

2 相关工作

在常识性问答任务中,通常需要借助广泛且多样化的常识背景知识来解答问题。预训练的LM和KG是获取这些背景知识的常用来源,研究者通常采用LM+KG的框架来解决常识性问答任务。目前结合LM和KG知识主要有两种方式:

1)将KG中的知识直接注入LM中,获取一个含有特定知识的预训练LM,来解决常识性问答任务。Yu等^[8]为了将KG中的信息融入LM中,提出了联合预训练JAKET框架,该框架中LM输出上下文感知的嵌入,初始化知识图谱中实体和关系的嵌入,使用图注意力网络(GAT)为LM输出基于知识的实体嵌入,以此来学习KG中的实体与关系信息。Liu等^[9]提出了TSTR方法,该方法采用双流BERT嵌入,将三元组分为头部实体及关系和尾部实体两个部分,通过并行BERT模型进行语义嵌入,构建实体候选集,学习三元组的结构信息。Meng等^[10]提出了MoP方法,该方法将KG划分为若干不相交的子图,对每一个子图使用一个单独的适配器进行学习,最终对所有适配器获得的知识进行融合,用于LM的下游任务。这些方法将KG中蕴含的知识注入LM中,并充分地利用了预训练LM中的隐式知识,但并没有发挥KG作为图结构知识的优势,限制了模型结构化推理的能力。

2)设计GNN对KG进行建模,结合LM来解决常识性问答任务。Lin等^[11]提出了KagNet框架,它首先将问题与答案的语义表示从语义空间映射到一个结构化的“模式图”(即KG子图)中,再利用基于路径的分层注意力机制增强了推理的透明性和可解释性。Feng等^[4]提出了MHGRN模型,它首先利用结构化关系注意力机制捕捉KG子图中实体间复杂的联系,再使用MHGRN对子图进行建模,有效地利用这些联系进行推理。Yasunaga等^[5]提出了QA-GNN模型,将问题上下文作为一个节点连接到子图,融合两种模态信息形成联合图,利用消息传递联合更新LM的表示和对子图的GNN

表示。这些方法利用了 KG 作为结构化知识的优点,但它们忽视了 KG 与文本之间的模态差距,这种差异导致了两者在数据分布上的不匹配,进而在组合跨模态表示时引发不一致性,阻碍了跨模态信息之间的有效对齐和交互,最终限制了模型在融合文本和 KG 信息时的性能表现。为了进一步探索两种模态的融合方式,Sun 等^[6]提出了 JointLK 模型,该模型设计了一个联合推理模块,在问题的每个 token 与 KG 节点之间使用了细粒度的双向注意力机制,不仅促进了两种模态之间的信息交互,还对两种模态的信息进行了动态更新,弥补了两者之间的差距。Zhang 等^[12]提出了 GreaseLM 模型,该模型在 LM 和 GNN 编码后,将各模态的信息通过双向传输的方式传输到另一方,让两种模态的信息都可以被另一种模态语境化。Dong 等^[13]提出了 HamQA 框架,设计了一个上下文感知图注意力网络来获取问题上上下文信息,并在双曲空间中通过最小化双曲测地线距离,保留 KG 中的结构信息,再对两种模态进行联合训练。类似的工作^[14-16]分别利用 LM 和 GNN 对问题上上下文和 KG 进行编码,并在最后一层或中间层进行自然语言和 KG 结构化知识的浅层融合,来实现对答案的预测。

先前的研究工作尽管已经结合了 LM 和 KG 来解决常识

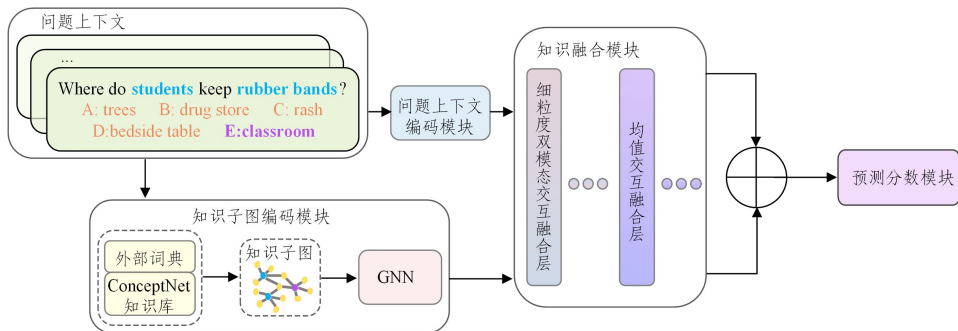


图1 基于图文知识融合的常识问答模型框架

Fig. 1 Framework of commonsense question answering model based on graph-text integrating

3.1 任务定义

在常识问答(CQA)任务中,问题与答案的形式如图1 QA context 部分所示,给定一个问题 q 和一组答案选项 $\{a_1, a_2, \dots, a_n\}$,需要评估各候选选项 a 与给定问题 q 之间的合理性得分。通过这一量化过程,本文选择出与问题含义最为契合的选项,即选出合理性得分最高的选项。但该任务的问题和选项内容在通常情况下,并不包含或仅含有少量的背景知识,因此采用从外部 KG 中检索相关知识的方法,结合问题与选项内容来进行最终答案的推理预测。

3.2 问题上上下文编码

首先,将问题 q 和候选答案 a 使用分隔符连接起来,形成一个新的序列 w 。接着,用 $\{w_1, w_2, \dots, w_n\}$ 表示问题上上下文的 token 集合,将这个集合作为预训练语言模型的输入,获取问题上上下文的文本向量序列 $\{s_1, s_2, \dots, s_n\}$:

$$w = (q \oplus a) \quad (1)$$

$$\{s_1, s_2, \dots, s_n\} = LM(w_1, w_2, \dots, w_n) \quad (2)$$

其中, n 表示 token 的总数, LM 表示本文使用的预训练语言模型。

3.3 知识子图编码

为更好地利用图结构的知识信息,本节介绍了知识子图

性问答任务,但仍没有充分发挥 KG 的结构化知识的潜力,并且两种模态的交互与融合也较为浅层。本文同样采用 LM+KG 框架,结合 GNN 解决常识性问答任务。但本文专注于优化知识子图的提取过程以及深化两种模态知识之间的交互与融合。在知识子图的提取阶段,本文强化关键实体识别并剔除噪声节点,构建精炼且价值高的子图。同时,使用一种新的融合模块,实现文本模态与 KG 模态之间细粒度、深层次的交互与融合。

3 本文方法

针对常识问答研究,本文提出了一个基于图文知识融合的常识问答模型——GTICQA 模型。如图1所示,GTICQA 主要由4个部分组成:问题上上下文编码模块、知识子图编码模块、知识融合模块和答案预测模块。本章安排如下:3.1节介绍了常识问答任务定义,阐述了任务的目标;3.2节使用预训练语言模型对问题上上下文进行文本编码处理;3.3节详细介绍了知识子图的生成、剪枝和编码过程,并引入 kFA 增强特征抑制噪声;3.4节探讨了如何对两种模态的信息进行深层次的融合和更新;3.5节描述了最终的答案预测。

的生成、剪枝、编码的过程以及引入的 k-Focus Attention 机制,增强对图相关特征的提取能力。首先,采用 Keybert 工具提取问题上上下文中的关键实体,并借助外部词典进行过滤,从而生成一个初始的实体集合。利用 ConceptNet,从中提取与这些实体相关的子图,并通过剪枝操作确保子图的质量和紧凑性。在子图生成和剪枝阶段完成后,引入 GNN 编码器对子图进行建模,生成每个节点的嵌入表示,从而捕捉子图的结构信息和语义关联。为进一步增强模型对相关特征的提取能力,引入 k-Focus Attention 机制,选择与关键节点最相关的 k 个节点进行注意力计算,过滤噪声信息,加速模型的训练。

3.3.1 子图的生成和剪枝

知识子图的质量决定了背景知识的质量,进而影响最终答案推理的准确性。在 KG 中,知识子图的提取依赖于关键实体的选择。因此,对关键实体的识别和利用,对于提取相关且高质量的知识子图十分重要。在本研究中,利用一种基于 BERT 模型的关键词提取工具 Keybert 从问题上上下文中提取实体。为了进一步筛选出关键实体,使用剑桥词典和维基词典作为短语库,对提取到的实体进行过滤,得到初始实体集合 V_{initial} 。这个过程去除了不相关和冗余的实体,从而保留与问题上上下文高度相关的关键实体,实现了对知识子图的有效剪枝。

在本文中,选取 ConceptNet 作为外部 KG。ConceptNet 是一个多语言的 KG,包含了丰富的实体和关系。本文从中提取所有英文实体节点及其关系,形成一个英文 KG,记为 G_{en} 。基于获得的初始实体集合 $V_{initial}$,从 G_{en} 中识别每个对应节点,并对这些节点进行扩展,包括它们的一跳和两跳邻居节点及其关系,生成子图 $G_{initial} = (V; E)$ 。其中 V 是子图中实体节点集合, $V_{initial} \subseteq V$; E 是子图中所有节点间的边的集合,代表实体间的关系。通过这种方式,确保了子图 $G_{initial}$ 包含了与初始实体集合 $V_{initial}$ 相关的所有必要信息,并通过剪枝去除了冗余的实体信息,提高了知识子图的质量。

3.3.2 子图编码

得到子图 $G_{initial}$ 后,对其进行建模获得向量表示。首先,获取由 RoBERTa-Large 生成的实体嵌入表示 $E = \{e_1, \dots, e_k\}$,然后使用 GNN 编码器^[17]在 $G_{initial}$ 中的各个相邻节点之间传递消息,生成新的节点嵌入表示:

$$\{e'_1, e'_2, \dots, e'_k\} = GNN(e_1^{-1}, e_2^{-1}, \dots, e_k^{-1}) \quad (3)$$

这里 GNN 表示联合图注意力网络 CAT,它在 GAT^[18] 的基础上提升了学习能力,增强了性能。GNN 通过计算每个节点及其邻居对应的特征的线性组合作为新一层的节点表示。

$$e'_i = (\alpha_i + \theta \cdot \frac{1}{|\mathcal{N}_i|}) e_i^{-1} + \sum_{j \in \mathcal{N}_i, j \neq i} \alpha_{ij} e'_j \quad (4)$$

其中, $\theta \in [0, 1]$ 表示一个可学习参数, \mathcal{N}_i 表示节点 i 相邻节点的并集, α_{ij} 表示节点 i 和节点 j 间相关性的加权平均注意力,其计算式如下:

$$f_{ij} = \frac{\exp(\text{LeakyReLU}(e'_i \parallel e'_j))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(e'_i \parallel e'_k))} \quad (5)$$

$$s_{ij} = \frac{\exp(\arg \min \mathcal{D}(A_{ij}, Y_{ij}^T))}{\sum_{k \in \mathcal{N}_i} \exp(\arg \min \mathcal{D}(A_{ik}, Y_{ik}^T))} \quad (6)$$

$$\alpha_{ij} = \frac{f_{ij} \cdot s_{ij}}{\sum_{k \in \mathcal{N}_i} f_{ik} \cdot s_{ik}} \quad (7)$$

其中,通过式(5)计算节点 i 和节点 j 之间的上下文相关性 f_{ij} ,这一过程捕捉量化了两节点在上下文环境中的相互依赖关系, \parallel 表示连接函数;通过式(6)计算节点 i 和节点 j 间的补充相关性 s_{ij} , Y_{ij} 表示节点 i 和节点 j 之间的结构干预矩阵, A_{ij} 表示节点 i 和节点 j 之间包含初始特征信息的先验特征矩阵, $\mathcal{D}(\cdot)$ 表示欧氏距离公式。

3.3.3 k-Focus Attention

在 GNN 编码器对 $G_{initial}$ 进行编码后,本文引入了一个 k 稀疏注意力机制(k-Focus Attention, kFA)。对于第 i 个关键节点 e_i ,先计算它与子图中其他节点的距离,选取距离它最近的 k 个节点,然后计算注意力:

$$At_i = \text{softmax}\left(\mathcal{T}_k\left(\frac{QK^T}{\sqrt{d}}\right)\right) \quad (8)$$

其中, \mathcal{T}_k 表示选择最近的 k 个节点的运算, Q 表示节点的查询向量矩阵, K 表示节点的键向量矩阵。然后将注意力权重 At 和值向量 V 相乘得到新的 \hat{V} :

$$\hat{V} = AtV \quad (9)$$

kFA 机制着重考虑了与关键节点最为接近的前 k 个节点。这一方法增强了 GNN 对相关特征的捕捉能力,同时有

效抑制了大概率没用的节点所引入的潜在噪声。此外, kFA 机制还具备加速模型训练的优点,能够在不牺牲信息完整性的前提下,实现对特征的增强和噪声的过滤。

3.4 知识融合模块

为了实现问题上上下文与知识子图两种不同模态的知识信息的有效交互与融合,将知识模块分为两个部分:细粒度双模态交互融合层(Fine-grained Dual-modal Interaction Fusion Layer, FDID layer)和均值交互融合层(Mean Interactive Fusion Layer, MIF layer)。

3.4.1 细粒度双模态交互融合层

如图 2 所示,在细粒度双模态交互融合层,使用细粒度注意力对文本向量序列 S 和子图向量序列 E 进行信息融合:首先拼接文本向量与子图向量作为联合输入向量,并通过一个全连接层处理捕获两种模态间的交互信息,得到图文联合特征向量 F 。接着使用 Sigmoid 激活函数对图文联合特征向量 F 进行处理,生成一个用于调整原始向量权重的条件向量 R_F 。

$$F' = \tanh(W_F(S^{-1} \oplus E^{-1}) + b_F) \quad (10)$$

$$R_F = \text{Sigmoid}(F') \quad (11)$$

其中, W_F 为权重矩阵, \oplus 表示向量拼接, b_F 表示偏置项, R_F 反映了图文联合特征 F 中各元素的重要性。我们利用条件向量 R_F 分别与 S 和 E 对应元素相乘,对其中的特征进行加权,得到特征增强后的文本向量和图向量。再将两者进行矩阵乘法获取一个包含文本和子图交互信息的图文联合矩阵 C_F 。接着计算 C_F 的概率分布 a_F ,并将两者相乘来强化关键信息,得到图文交互注意力矩阵 B 。

$$C_F = (R_F \odot S^{-1}) \otimes (R_F \odot E^{-1}) \quad (12)$$

$$a_F = \text{softmax}(C_F) \quad (13)$$

$$B' = a_F \otimes F' \quad (14)$$

其中, \odot 表示对应元素相乘, \otimes 表示矩阵乘法。在细粒度双模态交互融合层的最后阶段,将图文交互注意力矩阵与原始向量序列连接,并对文本向量和图向量进行更新。

$$S' = W_S(B'; S^{-1}; E^{-1}) \quad (15)$$

$$E' = W_E(B'; E^{-1}; S^{-1}) \quad (16)$$

其中, W_S 和 W_E 是可学习的权重矩阵,在整合图文交互注意力矩阵和原始向量序列的同时调整维度,将更新后的向量作为第 $l+1$ 个细粒度双模态交互融合层的输入。

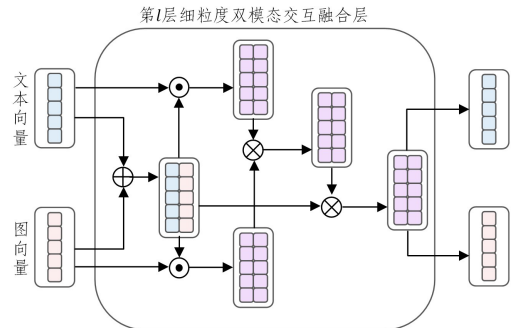


图 2 细粒度双模态交互融合层

Fig. 2 FDID layer

3.4.2 均值交互融合层

在经过 x 层细粒度双模态交互融合处理后,两种模态的信息已实现了深层次的交互与融合。在第二部分的均值交互

融合层,本文初始化两个均值交互向量 s_{ig} 和 e_{ig} ,分别用于表示文本和子图的均值特征,并作为双向文本-图编码器 f_{big} 的输入,对两种模态的信息进行进一步的提取与融合。

$$\begin{cases} s_{ig}^0 = \frac{1}{n} \sum_{i=1}^n s_i \\ e_{ig}^0 = \frac{1}{k} \sum_{j=1}^k e_j \end{cases} \quad (17)$$

$$s_{ig}^1, e_{ig}^1 = f_{big}(s_{ig}^0, e_{ig}^0) \quad (18)$$

这里使用两层多层感知器(MLP)作为本文的双向序列图编码器,对两个均值交互向量进行信息的交互和融合,生成新的均值交互向量。然后,将其添加到文本向量序列和子图向量序列中,再通过文本编码器和图编码器进行更新。

$$\begin{cases} S^x = \{s_1^x, s_2^x, \dots, s_n^x, \dots, s_{ig}^x\} \\ E^x = \{e_1^x, e_2^x, \dots, e_k^x, \dots, e_{ig}^x\} \end{cases} \quad (19)$$

$$S^{x+1} = LM(S^x) \quad (20)$$

$$E^{x+1} = GNN(E^x) \quad (21)$$

为了确保符号的一致性和清晰性,将初始的 s_{ig}^1 和 e_{ig}^1 统一表示为 s_{ig}^x 和 e_{ig}^x 。通过上述过程,实现了两种模态数据的有效融合。下文对基于融合后的数据进行答案预测。

3.5 答案预测

给定一个问题 q 和一个候选答案 a ,通过 x 层细粒度双模态交互融合层和 y 层均值交互融合层后,得到交互融合后的 S^z, E^z 。其中 $z=x+y$ 。计算 a 为正确答案的概率 P :

$$P = (a|q) = MLP([\text{AvgPool}(S^z) \oplus \text{AttnPool}(E^z)]) \quad (22)$$

其中, AvgPool 表示平均池化, AttnPool 表示注意力池化。最终使用交叉熵损失优化 GTICQA 模型进行梯度下降。

在 GTICQA 模型中,首先利用预训练的 LM 对输入的问题及其上下文进行深度编码,捕捉丰富的语义和句法信息。引入外部 ConceptNet 以及外部词典,从问题上下文中提取出相关的知识子图并使用 GNN 模块对其进行编码,学习图结构中的复杂关系与知识。在知识融合阶段,模型接收来自预训练 LM 的文本编码和 GNN 模块的图编码这两种模态的输入。通过细粒度双模态交互融合层和均值交互融合层融合两种模态的信息。经过多层交互与融合后,基于融合的信息对答案进行预测,并在第 4 章证明了 GTICQA 模型在常识性问答任务中的有效性和先进性。

4 实验结果及分析

4.1 数据集

本文针对 3 个具有代表性的选择问答数据——CommonsenseQA, OpenBookQA 和 MedQA-USMLE 进行了模型的评估。CommonsenseQA 数据集包含 12102 个问题,每个问题提供 5 个选项,它要求模型基于常识知识进行推理选择。由于无法获取官方数据的分割,本文参照先前研究的内部数据分割进行实验,确保实验的一致性和可比性。OpenBookQA 数据集包含 5957 个问题,每个问题提供 4 个选项,它更侧重于以基础科学知识进行问题的推理选择。本文参照 OpenBookQA 官方数据分割进行实验。MedQA-USMLE 数据集包含 12723 个问题,每个问题提供 4 个选项,它主要涉及生物医学领域的知识。本文使用官方的原始数据分割进行实验。通过在这 3 个数据集上的系统评估,本文模型在不同领域常

识选择问答任务中展现出有效性和泛化能力。具体统计细节如表 1 所列,其中 CSQA 表示数据集 CommonsenseQA, OBQA 表示数据集 OpenBookQA, MedQA 表示数据集 MedQA-USMLE。

表 1 数据集统计数据
Table 1 Dataset statistics

数据集	问题总数	选项数	训练集	验证集	测试集
CSQA	12102	5	8500	1221	1241
OBQA	5957	4	4957	500	500
MedQA	12723	4	10178	1272	1273

4.2 实验设置

本文针对上述 3 个数据集,采用了两种不同的 KG 作为结构化知识源。对于 CommonsenseQA 和 OpenBookQA 数据集,采用了广泛认可的常识知识图谱 ConceptNet 作为主要的结构化知识源,它包含了 799273 个节点和 2487810 条边,提供了丰富的常识性知识和关联信息。对于 MedQA-USMLE 数据集,使用针对生物医学领域自建的知识图谱,它整合了统一医学语言系统(UMLS)的疾病数据库以及 Drug Bank 的疾病数据库,包含了 9958 个节点和 44561 条边。在知识子图的初步剪枝过程中,本文采用了剑桥词典和维基词典作为短语库,以辅助筛选和优化知识子图中的信息。

在实验中,针对不同的数据集,使用了不同的预训练语言模型。对于 CommonsenseQA 数据集,使用 RoBERTa-large 作为 LM 的主干进行编码。对于 OpenBookQA 数据集,以 AristoRoBERTa 作为 LM 的主干。对于 MedQA-USMLE 数据集,采用了生物医疗领域的预训练语言模型 SapBERT 进行编码。在模型的训练中,使用 RAdam 作为模型的优化器。将 LM 编码器的学习率设置成 1×10^{-5} ,最大序列长度设置为 100,批大小设置为 128,GNN 编码器设置成 1×10^{-3} ,维度设置为 200。本实验所有模型采用 PyTorch 深度学习框架编写,利用一块显存为 24 GB 的 GPU (NVIDIA GeForce RTX4090)进行训练。

4.3 基准模型

本文模型侧重于提高 KG 的使用以及图结构知识与文本非结构化知识的交互与融合,所以选择 LM 和 LM+KG 的相关模型作为基准模型。首先,为了验证 KG 作为外部知识源的效果,选取了没有利用 KG 微调的 LM 进行比较。其次,在 CommonsenseQA 上,选取了微调的 RoBERTa-large 进行比较;对于 OpenBookQA,选取了微调的 RoBERTa-large 和 AristoRoBERTa 进行比较,其中 AristoRoBERTa 整合了额外的科学事实,仅适用于 OpenBookQA。最后,对于 MedQA-USMLE,选取微调的生物医学语言模型 SapBERT 进行对比。本文选取的现有 KG 增强的模型如下:

1) QAGNN (NAACL, 2021)^[5] 将问题上上下文信息映射为一个图节点,将其与知识子图相连接,对两种模态信息进行简单融合,实现 LM 与 KG 的联合推理。

2) JointLK (NAACL, 2022)^[6] 引入一个密集的双向注意力模块,聚焦于问题上上下文与知识子图节点,对两种模态的信息进行融合。

3) GreaseLM (ICLR, 2022)^[12] 将两种不同的态的信息通

过一个交互式模块传播到另一种模态中,从而实现两种模态信息的交互。

4)MKSQA(IPM,2023)^[19]利用多项选择问答任务中答案唯一性的优势和候选答案之间的交叉信息,从KG中检索并构建多个局部子图,实现跨子图联合推理。

5)HamQA(WWW,2023)^[13]在双曲空间中构建上下文感知的图注意力网络,通过联合优化方案来规范化问题上下文与知识子图之间的信息,实现信息的相互依赖。

6)FiTs(AAAI,2023)^[20]采用知识自适应来减小两种模态信息的差距,使用自监督学习方法识别知识,并强化相关知识的区分能力。

7)PipeNet(*SEM,2024)^[21]使用依赖解析工具对图进行剪枝,并提出了一个轻量化的图注意力网络进行两种模态知识的融合。

其中,QAGNN,GreaseLM,FiTs模型的主要工作在于建立两种模态间的交互和理解;JointLK,PipeNet模型更注重在两种模态融合过程中进行动态剪枝;MKSQA,HamQA模型则专注于采用不同策略对KG进行建模与融合。

4.4 结果及分析

本文将模型在数据集上的准确率作为评估模型性能指标,它衡量了正确预测的问题占总问题的比例。表2和表3分别列出了不同基准模型及本文提出的GTICQA模型在CommonsenseQA和OpenBookQA数据集上的性能。表2中IHdev一栏表示模型在CommonsenseQA内部分割的验证集上的准确率,IHtest一栏表示模型在内部分割的测试集上的准确率。表3中RoBERTa-Large一栏表示模型使用RoBERTa-Large作为LM的主干在OpenBookQA测试集上的准确率,AristoRoBERTa一栏表示模型使用AristoRoBERTa作为LM的主干在OpenBookQA测试集上的准确率。

表2 不同模型在CommonsenseQA上的性能表现

Table 2 Performance of different models on CommonsenseQA (%)

模型	IHDev	IHTest
RoBERTa-Large	73.07	68.69
QAGNN	76.54	73.41
JointLK	77.88	74.43
GreaseLM	78.50	74.20
MKSQA	—	74.53
HamQA	76.88	73.91
FiTs	—	75.60
PipeNet	78.95	74.75
GTICQA(Ours)	79.12	75.70

表3 不同模型在OpenBookQA上的性能表现

Table 3 Performance of different models on OpenBookQA (%)

模型	RoBERTa-large	Aristo RoBERTa
Fine-tuned LMs	64.80	78.40
QAGNN	67.80	82.77
JointLK	70.34	84.92
GreaseLM	—	84.80
MKSQA	71.80	—
HamQA	71.12	84.59
FiTs	—	86.00
PipeNet	69.60	87.80
GTICQA(Ours)	72.20	87.60

具体来说,在CommonsenseQA基准测试中,GTICQA模型在IHDev和IHTest上的准确率分别达到了79.12%和75.7%,相较于经过微调的RoBERTa-Large,分别实现了6.05个百分点和7.01个百分点的提升,相较于现有效果最好的模型实现了0.17个百分点和0.1个百分点的可比改进。对于OpenBookQA的基准测试,GTICQA模型在以RoBERTa-Large和AristoRoBERTa作为LM的主干上的准确率分别达到了72.2%和87.6%,相较于微调的LM分别实现了7.4个百分点和9.2个百分点的提升。在使用RoBERTa-Large作为LM主干的情况下,本文方法比现有效果最好的MKSQA模型实现了0.4个百分点的可比改进。值得注意的是,当使用AristoRoBERTa作为LM主干时,GTICQA落后于现有效果最好的PipeNet模型0.2个百分点,分析其原因是PipeNet精心设计的动态剪枝策略起到了作用,但在同样使用RoBERTa-Large的情况下,GTICQA与其相比提升了2.6个百分点,这表明GTICQA模型在不同的LM主干情况下具有适用性和有效性。

上述结果表明,GTICQA模型在处理常识性问答任务时,能够实现两种不同模态知识信息的高效交互与融合,从而展现出卓越的性能。相较于仅经过微调的LM,GTICQA模型通过引入外部知识源KG并采用LM+KG框架,证明了其框架的有效性;与QAGNN,GreaseLM,FiTs模型相比,GTICQA模型的知识融合模块更能促进两种模态间的信息交互与融合,进而提升性能;通过与JointLK,PipeNet模型的对比,证明了GTICQA模型采用的关键实体过滤的剪枝方法的有效性;通过与MKSQA,HamQA模型的对比,GTICQA模型的架构及其知识融合方法的有效性得到了进一步的验证。

除上述常识领域的多项选择问答基准之外,本文还在生物医学领域的MedQA-USMLE数据集上进行了测试,并在表4中列出了GTICQA的性能表现。结果表明,与在生物医学领域进行微调的语言模型以及其他结合知识图谱与语言模型的方法相比,GTICQA模型在性能上表现最佳。这一结果证明了本文模型的有效适用性和领域通用性。

表4 不同模型在MedQA-USMLE上的性能表现

Table 4 Performance of different models on MedQA-USMLE

模型	Test/%
SapBERT-base	37.20
QAGNN	38.00
GreaseLM	38.50
FiTs	39.20
GTICQA(Ours)	39.40

4.5 消融实验及分析

为了验证GTICQA各个模块的有效性,在CommonsenseQA的IHdev上对子图的生成剪枝,kFA模块、知识融合模块这3个主要模块进行了消融研究,实验结果如表5、图3、图4所示。

表5 各模块在 CommonsenseQA 上的消融测试

Table 5 Ablation tests of modules on CommonsenseQA

模型	IHDev/%
GTICQA	79.12
-子图生成剪枝	78.56
-kFA	79.05
-知识融合模块	75.94
-细粒度双模态交互融合层	76.76
-均值交互融合层	77.25

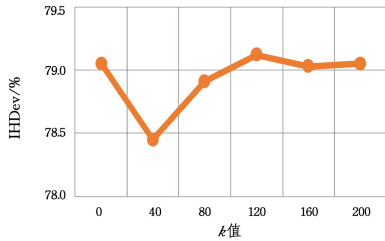


图3 kFA 模块的 k 值对性能的影响

Fig.3 k value effect of the kFA module

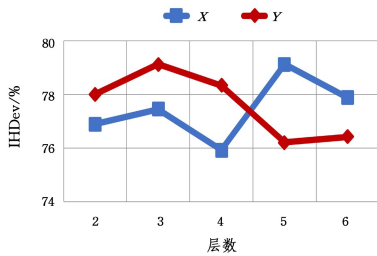


图4 不同交互融合层数对性能影响

Fig.4 Effect of different layer number

4.5.1 主要模块分析

当去除子图生成过程中的剪枝时,本文模型的性能下降了 0.56 个百分点,说明本文的剪枝策略可以提高知识子图的质量;当去除 kFA 模块后,本文模型的性能下降了 0.07 个百分点,说明 kFA 模块在增强 GNN 对关键实体特征的捕捉及抑制噪声节点方面发挥了作用;当去除整个知识融合模块后,本文模型的性能下降了 3.18 个百分点,说明有效地结合两种模态的知识信息可以更好地完成常识多选题问答任务,同时说明了该模块对两种模态知识信息融合的有效性。此外,将知识融合模块中的两个部分:细粒度双模态交互融合层和均值交互融合层,分别进行了消融。当去除细粒度双模态交互融合层时,本文模型的性能下降了 2.36 个百分点;当去除均值交互融合层时,本文模型的性能下降了 1.87 个百分点。这不仅说明了细粒度双模态交互融合层和均值交互融合层都对两种模态知识的融合的有效性,还表明了它们在增强知识融合效果方面的协同效应。

4.5.2 各模块参数分析

1)kFA 模块的 k 值影响。如图 3 所示,本文研究了不同 k 值情况下 GTICQA 的性能表现。随着 k 值减小,模型性能提升,当 k=120 时,GTICQA 获得了最佳性能。

2)知识融合模块中的细粒度双模态交互融合层数的影响。如图 4 所示,固定均值交互融合层层数 Y=3,观察到随着细粒度双模态交互融合层层数的增加,GTICQA 性能逐渐提升,在 X=5 时性能最佳。

3)知识融合模块中的均值交互融合层数的影响。如图 4 所示,固定细粒度双模态交互融合层层数 X=5,观察到随着细粒度双模态交互融合层层数的增加,GTICQA 性能逐渐提升,在 Y=3 时性能最佳。

4.6 实例分析

本文通过一个案例分析了子图的生成剪枝和 kFA 模块的作用。

如图 5 所示,将短语“rubber bands”看作一个整体,将“rubber”和“bands”作为不相关实体去除。因为“rubber”的含义对于解答当前问题并没有帮助,并且如果将它作为关键实体进行子图提取,还会引入如“drug store”“ballon”“tires”等一跳及两跳的邻居节点,这些节点与问题明显不相关,甚至可能会误导最终结果。所以,本文方法从 ConceptNet 中提取知识子图,通过外部词典的短语库,对问题上下文中的不相关实体进行了有效过滤。对知识子图进行编码后,通过 kFA 模块对与关键实体相关性较差的实体进行抑制。

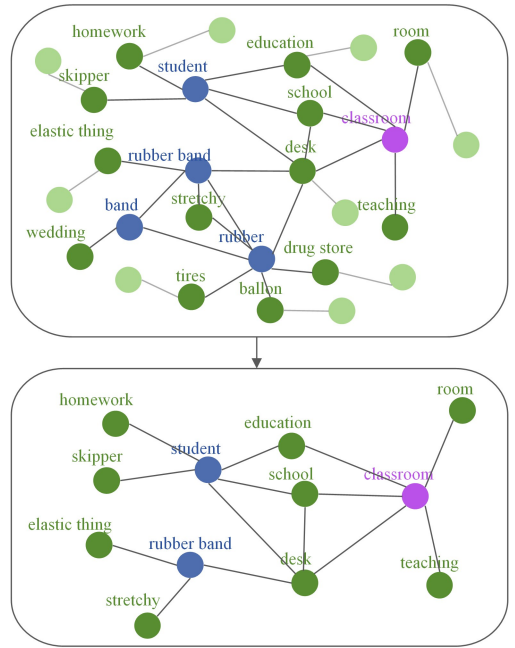


图5 子图的剪枝与 kFA 模块的作用

Fig.5 Role of subgraph pruning and kFA

结束语 本文提出了一种基于 KG+LM 框架的 GTICQA 模型,该模型通过外部词典精炼关键实体,引入一个 kFA 增强对子图全局特征的提取并抑制噪声节点干扰,同时提出一个由细粒度双模态交互融合层和均值交互融合层组成的知识融合模块,实现了对两种模态信息的有效交互与融合。通过在 3 个基准数据集上的实验,验证了 GTICQA 模型在子图提取与剪枝和知识交互融合方面的有效性和领域通用性。在未来的工作中,将进一步探索更高效地对两种模态的知识进行交互和融合,并探索更加有效的动态剪枝策略。

参考文献

[1] CHOWDHURY A, NARANG S, DEVLIN J, et al. Palm: Scaling language modeling with pathways [J]. Journal of Machine Learning Research, 2023, 24(240): 1-113.
 [2] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought

- prompting elicits reasoning in large language models[J]. Advances in Neural Information Processing Systems, 2022, 35: 24824-24837.
- [3] JI Z, LEE N, FRIESKE R, et al. Survey of hallucination in natural language generation[J]. ACM Computing Surveys, 2023, 55(12): 1-38.
- [4] FENG Y, CHEN X, LIN B Y, et al. Scalable Multi-Hop Relational Reasoning for Knowledge-Aware Question Answering[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 1295-1309.
- [5] YASUNAGA M, REN H, BOSSELUT A, et al. QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering[C]// Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021: 535-546.
- [6] SUN Y, SHI Q, QI L, et al. JointLK: Joint Reasoning with Language Models and Knowledge Graphs for Commonsense Question Answering[C]// Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022: 5049-5060.
- [7] REN H, DAI H, DAI B, et al. Lego: Latent execution-guided reasoning for multi-hop question answering on knowledge graphs[C]// International Conference on Machine Learning. PMLR, 2021: 8959-8970.
- [8] YU D, ZHU C, YANG Y, et al. Jacket: Joint pre-training of knowledge graph and language understanding[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2022: 11630-11638.
- [9] LIU S, QIN Y F, XU M, et al. Knowledge graph completion with triple structure and text representation[J]. International Journal of Computational Intelligence Systems, 2023, 16(1): 95.
- [10] MENG Z, LIU F, CLARK T, et al. Mixture-of-Partitions: Infusing Large Biomedical Knowledge Graphs into BERT[C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021: 4672-4681.
- [11] LIN B Y, CHEN X, CHEN J, et al. KagNet: Knowledge-Aware Graph Networks for Commonsense Reasoning[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 2829-2839.
- [12] ZHANG X, BOSSELUT A, YASUNAGA M, et al. GreaseLM: Graph REASONing Enhanced Language Models for Question Answering[C]// International Conference on Representation Learning (ICLR). 2022.
- [13] DONG J, ZHANG Q, HUANG X, et al. Hierarchy-aware multi-hop question answering over knowledge graphs[C]// Proceedings of the ACM Web Conference 2023. 2023: 2519-2527.
- [14] JIANG J, ZHOU K, WEN J R, et al. Great Truths are Always Simple: A Rather Simple Knowledge Encoder for Enhancing the Commonsense Reasoning Capacity of Pre-Trained Models[C]// Findings of the Association for Computational Linguistics: NAACL 2022. 2022: 1730-1741.
- [15] ZHENG C, KORDJAMSHIDI P. Dynamic Relevance Graph Network for Knowledge-Aware Question Answering[C]// Proceedings of the 29th International Conference on Computational Linguistics. 2022: 1357-1366.
- [16] QIAO S J, YANG G P, YU Y, et al. QA-KGNet: a language model-driven knowledge graph question-answering model. [J]. Ruan Jian Xue Bao/Journal of Software, 2023, 34(10): 4584-4600.
- [17] HET T, ONG Y S, BAI L. Learning conjoint attentions for graph neural nets[J]. Advances in Neural Information Processing Systems, 2021, 34: 2641-2653.
- [18] WANG K, SHEN W, YANG Y, et al. Relational Graph Attention Network for Aspect-based Sentiment Analysis[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 3229-3238.
- [19] ZHANG Q, CHEN S, FANG M, et al. Joint reasoning with knowledge subgraphs for Multiple Choice Question Answering [J]. Information Processing & Management, 2023, 60(3): 103297.
- [20] YE Q, CAO B, CHEN N, et al. Fits: Fine-grained two-stage training for knowledge-aware question answering[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2023: 13914-13922.
- [21] SU Y, ZHANG J, SONG Y, et al. PipeNet: Question Answering with Semantic Pruning over Knowledge Graphs[C]// Proceedings of the 13th Joint Conference on Lexical and Computational Semantics (*SEM 2024). 2024: 360-371.



CAI Ruixiang, born in 2000, postgraduate. His main research interests include machine learning and intelligent information processing.



ZHAO Shuliang, born in 1967, Ph. D., professor, Ph.D supervisor, is a member of CCF (No. 62875M). His main research interests include machine learning and intelligent information processing.

(责任编辑:柯颖)