

基于组合上下文提示的大型语言模型领域知识问答研究

方全, 张金龙, 王冰倩, 胡骏

引用本文

方全, 张金龙, 王冰倩, 胡骏. 基于组合上下文提示的大型语言模型领域知识问答研究[J]. 计算机科学, 2025, 52(11): 13-21.

FANG Quan, ZHANG Jinlong, WANG Bingqian, HU Jun. [Research on Domain Knowledge Question Answering via Large Language Models with Compositional Context Prompting](#) [J]. Computer Science, 2025, 52(11): 13-21.

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>

[基于实例级提示生成的多源域泛化故障诊断方法](#)

Multi-source Domain Generalization Fault Diagnosis Method Based on Instance-level Prompt Generation

计算机科学, 2025, 52(11): 213-222. <https://doi.org/10.11896/jsjcx.250300117>

[基于自注意力机制的图对比学习推荐算法](#)

Self-attention-based Graph Contrastive Learning for Recommendation

计算机科学, 2025, 52(11): 82-89. <https://doi.org/10.11896/jsjcx.240900134>

[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>

基于组合上下文提示的大型语言模型领域知识问答研究

方全¹ 张金龙² 王冰倩¹ 胡骏³

1 北京邮电大学人工智能学院 北京 100876

2 郑州大学河南先进技术研究院 郑州 450002

3 新加坡国立大学计算机学院 新加坡 117417

摘要 近年来,大型语言模型的快速发展引起了社会各界的广泛关注。大型语言模型虽然天然适应各种自然语言处理任务,但是在特定领域的问答任务中,由于缺少针对垂直领域的训练,生成答案的可靠性和适用性往往不尽如人意。为提升领域知识问答系统的性能,提出了一种新的基于组合上下文提示的大型语言模型领域知识问答方法。组合上下文提示包括了领域知识上下文和问答示例上下文两部分。领域知识上下文由采用基于对比学习的密集检索器从领域知识库中检索得到,能够增强大型语言模型的领域专业知识处理能力。问答示例上下文则通过语义相似检索从训练集中取得,能够提升大型语言模型对问题意图的理解能力。最后,将得到的组合上下文提示输入经过领域知识微调后的大型语言模型中,生成最终的领域答案。通过充分的实验和与基线模型的综合比较证明,所提方法在 BERTScore 指标上精确度和召回率分别比 ChatGPT 提高了 15.91% 和 16.14%,F1 Score 比 ChatGPT 提高了 15.87%。

关键词: 大语言模型; 领域知识问答; 组合上下文提示; 对比学习; 检索

中图分类号 TP391

Research on Domain Knowledge Question Answering via Large Language Models with Compositional Context Prompting

FANG Quan¹, ZHANG Jinlong², WANG Bingqian¹ and HU Jun³

1 School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China

2 Henan Institute of Advanced Technology, Zhengzhou University, Zhengzhou 450002, China

3 School of Computing, National University of Singapore, Singapore 117417, Singapore

Abstract In recent years, the rapid development of large language models has garnered widespread attention across various sectors. While these models naturally excel at various natural language processing tasks, their performance in domain-specific question answering tasks often falls short due to a lack of specialized training in vertical domains, leading to unreliable and less applicable answers. To improve the performance of domain knowledge question answering systems, this paper proposes a novel approach based on compositional context prompting for large language models. Compositional context prompting consists of domain knowledge context and question-answer example context. The domain knowledge context is retrieved from the domain knowledge base using a contrastive learning based dense retriever, which can enhance the domain expertise processing ability of large language models. The question-answer example context is obtained through semantic similarity retrieval from the training set, which improves the large language model's understanding of question intent. Finally, the obtained composite context prompts are inputted into the large-scale language model fine-tuned with domain knowledge to generate the final domain answers. Through extensive experiments and comprehensive comparisons with baseline models, the proposed method achieves an improvement of 15.91% in precision and 16.14% in recall on the BERTScore metric compared to ChatGPT, with an F1 Score improvement of 15.87%.

Keywords Large language models, Domain knowledge question answering, Compositional context prompting, Contrastive learning, Retrieval

到稿日期:2024-12-30 返修日期:2025-04-13

基金项目:北京市自然科学基金(JQ24019);空地一体新航行系统技术全国重点实验室开放课题(2024B31);国家自然科学基金(62036012)

This work was supported by the Beijing Natural Science Foundation(JQ24019), Open Project Program of State Key Laboratory of CNS/ATM (2024B31) and National Natural Science Foundation of China(62036012).

通信作者:方全(qfang@bupt.edu.cn)

1 引言

过去几十年中,信息技术的迅速发展推动了数据的爆炸式增长。然而,面对海量的数据,传统的数据库^[1]和搜索引擎^[2]主要依赖于基于关键词匹配的简单查询,这种方式常常难以满足用户对深度、专业领域知识的高级需求。在这一背景下,领域知识问答系统成为解决领域信息检索问题的一种重要工具,它可以通过自然语言交互,结合领域知识,为用户提供精准、即时的答案,使用户能够快速获取所需的领域信息。目前,已经出现了诸如面向医疗、金融、法律等特定领域的典型问答系统,它们在提供精准、可信赖的领域信息方面发挥着重要作用。

为了设计高效的领域知识问答系统,早期的工作主要采用符号化的方法。这些方法通过人工设计的规则模板,将自然语言问题转换为数据库查询语句,从而获取所需的信息,如 Baseball^[3]和 Lunar^[4]等专用领域问答系统。随着深度学习和知识图谱技术的兴起,基于深度学习模型并结合知识图谱的问答系统成为研究的热点。例如,Pei 等^[5]设计了基于 TransE 的中文领域知识图谱问答通用框架。目前,大型语言模型(Large Language Models, LLMs)^[6]的崛起为解决领域专业性问题提供了新的可能性。大型语言模型能够更高效、准确地理解人类提出的复杂语义问题,为用户提供智能问答和文本生成的技术支持。探索将 LLMs 应用于特定领域的问答系统,成为当前学术界和工业界的研究热点。

然而,尽管 LLMs 可以更为高效、准确地理解人类提出的复杂语义问题,并给予较为准确的回答,但是研究发现,由于其自身的解码器技术架构,通用型大语言模型生成的结果目前面临严重的幻觉问题^[7],如图 1 所示。

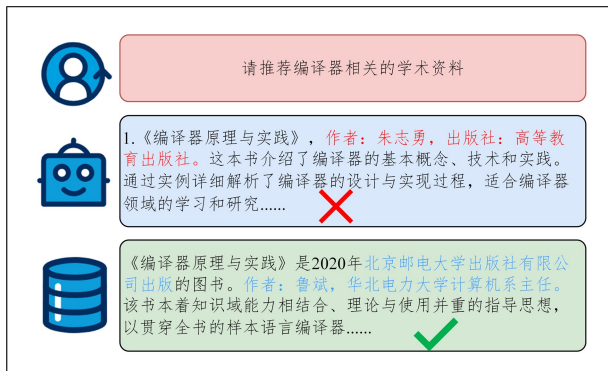


图 1 基于大型语言模型的领域问答幻觉示意图

Fig. 1 Schematic of domain question-answering illusion based on large language model

对于事实性问题,尽管 LLMs 能够自我修正,但前提是用户需具备一定的知识,能够明白 LLMs 的回答所存在的问题,因此其在大多数场景下会失去使用价值。导致幻觉问题的主要原因有:

1) 大型语言模型的领域专业知识处理能力弱。大型语言模型虽然具有通用领域问答的能力,但由于缺乏对特定垂直领域的训练,因此不能够有效地回答垂直领域的问题。

2) 大型语言模型有效理解领域问题的意图。尽管大型语

言模型在语言理解方面取得了巨大进展,但如果问题的上下文信息不够充分,会使其很难理解领域问题的真正意图。

针对以上问题,本文提出了一个组合上下文提示的大型语言模型领域知识问答的方法 CCP (Domain Knowledge Question Answering for Large Language Models Based on Compositional Context Prompts)。组合上下文提示包含了领域知识上下文和问答示例上下文。通过采用弱监督对比学习进行训练的领域知识上下文检索器,从线上数据库中检索出领域知识上下文,其主要用来增强 LLMs 的领域知识处理能力。同时,从训练集中获取问答示例上下文,以帮助 LLMs 更好地理解领域问题的意图。最后,将得到的组合上下文提示输入经过领域知识微调后的 LLMs 中,使其生成最终的领域答案。对于 CCP,在学术领域问答开展了实验,并且在面向通用领域的问答上也证明了其有效性。

本文的主要贡献如下:

- 1) 提出了一种新的基于组合上下文提示的大型语言模型领域知识问答方法 CCP,通过将领域知识上下文和问答示例上下文进行组合提示,引导大型语言模型更准确地回答问题;
- 2) 采用弱监督对比学习的方式训练了一个领域知识上下文检索器,来提高检索领域知识上下文的准确性;
- 3) 构建了一个包含 10 万条学术领域问答的数据集,并通过充分的实验和与基线模型的综合比较,证明了所提方法在提高特定领域问答系统性能方面具有显著优势。

2 相关工作

问答(Question Answering, QA)作为一个重要的研究问题,经常被研究人员从不同的方向进行研究。在开放领域知识问答中,已经涌现了多种系统和技术,它们依赖于不同形式的背景知识。一些基于管道的系统,如 OpenQA^[8],采用了模块化框架,使用标准化组件在结构化知识图谱(如 DBpedia^[9])上构建问答系统。Frankenstein^[10]基于自然语言输入问题,从社区创建的组件中创建了最合适的问答管道。QAnswer^[11]是一个多语言问答系统,通过查询不同链接的开放数据集获取正确答案。此外,Diefenbach 等^[12]在 QALD“关联数据问答”挑战^[13]的背景下讨论并比较了其他 QA-over-KG 系统(如 DEANNA^[14]和 SINA^[15])。

在上下文学习(In-Context Learning, ICL)方面,Brown 等^[16]首次发现预训练的大型语言模型具有通过少量示例在上下文中学习的显著能力。对于上下文能力的早期研究中,Xie 等^[17]也试图研究 ICL 是如何出现在预训练过程中的。Wie 等^[18]提出了首个基于 PAC 的上下文可学习框架,指出当预训练分布是潜在混合任务时,这些任务可以通过 ICL 进行有效的学习。

另一方面,检索增强生成首先被 Lewis 等^[19]提出,旨在通过整合检索到的相关信息来提高大型语言模型在问答任务中的性能。Luo 等^[20]提出了一种名为 ChatKBQA 的方法,这是一种新型生成再检索的知识图谱问答(Knowledge-based Question Answering, KBQA)框架,旨在结合 LLMs 和知识库检索,以提高问答系统的性能和效率。Lin 等^[21]提出了一种更为复杂的技术,其同时对 LLMs 和检索器进行微调,这种

微调基于查询、上下文和答案的三元组进行。Liu等^[22]通过Web搜索和检索功能来增强预训练大语言模型的通用能力。

在垂直领域中,检索增强生成方法也得到了广泛应用。例如,Ma等^[23]提出了一种基于检索的梯度提升决策树(RB-GBDT)模型,通过跨样本特征交互提升疾病风险评估的准确性。在法律问答领域,Wiratunga等^[24]通过动态检索法律条文和案例,提高了生成结果对法规更新的响应能力。Sarmah等^[25]则聚焦于金融领域,提出了以知识图谱为核心的增强生成框架。

然而,这些方法并没有同时考虑大型语言模型缺乏领域知识和不能有效理解领域问题意图的能力。因此,本文提出了一种组合上下文提示的大型语言模型领域知识问答的方法,通过组合领域知识上下文和问答示例上下文来提高大型语言模型领域知识问答的能力。

3 问题描述

给定一个领域知识库 $Z = \{z_i\}_{i=1}^I$, 每个 z_i 代表一个具体的领域知识条目。同时,给定一个问答集 $D = (x_i, a_i)_{i=1}^M$,

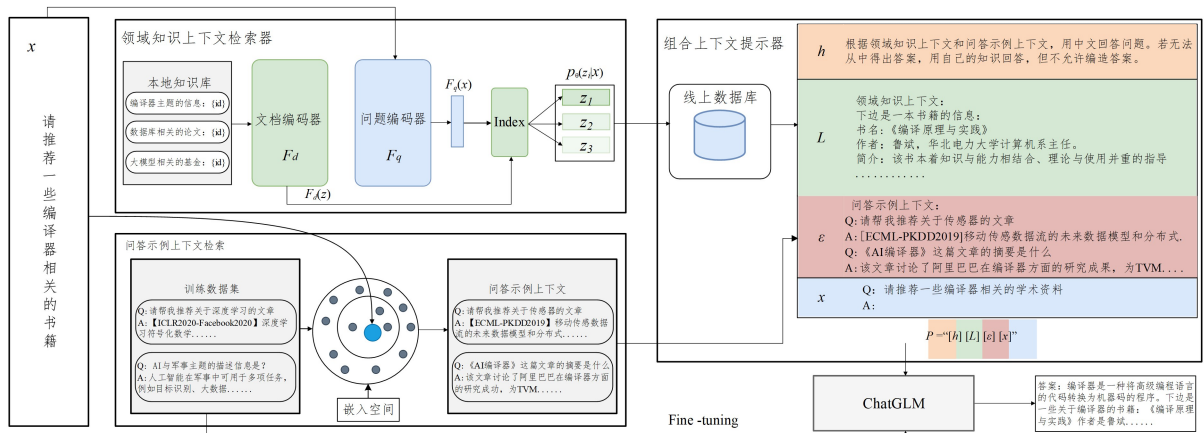


图2 CCP 框架

Fig. 2 Framework of CCP

基于组合上下文提示的大型语言模型领域知识问答方法可以分解为3个模块,分别是领域知识上下文检索器、问答示例上下文、组合上下文提示器。

4.1 领域知识上下文检索器

本节将介绍通过弱监督对比学习训练领域知识上下文检索器,以及利用训练好的检索器从本地知识库中进行检索。

在CCP中,领域知识上下文检索器采用了问题编码器 F_q 和文档编码器 F_d ,二者均为类似 Transformer^[27] 的编码器。其目标是根据给定的问题 x ,从构建的本地知识库 $Z = \{z_i\}_{i=1}^I$ 中检索出 K 个知识条目,这些知识预计能够为回答问题提供有用的信息。领域知识上下文将问题和知识条目分别编码为密集特征向量 $F_q(x) \in R^h$ 和 $F_d(z) \in R^h$ 。评分函数用于检索每个问题的知识条目,作为 x 和 z 表示之间的内积:

$$r(x, z) = F_q^T(x) F_d(z) \quad (1)$$

领域知识上下文检索器的训练目标是当知识条目 z 与回答问题 x 相关时,最大化得分 $r(x, z)$ 。然而,由于问题 x 和知识条目 z 之间的真实相关性并不容易获得,因此引入“伪相关

其中 x_i 表示具体的自然语言问题, a_i 是对应问题的答案。问答集提供了具体的问答示例,可以帮助模型学习如何利用领域知识库中的信息来回答问题。

设 f_{LLMs} 表示一个生成式大型语言模型,这种模型具有从给定上下文中生成连贯自然语言文本的能力。在领域知识问答系统中,本文的目标是让 f_{LLMs} 根据给定的领域知识库 Z 和问答集 D ,为每一个自然语言的测试问题 x 生成答案 a 。

4 整体框架

CCP的整体框架如图2所示。本文的目的是构建一个学术领域的知识问答系统,因此以专知平台线上数据库中存储的学术文档作为数据驱动,构造了一个学术领域的问答数据集,并使用该数据集对大型语言模型进行 Lora^[26] 微调。同时,为了准确检索领域知识上下文,构造了一个领域知识上下文检索器,其使用对比学习的方式进行了弱监督训练。此外,还从问答数据集中获取问答示例上下文,以供 LLMs 进行少样本学习。最终,通过组合上下文提示器将领域知识上下文和问答示例上下文设计为提示,并输入 LLMs 中生成答案。

标签”作为训练的代理。

为了构建伪相关性标签,引入了伪相关函数 $H(z, a)$,当知识条目 z 包含实际答案 a 中的关键词时,其值为1,否则其值为0。可以通过简单的字符串匹配,来确定知识条目 z 中是否包含实际答案 a 中的关键词。

对于每个问题-答案对 (x, a) ,为训练提取一个正面知识条目 $z^+(x)$ 。然后,使用批量负采样,将批次中除 $z^+(x)$ 之外的所有知识条目视为对于 (x, a) 是负面的。

将负面知识条目表示为 $N(x, a)$,正面知识条目的分数表示为 $r^+(x)$,则领域知识上下文检索器的损失函数 L_{DPR} 为:

$$-\sum_{(x,a) \in \mathcal{T}} \log \frac{\exp(r^+(x))}{\exp(r^+(x)) + \sum_{z \in N(x,a)} \exp(r(x,z))} \quad (2)$$

最终的领域知识上下文检索器会根据给定的问题 x 返回得分最高的 K 个知识条目,表示为 $\{z_k\}_{k=1}^K$ 。检索得分定义如下:

$$p_\theta(z_k | x) = \frac{\exp(r^+(x, z_k))}{\sum_{j=1}^K \exp(r^+(x, z_j))} \quad (3)$$

其中, θ 表示文档检索器的模型参数。

4.2 问答示例上下文

问答示例上下文的作用是帮助 LLMs 更好地理解领域问题的意图。已有研究表明,问答示例上下文的选择对于大型语言模型的少样本学习性能非常重要^[16]。其研究结果激发了本文设计出一种问答示例上下文选择策略。

给定一个测试输入 x 和任何训练输入 x_i , 可以使用向量化模型(如 text2vec-large-chinese^[28])来获取它们对应的向量 e 和 e_i 。如果 e 和 e_i 在嵌入空间中的距离很接近,它们更有可能共享类似的答案和问题输入。

首先计算测试输入与每个训练输入之间的融合特征的余弦相似度,然后在嵌入空间中选择前 N 个最近邻作为问答示例上下文。

$$J = \arg \operatorname{Top} N_{i \in \{1, 2, \dots, M\}} \frac{e^T e_i}{\|e\|_2 \|e_i\|_2} \quad (4)$$

其中, J 是训练集中前 N 个相似样本的索引集合。

问答示例上下文定义如下:

$$\mathcal{E} = \{(x_i, a_i) \mid i \in J\} \quad (5)$$

同时,训练输入的向量可以事先计算并存储,从而实现高效的问答示例上下文选择。

4.3 组合上下文提示器

大型语言模型在会话任务中取得了显著的性能。然而,它们的输出可能是不可靠的和带欺骗性的。因此,在这个阶段,使用从本地知识库中检索的知识条目和问答示例上下文,通过组合上下文提示器得到一个有助于回答用户问题的提示。

一个提示 P 由一个提示头 h 、一组问答示例上下文 E 、一组相关的知识文档 L 和一个用户输入 x 组成。因此,生成响应的整体输入提示如下:

$$P = "[h][\mathcal{E}][L][x]" \quad (6)$$

提示头用自然语言描述问答任务。虽然鼓励大语言模型根据检索到的相关知识生成答案,但也允许它进行广泛的探索,生成超出提示信息的答案。提示头部的完整格式如图 2 中的黄色框所示。

相关的知识文档 L 是组合上下文提示器模块从检索器检索到的各知识条目中提取相关的数据 ID 号,然后利用这些 ID 号去线上数据库中查询得到的。对应的知识文档表示为:

$$L = \{d_1, d_2, \dots, d_n\} \quad (7)$$

其中, $n \leq K$, 因为 K 个知识条目中可能存在相同的 ID 号。

问答示例上下文是从获得的 N 个问答示例上下文 $E = \{c_1, c_2, \dots, c_N\}$ 中得出的。对于测试输入,它的模板与上下文中的示例类似,只是答案槽是空白的,供大型语言模型填充。

将生成的提示与原始输入一起集成到大型语言模型中,以提高大语言模型回答的准确性和可靠性。最终,LLMs 根据问题 x 得到的提示 P 生成最终的答案 a 。

$$a = f_{\text{LLMs}}(a \mid P) \quad (8)$$

5 实验与结果分析

5.1 数据集

本文使用了 3 个问答数据集进行实验。第一个是本文构建的一个面向学术领域的问答数据集。第二个是面向通用领域的知识图谱问答数据集 NLPCC-ICCPOL 2016 KBQA。面向学术领域的问答数据集和通用领域的知识图谱问答数据集都是中文数据集。为了验证方法在其他语言上的领域知识问题处理能力,还使用了一个政治领域立场检测的大型英文数据集 P-STANCE 进行实验。

5.1.1 领域数据集的构建

本小节将介绍如何根据线上数据库中存储的学术文档构建本地知识库和领域问答数据集。

本地知识库主要是用来和线上数据库进行连接。关于本地知识库的构建,基于线上数据库中存储的学术文档,通过枚举的形式将每条文档根据不同的字段进行拆分,然后将拆分出来的字段分别和它们所对应的文档在数据库中的唯一标识符(ID号)进行关联,从而使本地知识库与线上数据库之间建议起一种紧密且精确的联系。后续通过组合上下文提示器,根据 ID 号从数据库中查找到对应的文档。例如,给定一个学术主题“大模型+知识图谱”,将该主题文档中相应的描述信息、相关主题等字段进行拆分,再和该主题对应的 ID 号进行组合。具体的形式如表 1 所列。

表 1 本地知识库数据形式

Table 1 Format of the local knowledge base data

大模型+知识图谱主题
大模型+知识图谱主题的描述信息: {2001376831669653}
大模型+知识图谱主题的相关主题: {2001376831669653}
大模型+知识图谱主题的相关人物: {2001376831669653}
大模型+知识图谱主题的资源链接: {2001376831669653}

领域问答数据集以问答对的形式组成。同样地,将线上数据库中存储的学术文档根据不同的字段进行拆分,构建了约 10 万条问答数据。同时,数据经过了自动和手动两步过滤。具体而言,我们自动过滤掉了答案过短的数据,其中大多数并没有回答任何实际的内容;并且手动过滤掉了一些答案质量较低的数据。后续直接使用这个领域问答数据集对大型语言模型进行混合微调,随机打乱数据顺序后输入到模型中微调。具体的数据集格式如表 2 所列。

表 2 学术领域问答数据集的数据形式

Table 2 Format of the academic domain Q & A dataset

问题	答案
《深度学习理论》综述这篇论文的摘要是什么?	近年来,深度学习已经成为机器学习和计算机视觉、自然语言处理等……
《基于大语言模型的问答技术研究进展综述》这篇论文的相关主题是什么?	大语言模型、问答技术、向量检索
数据挖掘主题的描述信息是什么?	数据挖掘一般是指从大量的数据中自动搜索隐藏于其中的有着特殊关系……
《人工智能简史》这本书的子主题有哪些?	机器学习、神经网络、深度学习……

方便起见,将本地知识库和领域问答数据集称为 ZhuanzhiData-100k。同时,从领域问答数据集中分别抽取 1 000 条问答对作为验证集和测试集。最终的数据信息如表 3 所列。

表 3 ZhuanzhiData-100k 数据集的总体信息

Table 3 Overall information about the ZhuanzhiData-100k dataset

Train	Valid	Test	Knowledge Item
103 800	1 000	1 000	926 972

5.1.2 通用领域问答数据集

除了使用学术领域问答数据集进行实验外,本文还选择了一个通用领域的知识图谱问答数据集 NLPCC-ICCPOL 2016 KBQA^[29]来验证所提方法。

NLPCC-ICCPOL 2016 KBQA 是由 NLPCC-ICCPOL 2016 的 KBQA 评估任务提供的中文知识图谱问答数据集。它提供了一个包含 14 609 个问答对的训练集和一个包含 9 870 个问答对的测试集;同时提供了一个中文知识库,包含 6 502 738 个实体,587 875 个属性以及 43 063 795 个(头实体、关系、尾实体)三元组。最终的总体数据信息如表 4 所列。

表 4 NLPCC-ICCPOL 2016 KBQA 数据集的总体信息

Table 4 Overall information about the NLPCC-ICCPOL 2016

KBQA dataset				
Train	Test	Entities	Attributes	Triples
14 609	9 870	6 502 738	587 875	43 063 795

KBQA 任务的流程是判别式,而大语言模型是生成式模型,这样会导致数据集中的某些问题非常依赖于知识库的检索,不适合大语言模型进行回答。例如,“告诉我张娜的体重是多少?”,这类问题一旦检索不到对应的三元组,大语言模型就没法利用自身的知识进行准确回答。因此,本文从测试数据集中挑选出了 3 000 条适合大语言模型进行回答的问答对来进行判定。

在评估 NLPCC-ICCPOL 2016 KBQA 数据集时同样遵循 CCP 框架,使用训练集对模型进行微调,同时从训练集中获取每个测试数据的问答示例上下文,并使用领域知识上下文检索器对三元组进行检索。

5.1.3 政治领域数据集

除了使用中文数据集外,还选择了一个政治领域的英文数据集 P-STANCE 来进一步验证所提方法。同样以问答任务的形式,使用本文框架进行立场检测,旨在让大型语言模型根据英文推文进行立场回答。

P-STANCE^[30]是一个开源的大型政治领域立场检测数据集,旨在促进自然语言处理中的立场分析研究。与社交媒体上构建的数据集相比,P-STANCE 数据集更具有挑战性,因为目标人物的提及更隐晦,包含更长的句子和更少的表面词汇线索,这意味着需要更深层次的语义理解来检测立场。该数据集能够更有效地检验模型系统在深层次语义理解和立场检测方面的性能。

P-STANCE 数据集包含 21 574 条与 2020 年美国总统选举相关的英文推文,专注于分析公众对 3 位政治人物——Donald Trump, Joe Biden 和 Bernie Sanders——的立场。每条推文都标注了公众对其中一位人物的立场。该数据集利用

亚马逊机械土耳其(AMT)平台进行众包标注,标注者被要求使用“Favor”(支持)、“Against”(反对)、“None”(无立场)或“I don’t know”(不确定)来标注每条推文。为了确保标注质量,对标注者做了严格的要求,包括完成的任务数量、地理位置和接受率。每条推文由 3 名随机标注者标注,并通过多数投票解决标签不一致的问题。

最终的数据集包含了针对“Donald Trump”的 7 953 条标注推文,针对“Joe Biden”的 7 296 条,以及针对“Bernie Sanders”的 6 325 条。每个人物的标签分布如表 5 所列。每条推文都标注有一个立场标签,即“支持”或“反对”。数据集按照 8:1:10 的比例划分了训练集、验证集和测试集。

表 5 P-STANCE 在不同对象上的标签分布

Table 5 P-STANCE label distribution on different objects

		Trump	Biden	Sanders
Train	Favor	2 937	2 552	2 858
	Against	3 425	3 254	2 198
Val	Favor	365	328	350
	Against	430	417	284
Test	Favor	361	337	343
	Against	435	408	292
Total		7 953	7 296	6 325

5.2 对比方法

为了验证本文模型的有效性,将其与以下几种问答模型进行对比实验。

Atom-7B-chat^[31];Llama2^[32]的完全开源版本,具有中文词汇扩展。

Vicuna-7B-v1.5^[33];在 Llama2 的基础上进行了监督指令微调。训练数据是从 ShareGPT.com 收集的大约 125 000 个对话。

BaiChuan-7B^[34];由百川智能开发的一款开源可商用的大规模预训练语言模型。基于 Transformer 结构,在大约 1.2 万亿 tokens 上训练的 70 亿参数模型,支持中英双语,上下文窗口长度为 4 096。

Chinese-Alpaca-2-7B^[35];基于 Llama2 的中文 LLaMA-2 以及 Alpaca-2 系列的模型。

ChatGPT^[36];由 OpenAI 开发的一个基于 GPT(Generative Pre-trained Transformer)^[37]架构的语言模型。GPT 系列模型旨在理解和生成自然语言文本,而 ChatGPT 专注于生成对话。

ChatGLM2-6B^[38];智谱 AI 和清华大学 KEG 实验室联合发布的对话预训练模型,具有 62 亿参数。

同时,在 KBQA 任务上,由于任务涉及知识图谱问答,因此本文采用了与该任务相关的基准模型。将 CCP 与 NLPCC-ICCPOL 2016 KBQA 任务中发布的基准系统(CDSSM)以及其他一些基准系统进行了比较,包括 MF-SMM^[39],DFM^[40],CGRM^[41],BT-CKBQA^[42],SUM^[43],BERT-QA^[44]。

此外,由于 P-STANCE 数据集专注于立场检测任务,因此采用了以下几种针对该任务的基准模型进行比较。

BiCE^[45];一种使用条件编码进行立场检测的模型。它通过编码目标信息来捕捉立场,并将这些信息用于跨目标立场

检测任务。

CNN^[46]:一种卷积神经网络模型。普通的 CNN 只将推文文本作为输入,不考虑目标信息。

TAN^[47]:一种基于注意力的 LSTM 模型。它提取特定于目标的特征。

KPT^[48]:一种基于知识提示的模型。它通过设计特定的提示模板,将外部知识与任务语境相结合,从而提升语言模型在少样本任务中的性能。

GCAE^[49]:一种 CNN 模型。它利用门控机制来阻断与目标无关的信息,是面向方面情态分析的一个强有力的基线。

PGCNN^[50]:与 GCAE 类似,基于门控卷积网络,并通过生成目标敏感的过滤器来编码信息。

BERT^[51]:一种预训练的语言模型。它通过向[CLS]令牌的隐藏表示中附加一个线性分类层来进行立场检测。

另外,为了验证弱监督对比学习训练领域知识上下文检索器的效果,采用了几种检索策略进行对比分析:余弦相似度(CosSim)、BM25、TF-IDF、局部敏感哈希(LSH)以及 LDA。其中,CosSim 和 BM25 是常见的检索策略;TF-IDF 通过词频与逆文档频率评估词语的重要性;LSH 通过将相似的文本映射到相同的哈希桶中,减少计算复杂度;LDA 则通过推断文档和词语的潜在主题分布,进行主题建模和相似度计算。

5.3 实现细节

实验选择 ChatGLM2-6B 作为骨干模型,该模型基于 GLM(General Language Model)^[52] 架构,采用了 GLM 的混合目标函数,并进行了 1.4 T 中英标识符的预训练以及人类偏好对齐训练。这个训练过程使得模型能够更好地理解语言结构和上下文。同时,ChatGLM2-6B 首先在 ZhuanzhiData-100k 的训练数据集上进行领域知识微调,这是一个领域语料库,使得 LLMs 具有对领域概念的基本理解。

训练过程中,在 4 个 Tesla V100-DGXS GPU 上对 ChatGLM2-6B 模型进行 Lora 微调,其仅微调低秩适配器。训练过程遵循以下超参数:秩的大小为 8,学习率固定为 3×10^{-4} ,1 个 epoch,最大序列长度为 512,无权重衰减。

对于基线模型,本文选择了一些不同的 LLMs 用于零样本学习,在运行基线模型时,全部使用的是基线模型的默认参数。在 ZhuanzhiData-100k 数据集上,对于领域知识上下文,分别从本地知识库中检索 1,3,5,7 个知识条目作为演示。同时,对于问答示例上下文,也分别从训练集中抽取 1,3,5,7 个问答对作为演示。若无特殊说明,CCP 方法均采用 $N=5$ 和 $K=3$ 。

5.4 实验评价指标

为了全面评估实验结果,针对 3 个数据集采用了不同的评估指标。

对于 ZhuanzhiData-100k 数据集,评估指标重点在于生成文本的语义和文本级别相似性,因此选择采用传统文本生成指标和基于模型的指标来进行评估。

1)传统文本生成指标。本文选择了两个传统文本生成指标,即 BLEU 和 ROUGE,来评估生成的答案。BLEU 主要是通过比较大语言模型生成的文本和参考文本中 n -gram(词

组)的重叠来工作, n -gram 的值通常为 $1 \sim 4$ 。BLEU 还考虑了短句的惩罚,以避免模型生成过短而不完整的句子。一个较高的 BLEU 分数,表明生成答案和标准答案之间的文本在词汇和短语层面更为接近。ROUGE 用于衡量生成答案和标准答案之间的重叠程度,同样,ROUGE 的分数越高,表示生成的文本在语义上与参考文本越接近。然而,这些评估指标主要用于衡量生成答案和标准答案之间的文本级相似性,这意味着它们并不能全面反映文本的语义相关性。

2)基于模型的指标。为了更全面地评估生成答案与标准答案之间的语义相似性,引入了 BERTScore 这一模型驱动的评估指标对其进行评估。BERTScore 利用预训练的 BERT 计算两个句子之间的语义相似性,它通过计算句子对之间的精确度、召回率和 F1 分数来评估文本的语义相似度,得分越高表示生成的答案和标准答案在语义层面上的相似度越大。这种评估方法在句子和系统级别上与人类判断密切一致。此外,相比其他评估指标,BERTScore 在处理对抗性示例时表现出了更强的鲁棒性。

对于 NLPCC-ICCPOL 2016 KBQA 数据集,由于 KBQA 任务主要关注生成答案和标准答案在字符级别的匹配度,而非复杂的文本生成或语义相似性评价,因此更适合采用字符级别的 F1 分数来进行评估。它可以衡量预测和真实之间的字符的重叠程度,分数越高,表示生成文本和标准答案在字符级别上的匹配度越高。

$$AveragedF1 = \frac{1}{|Q|} \sum_{i=1}^{|Q|} F_i \quad (9)$$

其中, F_i 是 Q_i 问题产生答案的 F1 值。如果标准答案和生成答案无交集,则 F1 为 0。

对于 P-STANCE 数据集,由于目标检测任务涉及多标签分类问题,因此实验采用 F_{avg} (平均 F 分数)和 F1-score 的宏观平均值(F_{macro})来评估基线模型的性能。 F_{macro} 避免了对标签分布不均衡的偏倚,确保模型在处理每个标签时都能保持均衡的表现。首先,按照式(10)和式(11)计算标签“Against”和“Favor”的 F1-score。

$$F_{favor} = \frac{2P_{favor}R_{favor}}{P_{favor} + R_{favor}} \quad (10)$$

$$F_{against} = \frac{2P_{against}R_{against}}{P_{against} + R_{against}} \quad (11)$$

其中, P 和 R 分别代表精确度和召回率。

之后,再计算 F_{avg} :

$$F_{avg} = \frac{F_{favor} + F_{against}}{2} \quad (12)$$

实验为每个目标计算 F_{macro} 。然后,通过所有目标上 F_{avg} 的平均值来计算 F_{macro} 。

5.5 实验结果与分析

表 6 和表 7 列出了所提方法和其他基线模型在 ZhuanzhiData-100k 数据集上的结果对比。本文采用了 3 种消融实验,其中 Domain-Knowledge Context 代表在 LLMs 的基础上增加领域知识上下文模块;QA-Example Context 代表在 LLMs 的基础上增加问答示例上下文模块;Fine-tuning 代表在 LLMs 的基础上进行领域知识微调;CCP 表示将上述 3 种方法结合在一起。

表6 各模型在 ZhuanzhiData-100k 数据集上的传统文本生成度量的结果对比

Table 6 Comparison of results of traditional text generation metrics for each model on the ZhuanzhiData-100k dataset

Type	Method	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-1	ROUGE-2	ROUGE-L
Zero-shot Reasoning	Atom	6.97	2.55	1.24	0.70	7.40	1.04	5.45
	Vicuna	9.04	3.64	1.88	1.11	10.16	1.76	7.33
	Baichuan	8.17	2.56	1.01	0.53	7.71	0.93	5.28
	ChatGLM	8.11	3.34	1.77	1.09	9.65	1.73	6.55
	ChatGPT	12.21	5.50	2.97	1.66	13.19	2.52	10.87
Domain-Knowledge Context	Chinese-Alpaca	10.25	4.25	2.18	1.24	11.26	1.99	8.74
	ChatGLM(K=1)	20.78	16.44	14.12	12.53	26.35	16.53	21.96
	ChatGLM(K=3)	21.43	17.48	15.25	13.67	27.58	17.91	22.94
	ChatGLM(K=5)	20.69	16.76	14.6	13.10	26.83	17.33	22.09
Fine-tuning	ChatGLM(K=7)	21.01	16.94	14.72	13.18	26.37	16.88	21.53
	ChatGLM	10.32	4.45	2.44	1.55	10.61	2.30	7.88
QA-Example Context	ChatGLM(N=1)	10.88	4.90	2.77	1.74	12.18	2.80	9.03
	ChatGLM(N=3)	12.49	6.20	3.87	2.62	14.08	3.85	10.79
	ChatGLM(N=5)	12.50	6.22	3.89	2.67	14.33	4.01	11.03
	ChatGLM(N=7)	12.19	6.19	3.83	2.61	14.29	3.98	10.79
CCP		27.61	22.78	20.14	18.61	33.2	23.07	28.72

表7 各模型在 ZhuanzhiData-100k 数据集上的基于模型的度量的结果对比

Table 7 Comparison of results of model-based metrics for each model on the ZhuanzhiData-100k dataset

Method	Precision	Recall	F1 Score
Atom	0.4902	0.5468	0.5118
Vicuna	0.5110	0.6103	0.5514
Baichuan	0.5443	0.5838	0.5589
ChatGLM	0.5030	0.6125	0.5476
ChatGPT	0.5568	0.6258	0.5855
Chinese-Alpaca	0.5444	0.6020	0.5664
CCP	0.6454	0.7268	0.6784

传统指标的主要结果如表6所列。从结果中可以看出,通过引入领域知识上下文,可以显著提高基线模型的表现。ChatGLM在不同的 K 值下都获得了远高于其他模型的BLEU和ROUGE指标,特别是在 $K=3$ 时表现最佳。在问答示例上下文方面,ChatGLM在不同的 N 值下都获得了较高的指标,且在 $N=5$ 时性能最好。同时,相比于问答示例上下文,领域知识上下文对模型的提升更加明显,这是因为领域知识上下文提供了更多有价值的信息。此外,领域知识微调也有效改善了模型的领域问答能力。最后,通过采用组合上下文提示(CCP),模型性能得到了进一步提升。

从表7中可以清楚地看出,在基于模型的指标BERTScore上,CCP的精确度和召回率分别比ChatGPT提高了15.91%和16.14%,而F1 Score比ChatGPT提高了15.87%。这意味着CCP在学术领域知识问答场景中生成的结果更受语言模型的接受。

同时,表8列出了所提方法和其他知识库问答(KBQA)任务模型在NLPCC-ICCPOL 2016 KBQA数据集上的结果对比,结果显示所提方法优于其他比较模型。另外,表9列出了所提方法和其他模型在P-STANCE数据集上的结果对比,同样可以看出所提方法优于其他比较模型。此外,表10列出了所提方法与不同检索策略的对比分析,其中,CCP-Retriever代表领域知识上下文检索器模块。结果表明,采用弱监督对比学习训练的领域知识上下文检索器在性能上优于其他检索策略。

表8 在 NLPCC-ICCPOL 2016 KBQA 数据集上的结果对比

Table 8 Comparison of results on the NLPCC-ICCPOL 2016 KBQA dataset

Method	平均 F1
CDSSM(baseline system)	52.47
MFSMM ^[39]	80.35
DFM ^[40]	83.45
CGRM ^[41]	85.77
SUM ^[43]	85.94
BT-CKBQA ^[42]	82.68
BERT-QA ^[44]	84.12
CCP	86.53

表9 在 P-STANCE 数据集上的结果对比

Table 9 Comparison of results on the P-STANCE dataset

Method	Trump	Biden	Sanders	F_{macro}
BiCE	77.15	77.69	71.24	73.36
CNN	76.80	77.22	71.40	75.14
TAN	77.10	77.64	71.60	75.45
KPT	80.20	80.40	77.10	79.23
PGCNN	76.87	76.60	72.13	75.20
GCAE	78.96	77.95	71.82	76.24
BERT	78.28	78.70	72.45	76.48
CCP	82.44	81.01	78.07	80.47

表10 不同检索策略在 ZhuanzhiData-100k 数据集上的基于模型的度量结果的对比

Table 10 Comparison of results of model-based metrics for different retrieval strategies on the ZhuanzhiData-100k dataset

Method	Precision	Recall	F1 Score
LSH	0.4969	0.6128	0.5440
LDA	0.5160	0.6001	0.5510
BM25	0.5829	0.6743	0.6202
TF-IDF	0.5929	0.6944	0.6341
CosSim	0.6297	0.5230	0.5659
CCP-Retriever	0.6959	0.6974	0.6372

综上,CCP方法在3个数据集上均优于所有基准模型,且消融实验结果均优于各对比实验结果,结果表明,基于组合上下文提示中的领域知识上下文和问答示例上下文对提高大型语言模型在领域知识问答中的表现都是有效的。

结束语 本文提出了一种基于组合上下文提示的大型语言模型领域知识问答方法 CCP, 以应用于领域知识问答任务。组合上下文提示主要由领域知识上下文和问答示例上下文两部分构成。其中, 领域知识上下文通过检索到的领域知识, 有效增强了大型语言模型领域专业知识处理的能力; 问答示例上下文通过相似语义匹配获得, 充分利用了大型语言模型的少样本学习能力, 提升了大型语言模型对于问题意图的理解能力。

在 3 个数据集上进行的对比实验证明, 所提出的方法相对于现有方法具有更好的性能; 消融实验的结果也表明, 领域知识上下文和问答示例上下文以及微调对于提高领域知识问答的性能均是有效的。实验结果表明, 本文方法在提高特定领域问答系统性能方面展现出了显著的优势, 为特定领域问答任务提供了更为创新和可靠的解决方案。

展望未来, 我们考虑将多模态信息(如图像和音频)纳入问答系统, 以实现更全面的信息检索。同时, 也期待在更多领域中应用 CCP 方法, 并开展更广泛的实证研究, 以验证其在不同领域知识问答任务中的有效性和适用性, 并希望这种方法能够为未来的相关研究提供参考和启示。

参 考 文 献

- [1] LIN Z Y, YANG D Q, WANG T J, et al. Keyword Query Based on Relational Databases[J]. *Journal of Software*, 2010, 21(10): 2454-2476.
- [2] SU Y, YANG S, SUN H, et al. Exploiting relevance feedback in knowledge graph search[C]// *Proceedings of the 21th ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*. 2015; 1135-1144.
- [3] GREEN J R B F, WOLF A K, CHOMSKY C, et al. Baseball: an automatic question-answerer[C]// *Western Joint IRE-AIEE-ACM Computer Conference*. 1961; 219-224.
- [4] WOODS W A. Progress in natural language understanding: an application to lunar geology[C]// *National Computer Conference and Exposition*. 1973; 441-450.
- [5] PEI Z, ZHANG J, XIONG W, et al. A general framework for Chinese domain knowledge graph question answering based on TransE[C]// *Journal of Physics: Conference Series*. IOP Publishing, 2020; 012136.
- [6] CHANG Y, WANG X, WANG J, et al. A survey on evaluation of large language models[J]. *ACM Transactions on Intelligent Systems and Technology*, 2024, 15(3): 39. 1-39. 45.
- [7] WEN S, QIAN L, HU M D, et al. A Review of Research Progress on Question Answering Technology Based on Large Language Models[J]. *Data Analysis and Knowledge Discovery*, 2024, 8(6): 16-29.
- [8] MARX E, USBECK R, NGOMO A C N, et al. Towards an open question answering architecture[C]// *Proceedings of the 10th International Conference on Semantic Systems*. 2014; 57-60.
- [9] AUER S, BIZER C, KOBILAROV G, et al. Dbpedia: A nucleus for a web of open data[C]// *International Semantic Web Conference*. Berlin; Springer, 2007; 722-735.
- [10] SINGH K, RADHAKRISHNA A S, BOTH A, et al. Why reinvent the wheel: Let's build question answering systems together [C]// *Proceedings of the 2018 World Wide Web Conference*. 2018; 1247-1256.
- [11] DIEFENBACH D, MIGLIATTI P H, QAWASMEH O, et al. QAnswer: A Question Answering prototype bridging the gap between a considerable part of the LOD cloud and end-users [C]// *The World Wide Web Conference*. 2019; 3507-3510.
- [12] DIEFENBACH D, LOPEZ V, SINGH K, et al. Core techniques of question answering systems over knowledge bases: a survey [J]. *Knowledge and Information Systems*. 2018, 55; 529-569.
- [13] ZOU L, HUANG R, WANG H, et al. Natural language question answering over RDF: a graph data driven approach[C]// *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*. 2014; 313-324.
- [14] YAHYA M, BERBERICH K, ELBASSUONI S, et al. Natural language questions for the web of data[C]// *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. 2012; 379-390.
- [15] SHEKARPOUR S, MARX E, NGOMO A C N, et al. Sina: Semantic interpretation of user queries for question answering on interlinked data[J]. *Journal of Web Semantics*, 2015, 30; 39-51.
- [16] BROWN T, MANN B, RUDER N, et al. Language models are few-shot learners[C]// *Advances in Neural Information Processing Systems*. 2020; 1877-1901.
- [17] XIE S M, RAGHUNATHAN A, LIANG P, et al. An explanation of in-context learning as implicit bayesian inference[J]. *arXiv*; 2111. 02080, 2021.
- [18] WIES N, LEVINE Y, SHASHUA A. The learnability of in-context learning[C]// *Advances in Neural Information Processing Systems*. 2024.
- [19] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks[C]// *Advances in Neural Information Processing Systems*. 2020; 9459-9474.
- [20] LUO H, TANG Z, PENG S, et al. Chatkbqa: A generate-then-retrieve framework for knowledge base question answering with fine-tuned large language models[J]. *arXiv*; 2310. 08975, 2023.
- [21] LIN X V, CHEN X, CHEN M, et al. Ra-dit: Retrieval-augmented dual instruction tuning[J]. *arXiv*; 2310. 01352, 2023.
- [22] LIU X, LAI H, YU H, et al. Webglm: Towards an efficient web-enhanced question answering system with human preferences [C]// *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2023; 4549-4560.
- [23] MA H, CAO J, FANG Y, et al. Retrieval-based gradient boosting decision trees for disease risk assessment[C]// *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2022; 3468-3476.
- [24] WIRATUNGA N, ABEYRANTNE R, JAYAWARDENA L, et al. CBR-RAG: case-based reasoning for retrieval augmented generation in LLMs for legal question answering[C]// *International Conference on Case-Based Reasoning*. Cham: Springer, 2024; 445-460.
- [25] SARMAH B, MEHTA D, HALL B, et al. Hybridrag: Integrating knowledge graphs and vector retrieval augmented genera-

- tion for efficient information extraction[C]//Proceedings of the 5th ACM International Conference on AI in Finance. 2024:608-616.
- [26] HU E J, SHEN Y, WALLIS P, et al. Lora: Low-rank adaptation of large language models[J]. arXiv:2106.09685, 2021.
- [27] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017.
- [28] XU M. Text2vec: Text to vector toolkit (Version 1.1.2) [EB/OL]. <https://github.com/shibing624/text2vec>.
- [29] ZHAO H, YUAN S, LENG J, et al. Calculating Question Similarity is Enough: A New Method for KBQA Tasks[J]. arXiv: 2111.07658, 2021.
- [30] LI Y, SOSEA T, SAWANT A, et al. P-stance: A large dataset for stance detection in political domain[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021:2355-2365.
- [31] LlamaFamily. Llama Chinese[EB/OL]. [2024-02-14]. <https://github.com/LlamaFamily/Llama-Chinese/tree/main>.
- [32] TOUVRON H, LAVRIL T, IZACARD G, et al. Llama: Open and efficient foundation language models [J]. arXiv: 2302.13971, 2023.
- [33] CHIANG W L, LI Z, LIN Z, et al. Vicuna: An open-source chatbot impressing gpt-4 with 90% * chatgpt quality[EB/OL]. <https://vicuna.lmsys.org>.
- [34] YANG A, XIAO B, WANG B, et al. Baichuan 2: Open large-scale language models[J]. arXiv:2309.10305, 2023.
- [35] CUI Y, YANG Z, YAO X. Efficient and effective text encoding for chinese llama and alpaca[J]. arXiv:2304.08177, 2023.
- [36] WU T, HE S, LIU J, et al. A brief overview of ChatGPT: The history, status quo and potential future development[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(5):1122-1136.
- [37] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training [EB/OL]. <https://gwern.net/doc/www/s3-us-west-2.amazonaws.com/d73fdc5ffa8627bce44dca2fc012da638ffb158.pdf>.
- [38] ZENG A, LIU X, DU Z, et al. Glm-130b: An open bilingual pre-trained model[J]. arXiv:2210.02414, 2022.
- [39] ZHANG Y, XU G, FU X, et al. Adversarial training improved multi-path multi-scale relation detector for knowledge base question answering[J]. IEEE Access, 2020, 8:63310-63319.
- [40] ZHOU G, XIE Z, YU Z, et al. DFM: A parameter-shared deep fused model for knowledge base question answering[J]. Information Sciences, 2021, 547:103-118.
- [41] ZHAO H, YUAN S, LENG J, et al. Calculating Question Similarity is Enough: A New Method for KBQA Tasks[J]. arXiv: 2111.07658, 2021.
- [42] YANG E, HAO F, SHANG J, et al. BT-CKBQA: An efficient approach for Chinese knowledge base question answering[J]. Data & Knowledge Engineering, 2023, 147:102204.
- [43] HAO H, SUN X, WEI J. A semantic union model for open domain Chinese knowledge base question answering[J]. Scientific Reports, 2023, 13(1):11903.
- [44] PRASANNA K S L, LOKESH S, CHANDRAMOULI G, et al. BERT-QA: Empowering Intelligent Question Answering with NLP and Entity Recognition[C]//2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAIC). IEEE, 2024:1006-1010.
- [45] AUGENSTEIN I, ROCKTASCHEL T, VLACHOS A, et al. Stance Detection with Bidirectional Conditional Encoding[C]//EMNLP 2016. The Association for Computational Linguistics. 2016:876-885.
- [46] YOON K. Convolutional neural networks for sentence classification[C]//EMNLP. 2014:1746-1751.
- [47] DU J, XU R, HE Y, et al. Stance classification with target-specific neural attention networks[C]//26th International Joint Conference on Artificial Intelligence (IJCAI 2017). International Joint Conferences on Artificial Intelligence, 2017:3988-3994.
- [48] SHIN T, RAZEGHI Y, LOGAN IV R L, et al. Autoprompt: Eliciting knowledge from language models with automatically generated prompts[J]. arXiv:2010.15980, 2020.
- [49] XUE W, LI T. Aspect based sentiment analysis with gated convolutional networks[J]. arXiv:1805.07043, 2018.
- [50] HUANG B, CARLEY K M. Parameterized Convolutional Neural Networks for Aspect Level Sentiment Classification[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:1091-1096.
- [51] KENTON J D M W C, TOUTANOVA L K. Bert: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of naacL-HLT. 2019:2.
- [52] DU Z, QIAN Y, LIU X, et al. Glm: General language model pre-training with autoregressive blank infilling [J]. arXiv: 2103.10360, 2021.



FANG Quan, born in 1988, professor. His main research interest is multimedia knowledge computing.

(责任编辑:柯颖)