基于句粒度提示的大语言模型时序知识问答方法*

李志东，罗琪彬，乔思龙
（华北计算技术研究所　大数据研发中心，北京　100083）

摘　要：知识问答是自然语言处理领域的研究热点之一，而时序知识问答还需考虑知识的时序关系，更是研究难点所在。当前时序知识问答方法通常将知识和问题的词向量相似度作为回答的重要依据，忽略了知识所蕴含的句粒度语义信息。对此，提出了一种基于句粒度提示的大语言模型时序知识问答方法，首先通过对句粒度提示的改进，让大语言模型高效学习句粒度语义信息，同时验证大语言模型在 Zero-shot、Few-shot 及弱监督微调下时序知识问答能力。在 ICEWS05-15 数据集上进行实验，所提方法回答正确准确率得到可观提升，体现了基于句粒度提示的大语言模型时序知识问答方法的有效性。

关键词：时序知识问答；大语言模型；提示学习；自然语言处理
中图分类号：TP391.1　　**文献标识码：**A　　**DOI：**10.19358/j.issn.2097-1788.2023.12.002
引用格式：李志东，罗琪彬，乔思龙. 基于句粒度提示的大语言模型时序知识问答方法[J]. 网络安全与数据治理，2023，42(12)：7-13.

Large language model based on sentence granularity prompts for temporal knowledge Q&A approach

Li Zhidong，Luo Qibin，Qiao Silong
(Big Data R&D Center，North China Institute of Computing Technology，Beijing 100083，China)

Abstract：Knowledge Q&A is one of the hot research topics in the field of natural language processing，and temporal knowledge Q&A is a difficult area of Q&A reasoning because it also needs to consider the temporal relationship of knowledge. Today's research usually focuses on the word vector similarity between knowledge and questions as an important basis for answering，while ignoring the sentence granularity semantic information embedded in the knowledge. In this paper，we propose a method of temporal knowledge Q&A for large language models based on sentence granularity prompts. Firstly，by improving the sentence granularity prompts，the large language models can learn the sentence granularity semantic information efficiently，and then the temporal knowledge Q&A ability of large language models under Zero-shot，Few-shot and weakly-supervised fine-tuning is verified. The experiments are conducted on the ICEWS05-15 dataset，and the accuracy of answers is significantly improved，which demonstrates the effectiveness of the temporal knowledge Q&A method for large language models based on sentence granularity prompts.

Key words：temporal knowledge graph question-answering；large language models；prompt learning；natural language processing

0　引言

业务系统中具有多种不同时间序列的数据信息，将这些数据通过相关性和因果关系相联系形成知识图谱有助于快速深入地掌握时序信息。此外，数据信息在时间维度上的语义表达不同，包括年、月、日等不同粒度，跨时间粒度的语义表达会对问答结果产生影响。由此，时序知识图谱（Temporal Knowledge Graph，TKG）的产生可以对不同的时间序列数据生成一个多层的、多粒度的知识图谱，使得时序之间的关系得以清晰描述。

基于知识图谱的问答系统（Question Answering System based on Knowledge Graphs，KGQA）最早被用于提高企业的核心竞争力，由于企业经营过程中沉淀了许多知识但并不能得到很好的利用，KGQA 的出现使得知识的完全利用成为了可能。而 TKG 是在传统的知识图谱上对时间

*基金项目：173 基础加强计划（2022-JCJQ-JJ-0935）

进行延伸，在三元组中加入时间维度，格式为“[头实体 关系 尾实体 时间]”。这样不仅仅描述了各实体间的关系，也包含了关系成立的时间点或者时间范围的信息。

时序知识图谱问答系统（Temporal Knowledge Graph Question Answering，Temporal KGQA）往往是基于时序知识库构建，需要经历复杂的推理，并且答案是实体或者时间。通常可以把问题分为两大类：简单问题和复杂问题。其中简单问题都是基于一跳关系的问答，答案是四元组内缺失的时间或者实体，需要问答系统找出或者推测出正确的答案；复杂问题中往往包含了“Before/After”“First/Last”等约束条件，大多需要复杂的时间推理，会比较难以解决。

传统的 Temporal KGQA 方法往往选择从特定的知识库中抽取实体和关系，将获取的内容填入模板生成答案；或通过特定规则和方法，与问题进行匹配产生答案；或是通过深度学习模型进行答案推理。但这些方法通常是在词粒度上进行匹配的，忽略了 TKG 本身所蕴含的句粒度语义信息，可能导致回答不准确。

针对上述问题，本文提出了一种基于句粒度提示的大语言模型 Temporal KGQA 方法，该方法分为三个部分。一是使用嵌入模型从 TKG 中提取出与问题相关的句粒度知识。具体来说，先把 TKG 中的四元组进行句粒度提示改造，再利用嵌入模型把问题和 TKG 向量化，最后通过向量的相似度匹配来从 TKG 中提取与问题高度相关的句粒度知识。二是通过提示学习、句粒度知识和问题来构建大语言模型需要的句粒度提示。三是基于大语言模型超强的语义理解能力，让其对句粒度知识进行理解和推理并得到答案。

本文的贡献总结如下：

（1）本文提出了一种基于句粒度提示的大语言模型来解决时序知识图谱问答的方法。

（2）本文通过结合提示学习、句粒度知识和问题，构建了多种提示模板，验证了大语言模型在无监督或弱监督下的时序知识问答能力。

（3）本文通过 LoRA 方法微调大语言模型，提升了大语言模型在时序知识问答任务上的性能。

（4）实验表明在 ICEWS05-15 数据集上，本文提出的方法最高可以达到 36% 的准确率，是一种科学可行的方法。

1　相关工作

1.1　时序知识图谱问答

KGQA 核心是机器对用户提出的自然语言的理解[1]，常规的 KGQA 对于处理多粒度时序关系和复杂的上下文关系稍有欠缺，同时，这些系统往往使用静态的知识库来处理自然语言问题，难以满足现实业务的需要。针对上述不足，基于句粒度提示的大语言模型时序知识问答的开发对其进行了完善和处理，提高了系统在多时序约束条件下的问答推理能力。为了解决时序知识图谱上的问题，《知识图谱：认知智能理论与实战》[2]介绍了基于知识图谱的智能问答系统的 Z 型框架，如图 1 所示，传统的解决方法主要依据这个框架从规则模板、语义解析和信息检索方面展开分析。

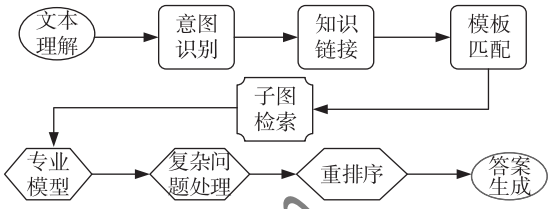

图 1　智能问答 Z 字框架

首先，基于规则和模板的方法通过定义一些预定义规则将问题和答案进行匹配，实现对知识图谱的问答。2017 年孙振[3]提出的基于人工智能标记语言（Artificial Intelligence Markup Language，AIML）规则的问答机器人系统使用 AIML 建立了问答知识库，实现了 AI 机器人对话系统的智能性和知识性。2021 年罗玲等[4]提出了基于知识图谱、词频-逆文本频率指数（Term Frequency-Inverse Document Frequency）和自注意力机制的双向编码表示（Bidirectional Encoder Representation From Transformers）的冬奥问答系统模型，用户可以精准获取冬奥会相关问答。此类方法的应用关键在于模板库的构建，事先需要很大工作量，因此该方法通常适用在一些简单、结构化的问答任务上，在多粒度时序知识问答上可用性较差。

其次，基于语义解析的方法也有许多研究，在简单 Temporal KG 中，Lan 等[5]2019 年提出将匹配-聚合模型以及特定上下文关系进行知识问答的方法用在知识库中，把问题和答案进行匹配，同时在 Saxena 等[6]2020 年提出的多跳问题问答（Multi-hop QA）解决方案中，使用了知识图谱嵌入（Knowledge Graph Embedding）技术来解决知识图谱上的 Multi-hop QA 问题。在复杂应用中，Luo 等[7]使用了复杂查询图（Complex Query Graph）来完成 Temporal KGQA 过程，并使用了三个知识库进行评估，为解决复杂 Temporal KGQA 问题提供了有效的解决办法。语义解析的方法避免了大量模板库的构建，主要在于自然语言到机器所能理解的语言的转化，容易受到语义鸿沟的影响，在解决 Temporal KGQA 问题上缺乏灵活性和通用性[8]。

在信息检索研究方法方面，Bordes 等[9]提出了利用深度学习模型从大规模无标签数据中生成嵌入向量，将

答案映射到同一空间并根据相似度进行答案匹配，处理了自然语言表述的复杂性问题。CRONKGQA 模型由 Saxena 等[10]在 2021 年提出，是一种基于 Transformer 的解决方案，利用最新的 TKG 嵌入方法使得问答在简单的时间推理问题上准确率很高，但在复杂问题的回答上可能稍有欠缺，且这种方法仅在词粒度上比较相关性，忽略了句粒度蕴含的语义信息。

1.2 大语言模型

自然语言处理（Natural Language Processing，NLP）一直以来都是人工智能领域的一个重要分支，解决 NLP 任务的模型统称为语言模型。语言模型最早来自于统计模型[11]，但随着神经网络和深度学习研究的发展，循环神经网络、长短时记忆网络也开始作为语言模型来应用。随着谷歌公司在 2018 年发布的 BERT[12]证明了预训练模型和 Transformer[13]架构的优越性后，语言模型的参数量开始爆发式增长，NLP 也就进入大语言模型时代。大语言模型通常是基于通识知识进行预训练，因此在面对特定场景时，常常需要借助模型微调或提示学习来提升大语言模型对下游任务的应用效果。

1.2.1 低秩适应微调

低秩适应[14]（Low-Rank Adaptation，LoRA）是大语言模型的高效微调方法，旨在解决大语言模型微调速度慢、计算开销大等问题。如图 2 所示，LoRA 微调时会冻结预训练模型的权重，并在每个 Transformer 块中注入可训练层，即秩分解矩阵。秩分解矩阵由降维矩阵 $\boldsymbol{A}$ 和升维矩阵 $\boldsymbol{B}$ 构成，$\boldsymbol{A}$ 矩阵由随机高斯分布初始化，$\boldsymbol{B}$ 矩阵由零矩阵初始化从而保证训练开始时秩分解矩阵是零矩阵。LoRA 微调时模型输入输出维度不变，只需要将秩分解矩阵的输出和预训练模型的输出相加作为最终的输出。这样，仅通过修改秩分解矩阵的参数而不需要修改预训练模型，就能快速提升大语言模型在特定任务上的性能。LoRA 微调的参数量较全参数微调显著降低，且性能与全参数微调基本持平。

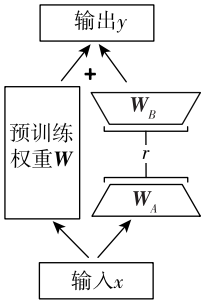

图 2　LoRA 微调

1.2.2 提示学习

提示学习是通过设计自然语言提示或指令来指导语言模型执行特定任务的方法[15]。早在 2018 年 Radford 等人[16]就已经在 GPT-1（Generative Pre-trained Transformer 1）中探索提示学习的应用。提示学习的目的是将下游任务通过提示模板转换为预训练的任务，如图 3 所示。

当使用掩码语言模型（Masked Language Model，MLM）来解决文本情感分类任务时，对于“I love this poem.”这句输入，可以在输入后面加上“The poem is_”这样的提示模板，然后让 MLM 用表示情感的词汇进行填空，如“wonderful”“terrible”等，最后再将该词汇转化成情感分类的标签。这样一来，通过选取合适的提示模板，便可控制模型的输出，从而使一个在通识数据集上训练的 MLM 可以被用来处理各种各样的下游任务。

1.2.3 上下文学习

作为一种特殊的提示学习形式，上下文学习在 GPT3[17]中首次得到应用，其核心思想是从类比中学习。基于大语言模型的泛化能力，上下文学习仅需要一些示例就能使模型快速适应所需要做的下游任务。如图 4 所示，仅通过 K 个下游问题的实例，就能让在通识数据上预训练的大语言模型快速适应文本情感分类任务。

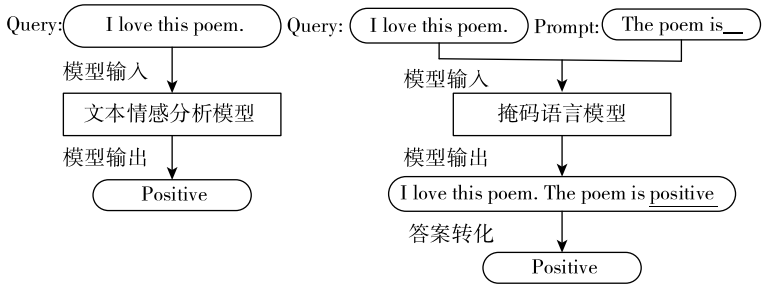

图 3　提示学习

根据提示学习中示例的个数也可以将上下文学习大致分类为三类：Zero-shot learning，One-shot learning 和 Few-shot learning。Zero-shot learning 只允许输入一则任务说明，不允许输入任何示例，即最原始的提示形式。One-shot learning 和 Few-shot learning 是在前者的基础上增加一条示例和多条示例。

2 关键技术

基于句粒度提示的大语言模型 Temporal KGQA 的原理如图 5 所示，在载入本地知识图谱文档后，首先会对文档进行切割，并使用嵌入模型将切割后的文档块向量化，所有向量依次存储到向量数据库中。然后使用相同的嵌入模型将问题向量化，将问题向量与向量数据库中的文档块向量进行匹配，选取出最高相似度的 K 个文档块向量并根据索引检索出对应的原文。根据已知答案的数据依照模板构建上下文学习的示例，最后将示例和匹配出的 K 段原文与问题一起添加到句粒度提示（prompt）中并输入到大语言模型。

2.1 本地知识文档的向量化

当输入本地知识文档后，程序首先会根据输入路径判断输入是文档还是文件夹，若是文档则直接读取，否则读取文件夹下的所有文档并将其连接到一起。读取文档后，程序会对文档进行切片并生成文档块。在切片的过程中如果遇到标点符号会直接切片，否则就等到切片长度达到限制时再切，且其长度可以通过参数来调节。

将文档切片后，使用嵌入模型“text2vec-large-chinese”对每个文档块进行向量化。如图 6 所示，嵌入模型首先会对文档块中的每个词进行向量化，然后将每个词的向量在对应维度求均值，最后将均值向量作为文档块的向量。

2.2 句粒度知识召回

将本地知识图谱文档向量化后，程序使用相同的嵌入模型将问题向量化，并计算问题向量与所有文档块向量的余弦相似度。然后程序会选择 K（个数可以通过参数来调节）个相似度最高的文档块作为已知信息添加到 prompt 中。值得强调的是，当本地知识图谱文档包含非结构化数据时，程序可能会把包含完整语义的一句话切成多个文档块，单个文档块可能无法表达完整的意思，所以在将召回的 K 个文档块添加到 prompt 中时，可以选择是否选取其邻近的文档块一起添加到 prompt 中。如图 7 所示，文档块 2 为匹配到的最佳文档块，文档块 1 和文档块 3 是其邻近文档块，程序会将最佳匹配文档块、邻近文档块和问题一起构建到 prompt 中，之后再输入到大语言模型中。

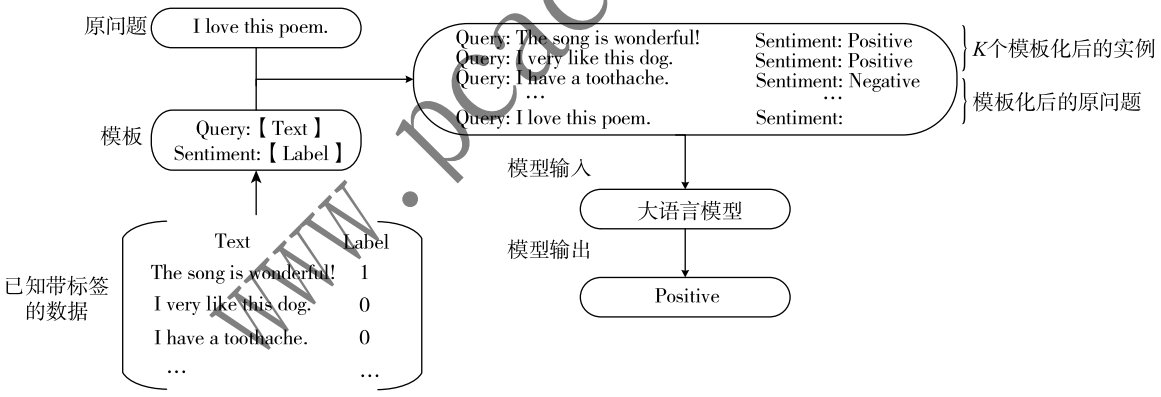

图 4 上下文学习

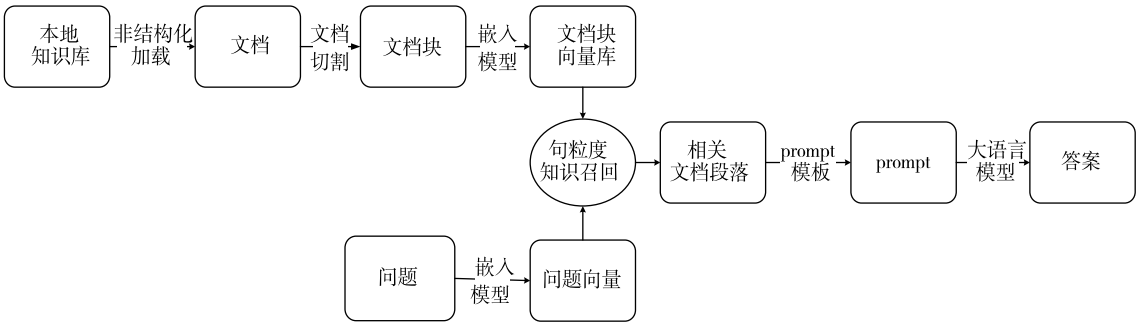

图 5 基于句粒度提示的大语言模型时序知识问答流程图

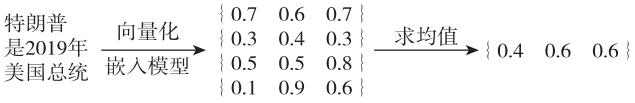

图 6　文档块向量化

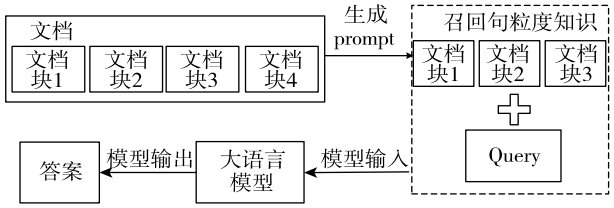

图 7　相关文档匹配流程

2.3　Prompt 构建

构建 prompt 时，本文根据上下文学习和对模型输出个数的限制来构建两大类 prompt。

如表 1 所示，本文首先会构建基础的 prompt，即不提供示例且限制模型只输出一个答案。在基础 prompt 上，再根据上下文学习来构建具有一个示例和具有五个示例的 prompt，也就是 One-shot learning 和 Few-shot learning。在构建示例时，本文会根据问题中的实体或时间，以字符串匹配的方式从知识图谱中获得该问题所对应的句粒度知识，之后会从中选取 5 条作为回答问题的已知信息，其中包含答案所对应的句粒度知识和用于干扰的句粒度知识。此外在 Few-shot learning 中，本文所选示例中有针对人物、时间、地点等多角度的问题，以此让示例包含尽可能多的信息。最后，本文会对模型的输出个数进行放开，以此来构建鼓励模型输出所有可能答案的 prompt。

在句粒度知识召回后，本文用召回的句粒度知识来替换模板中的“{context}”，再用问题来替换模板中的“{question}”。这样就生成了最终要输入到大语言模型中的 prompt。

2.4　大语言模型

本实验中，选用 ChatGLM-6B 作为 Temporal KGQA 的基础语言模型。该模型基于 General Laungage Model (GLM)[18] 架构，具有 62 亿参数。相比于 ChatGPT、MOSS、PaLM 等大语言模型，ChartGLM-6B 具有轻量运行的显著优势，这也是本实验选取它作为基础语言模型的主要原因。模型的 FP16 精度版本，仅需要大概 13 GB 显存即可运行。

3　实验

3.1　数据集

本实验采用数据集为 ICEWS05-15，TKG 的时序跨度为 2005 ~ 2015 年，数据集共包含 461 329 条数据，每条格式为“[头实体 关系 尾实体 时间]”。其中训练集 345 362条，验证集 24 683 条。本实验从训练集中选取前 10 000 条数据用于模型微调，从测试集中选取前 3 000 条用于模型测试。训练集和测试集都包含问题编号（quid），问题文本（question）和答案（answers）三项内容。问题与答案中包含多种粒度的时间信息，包括年、月和日，且问题中涉及多种时序逻辑约束，如之前、之后、首个等，问题答案限定于图谱中的实体和时间。

表 1　Prompt 模板

类型	Prompt 模板
基础模板	Respond to the user's question based on the content of the known information. The answer must be 1 word from the known information. Known information: {context}, Question is: {question}, Answer:
One-shot learning 模板	Respond to the user's question based on the content of the known information. The answer must be 1 word from the known information. Known information 1: …. Question 1: …. The answers 1: … Now we give the last known information and ask you to answer the questions based on the known information Known information: {context}, Question is: {question}, Answer:
Few-shot learning 模板	Respond to the user's question based on the content of the known information. The answer must be 1 word from the known information. There are 5 examples: Known information 1: …. Question 1: …. Answers 1: … Known information 2: …. Question 2: …. Answers 2: … …… Known information 5: …. Question 5 : … . Answer 5: …. Now we give the last known information and ask you to answer the questions based on the known information Known information: {context}, Question is: {question}, Answer:
不限制模型输出个数模板	Respond to the user's question based on the content of the known information. The answers must come from known information. Please output all possible answers. Known information: {context}, Question is: {question}, Answers:

鉴于大语言模型是以对话语料来进行训练的，为了让模型可以更好地理解 TKG 所蕴含的信息，本文对 TKG 进行了句粒度提示改造。在“［头实体 关系 尾实体 时间］”四元组中的尾实体和时间之间加入了“on”，并把四元组改成了字符串形式，删去了如“-”“_”等特殊字符。

3.2 评价指标

传统 KGQA 任务常用的评价指标为 MRR（Mean Reciprocal Rank），计算公式如下：

$$\mathrm{MRR} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{1}{\mathrm{rank}_i} \tag{1}$$

其中，Q 是问题数，rank_i 为第 i 个问题的第一个对应答案的排序值，如果正确答案未出现在候选答案集合中，则$\frac{1}{\mathrm{rank}_i}$取值为 0。

但 MRR 方法仅针对列表形式的答案进行评估，大语言模型以字符串的形式返回答案，因此本文采用一种松弛的 MRR 方法，记为 MRR-relax，计算公式如下：

$$\text{MRR-relax} = \frac{1}{Q}\sum_{i=1}^{Q}\mathrm{rank}_i \tag{2}$$

检测时序问答模型返回的字符串中是否包含答案池中的任意答案，若包含则 rank_i 得分为 1，否则得分为 0。

3.3 实验结果与分析

本实验使用“text2vec-large-chinese”作为嵌入模型，默认句子切分的最大长度为 30，在句粒度知识召回时只召回最佳文档块。LoRA 微调时使用 10 000 个训练样本，在预训练模型上训练 3 个 epoch。本实验采用交叉熵作为损失函数，LoRA 微调的损失曲线如图 8 所示。

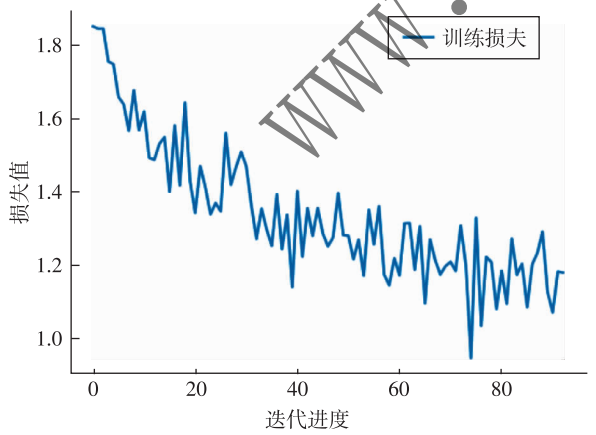

图 8 LoRA 微调损失曲线

本实验在预训练模型和 LoRA 微调模型上都进行了测试，并以句粒度知识召回数量为 5、上下文学习示例个数为 0、模型输出个数限制为 1 为基线方法，在上述三个方面设计了对比实验，详细实验结果如表 2 和表 3 所示。

表 2 预训练模型的实验结果

句粒度知识召回数	上下文学习示例数	模型输出个数限制	MRR-relax
5	0	1	0. 291 6
10	0	1	0. 313 6
20	0	1	0. 337 3
5	1	1	0. 312 6
5	5	1	0. 329 3
5	0	不限	0. 337 6
20	0	不限	0. 362

通过预训练模型的实验结果可以看到，随着句粒度知识召回数量的提升，大语言模型可以得到更多知识图谱的语义信息，模型的准确率也逐渐上升。此外基于上下文学习方法在 prompt 中添加示例，可以让大语言模型更好地适应 KGQA 任务，因此得到了更好的效果。最后，当不再限制大语言模型的输出个数时，模型会根据句粒度的提示信息输出所有可能的答案，相比于只输出一个答案，这显著提升了模型的性能。

表 3 LoRA 微调模型的实验结果

句粒度知识召回数	上下文学习示例数	模型输出个数限制	MRR-relax
5	0	1	0. 295 3
10	0	1	0. 325 0
20	0	1	0. 358 6
5	1	1	0. 309 6
5	5	1	0. 315 0
5	0	不限	0. 316 6
20	0	不限	0. 366 6

通过 LoRA 微调模型的实验结果可以看到，相比于预训练的大语言模型，仅使用 10 000 个数据集微调 3 个 epoch 后的模型在性能上也有明显提升，这也再次证明了基于句粒度提示的大语言模型时序知识问答方法的有效性。

4 结论

本文提出了一种基于句粒度提示的大语言模型时序知识问答方法，通过嵌入模型从时序知识图谱中提取与问题高度相关的句粒度知识，再根据提取出的句粒度知识、问题和上下文学习的内容来构建 prompt，最后依赖大语言模型超强的语义理解能力从句粒度的提示中得到答案。在时序知识图谱数据集 ICEWS05-15 上进行实验，取得了可观的效果，验证了该方法的有效性。在后续的

工作中，将基于知识图谱问答对来构建对话语料并使用多种方式对大语言模型进行微调，继续进行句粒度提示的大语言模型问答研究，为时序知识问答提供科学可行的解决方案。

参考文献

[1] 叶蕾，张宇迪，杨旭华. 利用知识图谱的多跳可解释问答[J/OL]. 小型微型计算机系统：1－11 [2023－06－16]. http://kns.cnki.net/kcms/detail/21.1106.TP.20230519.0923.002.html.

[2] 王文广. 知识图谱：认知智能理论与实战 [M]. 北京：电子工业出版社，2022.

[3] 孙振. 基于AIML的智能助理机器人系统 [D]. 合肥：安徽大学，2017.

[4] 罗玲，李硕凯，何清，等. 基于知识图谱、TF－IDF和BERT模型的冬奥知识问答系统 [J]. 智能系统学报，2021，16(4)：819－826.

[5] LAN Y S，WANG S H，JIANG J. Knowledge base question answering with a matching-aggregation model and question-specific contextual relations [J]. IEEE/ACM Transactions on Audio，Speech，and Language Processing，2019 (99)：1－1.

[6] SAXENA A，TRIPATHI A，TALUKDAR P. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics，2020.

[7] LUO K，LIN F，LUO X，et al. Knowledge base question answering via encoding of complex query graphs [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing，2018.

[8] 王智悦，于清，王楠，等. 基于知识图谱的智能问答研究综述 [J]. 计算机工程与应用，2020，56(23)：1－11.

[9] BORDES A，WESTON J，USUNIER N. Open question answering with weakly supervised embedding models [C]//Machine Learning and Knowledge Discovery in Databases：European Conference，ECML PKDD 2014. Springer Berlin Heidelberg，2014：165－180.

[10] SAXENA A，CHAKRABARTI S，TALUKDAR P. Question answering over temporal knowledge graphs [J]. arXiv preprint arXiv：2106.01515，2021.

[11] 钱学胜. 从ChatGPT看迈向通用人工智能的4种不同路径 [J]. 张江科技评论，2023，37(2)：11－13.

[12] DEVLIN J，CHANG M W，LEE K，et al. Bert：pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv：1810.04805，2018.

[13] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need [J]. Advances in Neural Information Processing Systems，2017，30.

[14] HU E J，SHEN Y，WALLIS P，et al. LoRA：low-rank adaptation of large language models [J]. arXiv preprint arXiv：2016.09685，2021.

[15] LIU P，YUAN W，FU J，et al. Pre-train，prompt，and predict：a systematic survey of prompting methods in natural language processing [J]. arXiv preprint arXiv：2107.13586，2021.

[16] RADFORD A，NARASIMBAN K，SALIMANS T，et al. Improving language understanding by generative pre-training [J]. 2018.

[17] DU Z X，QIAN Y J，LIU X，et al. GLM：general language model pretraining with autoregressive blank infilling [J]. arXiv preprint arXiv：2103.10360，2021.

[18] BROWN T B，MANN B，RYDER N，et al. Language models are few-shot learners [J]. arXiv preprint arXiv：2005.14165，2020.

（收稿日期：2023－09－30）

作者简介：

李志东（2000－），男，硕士，工程师，主要研究方向：知识图谱、提示学习。

罗琪彬（1996－），男，硕士，工程师，主要研究方向：态势认知、知识图谱。

乔思龙（1989－），男，硕士，工程师，主要研究方向：自然语言处理、大语言模型。

（上接第6页）

[12] LUKAS N，ZHANG Y，KERSCHBAUM F. Deep neural network fingerprinting by conferrable adversarial examples [J]. arXiv preprint arXiv：1912.00888，2019.

（收稿日期：2023－11－01）

作者简介：

屈详颜（1999－），男，博士研究生，主要研究方向：大模型水印、零样本图像分类。

于静（1989－），通信作者，女，博士，副研究员，主要研究方向：人工智能安全、跨模态人工智能。E-mail：yujing02@iie.ac.cn。

熊刚（1977－），男，博士，研究员，主要研究方向：网络测量与行为分析、信息对抗、信息安全。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部