



# 计算机科学

COMPUTER SCIENCE

## **DF-RAG:基于查询重写和知识选择的检索增强生成方法**

张浩然, 郝文宁, 靳大尉, 程恺, 翟颖

引用本文

张浩然, 郝文宁, 靳大尉, 程恺, 翟颖. [DF-RAG:基于查询重写和知识选择的检索增强生成方法](#)[J]. 计算机科学, 2025, 52(11): 30-39.

ZHANG Haoran, HAO Wenning, JIN Dawei, CHENG Kai, ZHAI Ying. [DF-RAG:A Retrieval-augmented Generation Method Based on Query Rewriting and Knowledge Selection](#) [J]. Computer Science, 2025, 52(11): 30-39.

---

## **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>

### [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>

### [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>

### [基于组合上下文提示的大型语言模型领域知识问答研究](#)

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

计算机科学, 2025, 52(11): 13-21. <https://doi.org/10.11896/jsjcx.241200198>

### [大语言模型与谣言:生成与检测的综述](#)

Large Language Models and Rumors:A Survey on Generation and Detection

计算机科学, 2025, 52(11): 1-12. <https://doi.org/10.11896/jsjcx.250700034>

# DF-RAG: 基于查询重写和知识选择的检索增强生成方法

张浩然 郝文宁 靳大尉 程恺 翟颖

陆军工程大学指挥控制工程学院 南京 210000

(ycdfzhr919@163.com)

**摘要** 大语言模型在会话任务中展现出强大的理解能力,但仍存在数据时效性及处理特定知识效率低等问题。检索增强生成(Retrieval-augmented Generation,RAG)成为解决上述问题的一种有效方案。然而,现有 RAG 仍面临查询理解偏差大、外部知识检索策略固化以及检索结果相关性低等挑战。针对上述挑战,提出动态细粒度检索增强生成(DF-RAG)方法。该方法由查询理解器、知识选择器和响应生成器 3 个模块构成,通过查询重写以及将外部相关文档纳入响应生成来改进基于检索增强的大语言模型管道,进而实现动态细粒度的检索增强。在 4 个开放域问答数据集上与 4 种不同类型的基准进行对比实验分析,结果表明,DF-RAG 在处理复杂模糊查询时能更有效地将外部知识与模型固有知识相结合,对于提高模型在复杂任务中的文本检索和响应生成能力具有重要意义。

**关键词:** 大语言模型;检索增强生成;知识问答;提示工程;信息检索

**中图分类号** TP391

## DF-RAG: A Retrieval-augmented Generation Method Based on Query Rewriting and Knowledge Selection

ZHANG Haoran,HAO Wenning,JIN Dawei,CHENG Kai and ZHAI Ying

College of Command & Control Engineering, Army Engineering University of PLA, Nanjing 210000, China

**Abstract** Large language models have demonstrated formidable comprehension abilities in conversational tasks, yet they still face issues such as data timeliness and inefficiency in handling specific knowledge. To address these challenges, Retrieval-augmented Generation(RAG) has emerged as an effective solution. However, existing RAG systems still encounter significant challenges, including query understanding bias, inflexible external knowledge retrieval strategies, and low relevance of retrieval results. In response to these issues, this paper proposes a Dynamic Fine-grained Retrieval-augmented Generation(DF-RAG) approach. This method comprises three modules: a query understander, a knowledge selector, and a response generator. By rewriting queries and incorporating externally relevant documents into response generation, it enhances the retrieval-augmented large language model pipeline, achieving dynamic fine-grained retrieval augmentation. Comparative experiments and analyses are conducted on four open-domain question answering datasets against four different types of benchmarks. The results indicate that DF-RAG can more effectively integrate external knowledge with the model's inherent knowledge when handling complex and ambiguous queries. This study holds significant importance for improving the model's text retrieval and response generation capabilities in complex tasks.

**Keywords** Large language models, Retrieval-augmented generation, Knowledge question answering, Prompt engineering, Information retrieval

### 1 引言

大语言模型(Large Language Models, LLMs)在文本生成领域已经展现出卓越的能力<sup>[1]</sup>。近年来, GLM-4 和 GPT-4<sup>[2]</sup>等经过海量训练数据预训练并经指令微调的 LLMs, 能够进行更细粒度的文本理解和生成, 在会话问答任务中彰显出巨大的潜力<sup>[3-4]</sup>。然而, LLMs 所存储的数据存在时效性问题, 致使其在处理实时更新的领域知识和研究进展时缺乏

准确性与相关性; 同时, 特定领域有其独特的术语、概念和知识体系, 导致 LLMs 在缺乏足够的特定领域数据的情况下, 难以准确理解和处理相关信息。

检索增强生成是当前辅助 LLMs 生成更准确、更丰富文本的主流技术。RAG 通过将外部检索到的相关信息作为有效线索整合到模型的提示中来增强 LLMs 的文本生成能力, 有效减少了知识密集型任务中的事实错误<sup>[5]</sup>。基本的 RAG 框架都遵循先检索再生成的方式, 利用 LLMs 的初始输入进

到稿日期: 2024-10-21 返修日期: 2025-02-19

基金项目: 国防工业技术发展计划(JCKY2020601B018)

This work was supported by the National Defense Industrial Technology Development Program(JCKY2020601B018).

通信作者: 郝文宁(hwnbox@163.com)

行相关性决策并检索相应的证据,虽能在简单下游任务中有效给出准确答案,但面对长上下文、多跳推理的复杂任务时会生成语义精确度低的内容。为提高生成文本的质量,Ma等<sup>[6]</sup>在重写-检索-阅读框架内引入问题重写器,将用户的问题与相应的知识库对齐,以增强信息检索的相关性。SELF-RAG<sup>[7]</sup>结合外部知识对查询问题进行检索标记,并整合检索和评价策略来提高响应的准确性和相关性。此外,一些方法对检索到的段落的相关性<sup>[8]</sup>以及 LLMs 现有知识和外部知识的结合问题<sup>[9]</sup>进行了研究,从而提升了模型生成质量。上述工作明显提升了 RAG 的性能,但所形成的复杂架构和交互模式降低了其效率。

然而,目前的 RAG 面临着几个关键挑战。1) 查询理解偏差大:在处理复杂或模糊查询时,易将查询过度泛化,检索到大量不相关信息,致使检索得到的信息与用户实际意图不符。此外,当查询涉及专业术语、缩写以及特定领域行话时,LLMs 检索生成的响应准确度非常低。2) 外部知识检索策略固化:当前检索策略多依赖固定规则来决定启用外部检索的时机,这会引发不必要的信息检索,不但增加了计算成本,还引入了与事实不相关或含有噪声的数据。此外,现有的检索策略通常仅局限于最近输出或标签,难以准确捕捉模型在整个上下文中的实时信息需求。3) 检索结果相关性低:以往的模型在进行预训练时并未对特定领域知识进行针对性训练微调,当检索到与任务不相关的文本时会分散注意力,进而导致生成质量低下。

为突破上述挑战的局限性,本文提出动态细粒度检索增强生成(DF-RAG)方法。该方法的框架由 3 个模块组成,包含查询理解器、知识选择器和响应生成器。首先,针对过度泛化及含有专业术语和缩写的查询,查询理解器对任务查询进行语义分析,通过扩充和提出相关问题的方式生成细粒度的重写查询,以解决模型对用户需求理解偏差大的问题。其次,利用知识选择器为问题生成特殊标签,判断是否需要触发检索机制来检索外部相关文档,能够避免不必要的信息检索及噪声数据的引入,并评估相关段落之间的相关性并进行相关性排序,分析不同段落与当前问题的关联程度。最后,响应生成器借助提示指令和问题相关段落生成最终响应,确保所提供的相关段落和响应之间的一致性,以解决检索结果相关性低的问题。

为验证 DF-RAG 的有效性,在 4 个开放领域问答数据集上将其与 4 种不同类型的基准模型进行比较实验。结果显示,在多数情况下,DF-RAG 在处理语义复杂的查询时表现出更为优越的性能,与现有工作相比,细粒度语义检索能更高效地将外部知识与模型固有知识相结合。

综上所述,本研究的贡献主要有以下几点:

- 1) 提出了一种新的 DF-RAG 框架,该框架由查询理解器、知识选择器和响应生成器 3 个不同模块组成;
- 2) 使用特定的提示指令,增强 LLMs 处理带有专业行话和领域知识的查询任务的能力,重写出语义丰富且更为标准化的查询任务;
- 3) 通过对不同标签的评估策略研究,以熵的阈值作为检索标签的判定依据,能够更精准地确定模型的信息需求;
- 4) 在抽取式阅读理解任务和复杂多跳问答任务上对模型

的性能进行全面评估,结果表明所提出的 DF-RAG 显著提高了各项指标的性能,充分展示出框架的潜力和鲁棒性。

## 2 相关工作

### 2.1 模块化检索增强生成

RAG 通过从大规模语料库中检索相关信息,辅助模型生成更为准确且丰富的输出<sup>[10]</sup>,是突破 LLMs 幻觉<sup>[11]</sup>、真实性<sup>[12]</sup>、长期记忆<sup>[13]</sup>等局限性的有效方法。

早期的 RAG 遵循索引、检索和生成的传统流程<sup>[14]</sup>。然而,其质量面临精度低、召回率低以及过程冗余等问题。随着 RAG 技术的持续发展和演变,新技术突破了传统的检索-生成框架,派生出模块化 RAG 的概念。Self-RAG 提出检索模块,旨在识别进行检索的最佳时机;Uprise<sup>[15]</sup>结合提示工程提出任务适配器模块,通过自动检索给定零样本任务输入的提示,提高跨任务和模型的通用性。

为提高模型的检索与理解能力,本研究引入了几个创新的模块化组件。

- 1) 查询理解器:全新的查询重写框架,借助 LLMs 的反馈,生成语义丰富且更为标准化的查询任务。
- 2) 知识选择器:通过生成反馈标签,判断检索时机并生成段落与问题之间的相关性。
- 3) 响应生成器:依据对话问题上下文和检索到的相关文档,为用户生成基于知识的自然语言响应。

### 2.2 黑盒语言模型的检索增强

目前广泛使用的大语言模型都具备良好的语言生成及上下文理解能力。然而,由于模型内部结构复杂,运作机制不透明,这些模型通常为非开源状态,以黑盒 API 的形式仅供用户调用来实现响应生成。现有研究证明,通过专门设计的提示词能够更大程度地激发 LLMs 的性能。Gao 等<sup>[16]</sup>利用 LLMs 为用户查询生成假设文档,随后通过这些假设文档来检索具有相似向量的真实文档,从而提高检索的准确性。RERLUG LST<sup>[17]</sup>借助语言模型产生监督信号以优化检索组件,使检索组件倾向于挑选出能降低生成文本困惑度的内容。

DF-RAG 通过提示黑盒大模型以标准术语替换原始表达并生成多种反馈标签来创建监督数据,随后将其知识提取至框架中,生成更标准化的查询任务并减轻不相关段落对模型的影响。

### 2.3 在反馈中学习

近期研究表明<sup>[18]</sup>,基于人类反馈(Human Feedback)对语言模型进行的强化学习,如近端策略优化(PPO<sup>[19]</sup>),极大地推动了 LLMs 的创建。RLHF(Reinforcement Learning from Human Feedback)<sup>[18]</sup>使用人类偏好学习奖励函数,并经由强化学习优化奖励以对齐模型,确保模型生成的文本质量,在提高输入提示实用性和相关性的同时,最大限度地减少潜在误差。Wu 等<sup>[20]</sup>引入细粒度 RLHF,以细粒度的人类反馈作为明确的训练信号,通过训练评估连贯性、事实准确性和信息完整性的奖励模型,提升语言模型生成内容的质量。Xu 等<sup>[21]</sup>提出对比非似然训练,通过语言误差反馈判断进行细粒度的不适当内容检测和纠正,展现了其在提升模型性能和与

人类偏好对齐方面的潜力。

DF-RAG 通过提示 LLMs 生成特殊标签的方法进行反馈训练,在进行检索时将反馈标签添加到相关文档中,用于实现可以控制的检索生成和更精确的答案生成。

### 3 DF-RAG 方法

DF-RAG 通过查询重写和提示标签反馈的方法改进基于检索增强的 LLMs 的管道,模型框架如图 1 所示。

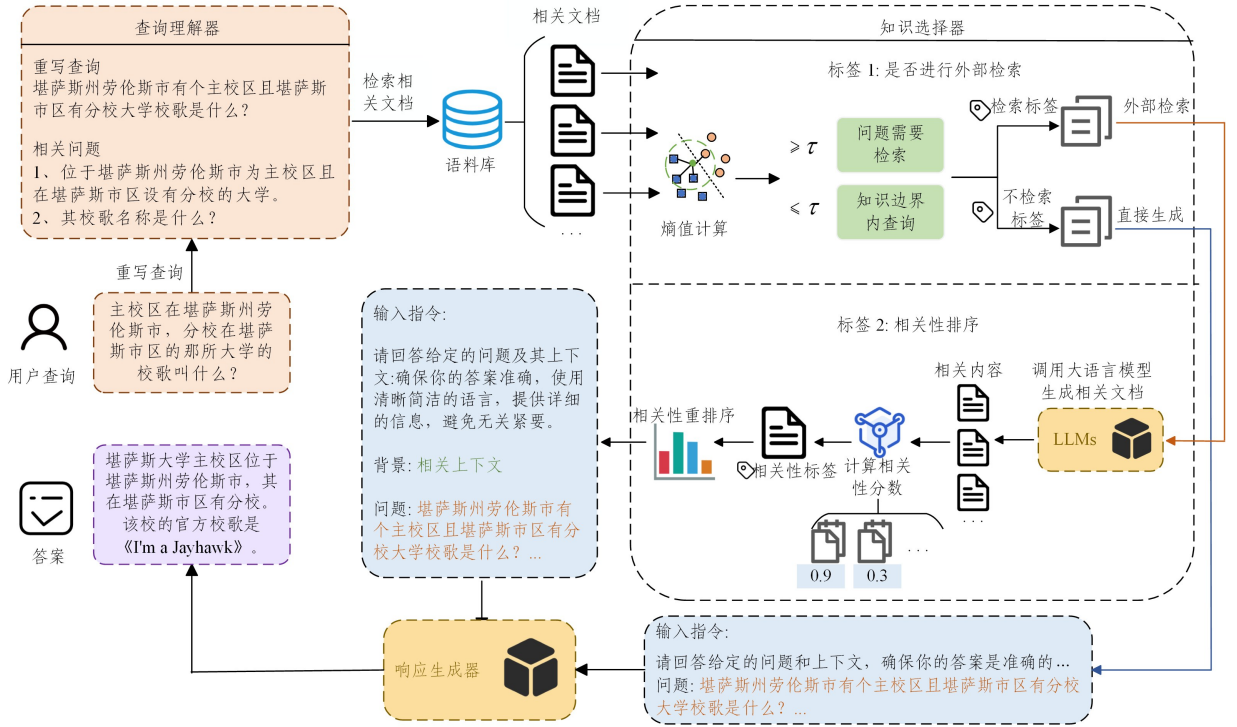


图 1 DF-RAG 框架概述

Fig. 1 Overview of the DF-RAG framework

#### 3.1 任务定义

本文所使用的符号及介绍如表 1 所列。

表 1 符号及其含义

Table 1 Symbols and their definitions

符号	定义
$W$	外部知识库
$S$	检索模型
$M_Q$	查询理解器
$M_R$	知识选择器
$M_G$	响应生成器
$M$	黑盒大语言模型
$D_n$	相关文档
$A_n$	相关片段
$A_n'$	最相关片段
$S_{acc}(q', a)$	查询质量分数
$q$	用户查询
$a$	答案
$Q_{train}$	重写查询数据集
$R(s_t, a_t)$	奖励函数
$A(s_t, a_t)$	优势函数
$\beta_{KL}$	发散系数
$V_\beta$	价值网络
$L_\theta$	价值损失
$L_\beta$	策略损失
$L_{PPO}$	最终损失函数
$H(A q)$	答案分布的熵值
$S_{true}$	检索相关性分数

在检索增强生成过程中,通常将给定的查询  $q$  输入检索模型  $S$ ,然后使用黑盒大语言模型  $M$  从外部知识库  $W$  中搜索

到的相关文档  $D_n = [d_1, d_2, \dots, d_n]$  生成对  $q$  更准确且有效的响应。DF-RAG 依托检索-生成的基本流程,将该流程模块化分解为查询理解器、知识选择器和响应生成器,以实现更准确、高效和个性化的响应,展现出高度的灵活性与优异性能。因此,本研究对各模块的任务定义如下。

**查询理解器:**当收到用户输入的原查询  $q$  时,模型根据设定好的提示指令  $I$ ,进行查询的知识扩充及相关问题的生成,最终得到重写查询  $q'$ 。在模型训练阶段,通过提示 LLMs 对给定知识密集型任务中的原查询  $q$  进行扩充重写,并计算生成查询  $q'$  的质量分数  $S_{acc}(q', a)$ ,随后利用 PPO 算法实现在线反馈训练。

**知识选择器:**对给定重写查询  $q'$  检索得到的相关文档  $D_n = [d_1, d_2, \dots, d_n]$  进行答案分布的熵  $H(A|q)$  的计算,当熵值大于检索阈值  $\tau$  时,添加外部检索标签 **Retrieve** = {Yes, No},以实现对外部检索时机的控制。在调用 LLMs 进行外部检索后,对重写问题  $q'$  的相关段落  $A_n$  进行相关性重排序,并添加相关性标签 **Relevant** = {0, 0.1, ..., 1},得到与查询最匹配的上下文  $A_n'$ 。

**响应生成器:**根据检索标签的不同,使用提示指令  $I$ ,结合重写查询  $q'$  和重排序后的最相关片段  $A_n' = [a_1, a_2, \dots, a_n]$  生成适当响应  $a$ 。最终生成的响应是基于细粒度检索增强的,其内容更加准确且符合用户需求。

算法 1 概述了 DF-RAG 的整个工作过程。当查询出现时,首先利用查询理解器将语义模糊的原查询重写为易于

理解的查询问题;接着,从语料库中检索出重写查询的一组相关文档,并根据查询问题是否需要外部知识解答,使用知识选择器对文档进行检索标签的标注,若需要进行检索,则调用黑盒大语言模型生成重写查询的相关段落,并用知识选择器评估段落相关性,根据相关性分数对段落重新排序;最后,使用响应生成器,根据集成的问题和相关片段生成最终答案。若检索到的相关知识足够支撑生成答案,则使用响应生成器直接输出问题答案。

### 算法1 DF-RAG的推理过程

需求:大规模语料库  $W$ ,检索器  $S$ ,查询理解器  $M_Q$ ,知识选择器  $M_R$ ,

响应生成器  $M_G$ ,黑盒大语言模型  $M$

输入:带有提示的查询  $q$

输出:答案  $a$

1.  $q' \leftarrow$  使用  $M_Q$  将查询  $q$  重写; // 查询重写
2.  $D_n \leftarrow$  使用  $S$  从  $W$  中检索  $q'$  的相关文档;
3. for  $d_n \in D_n$  do // 生成检索标签
4. 使用  $M_R$  生成 Retrieve 标签;

5. end

6. if Retrieve == Yes then

7.  $A_n \leftarrow$  使用  $M$  生成  $(q', D_n)$  的相关段落;

8.  $M_R$  生成  $A_n$  的 Relevant 标签; // 生成相关性标签

9.  $A_n' \leftarrow$  基于 Relevant 标签对  $A_n$  进行重排序;

10.  $a \leftarrow M_G$  根据提示指令输入  $(q', A_n')$  生成答案; // 生成响应

11. else if Retrieve == No then

12.  $a \leftarrow M_G$  根据提示指令输入  $q'$  生成答案; // 生成响应

13. end

## 3.2 任务理解器

该模块的关键任务在于对原始查询  $q$  进行语义增强,将其重写成易于理解的标准化查询  $q'$ ,并通过对问题进行细致的理解来扩充查询,以便更好地检索相关段落。重写查询的具体过程如图2所示。本研究旨在构建一个精准的重写模型  $M_Q$ ,其能够将  $q$  重写为:

$$q' = M_Q(q) \quad (1)$$

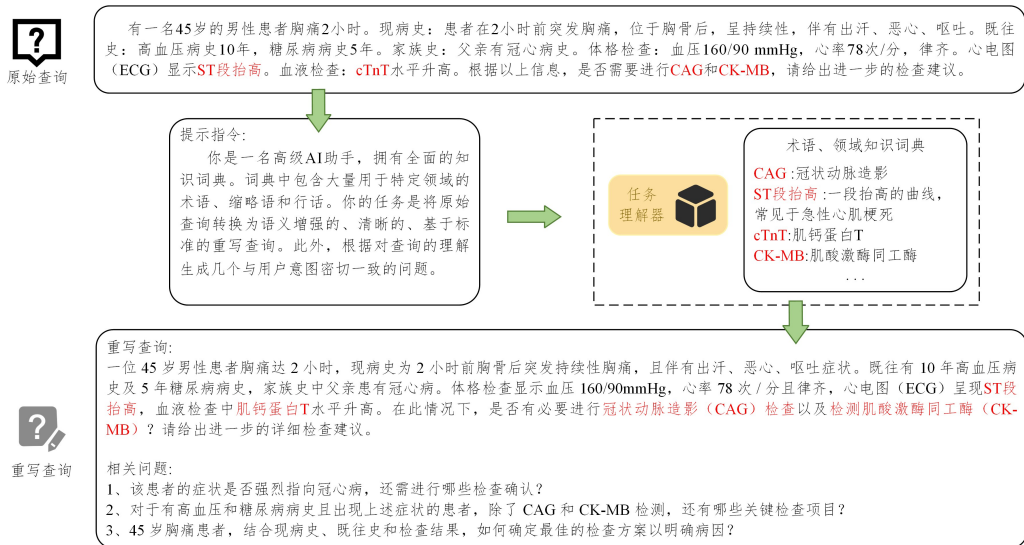


图2 重写查询过程

Fig. 2 Query rewriting process

### 3.2.1 监督微调重写器

调用黑盒 LLMs 进行查询重写会带来一些问题,在面对专业领域知识的查询重写任务时,若完全调用黑盒 LLMs 将会产生机器幻觉的现象,将问题过度泛化并引入噪声数据。此外,在重写任务过程中,推理错误或无效搜索也会阻碍模型的性能。采用 PPO 算法和监督微调相结合的方法来改进重写模型可以进一步提高改写效果,使其更适应特定的任务和检索需求。

本研究使用 T5-large<sup>[22]</sup> 作为训练模型,由于 T5 模型不具备查询重写能力,首先对重写模型进行初始化。在生成重写数据过程中,将 LLMs 视为一个能够涵盖广泛术语和特定领域知识的术语词典。通过制定恰当的提示指令  $I$  来提示 LLMs 识别并使用标准术语重写原始查询  $q$ ,以提高查询重写数据集的数据质量。具体而言,通过将当前查询  $q_t$  及其相关联的上下文  $a_n$  附加到指令  $I$  中,来提示 LLMs 进行查询重写:

$$Q_t' \sim LLM(I, a_n, q_t) \quad (2)$$

其中,  $Q_t'$  为重写查询  $q_t'$  的总和,本研究使用重写数据集  $Q_{\text{train}} = \{(q_t, q_t') \mid q_t' \in Q_t'\}$  进行  $M_Q$  的监督微调。

查询理解器  $M_Q$  以交叉熵损失为训练目标,在  $Q_{\text{train}}$  数据集上进行微调,表示为:

$$L_{\text{ce}} = - \sum_t \log M_Q(q_t' \mid q_{<t}, q) \quad (3)$$

训练模型的性能在很大程度上依赖于人工书写提示指令的质量,提示指令应遵循逻辑正确、高清晰度和低冗余度的原则<sup>[23]</sup>进行编写,将查询的模糊表达清晰化,以提高重写效率。

### 3.2.2 反馈训练

对于查询重写的反馈训练,DF-RAG 使用策略 PPO<sup>[19]</sup> 进一步微调重写器,具体训练过程如图3所示。在每个步骤  $t$  中,动作  $a_t$  是根据当前状态  $S_t = (q, \hat{q}_{<t})$  来生成下一个令牌  $q_t'$ ,其中状态  $S_t = (q, \hat{q}_{<t})$  由原始查询和生成的重写查询组成。在奖励函数设计中,将生成查询的质量作为评价标准,并加入了 KL 发散正则化<sup>[24]</sup>策略,以防止数值偏离过大。奖励计算如下:

$$R(s_t, a_t) = S_{\text{acc}}(q', a) - \beta_{\text{KL}}(M_Q \| M_\theta) \quad (4)$$

其中,  $S_{\text{acc}}$  是  $q'$  生成答案  $a'_x$  与准确答案  $a_x$  的准确率,  $M_\theta$  是训练阶段的固定模型。

其次, 对  $M_Q$  的价值网络  $V_\phi$  进行初始化, 从动作价值中减去状态价值来降低学习过程中的方差, 从而提高整个学习过程的稳定性。然后, 使用价值网络  $V_\phi$  构建优势函数  $A(s_t, a_t)$  (Advantage Function):

$$\begin{cases} \delta_t = R(s_t, a_t) + V_\phi(s_t + 1) - V_\phi(s_t) \\ A(s_t, a_t) = \sum_{t'=0}^{\infty} \lambda^{t'} \delta_{t+t'} \end{cases} \quad (5)$$

其中,  $\delta_t$  是时间步  $t$  的时间差分误差,  $A(s_t, a_t)$  是遵循广义

$$\begin{cases} \mathcal{L}_\theta = \mathbb{E}_{(s_t, a_t) \sim M_\theta} \left[ \min \left( \frac{M_G(s_t, a_t)}{M_\theta(s_t, a_t)} A(s_t, a_t) \text{clip} \left( \frac{M_G(s_t, a_t)}{M_\theta(s_t, a_t)}, 1 - \epsilon, 1 + \epsilon \right) A(s_t, a_t) \right) \right] \\ \mathcal{L}_\phi = \mathbb{E}_{(s_t, a_t) \sim M_\theta} (V_\phi(s_t) - R_t)^2 \\ \mathcal{L}_{\text{ppo}} = \mathcal{L}_\theta + \mathcal{L}_\phi \end{cases} \quad (6)$$

其中,  $\text{clip}$  函数表示裁剪的概率比率, 其限制了概率比率的变化范围, 防止学习中更新步骤过大;  $\epsilon$  为可调节的超参数, 用于限制策略更新的幅度。

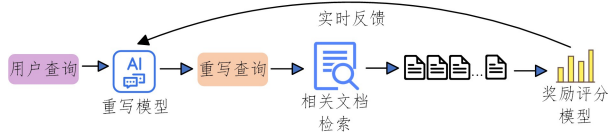


图3 反馈训练过程

Fig. 3 Feedback-driven training process

### 3.3 目标知识选择

为提高细粒度检索的效率, 评估查询是否超出当前的知识边界, 知识选择器  $\mathcal{M}_R$  定义了两种不同的反馈标记, 如表 2 所列。

表2 两种反馈标记

Table 2 Two types of feedback markers

标记类型	输入	输出	定义
Retrieve	$q', d_n$	{Yes, No}	决定是否需要检索
Relevant	$q', a_x$	{0, 0, 0, 1, ..., 1, 0}	$a_x$ 提供了多少有用信息

#### 3.3.1 检索标记

对于给定查询  $q$ , 通过判断模型是否需要使用外部文档来回答该查询, 即识别在知识边界内是否需要检索。近期研究表明<sup>[27]</sup>, LLMs 能够根据答案进行置信水平的自我评估, 知识选择模块利用自我评估机制来确定是否需要继续检索。

具体而言, 假设模型能够给出标准答案, 即对于特定输入进行多次采样后产生一个熵值较低的答案分布; 反之, 当模型不确定响应时, 则在答案分布中熵值会较高。评估答案分布的熵值可以表示如下:

$$\begin{aligned} H(A|q) &= \mathbb{E}_{a \sim \mathcal{M}_R} [-\log \text{Pr}(A|q)] \\ &= \mathbb{E}_{a \sim \mathcal{M}_R} \left[ -\sum_{a_x \in A} \log \text{Pr}(a_x | q, a_{<x}) \right] \end{aligned} \quad (7)$$

其中, 设置熵的阈值为  $\tau$ , 如果  $H(A|q)$  高于  $\tau$ , 则需要检索并生成检索标记; 反之, 则无需检索的标记。

#### 3.3.2 相关性标记

关于获得上下文与生成证据之间相关性软标签的方法, 受序列到序列模型<sup>[28]</sup>的启发, 根据文档与查询的相关程度来

优势估计 (Generalized Advantage Estimation, GAE)<sup>[25]</sup> 的优势函数。

最后, 对整个反馈训练的总损失  $L_{\text{ppo}}$  进行计算, 最终的损失函数由价值损失  $L_\phi$  和策略损失<sup>[26]</sup>  $L_\theta$  组成。在价值损失  $L_\phi$  的设计中, 使用均方误差 (Mean Squared Error, MSE) 来衡量预测值与实际回报之间的差异; 在策略损失  $L_\theta$  中, 通过比较新旧策略的概率比值与优势函数  $A(s_t, a_t)$  的乘积, 来反映在相同状态下, 新策略相对于旧策略采取某个动作的概率变化。通过将上述两种损失相结合的方式训练, 能够增强模型对语义理解的长期记忆功能, 具体计算式如式 (6) 所示:

生成目标词 “true” 和 “false” 的推理概率。

目标知识选择器  $\mathcal{M}_R$  根据输入查询与上下文的相关性分类标记 (“true” 和 “false”), 对生成 “true” 和 “false” 标记的概率进行归一化, 以获得最终相关性概率分数。将 “true” 标记的生成概率作为相关性标记的输出, 按式 (8) 计算分数:

$$S_{\text{true}} = \frac{P(\text{“true”} | q, a_x)}{P(\text{“true”} | q, a_x) + P(\text{“false”} | q, a_x)} \quad (8)$$

### 3.4 响应生成

响应生成器的功能是根据给定的查询段落对  $(q', A_x)$  生成正确响应。将任务理解和知识选择得到的结果, 包括重写的查询  $q'$ 、检索到的相关段落  $A_x$  以及相关性标记, 加入提示模板中, 以生成最终响应:

$$a_x = \mathcal{M}_G(q', A_x) \quad (9)$$

提示模板的设计对响应生成的性能有着重要的影响。本文设计的响应生成提示模板如图 4 所示。在响应生成模块中, 将查询指令分为两部分: 1) 明确任务和要求: 强调  $\mathcal{M}_G$  需仔细阅读并理解查询问题的含义, 避免出现响应偏离用户意图的情况。2) 引用相关上下文支撑回答: 要求  $\mathcal{M}_G$  参考相关片段进行回答, 使生成答案不仅是主观推测, 更是有客观依据的。

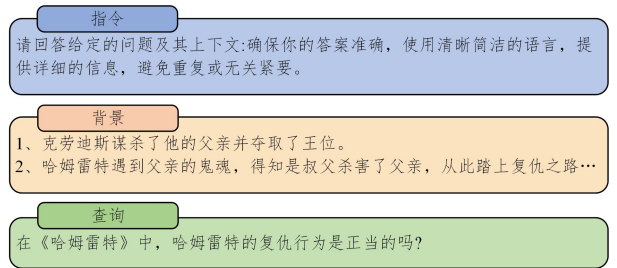


图4 响应生成使用的提示模板

Fig. 4 Prompt templates used for response generation

## 4 实验设置

### 4.1 数据集构建

为验证所提方法的有效性和通用性, 在抽取式阅读理解数据集和复杂多跳问答数据集上对 DF-RAG 和各类基准进行评估。

抽取式阅读理解数据集包含了 Natural Question (NQ)<sup>[29]</sup>。该数据集是一个开放域长尾分布的问答数据集,由 Google 收集的 30 万个自然产生的问题答案对构成,要求模型具备处理长文本以及理解其中逻辑关系和语境的能力。

复杂多跳问答数据集涵盖 PopQA 数据集<sup>[30]</sup>、TriviaQA (TQA)数据集<sup>[31]</sup>和 HotpotQA 数据集<sup>[32]</sup>。

TriviaQA(TQA)是华盛顿大学发布的大规模阅读理解数据集,包含超过 65 万个问题-答案-证据三元组。其中 9.5 万个问题-答案对由琐事爱好者编写,每个问题平均有 6 个独立收集的证据文件,为问答提供了高质量的保障。

PopQA 是一个开放领域的问答数据集,重点聚焦于大型语言模型在处理受欢迎和不受欢迎实体方面的能力,为探索和提升语言模型在复杂多变信息环境中的问答能力提供了独特视角。该数据集包含 14 000 个问题,在覆盖长尾实体的同时考虑了问题的流行度得分。

HotpotQA 由哈佛大学和斯坦福大学联合构建,是一个大型多跳问答数据集,旨在推动机器智能在理解和解答复杂问题方面的进步,是目前主流的用于提升模型理解和解答复杂问题能力的数据集。该数据集由约 10 万个多跳复杂问题以及对应的答案对组成,同时包含佐证证据用于解释答案的来源。

## 4.2 实验基准方法

在 TQA, PopQA, HotpotQA 和 NQ 数据集上对 DF-RAG 进行了实验,并将其与以下 4 种不同类型的基准进行比较:

1) 不具备查询重写策略和检索策略,只能直接处理查询问题的基准。Llama2-chat<sub>13B</sub><sup>[9]</sup>与 ChatGPT 两种模型均能够理解和回应复杂的查询,在处理和生成文本方面有着出色表现。

2) 具备动态检索策略,可根据查询内容确定检索时机的基准。FLARE<sup>[33]</sup>是一种多轮动态检索增强方法,通过训练 LLMs 学习特定的反馈标记,从而在推理过程中控制检索时机。SKR<sup>[34]</sup>能够利用内部知识判断模型能否回答给定查询。

3) 具备查询重写策略,能针对原始查询进行改进的基准。RaFe<sup>[35]</sup>利用一个公开可用的重新排序器对模型生成的重写查询进行排序,选择最优查询进行检索。GenRead<sup>[36]</sup>根据给定问题构建相应的上下文文档,然后利用语言模型基于原始问题和文档生成最终答案。

4) 具备多种策略的基准,能够有效增强模型检索性能。RA-ISF<sup>[37]</sup>通过迭代处理和自反馈机制,利用文本相关性评估、问题分解等策略,经由多轮对话优化搜索结果。RE-RAG<sup>[38]</sup>引入了一个相关性估计器,能够提供上下文之间的相对相关性和置信度,并且提出了一种新的解码策略,利用建议置信度选择依赖 LLMs 的参数知识而不是不相关的上下文。

## 4.3 评估标准

本文采用两种在检索增强生成实验中常用的评估指标对模型进行评估。对比实验以模型生成的答案是否涵盖标准答案为依据,使用精准匹配(Exact Matching, EM)作为标准指

标来评估模型的准确性;同时采用 Recall@K 作为评估指标,通过评估检索到的 K 个相关文档中包含标准答案的比例,检验模型的检索性能。

## 4.4 实验环境设置

本文实验均在一台配备 4 块 V100 显卡和共计 128GB 显存的机器上进行。将 2020 维基百科数据集<sup>[39]</sup>作为大规模语料库进行检索,以获取不同主题的文档,并采用 Contriever-MS-MARCO<sup>[40]</sup>进行相关文档检索。将 GLM-4 作为实验的黑盒大语言模型,且其在查询理解器的强化训练中充当重写器和阅读器。在模型预训练阶段,分别从每个数据集中随机抽取 2 000 个问答对,使用 GLM-4 进行查询重写和标签标注(相关性和是否检索)。使用 T5-large 模型作为查询理解器的基础模型,通过 PPO 算法进行反馈训练。将 epoch 设置为 2,批量大小设为 32,学习率设为  $5 \times 10^{-5}$ ,限幅范围参数  $\epsilon$  和 KL 发散系数  $\beta_{KL}$  均默认设置为 0.2。使用黑盒大语言模型生成的标签进行微调,并测试了 Llama2<sub>13B</sub> 作为知识选择器和响应生成器基础模型的性能。训练期间采用  $3 \times 10^{-4}$  的学习率,将 epoch 设置为 2,批量大小设置为 16,为使检索更加地频繁,将检索熵的阈值  $\tau$  设置为 0.2。具体的实验参数设置如表 3 所列。

表 3 实验参数设置

Table 3 Experimental parameter configuration

模型	参数	数值
T5-large	BatchSize	32
	学习率	$5 \times 10^{-5}$
	epoch	2
	KL 发散系数	0.2
Llama2 <sub>13B</sub>	BatchSize	16
	学习率	$3 \times 10^{-4}$
	epoch	2
	检索阈值	0.2

## 5 实验与结果分析

### 5.1 主要结果

#### 5.1.1 基线模型实验对比分析

为验证 DF-RAG 的有效性,将其与近 3 年内带有检索模块及重写模块的模型进行对比。对比实验结果如表 4 所列。

表 4 在 4 种不同数据集上的基线对比实验结果

Table 4 Baseline comparison results on four different datasets

方法	TriviaQA	PopQA	HotpotQA	NQ
Llama2-chat <sub>13B</sub>	60.9	20.9	15.8	22.8
ChatGPT	67.5	35.4	30.2	29.2
FLARE	56.7	—	14.9	—
SKR	55.4	37.2	18.9	20.8
RaFe	53.0	—	19.8	32.7
GenRead	68.0	34.7	—	—
RA-ISF	<b>71.4</b>	36.1	26.9	31.3
RE-RAG	46.7	—	—	33.9
DF-RAG	69.4	<b>37.8</b>	<b>31.5</b>	<b>34.9</b>

从表 4 能够看出,所提方法在 3 个数据集上的表现要明显优于其他方法。在与直接检索基准进行比较时,DF-RAG

的准确率平均提高了 7.1 个百分点,这表明有效的策略能够帮助模型更准确地预测出正确答案,充分证明了 DF-RAG 的有效性。

与具备动态检索策略的基准相比,DF-RAG 在性能上平均提高了 5.5 个百分点。实验结果表明,不同的检索时机选择策略对模型的准确率有着极大影响。若检索增强在错误的地方进行,不但不会提高检索生成的质量,反而还会影响模型的判断。实验发现,基于固定规则的检索方法(如 FLARE 和 SKR)并不总是优于直接检索问答的方法,因此应针对大语言模型的信息需求来激活对外部知识的调用。

与具备重写策略的基准相比,DF-RAG 展现出了更高的性能,准确率平均提升了 5 个百分点。这表明在处理模糊查询和复杂查询时,本文方法能够更有效地识别并查询每个问题的完整语义。

与具备多种策略的基准相比,DF-RAG 在 PopQA, NQ 和 HotpotQA 数据集上的表现均更优,但在 TriviaQA 数据集上的性能比 RA-ISF 差。经分析发现,当进行查询重写时,复杂问题无法补偿全面的参数知识,反而会带来噪声影响。大语言模型对查询进行重写后,重写语句的意思可能会偏离原始意思。此外,在把控检索时机时,复杂多跳问题往往会增加模型对不相关内容筛查的负担,极易产生幻觉。

综上所述,本文提出的动态细粒度检索增强生成方法 DF-RAG 在 4 种不同开放领数据集上均表现出色,实验结果有力证明了该方法在查询重写和检索增强任务中的有效性。

### 5.1.2 反馈标签效果分析

本节实验在模型检索任务中加入两种反馈标签(外部检索标签和相关性标签)的标注,根据反馈标签在不同模型中的运用情况做出相应的判断,以探究每个反馈标签对模型性能的影响。具体而言,采用了直接进行检索生成、仅采用外部检索标签、仅采用相关性标签和同时采用两种反馈标签 4 种不同的模型进行了对比实验。为了探索每个标签的效果,在 TriviaQA 数据集上进行对比实验,评估检索到的文档的召回率和根据文档回答的准确率,以比较它们在检索生成任务中的效果。这种细致的对比分析有助于理解不同检索标签在检索生成任务中的作用和效果,为未来的研究和应用提供帮助。

根据图 5 和图 6 的实验结果,在检索生成任务中选用基于两种反馈标签的模型表现最佳,验证了两种反馈标签的有效性。在模型召回率对比分析中,在采用 Top-100 相关文档情况下,外部检索标签与仅使用直接检索相比,召回率分数提升了 13.8%,而与使用两种反馈标签相比,召回率分数略低。实验结果体现出,采用外部检索的方式进行响应生成能有效提高提示指令中上下文的相关性。对于仅使用相关性标签,与其他 3 种基线模型相比,召回率比直接检索高出 1.6 个百分点,证明了相关性标签的有效性;但与其他两种基线模型的召回率相差过大,造成这种现象的原因是初次检索的维基百科语料库与 TriviaQA 开放数据集上的文档相关性较低。在模型准确率对比分析中,3 种使用反馈标签的模型比直接检索生成模型在准确率上分别提升了 13.9%,12.1%和 3.6%,进一步验证了将两种反馈标签作用在模型上对模型性能的提升有显著贡献。

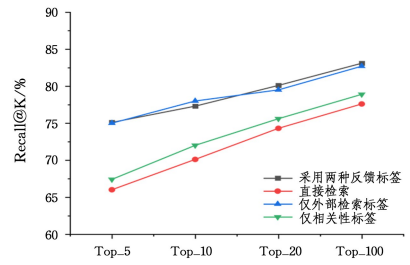


图 5 模型召回率对比分析

Fig. 5 Comparison and analysis of model recall rates

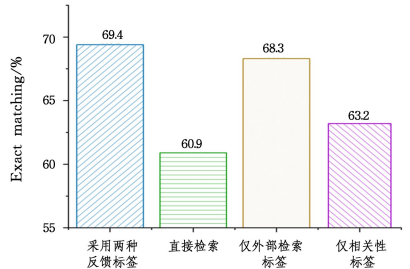


图 6 模型生成准确率对比分析

Fig. 6 Comparison and analysis of model generation accuracy

### 5.1.3 重写模型效率对比分析

为验证本研究提出的重写模型的有效性,在查询重写任务上将其与基于重写查询方法的模型进行对比。在 HotpotQA 和 PopQA 数据集上进行对比实验的结果如表 5 所列。实验数据显示,查询理解器和调用 GLM-4 黑盒 LLMs 在进行查询重写任务后,模型在响应生成的准确率和召回率上都有所提升,表明该方法在提升模型性能方面是有效的。

表 5 重写查询模型对比分析

Table 5 Comparison and analysis of query rewriting models

模型	方法步骤	HotpotQA		PopQA	
		EM	Recall	EM	Recall
DF-RAG	直接检索生成	27.3	81.2	32.4	81.1
	查询重写后检索生成	29.6	84.7	35.2	82.3
GLM-4	直接检索生成	28.6	83.5	34.2	83.9
	查询重写后检索生成	29.3	84.0	37.3	85.7

具体而言,无论是调用黑盒 LLMs 作为重写器还是使用训练过的重写器,在查询重写任务上都能带来模型性能的提升。在 HotpotQA 和 PopQA 数据集上,两种模型在重写查询后的响应生成的准确率均比直接查询高,分别提升了 5.4% 和 8.8%,召回率分别提升了 2.4% 和 1.7%。在 PopQA 数据集上,采用微调训练过的 T5-large 重写模型性能略低于调用黑盒 LLMs 模型,造成此结果的原因有两方面:一方面,在检索相关文档上本文方法使用的是维基百科语料库,相关文档数量要远低于调用 LLMs;另一方面,黑盒 LLMs 的训练规模和模型数据都是远超轻量化的预训练模型的。在 HotpotQA 数据集上,查询理解器的效率与调用黑盒 LLMs 进行查询重写相差不大,甚至还略高,这表明在处理复杂查询任务时,黑盒 LLMs 对问题的过度泛化和噪声的引入,影响了模型对问题的语义理解能力。因此,在处理特定领域和特定任务的查询重写时,使用结合强化学习训练和监督微调的查询

重写模型能提高模型对用户需求的理解能力。

#### 5.1.4 模型超参数调优实验分析

本小节实验中,对于检索模块,对调用外部检索的阈值超参数 $\tau$ 的调整采用逐步优化的方法,对 $\tau \in \{0, 0.1, 0.2, \dots, 1\}$ 进行调整,最终找出性能最佳的阈值超参数,并对模型的检索效率和准确性进行平衡分析。

具体而言,在 HotpotQA 数据集上评估了模型在不同阈值 $\tau$ ( $\tau$ 值越大表明进行外部检索的次数越少)下的性能,以此分析这种根据熵的阈值来判定是否进行外部检索的方法是如何影响模型整体准确性和检索频率的。图 7 的实验结果显示,随着检索阈值的增加,检索频率和精准匹配率都呈现出下降趋势,这表明,在更高阈值下进行检索时,模型更倾向于能够完全理解查询的语义并有相应的上下文能够帮助回答问题,从而减少了不必要的外部检索。当检索阈值在 0~0.2 时,精准匹配率的下降趋势相对缓慢,而检索频率下降较快;当检索阈值为 0.2 时两者的差距最大,即使用相对小的检索频率能让模型得到更高的效率;之后随着检索阈值的进一步增加,精准匹配率与检索频率的差距逐渐减小。在 0.4~0.8 的区间时,精准匹配率甚至超过检索频率的下降速度,这对模型效率来说是负增长的。因此,在检索阈值的设置上,过低的阈值会导致产生无效的外部检索,也会延长模型的响应时间;而过高的阈值则会导致错过一些有用的信息。因此,需要在检索频率和准确率之间找到一个平衡点,以提高模型的整体性能。

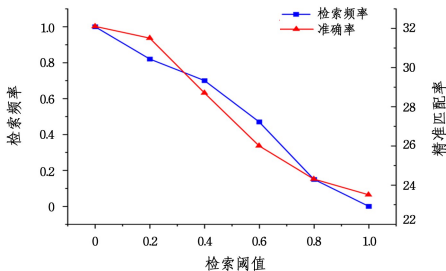


图 7 检索效率与生成准确性的对比分析

Fig. 7 Comparison and analysis of retrieval efficiency and generation accuracy

## 5.2 消融实验

本文对模型主要进行了两个方面的改进:首先引入查询重写模块帮助模型更全面地理解复杂查询中所包含的问题,从而增强模型对用户需求的理解;随后,使用按需调用外部检索的方法,以一种更准确且灵活的方式来提高模型面对超出知识边界问题的解决能力。为了评估 DF-RAG 不同组成部分对整体性能的影响,将其拆分为以下两种模型。

**查询理解器+响应生成器:**仅进行查询重写,不进行检索时机的选择。该模型旨在研究针对模块动态检索的效率及其对模型性能的影响。

**知识选择器+响应生成器:**不对查询进行重写,直接进行检索时机的选择。通过实验研究重写模型对整体性能的影响以及其在语义分析方面的能力。

所有消融实验均使用 T5-large 和 Llama2<sub>13B</sub> 作为基础模型。在 PopQA 和 HotpotQA 数据集上进行了测试,并使用文

档的召回率和从文档中检索到答案的精准匹配率来评估两个模块的效果,最终将模型评估结果与直接检索和 DF-RAG 进行了比较。

#### 5.2.1 任务理解效果对比分析

图 8 展示了查询理解器+响应生成器模型在 HotpotQA 数据集上与不同架构模型的 Recall@K 比较结果。与直接检索模型相比,重写查询后的检索通过对查询进行补充和问题的理解提升了模型的性能。实验结果充分验证了对用户查询进行全面语义分析的方法能够帮助 LLMs 有效地处理模棱两可和复杂查询。在与采用重写模块和检索模块的模型比较中,仅进行查询后重写方法的 Recall@K 结果均低于 DF-RAG,验证了查询理解器在模型框架中的有效性。

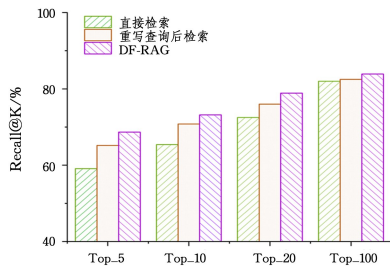


图 8 HotpotQA 数据集上前 100 个文档的召回率

Fig. 8 Recall rate of the top 100 documents on the HotpotQA dataset

本节实验中,将查询理解器的提示方法与删除信息性要求的精简重写指令、RaFe 方法中使用的提示指令进行比较,以分析本文方法的提示指令在查询重写方面的能力。实验结果如图 9 所示,相比于精简指令和其他重写方法中的指令,查询理解器所用的提示指令对整个模型有着更高的性能提升。这表明,在提示中使用更符合问题需求和任务场景的重写提示,会显著提升查询的优化效果。本文方法的提示侧重于解析查询中的专业术语以及尽可能多地提供有用信息,将查询转化为语义上增强、清晰和基于标准术语的重写查询问题,并生成多个不应相互重叠的查询,做到从不同粒度进行分析,以此提高模型的整体性能。

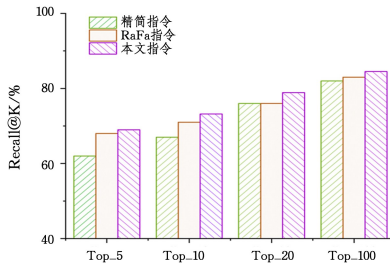


图 9 采用不同提示指令的召回率对比分析

Fig. 9 Comparison and analysis of recall rates using different prompt instructions

#### 5.2.2 知识选择作用分析

仅采用知识选择器和响应生成器的消融实验结果如表 6 所列。与直接检索相比,其进行检索时机选择的精确率提高约 13.95 个百分点,表明利用外部知识来解决超出当前知识边界的问题,可显著提升响应生成的性能。与 DF-RAG 相比,其精确度降低了 2.35 个百分点。结合上一节实验,充分

证明查询理解器与知识选择器能提升模型性能,而两者结合使用时,效果更加显著。实验结果有力证实,查询理解器和知识选择器两个子模块不仅增强了模型对复杂查询的理解能力,还在提高检索生成质量的同时减轻了不相关检索段落引起的负面影响。

表 6 3 种模型的精准匹配结果

Table 6 Exact matching results of three models

(%)		
方法	HotpotQA	PopQA
直接检索	15.8	20.9
动态检索策略	28.4	36.2
DF-RAG	31.5	37.8

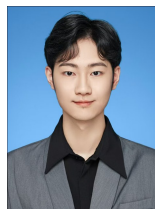
**结束语** 如何更好地让模型理解模糊复杂的查询以及动态选择检索的时机,进而实现细粒度检索,一直是极具应用价值的研究课题。本文提出一种新的模块化检索增强生成框架,即动态细粒度检索增强生成(DF-RAG),通过增强查询重写和检索时机的选择来提高响应生成的准确率。DF-RAG 借助黑盒大语言模型对查询中的专业术语进行扩充,将其重写为更容易理解的查询,提高了模型检索的有效性;同时,在检索过程中充分利用外部知识来解决超出知识边界的信息,提高了模型的性能。实验结果表明,它在不同数据集和任务中均展现出优于其他各类基准的性能,但仍不可避免地存在幻觉问题。

在未来的工作中,为了全面研究 DF-RAG 框架的性能以提高模型对领域知识理解的性能,研究数据集将不再局限于开放域数据集,而会扩展到更多如数学推理、医学等垂直领域数据集,以探寻提升模型对查询语义理解和知识选择等能力的更好方法。另外,还计划研究不同的动态检索方法,以满足现阶段大语言模型的各种需求,最终实现对用户问题的快速检索。

## 参 考 文 献

- [1] TOM B, BENJAMIN M, et al. Language models are few-shot learners[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 1877-1901.
- [2] JOSH A, STEVEN A, SAND H A, et al. GPT-4 Technical Report[R]. OpenAI, 2023.
- [3] ADAM R, COLIN R, NOAM S. How Much Knowledge Can You Pack Into the Parameters of a Language Model[C]// 2020 Association for Computational Linguistics. 2020: 5418-5426.
- [4] ORI R, YOAV L, ITAY D, et al. In-context retrieval-augmented language models[J]. *Transactions of the Association for Computational Linguistics*, 2023, 11: 1316-1331.
- [5] ALEX M, AKARI A, VICTOR Z. When Not to Trust Language Models: Investigating Effectiveness of Parametric and Non-Parametric Memories[C]// Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023: 9802-9822.
- [6] MA X B, GONG Y Y, HE P C, et al. Query Rewriting for Retrieval-Augmented Large Language Models[C]// Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023: 5303-5315.
- [7] AKARI A, SEWON M, ZHONG Z X. Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection[C]// Twelfth International Conference on Learning Representations. 2023: 41-46.
- [8] YU W H, ZHANG H M, PAN X M, et al. Chain-of-Note: Enhancing Robustness in Retrieval-Augmented Language Models [J]. arXiv:2307.09288, 2023.
- [9] HUGO T, LOUIS M, KEVIN S, et al. Llama 2: Open foundation and fine-tuned chat models[J]. arXiv:2307.09288, 2023.
- [10] ZHAO W Y X, ZHOU K, LI J Y, et al. A Survey on Evaluation of Large Language Models[J]. *Association for Computing Machinery*, 2024, 39: 2157-6904.
- [11] KURT S, SPENCER P, CHEN M Y, et al. Retrieval Augmentation Reduces Hallucination in Conversation [C] // Empirical Methods in Natural Language Processing 2021. 2021: 3784-3803.
- [12] XU X C, GOU Z B, WU W Q, et al. Long Time No See! Open-Domain Conversation with Long-Term Persona Memory[C]// Findings of the Association for Computational Linguistics: ACL 2022. 2022: 2639-2650.
- [13] WANG H R, HU M D, DENG Y, et al. Large Language Models as Source Planner for Personalized Knowledge-grounded Dialogues[C]// Findings of the Association for Computational Linguistics: EMNLP 2023. 2023: 9556-9569.
- [14] PATRICK L, ETHAN P, ALEKSANDRA P, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 9459-9474.
- [15] CHENG D X, HUANG S H, BI J Y, et al. UPRISE: Universal Prompt Retrieval for Improving Zero-Shot Evaluation[C]// Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023: 12318-12337.
- [16] GAO L Y, MA X G, LIN J M, et al. Precise Zero-Shot Dense Retrieval without Relevance Labels [C] // Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023: 1762-1777.
- [17] SHI W J, SEWON M, MICHHIRO Y, et al. REPLUG: Retrieval-Augmented Black-Box Language Models [C] // Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2024: 8371-8384.
- [18] OUYANG L, JEFFREY W, XU J, et al. Training language models to follow instructions with human feedback [C] // Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022. 2022: 27730-27744.
- [19] JOHN S, FILIP W, PRAFULLA D, et al. Proximal Policy Optimization Algorithms[R]. OpenAI, 2023.
- [20] WU Z Q, HU Y S, SHI W J, et al. Fine-Grained Human Feedback Gives Better Rewards for Language Model Training [C] // 37th Conference on Neural Information Processing Systems. 2023: 918-944.
- [21] XU W W, CAI D, ZHANG Z S, et al. Reasons to Reject? Aligning Language Models with Judgments [R]. Tencent AI Lab, 2023.
- [22] COLIN R, NOAM S, ADAM R, et al. Exploring the limits of

- transfer learning with a unified text-to-text transformer[J]. The Journal of Machine Learning Research, 2020, 67: 1532-4435.
- [23] YE F H, FANG M, LI S H, et al. Enhancing Conversational Search; Large Language Model-Aided Informative Query Rewriting[C] // Findings of the Association for Computational Linguistics; EMNLP 2023. 2023; 5985-6006.
- [24] RAJKUMAR R, PRITHVIRAJ A, KIANTE B, et al. Is reinforcement learning(not) for natural language processing; Benchmarks, baselines, and building blocks for natural language policy optimization[C] // The Eleventh International Conference on Learning Representations. 2023; 5687-5711.
- [25] JOHN S, PHILIPP M, SERGEY L, et al. High-dimensional continuous control using generalized advantage estimation[C] // 4th International Conference on Learning Representations. 2016; 571-584.
- [26] ZHENG R, DOU S, GAO S Y, et al. Secrets of RLHF in large language models part1; PPO[R]. Fudan NLP Group, 2023.
- [27] SAURAV K, CONERLY T, ASKELL A, et al. Language models(mostly) know what they know[J]. Transactions of the Association for Computational Linguistics, 2022, 8: 423-438.
- [28] NOGUEIRA R, JIANG Z Y, PRADEEP R, et al. Document ranking with a pretrained sequence-to-sequence model [C] // Findings of the Association for Computational Linguistics; EMNLP 2020. 2020; 708-718.
- [29] TOM K, JENNIMARIA P, OLIVIA R, et al. Natural Questions; A Benchmark for Question Answering Research [C] // Transactions of the Association for Computational Linguistics. 2019; 452-466.
- [30] ALEX M, AKARI A, VICTOR Z, et al. When Not to Trust Language Models; Investigating Effectiveness of Parametric and Non-Parametric Memories[C] // Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023; 9802-9822.
- [31] MANDAR J, EUNSOL C, DANIEL W, et al. TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension[C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017; 1601-1611.
- [32] YANG Z L, QI P, ZHANG S Z, et al. Manning. HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering [C] // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018; 2369-2380.
- [33] HARSH T, NIRANJAN B, TUSHAR K, et al. Interleaving retrieval with chain-of-thought reasoning for knowledge-intensive multi-step questions[C] // 2023 Association for Computational Linguistics. 2023; 10014-10037.
- [34] WANG Y L, LI P, SUN M S, et al. Self-knowledge guided retrieval augmentation for large language models[C] // Findings of the Association for Computational Linguistics; EMNLP 2023. 2023; 10303-10315.
- [35] MAO S Y, JIANG Y, CHEN B L, et al. RaFe: Ranking Feedback Improves Query Rewriting for RAG[C] // Findings of the Association for Computational Linguistics; EMNLP 2024. 2024; 884-901.
- [36] YU W H, DAN I, WANG S H, et al. Generate rather than Retrieve; Large Language Models are Strong Context Generators [C] // International Conference on Learning Representations. 2023.
- [37] LIU Y M, PENG X Y, ZHANG X H, et al. RA-ISF: Learning to Answer and Understand from Retrieval Augmentation via Iterative Self-Feedback[C] // Association for Computational Linguistics. 2024; 4730-4749.
- [38] KISEUNG K, JAY-YOON L. RE-RAG: Improving Open-Domain QA Performance and Interpretability with Relevance Estimator in Retrieval-Augmented Generation[C] // Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. 2024; 22149-22161.
- [39] VLADIMIR K, BARLAS O, SEWON M, et al. Dense passage retrieval for open-domain question answering[C] // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020; 6769-6781.
- [40] GAUTIER I, PATRICK L, MARIA L, et al. Atlas: Few-shot learning with retrieval augmented language models[J]. Journal of Machine Learning Research, 2023, 24: 1-43.



**ZHANG Haoran**, born in 2001, post-graduate. His main research interests include natural language processing and data mining.



**HAO Wenning**, born in 1971, Ph.D, professor, Ph.D supervisor. His main research interests include data mining and machine learning.

(责任编辑:柯颖)