

在知识密集环境中的引导式导航：使用引导图的结构化语义探索

Dehao Tao¹, Guangjie Liu², weizheng¹, Yongfeng Huang¹, Minghu jiang¹

¹ Tsinghua University

² Xinjiang University

Abstract

尽管大型语言模型 (LLMs) 展现出强大的语言能力, 但他们对静态知识的依赖和不透明的推理过程限制了其在知识密集型任务中的表现。知识图谱 (KGs) 提供了一种有前途的解决方案, 但当前的探索方法面临一个基本的权衡: 问题引导的方法由于粒度不匹配而导致冗余探索, 而线索引导的方法未能在复杂场景中有效利用上下文信息。为了解决这些限制, 我们提出了引导图引导的知识探索 (GG-Explore), 一种创新框架, 通过引入中介引导图来桥接非结构化查询和结构化知识检索。引导图通过抽象目标知识的结构而定义了检索空间, 同时保持更广泛的语义上下文, 从而实现精确和高效的探索。基于引导图, 我们开发了: (1) 结构对齐, 可以在不增加 LLM 负担的情况下过滤不兼容的候选项, 以及 (2) 情境感知剪枝, 确保与图约束的语义一致性。大量实验表明, 我们的方法在复杂任务上实现了卓越的效率和优于现有技术表现, 同时在较小的 LLM 上保持强大的性能, 展示了实际价值。

1 介绍

大型语言模型 (LLM) 在各种自然语言任务中展示了卓越的能力, 包括问答 (Wang et al. 2024a; Li, Zhou, and Dou 2024; Zhao et al. 2024)、文本生成 (Ji et al. 2024; Chen et al. 2023, 2022; Gong and Sun 2024) 和推荐系统 (Zhang et al. 2023, 2024; Wu et al. 2024; Wang et al. 2024c)。通过利用深度学习和庞大的训练数据, 它们实现了类似人类的语言流畅性。然而, 它们在训练后知识是静态的, 限制了实时更新; 它们可能会生成似是而非但不正确的回答 (“幻觉”) (Bang et al. 2023; Ji et al. 2023; Luo et al. 2023b)。整合外部知识源 (例如, 知识图谱) 提供了有前途的解决方案 (Zhang et al. 2019a; Yao, Mao, and Luo 2019; Wang et al. 2021; Luo et al. 2023a), 强调了在这一不断发展的领域中进一步研究的必要性。

一些研究人员尝试直接采用大型语言模型 (LLM) 将问题转换为数据库查询语句, 以从知识图谱中检索相关知识 (Hu et al. 2023; Wang et al. 2023)。虽然这种方法很直接, 但生成的查询可能并不总是与知识图谱的结构良好对齐。一种可行的替代方法是为 LLM 提供知识图谱上下文, 使用问题引导的多轮迭代逐步识别相关知识 (Li et al. 2023d; Zhao et al. 2023; Sun et al. 2024)。这种问题引导的方法从根本上遭受自然语言查询与知

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

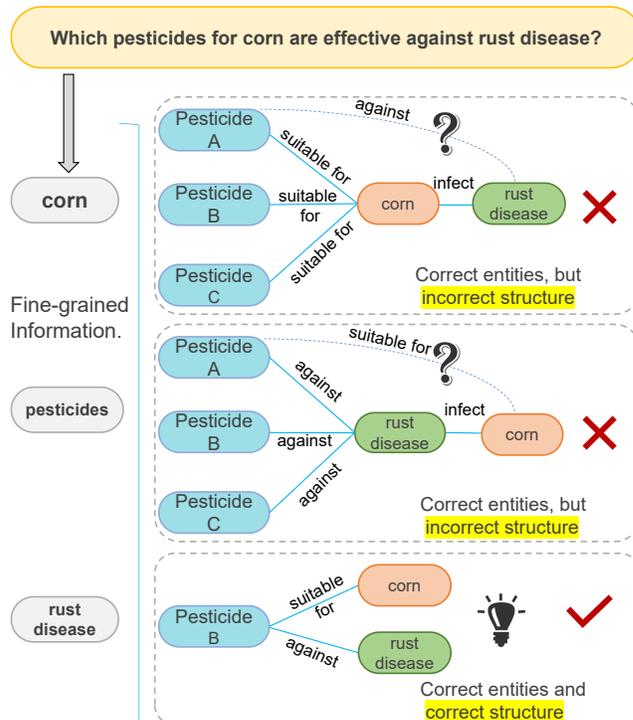


Figure 1: 仅依赖于细粒度信息可能无法区分包含相同实体但具有不同关系结构的路径。

识图谱实体之间的粒度不匹配之苦, 导致两个关键的限制: (1) 由于语义不对齐导致的知识定位不精确, 以及 (2) 由于冗余图探索造成的计算低效。

FiSKE 从查询中提取细粒度关键词用于有状态知识检索, 避免了无状态方法中的冗余 (Tao et al. 2025)。然而, FiSKE 会丢弃问题中的其他上下文信息, 以提取细粒度关键词。虽然这种简化对于结构简单的场景证明有效, 当处理知识密集型环境中复杂的结构关系时则显得不足, 正如图中所示 1。

在本文中, 我们提出了引导图引导的知识探索 (GG-Explore), 这是一种将非结构化查询与结构化知识检索之间的差距弥合的新范式。该新范式引入了一个中介的引导图构建阶段。引导图定义了检索空间, 要求所有提取的知识与其结构和语义框架对齐。它抽象化了目标

知识的结构，同时保留了更广泛的语义，从而以精确的方式指导检索。与现有的方法仅基于关系和实体进行修剪不同，我们的方法利用指导图提供的结构和上下文信息，设计了结构对齐和上下文感知的修剪机制。这些组件从结构和语义视角选择性地过滤和匹配目标知识。具体来说，结构对齐有效地消除了与指导图结构不一致的候选项，而无需模型计算，而上下文感知修剪则精确地识别与指导图中上下文信息一致的知识。

本文的贡献如下：

1. 我们提出了一种新颖的问题引导图知识范式，该范式引入了一个中间的引导图构建阶段，以连接非结构化查询和结构化知识检索。引导图通过抽象目标知识的结构而定义检索空间，同时保留更广泛的语义。
2. 我们开发了基于指导图的两个核心机制：结构对齐和上下文感知剪枝。这些组件协同工作，从结构和语义两个角度精确过滤和匹配目标知识，显著提高检索效率和准确性。
3. 大量实验证明了我们方法的有效性，具有三个主要优势：(1) 与现有方法相比，答案准确率显著更高，特别是对于复杂问题；(2) 高效，仅需最少的 LLM 调用和符号使用；(3) 在小参数 LLM 上表现优异，突显出强大的实用价值。

2 相关工作

2.1 使用提示进行 LLM 推理

近年来，在 LLMs 的推理能力方面取得的进展主要源于各种提示技术，这些技术旨在提高其在复杂任务中的表现。Decomp (He et al. 2021) 将推理任务分解为可管理的子任务，从而实现逐步解决问题。链式思维 (CoT) (Wei et al. 2022a) 及其衍生方法，例如思维树 (ToT) (Yao et al. 2023)、思维图 (GoT) (Besta et al. 2024) 和思维记忆 (MoT) (Li and Qiu 2023)，在鼓励 LLMs 生成中间推理步骤方面起到了重要作用，从而增强其认知过程。计划与解决 (Wang et al. 2024b) 提示 LLMs 根据这些计划制定计划并进行推理。

2.2 基于知识图谱增强的大型语言模型的问答

知识图谱问答 (KGQA) 随着大语言模型和知识图谱的整合而显著发展。早期的方法在预训练或微调期间嵌入知识图谱知识至大语言模型中 (Zhang et al. 2019b)，但这限制了解释能力和更新效率。为了增强推理能力，检索增强技术从知识图谱中检索相关事实 (Li et al. 2023a)。UniKGQA (Jiang et al. 2022) 在单一的大语言模型中统一了图谱检索和推理，实现了最先进的结果。最近的方法将知识图谱知识翻译成用于大语言模型的文本提示，在不牺牲知识图谱优势的情况下增强推理能力 (Li et al. 2023b)。例如，生成 SPARQL 查询或采样相关的三元组有助于大语言模型的推断 (Baek, Aji, and Saffari 2023)。知识图谱增强的大语言模型范式将大语言模型视为探索知识图谱的互动代理，从而提高推理能力 (Jiang et al. 2023a)。

3 方法

3.1 问题定义

知识图谱问答涉及查询结构化的知识图谱，以生成对自然语言问题的准确回答。在我们提出的方法中，我们首

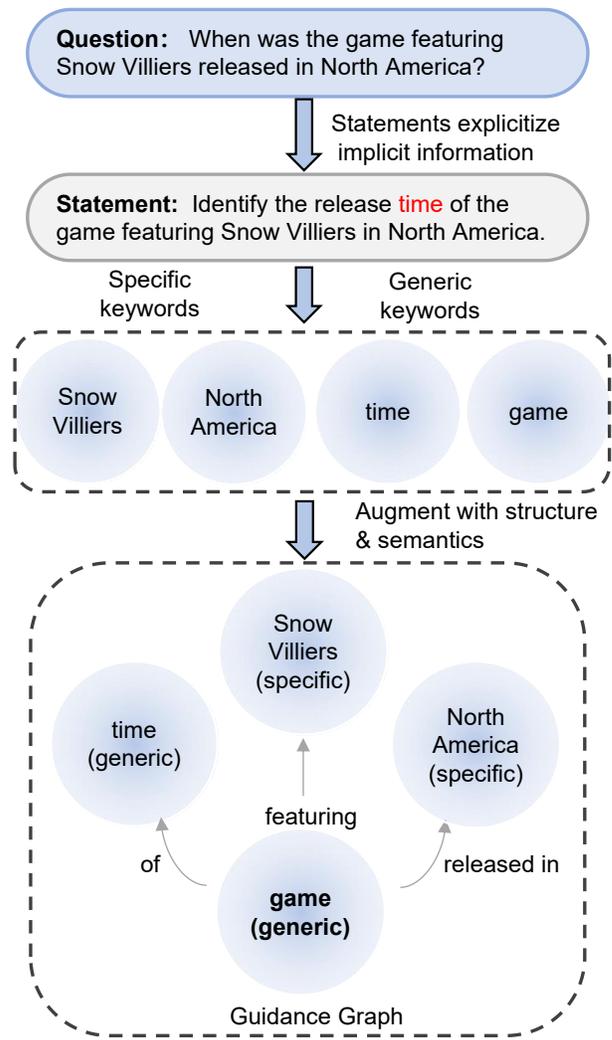


Figure 2: 指导图构建流程

先从问题中提取信息以构建一个指导图 $GG = (e, v)$ 。随后，在知识探索过程中，我们识别从 GG 到知识图谱中的实体和关系的所有节点 $clue_e$ 和边 $clue_r$ 的映射关系。最后，我们从这些映射中检索相应的知识，并提供给 LLM 进行问题解答。

3.2 指导图构建

我们的框架构建了图 2 所示的指导图。

我们使用大型语言模型来对问题进行层次化处理。首先，我们将省略性问题转化为完整的陈述句以解决信息不完整性问题。其次，我们从这些陈述中提取细粒度的元素，并将其分类为：(1) 特定关键词（如“北美”这样的命名实体）和 (2) 通用关键词（如“国家”这样的广义术语）。第三，我们挖掘关键词之间的语义关系，以加强指导图的结构表示。

为了建立图结构同时丰富语义信息，我们基于特定和通用关键词设计了以下生成规则：

1. 特定的关键词专门作为图节点（即三元组中的实体），而不是边缘。

2. 当一个通用关键字与一个特定关键字共同指代同一实体时，它作为三元组的关系而不是一个实体。
3. 对于引用同一实体的两个不同通用关键词，其中一个必须被指定为三元组中的关系。
4. 对于引用不同实体的相关关键词，我们构建三元组，其中关键词作为头/尾实体，它们的关联形成关系边。

这些规则确保指导图的结构和语义与问题的原始意图紧密对齐。使用的提示显示在附录中。

为了避免与知识图谱中的实体和关系混淆，我们将指导图中的实体和关系称为线索实体和线索关系。我们通过反复的知识探索逐步将线索实体和线索关系映射到知识图谱，最终识别出在结构和语义上都与指导图对齐的知识子图。所使用的提示在附录中显示。

起始点 知识探索从起点开始，并逐渐向外扩展。问题引导的方法倾向于将这些起点视为先验知识，而线索引导的方法尝试利用所有细粒度的线索作为起点。前者缺乏实用性，而后者可能会错误地将概念实体（例如，国家）视为起点，直接导致探索失败。

在我们的方法中，我们选择指导图中的特定关键词作为起始点。知识图谱通常包含与这些特定关键词同名的实体，可以通过基本的数据库操作检索到。一旦检索到任何一个起始点，知识探索就可以继续进行，而无需识别所有的特定关键词。在实验部分，我们研究了起始点检索成功对我们方法的影响。指导图中的线索实体与知识图谱中相应实体之间的映射保持在 $clue_mapping = \{clue : entities\}$ 中，其中包括作为该映射一部分的已检索到的起始点。

为确保生成的知识子图在结构和语义上与指导图一致，我们从指导图中选择线索实体作为目标 $next_e$ ，并在每次探索迭代之前确定其上下文 $(next_e, next_r, current_e)$ 或 $(current_e, next_r, next_e)$ 。选择标准如下：

- 在 Guidance Graph 中，存在 $(next_e, next_r, current_e)$ 或 $(current_e, next_r, next_e)$ ，其中 $current_e \in \{c | (c : e) \in clue_mapping\}$
- $clue_e \notin \{c | (c, e) \in clue_mapping\}$

多个有效的目标知识候选可能存在，它们的映射顺序对置换不变——随机选择就足够了。

为了确保检索到的知识子图与引导图之间的结构一致性，我们在每次知识探索迭代中对当前候选子图进行结构对齐。候选子图由两个部分组成：映射子图和候选实体集。映射子图包括来自 $clue_mapping = \{clue : entities\}$ 的所有映射实体，而候选实体集 E_{cand} 则包含知识图中与 $current_e$ 的映射实体相关联的所有实体，这些实体可以通过简单的数据库操作获得。

在只有 $current_e$ 拥有指向目标实体 $next_e$ 的边的情况下，候选子图自然与指导图的结构对齐。为了说明这点，我们假设指导图还包含一个额外的线索实体 $related_e$ ，它也与 $next_e$ 和 $current_e$ 相连。

我们从 $clue_mapping = \{clue : entities\}$ 中检索出两组映射实体：对应于 $current_e$ 的实体集 E_{curr} 和对应于 $related_e$ 的实体集 E_{rel} 。

对于候选集合 E_{cand} 中的每个实体，如果它在知识图中满足以下两个连接条件，我们认为它在结构上是有效的：(1) 与集合 E_{curr} 中至少一个实体有连接边，以及

(2) 与集合 E_{rel} 中至少一个实体有连接边。这个双重连接要求确保选择的候选在指导图中指定的同时维持与两个线索实体的结构关系。在结构上有效的候选实体集和已映射的子图共同形成一个结构上有效的候选子图。

在完成这轮知识探索并获得新的映射 $\{next_e : new_entities\}$ 后，将进行整体结构对齐。该对齐从新发现的映射向外传播，以涵盖所有现有映射。对于实体集 E_{curr} 或 E_{rel} 中与 $new_entities$ 中的实体不共享的任何实体，该实体将从映射集中移除。然后，将相同的连接性检查应用于连接到 $current_e$ 和 $related_e$ 的其他线索实体，以及之后完成的所有映射。此过程确保所有映射之间的相互连通性，从而维护指导图的整体连接结构。

上下文感知剪枝 结构上对齐的候选实体集可能仍包含无关的实体。为了解决这个问题，我们设计了一种上下文感知的修剪方法来优化候选选项，确保所选实体与指导图保持一致的上下文语义。

候选实体集可能非常庞大，因此逐一评估每个候选实体的上下文在计算上代价很高。由于知识图谱中的关系数量通常远小于实体数量，我们转而操作候选关系集 R 。这个集合由候选实体集和实体集 E_{curr} 之间的所有关系组成。

对于目标选择阶段确定的目标的 $(next_e, next_r, current_e)$ 或 $(current_e, next_r, next_e)$ ，我们通过顺序连接其组件来构建上下文短语。然后，指示 LLM 选择最合适的关系 $r \in R$ ，可以替换 $next_r$ ，同时保留目标短语的原始语义。这种替代方法确保了与 $next_r$ 的最佳语义匹配，同时保持上下文实体语义。

在获得映射关系 r 后，我们将 $E_{mapping}$ 定义为候选实体集中与 r 相连的实体集合，并用新的映射 $next_e : E_{mapping}$ 更新 $clue_mapping$ 。然后，我们执行上述整体结构对齐，去除 $clue_mapping$ 中的不连通实体。

我们的剪枝方法相较于现有的关系剪枝方法提供了两个关键优势：目标感知和上下文感知，使其在理论上更为精确。所提出的结构对齐方法不仅确保了已探索子图与指导图之间的结构一致性，而且在整体结构对齐过程中，无关的实体自动被消除，从而完全取代了传统实体剪枝方法的作用。

3.3 知识不完整下的动态分支选择

知识图通常表现出不完整性，其中并非所有特定关键字派生的线索实体都能映射到相应的节点，或者它们的关联关系可能无法映射。当由于缺少关系而识别了 $related_e$ 但直接约束修剪失败时，我们从 $(current_e, next_r, next_e)$ 和 $(related_e, related_r, next_e)$ 生成语义短语。LLM 选择更符合查询上下文的短语。选择 $related_e$ 会触发从它开始的新探索，而非选择分支的起点将在后续探索路径中被永久修剪。

4 实验

4.1 实验设置

数据集和评估指标。为了评估我们所提出范式的性能，我们选择了一个开源知识图谱和一个自主构建的图谱作为实验的外部知识库。开源知识图谱是 Freebase (Bollacker et al. 2008)。我们自主构建的图谱将很快公开发布，在本文中称为农业知识图谱。

Method	WebQSP		CWQ	
	partial match	complete match	partial match	complete match
without external knowledge				
IO prompt (Brown et al. 2020) w/DeepSeek-V3	63.3	35.3	44.8	44.8
COT (Wei et al. 2022b) w/DeepSeek-V3	70.5	41.4	46.7	46.7
with external knowledge				
ToG (Sun et al. 2024) w/Llama3-8B-Instruct	55.6	32.3	-	-
PoG (Chen et al. 2024) w/Llama3-8B-Instruct	63.4	34.4	-	-
FiSKE (Tao et al. 2025) w/Llama3-8B-Instruct	70.8	40.4	-	-
StructGPT (Jiang et al. 2023b) w/GPT-3.5	72.6	-	54.3	54.3
KB-BINDER (Li et al. 2023c) w/GPT-3.5	74.4	-	-	-
ToG (Sun et al. 2024) w/GPT-3.5	76.2	-	57.1	57.1
PoG (Chen et al. 2024) w/DeepSeek-V3	81.9	60.7	55.7	55.7
FiSKE (Tao et al. 2025) w/DeepSeek-V3	82.5	61.1	50.2	50.2
Ours w/Llama3-8B-Instruct	79.3	54.1	56.7	56.7
Ours w/DeepSeek-V3	81.8	64.5	71.8	71.8

Table 1: WebQSP 和 CWQ 的结果。在 CWQ 中，每个问题只有一个答案，因此部分匹配等同于完全匹配。

Method	1-Hop Query		multi-Hop Query	
	partial match	complete match	partial match	complete match
IO prompt (Brown et al. 2020) w/DeepSeek-V3	35.8	22.1	12.4	6.6
CoT (Wei et al. 2022b) w/DeepSeek-V3	37.2	23.3	12.8	7.1
PoG (Sun et al. 2024) w/DeepSeek-V3	67.7	62.6	23.9	13.0
FiSKE (Tao et al. 2025) w/DeepSeek-V3	71.1	62.8	28.3	12.9
Ours w/DeepSeek-V3	89.9	75.8	81.8	56.8

Table 2: 农业图的结果。

Dataset	Method	LLM Call	Input Token	Output Token	Total Token
WebQSP	PoG	11.3	6590.2	427.0	7017.2
	FiSKE	7.9	3079.0	1379.7	4458.7
	Ours	8.6	3264.5	584.0	3848.5
CWQ	PoG	23.4	15483.4	694.8	16178.2
	FiSKE	9.4	3578.8	1828.7	5407.5
	Ours	10.2	4052.0	708.6	4760.6
agr-one-hop	PoG	5.6	2233.9	217.4	2451.4
	FiSKE	5.9	-	-	-
	Ours	4.6	1000.0	294.4	1294.4
agr-multi-hop	PoG	6.9	3243.6	379.6	3623.3
	FiSKE	13.1	-	-	-
	Ours	4.1	975.5	380.7	1356.2

Table 3: 不同数据集的效率比较。agr 表示农业数据集。

Dataset	Method	LLM Call	Input Token	Output Token	Total Token	Partial match	Complete match
CWQ	Ours	10.2	4052.0	708.6	4760.6	0.7179	0.7179
Filtered-CWQ	Ours	12.0	4875.4	769.9	5645.3	0.7460	0.7460

Table 4: 关于起始点的实验结果。在 Filtered-CWQ 中，没有可识别起始点的问题被移除。在我们的方法中，这些问题直接由 LLM 回答。

Variant	partial match	LLM Call	Input Token	Output Token	Total Token
Ours	71.8	10.2	4052.0	708.6	4760.6
w/o Context-Aware Pruning	53.8	7.9	2908.5	591.1	3499.6
w/o Structural Alignment	67.9	11.6	4761.9	829.5	5591.4
w/o Dynamic Branch Selection	69.2	10.0	3983.8	701.4	4685.2

Table 5: 关于 CWQ 数据集的消融研究。

Freebase 是一个由 Google 支持的大规模半结构化数据库，旨在收集和连接有关全球数百万实体及其关系的信息。凭借其异常庞大的关系和实体量，Freebase 完全契合我们的知识密集场景。我们在两个以 Freebase 作为外部知识库的 QA 集上进行实验：WebQSP (Yih et al. 2016) 和 CWQ (Talmor and Berant 2018)。

我们构建的农业知识图谱包含超过 100,000 个实体和 100 万个三元组。虽然与 Freebase 等通用知识图谱相比，实体数量相对较少，但其关系结构仍然非常复杂。在此基础上，我们构建了单跳和多跳问答集。

我们通过两个不同的指标来评估多答案问题：(1) 部分匹配，其中成功要求检索到至少一个正确答案，以及 (2) 完全匹配，其需要识别出所有正确答案。

基线方法。我们将我们的方法与五种基线方法进行了比较：标准提示 (IO 提示) (Brown et al. 2020)，思维链提示 (CoT 提示) (Wei et al. 2022b)，ToG (Sun et al. 2024)，PoG (Chen et al. 2024)，StructGPT (Jiang et al. 2023b)，KB-BINDER (Li et al. 2023c) 和 FiSKE (Tao et al. 2025)。IO 提示和 CoT 提示是两种无知识的方法，用于衡量大型语言模型仅基于其内部知识可以回答多少问题。ToG、PoG、StructGPT 和 KB-BINDER 代表了以前在知识库问答中的最新方法，作为基线来评估我们知识检索方法的有效性。

实验细节。为了确保实验的可靠性和可重复性，我们将所有大语言模型的温度参数设为 0。对于英文 QA 数据集 (WebQSP 和 CWQ)，我们采用了基线方法代码库中提供的原始提示。对于中文农业 QA 数据集，我们通过翻译英文模板来调整中文版本的提示，同时保留每个基线方法的结构格式。

4.2 性能比较

WebQSP 和 CWQ 的结果如表 1 所示。在 WebQSP 数据集上，我们采用 Llama3-8B-Instruct 的方法显著优于使用相同模型的现有方法，取得了可与或甚至超越大型模型 (如 GPT-3.5 或 DeepSeek-V3 (DeepSeek-AI et al. 2024)) 的结果。这不仅证明了我们设计的有效性，还突出了它对较小的 LLM 特别适合，表明其具有实质性的实际价值。在 DeepSeek-V3 上的评估中，我们的方法在部分匹配性能上可媲美最先进的 FiSKE 和 PoG 方法，同时优于所有其他方法，并在完整匹配结果中超过所有方法。这表明我们的方法更倾向于检索更全面的知

识。

与 WebQSP 相比，CWQ 数据集包含更多的多跳问题。在 CWQ 上，FiSKE 表现不如 PoG，这验证了我们的主张，即仅依赖细粒度信息难以应对复杂问题。我们的方法在 FiSKE 的基础上提高了 20 多个百分点，展示了其在知识密集型场景中解决复杂问题的强大能力。

在表格 2 中展示了农业数据集上的实验结果。为了公平比较，我们为 PoG 配备了与我们相同的细粒度信息提取方法，因为其原始实现依赖于数据集提供的主题实体，而这些在农业 QA 集合中不可用。我们的方法显示出比现有方法显著的改进，特别是在多跳问答方面，优势达到了显著水平。值得注意的观察是，尽管现有方法在单跳和多跳 QA 任务之间表现出显著的性能差距，我们的方法在这两种情况下始终维持高性能。这显示了我们的方法在处理知识密集型图中的复杂问题方面的卓越能力，这也在表格 1 中得到了证实。

总体而言，我们的方法显示出优于现有方法的性能，尤其在知识密集型场景中解决复杂问题时表现出色，同时保持与小规模 LMs 的强兼容性。

4.3 计算成本

我们对我们的方法进行了效率研究，结果如表 3 所示。总体而言，我们的分析揭示了我们的方法在 LLM 调用次数和总 token 消耗方面相比 PoG 展示了全面的优势。在 WebQSP (以单跳问题为主) 上，我们的方法显示出适度的改善，而在 CWQ (主要包含多跳问题) 上，则取得了显著的提升，突显了其在处理复杂问题方面的优越能力。

在农业数据集上，我们的方法所需的 LLM 调用比 PoG 略少，同时 LLM 的令牌消耗则显著减少。我们将此归因于 PoG 依赖于有严格规则限制的提示来管理 LLM 输出，而我们的方法主要是在设计策略中嵌入约束而不是提示。这一区别可能解释了为什么我们的方法对较小的 LLM 更友好。FiSKE 在 WebQSP 和 CWQ 上表现出高效率，但在处理结构复杂的农业数据集时会产生显著的计算开销。相反，我们的方法在所有数据集上始终实现稳定且优越的性能。

我们进行了实验来评估起始点检索成功对我们提出的方法的影响，结果如表 4 所示。当起始点检索失败时，我们的方法默认由 LLM 直接回答问题。需要注意的是，像 ToG 和 PoG 这样的引导性问题方法直接从数

据集中检索主题实体作为起始点。

结果表明, 绕过没有明确起始点的问题可以在效率略降低的情况下提升答案准确性。此观察是合理的: 虽然我们的方法比直接使用 LLM 回复显著提高了准确性, 但它自然会带来额外的计算开销。

总体而言, 尽管无法识别起始点影响了我们方法的表现, 但这一限制并不掩盖其显著优势。

4.4 消融研究

我们进行了消融研究, 结果如表格 5 所示。起始点的影响在上一小节中进行了讨论。因此, 我们对上下文感知剪枝、结构对齐和动态分支选择进行了消融研究。对于上下文感知剪枝, 我们用单一线索替换了原本语义丰富的目标短语。对于后两个组件, 我们在消融过程中将其完全移除。

消融研究证明了每个组件的必要性, 因为去除任何模块都会导致性能下降。去掉上下文感知剪枝会导致显著的性能下降, 突出了上下文语义在基于 LLM 的剪枝中的关键作用。去除结构对齐不仅降低了准确性, 还降低了效率, 证实了其所设计的双重优势。禁用动态分支选择虽然略微减少了计算开销, 但却以性能为代价, 为了显著提高准确性, 其最小化 token 消耗的权衡是合理的。

我们的方法的另一个关键特征是其顺序性, 具体体现在目标选择过程中。三个模块(情境感知修剪、结构对齐和动态分支选择)只能在这种顺序框架内有效运行。这些模块的综合效果证明了保持这个有序执行流程的必要性。

在本文中, 我们提出了一种新的范式, 称为导图引导的知识探索, 该范式在非结构化查询和结构化知识检索之间建立了桥梁。通过引入导图, 我们的方法解决了现有方法面临的难题, 这些方法要么遭遇冗余探索, 要么难以区分复杂结构。我们的方法从问题中提取多维信息, 并为导图建立系统的构建规则, 赋予其丰富的结构和语义信息。我们提出了上下文感知修剪和结构对齐, 以高效准确地识别与导图保持结构和语境语义一致的知识子图。大量实验表明, 我们的方法在多个数据集上取得了显著的改进, 特别是在知识密集型场景中展现了卓越的区分复杂结构的能力。结果进一步揭示了我们的方法与小规模 LLM 的强大兼容性, 突显其实用价值。全面的消融研究验证了每个提出模块的贡献。

References

Baek, J.; Aji, A. F.; and Saffari, A. 2023. Knowledge-augmented language model prompting for zero-shot knowledge graph question answering. arXiv preprint arXiv:2306.04136.

Bang, Y.; Cahyawijaya, S.; Lee, N.; Dai, W.; Su, D.; Wilie, B.; Lovén, H.; Ji, Z.; Yu, T.; Chung, W.; et al. 2023. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. arXiv preprint arXiv:2302.04023.

Besta, M.; Blach, N.; Kubicek, A.; Gerstenberger, R.; Podstawski, M.; Gianinazzi, L.; Gajda, J.; Lehmann, T.; Niewiadomski, H.; Nyczyk, P.; et al. 2024. Graph of thoughts: Solving elaborate problems with large language models. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 17682–17690.

Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD international conference on Management of data, 1247–1250.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners advances in neural information processing systems 33.

Chen, L.; Li, Z.; He, W.; Cheng, G.; Xu, T.; Yuan, N. J.; and Chen, E. 2022. Entity summarization via exploiting description complementarity and salience. IEEE transactions on neural networks and learning systems, 34(11): 8297–8309.

Chen, L.; Tong, P.; Jin, Z.; Sun, Y.; Ye, J.; and Xiong, H. 2024. Plan-on-Graph: Self-Correcting Adaptive Planning of Large Language Model on Knowledge Graphs. arXiv preprint arXiv:2410.23875.

Chen, X.; Lu, X.; Xin, H.; Peng, W.; Duan, H.; Jiang, F.; Zhou, J.; and Xiong, H. 2023. A table-to-text framework with heterogeneous multidominance attention and self-evaluated multi-pass deliberation. In Findings of the Association for Computational Linguistics: EMNLP 2023, 607–620.

DeepSeek-AI; Liu, A.; Feng, B.; Xue, B.; Wang, B.-L.; Wu, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; Ruan, C.; Dai, D.; Guo, D.; Yang, D.; Chen, D.; Ji, D.-L.; Li, E.; Lin, F.; Dai, F.; Luo, F.; Hao, G.; Chen, G.; Li, G.; Zhang, H.; Bao, H.; Xu, H.; Wang, H.; Zhang, H.; Ding, H.; Xin, H.; Gao, H.; Li, H.; Qu, H.; Cai, J. L.; Liang, J.; Guo, J.; Ni, J.; Li, J.; Wang, J.; Chen, J.; Chen, J.; Yuan, J.; Qiu, J.; Li, J.; Song, J.-M.; Dong, K.; Hu, K.; Gao, K.; Guan, K.; Huang, K.; Yu, K.; Wang, L.; Zhang, L.; Xu, L.; Xia, L.; Zhao, L.; Wang, L.; Zhang, L.; Li, M.; Wang, M.; Zhang, M.; Zhang, M.; Tang, M.; Li, M.; Tian, N.; Huang, P.; Wang, P.; Zhang, P.; Wang, Q.; Zhu, Q.; Chen, Q.; Du, Q.; Chen, R. J.; Jin, R.; Ge, R.; Zhang, R.; Pan, R.; Wang, R.; Xu, R.; Zhang, R.; Chen, R.; Li, S. S.; Lu, S.; Zhou, S.; Chen, S.; Wu, S.-P.; Ye, S.; Ma, S.; Wang, S.; Zhou, S.; Yu, S.; Zhou, S.; Pan, S.; Wang, T.; Yun, T.; Pei, T.; Sun, T.; Xiao, W. L.; Zeng, W.; Zhao, W.; An, W.; Liu, W.; Liang, W.; Gao, W.; Yu, W.-X.; Zhang, W.; Li, X. Q.; Jin, X.; Wang, X.; Bi, X.; Liu, X.; Wang, X.; Shen, X.-C.; Chen, X.; Zhang, X.; Chen, X.; Nie, X.; Sun, X.; Wang, X.; Cheng, X.; Liu, X.; Xie, X.; Liu, X.; Yu, X.; Song, X.; Shan, X.; Zhou, X.; Yang, X.; Li, X.; Su, X.; Lin, X.; Li, Y. K.; Wang, Y. Q.; Wei, Y. X.; Zhu, Y. X.; Zhang, Y.; Xu, Y.; Huang, Y.; Li, Y.; Zhao, Y.; Sun, Y.; Li, Y.; Wang, Y.; Yu, Y.; Zheng, Y.; Zhang, Y.; Shi, Y.; Xiong, Y.; He, Y.; Tang, Y.; Piao, Y.; Wang, Y.; Tan, Y.; Ma, Y.-B.; Liu, Y.; Guo, Y.; Wu, Y.; Ou, Y.; Zhu, Y.; Wang, Y.; Gong, Y.; Zou, Y.; He, Y.; Zha, Y.; Xiong, Y.; Ma, Y.; Yan, Y.; Luo, Y.-W.; mei You, Y.; Liu, Y.; Zhou, Y.; Wu, Z. F.; Ren, Z.; Ren, Z.; Sha, Z.; Fu, Z.; Xu, Z.; Huang, Z.; Zhang, Z.; Xie, Z.; guo Zhang, Z.; Hao, Z.; Gou, Z.; Ma, Z.; Yan, Z.; Shao, Z.; Xu, Z.; Wu, Z.;

- Zhang, Z.; Li, Z.; Gu, Z.; Zhu, Z.; Liu, Z.; Li, Z.-A.; Xie, Z.; Song, Z.; Gao, Z.; and Pan, Z. 2024. DeepSeek-V3 Technical Report. ArXiv, abs/2412.19437.
- Gong, Z.; and Sun, Y. 2024. Graph reasoning enhanced language models for text-to-sql. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2447–2451.
- He, G.; Lan, Y.; Jiang, J.; Zhao, W. X.; and Wen, J.-R. 2021. Improving multi-hop knowledge base question answering by learning intermediate supervision signals. In Proceedings of the 14th ACM international conference on web search and data mining, 553–561.
- Hu, C.; Fu, J.; Du, C.; Luo, S.; Zhao, J.; and Zhao, H. 2023. Chatdb: Augmenting llms with databases as their symbolic memory. arXiv preprint arXiv:2306.03901.
- Ji, B.; Liu, H.; Du, M.; and Ng, S.-K. 2024. Chain-of-thought improves text generation with citations in large language models. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 18345–18353.
- Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y. J.; Madotto, A.; and Fung, P. 2023. Survey of hallucination in natural language generation. *ACM computing surveys*, 55(12): 1–38.
- Jiang, J.; Zhou, K.; Dong, Z.; Ye, K.; Zhao, W. X.; and Wen, J.-R. 2023a. Structgpt: A general framework for large language model to reason over structured data. arXiv preprint arXiv:2305.09645.
- Jiang, J.; Zhou, K.; Dong, Z.; Ye, K.; Zhao, W. X.; and Wen, J.-R. 2023b. Structgpt: A general framework for large language model to reason over structured data. arXiv preprint arXiv:2305.09645.
- Jiang, J.; Zhou, K.; Zhao, W. X.; and Wen, J.-R. 2022. Unikgqa: Unified retrieval and reasoning for solving multi-hop question answering over knowledge graph. arXiv preprint arXiv:2212.00959.
- Li, S.; Gao, Y.; Jiang, H.; Yin, Q.; Li, Z.; Yan, X.; Zhang, C.; and Yin, B. 2023a. Graph reasoning for question answering with triplet retrieval. arXiv preprint arXiv:2305.18742.
- Li, T.; Ma, X.; Zhuang, A.; Gu, Y.; Su, Y.; and Chen, W. 2023b. Few-shot in-context learning for knowledge base question answering. arXiv preprint arXiv:2305.01750.
- Li, T.; Ma, X.; Zhuang, A.; Gu, Y.; Su, Y.; and Chen, W. 2023c. Few-shot in-context learning for knowledge base question answering. arXiv preprint arXiv:2305.01750.
- Li, X.; and Qiu, X. 2023. Mot: Memory-of-thought enables chatgpt to self-improve. arXiv preprint arXiv:2305.05181.
- Li, X.; Zhao, R.; Chia, Y. K.; Ding, B.; Bing, L.; Joty, S.; and Poria, S. 2023d. Chain of knowledge: A framework for grounding large language models with structured knowledge bases. arXiv preprint arXiv:2305.13269, 3.
- Li, X.; Zhou, Y.; and Dou, Z. 2024. Unigen: A unified generative framework for retrieval and question answering with large language models. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 8688–8696.
- Luo, L.; Li, Y.-F.; Haffari, G.; and Pan, S. 2023a. Reasoning on graphs: Faithful and interpretable large language model reasoning. arXiv preprint arXiv:2310.01061.
- Luo, Z.; Xu, C.; Zhao, P.; Geng, X.; Tao, C.; Ma, J.; Lin, Q.; and Jiang, D. 2023b. Augmented large language models with parametric knowledge guiding. arXiv preprint arXiv:2305.04757.
- Sun, J.; Xu, C.; Tang, L.; Wang, S.; Lin, C.; Gong, Y.; Ni, L.; Shum, H.-Y.; and Guo, J. 2024. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. In The Twelfth International Conference on Learning Representations.
- Talmor, A.; and Berant, J. 2018. The web as a knowledge-base for answering complex questions. arXiv preprint arXiv:1803.06643.
- Tao, D.; Wang, C.; Huang, F.; Chen, J.; Huang, Y.; and Jiang, M. 2025. Fine-grained Stateful Knowledge Exploration: Effective and efficient graph retrieval with Large Language Models. *Knowledge-Based Systems*, 326: 114011.
- Wang, L.; Hu, Y.; He, J.; Xu, X.; Liu, N.; Liu, H.; and Shen, H. T. 2024a. T-sciq: Teaching multimodal chain-of-thought reasoning via large language model signals for science question answering. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, 19162–19170.
- Wang, L.; Xu, W.; Lan, Y.; Hu, Z.; Lan, Y.; Lee, R.; and Lim, E. 2024b. Plan-and-solve prompting: improving zero-shot chain-of-thought reasoning by large language models (2023). arXiv preprint arXiv:2305.04091.
- Wang, S.; Sui, Y.; Wu, J.; Zheng, Z.; and Xiong, H. 2024c. Dynamic sparse learning: A novel paradigm for efficient recommendation. In Proceedings of the 17th ACM international conference on web search and data mining, 740–749.
- Wang, X.; Gao, T.; Zhu, Z.; Zhang, Z.; Liu, Z.; Li, J.; and Tang, J. 2021. KEPLER: A unified model for knowledge embedding and pre-trained language representation. *Transactions of the Association for Computational Linguistics*, 9: 176–194.
- Wang, X.; Yang, Q.; Qiu, Y.; Liang, J.; He, Q.; Gu, Z.; Xiao, Y.; and Wang, W. 2023. Knowledgpt: Enhancing large language models with retrieval and storage access on knowledge bases. arXiv preprint arXiv:2308.11761.
- Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022a. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022b. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.

Wu, W.; Wang, C.; Shen, D.; Qin, C.; Chen, L.; and Xiong, H. 2024. Afdgcf: Adaptive feature de-correlation graph collaborative filtering for recommendations. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1242–1252.

Yao, L.; Mao, C.; and Luo, Y. 2019. KG-BERT: BERT for knowledge graph completion. *arXiv preprint arXiv:1909.03193*.

Yao, S.; Yu, D.; Zhao, J.; Shafran, I.; Griffiths, T.; Cao, Y.; and Narasimhan, K. 2023. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, 36: 11809–11822.

Yih, W.-t.; Richardson, M.; Meek, C.; Chang, M.-W.; and Suh, J. 2016. The value of semantic parse labeling for knowledge base question answering. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 201–206.

Zhang, S.; Chen, L.; Wang, C.; Li, S.; and Xiong, H. 2024. Temporal graph contrastive learning for sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 9359–9367.

Zhang, Y.; Sun, Y.; Zhuang, F.; Zhu, Y.; An, Z.; and Xu, Y. 2023. Triple dual learning for opinion-based explainable recommendation. *ACM Transactions on Information Systems*, 42(3): 1–27.

Zhang, Z.; Han, X.; Liu, Z.; Jiang, X.; Sun, M.; and Liu, Q. 2019a. ERNIE: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*.

Zhang, Z.; Han, X.; Liu, Z.; Jiang, X.; Sun, M.; and Liu, Q. 2019b. ERNIE: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*.

Zhao, L.; Liu, Q.; Yue, L.; Chen, W.; Chen, L.; Sun, R.; and Song, C. 2024. Comi: Correct and mitigate shortcut learning behavior in deep neural networks. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 218–228.

Zhao, R.; Li, X.; Joty, S.; Qin, C.; and Bing, L. 2023. Verify-and-edit: A knowledge-enhanced chain-of-thought framework. *arXiv preprint arXiv:2305.03268*.