o1再升级！人大&清华提出Search-o1：赋予推理模型主动搜索的能力

原创 PaperAgent PaperAgent 2025年01月10日 13:30 湖北

近年来，推理模型如OpenAI-o1和千问QwQ等，展示出了令人印象深刻的逐步推理能力。然而，这些模型在进行长链式推理时，常常面临知识不足的问题，导致推理过程中出现不确定性和潜在错误。为了解决这一挑战，本文提出了一种新的框架——**Search-o1**，旨在通过自主知识检索，提升大型推理模型的可靠性和适用性。

Search-o1: Agentic Search-Enhanced Large Reasoning Models

Xiaoxi Li[1], Guanting Dong[1], Jiajie Jin[1], Yuyao Zhang[1], Yujia Zhou[2], Yutao Zhu[1], Peitian Zhang[1], Zhicheng Dou[1*]
[1]Renmin University of China [2]Tsinghua University
{xiaoxi_li, dou}@ruc.edu.cn

Project Page: https://search-o1.github.io/

Paper: https://arxiv.org/abs/2501.05366
HuggingFace:
https://huggingface.co/papers/2501.05366
Github:
https://github.com/sunnynexus/Search-o1

推理模型的现状与挑战

大型推理模型通过大规模的强化学习，能够进行长步骤的逐步推理，适用于科学、数学、编码等复杂领域。这种“慢思考”模式不仅增强了推理的逻辑连贯性和可解释性，但也带来了一个显著的问题：**知识不足**。在推理过程中，模型可能会遇到无法确定的知识点，导致整个推理链条的错误传播，影响最终的答案质量。

研究动机

在初步实验中，本文发现，类似OpenAI-o1的推理模型在处理复杂问题时，平均每个推理过程中会出现超过30次的不确定词汇，如“或许”、“可能”等。这不仅增加了推理的复杂性，还使得手动验证推理过程变得更加困难。因此，如何在推理过程中自动补充所需知识，成为提升大型推理模型可信度的关键。

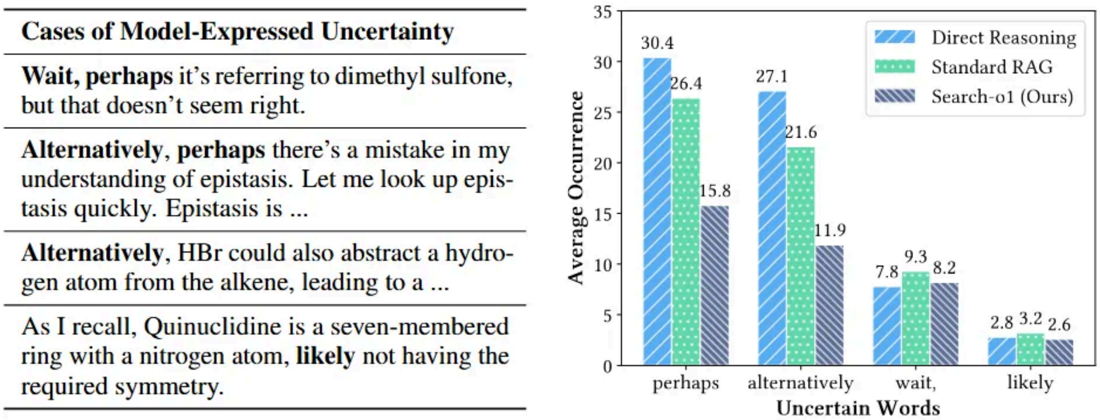

Figure 1: Analysis of reasoning uncertainty with QwQ-32B-Preview. **Left**: Examples of uncertain words identified during the reasoning process. **Right**: Average occurrence of high-frequency uncertain words per output in the GPQA diamond set.

Search-o1：自主知识检索增强的推理框架

为了解决上述问题，本文提出了**Search-o1**框架。该框架通过集成**自主检索增强生成（Agentic Retrieval-Augmented Generation）机制**和**文档内推理模块（Reason-in-Documents）**，实现了在推理过程中动态获取和整合外部知识的能力。

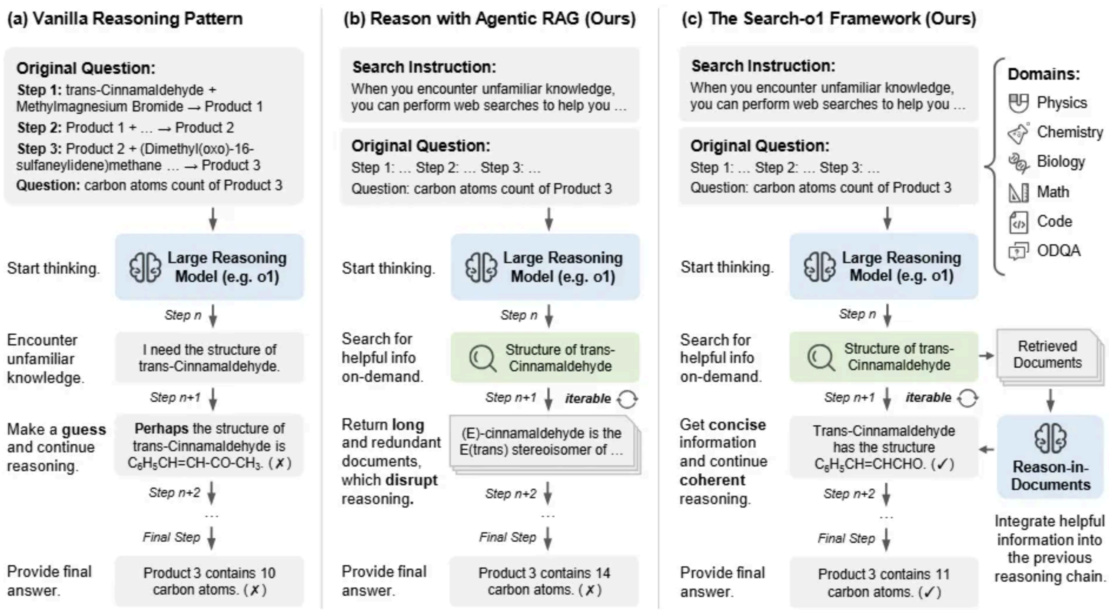

Figure 2: Comparison of reasoning approaches: (a) Direct reasoning without retrieval often results in inaccuracies due to missing knowledge. (b) Our agentic retrieval-augmented reasoning approach improves knowledge access but usually returns lengthy, redundant documents, disrupting coherent reasoning. (c) Our Search-o1 integrates concise and accurate retrieved knowledge seamlessly into the reasoning process, enabling precise and coherent problem-solving.

核心组件

1.自主检索增强生成机制：Search-o1 使模型能够在推理过程中自主决定何时检索外部知识。当模型在推理中遇到不确定的知识点时，会自动生成检索查询，获取相关的外部文档。这种动态检索方式相比传统的静态检索，更加灵活和高效。

2.文档内推理模块：为了避免直接插入冗长且可能含有噪音的检索文档，Search-o1 引入了知识精炼模块。该模块能够对检索到的文档进行筛选和精炼，提取出与当前推理步骤高度相关的关键信息，确保推理过程的连贯性和逻辑一致性。

推理过程

在Search-o1的推理过程中，模型会在生成推理链条的过程中，自动检测是否需要检索外部知识。当需要时，模型会生成特定的检索查询，获取相关文档，并通过文档内推理模块精炼这些文档，将精炼后的知识无缝整合到推理链条中。这一过程能够反复进行，确保模型在整个推理过程中都能获得所需的外部知识支持。

Algorithm 1 Search-o1 Inference

Require: Reasoning Model $\mathcal{M}$, Search function `Search`

Input: Questions $\mathcal{Q}$, Task instruction I, Reason-in-documents instruction I_{docs}
Initialize set of unfinished sequences $\mathcal{S} \leftarrow \{I \oplus q \mid q \in \mathcal{Q}\}$
Initialize set of finished sequences $\mathcal{F} \leftarrow \{\}$
while $\mathcal{S} \neq \emptyset$ **do**
  Generate next tokens for all sequences in $\mathcal{S}$: $\mathcal{T} \leftarrow \mathcal{M}(\mathcal{S})$ ▷ Batch Generate
  Initialize empty set $\mathcal{S}_r \leftarrow \{\}$ ▷ Reason-in-documents Inputs
  for each sequence Seq $\in \mathcal{T}$ **do**
    if Seq ends with <|end_search_query|> **then**
      Pause generation for Seq ▷ Pause Generation
      Extract search query: $q_{\text{search}} \leftarrow$ Extract(Seq, <|begin_search_query|>, <|end_search_query|>)
      Retrieve documents: $\mathcal{D} \leftarrow$ `Search`(q_{search}) ▷ Retrieval
      Construct input for Reason-in-documents: $I_{\mathcal{D}} \leftarrow I_{\text{docs}} \oplus q_{\text{search}} \oplus$ Seq
      Append the tuple $(I_{\mathcal{D}}, \text{Seq})$ to $\mathcal{S}_r$
    else if Seq ends with EOS **then**
      Remove Seq from $\mathcal{S}$, add Seq to $\mathcal{F}$ ▷ Sequence Finished
  if $\mathcal{S}_r \neq \emptyset$ **then**
    Prepare batch inputs: $\mathcal{I}_r \leftarrow \{I_{\mathcal{D}} \mid (I_{\mathcal{D}}, \text{Seq}) \in \mathcal{S}_r\}$
    Reason-in-documents: $\mathcal{T}_r \leftarrow \mathcal{M}(\mathcal{I}_r)$ ▷ Batch Generate
    for $i \leftarrow \{1, \dots, |\mathcal{T}_r|\}$ **do**
      Let $r \leftarrow \mathcal{T}_r[i]$, Seq $\leftarrow \mathcal{S}_r[i]$.Seq
      Extract knowledge-injected reasoning step: $r_{\text{final}} \leftarrow$ Extract(r)
      Update sequence in $\mathcal{S}$: Seq $\leftarrow$ Insert (<|begin_search_result|>, r_{final}, <|end_search_result|>)
Output: Finished Sequences $\mathcal{F}$

实验结果

为了验证Search-o1的有效性，本文在多个复杂推理任务和开放域问答基准上进行了广泛的实验。以下是主要的实验结果：

复杂推理任务

Table 1: Main results on challenging reasoning tasks, including PhD-level science QA, math and code benchmarks. We report Pass@1 metric for all tasks. For models with 32B parameters, the best results are in **bold** and the second-best are underlined. Results from larger or non-proprietary models are in gray color for reference. Symbol "†" indicates results from their official releases.

Method	GPQA (PhD-Level Science QA)				Math Benchmarks			LiveCodeBench			
	Physics	Chemistry	Biology	Overall	MATH500	AMC23	AIME24	Easy	Medium	Hard	Overall
Direct Reasoning (w/o Retrieval)											
Qwen2.5-32B	57.0	33.3	52.6	45.5	75.8	57.5	23.3	42.3	18.9	14.3	22.3
Qwen2.5-Coder-32B	37.2	25.8	57.9	33.8	71.2	67.5	20.0	61.5	16.2	12.2	25.0
QwQ-32B	75.6	39.8	68.4	58.1	83.2	82.5	53.3	61.5	29.7	**20.4**	**33.0**
Qwen2.5-72B	57.0	37.6	68.4	49.0	79.4	67.5	20.0	53.8	29.7	24.5	33.0
Llama3.3-70B	54.7	31.2	52.6	43.4	70.8	47.5	36.7	57.7	32.4	24.5	34.8
DeepSeek-R1-Lite†	-	-	-	58.5	91.6	-	52.5	-	-	-	51.6
GPT-4o†	59.5	40.2	61.6	50.6	60.3	-	9.3	-	-	-	33.4
o1-preview†	89.4	59.9	65.9	73.3	85.5	-	44.6	-	-	-	53.6
Retrieval-augmented Reasoning											
RAG-Qwen2.5-32B	57.0	37.6	52.6	47.5	82.6	72.5	30.0	61.5	24.3	8.2	25.9
RAG-QwQ-32B	76.7	38.7	73.7	58.6	84.8	82.5	50.0	57.7	16.2	12.2	24.1
RAgent-Qwen2.5-32B	58.1	33.3	63.2	47.0	74.8	65.0	20.0	57.7	24.3	6.1	24.1
RAgent-QwQ-32B	76.7	46.2	68.4	61.6	85.0	**85.0**	**56.7**	**65.4**	18.9	12.2	26.8
Retrieval-augmented Reasoning with Reason-in-Documents											
Search-o1 (Ours)	**77.9**	**47.3**	**78.9**	**63.6**	**86.4**	**85.0**	**56.7**	57.7	**32.4**	**20.4**	**33.0**

在复杂推理任务中，包括PhD级别的科学问答（GPQA）、数学（MATH500、AMC2023、AIME2024）和编码能力（LiveCodeBench），Search-o1均显著优于传统的直接推理方法和标准RAG方法。

1. **大型推理模型的优势**：即使在没有检索增强的情况下，QwQ-32B-Preview模型在多个任务上也表现优异，甚至超过了一些更大规模的模型，如Qwen2.5-72B和Llama3.3-70B。这展示了大型推理模型在推理任务中的强大能力。
2. **自主检索增强的效果**：使用自主RAG机制的RAgent-QwQ-32B在大多数任务上超越了标准RAG和直接推理的QwQ-32B，表明自主检索能够有效提升推理模型的知识获取能力。
3. **Search-o1的卓越表现**：进一步引入文档内推理模块后的Search-o1，在大多数任务上超越了RAgent-QwQ-32B，尤其在GPQA、数学和编码任务上取得了显著的性能提升。

检索文档数量的影响

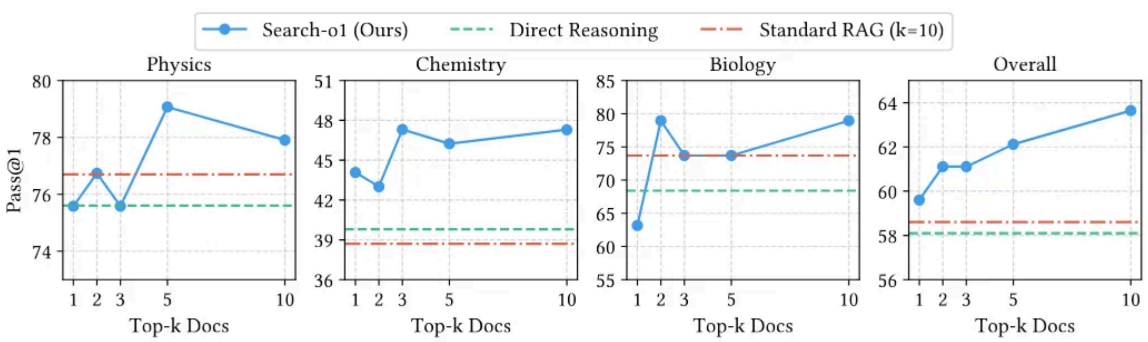

Figure 3: Scaling analysis of top-k retrieved documents utilized in reasoning. All results are based on QwQ-32B-Preview model.

研究发现，Search-o1能够有效利用增加的检索文档数量，进一步提升复杂推理任务的处理能力。即使只检索一篇文档，Search-o1也能够超过直接推理和标准RAG模型，显示出自主检索和文档精炼策略的高效性。

开放域问答任务

Table 3: Performance comparison on open-domain QA tasks, including single-hop QA and multi-hop QA datasets. For models with 32B parameters, the best results are in **bold** and the second-best are underlined. Results from larger models are in gray color for reference.

Method	Single-hop QA				Multi-hop QA							
	NQ		TriviaQA		HotpotQA		2WIKI		MuSiQue		Bamboogle	
	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1
Direct Reasoning (w/o Retrieval)												
Qwen2.5-32B	22.8	33.9	52.0	60.3	25.4	34.7	29.8	36.3	8.4	18.0	49.6	63.2
QwQ-32B	23.0	33.1	53.8	60.7	25.4	33.3	34.4	40.9	9.0	18.9	38.4	53.7
Qwen2.5-72B	27.6	41.2	56.8	65.8	29.2	38.8	34.4	42.7	11.4	20.4	47.2	61.7
Llama3.3-70B	36.0	48.7	68.8	76.8	37.8	49.1	46.0	54.2	14.8	23.6	54.4	67.8
Retrieval-augmented Reasoning												
RAG-Qwen2.5-32B	33.4	49.3	**65.8**	**79.2**	38.6	50.4	31.6	40.6	10.4	19.8	52.0	66.0
RAG-QwQ-32B	29.6	44.4	65.6	77.6	34.2	46.4	35.6	46.2	10.6	20.2	55.2	67.4
RAgent-Qwen2.5-32B	32.4	47.8	63.0	72.6	44.6	56.8	55.4	69.7	13.0	25.4	54.4	66.4
RAgent-QwQ-32B	33.6	48.4	62.0	74.0	43.0	55.2	**58.4**	71.2	13.6	25.5	52.0	64.7
Retrieval-augmented Reasoning with Reason-in-Documents												
Search-o1 (Ours)	**34.0**	**49.7**	63.4	74.1	**45.2**	**57.3**	58.0	**71.4**	**16.6**	**28.2**	**56.0**	**67.8**

在开放域问答任务中，尤其是多跳问答任务，Search-o1表现尤为突出，平均准确率提升了近30%，充分展示了其在知识密集型任务中的优势。而在单跳任务中，虽然提升不显著，但这也表明多跳任务更需要动态知识检索的支持。

结语：迈向更可信赖的智能系统

Search-o1 不仅提升了大型推理模型在复杂任务中的表现，更为智能系统的可靠性和适用性奠定了坚实的基础。通过自主知识检索和精炼整合，Search-o1有效解决了知识不足的问题，显著增强了推理模型的可信度和实用性。未来，随着这一框架的进一步优化和推广，我们可以赋予类o1的推理模型更多的工具，而不仅局限于Search这一个工具，在更多复杂问题的解决中展现出更强大的能力。

推荐阅读

- 对齐LLM偏好的直接偏好优化方法：DPO、IPO、KTO
- 2024：ToB、Agent、多模态
- RAG全景图：从RAG启蒙到高级RAG之36技，再到终章Agentic RAG！
- Agent到多模态Agent再到多模态Multi-Agents系统的发展与案例讲解（1.2万字，20+文献，27张图）

欢迎关注我的公众号“**PaperAgent**”，每天一篇大模型（LLM）文章来锻炼我们的思维，简单的例子，不简单的方法，提升自己。