arxiv.org/abs/2508.12800

Atom-Searcher: Enhancing Agentic Deep Research via Fine-Grained Atomic Thought Reward

Yong Deng*, **Guoqing Wang***, **Zhenzhe Ying***, **Xiaofeng Wu***, **Jinzhen Lin, Wenwen Xiong, Yuqin Dai, Shuo Yang, Zhanwei Zhang, Qiwen Wang, Yang Qin, Changhua Meng**

Ant Group

Core Contributors

代码: https://github.com/antgroup/Research-Venus

简介

本文提出Atom-Searcher，一个以search engine为tool的TIR模型，或者说面向deep research场景的web agent。核心创新是将lrm的<think>...</think>过程细分为更小粒度的Atomic Thoughts(如<plan>、<reflection> 等)，并利用Qwen3对这些原子思维过程进行评分，从而构建细粒度的Atomic Thought Reward (ATR)再结合answer reward组成最终的reward值。Atom-Searcher采用两阶段训练策略：1) 通过SFT让lrm具备生成Atomic Thoughts的能力，简单说就是生成的trajectory要符合格式；2) RLVR，特别的一点是用课程学习思想，对ATR的权重进行线性递减。

背景

本文属于deep research/web agent/TIR方向的工作，目前主流训练方法是基于Outcome Reward Model (ORM)的RLVR，但是ORM的reward既稀疏又粗糙，难以引导llm学会正确的中间推理过程。本文将reasoning trajectory中<think>...</think>内的内容再细化为多组tag，比如<plan>...</plan>，<reflection>...</reflection>，作者把这样更细粒度的tag内容称为Atomic Thoughts，用现成的llm赋予reward值，然后结合ORM作为trajectory最终的reward值用RL训练。

实验设置

- 模型：Qwen2.5-7B-Instruct，框架：VeRL，RL算法：GRPO，tool：搜索引擎
- Reasoning Reward Model：Qwen3-30B-A3B
- reward包含两项，并且atom项权重现行递减

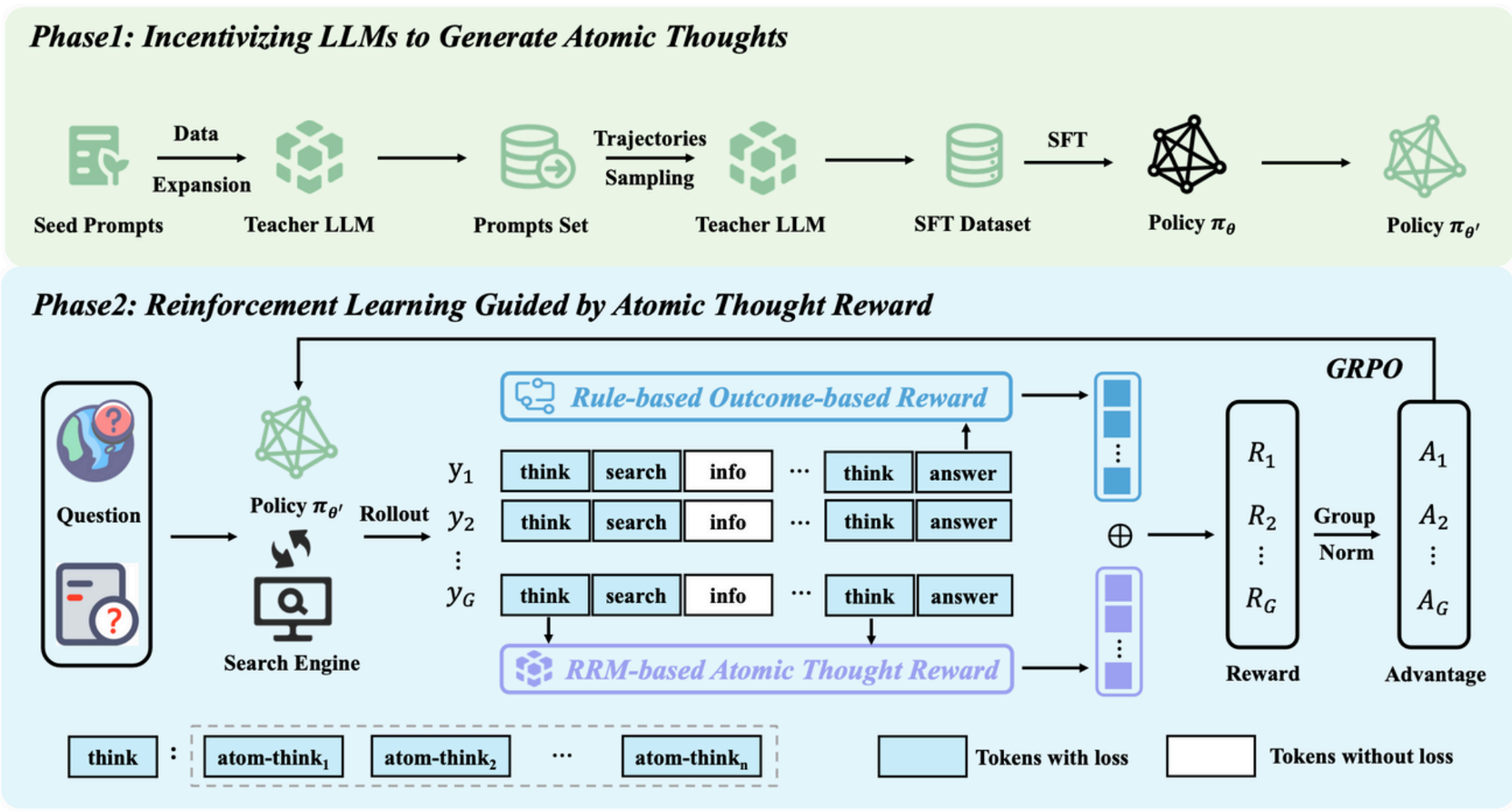
$$\alpha = 0.5 \times (1 - \frac{T}{T_{MAX}})$$

$$R = \begin{cases} \alpha R_{atom} + (1-\alpha) R_{f1} & \text{if format is correct} \\ -1 & \text{if format is incorrect} \end{cases}$$

$$R_{f1} = \frac{2 \times IN}{PN + RN}$$

Atom-Searcher之两阶段训练

- SFT阶段：首先是创建sft数据集，人写几条seed数据+llm大量合成
- RL阶段：重点是reward的设计，包含两项：answer reward 和关注过程的atom reward



思考

1.我看了下Github repo，目前还没有上传代码，希望尽快整理出来吧
2.我很关注到底atom thought对应哪些tag，这样才能理解其定义，而论文中不但没有给出列表，甚至附录中也没有给出system prompt，只提供了RRM用来打分的prompt
3.PRM的难点是定义中间过程的“step”以及如何对其客观打分，本文做法是直接用qwen3打分