arxiv.org/abs/2508.13167

Chain-of-Agents: End-to-End Agent Foundation Models via Multi-Agent Distillation and Agentic RL

Chain-of-Agents：通过multi-agent蒸馏 + RLVR 做TIR

OPPO AI Agent Team

代码地址: github.com/OPPO-PersonalAI/Agent_Foundation_Models

简介

本文提出Chain-of-Agents，一个面向TIR的两阶段训练范式：1) 基于multi-agent system (MAS)的蒸馏SFT，将MAS解决任务过程中生成的推理轨迹(reasoning trajectory) 作为训练数据，对lrm进行SFT，让它学会agent role-playing和tool use；2) 用RLVR进一步强化模型的工具推理能力。作者把最终训练得到的模型称为Agent Foundation Model (AFM)，这是一个具备TIR能力并且是在推理层面通过模拟MAS来做TIR的lrm。

简单点说，就是把multi-agent system中的各个agent看做一个个的function，通过system prompt让lrm知道有哪些function和外部tool可以调用，所以这也是在做multi-tool TIR。

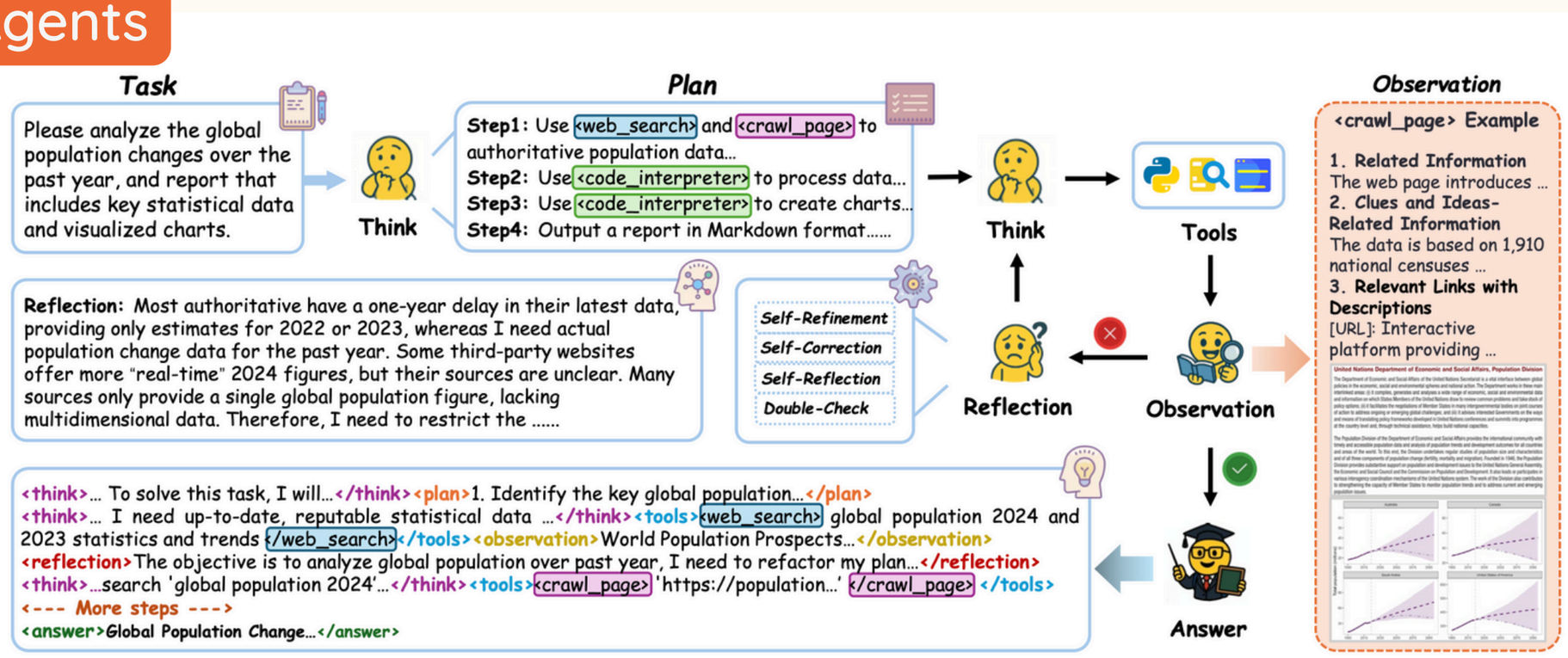
背景

本文属于TIR (tool-integrated reasoning)方向的工作，TIR最大特点是让lrm具备借助tool辅助推理的能力，还有另一个研究llm + tool use的方向，就是agent，尤其是multi-agent system效果比较显著，缺点是成本(token消耗)和执行时间都太高。本文提出了Chain-of-Agents框架，通过蒸馏multi-agent的reasoning trajectory + RLVR两阶段训练，将系统级别的解决问题能力压缩进一个llm，简单说，这个lrm可以模拟multi-agent system来解决问题，作者把这样的lrm称为Agent Foundation Model。

实验设置

- 两类实验场景：Web QA任务和Code任务，每类任务需要分别做sft+rlvr
- 模型：Qwen2.5 instruct系列
- 框架：LLaMA-Factory和VeRL
- RL算法：DAPO
- OAgents提供MAS reasoning trajectory

Chain-of-Agents



把MAS中的agent看作function，这样MAS的reasoning trajectory就和TIR中ReAct格式的三元组形式（think_i, function_i, obsversation_i）一致了

思考

1.从GAIA评测结果看，比常见的TIR lrm好一些，和MAS还是有不小差距的
2.MAS一般支持的tool比较多，本文得到的AFM比一般TIR支持的tool稍微多几个，但和MAS比还有差距
3. 把agent看作function，本文工作定位于multi-tool TIR或许也很合适