

LONGWRITER: UNLEASHING 10,000+ WORD GENERATION FROM LONG CONTEXT LLMs

LongWriter: 通过生成长输出数据做sft提升llm的长文生成能力

Yushi Bai^{1†}, Jiajie Zhang^{1†}, Xin Lv², Linzhi Zheng¹, Siqi Zhu¹,
Lei Hou¹, Yuxiao Dong¹, Jie Tang¹, Juanzi Li¹

¹Tsinghua University ²Zhipu AI 开源代码: github.com/THUDM/LongWriter

简介

本文提出LongWriter框架以提升llm的长文生成能力。作者发现，llm输出长度的天花板主要源于sft阶段所使用response数据的输出长度限制。为突破限制，需要使用更长输出的sft数据训练，如何创建这样的sft数据集呢？作者提出 AgentWrite workflow将长文生成拆解为两步：先让llm生成段落级别的写作计划(plan)，然后让llm按计划生成每个段落。此处llm用的是GPT-4o，创建了包含6k条数据的LongWriter-6k数据集，然后与通用sft数据混合，对base model进行sft，为进一步优化sft模型的指令遵循能力(包括对长文生成指令的遵循)，作者又对sft模型进行了dpo对齐，同样的dpo数据也是混合了通用偏好数据与长文生成相关数据，进一步提升了模型的长文生成质量与长度。

背景

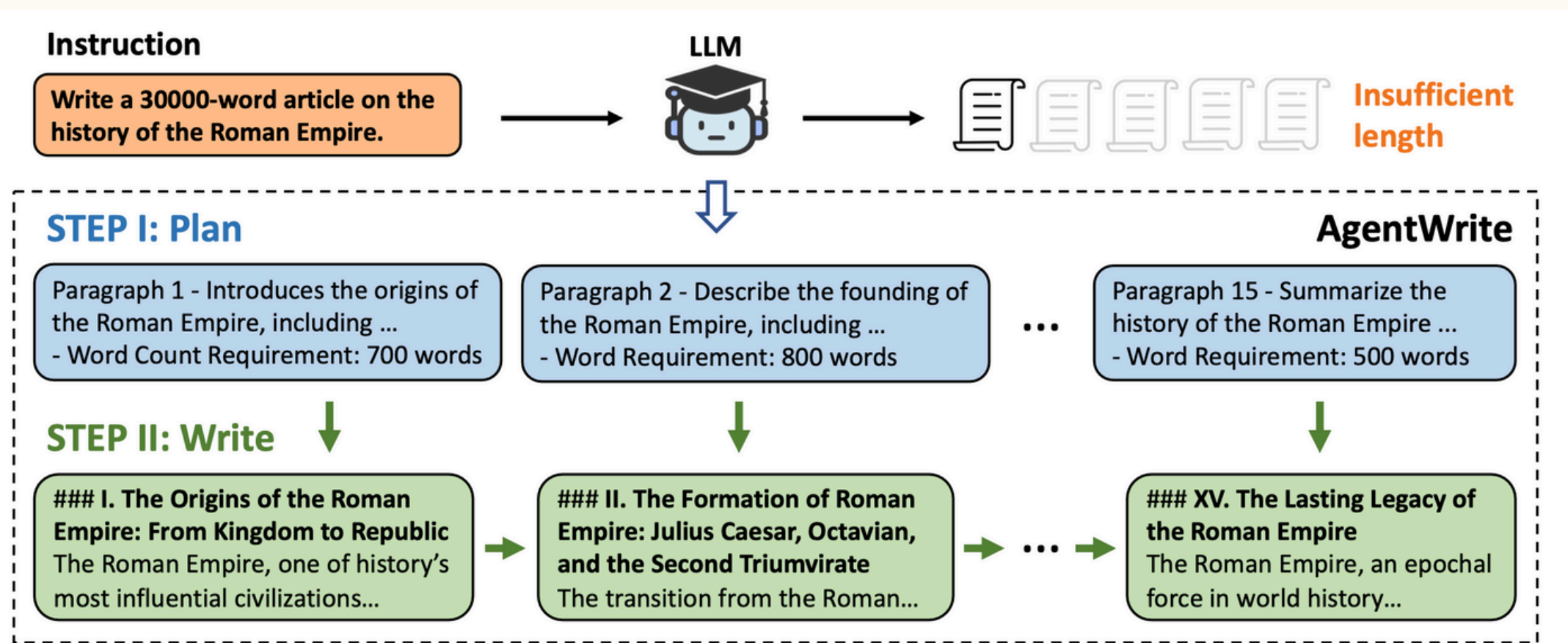
目前llm生成很长response的能力不足，作者发现不论在prompt中明确设置response长度是多少，最多也只能生成2k词左右。为什么会这样呢？作者认为是因为在llm的sft阶段缺乏长response数据，即使在pre-training阶段的sequence length很大也没用，因为在sft阶段大部分sft数据的response长度不足2k。

实验设置

- AgentWrite用的llm: GPT-4o
- AgentWrite第二步让llm生成每个段落，prompt包含plan和前面已经生成的所有段落内容
- sft用的模型: GLM-4-9B和Llama-3.1-8B，都是base model
- sft数据集大小: 180k general sft 数据 + 6k涉及长文生成的sft数据
- dpo数据集大小: 50k general dpo数据 + 4k涉及长文生成的dpo数据

AgentWrite流程

一个并不复杂的长文生成workflow，先生成全文plan，plan包含所有段落描述和字数要求，再依次生成每个段落，这一步无法并行，因为作者要求prompt中包含plan和前面已经生成的所有段落内容，以提升生成内容的连贯性，因此这一流程无法并行执行。



思考

为什么作者认定是sft影响了llm的response长度天花板呢？作者做了如下的控制实验：对于一个base model，通过只改变sft数据中最大输出长度(比如限制为500, 1000, 2000)的方式创建几个sft数据集，然后对base model做sft，在相同测试集上测生成response的长度。结果发现llm最大输出长度随sft数据中最大response长度成正比，于是验证了猜想。我第一个想法是，rlhf就不会影响吗？为什么不考虑rlhf呢，毕竟它才是post-training的最后一个阶段，或许可以补一个正常sft+包含长文数据的DPO实验。另外再仔细想想，不论是sft还是rlhf，追根溯源用的都是human written response，让人去写长达几千几万的response太不现实了，这或许是当前训练长输出模型面临的根本数据瓶颈。