

PR #2036 完整报告

THUDM/slime

fix(search-r1): stop generation at and

合并时间: 2026-06-13 10:57

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2036>

执行摘要

- 一句话: 修复 search-r1 生成中停止标签缺失问题
- 推荐动作: 该 PR 值得合并, 修复逻辑清晰且风险低。建议确认 no_stop_trim=True 的行为与预期一致, 并可考虑为类似示例提供通用停止标签配置模式。

功能与动机

PR body 指出: 多轮 rollout 未告知推理引擎在工具 / 答案边界停止, 导致模型在 `</search>` / `</answer>` 后持续生成垃圾内容 (甚至伪造新的 `Question:`)。当 `return_logprob=True` (TIS 需要) 时, `postprocess_responses` 被禁用以保持 token/logp 对齐, 因此垃圾内容留在轨迹中并被训练 (`loss_mask=1`), 同时破坏了 `is_valid_sequence` 校验 (`</answer>` 后内容导致格式无效、奖励降低)。

实现拆解

1. 定位问题: 在 `examples/search-r1/generate_with_search.py` 的 `generate` 函数中, `sampling_params` 未设置 `stop` 字段, 导致模型自由生成。
2. 添加停止标签: 在进入多轮循环前, 构造 `_stop_tags = ["</search>", "</answer>"]`, 并与已有的 `stop` 参数 (如果有) 合并去重。
3. 更新 `sampling_params`: 通过 `sampling_params = {**sampling_params, "stop": list(dict.fromkeys([*_existing_stop, *_stop_tags]))}` 原地更新, 保留原有参数并新增 `stop` 标签。
4. 备注: `slime` 已默认设置 `no_stop_trim=True`, 因此停止标签会保留在输出中, 不会影响 token/logp 对齐。

关键文件:

- `examples/search-r1/generate_with_search.py` (模块 搜索生成示例; 类别 `source`; 类型 `core-logic`): 唯一修改文件, 核心修复位置, 在 `sampling_params` 中注入停止标签。

关键符号: 未识别

关键源码片段

`examples/search-r1/generate_with_search.py`

唯一修改文件, 核心修复位置, 在 `sampling_params` 中注入停止标签。

```

async def generate(args, sample: Sample, sampling_params) -> Sample:
    # ... 初始化代码 ...
    rollout_log_probs = [] if SEARCH_R1_CONFIGS["return_logprob"] else None

    # BUGFIX: make the inference engine STOP at the tool/answer boundary.
    # Without a stop, sglang keeps emitting tokens after </search> / </answer>
    # (junk, even fabricated new "Question:"s). The example only trimmed that junk
    # via postprocess_responses when return_logprob=False; with return_logprob=True
    # (TIS) trimming is disabled to keep token/logp aligned, so the junk stayed in
    # the trajectory and got trained on (loss_mask=1) AND broke is_valid_sequence
    # (trailing content after </answer> -> format invalid -> lower reward).
    # Stopping at the tag avoids all of that and keeps token/logp aligned natively.
    # slime already sets no_stop_trim=True, so the closing tag is kept in the output.
    _stop_tags = ["</search>", "</answer>"]
    _existing_stop = sampling_params.get("stop") or []
    if isinstance(_existing_stop, str):
        _existing_stop = [_existing_stop]
    sampling_params = {**sampling_params, "stop": list(dict.fromkeys([*_existing_stop, *_stop_tags]))}

    for _turn_idx in range(SEARCH_R1_CONFIGS["max_turns"]):
        payload = {
            "text": prompt_text + response,
            "sampling_params": sampling_params,
        }
        # ... 后续请求和响应处理 ...

```

评论区精华

无 review 讨论。作者在 PR body 中详细解释了问题根因和修复方案的合理性，并附上了修复前后的训练曲线对比，显示修复后效果明显更好。

- 暂无高价值评论线程

风险与影响

- 风险：低风险。变更仅限于 generate_with_search.py 单文件，仅修改 sampling_params 的 stop 列表，不涉及核心框架。由于 slime 默认 no_stop_trim=True，停止标签会保留在输出文本中，不会影响下游训练逻辑。唯一潜在影响是如果已有其他 stop 标签，合并后可能改变生成行为，但代码使用去重合并，不会覆盖已有标签。
- 影响：直接影响 examples/search-r1 中的多轮搜索生成流程，显著提升生成质量和训练数据质量。间接影响依赖该示例的用户，需要更新代码获得修复。不影响 slime 核心库或其他示例。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR