

PR #2167 完整报告

THUDM/slime

Always requires rollout_top_p_token_ids when rollout_top_p is not 1.0

合并时间: 2026-07-02 11:44

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2167>

执行摘要

- 一句话: 修复 rollout_top_p 条件检查, 确保非 1.0 时强制要求 token ids
- 推荐动作: 小范围正确性修复, 值得合并, 无需精读。

功能与动机

当 rollout_top_p 配置不为 1.0 时, rollout 阶段必须产生 top_p token ids 和 offsets 数据用于 off-policy 修正; 原逻辑仅查看 sample 对象中字段是否为 None, 若未正确设置则可能静默跳过, 造成训练数据不完整。PR 标题对动机有明确说明。

实现拆解

1. 在 slime/ray/rollout.py 的 _convert_samples_to_train_data 方法中, 将原本的条件判断从 samples[0].rollout_top_p_token_ids is not None 改为 getattr(self.args, "rollout_top_p", 1.0) != 1.0。
2. 这一改动使得当 rollout_top_p 参数不是默认值 1.0 时, 强制进入数据准备分支, 并在内部通过断言确保相关字段非空。
3. 如果 rollout_top_p 为 1.0 (即不使用 top-p 采样), 则跳过该分支, 与之前行为一致。
4. 没有测试、配置或部署配套改动。

关键文件:

- slime/ray/rollout.py (模块 训练引擎; 类别 source; 类型 core-logic): 修改了条件判断逻辑, 从检查字段 None 改为检查配置参数, 是本次变更的核心文件。

关键符号: _convert_samples_to_train_data

关键源码片段

slime/ray/rollout.py

修改了条件判断逻辑, 从检查字段 None 改为检查配置参数, 是本次变更的核心文件。

```
# slime/ray/rollout.py
# 原逻辑: 根据 sample 字段是否为 None 决定是否要求 top-p 数据
# 新逻辑: 根据 rollout_top_p 配置 (默认 1.0) 决定是否强制要求
if getattr(self.args, "rollout_top_p", 1.0) != 1.0:
    for sample in samples:
```

```
# 断言确保配置启用时数据必须存在
assert sample.rollout_top_p_token_ids is not None
assert sample.rollout_top_p_token_offsets is not None
assert len(sample.rollout_top_p_token_offsets) == sample.response_length + 1, (
    f"top-p token offsets length {len(sample.rollout_top_p_token_offsets)} "
    f"!= response length + 1 {sample.response_length + 1}"
)
offset_end = int(sample.rollout_top_p_token_offsets[-1])
assert offset_end == len(sample.rollout_top_p_token_ids), (
    f"top-p token offsets[-1] {offset_end} "
    f"!= token ids length {len(sample.rollout_top_p_token_ids)}"
)
train_data["rollout_top_p_token_ids"] = [sample.rollout_top_p_token_ids for sample in samples]
train_data["rollout_top_p_token_offsets"] = [sample.rollout_top_p_token_offsets for sample in samples]
```

评论区精华

该 PR 没有 review 评论或讨论。

- 暂无高价值评论线程

风险与影响

- 风险：风险较低，仅修改了一个条件判断，逻辑上更符合配置驱动。但如果某些场景下 `samples[0].rollout_top_p_token_ids` 为 `None` 但配置 `rollout_top_p` 也不为 `1.0`（例如 rollout 阶段未正确生成数据），会触发断言错误，暴露已有问题而非静默失败，属于预期行为。
- 影响：影响范围小，仅影响启用 `rollout_top_p` 的 PPO 训练流程，确保数据完整性。对不使用该功能的用户无影响。
- 风险标记：核心路径变更

关联脉络

- 暂无明显关联 PR