

PR #2121 完整报告

THUDM/slime

Fix bug on non-float reward

合并时间: 2026-06-23 10:45

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2121>

执行摘要

- 一句话: 修复非浮点型 reward 导致的 crash
- 推荐动作: 建议合并。修复明确, 改动安全。若需保留平均 reward 展示, 应在后续 PR 中引入更健壮的类型处理, 例如统一转换为 float。

功能与动机

当 reward 不是 float 类型 (例如 int 或 dict) 时, 原代码中的 reward 累积逻辑会触发 TypeError, 导致 rollout 中断。作者通过删除该统计逻辑来修复此问题。

实现拆解

1. 在 generate_rollout_async 函数中, 删除了 total_reward 和 n_reward_samples 两个局部变量的初始化和更新逻辑。
2. 删除了每个样本组处理完成后对 reward 的遍历、类型判断 (isinstance(s_.reward, (int, float)) 或 dict 取值) 及累加计算。
3. 删除了 tqdm 进度条上显示平均奖励的 set_postfix 调用。
4. 无其他文件变更, 未引入测试代码。

关键文件:

- slime/rollout/sglang_rollout.py (模块 Rollout; 类别 source; 类型 core-logic): 唯一变更文件, 删除了非浮点 reward 导致 crash 的累积统计逻辑。

关键符号: generate_rollout_async

关键源码片段

[slime/rollout/sglang_rollout.py](#)

唯一变更文件, 删除了非浮点 reward 导致 crash 的累积统计逻辑。

```
# 以下为变更后的关键片段, 删除了 reward 累积统计和进度条更新
async def generate_rollout_async(
    args: Namespace, rollout_id: int, data_source: Callable[[int], list[list[Sample]]]
) -> tuple[RolloutFnTrainOutput, list[list[Sample]]]:
    # ...
    data = []
```

```

all_data = []
do_print = True
# 删除了 total_reward = 0.0 和 n_reward_samples = 0
pbar = tqdm(total=target_data_size * args.n_samples_per_prompt, desc="Rollout generation")
while len(data) < target_data_size:
    # ...
    for task in done:
        group: list[Sample] = task.result()
        # ...
        all_data.append(group)

    # 删除了 reward 累积和平均值的计算与显示
    # 直接进入动态过滤和收集逻辑
    dynamic_filter_output = call_dynamic_filter(dynamic_filter, args, group)
    if not dynamic_filter_output.keep:
        metric_gatherer.on_dynamic_filter_drop(reason=dynamic_filter_output.reason)
        state.remaining_batch_size -= 1
        continue
    if len(data) < target_data_size:
        data.append(group)
        pbar.update(args.n_samples_per_prompt)
pbar.close()
# ...

```

评论区精华

无 review 讨论。

- 暂无高价值评论线程

风险与影响

- 风险：删除平均 reward 统计后，进度条不再显示平均奖励，可能影响训练过程中的实时监控。但该统计仅用于日志展示，不影响核心 rollout 逻辑或训练正确性。如果后续需要类似监控，需在其他位置或更高层级实现。
- 影响：影响范围极小：仅移除一处非关键统计日志功能，所有依赖该统计的代码（如有）将无法获得此信息。不涉及 API 变更、配置变更或行为变更。
- 风险标记：移除监控指标

关联脉络

- 暂无明显关联 PR