

PR #1965 完整报告

THUDM/slime

Don't use sample.index as default rollout_id

合并时间: 2026-05-27 20:19

原文链接: <http://prhub.com.cn/THUDM/slime/pull/1965>

执行摘要

- 一句话: 修复默认 rollout_id 使用 sample.index 的问题
- 推荐动作: 该 PR 修复了一个影响训练正确性的隐蔽 bug, 值得精读。关注点在于 `_convert_samples_to_train_data` 中 rollout_ids 的生成逻辑变化, 以及 sample_indices 与 rollout_ids 的分离设计。建议合入, 并添加对应测试以覆盖 compact 路径。

功能与动机

原先使用 sample.index 作为默认 rollout_id, 在 compact 或 subagent 路径下, 一个 rollout 生成多个 training sample, 且未显式设置 rollout_id 时, 这些样本会获得不同的全局唯一 index, 导致 loss reducer 无法正确聚合。PR 描述虽简短, 但明确了问题: rollout_id 用于标识一次 rollout 执行, 同一 rollout 的多个样本应共享同一 id。

实现拆解

1. 提前计算 rollout_ids: 在构造 train_data 字典前, 先判断 samples[0].rollout_id 是否为 None。
2. 条件分支: 如果为 None, 则使用 list(range(len(samples))) 生成本地索引; 否则使用每个样本的显式 rollout_id。
3. 替换字典中的 rollout_ids 字段: 将原先内联的三元表达式替换为预先计算的 rollout_ids 变量。
4. 保留 sample_indices: train_data 中仍保留 sample_indices 字段, 使用 sample.index, 以便需要全局索引时仍可获取。

关键文件:

- slime/ray/rollout.py (模块 rollout; 类别 source; 类型 core-logic; 符号 `_convert_samples_to_train_data`): 核心变更文件, 修改了 `_convert_samples_to_train_data` 方法中 rollout_ids 的默认生成逻辑, 直接修正了训练数据聚合的关键 bug。

关键符号: `_convert_samples_to_train_data`

关键源码片段

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

核心变更文件，修改了 `_convert_samples_to_train_data` 方法中 `rollout_ids` 的默认生成逻辑，直接修正了训练数据聚合的关键 bug。

```
# file: slime/ray/rollout.py (modified)
# 在 _convert_samples_to_train_data 方法中，提前计算 rollout_ids
# 而非在字典内联中 fallback 到 sample.index

# Rollout id (one per rollout execution). Default rollouts emit one
# sample per rollout, so we fall back to local indices (unique within batch).
# Compact / subagent paths that emit multiple training samples per
# rollout set ``rollout_id`` explicitly so all siblings share a
# value; the loss reducer then aggregates them as one rollout.
if samples[0].rollout_id is None:
    # No explicit rollout_id set: use local index (0, 1, 2, ...)
    # This guarantees that samples from the same rollout (e.g., compact mode)
    # are assigned the same id only when explicitly set.
    rollout_ids = list(range(len(samples)))
else:
    # Use the explicitly set rollout_id (shared among siblings)
    rollout_ids = [sample.rollout_id for sample in samples]

train_data = {
    "tokens": [sample.tokens for sample in samples],
    "response_lengths": [sample.response_length for sample in samples],
    "rewards": rewards,
    "raw_reward": raw_rewards,
    "truncated": [1 if sample.status == Sample.Status.TRUNCATED else 0 for sample in samples],
    "sample_indices": [sample.index for sample in samples], # keep original index for reference
    "rollout_ids": rollout_ids, # now uses precomputed list
}
```

评论区精华

该 PR 无 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。变更仅影响 `_convert_samples_to_train_data` 中的 `rollout_id` 默认值逻辑。原先依赖 `sample.index` 作为 `rollout_id` 的代码会切换到本地索引，但 `sample.index` 仍保留在 `sample_indices` 字段中，不会丢失信息。需注意若其他地方依赖 `rollout_id` 的唯一性（如用作数据键），本地索引可能与其他批次重复，但 `rollout` 处理通常按批次进行，风险可控。
- 影响：主要影响 agent RL 训练流程中的 loss 聚合。修复后，compact/subagent 模式下多样本的 `rollout_id` 将正确共享，loss reducer 能正确归一化，训练更为准确。对其他模式（每个 rollout 仅一个 sample）无影响，因为此时 `len(samples) == 1`，本地索引与原先逻辑行为一致。

- 风险标记：核心路径变更，缺少测试覆盖

关联脉络

- PR #1963 Fix trajectory merging logic: 同属 agent RL 训练数据处理的 bugfix 系列，涉及 loss mask 与 rollout 聚合逻辑。
- PR #1960 Extract more util code from coding_agent_rl example: 重构了 trajectory 和 parsing 模块，与 rollout_id 的使用场景相关。
- PR #1954 [coding_agent_rl] middleware: shutdown_session drains in-flight handl…: 修复 rollout 关闭时的竞态问题，与 rollout 生命周期相关。
- PR #1923 [examples] add coding_agent_rl: agent-in-sandbox RL minimal demo: 引入了 compact 和 subagent 路径，暴露了此 bug。