

PR #2035 完整报告

THUDM/slime

Don't upload 'count' to wandb

合并时间: 2026-06-09 09:36

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2035>

执行摘要

- 一句话: 移除 wandb 中 count 指标的上传
- 推荐动作: 该 PR 变更简单、风险低, 无需详细 Review。对于 wandb 日志量敏感的用户, 此项清理可适当降低存储成本。

功能与动机

PR 标题为 'Don't upload 'count' to wandb', 主要目的是去除向 wandb 上报的 count 类指标, 避免无意义的重复统计, 降低 wandb 日志传输量。

实现拆解

1. 移除 metrics 字典中预置的 request/count 和 request/profiled_count 键值对 (原用于统计请求总数和成功采集性能指标的请求数)。
2. 删除遍历 values_by_metric 后不再设置 {key}/count 字段 (原用于记录每个性能指标的实际采集样本数)。
3. 改动集中在 slime/ray/rollout.py 文件的 _compute_sglang_request_perf_metrics 函数, 共减少 5 行、增加 1 行。

关键文件:

- slime/ray/rollout.py (模块 Rollout; 类别 source; 类型 core-logic; 符号 _compute_sglang_request_perf_metrics): 核心变更文件, 修改了 wandb 指标计算的函数, 移除 count 相关字段。

关键符号: _compute_sglang_request_perf_metrics

关键源码片段

slime/ray/rollout.py

核心变更文件, 修改了 wandb 指标计算的函数, 移除 count 相关字段。

```
# 修改前: metrics 中包含 request/count、request/profiled_count 以及各子指标 count
# 修改后: 仅保留统计量, 移除所有 count 字段
```

```
def _compute_sglang_request_perf_metrics(all_samples: list[Sample]):
    attrs_by_request = list(_iter_sglang_generate_attrs(all_samples))
```

```
if not attrs_by_request:
    return {}

values_by_metric: dict[str, list[float]] = {}
profiled_request_count = 0

# ... 略去 add_value 和循环收集 values_by_metric 逻辑 ...

metrics: dict[str, float] = {} # 原为 {"request/count": ..., "request/profiled_count": ...}

for key, values in values_by_metric.items():
    if not values:
        continue
    # 合并统计量, 原还设置了 metrics[f"{key}/count"] = len(values)
    metrics |= dict_add_prefix(compute_statistics(values), f"{key}/")

return metrics
```

评论区精华

该 PR 未产生 review 评论，无讨论记录。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。仅移除 wandb 日志中 count 类冗余字段，不影响核心训练 / 推理逻辑；若下游依赖了这些 count 指标进行监控，可能造成缺失，但 count 通常可通过其他字段推算，影响有限。
- 影响：影响范围小，仅涉及 wandb 日志输出：去除了 request/count、request/profiled_count 和各 {key}/count 字段，减少 wandb 传输数据量。对用户无行为影响，对系统无性能影响。
- 风险标记：暂无

关联脉络

- PR #2027 Only upload per sample stats to wandb: 同为 wandb 指标上传相关的清理与优化，是本次 PR 的延续和补充。