

PR #2189 完整报告

THUDM/slime

[Doc] Clarify PPO/Critic docs after #1856

合并时间: 2026-08-12 13:41

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2189>

执行摘要

纯文档 PR，用于对齐 PPO/Critic 文档与 #1856 之后的实际行为：critic 不再独立申请 GPU，而是与 actor 共享同一 train placement group，资源规模跟随 actor，角色级参数通过 `--megatron-config-path` 的 YAML 覆盖。中英文 `usage.md` 与 `megatron-config.md` 同步更新，共 4 个文件、+20/-58 行。

功能与动机

PR body 指出要 "Clarify PPO/Critic docs after #1856"。旧文档仍描述 critic 独立占用一组 GPU、可通过 `--critic-num-nodes` / `--critic-num-gpus-per-node` 单独配置资源，并给出 $4 \text{ (actor)} + 4 \text{ (critic)} + 8 \text{ (rollout)} = 16 \text{ GPU}$ 的示例；而 #1856 之后 critic 与 actor 共享训练 GPU，继续按旧文档配置会导致用户多预留 GPU、产生困惑。同时 `--critic-load`、`--critic-save`、`--critic-lr` 等 CLI 参数已被角色化 YAML 覆盖取代，需要同步更新。

实现拆解

- 重写 `usage.md` 的 PPO 章节 (`docs/en` 与 `docs/zh`)：删除 critic 独立 GPU 及资源示例，改为说明以下事实：
 - PPO 创建 actor 和 critic 两套训练进程组，但放到同一 train placement group；
 - critic 训练规模跟随 actor 配置，当前 actor / critic 的 Megatron 并行拓扑必须一致；
 - PPO 强制 train 侧 offload，actor 和 critic 在同一批 GPU 上轮流唤醒和释放显存；
 - 无单独配置 critic 训练资源的 CLI 参数，节点数与每节点 GPU 数由 actor 派生。同时将 `--critic-*` 参数替换为 `--megatron-config-path` 及其角色覆盖用法。
- 更新 `megatron-config.md` (`docs/en` 与 `docs/zh`)：示例中移除 `--critic-num-nodes` / `--critic-num-gpus-per-node`；注意事项新增“actor 和 critic 共享同一 train placement group，critic 资源由 actor 配置派生”；FAQ 改为“能否把资源配置写进 YAML？”，答复为资源分配仍由 CLI 控制，YAML 字段会被忽略。
- 同步中英文：两个语言版本逐条对应，保持信息一致。
- 无测试、配置或部署配套改动：纯文档变更。

关键源码片段

本 PR 为纯文档变更，无源码片段可展示。

评论区精华

本 PR 没有 review 评论或讨论线程，属于直接合并的文档修正。

风险与影响

- 风险：文档描述与实现强绑定，若后续训练资源分配逻辑再次变化需同步更新；中英文文档需持续保持一致。
- 影响：用户在配置 PPO 时不再需要为 critic 预留独立 GPU，可避免资源浪费和初始化失败；对源码无影响，团队维护成本低。

关联脉络

本 PR 是 #1856 的文档跟进（#1856 不在本次提供的关联列表中）。结合近期历史 PR（如 #2249 移除 `--train-memory-margin-bytes`、#2250 新增 rollout 采样钩子与 `slime/ray/rollout.py` 相关调整），可以看到 PPO 训练配置正逐步从零散 CLI 参数收敛到角色化 YAML 覆盖和共享资源模型，本 PR 是对这一演进方向的文档固化。