

PR #7492 完整报告

verl-project/verl

[ci] fix: revert megatron version on ascend ci

合并时间: 2026-08-20 20:30

原文链接: <http://prhub.com.cn/verl-project/verl/pull/7492>

执行摘要

本 PR 将 Ascend CI 测试环境中使用的 Megatron-LM 和 MindSpeed 依赖版本进行回退，以修复由新版本引入的测试失败，恢复 CI 稳定性。变更仅涉及一个 GitHub Actions workflow 文件，影响范围严格限于 Ascend 平台的自动化测试流程。

功能与动机

动机明确且紧急：恢复 Ascend CI 的正常运行。根据 PR 描述，“Revert megatron version. Only affect ascend ci.”，这是对之前引入的 Megatron `core_r0.18.0` 版本进行回退，表明该版本在 Ascend NPU 环境下存在兼容性问题，导致端到端测试失败。仓库近期历史中多个 Ascend CI 修复 PR（如 #7584， #7558）印证了该环境维护的持续性挑战。

实现拆解

1. 变更入口：直接修改 CI workflow 定义文件 `.github/workflows/e2e_ppo_trainer_megatron_sglang_ascend.yml`。
2. 核心逻辑移除：在工作流的 `test_fsg` 和 `test_e2e` 两个 job 中，完全删除了“Configure related dependencies”步骤。该步骤原先负责将 Megatron-LM 和 MindSpeed 的 `core_r0.18.0` 版本克隆到容器内固定路径并注入 `PYTHONPATH`。移除后，测试将依赖容器镜像或 `pip` 安装步骤中已准备好的依赖版本。
3. 无配套变更：本次修改未涉及任何业务代码、单元测试或文档。其影响纯粹是 CI 测试环境的依赖版本配置。

(无需展示，变更内容为配置步骤的删除)

评论区精华

PR 未产生实质性讨论。评审人 `wucong25` 的 approve 评论为空，表明该变更被视为一个直接、必要的运维操作，无需技术辩论。

风险与影响

- 技术风险：主要风险是测试环境依赖版本回退后，可能漏掉对 Megatron `core_r0.18.0` 在 Ascend 上表现的验证。若该版本包含重要修复，则修复不会在 Ascend 上得到测试。
- 兼容性风险：使用“预装”版本替代“显式指定”版本，需确保预装版本与 verl 代码库（特别是 Megatron 适配层）保持兼容。

- 维护风险：这只是临时性解决方案，根本原因（版本兼容性）未解决，未来升级版本时问题可能重现。
- 系统影响：对用户无影响。对团队而言，恢复了 Ascend CI 的可靠性，但需投入额外关注以跟踪上游 Megatron 变化。

关联脉络

本 PR 是近期一系列 Ascend CI 维护工作（#7584, #7558, #7606）的一部分，共同体现了团队在保证多平台（GPU, NPU）测试稳定性上所做的努力。它与清理过时配置（如 `enable_chunked_prefill=False`）的 PR 处于同一问题域，反映了上游依赖变更对不同硬件平台测试环境的连带影响。

建议

该 PR 应作为维护性修复快速合入，以恢复 CI。同时，建议团队：

1. 记录本次回退使用的具体 Megatron 版本（如镜像中的版本号），便于未来参考。
2. 在内部任务中跟踪“Ascend 平台 Megatron 兼容性”问题，评估是升级镜像、寻找稳定版本，还是向 MindSpeed/NVIDIA 反馈兼容性问题。