

PR #6586 完整报告

verl-project/verl

[ci] fix: chang sglang8.5.0 a3

合并时间: 2026-06-04 10:15

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6586>

执行摘要

- 一句话: 升级 Ascend Docker 中 MindSpeed/Megatron-LM 版本至 core_r0.16.0
- 推荐动作: 建议关注后续 NPU CI 稳定性, 确认新版本兼容性。本次为常规依赖升级, 无需深入精读。

功能与动机

MindSpeed 和 Megatron-LM 的旧版本可能存在兼容性问题或不再被维护, 升级至 core_r0.16.0 是后续 NPU CI 和功能开发的基础。显式指定 PyTorch 版本避免 SGLang 安装过程中依赖解析导致的版本冲突。

实现拆解

1. 版本升级: 在 Dockerfile.ascend_9.0.0_a2 和 Dockerfile.ascend_9.0.0_a3 中, 将 MindSpeed 的 checkout 分支从 2.3.0_core_r0.12.1 改为 core_r0.16.0, 将 Megatron-LM 的 clone 分支从 core_v0.12.1 改为 core_r0.16.0。
2. SGLang 镜像调整: 在 Dockerfile.ascend.sglang_8.5.0_a3 中, 安装 SGLang 后添加 `pip install torch==2.8.0 torchvision==0.23.0` 以锁定 PyTorch 版本; 同时清除末尾空行。
3. 文档同步: 在 docs/ascend_tutorial/get_start/dockerfile_build_guidance.rst 中将 MindSpeed 和 Megatron-LM 版本号更新为 core_r0.16.0, 保持文档与镜像版本一致。

关键文件:

- docker/ascend/Dockerfile.ascend_9.0.0_a2 (模块 Docker 构建; 类别 infra; 类型 infrastructure) : 核心 Dockerfile, 包含 MindSpeed 和 Megatron-LM 版本升级。
- docker/ascend/Dockerfile.ascend_9.0.0_a3 (模块 Docker 构建; 类别 infra; 类型 infrastructure) : 与 a2 相同的版本升级, 用于不同的镜像变体。
- docker/ascend/Dockerfile.ascend.sglang_8.5.0_a3 (模块 Docker 构建; 类别 infra; 类型 infrastructure) : 针对 SGLang 镜像的特殊调整, 添加了显式 PyTorch 版本锁定。
- docs/ascend_tutorial/get_start/dockerfile_build_guidance.rst (模块 文档; 类别 docs; 类型 documentation) : 同步更新文档中的版本记录。

关键符号: 未识别

评论区精华

自动化代码审查工具 `gemini-code-assist[bot]` 指出多个临时调试语句 (`echo`、`pip list`) 及冗余的 `pip install torch==2.8.0`，建议移除以避免不必要的镜像层。这些语句在最终合并前被清除，最终由 `wucong25` 审批通过。

- 临时调试语句 (style): 这些临时语句在最终合并前被清除，保留必要的版本变更。

风险与影响

- 风险：升级 MindSpeed 和 Megatron-LM 版本可能引入新行为，需确保 NPU CI 测试通过。显式锁定 PyTorch 版本可能与其他依赖产生冲突，但镜像基于特定版本构建，风险可控。整体影响局限于 Ascend 镜像构建环境。
- 影响：仅影响使用 Ascend NPU 镜像的用户，需重新构建镜像以获取更新。对系统其他部分无影响。
- 风险标记：依赖版本变更，Docker 构建可能受影响

关联脉络

- PR #6581 [docker] fix: align stable sglang image for GB200: 同样修改了 SGLang Dockerfile，涉及 Gemini NPU 镜像调整。
- PR #6520 [ci] chore: npu ci use cann9.0.0: 也与 NPU 镜像依赖版本升级相关，升级了 CANN 和模型支持。