

PR #6568 完整报告

verl-project/verl

[ci] fix: stablize main CIs

合并时间: 2026-06-03 15:51

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6568>

执行摘要

- 一句话: 稳定主 CI 并修复 FSDP2 全异步策略 OOM
- 推荐动作: 该 PR 值得浏览, 特别是 CI 配置改动和 FSDP 策略内存调整的技巧。对于希望维护 CI 稳定的团队有借鉴意义。

功能与动机

PR 作者在 CI 中发现不稳定问题, 需要修复以避免阻断开发流程。第二个 commit 针对 fully async fsdp2 策略显存不足问题进行调整, 防止 OOM 导致 CI 失败。

实现拆解

1. 修复工作流依赖安装: 在 model.yml、reward_model_sglang.yml、reward_model_vllm.yml、e2e_one_step_off_policy.yml、e2e_ppo_grpo_trainer_trtlm.yml、sgl.yml、vllm.yml、sanity.yml 等多个工作流中, 在安装阶段补上了 pip3 install -r requirements.txt 步骤, 并增加了 pip3 list 调试命令用于验证依赖状态。
2. 防止 fully async fsdp2 OOM: 修改 tests/special_e2e/run_fully_async_policy.sh 中 FSDP2 分支的默认值, 将 gen_tp 从 1 改为 2, fsdp_size 从 1 改为 2, 从而增大并行度降低每 GPU 显存占用, 避免 OOM。
3. 重写全异步策略 CI 配置: 重构 .github/workflows/e2e_fully_async_policy.yml, 清理过长的注释头部, 调整触发路径和策略, 简化矩阵定义, 并添加并发取消组等标准字段。

关键文件:

- .github/workflows/e2e_fully_async_policy.yml (模块 CI 工作流; 类别 infra; 类型 infrastructure): 全异步策略 CI 配置被全面重写, 修改幅度最大 (+233/-169), 清理注释、调整触发条件和策略, 是本次 PR 的核心基础设施变更。
- tests/special_e2e/run_fully_async_policy.sh (模块 测试脚本; 类别 test; 类型 test-coverage): 修改了 FSDP2 分支的并行参数解决 OOM, 是影响 fully async 测试实际运行的关键修复。
- .github/workflows/model.yml (模块 CI 工作流; 类别 infra; 类型 infrastructure): 作为模型测试工作流, 添加了 requirements.txt 安装和 pip3 list 调试, 是 CI 稳定性修复的代表性变更之一。

- .github/workflows/reward_model_sglang.yml (模块 CI 工作流; 类别 infra; 类型 infrastructure) : 同样增加了 requirements.txt 安装步骤和 pip3 list 调试, 是多个工作流中一致变更的一部分。

关键符号: 未识别

关键源码片段

tests/special_e2e/run_fully_async_policy.sh

修改了 FSDP2 分支的并行参数解决 OOM, 是影响 fully async 测试实际运行的关键修复。

```
# tests/special_e2e/run_fully_async_policy.sh (FSDP2 分支关键片段)
if [ "${ACTOR_STRATEGY}" == "fsdp2" ]; then
    echo "Running fully async training with FSDP2 strategy..."
    # FSDP2 specific parameters
    # trtllm: one replica uses all rollout GPUs as a single TP group.
    # vllm/sglang: TP=1, rely on data parallelism across replicas.
    if [ "${rollout_name}" = "trtllm" ]; then
        gen_tp=${GEN_TP:-${n_gpus_rollout}}
    else
        gen_tp=2 # 改为 2 以避免 OOM, 通过增大 TP 并行降低每 GPU 模型内存占用
    fi
    sp_size=1
    fsdp_size=2 # 同样增大 FSDP 分片数, 减少每 GPU 内存使用
    ref_offload=True
    actor_offload=False
    # ... 设备检测和 NPU 特殊处理部分
    if [ -n "$device_name" ] && [ "$device_name" == "npu" ]; then
        common_params+=(
            actor_rollout_ref.rollout.gpu_memory_utilization=0.70
        )
        actor_offload=True
    fi
    # ... 启动训练命令
fi
```

评论区精华

无 review 评论。

- 暂无高价值评论线程

风险与影响

- 风险: 本次变更为 CI 配置和测试脚本调整, 风险较低。主要风险点在于: 修改了 CI 触发条件 (paths 过滤), 可能导致部分分支或路径不被触发; 添加的 pip3 install -r requirements.txt 可能与已有依赖冲突, 但概率较小; fully async 测试参数调整增加了 GPU 使用量, 可能影响并发排队。

- 影响：影响所有开发者的 CI 体验：修复后主分支 CI 应更稳定，减少因依赖缺失或 OOM 导致的失败。全异步策略测试使用更多并行度，虽增加资源消耗但避免了崩溃。对于 PR 测试，增加了调试信息（pip3 list）有助于快速定位依赖问题。
- 风险标记：CI 触发策略变更，测试并行参数调整，依赖安装顺序变动

关联脉络

- 暂无明显关联 PR