

PR #6956 完整报告

verl-project/verl

[ci] feat: add GSPO Qwen3-8B FSDP2 vLLM nightly test for Ascend

合并时间: 2026-07-07 20:02

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6956>

执行摘要

- 一句话: 新增 Ascend GSPO Qwen3-8B FSDP2 nightly 测试
- 推荐动作: 该 PR 是标准的 CI 增强, 值得关注其路径兼容性问题, 可作为后续改进点。建议在需要非 root 环境运行时修复。

功能与动机

PR 旨在增加 CI 覆盖, 确保持续集成覆盖 GSPO Qwen3-8B FSDP2 训练在 Ascend NPU 上的稳定性。PR body 明确指出 'Add a new nightly CI workflow for GSPO Qwen3-8B FSDP2 training on Ascend using the vLLM rollout backend.'

实现拆解

实现分为两步:

1. 新增测试脚本 `tests/special_npu/nightly_ci_ascend/run_gspo_qwen3_8b_fsdp2_npu.sh`, 设置环境变量、模型路径、训练超参数, 并定义数据、模型、actor、ref、rollout 等配置数组, 最终调用 `verl trainer` 启动训练。
2. 修改 CI 配置文件 `.github/workflows/nightly_ascend.yml`, 新增一个名为 `nightlyCI_gspo-qwen3-8b-fsdp2-vllm_ascend` 的 job, 指定容器镜像、环境变量、步骤, 并调整 cron 时间至 UTC 19:00 以避免与现有 job 冲突。

关键文件:

- `tests/special_npu/nightly_ci_ascend/run_gspo_qwen3_8b_fsdp2_npu.sh` (模块 NPU 测试; 类别 test; 类型 test-coverage): 新增的 GSPO Qwen3-8B FSDP2 vLLM nightly 测试脚本, 包含所有训练参数和环境配置。
- `.github/workflows/nightly_ascend.yml` (模块 CI 配置; 类别 infra; 类型 infrastructure): CI 配置文件, 新增 GSPO Qwen3-8B FSDP2 vLLM nightly job 并调整 cron 调度。

关键符号: 未识别

关键源码片段

`tests/special_npu/nightly_ci_ascend/run_gspo_qwen3_8b_fsdp2_npu.sh`

新增的 GSPO Qwen3-8B FSDP2 vLLM nightly 测试脚本, 包含所有训练参数和环境配置。

```
#!/usr/bin/env bash
# GSPO | Qwen3-8B | GSM8K | vLLM rollout | FSDP2 training | Ascend NPU

set -xeuo pipefail

# 环境变量: RAY、HYDRA、HCCL、VLLM
export RAY_DEDUP_LOGS=0
export HYDRA_FULL_ERROR=1
export TASK_QUEUE_ENABLE=1
export HCCL_EXEC_TIMEOUT=3600
export VLLM_USE_V1=1

# 模型路径, 可通过环境变量覆盖
MODEL_ID=${MODEL_ID:-Qwen/Qwen3-8B}
MODEL_PATH=${MODEL_PATH:-${HOME}/.cache/models/${MODEL_ID}}

# 数据文件: GSM8K 训练 / 验证集
TRAIN_FILE=${TRAIN_FILE:-$HOME/data/gsm8k/train.parquet}
TEST_FILE=${TEST_FILE:-$HOME/data/gsm8k/test.parquet}

# 训练超参数 (CI 默认值保持短运行时间)
NGPUS_PER_NODE=${NGPUS_PER_NODE:-16}
TRAIN_BATCH_SIZE=${TRAIN_BATCH_SIZE:-32}
PPO_MINI_BATCH_SIZE=${PPO_MINI_BATCH_SIZE:-16}
MAX_PROMPT_LENGTH=${MAX_PROMPT_LENGTH:-2048}
MAX_RESPONSE_LENGTH=${MAX_RESPONSE_LENGTH:-2048}
ROLLOUT_N=${ROLLOUT_N:-4}
SP_SIZE=${SP_SIZE:-2}
TOTAL_TRAINING_STEPS=${TOTAL_TRAINING_STEPS:-15}

# GSPO 算法专用参数
CLIP_RATIO_LOW=${CLIP_RATIO_LOW:-3e-4}
CLIP_RATIO_HIGH=${CLIP_RATIO_HIGH:-4e-4}
ACTOR_LR=${ACTOR_LR:-1e-6}
SEED=${SEED:-42}

# 数据配置数组
DATA=(
    algorithm.adv_estimator=grpo
    algorithm.use_kl_in_reward=False
    data.train_files="${TRAIN_FILE}"
    data.val_files="${TEST_FILE}"
    data.train_batch_size=${TRAIN_BATCH_SIZE}
    data.max_prompt_length=${MAX_PROMPT_LENGTH}
    data.max_response_length=${MAX_RESPONSE_LENGTH}
    data.truncation='left'
)

# 模型配置数组
```

```
MODEL=(
  actor_rollout_ref.model.path="${MODEL_PATH}"
  actor_rollout_ref.model.use_remove_padding=True
  actor_rollout_ref.model.enable_gradient_checkpointing=True
)

# Actor 配置数组（关键：使用 GSPO 损失、FSDP2 策略）
ACTOR=(
  actor_rollout_ref.actor.strategy=fsdp2
  actor_rollout_ref.actor.policy_loss.loss_mode=gspo
  actor_rollout_ref.actor.loss_agg_mode=seq-mean-token-mean
  actor_rollout_ref.actor.clip_ratio_low=${CLIP_RATIO_LOW}
  actor_rollout_ref.actor.clip_ratio_high=${CLIP_RATIO_HIGH}
  # ... 余下包含梯度裁剪、动态 batch、优化器、FSDP 配置等
)
```

评论区精华

评论中指出测试脚本硬编码了 `/root/.cache` 作为日志目录，建议使用 `${HOME}/.cache` 以确保非 root 环境兼容。该建议未被采纳，但 PR 最终被批准，表明在 CI root 环境下该问题可接受。

- 硬编码日志目录路径兼容性 (correctness): 评论被提出，但作者未回应；PR 最终被批准，表明在当前的 CI root 环境下该问题不阻塞合并。

风险与影响

- 风险：主要风险：脚本硬编码 `/root/.cache` 路径，在非 root 环境中会因权限失败。但当前 CI 使用 root 容器，因此风险有限。此外，新增 CI job 可能增加队列等待时间。
- 影响：影响范围：仅限夜间 CI 流程，新增一个 job。用户无直接影响。团队需关注该测试的稳定性。
- 风险标记：硬编码日志路径，权限问题，仅限 root 环境

关联脉络

- PR #6924 [doc] chore: update ascend Retool/DAPO/GSPO best practice version compatibility: 该 PR 更新了 Ascend 上 GSPO 最佳实践的版本兼容性，本 PR 新增了 GSPO 的 CI 测试，两者同属 Ascend GSPO 功能线。