

PR #6012 完整报告

verl-project/verl

[fully_async] fix: add fully async grpo qwen3-235b npu script in main branch

合并时间: 2026-04-16 21:03

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6012>

执行摘要

- 一句话: 新增 Qwen3-235B 完全异步 GRPO NPU 训练脚本并修复引擎补丁问题。
- 推荐动作: 建议工程师在类似 NPU 环境部署时参考此脚本的参数配置, 并关注 `override_transformer_config` 修复以确保补丁应用正确性; 该 PR 值得快速浏览以了解 `fully_async` 训练扩展。

功能与动机

根据 PR body, 动机是修复 PR #5583 导致的补丁未正确应用问题, 并为主分支添加 NPU 上的 Qwen3-235B 完全异步 GRPO 训练脚本, 以支持更多硬件和模型组合。

实现拆解

1. 修复 `engine_workers.py`: 重新应用 `override_transformer_config` 函数, 确保 NPU 特定 MindSpeed 补丁能正确应用, 修复了因先前 PR 导致的补丁失效问题。
2. 添加 shell 脚本: 在 `verl/experimental/fully_async_policy/shell/` 目录下创建 `grpo_qwen3_235b_megatron_npu.sh`, 脚本设置了训练参数如模型路径、批次大小、并行配置, 并调用 `fully_async_main` 启动训练。
3. 配置参数调整: 脚本中定义了 GRPO 算法参数、响应长度、性能相关参数 (如 TP/PP/EP 设置) 和 `fully_async` 特定参数 (如节点数和 staleness 阈值), 确保在 NPU 环境中的兼容性。
4. 测试与部署配套: 无显式测试添加, 但脚本直接集成到现有 `fully_async` 训练流程中, 依赖现有 CI 和工作流。

关键文件:

- `verl/experimental/fully_async_policy/shell/grpo_qwen3_235b_megatron_npu.sh` (模块 完全异步; 类别 other; 类型 configuration) : 新增完全异步 GRPO 训练的 NPU 脚本, 配置 Qwen3-235B 模型的训练参数, 是 PR 的核心变更入口。

关键符号: 未识别

关键源码片段

[verl/experimental/fully_async_policy/shell/grpo_qwen3_235b_megatron_npu.sh](#)

新增完全异步 GRPO 训练的 NPU 脚本，配置 Qwen3-235B 模型的训练参数，是 PR 的核心变更入口。

```
#!/usr/bin/env bash
set -xeuo pipefail

# 项目与实验名称配置
project_name='GRPO'
exp_name='GRPO-qwen3-235b-megatron-fully-async'

# 数据与模型路径设置
RAY_DATA_HOME=${RAY_DATA_HOME:-"${HOME}/verl"}
MODEL_PATH=Qwen3-235B-A22B
CKPTS_DIR=${CKPTS_DIR:-"${RAY_DATA_HOME}/ckpts/${project_name}/${exp_name}"}
TRAIN_FILE=gsm8k/train.parquet
TEST_FILE=gsm8k/test.parquet

# 滚动模式与算法参数
rollout_mode="async"
rollout_name="vllm" # 使用 vLLM 引擎
if [ "$rollout_mode" = "async" ]; then
    export VLLM_USE_V1=1
    return_raw_chat="True"
fi
adv_estimator=grpo # 使用 GRPO 优势估计器
use_kl_in_reward=False

# 响应长度配置
max_prompt_length=$((1024 * 8)) # 最大提示长度 8192
max_response_length=$((1024 * 4)) # 最大响应长度 4096

# 性能相关参数
use_dynamic_bsz=False
train_tp=4 # 训练张量并行大小
train_ep=4 # 训练专家并行大小
train_pp=8 # 训练流水线并行大小

# Fully async 特定参数
NNODES_ROLLOUT=${NNODES_ROLLOUT:-8} # 滚动节点数
NNODES_TRAIN=${NNODES_TRAIN:-8} # 训练节点数
NGPUS_PER_NODE=${NGPUS_PER_NODE:-16} # 每节点 GPU 数

# 调用 fully_async_main 启动训练
python -m verl.experimental.fully_async_policy.fully_async_main \
    --config-path=config \
    --config-name='fully_async_ppo_megatron_trainer.yaml' \
    algorithm.adv_estimator=${adv_estimator} \
    data.train_files="${TRAIN_FILE}" \
    data.val_files="${TEST_FILE}" \
```

```
actor_rollout_ref.rollout.free_cache_engine=True \  
data.max_prompt_length=${max_prompt_length} \  
data.max_response_length=${max_response_length} \  
actor_rollout_ref.model.path="${MODEL_PATH}" \  
actor_rollout_ref.actor.megatron.pipeline_model_parallel_size=${train_pp} \  
actor_rollout_ref.actor.megatron.tensor_model_parallel_size=${train_tp} \  
actor_rollout_ref.actor.megatron.expert_model_parallel_size=${train_ep}
```

评论区精华

Review 中仅有 `gemini-code-assist[bot]` 指出引入新脚本和补丁修复，`wucong25` 批准；Issue 评论中 `wuxibin89` 提醒与 PR #5967 重复，作者回应检查，但未深入讨论技术细节。无实质性争议或设计权衡。

- 重复 PR 提醒 (question): 未明确解决重复问题，但 PR 已合并。

风险与影响

- 风险：技术风险：1. 脚本兼容性风险：脚本中的硬件特定参数（如 NPU 配置）可能不适用于所有环境，导致训练失败。2. 补丁回归风险：重新应用 `override_transformer_config` 可能引入新的补丁冲突或影响其他引擎组件。3. 依赖未验证：脚本假设 NPU 驱动和库已正确安装，未在 CI 中全面测试。
- 影响：影响范围：主要针对使用 NPU 进行完全异步 GRPO 训练的用户，提供了新的训练脚本选项。影响程度：中等，扩展了模型和硬件支持，但对核心系统逻辑影响较小。团队需注意脚本参数调整和维护。
- 风险标记：脚本兼容性风险，补丁回归风险，缺少测试覆盖

关联脉络

- PR #5583 未知：此 PR 部分回滚了 PR #5583 的变更，以修复补丁应用问题。
- PR #5967 未知：Issue 评论指出与此 PR 重复，可能涉及类似脚本添加。
- PR #6006 [fully_async] feat: add Qwen3-VL-8B fully async GRPO training script on geo3k: 类似地添加了 `fully_async` 训练脚本，扩展了硬件和模型支持。