PR #6006 完整报告

verl-project/verl

[misc, fully_async] feat: add Qwen3-VL-8B fully async GRPO training script on geo3k

合并时间：2026-04-15 10:26

原文链接：http://prhub.com.cn/verl-project/verl/pull/6006

执行摘要

- 一句话：新增 Qwen3-VL-8B 模型在 geo3k 数据集上的完全异步 GRPO 训练脚本。
- 推荐动作：该 PR 值得快速浏览，了解异步训练配置和参数组织方式。建议关注异步特定参数如 staleness_threshold 和 rollout_correction 的设计，以及配置块的组织模式，以复用最佳实践。

功能与动机

PR body 中提到：'improving GPU utilization by overlapping training and inference'，即通过重叠训练和推理提高 GPU 利用率，解决同步训练中资源未充分利用的问题，为 geo3k 数据集提供一个异步训练示例。

实现拆解

1. 新增脚本文件：在 verl/experimental/fully_async_policy/shell/ 目录下创建 geo3k_qwen3vl_8b_fsdp2_16_16_npu.sh 文件，作为训练入口。
2. 环境与路径设置：设置环境变量如 CUDA_DEVICE_MAX_CONNECTIONS 和数据路径（硬编码为 $HOME/data/geo3k），配置 rollout 模式为异步。
3. GPU 分配配置：定义 n_gpus_rollout=16 和 n_gpus_training=16 等参数，用于 Ray 资源分配，但评论指出可能需调整为每节点 8 个 GPU。
4. 配置块定义：组织参数到命名配置块（如 DATA_CONFIG、ACTOR_CONFIG 等），包含数据、模型、优化器和异步特定参数（如 staleness_threshold、rollout_correction）。
5. 异步参数集成：添加 trigger_parameter_sync_step、require_batches 等参数，支持完全异步训练模式。 无测试、配置或部署配套改动，仅新增示例脚本。

关键文件：

- verl/experimental/fully_async_policy/shell/geo3k_qwen3vl_8b_fsdp2_16_16_npu.sh（模块 异步策略；类别 other；类型 entrypoint）：这是 PR 的唯一变更文件，新增了完全异步 GRPO 训练的 shell 脚本，定义了环境、GPU 分配和训练配置，是异步训练的实现入口。

关键符号：未识别

关键源码片段

verl/experimental/fully_async_policy/shell/geo3k_qwen3vl_8b_fsdp2_16_16_npu.sh

这是 PR 的唯一变更文件，新增了完全异步 GRPO 训练的 shell 脚本，定义了环境、GPU 分配和训练配置，是异步训练的实现入口。

```
# ================================== GPU Allocation ==========================
============
n_gpus_rollout=16 # 分配给rollout任务的GPU数量，用于异步推理
n_gpus_training=16 # 分配给训练任务的GPU数量，用于模型更新
n_nodes_rollout=1 # rollout节点数，当前配置为单节点
n_nodes_train=1 # 训练节点数，当前配置为单节点
# 注意：评论指出此配置可能导致标准硬件（每节点8加速器）上作业挂起，建议调整为每节点8 GPU。

# ================================== Async Config ============================
==========
ASYNC_CONFIG="
  async.staleness_threshold=10 \
  async.trigger_parameter_sync_step=5 \
  async.require_batches=2 \
  async.partial_rollout=True \
  async.rollout_correction=sequence_tis_geometric_rs"
# 注释：ASYNC_CONFIG
定义了异步训练的关键参数，包括陈旧度阈值、参数同步步长和rollout校正机制，以处理训练与推理间的延迟。
```

评论区精华

review 评论中，gemini-code-assist[bot] 指出两个关键问题：

- 数据路径可移植性：硬编码 $HOME/data/geo3k 路径可能导致脚本在不同环境（如 CI/CD）中失效，建议使用环境变量默认值。
- GPU 资源配置：配置每节点 16 个 GPU 可能与标准硬件（如 8 加速器节点）不兼容，导致 Ray 作业挂起，建议调整为每节点 8 个 GPU 并使用多节点。 评论未显示是否采纳建议，PR 被合并，可能存在未解决疑虑。
- 数据路径可移植性 (design): 评论未显示是否采纳建议，PR 被合并，可能未解决可移植性问题。
- GPU 资源配置 (correctness): 评论未显示是否采纳建议，PR 被合并，存在资源配置风险。

风险与影响

- 风险：技术风险：
- 环境依赖硬编码：脚本中的数据路径 $HOME/data/geo3k 缺乏灵活性，在不同用户环境或 CI 中可能无法运行。
- 资源配置不当：GPU 配置（16 个每节点）若与硬件不匹配，可能导致作业无限挂起，影响训练可执行性。

- 兼容性风险：脚本依赖特定模型路径 HF_MODEL_PATH 和 Ascend NPU 环境，缺少跨平台验证。
- 影响：对用户：提供了新的训练示例，帮助用户在 Ascend NPU 上运行 Qwen3-VL-8B 的异步 GRPO 训练，提升资源利用率。对系统：无直接影响，仅添加示例脚本。对团队：扩展了 fully_async 模块的支持范围，增强了多模态模型训练能力，促进异步训练实践。
- 风险标记：环境依赖硬编码，资源配置风险

关联脉络

- PR #5988 [fully_async] feat: enable fully async to log_val_generations: 同属 fully_async 模块，扩展了异步训练功能，涉及类似配置和日志增强。
- PR #5950 [doc] chore: add rloo advantage estimator example script for npu: 类似地添加了 NPU 训练示例脚本，展示了硬件特定训练配置的模式。
- PR #5961 [rollout, vllm] fix: auto-convert disable_mm_preprocessor_cache to mm_processor_cache_gb for vllm >= 0.13.0: 涉及 rollout 和 vLLM 配置，与本脚本中可能使用的 vLLM rollout 模式相关。