PR #5864 完整报告

verl-project/verl

[fully_async] chore: Update fully async dapo qwen3-30b npu script

合并时间：2026-04-03 01:11

原文链接：http://prhub.com.cn/verl-project/verl/pull/5864

执行摘要

本次 PR 更新了完全异步策略下 Qwen3-30B 模型在 NPU 上运行的 DAPO 训练脚本，主要调整了多个超参数配置，包括禁用过长缓冲区、减小批次大小、增加陈旧度阈值、修改 PPO token 长度计算等，旨在使奖励进展与对应的同步脚本保持一致。变更仅影响实验性脚本，风险较低但需注意内存需求和训练稳定性。

功能与动机

根据 PR 描述，本次变更的目的是 " 确保奖励进展与对应的同步脚本对齐 "（ensuring a proper reward progression aligned with the corresponding sync script）。这表明脚本调整是为了解决完全异步训练模式下可能存在的奖励曲线不一致问题，通过超参数调优使异步训练的行为更接近同步训练，提高训练的可预测性和稳定性。

实现拆解

仅修改了 verl/experimental/fully_async_policy/shell/dapo_30b_a3b_math_fsdp_npu.sh 一个文件，主要调整包括：

参数	原值	新值	影响
enable_overlong_buffer	True	False	禁用过长缓冲区处理
train_prompt_mini_bsz	32	16	训练批次大小减半
total_rollout_steps	(64*100)	(32*100)	rollout 总步数减半
staleness_threshold	0.45	0.75	陈旧度阈值大幅增加
PPO token 长度计算分母	8	2	内存需求显著增加
fsdp_size	16	-1	使用自动 FSDP 大小

参数	原值	新值	影响
新增参数	-	recompute=True, max_num_seqs=128, loss_agg_mode="token-mean"	添加梯度检查点和优化设置
模型 dropout 配置	-	添加 attention_dropout=0., embd_pdrop=0., resid_pdrop=0.	禁用 dropout 以稳定训练

评论区精华

review 中仅有一条来自 gemini-code-assist[bot] 的评论，指出脚本中 fsdp_size 变量定义但未使用的问题：

> "The variable fsdp_size is defined at line 70 but is not utilized here. Instead, a hardcoded value of -1 is used. This reduces the maintainability of the script... It is better to use the variable to ensure consistency."

该建议未被采纳，PR 最终以硬编码方式合并，这可能导致未来脚本维护时出现不一致。

风险与影响

技术风险：

1. 训练稳定性风险：staleness_threshold 从 0.45 增至 0.75，可能显著改变异步训练中的参数同步频率，影响收敛行为。
2. 内存风险：PPO token 长度计算分母从 8 改为 2，使最大 token 长度增加 4 倍，在 NPU 环境下可能引发内存不足错误。
3. 维护风险：fsdp_size 硬编码而非使用变量，降低了脚本的可维护性。

影响范围：

- 仅影响使用该特定脚本进行完全异步 DAPO 训练的实验人员。
- 对系统其他模块无直接影响。
- 需要实际训练验证调整效果。

关联脉络

从近期历史 PR 看，完全异步策略相关变更呈现以下趋势：

1. 实验性模块持续优化：PR 5816 同样修改了完全异步策略的脚本，移除硬编码的工具代理循环，与本 PR 同属实验模块的维护工作。
2. NPU 设备支持加强：PR 5795 为 NPU 启用 expandable segment 支持，与本 PR 的 NPU 环境优化相呼应。

3. 训练脚本调优常态化：多个 PR 涉及训练脚本的超参数调整（如 PR 5679、5826），表明项目处于密集的实验调优阶段。

本次 PR 是实验性训练脚本调优的典型代表，反映了团队在异步训练稳定性方面的探索。