

PR #6286 完整报告

verl-project/verl

[misc] refactor: re-format npu examples

合并时间: 2026-05-12 14:26

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6286>

执行摘要

- 一句话: 重构 NPU 示例并新增两个 GRPO 训练脚本
- 推荐动作: 建议读者在使用前先根据 Review 建议修正脚本中的关键问题 (设置 `use_mbridge=True`、定义 `WORK_DIR` 默认值、启用 `set -e`、修正 Hydra 前缀)。该 PR 的价值在于补充了大规模 NPU 训练参考配置, 但质量审核不足, 可作为反面案例提醒开发者注意配置的完整性和正确性。

功能与动机

PR 描述明确说明目的为 're-format npu examples', 补充和重排 Ascend NPU 上的训练示例配置。从内容看, 主要解决 Qwen3 大模型在 NPU 上的 GRPO 训练参考脚本缺失问题。

实现拆解

分两个步骤:

1. 新增 256K 超长上下文 Megatron 脚本: 文件 `examples/ascend_extras/grpo_trainer/run_qwen3_235b_256k_megatron.sh` 提供 Qwen3-235B 模型在 16 节点 256 卡 NPU 上的 GRPO 训练配置, 包含 Megatron 的 TP/EP/PP/CP 设置、vLLM 推理引擎配置, 以及 KL 正则等算法参数。
2. 新增 FSDP 分布式脚本: 文件 `examples/ascend_extras/grpo_trainer/run_qwen3_32b_fsdp.sh` 提供 Qwen3-32B 模型在 4 节点 32 卡上的 GRPO 训练配置, 使用 FSDP 混合精度和参数卸载。两个脚本均通过 `python -m verl.trainer.main_ppo` 启动, 以 Hydra 配置覆盖方式传递参数。

关键文件:

- `examples/ascend_extras/grpo_trainer/run_qwen3_235b_256k_megatron.sh` (模块 示例脚本; 类别 other; 类型 configuration): 新增的 Megatron 后端训练脚本, 支持 256K 超长上下文, 是 PR 核心文件
- `examples/ascend_extras/grpo_trainer/run_qwen3_32b_fsdp.sh` (模块 示例脚本; 类别 other; 类型 configuration): 新增的 FSDP 后端训练脚本, 覆盖不同模型规模的配置需求

关键符号: 未识别

关键源码片段

examples/ascend_extras/grpo_trainer/run_qwen3_235b_256k_megatron.sh

新增的 Megatron 后端训练脚本，支持 256K 超长上下文，是 PR 核心文件

```
#!/bin/bash
# 该脚本为 Qwen3-235B 在 Megatron 后端上的 GRPO 训练示例
# 注意：默认注释了 set -e，实际部署建议启用：
# set -xeuo pipefail

# 项目与实验名称
project_name='GRPO-Qwen3-235B-A22B-Instruct-MATH'
exp_name='GRPO-Qwen3-235B-A22B-Instruct-Megatron-vLLM'

# 节点资源
NNODES=${NNODES:-16}
NPUS_PER_NODE=${NPUS_PER_NODE:-16}

# 模型权重路径 - 注意 WORK_DIR 需在环境中定义或显式赋值
MODEL_PATH=${WORK_DIR}/Qwen3-235B-A22B-Instruct-2507
MCORE_MODEL_PATH=${WORK_DIR}/Qwen3-235B-A22B-Instruct-2507-Mcore

# 训练与测试数据
TRAIN_FILE=${WORK_DIR}/gsm8k/train.parquet
TEST_FILE=${WORK_DIR}/gsm8k/test.parquet

# 数据参数：最长 1K prompt 和 255K response
max_prompt_length=$((1024 * 1))
max_response_length=$((1024 * 255))

# 算法配置：GRPO，不使用 KL in reward，使用 KL loss
adv_estimator=grpo
use_kl_in_reward=False
kl_coef=0.0
use_kl_loss=True
kl_loss_coef=0.001

# 训练批次
train_prompt_bsz=4
n_resp_per_prompt=4
train_prompt_mini_bsz=4

# 性能：全 offload，关闭动态 bsz
all_offload=True
use_dynamic_bsz=False

# 以下 actor 配置部分存在需要特别关注的问题：
# 1. use_dynamic_bsz 关闭
# 2. ppo_epochs=1 避免过拟合
# 3. ** use_mbridge=False 有风险 **: MegatronEngine 要求必须为 True，否则初始化时断言失败
# 4. ** Hydra 语法 ++ 应为 + **
```

```

ACTOR_ARGS=(
  actor_rollout_ref.actor.use_torch_compile=False
  actor_rollout_ref.actor.use_dynamic_bsz=${use_dynamic_bsz}
  actor_rollout_ref.actor.use_kl_loss=${use_kl_loss}
  actor_rollout_ref.actor.kl_loss_coef=${kl_loss_coef}
  actor_rollout_ref.actor.entropy_coeff=0
  actor_rollout_ref.actor.ppo_epochs=1
  actor_rollout_ref.actor.ppo_micro_batch_size_per_gpu=1
  actor_rollout_ref.actor.ppo_max_token_len_per_gpu=${actor_ppo_max_token_len}
  actor_rollout_ref.actor.ppo_mini_batch_size=${train_prompt_mini_bsz}
  actor_rollout_ref.actor.kl_loss_type=low_var_kl
  # 此处应使用 use_mbridge=True, 但脚本错误地用了 False
  # actor_rollout_ref.actor.megatron.use_mbridge=False
)

```

评论区精华

Review 由 [gemini-code-assist\[bot\]](#) 提出五点严重问题：

- `use_mbridge` 必须为 `True` (2 处)：脚本中均设为 `False`，但 `MegatronEngine` 内部有 `assert use_mbridge`，将导致初始化崩溃。
- `WORK_DIR` 未定义：模型路径依赖 `WORK_DIR` 环境变量，但无默认值，若未设置则路径异常。
- 缺少错误处理：shell 选项 `set -e` 被注释，开启后应尽早暴露失败。
- Hydra 语法错误：在 `++` 前缀应为 `+`，可能导致配置解析异常。这些评论均未得到作者回应，但 PR 获得批准合并。
- `actor_rollout_ref.actor.megatron.use_mbridge` 必须为 `True` (correctness): 应改为 `True`，但 PR 未修改
- `actor_rollout_ref.ref.megatron.use_mbridge` 同样必须为 `True` (correctness): 应改为 `True`
- `WORK_DIR` 未初始化 (correctness): 建议添加默认值
`WORK_DIR=${WORK_DIR:-${HOME}/verl}`
- 缺少 `set -e` 等错误处理 (style): 建议启用以提高稳健性
- Hydra 语法错误：`++` 应为 `+` (correctness): 应改为 `+`

风险与影响

- 风险：潜在问题包括：
 1. 配置断言失败：`examples/ascend_extras/grpo_trainer/run_qwen3_235b_256k_megatron.sh` 中 `actor_rollout_ref.actor.megatron.use_mbridge=False` 与 `ref.megatron.use_mbridge=False` 违反 `MegatronEngine` 约束，导致训练无法启动（见 Review 评论）。
 2. 环境变量依赖：`WORK_DIR` 未设置默认值，若运行环境未预定义则路径为根目录，造成数据 / 模型加载失败。

3. 缺少错误穿透: `set -xeuo pipefail` 被注释, 脚本在子命令失败时继续执行, 掩盖错误。
4. Hydra 覆盖语法: `++` 前缀错误使用, 可能被 Hydra 忽略或导致意外行为。- 影响: 直接影响使用 NPU 示例的用户: 若直接运行脚本将遇到 `use_mbridge` 断言错误或路径问题, 无法完成训练。需手动修改配置后方可正常执行。对系统无性能或安全问题, 但影响易用性。团队需尽快修复这些关键问题并更新脚本。- 风险标记: 配置断言错误, 环境变量未初始化, 缺少错误处理, Hydra 语法错误

关联脉络

- PR #6285 [model] chore: refactor npu scripts: 该 PR 重构了 NPU 目录结构, 本 PR 在新目录下补充了示例脚本
- PR #6291 [doc] chore: update vllm and vllm ascend from 0.13.0 to 0.18.0 in docs and dockerfile: 本 PR 脚本中 vLLM 版本依赖需与升级保持一致