PR #6807 完整报告

verl-project/verl

[model] feat: add qwen3-122b long seq launch script for ascend

合并时间：2026-06-22 11:39

原文链接：http://prhub.com.cn/verl-project/verl/pull/6807

执行摘要

- 一句话：新增 Ascend Qwen3.5-122B 长序列 GRPO 启动脚本
- 推荐动作：对于需要在大规模 Ascend 集群上运行 Qwen3.5-122B GRPO 训练的开发者，可直接参考此脚本进行调整。脚本中的环境标志和并行配置值得学习。

功能与动机

在 Ascend NPU 上支持 Qwen3.5-122B 模型的长序列 GRPO 训练，提供一个可复用的启动脚本，方便用户快速部署实验。

实现拆解

1. 环境变量设置：定义兼容 Ascend 的 vLLM 和 Megatron 环境标志，如 VLLM_ASCEND_ENABLE_PREFETCH_MLP、CPU_AFFINITY_CONF 等。
2. 快速配置区：允许用户通过环境变量覆盖节点数、模型路径、数据路径、并行策略（TP/PP/CP/EP）等关键参数。
3. 参数数组定义：按模块组织训练参数，包括 DATA、MODEL、ALGORITHM、ACTOR、ROLLOUT、REFERENCE、REWARD 等数组，每个数组封装了对应模块的配置键值对。
4. 运行命令：通过 torchrun 启动 verl 训练，引用之前定义的数组，并传递节点信息和路径变量。设置奖励模型为同一个 checkpoint，开启 GPU 内存统计等。
5. 修正：review 中指出的 NDEVICES_PER_NODE 变量错误在合并前已修正为 NPUS_PER_NODE。

关键文件：

- *examples/ascend_extras/grpo_trainer/run_qwen3_5_122b_a10b_32k_megatron.sh*（模块 启动脚本；类别 config；类型 configuration）：新增 Qwen3.5-122B 长序列 GRPO 训练启动脚本，定义了完整的训练配置和环境变量，是此 PR 的唯一变更。

关键符号：未识别

关键源码片段

examples/ascend_extras/grpo_trainer/run_qwen3_5_122b_a10b_32k_megatron.sh

新增 Qwen3.5-122B 长序列 GRPO 训练启动脚本，定义了完整的训练配置和环境变量，是此 PR 的唯一变更。

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

# 设置 Ascend 专用环境变量，优化性能
export VLLM_USE_V1=${VLLM_USE_V1:-1} # 启用 vLLM V1
export VLLM_ALLREDUCE_USE_SYMM_MEM=${VLLM_ALLREDUCE_USE_SYMM_MEM:-0} # 关闭对称内存 allreduce
export VLLM_ASCEND_ENABLE_PREFETCH_MLP=${VLLM_ASCEND_ENABLE_PREFETCH_MLP:-1} # MLP 预取
export VLLM_ASCEND_ENABLE_TOPK_OPTIMIZE=${VLLM_ASCEND_ENABLE_TOPK_OPTIMIZE:-1} # TopK 优化
export VLLM_ASCEND_ENABLE_FLASHCOMM1=${VLLM_ASCEND_ENABLE_FLASHCOMM1:-1} # FlashComm1 通信
export CPU_AFFINITY_CONF=${CPU_AFFINITY_CONF:-1} # CPU 亲和性

# 快速配置区：用户可覆盖变量，默认适合 4 节点 64 卡
NNODES=${NNODES:-4}
NPUS_PER_NODE=${NPUS_PER_NODE:-16}
TP=${TP:-2}; PP=${PP:-4}; CP=${CP:-4}; EP=${EP:-16}; ETP=${ETP:-1}
GEN_TP=${GEN_TP:-16}

# 数据配置：32K 长序列，过滤过长 prompt
DATA=(
  data.train_files=${train_path}
  data.val_files=${test_path}
  data.train_batch_size=16
  data.max_prompt_length=$((1024 * 4))
  data.max_response_length=$((1024 * 32))
  data.truncation='error'
  data.filter_overlong_prompts=True
)

# 模型配置：信任远程代码，关闭 remove padding
MODEL=(
  actor_rollout_ref.model.path=${HF_MODEL_PATH}
  actor_rollout_ref.model.trust_remote_code=True
  actor_rollout_ref.model.use_remove_padding=False
)

# 算法：GRPO，KL 不作为奖励
ALGORITHM=(
  algorithm.adv_estimator=grpo
  algorithm.use_kl_in_reward=False
)

# ... 其他数组（ACTOR、ROLLOUT、REFERENCE、REWARD）类似
```

```
# 启动训练，展开所有配置数组
torchrun --nproc_per_node ${NPUS_PER_NODE} \
  --nnodes ${NNODES} \
  --rdzv_endpoint ${MASTER_ADDR}:${MASTER_PORT} \
  -m verl.trainer.ppo.v1.trainer \
  "${DATA[@]}" "${MODEL[@]}" "${ALGORITHM[@]}" \
  # ... 其他数组
  trainer.n_gpus_per_node=${NPUS_PER_NODE} \
  trainer.default_hdfs_dir=null \
  trainer.perf_collection=True
```

评论区精华

gemini-code-assist[bot] 指出脚本第 148 行引用了未定义的变量 NDEVICES_PER_NODE，由于 set -u 会导致脚本崩溃，建议替换为 NPUS_PER_NODE。该建议被采纳，最终脚本使用 NPUS_PER_NODE。

- 未定义变量 NDEVICES_PER_NODE (correctness): 该建议被采纳，最终脚本使用 NPUS_PER_NODE。

风险与影响

- 风险：新增脚本不触及现有逻辑，风险较低。但脚本依赖特定环境（Ascend NPU、vLLM、Megatron），若环境不匹配可能运行失败；变量覆盖和并行策略配置需要用户根据节点资源调整，否则可能 OOM 或性能下降。
- 影响：影响范围为 Ascend 平台用户。提供了 Qwen3.5-122B 模型 32K 长序列训练的参考配置，降低用户在该场景下的搭建成本。对团队而言，丰富了示例库，便于后续维护和推广。
- 风险标记：变量未定义风险，环境依赖强

关联脉络

- PR #6791 [megatron,doc] feat: support DeepseekV4, GLM5, KimiK2.5 via Megatron Lite: 同为添加启动脚本（Megatron Lite 后端示例），与本 PR 一起扩展了 verl 在不同硬件和模型上的示例覆盖。