PR #6982 完整报告

verl-project/verl

[trainer,doc] fix: workaround vLLM multimodal cache in verl 0.8.0, add Qwen3.5 397B

合并时间：2026-07-09 19:08

原文链接：http://prhub.com.cn/verl-project/verl/pull/6982

执行摘要

- 一句话：规避 vLLM 多模态缓存问题并新增 Qwen3.5 397B 支持
- 推荐动作：该 PR 为示例配置与文档更新，适合在 Ascend NPU 上部署 Qwen3.5 模型的用户参考。建议精读新脚本中的并行度选择、MoE 通信策略（alltoall）和张量归一化（use_naive_l2norm）的设置，理解超大模型训练的资源分配方式。

功能与动机

在 verl 0.8.0 分支上，vLLM 的多模态处理器缓存（mm_processor_cache_gb）在 Qwen3.5 系列模型上存在兼容问题，导致 rollout 失败。同时，社区需要支持更大规模的 Qwen3.5-397B 模型以扩展场景。该 PR 通过设置缓存为 0 绕过该问题，并补充 397B 的训练入口和文档。

实现拆解

1. 规避 vLLM 缓存问题：在所有现有 Qwen3.5 GRPO 脚本（run_qwen3_5_35b_fsdp.sh、run_qwen3_5_35b_megatron.sh、run_qwen3_5_122b_a10b_megatron.sh）的 NPU rollout 配置中添加 mm_processor_cache_gb=0，绕过 vLLM 多模态缓存引起的初始化失败。
2. 统一环境变量：将 n_devices_per_node 改为全大写 NDEVICES_PER_NODE，与社区惯例对齐，避免大小写歧义。
3. 回退 Megatron-LM 版本：将 Megatron-LM 从 0.16.1 回退至 0.16.0，确保与现有的 MindSpeed 和 Megatron-Bridge 组合兼容。
4. 新增 397B 训练脚本：创建 run_qwen3_5_397b_megatron.sh，包含 16 节点 A3（256 die）的完整并行配置（TP=2, PP=4, EP=64, GEN_TP=16, GEN_DP=16, GEN_EP=256），并启用 alltoall 和 use_naive_l2norm 优化 MoE 通信与归一化。
5. 更新文档：在 qwen3_5_megatron_npu.md 中添加 397B 的模型行、并行配置表（新增 GEN_DP、GEN_EP 列）、下载命令及启动示例说明。

关键文件：

- *examples/grpo_trainer/run_qwen3_5_397b_megatron.sh*（模块 训练脚本；类别 config；类型 configuration）：新增 Qwen3.5-397B 的 Megatron GRPO 训练脚本，包含 16 节点 A3 的精确并行配置和 MoE 优化参数。

- docs/ascend_tutorial/model_support/examples/qwen3_5_megatron_npu.md（模块 Ascend 文档；类别 docs；类型 documentation）：更新文档以包含 Qwen3.5-397B，添加硬件与并行配置表、下载命令和启动示例。
- examples/grpo_trainer/run_qwen3_5_35b_fsdp.sh（模块 训练脚本；类别 config；类型 configuration）：在 NPU rollout 中设置 mm_processor_cache_gb=0，绕过 vLLM 多模态缓存问题。
- examples/grpo_trainer/run_qwen3_5_122b_a10b_megatron.sh（模块 训练脚本；类别 config；类型 configuration）：添加 mm_processor_cache_gb=0、回退 Megatron-LM 版本到 0.16.0、标准化环境变量、添加 MoE token dispatch 配置。
- examples/grpo_trainer/run_qwen3_5_35b_megatron.sh（模块 训练脚本；类别 config；类型 configuration）：添加 mm_processor_cache_gb=0 和 MoE token dispatch 配置，回退 Megatron-LM 版本。

关键符号：未识别

关键源码片段

examples/grpo_trainer/run_qwen3_5_397b_megatron.sh

新增 Qwen3.5-397B 的 Megatron GRPO 训练脚本，包含 16 节点 A3 的精确并行配置和 MoE 优化参数。

```
#!/usr/bin/env bash
# Qwen3.5-397 MoE GRPO RL with Megatron
#
# Requirements on Ascend:
# - 16 nodes * A3 (16*8*2=256 die)
# - Megatron-LM==0.16.0, MindSpeed==0.16.0, Megatron-Bridge==de93536e
# - verl==0.8.0
#
# Qwen3.5 uses Gated Delta Net (GDN) linear attention which currently does
# NOT support packed sequences (THD format) in Megatron-LM. Therefore:
# - model.use_remove_padding=False (forces bshd compute format)
# - actor.megatron.use_remove_padding=False
# - actor.use_dynamic_bsz=False
#
# Tested parallelism config:
# TP=2 PP=4 CP=1 EP=64 ETP=1 GEN_TP=16 GEN_DP=16 GEN_EP=256

export CUDA_DEVICE_MAX_CONNECTIONS=1
export VLLM_USE_V1=1
export VLLM_ALLREDUCE_USE_SYMM_MEM=0
mkdir -p logs
set -xeuo pipefail

DEVICE=${DEVICE:-$(python3 -c 'import torch_npu' 2>/dev/null && echo npu || echo gpu)}
case "${DEVICE}" in
  gpu)
```

```
    export CUDA_VISIBLE_DEVICES=${CUDA_VISIBLE_DEVICES:-0,1,2,3,4,5,6,7}
    ;;
  npu)
    # NPU-specific optimizations
    export CPU_AFFINITY_CONF=1
    export OMP_NUM_THREADS=1
    export PYTORCH_NPU_ALLOC_CONF=garbage_collection_threshold:0.8
    export VLLM_ASCEND_ENABLE_NZ=0
    ;;
esac
# ... ( 后续 AI 模型路径、数据路径、并行度重载、训练启动命令 )
```

评论区精华

无实质人工讨论。Gemini Code Assist 自动生成代码审核摘要，无评论反馈；合并者 wucong25 直接批准。

- 暂无高价值评论线程

风险与影响

- 风险：
  - 性能风险：mm_processor_cache_gb=0 会禁用 vLLM 多模态处理器缓存，导致每次请求需重新处理媒体输入，可能降低推理吞吐。该 workaround 依赖于后续 vLLM 官方修复。
  - 兼容风险：Megatron-LM 回退至 0.16.0 可能丢失 0.16.1 中的 bugfix（如融合核修复）。
  - 配置依赖风险：新脚本的并行配置针对特定集群（16 nodes × A3），在其他硬件组合下未经测试，用户需自行调整。
  - 标准化风险：环境变量 NDEVICES_PER_NODE 的改名可能导致依赖旧变量名的外部用户脚本失效。
- 影响：
  - 用户影响：Qwen3.5 用户现可用 verl 0.8.0 分支进行 GRPO 训练（含 397B 超大规模模型），NPU 用户必须设置 mm_processor_cache_gb=0。
  - 系统影响：仅涉及示例脚本和文档变更，不触及核心训练器或 Rollout 逻辑。
  - 团队影响：维护 4 个模型特定脚本（35B FSDP、35B Megatron、122B Megatron、397B Megatron），未来建议通过模板化配置减少重复。
  - 风险标记：vLLM 缓存 workaround，版本回退兼容，环境变量标准化

关联脉络

- PR #6955 [env] chore: Adapt qwen3.5 to Docker containers for Verl release 0.8.0: 同一模型系列（Qwen3.5）的容器适配与示例脚本紧密配合。
- PR #6972 [trainer] fix: stabilize Qwen3 Next 80B FSDP rollout on NPU: 同为 NPU 上 Qwen 系列的训练脚本优化。