

PR #6890 完整报告

verl-project/verl

[doc, algo] feat: add Ascend NPU GSPO script for Qwen3-32B

合并时间: 2026-06-29 20:31

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6890>

执行摘要

- 一句话: 新增 Ascend NPU Qwen3-32B GSPO 训练脚本及文档链接
- 推荐动作: 对于需要在 Ascend NPU 上使用 GSPO 算法训练 Qwen3-32B 的用户, 此 PR 提供了可靠的参考实现。建议关注以下几点:
 - GSPO 的配置差异点 (对比 GRPO): `loss_mode=gspo`、`loss_agg_mode=seq-mean-token-mean` 以及不使用 KL 惩罚。
 - FSDP2 与 vLLM rollout 的协同配置, 特别是序列并行度 (SP_SIZE) 和 TP 的分配。
 - 环境变量对性能的影响, 可按需启用 jemalloc 等优化。整体而言, 该 PR 设计清晰, 适合作为在 Ascend 上运行 GSPO 的基础模板。

功能与动机

PR 提出提供一个规范化的 Ascend NPU 训练脚本用于 Qwen3-32B GSPO, 位于 [examples/ascend_extras/gspo_trainer/](#), 并在现有 GSPO 优化实践文档中链入该脚本。参照 PR body: 'Add a canonical Ascend NPU training script for Qwen3-32B GSPO under [examples/ascend_extras/gspo_trainer/](#), and link it from the existing GSPO optimization practice doc.'

实现拆解

1. 创建训练脚本: 在 [examples/ascend_extras/gspo_trainer/](#) 下新建 `run_qwen3_32b_fsdp.sh`。脚本包含完整的 Ascend NPU 环境变量导出、用户可调参数 (项目名、模型路径、数据路径、训练超参数等)、训练配置 (数据、Actor、Rollout 参数) 以及启动 Ray 集群和运行 `verl.trainer.main_ppo` 的命令。脚本采用 FSDP2 策略训练 actor, vLLM 作为 rollout 引擎, 并启用了 GSPO 损失模式 (`loss_mode=gspo`) 和序列级均值聚合 (`loss_agg_mode=seq-mean-token-mean`)。
2. 更新文档: 修改 [docs/ascend_tutorial/model_support/examples/gspo_optimization_practice.md](#), 将最后更新日期改为 2026/06/29, 在文档开头添加指向新脚本的相对链接, 并在正文入口函数说明处补充完整可运行示例的链接, 确保用户能从文档直接跳转到脚本。
3. 配置要点: 脚本中的 GSPO 相关关键参数包括 `clip_ratio_low/clip_ratio_high`、`use_kl_in_reward=False`、`adv_estimator=grpo` 等, 用户可根据实际需求调整。脚本内默认使用 DAPO-Math-17k 数据集和 AIME-2024 测试集。本 PR 未涉及测试或 CI 变更, 因为脚本依赖 Ascend NPU 环境, 无法在通用 CI 中运行。

关键文件：

- `examples/ascend_extras/gspo_trainer/run_qwen3_32b_fsdp.sh`（模块 示例脚本；类别 other；类型 core-logic）：新增的主要训练脚本，包含完整配置和启动命令，是 PR 的核心产出。
- `docs/ascend_tutorial/model_support/examples/gspo_optimization_practice.md`（模块 文档；类别 docs；类型 documentation）：更新了 GSPO 优化实践文档，添加指向新脚本的链接，保持文档与示例一致。

关键符号：未识别

关键源码片段

`examples/ascend_extras/gspo_trainer/run_qwen3_32b_fsdp.sh`

新增的主要训练脚本，包含完整配置和启动命令，是 PR 的核心产出。

```
#!/usr/bin/env bash
# GSPO | Qwen3-32B | DAPO-Math-17k | vLLM rollout | FSDP2 training | Ascend NPU
set -xeuo pipefail

export RAY_DEDUP_LOGS=0
export HYDRA_FULL_ERROR=1
export TASK_QUEUE_ENABLE=1
export HCCL_ASYNC_ERROR_HANDLING=0
export VLLM_USE_V1=1
export VLLM_ATTENTION_BACKEND=XFORMERS
export VLLM_ASCEND_ENABLE_FLASHCOMM=1
export VLLM_ASCEND_ENABLE_PREFETCH_MLP=1
export VLLM_ASCEND_ENABLE_DENSE_OPTIMIZE=1

PROJECT_NAME=${PROJECT_NAME:-GSPO-Qwen3-32B-BASE-MATH}
EXPERIMENT_NAME=${EXPERIMENT_NAME:-GSPO-Qwen3-32B-BASE-FSDP-vLLM}
NNODES=${NNODES:-4}
NPUS_PER_NODE=${NPUS_PER_NODE:-16}
MODEL_PATH=${MODEL_PATH:-Qwen/Qwen3-32B}
TRAIN_FILE=${TRAIN_FILE:-${RAY_DATA_HOME}/dataset/dapo-math-17k.parquet}
MAX_PROMPT_LENGTH=${MAX_PROMPT_LENGTH:-2048}
MAX_RESPONSE_LENGTH=${MAX_RESPONSE_LENGTH:-8192}
TRAIN_BATCH_SIZE=${TRAIN_BATCH_SIZE:-256}
ROLLOUT_N=${ROLLOUT_N:-16}
CLIP_RATIO_LOW=${CLIP_RATIO_LOW:-3e-4}
CLIP_RATIO_HIGH=${CLIP_RATIO_HIGH:-4e-4}
ACTOR_LR=${ACTOR_LR:-1e-6}

DATA=(
  algorithm.adv_estimator=grpo
  algorithm.use_kl_in_reward=False
  data.train_files=${TRAIN_FILE}
  data.val_files=${TEST_FILE}
```

```

data.prompt_key=prompt
data.train_batch_size=${TRAIN_BATCH_SIZE}
data.max_prompt_length=${MAX_PROMPT_LENGTH}
data.max_response_length=${MAX_RESPONSE_LENGTH}
data.truncation='left'
)

ACTOR=(
  actor_rollout_ref.actor.strategy=fsdp2
  actor_rollout_ref.actor.use_torch_compile=False
  actor_rollout_ref.actor.policy_loss.loss_mode=gspo
  actor_rollout_ref.actor.loss_agg_mode=seq-mean-token-mean
  actor_rollout_ref.actor.clip_ratio_low=${CLIP_RATIO_LOW}
  actor_rollout_ref.actor.clip_ratio_high=${CLIP_RATIO_HIGH}
  actor_rollout_ref.actor.learning_rate=${ACTOR_LR}
)

ROLLOUT=(
  actor_rollout_ref.rollout.n=${ROLLOUT_N}
  actor_rollout_ref.rollout.tensor_parallel_size=${ROLLOUT_TP}
  actor_rollout_ref.rollout.gpu_memory_utilization=${ROLLOUT_GPU_MEM_UTIL}
)

```

评论区精华

无实质性 review 讨论。维护者 wucong25 直接批准了 PR，仅有一条自动化代码审查机器人评论（Gemini Code Assist）但未提出具体问题，因此没有需要关注的争议点或决策权衡。

- 暂无高价值评论线程

风险与影响

- 风险：风险较低。主要风险在于：
 - 环境依赖：脚本通过大量 export 设置 Ascend NPU 特有环境变量（如 HCCL_*、VLLM_ASCEND_*），如果用户运行环境缺少对应硬件或软件栈（如 CANN、vLLM Ascend 插件），脚本将无法执行。
 - 配置硬编码：部分路径和参数（如数据路径、模型名称）以默认值形式出现，用户需自行覆盖；未检查路径有效性可能带来运行时错误。
 - 无测试覆盖：由于依赖特定硬件，未在 CI 中包含测试，脚本的正确性需用户实际运行验证。
 - 性能风险：VLLM_USE_V1=1 等优化选项在特定 Ascend 版本下可能不稳定，建议用户根据实际环境调整。整体而言，作为示例脚本，这些风险是可接受的。
- 影响：用户影响：为 Ascend NPU 用户提供了可直接复用的 GSPO 训练脚本，降低了使用门槛；文档更新使得用户能更方便地找到脚本。系统影响：未修改任何核心库代码，不影响现有功能。团队影响：新增一个需要维护的示例脚本，但由于文档中已建立链接，后续变更需保持同步。

- 风险标记：环境依赖敏感，示例脚本未在 CI 测试

关联脉络

- 暂无明显关联 PR