

PR #7788 完整报告

verl-project/verl

[recipe] feat: add GLM-5.2 GRPO training example on Ascend NPUs

合并时间: 2026-09-09 11:46

原文链接: <http://prhub.com.cn/verl-project/verl/pull/7788>

执行摘要

- 一句话: 新增 GLM-5.2 模型在 Ascend NPU 上的 32 节点 GRPO 训练示例脚本。
- 推荐动作: 该 PR 是一个有价值但无需精读代码的示例脚本提交。值得精读的程度较低, 因为变更仅为一个 Shell 脚本, 不涉及核心逻辑。但值得关注的设计决策和配置包括:
 1. 如何为 GLM-5.2 这类 MoE 模型在 32 节点规模下配置专家并行 (EP=64) 和推理并行 (INFER_EP=64)。
 2. 针对 Ascend NPU 优化的 Megatron 配置细节, 如 `use_fused_rmsnorm=True`, `experimental_attention_variant="dsa"`。
 3. GRPO 算法的具体参数设置 (如 `clip_ratio_low=0.2`, `clip_ratio_high=0.28`), 这反映了该训练配方的具体调优思路。

功能与动机

为 GLM-5.2 模型提供在 Ascend NPU 硬件上进行大规模 (32 节点) 强化学习训练的可运行示例。PR body 中明确指出脚本用于在 DAPO-Math-17k 数据集上, 结合 Megatron 和 vLLM-Ascend 进行 GLM-5.2 的 GRPO 训练, 并强调需要设置环境变量 `VERL_VLLM_ASCEND_GLM52_PATCH=1`。

实现拆解

1. 新增训练脚本: 在 `examples/ascend_extras/grpo_trainer/` 目录下新增 `run_glm5_2_megatron.sh` 脚本, 作为 PR 的唯一变更文件, 提供了完整的训练启动命令和配置。
2. 环境与硬件检查: 脚本开头通过 `python3` 检查 `torch.npu.is_available()`, 确保运行在 Ascend NPU 设备上, 并打印详细的依赖版本信息 (`vllm`, `megatron`, `mindspeed` 等)。
3. 训练配置定义: 脚本内部通过变量定义了关键的训练参数, 包括: 并行度 (`PP=8`, `TP=8`, `EP=64`, `INFER_TP=8`, `INFER_DP=8`, `INFER_EP=64`)、训练批次大小、序列长度 (2K prompt, 6K response)、以及 GRPO 算法参数 (KL 损失系数、裁剪比率等)。
4. 命令行参数构建: 将上述配置组装成 `DATA_ARGS`, `MODEL_ARGS`, `ALGORITHM_ARGS`, `ACTOR_ARGS` 等数组, 最终传递给 `verl-trainer` 命令。
5. 依赖环境说明: 脚本头部注释明确列出了运行所需的 `vllm`, `megatron`, `mindspeed` 等软件栈的具体版本和 Commit ID, 为用户复现环境提供精确指引。

关键文件:

- examples/ascend_extras/grpo_trainer/run_glm5_2_megatron.sh (模块 训练示例; 类别 other; 类型 core-logic) : PR 的唯一变更文件, 是完整的 GLM-5.2 在 Ascend 上的 32 节点 GRPO 训练脚本, 包含了所有环境、硬件、模型和算法配置。

关键符号: 未识别

关键源码片段

examples/ascend_extras/grpo_trainer/run_glm5_2_megatron.sh

PR 的唯一变更文件, 是完整的 GLM-5.2 在 Ascend 上的 32 节点 GRPO 训练脚本, 包含了所有环境、硬件、模型和算法配置。

```
#!/bin/bash

# Usage:
# 0. This script is designed to run on Ascend NPUs. Nvidia/AMD GPUs are not supported
# 1. Start a ray cluster of 32 nodes
# 2. Run this script to start the training job

# ... ( 依赖版本声明部分省略 ) ...

set -x

# Stop before configuring training if this machine has no available Ascend NPUs.
if ! python3 -c 'import torch; import torch_npu; raise SystemExit(not torch.npu.is_available())';
then
    echo "Error: This script requires available Ascend NPUs." >&2
    exit 1
fi

# Project Configuration
project_name='verl_megatron_dapo_math_17k_examples'
experiment_name='glm5_2-32nodes'

# Model Weights Paths
MODEL_PATH=${WORK_DIR}/glm52_weights
RAY_DATA_HOME=${RAY_DATA_HOME:-"${HOME}/verl"}
CKPTS_DIR=${CKPTS_DIR:-"${RAY_DATA_HOME}/ckpts/${project_name}/${experiment_name}"}

# ... ( 文件系统路径、训练配置省略 ) ...

NNODES=32

# Train config
PP=8
TP=8
EP=64
ETP=1
CP=1
```

```
INFER_TP=8
INFER_DP=8
INFER_EP=64

# ... ( 其他超参数省略 ) ...

# Actor Model Configuration - 关键并行配置示例
ACTOR_ARGS=(
  actor_rollout_ref.actor.megatron.pipeline_model_parallel_size=$PP
  actor_rollout_ref.actor.megatron.tensor_model_parallel_size=$TP
  actor_rollout_ref.actor.megatron.expert_model_parallel_size=$EP
  actor_rollout_ref.actor.megatron.expert_tensor_parallel_size=$ETP
  actor_rollout_ref.actor.megatron.context_parallel_size=$CP
  actor_rollout_ref.actor.megatron.override_transformer_config.sequence_parallel=True
  actor_rollout_ref.actor.megatron.param_offload=True
  actor_rollout_ref.actor.megatron.optimizer_offload=True
  # ... ( 更多 Megatron 和 GRPO 配置 ) ...
)
```

评论区精华

Review 中的核心讨论集中在提升脚本的易用性和健壮性：

1. 环境变量处理：评审者 wucong25 建议将 `export VERL_VLLM_ASCEND_GLM52_PATCH=1` 默认添加到脚本中，而不是依赖用户手动设置。作者 lxb007981 表示已修复此问题，从最终提交的脚本看，脚本开头确实添加了 `export VERL_VLLM_ASCEND_GLM52_PATCH=1`，并移除了原有对用户手动设置的依赖描述。
 2. 硬件平台声明：wucong25 建议在脚本注释中明确说明当前仅支持 Ascend NPU，不支持 NVIDIA 和 AMD GPU。作者同样表示已修复，最终脚本在 Usage 注释中明确标注了 “This script is designed to run on Ascend NPUs. Nvidia/AMD GPUs are not supported”。
 3. 讨论结论与状态：两个主要问题均已由作者在提交前修复，评审者 wuxibin89 最终批准了 PR。整个讨论过程高效，聚焦于用户友好性和文档清晰度。
- 脚本用户友好性改进 (documentation): 作者 lxb007981 接受建议并已修复。最终脚本在开头自动设置了环境变量，并在 Usage 注释中明确标注了硬件限制。

风险与影响

- 风险：
 1. 环境与版本依赖风险：脚本硬编码了非常具体的软件栈版本和 Commit ID（例如 `vllm-ascendreleases/v0.23.0dc55ef82`）。这些依赖的任何变动都可能导致脚本失效，维护成本较高。此风险在近期 PR #7604 升级 Ascend 栈至 Megatron 0.18.0 后尤为突出。
 2. 配置复杂度风险：脚本中包含大量且相互关联的并行与超参数配置（如 `EP=64`，`recompute_num_layers=1`，`dsa_indexer_loss_coeff=0`）。对于新用户而言，理解和调整这些配置存在较高门槛，若参数组合不当可能导致训练不稳定或性能低下。

3. 硬件特定逻辑风险：脚本依赖 `VERL_VLLM_ASCEND_GLM52_PATCH=1` 环境变量来激活 GLM-5.2 的特定补丁（参见近期 PR #7779）。这是一个隐式依赖，如果环境变量未正确设置，训练可能在运行时而非启动时失败，增加了调试难度。- 影响：用户影响：为希望在 Ascend NPU 集群上运行 GLM-5.2 强化学习训练的用户提供了直接可运行的参考实现，降低了上手门槛。系统影响：本身是独立的示例脚本，对 verl 核心代码库无直接影响。但其存在验证并展示了项目在 Ascend 生态上的集成能力，可能吸引更多用户。团队影响：为硬件 / 模型支持团队提供了 GLM-5.2 在 Ascend 上大规模训练的配置基准，可作为后续模型移植的模板。但同时也带来了维护该示例与上游依赖版本同步的长期责任。

- 风险标记：硬编码依赖版本，配置复杂度高，隐式环境变量依赖

关联脉络

- PR #7779 [vllm, hardware] fix: patch vllm-ascend for GLM-5.2 RL training on Ascend NPU: 本 PR 的训练脚本依赖于 PR#7779 引入的 vllm-ascend 补丁（通过环境变量 `VERL_VLLM_ASCEND_GLM52_PATCH` 激活），两者是功能上的配套关系。
- PR #7604 [megatron, ci] chore: upgrade Ascend stack to Megatron 0.18.0: 本脚本依赖的 Megatron 和 MindSpeed 版本与 PR#7604 升级后的 Ascend 技术栈一致，属于同一升级浪潮下的示例补充。
- PR #7724 [veomni] feat: add async_activation_offload config: 同属为 Ascend/NPU 硬件生态增加训练支持和优化能力的系列工作，虽然引擎不同（Megatron vs VeOmni），但方向一致。