

PR #7678 完整报告

verl-project/verl

[fsdp] feat: update Qwen3.5-35B GRPO example for 2K prompt and 6K response

合并时间: 2026-09-03 09:24

原文链接: <http://prhub.com.cn/verl-project/verl/pull/7678>

执行摘要

- 一句话: 更新 Qwen3.5-35B GRPO 示例脚本的序列长度和训练参数。
- 推荐动作: 该 PR 值得精读, 特别是对于负责维护训练示例和配置的工程师。虽然实现简单, 但它展示了一个针对特定模型 (Qwen3.5-35B)、特定后端 (FSDP-Turbo)、特定硬件 (NPU) 和特定任务 (长上下文 GRPO) 的完整训练配置调整范例。关键的设计决策在于参数间的耦合调整 (如增大序列长度后相应调整 `max_num_batched_tokens` 和 `val_before_train` 等流程控制)。

功能与动机

PR 的动机是调整示例脚本中的 `prompt` 和 `response` 长度设置, 并更新相应的训练配置以适应这个序列长度设置 (“Adjusts the prompt and response length settings... Updates the corresponding training configuration”)。

实现拆解

本次变更集中在一个 shell 脚本文件 `examples/grpo_trainer/run_qwen3_5_35b_fsdp_turbo.sh`, 主要涉及训练超参数和流程控制的调整, 具体步骤如下:

1. 调整核心序列长度参数: 将 `data.max_prompt_length` 从 1024 调整为 2048, `data.max_response_length` 从 2048 调整为 6144。这是本次变更的核心, 旨在支持更长的输入提示和生成的响应, 用于测试模型在长上下文场景下的训练表现。
2. 调整训练流程控制参数: 将 `trainer.val_before_train` 从 `False` 改为 `True`, 启用训练前的验证以确保初始模型状态; 将 `trainer.test_freq` 从 -1 改为 10, 即每 10 个 epoch 进行一次测试; 将 `trainer.total_epochs` 从 15 改为 100, 大幅增加训练轮数以支持更充分的收敛。
3. 补充 NPU 后端特定配置: 在 NPU (`DEVICE=npu`) 的分支中, 新增了 `actor_rollout_ref.rollout.engine_kwargs.vllm.mm_processor_cache_gb=0` (禁用多模态处理器缓存) 和 `actor_rollout_ref.rollout.max_num_batched_tokens=2048` (设置 vLLM 引擎的最大批处理 token 数)。这些配置针对 NPU 硬件和可能的 vLLM 推理后端进行了优化。
4. 同步专家并行配置: 在 NPU 分支开头, 新增了 `EP_SIZE=16` 的环境变量设置, 用于配置专家并行, 确保与 `ep_dispatcher=fused` 的设置匹配。

关键文件:

- `examples/grpo_trainer/run_qwen3_5_35b_fsdp_turbo.sh`（模块 训练示例；类别 `other`；类型 `configuration`）：本次 PR 唯一变更的文件，包含了 Qwen3.5-35B 模型使用 FSDP-Turbo 引擎进行 GRPO 训练的完整启动配置。其参数调整直接决定了训练的行为（序列长度、验证频率、总轮数等）和硬件适配性（NPU 分支配置）。

关键符号：未识别

关键源码片段

`examples/grpo_trainer/run_qwen3_5_35b_fsdp_turbo.sh`

本次 PR 唯一变更的文件，包含了 Qwen3.5-35B 模型使用 FSDP-Turbo 引擎进行 GRPO 训练的完整启动配置。其参数调整直接决定了训练的行为（序列长度、验证频率、总轮数等）和硬件适配性（NPU 分支配置）。

以下是脚本中关键配置参数的调整部分，展示了为支持更长上下文和优化训练流程所做的变更。

```
DATA=(
  data.train_files="${TRAIN_FILE}"
  data.val_files="${TEST_FILE}"
  data.train_batch_size=16
  # --- 核心变更：调整序列长度，支持更长的提示和生成响应 ---
  # 将最大提示长度从 1024 增加到 2048
  data.max_prompt_length=2048
  # 将最大响应长度从 2048 增加到 6144
  data.max_response_length=6144
  data.filter_overlong_prompts=True
  data.truncation='error'
  data.image_key=images
)
```

```
TRAINER=(
  trainer.nnodes=${NNODES}
  trainer.balance_batch=False
  # --- 变更：开启训练前验证 ---
  trainer.val_before_train=True
  trainer.save_freq=-1
  # --- 变更：开启并设置测试频率，增加总训练轮数 ---
  # 每 10 个 epoch 进行一次测试
  trainer.test_freq=10
  # 总训练轮数从 15 增加到 100
  trainer.total_epochs=100
)
```

```
case "${DEVICE}" in
  npu)
    # --- 变更：为 NPU 后端补充 vLLM 推理引擎的特定配置 ---
    ROLLOUT+=(
      # 禁用多模态处理器缓存，可能针对 NPU 内存优化
      actor_rollout_ref.rollout.engine_kwargs.vllm.mm_processor_cache_gb=0
    )
  ;;
esac
```

```
# 设置 vLLM 引擎单次处理的最大 token 数
actor_rollout_ref.rollout.max_num_batched_tokens=2048
)
;;
*)
;;
esac
```

评论区精华

PR 审核过程顺利，无具体讨论内容。审核者 wucong25 直接批准（APPROVED）了变更。

- 暂无高价值评论线程

风险与影响

- 风险：风险较低。本次变更仅修改了一个示例脚本文件，属于配置调整，不涉及核心代码逻辑。主要风险在于配置不当（如序列长度过大导致显存不足）可能导致示例训练无法正常运行，但这种风险仅影响使用该示例的用户，不影响 verl 核心功能。
- 影响：影响范围有限。主要影响希望使用 verl 在 NPU 硬件上运行 Qwen3.5-35B 模型 GRPO 训练的用户，特别是那些需要使用 2K 提示和 6K 响应等长上下文场景的用户。该变更更为用户提供了经过调整的、可直接参考的启动配置。
- 风险标记：配置依赖硬件

关联脉络

- PR #7633 [ci, trainer] refactor: switch fsdp_turbo e2e test from GPU to NPU: PR#7633 将 FSDP-Turbo 的 e2e 测试从 GPU 迁移到了 NPU，与本 PR 修改的 NPU 脚本属于同一功能线（NPU 上的 FSDP-Turbo 支持）。
- PR #7605 [fsdp] feat: add Qwen3.5-2B on-policy distillation FSDP script: PR#7605 新增了 Qwen3.5-2B 的 FSDP 蒸馏脚本，与本 PR 修改的 Qwen3.5-35B GRPO 脚本同属 examples/grpo_trainer 目录下的 FSDP 训练示例，展示了针对不同规模模型的配置。