

PR #6638 完整报告

verl-project/verl

[fsdp] feat: add Qwen3.5-4B on-policy distillation FSDP script

合并时间: 2026-06-06 12:00

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6638>

执行摘要

- 一句话: 新增 Qwen3.5-4B FSDP 蒸馏示例脚本
- 推荐动作: 建议用户在使用该脚本前仔细阅读 review 评论并手动调整默认参数 (特别是资源分配和 offload 配置)。该脚本可作为蒸馏配置的参考模板, 但直接运行时需评估自己的硬件环境。对于维护者, 应考虑后续修复评论中指出的问题, 尤其是资源默认值和变量命名。

功能与动机

提供 Qwen3.5-4B 在 Ascend NPU 上基于策略蒸馏的标准训练入口, 帮助用户快速复现蒸馏实验。PR 提及该脚本与已有 `run_qwen3_8b_fsdp.sh` 的主要区别在于视觉支持、MoE 教师和 FSDP2 配置。

实现拆解

该 PR 仅包含一个 Bash 脚本, 通过参数化配置实现训练启动。

1. 用户可调参数定义: 脚本开头的参数区域允许用户自定义模型路径、节点数、GPU 数、损失模式、批大小、学习率等, 所有参数均提供默认值。
2. 数据与模型配置: 使用 DATA 和 MODEL 数组组织数据文件路径、批大小、截断策略 (`data.truncation='error'`) 和图像键 (`data.image_key=images`), 以及模型路径、梯度检查点等基础设置。
3. Actor 与 Ref 策略配置: 通过 ACTOR 和 REF 数组指定训练策略为 `fsdp2`, 启用 `entropy_checkpointing`, 并设置参数 / 优化器 `offload`。注意此处使用了 `offload_policy=False` 结合手动 `offload` 的方式, 在 review 中被指出不够高效。
4. 蒸馏相关配置: 设置教师模型路径、张量并行度 (`tp`)、专家并行度 (`ep`, 教师为 8)、蒸馏损失模式 (`k1`)、`top-k` 采样数、是否使用策略梯度等。
5. Rollout 与日志输出: 配置 vLLM rollout 引擎的 TP 大小、内存利用率和最大生成长度, 最后启动 Ray trainer 并将日志输出到指定文件。

关键文件:

- `examples/on_policy_distillation_trainer/run_qwen3_5_4b_fsdp.sh` (模块 蒸馏脚本; 类别 other; 类型 core-logic): PR 的唯一变更文件, 实现了 Qwen3.5-4B 基于策略蒸馏的完整训练配置, 包括数据、模型、策略、rollout 等所有环节。

关键符号: 未识别

关键源码片段

examples/on_policy_distillation_trainer/run_qwen3_5_4b_fsdp.sh

PR 的唯一变更文件，实现了 Qwen3.5-4B 基于策略蒸馏的完整训练配置，包括数据、模型、策略、rollout 等所有环节。

```
#!/usr/bin/env bash
# On-policy distillation I vision (Qwen3.5) I vLLM rollout I FSDP training I Ascend NPU/GPU
# dependency: NPU vllm==0.18.0, vllm-ascend@<54879467>, transformers@<cc7ab9be>

set -xeuo pipefail

# ---- user-adjustable ----
STUDENT_MODEL=${STUDENT_MODEL:-Qwen/Qwen3.5-4B}
TEACHER_MODEL=${TEACHER_MODEL:-Qwen/Qwen3.5-35B-A3B}
NNODES=${NNODES:-1}
NGPUS_PER_NODE=${NGPUS_PER_NODE:-8}
TEACHER_WORLD_SIZE=${TEACHER_WORLD_SIZE:-8} # 注意：实际含义是 n_gpus_per_node，多节点下易误用
distillation_loss_mode=${DISTILLATION_LOSS_MODE:-k1}
use_policy_gradient=${USE_POLICY_GRADIENT:-True}
distillation_topk=${DISTILLATION_TOPK:-64}
# ... ( 其余参数省略 )

# Actor 策略配置
ACTOR=(
  actor_rollout_ref.actor.optim.lr=${actor_lr}
  actor_rollout_ref.actor.ppo_mini_batch_size=${ppo_mini_batch_size}
  actor_rollout_ref.actor.strategy=fsdp2
  actor_rollout_ref.actor.fsdp_config.reshard_after_forward=True
  actor_rollout_ref.actor.fsdp_config.entropy_checkpointing=True
  actor_rollout_ref.actor.fsdp_config.offload_policy=False # 推荐改为 True 以启用原生 offload
  actor_rollout_ref.actor.fsdp_config.param_offload=True
  actor_rollout_ref.actor.fsdp_config.optimizer_offload=True
)

# Ref 策略配置
REF=(
  actor_rollout_ref.ref.strategy=fsdp2
  actor_rollout_ref.ref.fsdp_config.offload_policy=False # 类似地推荐改为 True
  actor_rollout_ref.ref.fsdp_config.param_offload=True
)
```

评论区精华

Review 评论由 gemini-code-assist[bot] 提出，主要指出以下问题：

- 默认资源饥饿：单节点 8 卡场景下，学生 pool 请求 8 GPU，教师 pool 也请求 8 GPU，导致 Ray 无法分配资源而挂起。建议降低默认值或增加文档说明。

- 变量命名误导: TEACHER_WORLD_SIZE 实际传递给 distillation.n_gpus_per_node, 在多节点环境中容易误解为总 GPU 数而引发崩溃。建议改为 TEACHER_GPUS_PER_NODE。
- FSDP2 offload 配置次优: 使用 offload_policy=False 并手动设置 param_offload=True、optimizer_offload=True, 不如采用 offload_policy=True 让 FSDP2 原生管理 offload, 以获得更好的性能。
- 日志文件命名误导: 日志文件名包含 35b, 但训练的是 4B 学生模型, 容易造成混淆。建议修正。

以上问题在 PR 合并时均未被采纳或修复。

- 默认配置下资源饥饿导致单节点挂起 (correctness): 未解决, PR 合并时未修改默认值。
- 变量名 TEACHER_WORLD_SIZE 误导 (correctness): 未解决, PR 合并时未重命名。
- FSDP2 offload 配置应使用 offload_policy=True (performance): 未解决, PR 合并时未采纳。
- 日志文件名误导 (35b 应为 4b) (documentation): 未解决, PR 合并时未修正。

风险与影响

- 风险:
 1. 默认资源不足 (examples/on_policy_distillation_trainer/run_qwen3_5_4b_fsdp.sh): 默认 NNODES=1、TEACHER_WORLD_SIZE=8 会导致学生和教师各请求 8 GPU, 在单节点 8 卡上必然死锁。用户若直接运行将无法启动。
 2. 变量名误导 (同一文件): TEACHER_WORLD_SIZE 在多节点场景下 (如 NNODES=2) 若用户误设为 16, 会因 n_gpus_per_node=16 超过物理 GPU 数而崩溃。
 3. Offload 配置非最优: offload_policy=False + 手动 offload 不符合 FSDP2 推荐用法, 可能导致额外内存和性能开销。
 4. 日志混淆: 日志文件名中的 35b 误导用户, 且缺少 mkdir -p 日志目录和 2>&1 错误重定向, 可能导致日志丢失。- 影响: 该 PR 直接影响使用该脚本进行蒸馏训练的用户。由于是示例脚本, 影响范围限于尝试复现 Qwen3.5-4B 蒸馏实验的用户。如果用户不修改默认参数, 将遇到资源挂起等严重问题; 变量名误导可能导致多节点配置失败; offload 配置次优影响训练效率。总体影响程度中等, 需要用户主动调整才能顺利运行。- 风险标记: 默认资源设置不足, 变量名可能误导, offload 配置非最优, 日志文件名误导

关联脉络

- PR #6616 [rollout] chore: change default to calculate_log_probs=True: 该 PR 修改了 rollout 默认配置, 可能间接影响蒸馏脚本的 rollout 行为, 但无直接代码依赖。
- PR #6604 [fsdp] fix: do not manually move model to GPU: 同为 FSDP 相关修复, 涉及模型放置逻辑, 蒸馏脚本若使用 FSDP 可能受潜在影响, 但本脚本未使用该修复中的代码路径。