

PR #6131 完整报告

verl-project/verl

[fsdp, fully_async] feat: add Qwen3-VL-30B-A3B fully async GRPO training script on geo3k

合并时间: 2026-05-13 18:26

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6131>

执行摘要

- 一句话: 修复 Qwen3-VL MoE 路由与 vision pos_embed bug, 添加全异步 GRPO 脚本
- 推荐动作: 建议团队对 fast_pos_embed_interpolate 进行专项测试, 确保在各类输入尺寸下无 tensor/int 转换错误; 考虑添加 NPU 上 MoE 路由权重的单元测试; 确认 qwen3_5 模块中的实现是否需要同步修复。

功能与动机

在 Ascend NPU 上训练 Qwen3-VL-30B-A3B 时发现 reward 不上升、ppo-kl 比 GPU 低几个数量级, 以及 fast_pos_embed_interpolate 因设备设置错误导致训练异常。参照 Qwen3.5 的修复方式, 通过 monkey patch 解决 transformers 版本兼容问题。

实现拆解

步骤 1: 修复 NPU MoE 路由权重计算

- 文件: verl/models/transformers/npu_patch.py
- 变更: 在 NPUQwen3VLMoeTextSparseMoeBlock.forward 中移除 if not self.training 条件, 使 routing_weights = torch.zeros_like(router_logits).scatter_(1, router_indices, routing_weights) 在所有模式下执行。
- 原因: 原始代码在训练时跳过 scatter 操作, 导致路由权重矩阵不正确, 影响 reward 和 KL 指标。移除后训练时也能正确生成 one-hot 形式的路由权重。

步骤 2: 为 Vision Model 添加 fast_pos_embed_interpolate monkey patch

- 文件: verl/models/transformers/monkey_patch.py
- 变更: 1) 导入 Qwen3VLVisionModel 和 Qwen3VLMoeVisionModel; 2) 从 verl.models.transformers.qwen3_5 导入 fast_pos_embed_interpolate; 3) 将 fast_pos_embed_interpolate 赋值给两个 vision model 的对应方法。
- 原因: FSDP2 启用 cpu_offload 时, self.pos_embed 可能位于 CPU 上, 导致设备不匹配。该函数内部根据 grid_thw 获取设备, 避免此问题。

步骤 3: 添加全异步 GRPO 训练脚本

- 文件: verl/experimental/fully_async_policy/shell/geo3k_qwen3vl_30b_a3b_fsdp2_32_32.sh (新增 197 行)

- 内容：配置模型路径、训练数据（geo3k）、FSDP2 训练策略、vLLM 异步 rollout、UL2 序列并行（sp_size=8）、学习率调度等。
- 配套：无测试或配置文件变更，但需要 vllm 0.13.0 和 vllm-ascend 0.13.0 环境。

关键文件：

- verl/models/transformers/npu_patch.py（模块 模型补丁；类别 source；类型 bugfix；符号 NPUQwen3VLMoeTextSparseMoeBlock.forward）：核心 bugfix：移除训练模式条件，确保 MoE 路由权重在训练时正确计算
- verl/models/transformers/monkey_patch.py（模块 模型补丁；类别 source；类型 enhancement；符号 Qwen3VLMoeVisionModel.fast_pos_embed_interpolate, Qwen3VLVisionModel.fast_pos_embed_interpolate）：为 Qwen3-VL vision model 添加 fast_pos_embed_interpolate monkey patch，修复 FSDP2 cpu_offload 导致设备不一致
- verl/experimental/fully_async_policy/shell/geo3k_qwen3vl_30b_a3b_fsdp2_32_32.sh（模块 训练脚本；类别 other；类型 new-config）：新增全异步 GRPO 训练脚本，配置 FSDP2、vLLM 异步 rollout、序列并行等

关键符号：NPUQwen3VLMoeTextSparseMoeBlock.forward, fast_pos_embed_interpolate

关键源码片段

verl/models/transformers/npu_patch.py

核心 bugfix：移除训练模式条件，确保 MoE 路由权重在训练时正确计算

```
# verl/models/transformers/npu_patch.py (head)
# 修改后的 NPUQwen3VLMoeTextSparseMoeBlock.forward 方法

def forward(self, hidden_states: torch.Tensor) -> torch.Tensor:
    batch_size = hidden_states.shape[0]
    hidden_states = hidden_states.reshape(-1, self.hidden_size)
    router_logits = self.gate(hidden_states)
    # softmax 保持 float 高精度
    routing_weights = torch.nn.functional.softmax(router_logits, dim=-1, dtype=torch.float)
    routing_weights, router_indices = torch.topk(routing_weights, self.top_k, dim=-1)
    routing_weights = routing_weights / routing_weights.sum(dim=-1, keepdim=True)
    routing_weights = routing_weights.to(router_logits.dtype)
    hidden_states = hidden_states.reshape(batch_size, -1, self.hidden_size)
    # 关键修复：移除了 if not self.training 条件，始终执行 scatter 操作
    routing_weights = torch.zeros_like(router_logits).scatter_(1, router_indices, routing_weights)
    routed_out = self.experts(hidden_states, routing_weights, router_indices)
    return routed_out
```

verl/models/transformers/monkey_patch.py

为 Qwen3-VL vision model 添加 fast_pos_embed_interpolate monkey patch，修复 FSDP2 cpu_offload 导致设备不一致

```
# verl/models/transformers/monkey_patch.py (head)
# 在 qwen3_vl 分支中新增的导入与 patch
```

```

from transformers.models.qwen3_vl.modeling_qwen3_vl import (
    Qwen3VLForConditionalGeneration,
    Qwen3VLModel,
    Qwen3VLTextModel,
    Qwen3VLVisionModel, # 新增导入
)
from transformers.models.qwen3_vl_moe.modeling_qwen3_vl_moe import (
    Qwen3VLMoeForConditionalGeneration,
    Qwen3VLMoeModel,
    Qwen3VLMoeTextModel,
    Qwen3VLMoeVisionModel, # 新增导入
)
# 从 qwen3_5 模块复用 fast_pos_embed_interpolate
from verl.models.transformers.qwen3_5 import fast_pos_embed_interpolate

# ... 原有 patch 完成后追加以下两行
Qwen3VLMoeVisionModel.fast_pos_embed_interpolate = fast_pos_embed_interpolate
Qwen3VLVisionModel.fast_pos_embed_interpolate = fast_pos_embed_interpolate
print(f"Monkey patch {model.__class__.__name__} model forward")

```

评论区精华

- wucong25 对 npu_patch 的质疑：询问移除 if not self.training 后是否验证过之前场景需要该分支。zhihaofang1017 回复已确认 Qwen3-vl-moe 会进入此分支，移除后正确工作。
- wucong25 对 monkey_patch 的疑问：询问具体修复什么 issue。PR body 说明这是 transformers 已知问题，Qwen3.5 也采用相同方式避免，但对话中未展开。
- gemini-code-assist[bot] 对 fast_pos_embed_interpolate 的审查：指出函数内部使用 tensor 作为 torch.linspace 等的大小参数，可能因未转换为 int 而报错；同时 tolist() 调用导致 GPU-CPU 同步，可能影响性能。该问题来自 qwen3_5 模块，本 PR 只是导入且未修改，建议后续验证。
 - 移除 training 条件的影响 (correctness): 已确认移除条件不影响其他场景，仅在 Qwen3-vl-moe 路径中有效。
 - fast_pos_embed_interpolate 的 tensor/int 转换与性能问题 (performance): PR 作者未直接回应，但该函数从 qwen3_5 模块导入，可能已在其他场景验证。建议后续验证。
 - monkey_patch 修复的具体问题 (question): 未得到明确答复，但已合并，可能已知背景。

风险与影响

- 风险：
 1. fast_pos_embed_interpolate 的健壮性：从 qwen3_5 导入的函数存在 tensor/int 转换风险（如 torch.linspace 传入 tensor 大小），在特定输入下可能报错；频繁 tolist() 可能导致性能开销。需确认已在 Qwen3.5 场景充分测试。
 2. NPU MoE 补丁的回归风险：移除 if not self.training 后，eval 模式也执行 scatter 操作，可能产生与预期不同的路由行为。作者已验证 Qwen3-vl-moe 路径，但其他模型（如 Qwen3 Moe）若使用此 patch 可能受影响。

3. 脚本依赖版本锁定：训练脚本要求 vllm 0.13.0 和 vllm-ascend 0.13.0，升级这些依赖可能导致配置失效。

• 影响：

- 用户：NPU 用户可直接运行 Qwen3-VL-30B-A3B 的全异步 GRPO 训练，获得有效的 reward 曲线和精度。
- 系统：monkey_patch 扩展了 Qwen3-VL 的支持，不影响其他模型分支。npu_patch 的改动影响所有使用 NPU 补丁的模型，但已验证无副作用。
- 团队：新增脚本作为参考示例，降低后续类似模型训练的配置成本。
- 风险标记：fast_pos_embed_interpolate 可能 tensor/int 转换错误，缺失 fast_pos_embed_interpolate 单元测试，NPU MoE 补丁缺少 eval 回归验证

关联脉络

- PR #6076 [fully_async] feat: reuse trainer worker group for hybrid rollout to do validation: 全异步训练核心架构 PR，本 PR 在其基础上添加了具体模型的训练脚本。
- PR #6135 [megatron] fix: the NPU error that occurs after migrating from megatron worker to engine worker.: 修复了 NPU 引擎 worker 的 value_model 支持，与本 PR 的 NPU 场景直接相关。
- PR #6286 [misc] refactor: re-format npu examples: 重构 NPU 示例并新增 GRPO 脚本，与本 PR 新增脚本同属 NPU 训练示例改进。