

# PR #6046 完整报告

verl-project/verl

[fully\_async] fix: preserve per-iteration routed\_experts on partial rollout resume

合并时间: 2026-04-17 22:01

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6046>

## 执行摘要

- 一句话: 修复完全异步训练中部分 rollout 恢复时 routed\_experts 拼接错误, 确保路由专家与生成模型版本一致。
- 推荐动作: 该 PR 值得精读, 特别是对于从事异步训练和 MoE 模型开发的工程师。关注点包括: 1) 如何在部分 rollout 中处理模型版本差异; 2) 路由专家拼接的设计决策 (切片而非替换); 3) review 中关于张量类型的讨论, 展示了实际环境中数据类型的保证。

## 功能与动机

PR body 指出, 在完全异步训练中, 模型权重可能在部分 rollout 迭代间更新, 而之前 #6030 的修复简单替换了整个 routed\_experts 序列, 导致最新调用的路由专家反映的是新模型权重, 而非实际生成早期 token 的旧模型。这会导致训练不稳定, 表现为 actor/ppo\_kl 指标波动剧烈。作者通过实验验证, 保留每个迭代实际生成 token 的路由专家能使 PPO KL 散度更平滑。

## 实现拆解

1. 导入依赖调整: 在 verl/experimental/fully\_async\_policy/agent\_loop/agent\_loop.py 中添加 import torch, 为后续张量拼接提供支持。
2. 核心逻辑重构: 修改 generate 方法中 routed\_experts 的处理逻辑:
  - 当 output.routed\_experts 不为空且生成了新 token 时 ( $\text{len}(\text{output.token\_ids}) > 0$ ) 才进行处理。
  - 如果是首次迭代 ( $\text{final\_output.routed\_experts}$  is None), 则直接使用完整的 output.routed\_experts (覆盖 prompt 和第一批 token)。
  - 如果是后续迭代, 则切片  $\text{output.routed\_experts}[-\text{len}(\text{output.token\_ids}):]$  获取新生成 token 的路由专家, 并通过 torch.cat 与现有的 final\_output.routed\_experts 拼接。
  - 添加  $\text{len}(\text{output.token\_ids}) > 0$  的防护, 避免 [-0:] 切片错误。
3. 注释更新: 将原注释替换为说明模型版本差异和保留每个迭代路由的原因。
4. 无测试配套: 由于变更需要 MoE 模型、异步部分 rollout 和多 GPU 环境, 作者通过实验验证而非单元测试。

关键文件:

- verl/experimental/fully\_async\_policy/agent\_loop/agent\_loop.py (模块 异步策略; 类别 source; 类型 core-logic; 符号 generate): 这是唯一变更的文件, 包含了修复

routed\_experts 拼接逻辑的核心实现。

关键符号: generate

## 关键源码片段

[verl/experimental/fully\\_async\\_policy/agent\\_loop/agent\\_loop.py](#)

这是唯一变更的文件, 包含了修复 routed\_experts 拼接逻辑的核心实现。

```
async def generate(
    self,
    request_id: str,
    prompt_ids: list[int],
    sampling_params: dict,
    image_data: Optional[list[Any]] = None,
    video_data: Optional[list[Any]] = None,
) -> TokenOutput:
    # ... 初始化 final_output 等代码 ...
    while True:
        # 1. 生成token
        output = await super().generate(
            request_id=request_id,
            prompt_ids=prompt_ids + final_output.token_ids,
            sampling_params=sampling_params,
            image_data=image_data,
            video_data=video_data,
        )

        # 2. 合并输出到final_output
        final_output.token_ids.extend(output.token_ids)
        if output.log_probs is not None:
            final_output.log_probs.extend(output.log_probs)
        # 在部分rollout恢复时, 模型版本可能不同, 因此保留现有路由,
        # 仅追加新生成token的路由专家。
        if output.routed_experts is not None and len(output.token_ids) > 0:
            if final_output.routed_experts is None:
                # 首次迭代: 使用完整的routed_experts (覆盖prompt和第一批token)
                final_output.routed_experts = output.routed_experts
            else:
                # 后续迭代: 切片新生成token的路由专家并与现有路由拼接
                final_output.routed_experts = torch.cat(
                    [final_output.routed_experts, output.routed_experts[-len(output.token_ids) :]],
                    dim=0,
                )
        if output.num_preempted is not None:
            final_output.num_preempted += output.num_preempted
        final_output.stop_reason = output.stop_reason
        # ... 其余代码 (更新max_new_tokens、检查停止原因等) ...
```

## 评论区精华

reviewer [gemini-code-assist\[bot\]](#) 建议使用 `torch.as_tensor` 确保输入为张量，避免 `output.routed_experts` 可能返回 `numpy.ndarray` 或列表导致的 `TypeError`。作者 [NoonePauseferg](#) 回复指出，在当前上下文中 `routed_experts` 始终是 torch 张量（`sglang` 和 `vllm` 引擎均返回张量），因此未采纳该建议。最终 [wuxibin89](#) 批准了 PR，但未进一步讨论。

- `routed_experts` 拼接时的张量类型安全 (correctness): 作者未采纳建议，认为现有实现已保证类型安全。

## 风险与影响

- 风险：1. 回归风险：修改了 `routed_experts` 的拼接逻辑，如果切片索引计算错误或张量维度不匹配，可能导致下游路由重放 (router replay) 功能异常。2. 性能风险：使用 `torch.cat` 拼接张量可能增加内存开销，但影响较小，因为 `routed_experts` 通常规模不大。3. 兼容性风险：依赖 torch 导入，如果环境中未安装 torch 或版本不兼容，可能导致运行时错误。4. 测试覆盖不足：PR body 指出由于需要复杂环境，未添加单元测试，仅通过实验验证，可能存在未覆盖的边缘情况。
- 影响：1. 对用户影响：使用完全异步训练、部分 rollout 和 MoE 模型的用户将受益于更稳定的训练过程，PPO KL 散度波动减小。2. 对系统影响：修复了路由专家与生成模型版本不一致的问题，提升了异步训练中路由重放的准确性。3. 对团队影响：改进了 #6030 的修复，展示了在异步训练中模型权重更新对路由专家一致性的重要性，为后续类似问题提供参考。
- 风险标记：核心路径变更，缺少测试覆盖，依赖新增导入

## 关联脉络

- PR #6030 [fsdp, sft] feat: add Gemma4 FSDP SFT support: PR body 提到本 PR 是对 #6030 的改进，但 #6030 实际是关于 Gemma4 FSDP SFT 支持，可能引用错误。更相关的可能是历史 PR 中 #6029 (修复完全异步策略中 `routed_experts` 问题)，但未在上下文中明确提及。
- PR #6029 [fully\_async] fix: replace routed\_experts on partial rollout resume i...: 历史 PR 分析显示 #6029 修复了完全异步策略中部分 rollout 恢复时 `routed_experts` 重复拼接问题，与本 PR 主题高度相关，可能为同一问题线的后续修复。