# PR #5591 完整报告

verl-project/verl

[fsdp] fix: pass dp_group to prepare_dynamic_batch to fix CUDA deadlock

合并时间：2026-03-26 16:06

原文链接：http://prhub.com.cn/verl-project/verl/pull/5591

## 执行摘要

本 PR 修复了 FSDP 训练中因动态批处理 micro-batch 计数不同步导致的 CUDA 死锁问题，通过向 prepare_dynamic_batch 函数传递 dp_group 参数确保所有数据并行 rank 同步，提升训练稳定性。

## 功能与动机

动机源于 commit f5c34bb 在 verl/utils/seqlen_balancing.py 中添加了 dp_group is not None guard，导致 FSDP actor 调用 prepare_dynamic_batch 时未传递 dp_group 参数，跳过 all_reduce 操作。在动态批处理场景下，不同 rank 因序列长度分布差异计算不同 micro-batch 数量，进而引发 FSDP collectives 死锁。PR body 指出：" 不同序列长度分布导致不同 rank 计算不同微批次计数 ...FSDP collectives (AllGather/ReduceScatter) 死锁 "。

## 实现拆解

实现集中于单个文件 verl/workers/actor/dp_actor.py：

- 在 compute_log_prob 函数第 468 行附近，修改 prepare_dynamic_batch 调用，添加 dp_group=torch.distributed.group.WORLD。
- 在 update_policy 函数第 560 行附近，进行相同修改。

关键代码片段：

```
micro_batches, batch_idx_list = prepare_dynamic_batch(
    data, max_token_len=max_token_len, dp_group=torch.distributed.group.WORLD
)
```

这确保所有 rank 通过 all_reduce(MAX) 同步 micro-batch 计数，防止死锁。

## 评论区精华

Review 讨论简单，仅有 gemini-code-assist[bot] 的评论：

> "The fix, which involves passing torch.distributed.group.WORLD as the dp_group to prepare_dynamic_batch in both compute_log_prob and update_policy, is a robust solution."

评论肯定了修复的有效性，无争议点。wuxibin89 直接批准。

## 风险与影响

- 风险：变更引入 dp_group 参数，使用 WORLD 在所有 FSDP 并行配置下正确，但需确保其他调用 prepare_dynamic_batch 的地方也正确处理该参数。PR 提供了在 2-node EKS 集群的测试验证，但缺少单元测试覆盖。
- 影响：修复了 FSDP 动态批处理训练中的死锁 bug，影响所有使用该配置的用户，提升训练可靠性和性能。测试显示修复后 4/4 运行成功，而死锁前 100% 再现。

## 关联脉络

本 PR 是 #5451 的 FSDP 对应版本，#5451 修复了 megatron workers 中的相同 bug。这表明动态批处理同步问题在多个并行策略中普遍存在，需要跨模块统一处理。近期历史 PR 如 #5604 涉及 FSDP workers 重构，但本 PR 专注于具体 bugfix。