PR #6226 完整报告

verl-project/verl

[reward] fix: compute correct rollout world size

合并时间：2026-05-06 17:43

原文链接：http://prhub.com.cn/verl-project/verl/pull/6226

执行摘要

- 一句话：修复 reward model 并行度计算偏差
- 推荐动作：该 PR 为明确的 bugfix，变更量小、逻辑清晰，值得快速审阅合并。建议确认是否有必要添加单元测试或集成测试覆盖 DP > 1 和 PP > 1 的场景，以防止未来回归。

功能与动机

在 RewardModelManager._initialize_llm_servers 中，rollout_world_size 被错误计算为仅包含 tensor_model_parallel_size，忽略了 data_parallel_size 和 pipeline_model_parallel_size。这导致 num_replicas = world_size // rollout_world_size 在 DP > 1 或 PP > 1 时被高估，从而创建过多的 reward model 副本，资源池分割错误。

实现拆解

修改 verl/experimental/reward_loop/reward_model.py 文件中 RewardModelManager._initialize_llm_servers 方法第 51 行，将原有的一行赋值：rollout_world_size = self.config.rollout.tensor_model_parallel_size 替换为：rollout_world_size = self.config.rollout.tensor_model_parallel_size * self.config.rollout.data_parallel_size * self.config.rollout.pipeline_model_parallel_size。此修改确保 rollout_world_size 正确反映每个 rollout 副本实际使用的 GPU 数量，与项目中其他位置（如 llm_server.py:295-299 和 vllm_rollout.py:79-83）的计算方式保持一致。

关键文件：

- verl/experimental/reward_loop/reward_model.py（模块 reward 模型；类别 source；类型 data-contract；符号 RewardModelManager, _initialize_llm_servers）：核心修复文件，修改 _initialize_llm_servers 方法中 rollout_world_size 的计算方式，从仅使用 tensor_model_parallel_size 扩展为乘以 data_parallel_size 和 pipeline_model_parallel_size。

关键符号：_initialize_llm_servers

关键源码片段

verl/experimental/reward_loop/reward_model.py

核心修复文件，修改 _initialize_llm_servers 方法中 rollout_world_size 的计算方式，从仅使用 tensor_model_parallel_size 扩展为乘以 data_parallel_size 和 pipeline_model_parallel_size。

```
class RewardModelManager:
    # ... 省略 __init__ 等代码 ...

    def _initialize_llm_servers(self):
        # 修复：rollout_world_size 现在正确地将所有并行度维度相乘，
        # 而不是仅使用 tensor_model_parallel_size。
        # 这与 llm_server.py:295-299 和 vllm_rollout.py:79-83 中的计算一致。
        rollout_world_size = (
            self.config.rollout.tensor_model_parallel_size
            * self.config.rollout.data_parallel_size # 之前缺失
            * self.config.rollout.pipeline_model_parallel_size # 之前缺失
        )
        world_size = (
            self.resource_pool.world_size
            if self.resource_pool # colocate mode
            else self.config.n_gpus_per_node * self.config.nnodes # standalone mode
        )
        # num_replicas 现在基于完整的 rollout_world_size 计算，
        # 从而在每个副本使用正确数量的 GPU 时，副本数量不会过多。
        num_replicas = world_size // rollout_world_size
        # ... 省略后续代码 ...
```

评论区精华

该 PR 无 review 讨论。只有来自 gemini-code-assist[bot] 的自动评论，确认变更内容正确且无反馈，以及维护者 wuxibin89 的直接批准。

- 暂无高价值评论线程

风险与影响

- 风险：风险较低。变更仅涉及单行表达式，逻辑清晰且已有其他文件中的相同计算作为参考。但需注意：如果未来在其他地方也有类似的 rollout_world_size 计算，可能仍然存在遗漏；本次未包含相关测试，回归风险虽小但理论上存在。
- 影响：影响范围限于 verl/experimental/reward_loop/reward_model.py 中的 RewardModelManager._initialize_llm_servers 方法。只有使用 DP > 1 或 PP > 1 配置的 reward model 启动流程会受到影响。对于 DP = 1 且 PP = 1 的配置，行为不变。修复后，reward model 副本数量和资源池分割将正确反映实际并行度，避免资源浪费或分配错误。
- 风险标记：缺少测试覆盖，已被回滚

关联脉络

- PR #6258 Revert "[reward] fix: compute correct rollout world size": 此 PR (#6258) 正是对本 PR (#6226) 的回滚，说明后续可能发现问题或决定撤销该修复。需注意时序：本

PR 先被合并，而后被回滚。