

PR #6258 完整报告

verl-project/verl

Revert "[reward] fix: compute correct rollout world size"

合并时间: 2026-05-07 10:17

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6258>

执行摘要

- 一句话: 回滚 reward model 的 world_size 修复
- 推荐动作: 建议关注后续 PR #6265 对 RewardModelConfig 的修复, 该 PR 会为配置类添加缺失的字段, 使 #6226 的正确计算可以被安全地重新应用。当前回滚是临时解决方案。

功能与动机

PR body 和关联 Issue 指出, reward model 的配置对象 `RewardModelConfig` 不包含 `pipeline_model_parallel_size` 属性, 因此 #6226 的乘法修正会引发 `AttributeError`。PR 作者 wuxibin89 通过回滚恢复到之前仅使用 `tensor_model_parallel_size` 的逻辑。

实现拆解

1. 回滚计算逻辑: 将 `verl/experimental/reward_loop/reward_model.py` 中 `_initialize_llm_servers` 方法的 `rollout_world_size` 计算从三元乘积恢复为仅用 `tensor_model_parallel_size`。
2. 文件修改: 仅修改 `reward_model.py`, 删除 5 行、增加 1 行。
3. 无测试配套: 本次回滚无配套的测试或配置变更。

关键文件:

- `verl/experimental/reward_loop/reward_model.py` (模块 奖励模型; 类别 source; 类型 data-contract; 符号 `RewardModelManager._initialize_llm_servers`): 回滚的核心文件, 调整了 `rollout_world_size` 计算方式。

关键符号: `RewardModelManager._initialize_llm_servers`

关键源码片段

`verl/experimental/reward_loop/reward_model.py`

回滚的核心文件, 调整了 `rollout_world_size` 计算方式。

```
# verl/experimental/reward_loop/reward_model.py (head)
class RewardModelManager:
    def _initialize_llm_servers(self):
        # 回滚后: 仅使用 tensor_model_parallel_size
        # 若 DP>1 或 PP>1, 此值小于实际世界大小, 会导致 num_replicas 多算
```

```
rollout_world_size = self.config.rollout.tensor_model_parallel_size
world_size = (
    self.resource_pool.world_size
    if self.resource_pool
    else self.config.n_gpus_per_node * self.config.nnodes
)
num_replicas = world_size // rollout_world_size
# ... 后续创建 rollout_replicas ...
```

评论区精华

Review 中 [gemini-code-assist\[bot\]](#) 指出回滚是错误的，会导致 DP>1 或 PP>1 时资源分配失败，因为 `RolloutReplica` 内部仍使用三元乘积计算 `world_size`，两者不一致会导致 `num_replicas` 计算错误。原 PR 作者 [guillemgt](#) 在 Issue 评论中承认 `RewardModelConfig` 类型注解暗示字段存在，但实际上未被使用，已提交 #6265 修复配置问题。

- `rollout_world_size` 计算回滚的正确性 (correctness): 未解决，回滚被合并，但 reviewer 的疑虑未被充分回应。

风险与影响

- 风险：回滚后，若用户配置了 `data_parallel_size > 1` 或 `pipeline_model_parallel_size > 1`，`reward_model` 管理的分布式服务器数量计算将再次错误，可能导致资源分配过少或进程崩溃。这是一个回归风险。
- 影响：影响范围限于实验性模块 `verl/experimental/reward_loop` 中的 `reward model manager`。使用默认并行配置 (DP=1, PP=1) 的用户不受影响。
- 风险标记：回归风险，配置兼容性

关联脉络

- PR #6226 [reward] fix: compute correct rollout world size: 本 PR 直接回滚该 PR 的变更。
- PR #6265 [reward] fix: add missing parallel config fields to `RewardModelConfig`: Issue 评论中 [guillemgt](#) 提到该 PR 将修复配置类字段缺失问题，从而使 #6226 的正确计算可以被重新应用。