PR #5705 完整报告

verl-project/verl

[megatron, ckpt] fix: set dist_ckpt_optim_fully_reshardable default to False

合并时间：2026-03-23 12:25

原文链接：http://prhub.com.cn/verl-project/verl/pull/5705

执行摘要

- 一句话：将 dist_ckpt_optim_fully_reshardable 默认值设为 False，以避免检查点保存时的高 CPU 内存 OOM。
- 推荐动作：建议：此 PR 值得精读，特别是对于使用大模型训练的团队。重点阅读 verl/trainer/config/engine/megatron.yaml 的更改和文档中的警告部分，了解 fully reshardable 与 dp-reshardable 格式的权衡，以确保正确配置检查点策略。

功能与动机

根据 PR body，更改旨在 'Disable fully reshardable optimizer checkpoint format by default to avoid high CPU memory usage during checkpoint saving. Fully-reshardable format requires gathering optimizer states on DP rank 0, which may lead to CPU memory spikes and OOM issues.' 这与 issue #5670 相关，Issue 评论中引用 PR #5154 报告了模型大于 30B 时的稳定 CPU 内存 OOM 问题。

实现拆解

实现方案涉及三个文件：1) verl/trainer/config/_generated_ppo_megatron_trainer.yaml：将 dist_ckpt_optim_fully_reshardable 从 True 改为 False，影响 PPO 训练器配置；2) verl/trainer/config/engine/megatron.yaml：同样修改默认值，控制 Megatron 引擎的检查点行为；3) docs/advance/checkpoint.rst：更新文档，添加关于 optimizer checkpoint 格式的说明和警告，详细解释 dp-reshardable（默认）与 fully-reshardable 格式的内存和性能权衡。

关键文件：

- verl/trainer/config/engine/megatron.yaml（模块 megatron）：核心 Megatron 引擎配置文件，设置默认值为 False 直接影响检查点行为，属于关键路径变更。
- verl/trainer/config/_generated_ppo_megatron_trainer.yaml（模块 trainer）：训练器配置文件，影响 PPO 训练流程的检查点设置，覆盖多个 actor 和 critic 配置。
- docs/advance/checkpoint.rst（模块 doc）：文档更新提供关键信息，解释不同 checkpoint 格式的权衡和内存影响，帮助用户避免错误配置。

关键符号：未识别

评论区精华

review 中，gemini-code-assist[bot] 评论强调更改的重要性：'Setting dist_ckpt_optim_fully_reshardable to False by default is a significant change that impacts the memory usage during checkpointing. Ensure that this change is thoroughly tested across different model sizes and parallelism configurations to prevent regressions.' Begunner 批准了 PR，表明变更被接受，但测试建议未在讨论中明确验证。

- 更新日期验证 (style): PR 已合并，日期更改可能被接受，但未明确验证正确性。
- 测试建议以预防回归 (testing): 更改被批准，但测试建议未在评论中进一步讨论或验证，可能存在未解决的疑虑。

风险与影响

- 风险：风险包括：1) 默认值更改可能影响依赖 fully reshardable 格式的用户，特别是在恢复检查点时需要不同并行配置的场景；2) 如果没有充分测试，可能在特定模型大小或并行配置下引入兼容性问题或回归，尤其针对核心文件 verl/trainer/config/engine/megatron.yaml；3) 文档更新虽然详细，但用户可能忽略警告，导致配置错误。
- 影响：影响：1) 对用户而言，默认情况下检查点保存的 CPU 内存使用降低，减少 OOM 风险，特别是对于大模型训练，但牺牲了检查点恢复时的灵活性（仅支持 DP 重分片）；2) 系统方面，优化了内存性能，可能轻微影响保存速度；3) 团队需要更新配置并关注文档变更，影响范围集中在使用 Megatron 训练模块的用户。
- 风险标记：缺少测试覆盖，配置更改影响核心路径

关联脉络

- PR #5154 未知（从 Issue 评论推断为 CPU 内存 OOM 相关）：Issue 评论引用此 PR 报告了模型大于 30B 时的稳定 CPU 内存 OOM 问题，是本更改的直接触发点。