PR #5745 完整报告

verl-project/verl

[2/2][rollout,trainer] feat: Teacher colocate mode

合并时间：2026-03-26 11:52

原文链接：http://prhub.com.cn/verl-project/verl/pull/5745

执行摘要

- 一句话：添加教师模型 colocate 模式，支持在 rollout 后计算教师 logprobs。
- 推荐动作：建议技术管理者和工程师精读此 PR，特别关注教师 logprobs 计算路径的设计决策，如 stream_teacher_with_rollout 标志的使用和批处理实现。同时，检查 review 中指出的 bug 是否已在提交历史中妥善解决，并评估测试覆盖是否充分。

功能与动机

根据 PR body，此 PR 是 #5723 的延续，旨在解决在 student rollouts 完成后计算教师 logprobs 的问题，以避免教师模型在 rollout 期间占用资源。@wuxibin89 指出了从当前 colocate 模式切换到新模式的必要性，以优化蒸馏训练流程。

实现拆解

实现方案分为三个核心层次：首先，在 agent_loop.py 中引入 stream_teacher_with_rollout 标志，基于 distillation_config.teacher_model.enable_resource_pool 控制教师 server 的初始化；其次，在 teacher_manager.py 中添加 compute_teacher_logprobs_batch 函数支持批处理计算，并修改 _unpad_teacher_inputs 和 compute_teacher_logprobs_single 以优化输入格式；最后，在 teacher_model.py 和 ray_trainer.py 中集成 colocate 计算路径，通过 _compute_teacher_colocate 等方法在训练流程中调用。

关键文件：

- verl/experimental/teacher_loop/teacher_manager.py（模块 teacher_loop）：核心修改文件，添加 compute_teacher_logprobs_batch 函数和 _unpad_teacher_inputs 逻辑，是教师 logprobs 计算的关键模块，review 中讨论了多个 bug 风险。
- verl/experimental/agent_loop/agent_loop.py（模块 agent_loop）：引入 stream_teacher_with_rollout 标志，控制教师 server 的初始化和计算路径，影响蒸馏模式切换逻辑。
- verl/trainer/ppo/ray_trainer.py（模块 trainer）：集成 colocate 计算路径，添加 _compute_teacher_colocate 和 _should_compute_teacher_colocate 方法，直接影响训练流程中的教师 logprobs 计算时机。

关键符号：compute_teacher_logprobs_batch, _compute_teacher_colocate, _should_compute_teacher_colocate, _unpad_teacher_inputs

评论区精华

Review 讨论聚焦于三个关键点：gemini-code-assist[bot] 指出 _pad_teacher_outputs 函数可能因张量维度错误导致 RuntimeError，并建议 compute_teacher_logprobs_batch 添加空输入处理检查；wuxibin89 建议 compute_teacher_logprobs_single 接收 input_ids 而非 prompt_ids 和 response_ids 以简化逻辑，JacobHelwig 回应并提供 diff 表示已采纳；这些讨论帮助识别了潜在 bug 并优化了设计，但部分问题如 _pad_teacher_outputs 的修复状态未明确确认。

- _pad_teacher_outputs 张量维度错误 (correctness): 从 commit 历史看可能有修复（如提交 'Unpad teacher inputs'），但未在 review 中明确确认。
- compute_teacher_logprobs_batch 空输入处理 (correctness): review 中未明确采纳，但代码变更可能已隐含处理，需进一步验证。
- compute_teacher_logprobs_single 输入格式优化 (design): JacobHelwig 回应并提供 diff，表明已修改为接收 sequence_ids，优化了设计。

风险与影响

- 风险：技术风险包括：_pad_teacher_outputs 函数在 verl/experimental/teacher_loop/teacher_manager.py 中可能存在张量维度不匹配，导致运行时错误；compute_teacher_logprobs_batch 未处理空输入，可能引发 torch.cat 异常；stream_teacher_with_rollout 标志的引入增加配置复杂性，需确保在不同模式下正确设置；测试文件有修改但覆盖度有限，可能遗漏回归场景。
- 影响：对用户影响：提供了更灵活的教师模型计算模式，可根据资源池配置选择 colocate 或 standalone 模式，可能提升训练效率和资源利用率。对系统影响：扩展了蒸馏功能的核心路径，需确保与现有 rollout 和训练逻辑兼容，避免性能退化或崩溃。对团队影响：增强了代码模块化，但需跟进 review 中识别的 bug 修复，并可能影响后续蒸馏相关开发。
- 风险标记：张量维度错误风险，边界条件处理不足，新标志增加配置复杂性

关联脉络

- PR #5723 [1/2][rollout,trainer] refactor: Teacher colocate mode -- Move teacher logprob computation to AsyncTeacherLLMServerManager: 这是本 PR 的第一部分，共同实现教师 colocate 模式的重构，移动教师 logprob 计算到专用管理器。