PR #1788 完整报告

THUDM/slime

[WIP] fix loss oom

合并时间：2026-04-04 23:41

原文链接：http://prhub.com.cn/THUDM/slime/pull/1788

执行摘要

- 一句话：修复损失计算内存溢出，优化 PPO 熵计算和 Megatron 损失路径。
- 推荐动作：建议工程师精读此 PR，特别是熵梯度控制设计和 checkpointing 优化，这些是内存优化中的常见技巧。同时关注 Copilot 指出的潜在正确性问题，以确保变更不影响训练稳定性。

功能与动机

PR body 中展示了优化前后的内存使用对比图像，表明在损失计算过程中存在内存峰值问题。动机是修复 OOM，确保训练过程更加稳定，特别是在使用大模型或复杂配置时。

实现拆解

主要改动在两个文件：1) slime/backends/megatron_utils/loss.py：重构了 _allgather_cp_redistribute 函数，引入 _build_shifted_tokens 函数以优化 token 构建逻辑，并调整 get_log_probs_and_entropy 函数一次性计算完整 logits 的 log-probs 和熵，减少重复计算。2) slime/utils/ppo_utils.py：修改 calculate_log_probs_and_entropy 函数，添加 need_entropy_grad 参数，当熵系数为零时使用 torch.no_grad() 避免梯度跟踪，降低内存开销；同时调整代码顺序和 checkpointing 为 use_reentrant=False。

关键文件：

- slime/backends/megatron_utils/loss.py（模块 Megatron 损失模块）：包含损失计算的核心逻辑重构，优化内存使用和 allgather 操作。
- slime/utils/ppo_utils.py（模块 PPO 工具模块）：修改 PPO log-probs 和熵计算函数，添加熵梯度控制以减少内存开销。

关键符号：_allgather_cp_redistribute, _build_shifted_tokens, get_log_probs_and_entropy, calculate_log_probs_and_entropy

评论区精华

Review 中 Copilot 指出三个关键问题：1) 温度缩放缺失，导致 rollout_temperature != 1.0 时 PPO 行为不一致；2) 熵梯度处理在 allgather_cp 路径下可能无效，因为 _allgather_cp_redistribute 使用可微分操作；3) allgather_cp 配置仅支持 thd 格式，但代码未做检查。Zhuzilin 询问代码移动和 logits.clone() 的必要性。讨论未显示明确结论，但 PR 已

获批准，可能问题被接受或后续处理。

- 温度缩放缺失 (correctness): 未明确解决，PR 已批准，可能问题被接受或忽略。
- 熵梯度处理可能无效 (performance): 未明确解决。
- allgather_cp 配置检查 (design): 未明确解决。
- 代码移动和 clone 必要性 (design): 未明确回答，但 PR 已批准，可能变更被接受。

风险与影响

- 风险：风险包括：1) 温度缩放缺失可能改变 PPO 损失计算，影响训练收敛；2) 熵梯度优化在 allgather_cp 启用时可能不生效，内存减少有限；3) 配置不一致（如 allgather_cp 与 qkv_format 不匹配）可能导致运行时错误；4) 核心路径变更引入回归风险，需测试验证正确性。
- 影响：对用户：减少训练时内存使用，降低 OOM 风险，提升大模型训练体验。对系统：修改核心损失计算路径，影响所有使用 Megatron 和 PPO 的训练任务；性能优化可能提升整体训练效率。影响范围中高，需在集成后监控内存和收敛行为。
- 风险标记：温度缩放缺失，熵梯度处理可能无效，配置不一致风险

关联脉络

- PR #1775 [Fix] Fix duplicate Megatron LR scheduler resume when optimizer state is not loaded: 同样涉及 Megatron 模块的性能 bugfix，主题相关。
- PR #1764 Add host memory metrics to available_memory function: 与内存监控和优化相关，主题相似。
- PR #1769 Support FP8 conversion for Qwen3.5: 性能优化相关，都涉及训练效率改进。