PR #5669 完整报告

verl-project/verl

[fsdp, perf, doc] fix: fix Liger integration for VL models and RL training, allowing liger speed improvement

合并时间：2026-03-23 13:22

原文链接：http://prhub.com.cn/verl-project/verl/pull/5669

执行摘要

本 PR 修复了 Liger 内核在 verl 中与视觉语言模型和强化学习训练的兼容性问题，通过显式传递参数禁用冲突的 fused_linear_cross_entropy 并启用 SwiGLU 融合。修复后，Liger 可安全用于所有模型类型，带来显著速度提升和内存优化，文档已相应更新，无需用户 API 变更。

功能与动机

此 PR 旨在解决 issue #2609 中报告的问题：当 use_liger=True 时，_apply_liger_kernel_to_instance 默认设置 fused_linear_cross_entropy=True，导致与 verl 的前向补丁冲突，使 VL 模型崩溃。动机是允许 Liger 内核在 VL 模型和 RL 训练中正常工作，以提升训练性能，同时关闭相关 issue。

实现拆解

实现主要包括以下三个改动点：

1. 代码逻辑修复：在 verl/workers/fsdp_workers.py 和 verl/workers/engine/fsdp/transformer_impl.py 中，修改 _apply_liger_kernel_to_instance 调用，添加参数 fused_linear_cross_entropy=False 和 swiglu=True。 - 例如，在 verl/workers/fsdp_workers.py 中：python _apply_liger_kernel_to_instance(model=actor_module, fused_linear_cross_entropy=False, swiglu=True,)
2. 测试覆盖：新增 tests/models/test_liger_vl_compat.py 文件，创建一个微型的 Qwen3-VL 模型，验证 Liger 启用时前向传递成功，确保修复不会引入回归。
3. 文档更新：修改 docs/perf/perf_tuning.rst，澄清 Liger 的使用场景、兼容性和配置说明，帮助用户正确启用优化。

评论区精华

review 中仅有简短的正面反馈：

- gemini-code-assist[bot] 评论："The pull request effectively addresses the compatibility issues with Liger integration for VL models and RL training..." 强调了修复的有效性和测试验证。
- wuxibin89 直接批准，无进一步讨论。 讨论焦点集中在正确性和测试覆盖上，无争议或未解决疑虑。

风险与影响

风险分析：技术风险较低，因为修复通过显式禁用 fused_linear_cross_entropy 避免冲突，且该优化在 verl 中为惰性（不传递 labels）。但修改涉及核心训练路径，需确保无回归影响其他模型类型或训练场景；测试覆盖了 VL 模型，但可能未覆盖所有边缘情况。 影响分析：对用户影响积极，现有 use_liger=True 配置现可正确用于 VL 模型和 RL 训练，无需 API 变更。系统性能提升显著，基准测试显示速度提升 20-50%，内存使用减少高达 28%。团队需推广此修复以优化训练效率。

关联脉络

此 PR 直接关联 issue #2609（修复目标）和 issue #1720（相关前向冲突问题），但未在历史 PR 中找到直接相关 PR。从上下文看，它属于 verl 性能优化和 FSDP 工作流的一部分，可能与近期涉及 perf 或 fsdp 标签的 PR（如 PR 5627 关于 NUMA 亲和性优化）有间接关联，但无强功能联系。