

PR #1945 完整报告

THUDM/slime

[docker] upgrade to sglang v0.5.12.post1

合并时间: 2026-05-26 19:04

原文链接: <http://prhub.com.cn/THUDM/slime/pull/1945>

执行摘要

- 一句话: 升级 sglang 至 v0.5.12.post1, 更新 Docker 与 CI 配置。
- 推荐动作: 建议关注 sglang patch 中关于 bootstrap 超时和内存占用释放的实现, 这是提升分布式推理稳定性的关键改动。gpu_lock_exec 的重写模式可在其他类似工具中复用。

功能与动机

升级 sglang 版本以获取新功能 (如 DeepseekV32ForCausalLM 支持、disaggregation 改进等), 同时保持与 slime 框架的兼容性。现有的 patch 基于旧版本开发, 需要针对 v0.5.12.post1 进行适配, 因此同步更新补丁。此外, 修复了测试过程中发现的 checkpoint 保存 / 加载路径硬编码问题, 并增强 GPU 锁脚本的健壮性。

实现拆解

1. 版本升级与补丁同步: 修改 docker/Dockerfile 和 docker/version.txt 将 sglang 版本指向 v0.5.12.post1。重新生成 docker/patch/latest/sglang.patch 和 docker/patch/latest/megatron.patch, 主要变更包括新增 Kwargs.aux_buffer_names 字段以支持辅助缓冲区名称传递, 为 DecodePreallocQueue 添加 release_memory_occupation/resume_memory_occupation 方法以支持内存占用释放与恢复, 以及添加 bootstrap 超时机制 (SGLANG_DISAGGREGATION_TRANSFER_TIMEOUT 环境变量控制)。
2. Checkpoint 测试参数化: 修改 tests/test_qwen3_4B_ckpt.py, 将原先硬编码的 checkpoint 目录改为根据 --save-optimizer 和 --load-optimizer 参数动态生成, 新增 default_checkpoint_dir 和 optimizer_args 函数。execute 函数新增 optimizer 和 checkpoint_dir 参数, 支持四种优化器放置组合 (save cpu/load cpu 等)。
3. GPU 锁执行器重写: 重写 tests/ci/gpu_lock_exec.py, 将原有的 os.execvp 调用替换为 subprocess.Popen 管理的子进程。新增信号处理器 terminate_child, 在收到 SIGINT/SIGTERM/SIGHUP 时先尝试优雅终止子进程, 超时后发送 SIGKILL, 确保锁文件被 finally 块释放。新增 _normalize_returncode 正确转换被信号终止的返回码。同时移除了 FdLock.open 中对 os.set_inheritable 的调用。
4. CI 工作流调整: 在 .github/workflows/pr-test.yml 及模板 .github/workflows/pr-test.yml.j2 中, 为 e2e-test-ckpt 和 e2e-test-train 工作流增加了针对四种优化器组合的测试条目, 并移除了原有的单一 test_args 空测试。所有 docker run 指令添加 --pull=always 参数。

关键文件：

- docker/patch/latest/sglang.patch（模块 SGLang 补丁；类别 test；类型 test-coverage；符号 `_get_logits_head_gate`, `_sync_fwd_experts_buffer_DtoH`, `capture`, `get_routed_experts`）：升级 `sglang` 版本后必须同步更新补丁，以适配新版 API 变化；本文件是 patch 核心，包含 `bootstrap` 超时、`aux_buffer_names` 等关键改动。
- tests/test_qwen3_4B_ckpt.py（模块 检查点测试；类别 test；类型 test-coverage；符号 `prepare`, `default_checkpoint_dir`, `execute`, `optimizer_args`）：重构 checkpoint 测试以支持优化器放置参数化，覆盖四种组合，提升测试质量。
- tests/ci/gpu_lock_exec.py（模块 GPU 锁；类别 test；类型 test-coverage；符号 `_os_execvp`, `_run_command`, `terminate_child`, `_normalize_returncode`）：重写 GPU 锁执行器，将 `execvp` 替换为 `subprocess`，添加信号处理以确保锁释放。
- docker/patch/latest/megatron.patch（模块 Megatron 补丁；类别 test；类型 test-coverage）：同步 `megatron` 补丁以适配新版本，增加 `disable_grad_buffers_cpu_backup` 参数和 checkpoint 加载容错。
- .github/workflows/pr-test.yml（模块 CI 配置；类别 infra；类型 infrastructure）：CI 配置变更：添加 `--pull=always`，扩展 ckpt 测试矩阵涵盖四种优化器组合。
- .github/workflows/pr-test.yml.j2（模块 CI 模板；类别 infra；类型 infrastructure）：CI 模板同步更新，增加 ckpt 测试条目和 `--pull=always`。
- docker/Dockerfile（模块 Docker 构建；类别 infra；类型 infrastructure）：指定 `sglang` 新版本镜像标签。
- docker/version.txt（模块 版本记录；类别 docs；类型 documentation）：记录更新后的版本号。

关键符号：`_get_logits_head_gate`, `_sync_fwd_experts_buffer_DtoH`, `capture`, `get_routed_experts`, `SenderWrapper`, `init`, `init_weights_send_group_for_remote_instance`, `_unload_lora_adapter_locked`, `prepare`, `default_checkpoint_dir`, `execute`, `optimizer_args`, `_os_execvp`, `_run_command`, `terminate_child`, `_normalize_returncode`

评论区精华

本 PR 为单作者自主合入，无公开讨论记录。

- 暂无高价值评论线程

风险与影响

- 风险：
 1. `sglang` 版本升级可能引入与现有模型（如 DeepseekV3、GlmMoe）的兼容性问题，需通过 CI 验证。
 2. patch 中新增的 `bootstrap` 超时机制依赖于环境变量，若未设置默认 600 秒，可能过长或过短，需要用户自行调优。
 3. `gpu_lock_exec` 从 `execvp` 改为 `subprocess` 后，进程组管理变化，若存在遗留的锁文件可能干扰。

4. checkpoint 测试矩阵扩展增加了 CI 执行时间，但降低了遗漏风险。 - 影响：对用户：Docker 镜像将基于新版 sglang，所有依赖 sglang 的 rollout 将受影响；但补丁向后兼容性良好。对系统：GPU 锁脚本更健壮，不再因子进程被杀死而泄漏锁。对团队：CI 覆盖更全面的 checkpoint 场景，便于早期发现优化器放置相关 bug。 - 风险标记：底层版本升级，补丁兼容性，测试覆盖扩展，子进程管理模式变化

关联脉络

- PR #1946 Add backward compatibility to delta weight updation: 本 PR 升级 sglang 版本后需同步更新 delta weight 相关补丁，保持兼容性。
- PR #1806 feat: delta weight sync (disk + nccl transports): 该 PR 引入了 delta weight 同步功能，依赖特定的 sglang 补丁，本 PR 的升级需确保补丁一致。
- PR #1924 Reduce host memory with upgraded tms: 同为 Docker 镜像升级 PR，涉及 torch_memory_saver 版本管理，与本 PR 共同推进基础设施更新。