

PR #21655 完整报告

sgl-project/sglang

[Bug][VLM] Fix shared memory race condition in ShmPointerMMData broadcast for multi-GPU VLM serving

合并时间: 2026-04-01 16:15

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/21655>

执行摘要

本 PR 修复了 sglang 仓库中一个关键 bug: 在多 GPU VLM (视觉语言模型) 服务中, 当 `tp_size > 1` 且使用图像输入时, 共享内存广播的竞态条件导致 `FileNotFoundError` 崩溃。通过引入条件 barrier 和共享内存特征检测, 确保所有 GPU rank 在卸载共享内存前完成同步, 解决了由 PR #21465 引入的回归问题。变更影响范围仅限于特定 VLM 配置, 但对系统稳定性至关重要, 推荐团队关注此分布式同步设计。

功能与动机

为什么需要这个修复? 在 PR #21465 之后, 运行 VLM 模型 (如 Kimi-K2.5) 在 `tp_size > 1` 和多模态输入时崩溃, 错误如下:

```
FileNotFoundError: [Errno 2] No such file or directory: '/psm_...'
```

错误发生在 `ShmPointerMMData.__setstate__` 中, 根源于 PR #21465 移动了 `unwrap_shm_features()` 函数位置: 源 rank 在广播后立即取消链接共享内存段 (通过 `shm_unlink`), 而非源 rank 仍在尝试打开它 (通过 `shm_open`), 导致竞态条件。此 bug 仅影响非 DP-attention 路径的 VLM 服务, 纯文本输入不受影响。

实现拆解

实现分为两个关键文件:

文件路径	关键变更	目的
<code>python/sglang/srt/managers/mm_utils.py</code>	新增 <code>has_shm_features</code> 函数	检测请求列表是否包含 <code>ShmPointerMMData</code> 共享内存特征, 代码如下:

文件路径	关键变更	目的
<pre>def has_shm_features(recv_reqs): """Return True if any request in the list contains ShmPointerMMData.""" for req in recv_reqs: if hasattr(req, "batch"): if has_shm_features(req.batch): return True elif hasattr(req, "mm_inputs") and req.mm_inputs: for item in req.mm_inputs.get("mm_items", []): if isinstance(item.feature, ShmPointerMMData): return True return False</pre>		
python/sglang/srt/managers/scheduler.py	修改 <code>recv_requests</code> 函数	添加条件 <code>barrier</code> ，逻辑如下：
<pre>if recv_reqs: if (not self.server_args.enable_dp_attention and self.tp_size > 1 and self.model_config.is_multimodal and has_shm_features(recv_reqs)): barrier(group=self.tp_cpu_group) for req in recv_reqs: unwrap_shm_features(req)</pre>		
<p><code>barrier</code> 仅在非 DP-attention、<code>tp_size > 1</code>、多模态配置且有共享内存特征时触发，其他场景（如单 GPU 或 DP-attention 启用）跳过以优化性能。</p>		

评论区精华

review 讨论中最有价值的交锋围绕 `barrier` 的粒度和测试覆盖展开：

- `gemini-code-assist[bot]` 强调测试重要性：> "To prevent regressions of this race condition, it would be beneficial to add a test case that specifically covers the failing scenario: a VLM model with `tp_size > 1` and `enable_dp_attention=False`." 此建议指出未来需添加回归测试以巩固修复。
- `yuan-luo` 关注性能影响：> "Barrier is unconditional per-batch, adding latency to every scheduling round... A more targeted approach could check whether any request in `recv_reqs` actually contains `ShmPointerMMData`..." 此评论推动了 PR 优化，通过添加 `has_shm_features` 函数使 `barrier` 更精确，避免了不必要延迟。

最终，代码采纳了优化方案，barrier 仅在检测到共享内存特征时才执行，平衡了正确性和性能。

风险与影响

技术风险：

1. 条件逻辑复杂：barrier 触发依赖于四个条件（enable_dp_attention、tp_size、is_multimodal、has_shm_features），如果未来配置变更或检测逻辑不完善，可能导致竞态再现或误加延迟。
2. 性能轻微开销：尽管 barrier 已优化，但在高负载 VLM 场景中仍可能引入额外网络同步时间，需监控实际影响。
3. 缺乏测试覆盖：当前 PR 未添加针对此竞态的回归测试，未来代码变更可能意外破坏修复，建议团队后续补全。

影响评估：

- 用户：直接解决多 GPU VLM 模型的崩溃问题，提升服务可靠性，特别是对于图像输入的高并发场景。
- 系统：变更仅限于调度路径，在非 DP-attention 的 VLM 配置下添加同步，整体系统性能影响有限，但避免了关键故障。
- 团队：此修复突显了分布式共享内存处理的脆弱性，建议工程师在类似场景中谨慎设计同步机制。

关联脉络

本 PR 是 sglang 仓库 VLM 功能演进的一部分，直接关联以下 PR：

- PR #21465：作为 bug 根源，其移动 unwrap_shm_features 函数引发了竞态条件，本 PR 是修复该回归的关键补丁。
- 其他 VLM 相关 PR（如 #21671、#21458）：展示了团队对多模态模型的持续优化，本 bugfix 确保这些优化在分布式环境下稳定运行。更大的趋势是：随着 VLM 模型复杂度增加，共享内存和分布式同步成为关键挑战，本修复为后续性能优化（如内核融合）奠定了基础，避免竞态干扰。建议未来在 CI 中增加 tp_size > 1 的 VLM 测试，以增强 robustness。