PR #24694 完整报告

sgl-project/sglang

Fix NCCL deadlock in Ulysses SP when sequence length has remainder

合并时间：2026-05-09 11:05

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/24694

执行摘要

- 一句话：修复 Ulysses SP 下 NCCL 死锁
- 推荐动作：此 PR 修复了一个关键的并发死锁问题，变更简洁且经过讨论，值得合并。建议在合并后执行一次包含 LTX2 SP 模式的 CI 测试以确认无回归。

功能与动机

修复 LTX2 视频生成集成过程中遇到的 NCCL 超时问题。PR body 指出：当序列长度不能被 sp_world_size 整除时，_build_ltx2_sp_padding_mask 可能返回 None，导致不同 rank 执行不同的 NCCL 集合操作，造成死锁。

实现拆解

1. 修改 _build_ltx2_sp_padding_mask（ltx_2_denoising.py）：移除 valid <= 0 or valid >= seq_len 时返回 None 的早期分支。改为始终返回一个 torch.bool 类型的 mask 张量：全 True 表示无 padding，否则在无效位置设置 False。这确保了所有 rank 生成的 mask 类型一致。
2. 优化 mask 数据类型：将 mask 从 torch.float32 改为 torch.bool，因为 SDPA 的 additive mask 需要 bool 类型，_prepare_sdpa_mask 会将 True 转换为 0.0，False 转换为负无穷，而 float32 的 1.0/0.0 是语义错误的。
3. 添加警告注释（layer.py）：在 USPAttention.forward 中 sequence_model_parallel_all_gather 调用前插入注释，提示若在此处发生 NCCL 超时 / 死锁，应检查 attn_mask 是否跨 rank 不一致。

关键文件：

- python/sglang/multimodal_gen/runtime/pipelines_core/stages/ltx_2_denoising.py（模块 扩散管线；类别 source；类型 core-logic；符号 _build_ltx2_sp_padding_mask）：修复的核心文件：修改了 _build_ltx2_sp_padding_mask 方法，移除返回 None 的分支，改为始终返回 bool mask，解决了死锁根因。
- python/sglang/multimodal_gen/runtime/layers/attention/layer.py（模块 注意力层；类别 source；类型 documentation）：添加了警告注释，帮助未来调试 NCCL 死锁问题，指示排查 mask 跨 rank 不一致。

关键符号：_build_ltx2_sp_padding_mask

关键源码片段

python/sglang/multimodal_gen/runtime/pipelines_core/stages/ltx_2_denoising.py

修复的核心文件：修改了 _build_ltx2_sp_padding_mask 方法，移除返回 None 的分支，改为始终返回 bool mask，解决了死锁根因。

```
@staticmethod
def _build_ltx2_sp_padding_mask(
    batch: Req,
    *,
    seq_len: int,
    batch_size: int,
    key: str,
    device: torch.device,
) -> torch.Tensor | None:
    valid = getattr(batch, key, None)
    if valid is None:
        return None
    valid = int(valid)
    # 始终返回 bool 类型 mask，保证跨 rank 一致
    # 使用 bool 而非 float32，因为 _prepare_sdpa_mask 能正确处理 bool：
    # True -> 0.0，False -> -inf
    mask = torch.ones((batch_size, int(seq_len)), device=device, dtype=torch.bool)
    if valid < int(seq_len):
        # 只有存在 padding 时才将尾部设为 False
        mask[:, max(0, valid):] = False
    return mask
```

python/sglang/multimodal_gen/runtime/layers/attention/layer.py

添加了警告注释，帮助未来调试 NCCL 死锁问题，指示排查 mask 跨 rank 不一致。

```
        # If NCCL timeout/deadlock occurs here, check whether
        # attn_mask is inconsistent across SP ranks (None on some, Tensor on
        # others), which causes all_gather participant mismatch. Upstream
        # mask builders must ensure all ranks produce the same mask type.
        gathered_mask = sequence_model_parallel_all_gather(
            attn_mask.contiguous(), dim=1
        )
```

评论区精华

Review 中 gemini-code-assist[bot] 指出：当 valid <= 0 时仍返回 None，在 SP 环境下如果某些 rank 在其本地分片中无有效 token，仍可能触发死锁。作者 storyicon 采纳了建议，将 valid <= 0 的情况也改为返回 bool mask，并同时将数据类型从 float32 改为 bool，以兼容 SDPA 的正确语义。

- valid <= 0 时仍可能死锁 (correctness): 作者采纳建议，移除 valid <= 0 时返回 None 的分支，改为返回 bool mask。同时将 dtype 从 float32 改为 bool 以兼容 SDPA。

风险与影响

- 风险：变更范围小，仅涉及两个文件共 7 行增删。核心风险在于：全 True mask 在语义上等同于 None（不掩盖任何 token），理论上不影响模型输出。但需确认下游 USPAttention.forward 对 bool mask 的处理与 float32 mask 一致，尤其在 _prepare_sdpa_mask 中的转换逻辑是否符合预期。
- 影响：影响范围限于 LTX2 视频生成管线中使用 Ulysses SP 模式的场景。修复后，序列长度不能被 sp_world_size 整除时不再死锁，提升了系统的鲁棒性。对非 SP 模式或无 padding 的情况无影响。
- 风险标记：核心路径变更，并发 / 死锁风险

关联脉络

- 暂无明显关联 PR