

PR #25044 完整报告

sgl-project/sglang

Revert "Migrate Intel CPU cases to the test/registered."

合并时间: 2026-05-12 13:32

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/25044>

PR 分析报告 : Revert "Migrate Intel CPU cases to the test/registered."

执行摘要

此 PR 回滚了 #22670，目的是修复该 PR 意外修改 `docs_new/index.mdx` 的问题。回滚后删除了所有新增的 Intel CPU 测试文件（24 个文件），恢复了文档首页。虽然短期内降低了 CPU 测试覆盖率，但避免了文档错误，为后续以正确方式重新引入测试提供了清理基础。

功能与动机

PR author mingfeima 在评论中指出：

```
revert #22670 which unintentionally changed docs_new/index.mdx.
```

因此动机是消除意外变更，保证文档正确。

实现拆解

本 PR 仅包含一个 `git revert` commit，直接反向应用 #22670 的更改：

- 测试文件删除：`test/registered/cpu/` 下所有在 #22670 中新增的文件被整体移除。涉及的文件包括：
 - QKV projection with RoPE (`test_qkv_proj_with_rope.py`)
 - 归一化 (`test_norm.py`)
 - Mamba attention (`test_mamba.py`)
 - MoE (`test_moe.py`)
 - GEMM (`test_gemm.py`)
 - 因果卷积 (`test_causal_conv1d.py`)
 - RoPE (`test_rope.py`)
 - Flash attention (`test_flash_attn.py`)
 - 共享专家 (`test_shared_expert.py`)
 - 扩展注意力 (`test_extend.py`)
 - TopK (`test_topk.py`)
 - 及其他文件和工具模块 `utils.py`
- 文档恢复：`docs_new/index.mdx` 被恢复到 #22670 合并前的状态（+35 行变更）。

3. 无其他代码改动：没有引入新功能或修复其他问题。

本次 PR 不涉及新增或修改的逻辑代码，因此没有需要展示的源码片段。所有变更为文件删除和文档回退。

评论区精华

- mingfeima: 说明 revert 的原因——修复 docs_new/index.mdx 的意外修改。
- mintlify[bot]: 报告文档预览部署失败（可能与恢复过程中的临时状态有关，回滚后应已解决）。

风险与影响

- 风险：CPU 测试覆盖暂时降低，可能遗漏相关算子的回归问题。但原 PR 引入测试的方式不完善（附带文档误改），回滚是安全的选择。
- 影响：Intel CPU 的 CI 阶段不再运行这些特定测试，团队需计划后续以正确方式重新引入。用户无感知。

关联脉络

- #22670: 被回滚的原 PR，尝试将 CPU 测试迁移到 test/registered/ 下的统一框架，但因误改文档被回退。
 - 近期多个 PR（如 #25029、#25030）在清理投机解码和 Mamba 相关代码，与本 PR 无直接关联，但表明项目正在积极重构和优化。
 - 团队需要跟踪后续是否有新 PR 重新引入 CPU 测试（确保不修改 index.mdx）。