

PR #21562 完整报告

sgl-project/sglang

[CI] Relax several thresholds in flaky CIs

合并时间: 2026-03-28 04:16

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/21562>

执行摘要

此 PR 通过放宽多个测试文件的阈值（如 epsilon、问题数量、精度），减少 CI 不稳定性，但可能降低测试严格性。适合快速浏览以了解 CI 调整。

功能与动机

动机是解决 CI 测试不稳定（flaky）问题，通过放宽阈值来减少失败率。PR 标题直接表明 "[CI] Relax several thresholds in flaky CIs"。

实现拆解

修改涉及三个测试文件：

- test/registered/lora/test_lora_moe_vllm_sgl_logprob_diff.py: 将 epsilon 值从 $1e-4$ 增加到 $2e-4$ ，允许更大 logprob 差异。
- test/registered/quant/test_deepseek_v32_fp4_4gpu.py:
 - 将 num_questions 和 parallel 从 1319 减少到 500，降低测试负载。
 - 精度阈值从 0.935 降低到 0.93。
- test/registered/quant/test_deepseek_v32_fp4_mtp_4gpu.py: 精度阈值从 0.94 降低到 0.93。

评论区精华

review 评论来自 gemini-code-assist[bot]，建议改进代码结构：

"This epsilon value is hardcoded. To improve readability and make it easier to adjust in the future, consider defining it as a constant at the module level." "The test_a_gsm8k method is duplicated... consider refactoring the common logic into a helper function or a mixin." 未引发人类讨论，建议未采纳。

风险与影响

风险：精度阈值降低可能掩盖模型性能退化；问题数量减少可能影响测试覆盖率；重复代码增加维护成本。影响：对 CI 系统，提高构建成功率但降低测试质量；对团队，简化流程但需监控测试有效性。

关联脉络

与本 PR 相关的历史 PR 包括：

- 21485：删除冗余 DeepSeek V3 FP4 测试，同为优化 CI 负担。
- 21516：调整测试超时，类似 CI 稳定性改进。
- 21547：注册测试文件，涉及测试基础设施调整。这些 PR 共同反映团队在持续优化 CI 流程以减少不稳定因素。