

PR #34103 完整报告

sgl-project/sglang

[NPU] Increase the retry count to 3 for the GSM8K.

合并时间: 2026-08-10 10:55

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/34103>

执行摘要

- 一句话: NPU GSM8K 精度测试重试增至 3 次并放宽超时
- 推荐动作: 该 PR 值得快速浏览, 不建议精读——它只涉及测试工具与 CI 配置, 无运行时逻辑。值得关注的设计点是: 用全局 `MAX_RETRY_COUNT` 取代各测试基类的 `max_attempts` 属性, 是 "单一事实来源" 的合理收敛; 但 `DATASET_FLUCTUATION` 与 `DATASET_QUESTION_COUNTS` 不同步登记 `gsm8k` 的问题需要后续跟进, 否则新增配置形同虚设。对负责 NPU/CI 稳定性的同学, 建议顺带检查 `get_max_retries` 的完整实现, 确认 GSM8K 重试确实达到 3。

功能与动机

PR 标题与最终提交信息明确表达了动机: "Increase the retry count to 3 for the GSM8K." 和 "Extend the timeout duration for the single model test case script."。NPU 上的 GSM8K 精度测试经常因偶发性能波动或环境抖动在首次 / 第二次运行中达不到阈值而误报失败, 因此希望通过提高重试次数、放宽容差并延长超时来降低假阳性率, 稳定 NPU CI。PR body 本身只有模板内容, 没有补充额外背景。

实现拆解

本次变更共 2 个文件、+2/-4 行, 分三层拆解:

1. 测试工具层: 登记 GSM8K 波动容差 - 文件: `python/sglang/test/ascend/e2e/test_npu_accuracy_utils.py` - 在 `DATASET_FLUCTUATION` 中新增 `"gsm8k": 3`, 意图是为 GSM8K 提供 3 题的绝对波动空间。 - 注意: `DATASET_QUESTION_COUNTS` 并未同步登记 `gsm8k`, 而 `get_accuracy_threshold` 要求两个字典同时命中才走绝对容差分支, 因此该配置在当前实现下实际不会生效, GSM8K 仍走 `baseline_accuracy * ACCURACY_TOLERANCE` 的百分比容差路径——这是一个值得确认的配置不一致点。
2. 重试机制统一到 `MAX_RETRY_COUNT` - 删除 `TestNpuAccuracyTestCaseBase`、`TestNpuAccuracyMultiNodePdMixTestCaseBase`、`TestNpuAccuracyMultiNodePdSepTestCaseBase` 三个测试基类上的 `max_attempts = 2` 类属性。 - 此前重试次数由类属性硬编码为 2, 删除后重试逻辑统一由文件级常量 `MAX_RETRY_COUNT = 3` (经 `get_max_retries` 查询) 控制, 使 GSM8K 等数据集的实际重试次数从 2 提升到 3, 消除类属性与全局常量并存的分裂状态。

3. CI 层：延长单文件执行超时 - 文件：.github/workflows/_npu-single-node-test-stage.yml
- 将 run_suite.py --timeout-per-file 由 3600 秒调整为 7200 秒。 - 原因：3 次重试叠加服务器启动 (SERVER_INITIALIZATION_DELAY = 120 秒、DEFAULT_TIMEOUT_FOR_SERVER_LAUNCH = 3600 秒) 后，单用例执行时长可能超过 1 小时，原超时阈值容易误杀仍在重试中的测试。
4. 配套情况 - 没有新增正式单元测试或文档；该 PR 的 "测试" 本身就是 CI 精度测试的稳定性调整。 - CI 状态：PR Test (Base) 与 PR Test (Extra) 最新运行均为 🟢，说明合入前并未跑绿；最终由 sglang-npu-bot 以 /tag-and-rerun-ci 触发后自动合并。

关键文件：

- python/sglang/test/ascend/e2e/test_npu_accuracy_utils.py (模块 精度测试；类别 test；类型 test-coverage；符号 get_accuracy_threshold, TestNpuAccuracyTestCaseBase, TestNpuAccuracyMultiNodePdMixTestCaseBase, TestNpuAccuracyMultiNodePdSepTestCaseBase)：NPU 精度测试的核心工具文件：登记 GSM8K 波动容差、删除三个测试基类的 max_attempts=2，使重试次数统一由 MAX_RETRY_COUNT=3 控制，是本 PR 行为变更的主要载体。
- .github/workflows/_npu-single-node-test-stage.yml (模块 CI 编排；类别 infra；类型 infrastructure)：NPU 单节点测试的 CI 编排文件：将 run_suite.py 的 --timeout-per-file 从 3600 秒提升到 7200 秒，为 3 次重试提供时间预算，是本次变更的配套基础设施调整。

关键符号：get_accuracy_threshold, get_max_retries

关键源码片段

python/sglang/test/ascend/e2e/test_npu_accuracy_utils.py

NPU 精度测试的核心工具文件：登记 GSM8K 波动容差、删除三个测试基类的 max_attempts=2，使重试次数统一由 MAX_RETRY_COUNT=3 控制，是本 PR 行为变更的主要载体。

```
# 各数据集的题目总数与允许的绝对波动（单位：题数）
# 注意：gsm8k 只登记了 DATASET_FLUCTUATION，未同步登记 DATASET_QUESTION_COUNTS,
# 因此 get_accuracy_threshold 的绝对容差分支对 gsm8k 实际不会命中
dataset_question_counts = {
    "aime25": 30,
    "aime26": 30,
    "gpqa_diamond": 198,
}

dataset_fluctuation = {
    "aime25": 2,
    "aime26": 2,
    "gpqa_diamond": 5,
    "gsm8k": 3, # 本 PR 新增：为 GSM8K 预留 3 题波动
}
```

max_retry_count = 3 # 统一重试上限，删除各基类 max_attempts=2 后即以此为准

```
def get_accuracy_threshold(datasets, baseline_accuracy):
    """基于数据集波动容差计算精度阈值。

    对同时登记了 QUESTION_COUNTS 与 FLUCTUATION 的数据集 (aime*、gpqa_diamond)
    按题目数折算相对容差；其余数据集 (如 gsm8k、mmmu) 走百分比容差 (0.99) 。
    """
    dataset = datasets[0] if datasets else None
    if dataset in dataset_fluctuation and dataset in dataset_question_counts:
        fluctuation = dataset_fluctuation[dataset] / dataset_question_counts[dataset]
        return baseline_accuracy - fluctuation
    return baseline_accuracy * ACCURACY_TOLERANCE
```

评论区精华

本 PR 没有任何人工 review 评论或审核记录。唯一的动态是 [sglang-npu-bot](#) 机器人发出 [/tag-and-rerun-ci](#) 指令触发 CI 重跑。也就是说，这是一次由机器人自动合并的 CI 稳定性微调，缺少人工对 " 重试掩盖回归 " 风险的讨论。

- 暂无高价值评论线程

风险与影响

- 风险：
 1. 重试可能掩盖真实精度回归：GSM8K 由单次 2 次尝试改为 3 次尝试取最优结果，若模型精度确实退化，只要 3 次中偶有一次达标就可能被放行，降低测试的回归发现能力。这是此类 " 提高重试次数 " 改动最常见的隐患。
 2. `DATASET_FLUCTUATION["gsm8k"]` 配置可能不生效：`get_accuracy_threshold` 要求数据集同时出现在 `DATASET_FLUCTUATION` 与 `DATASET_QUESTION_COUNTS` 中才使用绝对容差，而 `DATASET_QUESTION_COUNTS` 未加入 gsm8k。若该配置目的是放宽 GSM8K 阈值，当前实现没有达到预期效果；若只是遗留声明，则属于无效配置，建议清理或补全。
 3. CI 资源占用翻倍：`--timeout-per-file` 从 3600 秒升到 7200 秒，意味着 NPU 单节点测试的 worst-case 时长翻倍，排队和占用成本上升，可能影响其他 NPU CI 任务。
 4. 无人工 review 即合并：PR 由 bot 自动合入，`MAX_RETRY_COUNT` 与类属性删除之间的行为变化缺乏人工确认。
- 影响：影响范围集中在 NPU 硬件后端的 CI 精度测试：
 - 对 GSM8K 用例：实际重试次数从 2 提升到 3，可显著降低偶发失败导致的 PR 阻塞，但可能弱化精度回归的检出灵敏度。
 - 对 NPU 单节点测试流水线：单文件执行超时上限从 1 小时翻倍至 2 小时，CI 任务的最长运行时间与资源占用相应增加。
 - 对用户与推理服务：无影响，完全不涉及 `sglang/srt` 运行时代码。

- 对团队：维护 NPU CI 的同学需要关注重试策略与阈值配置的一致性，避免 " 测试易绿但回归漏检 "。
- 风险标记：重试可能掩盖精度回归，gsm8k 容差配置疑似未生效，CI 超时翻倍增加资源占用，缺少人工 review 自动合并

关联脉络

- PR #34147 [AMD] [CI] Register the DeepSeek-V4-Pro-DSPark MI35x nightly job so its suite actually runs: 同为硬件加速器（AMD/NPU）CI 夜测任务注册与修复，方向一致，都是让硬件 CI 套件真正稳定运行。
- PR #34186 [CI] Key scheduled CUDA suites by runner_config instead of hand-written jobs: CI 套件调度正从手写 job 迁移到 run_suite.py 注册式驱动，本 PR 的超时调整恰好作用于该执行链，两者共同演进 CI 套件执行模型。