

# PR #23287 完整报告

sgl-project/sglang

[CI][LoRA] Drop flaky all-None batch from multi-LoRA parity test

合并时间: 2026-04-21 05:43

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/23287>

## 执行摘要

- 一句话: 移除多 LoRA 测试中不稳定的全 None 批次, 修复 CI 随机失败。
- 推荐动作: 该 PR 变更简单直接, 适合快速浏览以理解 CI 测试优化策略。重点关注测试用例设计的合理性: 移除不提供特定功能信号的测试, 避免 CI 脆弱性。

## 功能与动机

PR body 指出, 测试 `test_ci_lora_models_multi_batch` 在 CI 中因全 None 批次 (`lora_paths=[None, None, None]`) 的 ROUGE-L 分数漂移而随机失败。该批次仅验证基础模型生成对等性, 并非 LoRA 行为测试, 且已有 `test_generation_models.py` 覆盖相同功能。移除该批次可消除 CI 不稳定因素, 避免因 CUDA/ 驱动版本差异导致的脆弱性。

## 实现拆解

1. 移除全 None 批次生成逻辑: 在 `python/sglang/test/lora_utils.py` 的 `create_multiple_batch_test_samples` 函数中, 删除原本在 `_use_aiter` 标志控制下添加的 `[None, None, None]` 测试用例。
2. 清理无用导入: 移除不再使用的 `get_bool_env_var` 和 `is_hip` 导入。
3. 保留核心测试覆盖: 函数仍保留两个 LoRA 混合批次测试用例 (如 `[None, lora_paths[0], lora_paths[1]]` 和 `[lora_paths[0], lora_paths[1], None]`), 确保 LoRA 多批次功能测试完整性。
4. 无其他配套改动: 未修改测试配置、部署或文档, 仅调整测试数据生成逻辑。

关键文件:

- `python/sglang/test/lora_utils.py` (模块 LoRA 测试; 类别 test; 类型 test-coverage; 符号 `create_multiple_batch_test_samples`): 唯一变更文件, 移除了导致 CI 不稳定的测试用例生成逻辑。

关键符号: `create_multiple_batch_test_samples`

## 关键源码片段

`python/sglang/test/lora_utils.py`

唯一变更文件, 移除了导致 CI 不稳定的测试用例生成逻辑。

```
def create_multiple_batch_test_samples(
```

```

prompts: List[str], lora_adapter_paths: List[str]
):
    random.seed(42)
    # 移除了原本的导入: from sglang.srt.utils.common import get_bool_env_var, is_hip
    # 移除了 _use_aiter 标志及相关逻辑

    test_cases = [
        ([
            random.choice(prompts),
            random.choice(prompts),
            random.choice(prompts),
        ], [
            None,
            lora_adapter_paths[0],
            lora_adapter_paths[1],
        ]),
        # 保留其他 LoRA 混合批次用例, 例如:
        ([
            random.choice(prompts),
            random.choice(prompts),
            random.choice(prompts),
        ], [lora_adapter_paths[0], lora_adapter_paths[1], None]),
    ]
    # 移除了原本在此处添加的 [None, None, None] 批次
    # 理由: 该批次仅测试基础模型生成对等性, 非 LoRA 行为, 且已有其他测试覆盖
    return test_cases

```

## 评论区精华

Review 中无实质性讨论, 仅由 Fridge003 批准。PR body 已详细阐述移除理由: 全 None 批次不提供 LoRA 特定信号, 且基础模型对等性已有其他测试覆盖。

- 暂无高价值评论线程

## 风险与影响

- 风险: 低风险:
  - 回归风险: 移除的是非 LoRA 相关测试用例, 不影响 LoRA 功能正确性验证。
  - 测试覆盖缺口: 基础模型生成对等性测试仍由 `test_generation_models.py` 覆盖, 无实质性缺口。
  - 兼容性: 仅删除测试数据, 不影响运行时或 API。
  - 影响: 对 CI 系统: 减少因随机分数漂移导致的测试失败, 提升 CI 稳定性。对开发团队: 简化测试维护, 避免因环境差异 (如 CUDA 版本) 导致的脆弱断言。对用户: 无直接影响, LoRA 功能行为不变。
- 风险标记: 测试覆盖调整

## 关联脉络

- 暂无明显关联 PR