

PR #31056 完整报告

sgl-project/sglang

Fix MockDSV4ModelRunner missing spec_algorithm

合并时间: 2026-07-14 04:24

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/31056>

执行摘要

- 一句话: 修复 MockDSV4ModelRunner 缺少 spec_algorithm 属性
- 推荐动作: 值得精读, 以理解 mock 对象维护的重要性。建议同步检查其他 mock runner 是否也需要类似修复。

功能与动机

DSPark (#30261) 切换了 DSV4 注意力后端初始化时读取 spec_algorithm 的源, 导致所有 dsv4 注意力单元测试 (如 test/registered/attention/unittests/dsv4/test_deepseek_v4.py) 在构造后端时 AttributeError。该 PR 旨在修复 CI 阻塞。

实现拆解

1. 在 python/sglang/test/kits/attention_unittest/attention_methods/dsv4_attention.py 中添加 from sglang.srt.speculative.spec_info import SpeculativeAlgorithm。
2. 在 MockDSV4ModelRunner.__init__ 中设置 self.spec_algorithm = SpeculativeAlgorithm.NONE, 匹配 DSPark 变更前的行为 (server_args.speculative_algorithm is None 等价于 NONE)。

关键文件:

- python/sglang/test/kits/attention_unittest/attention_methods/dsv4_attention.py (模块 DSV4 注意力; 类别 test; 类型 test-coverage): 唯一的变更文件: 为 MockDSV4ModelRunner 添加 spec_algorithm 属性及对应 import, 修复单元测试崩溃。

关键符号: MockDSV4ModelRunner.init

关键源码片段

`python/sglang/test/kits/attention_unittest/attention_methods/dsv4_attention.py`

唯一的变更文件: 为 MockDSV4ModelRunner 添加 spec_algorithm 属性及对应 import, 修复单元测试崩溃。

```
# 文件: python/sglang/test/kits/attention_unittest/attention_methods/dsv4_attention.py
```

```
# 在文件头部新增 import
```

```
from sglang.srt.speculative.spec_info import SpeculativeAlgorithm

class MockDSV4ModelRunner:
    def __init__(self, ...):
        # ... 其他属性 ...
        self.is_draft_worker = False
        self.spec_algorithm = SpeculativeAlgorithm.NONE # 新增：默认为非推测模式
        self._kernel_warmed_up = True
```

评论区精华

PR 无 review 评论，变更直接合并。

- 暂无高价值评论线程

风险与影响

- 风险：低风险。仅影响测试 mock 对象，不涉及生产代码。spec_algorithm 设为 NONE 与 DSpark 变更前行为等价，不会引入回归。
- 影响：修复了所有 dsv4 注意力单元测试的 CI 崩溃，确保后续 PR 可通过这些测试。
- 风险标记：测试基础设施修复

关联脉络

- PR #30261 DSpark: 本 PR 修复 DSpark 引入的回归，DSpark 切换了 spec_algorithm 的读取源。