

PR #42388 完整报告

vllm-project/vllm

[UT][XPU] fix test_parallel_sampling due to global random state

合并时间: 2026-05-12 18:03

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/42388>

执行摘要

- 一句话: 修复 XPU 测试因全局 random 状态导致的 flaky 问题
- 推荐动作: 此类测试可重复性 bugfix 值得精读, 作为团队编写确定性测试的范例。

功能与动机

PR body 明确指出: `_get_test_sampling_params` 使用全局 random 模块, 当与其他测试文件在同一 pytest session 中运行时, 全局随机状态被污染, 导致生成的 n 值偏大, 增加重复 completion 的概率。问题定位清晰。

实现拆解

1. 定位问题: 在 `tests/v1/engine/test_llm_engine.py` 的 `_get_test_sampling_params` 函数中, `get_mostly_n_gt1` 内部通过 `random.randint(0, 28)` 获取随机数, 依赖全局 random 状态。
2. 修复方案: 在函数体开头创建局部 RNG: `rng = random.Random(seed)`, 并将 `get_mostly_n_gt1` 中的 `random.randint` 替换为 `rng.randint`。
3. 效果: 即使同一个 pytest session 中其他测试修改了全局 random 状态, `_get_test_sampling_params` 依然基于固定 seed 生成可预测的 `n_list`, 消除 flaky 问题。
4. 仅测试改动: 无产品代码变更。

关键文件:

- `tests/v1/engine/test_llm_engine.py` (模块引擎; 类别 test; 类型 test-coverage; 符号 `_get_test_sampling_params`, `get_mostly_n_gt1`): 唯一变更文件, 修复了测试辅助函数中全局 random 状态污染导致的 flaky 问题。

关键符号: `_get_test_sampling_params`

关键源码片段

`tests/v1/engine/test_llm_engine.py`

唯一变更文件, 修复了测试辅助函数中全局 random 状态污染导致的 flaky 问题。

```
# tests/v1/engine/test_llm_engine.py
```

```
def _get_test_sampling_params(
```

```

prompt_list: list[str],
seed: int | None = 42,
structured_outputs: bool = False,
) -> tuple[list[SamplingParams], list[int]]:
    """Generate random sampling params for a batch."""
    # 使用局部 RNG 实例以隔离全局 random 状态
    # 关键修复: rng 的种子固定为 seed, 保证同一 session 不同顺序均可复现
    rng = random.Random(seed)

    def get_mostly_n_gt1() -> int:
        r"""Mostly n \in [2,20], ~1/3 n=1"""
        # 使用 rng.randint 而非 random.randint, 避免受全局状态影响
        x = rng.randint(0, 28)
        if x < 10:
            return 1
        else:
            return x - 8

    n_list = [get_mostly_n_gt1() for _ in range(len(prompt_list))]
    # High temperature to maximize the chance of unique completions
    return [
        SamplingParams(
            temperature=0.95,
            top_p=0.95,
            n=n,
            seed=seed,
            structured_outputs=StructuredOutputsParams(regex="[0-9]+")
            if structured_outputs
            else None,
        )
        for n in n_list
    ], n_list

```

评论区精华

无实质 review 讨论。机器人评论均未提出异议，Maintainer jikunshang 直接批准合并。

- 暂无高价值评论线程

风险与影响

- 风险：无风险。仅修改测试辅助函数，不影响产品代码；局部 RNG 实例隔离性强，不会引入回归或副作用。
- 影响：影响范围局限于 XPU 及 test_parallel_sampling 测试的稳定性，提升 CI flaky 指标。用户无感知。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR