

PR #25429 完整报告

sgl-project/sglang

[Test] Set default temperature to 0.0 in kl_test_utils

合并时间: 2026-05-16 09:20

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/25429>

执行摘要

- 一句话: KL 测试工具默认温度改为 0.0
- 推荐动作: 值得合并, 测试稳定性改进。建议确认所有调用方均无依赖旧默认值的意外行为。

功能与动机

原默认温度 `temperature=1` 会引入随机性, 导致 KL 散度测试结果不稳定。PR 旨在通过温度设为 0.0 使输出确定性, 便于 KL 测试通过。

实现拆解

1. 在 `python/sglang/test/kl_test_utils.py` 中: - `_generate` 函数签名 `temperature=1` → `temperature=0.0` - `_get_input_logprobs` 函数签名 `temperature=1` → `temperature=0.0`
2. 两个函数内部逻辑不变, 仅默认值变更, 调用方无需修改。

关键文件:

- `python/sglang/test/kl_test_utils.py` (模块 测试工具; 类别 `test`; 类型 `test-coverage`; 符号 `_generate`, `_get_input_logprobs`): 核心变更文件, 修改默认温度参数以稳定 KL 测试。

关键符号: `_generate`, `_get_input_logprobs`

关键源码片段

`python/sglang/test/kl_test_utils.py`

核心变更文件, 修改默认温度参数以稳定 KL 测试。

```
# python/sglang/test/kl_test_utils.py
# 变更: 默认温度从 1 改为 0.0, 使测试结果确定
def _generate(
    base_url,
    input_ids,
    max_new_tokens,
    return_logprob=False,
    logprob_start_len=-1,
    temperature=0.0, # 原为 1
):
    """Send generate request and return results."""
```

```
json_data = {
    "input_ids": input_ids,
    "sampling_params": {
        "temperature": temperature,
        "max_new_tokens": max_new_tokens,
        "ignore_eos": True,
    },
}
# ... 其余代码不变 ...
```

```
def _get_input_logprobs(base_url, new_input_ids, output_logprobs, temperature=0.0):
    """Run prefill to get input logprobs matching output logprobs."""
    # ... 内部调用 _generate, 温度传递 .....
```

评论区精华

无 review 评论，仅作者在 PR body 中说明动机：为确定性测试结果而修改。

- 暂无高价值评论线程

风险与影响

- 风险：影响范围极小：仅测试辅助函数默认值变更，不影响生产代码。若已有显示传入 `temperature` 的调用，则不受影响；若依赖默认值 1 的调用，行为会变，但测试套件本应保证确定性，风险低。
- 影响：仅影响 KL 测试工具用户。变更后 KL 测试结果更稳定，减少因随机性导致的 flaky 测试。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR