

PR #44472 完整报告

vllm-project/vllm

[XPU] Enable v1/sample tests on XPU CI

合并时间: 2026-07-10 15:40

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/44472>

执行摘要

- 一句话: XPU CI 启用 v1/sample 全量测试
- 推荐动作: 该 PR 变更简单, 适合快速合并。XPU 团队后续应关注 CI 执行情况, 确保所有测试通过。

功能与动机

PR body 明确说明 "Enable the full **v1/sample** test suite on XPU CI", 目的是在 XPU 持续集成中运行完整的 v1/sample 测试套件, 确保 XPU 设备上 sampler 相关功能的正确性。

实现拆解

1. 修改 CI 配置: 在 .buildkite/intel_jobs/misc_intel.yaml 的 commands 段中, 原先用 3 行 pytest 分别运行 test_topk_topp_sampler.py、test_logprobs.py 和 test_logprobs_e2e.py, 现在合并为 1 行 pytest -v -s v1/sample', 即运行 tests/v1/sample 目录下所有测试。
2. 不涉及源码变更: 本次只改动了 CI 配置, 无 Python 或其他语言代码改动。

关键文件:

- .buildkite/intel_jobs/misc_intel.yaml (模块 CI 配置; 类别 config; 类型 configuration) : 唯一的变更文件, 将原先逐个指定 v1/sample 测试文件的方式改为运行整个目录, 简化配置并提升测试覆盖率。

关键符号: 未识别

关键源码片段

.buildkite/intel_jobs/misc_intel.yaml

唯一的变更文件, 将原先逐个指定 v1/sample 测试文件的方式改为运行整个目录, 简化配置并提升测试覆盖率。

```
# .buildkite/intel_jobs/misc_intel.yaml (部分)
# 变更前: 单独运行 3 个测试文件
# pytest -v -s v1/sample/test_topk_topp_sampler.py &&
# pytest -v -s v1/sample/test_logprobs.py &&
# pytest -v -s v1/sample/test_logprobs_e2e.py
# 变更后: 直接运行整个 v1/sample 测试目录
```

commands:

```
- >-  
bash .buildkite/scripts/hardware_ci/run-intel-test.sh  
'pip install lm_eval[api]>=0.4.12 &&  
export VLLM_WORKER_MULTIPROC_METHOD=spawn &&  
cd tests &&  
pytest -v -s v1/logits_processors --ignore=v1/logits_processors/test_custom_online.py --  
ignore=v1/logits_processors/test_custom_offline.py &&  
pytest -v -s v1/test_oracle.py &&  
pytest -v -s v1/test_request.py &&  
pytest -v -s v1/test_outputs.py &&  
pytest -v -s v1/sample'
```

评论区精华

无实质性讨论。只有 Mergify 机器人自动提示存在合并冲突，需 rebase；另外 Claude 机器人评论告知 fork PR 无法自动审查。最终由 jikunshang 批准合并。

- 暂无高价值评论线程

风险与影响

- 风险：低风险。变更仅涉及 CI 配置，不修改任何功能代码。如果新增测试中有不兼容 XPU 的用例，可能导致 CI 任务失败，但可通过后续调整快速回滚。
- 影响：
 - 用户：无直接影响。
 - 系统：XPU CI 的测试覆盖范围扩大，v1/sample 目录下所有测试用例都会在 XPU 上运行，有助于提前发现 XPU 平台上的 sampler 回归问题。
 - 团队：XPU 维护者需关注 CI 测试结果，若新增测试失败需及时排查。
 - 风险标记：仅 CI 配置变更，低风险

关联脉络

- PR #48126 Add XPU nightly and release image publishing to DockerHub: 同为 XPU 基础设施相关 PR，涉及 XPU CI 持续集成和发布流程。