

PR #47015 完整报告

vllm-project/vllm

Fix transient dependency issues caused by `requirements/common.txt`

合并时间: 2026-06-30 05:20

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/47015>

执行摘要

- 一句话: 修复测试依赖固定与 xgrammar 兼容性
- 推荐动作: 建议技术管理者关注此 PR 采用的依赖管理策略: 将公共依赖固定在 common.txt 中, 各平台测试依赖通过 -r 引用并用 pip-compile 生成锁定文件。这种方法值得在项目其他部分推广。测试断言的语义化改进也值得参考。

功能与动机

根据 PR 描述, 测试可能因 requirements/common.txt 中的依赖未在测试依赖文件中固定而导致临时依赖版本浮动, 引发 CI 失败。同时, xgrammar 0.2.3 在 json_schema_content 中新增了 any_order 字段, 导致 test_get_model_structural_tag_supports_vllm_hermes 中的精确字典比较断言失败。因此需要同时修复依赖固定和测试断言方式。

实现拆解

1. 测试断言语义化: 修改 tests/tool_parsers/test_structural_tag_registry.py, 将 test_get_model_structural_tag_supports_vllm_hermes 中的 assert tag.model_dump() == {...} 逐字段比较结构 (type、format、tags 内容等), 避免因 xgrammar 版本升级增加新键 (如 any_order) 导致断言失败。
2. 测试依赖固定: 在 requirements/test/cuda.in 和 requirements/test/xpu.in 顶部添加 -r ./common.txt, 使 pip-compile 自动将 common.txt 中的依赖纳入锁定文件。requirements/test/rocm.in 已存在该引用, 保持不变。
3. 剔除重复依赖: 从 cuda.in、rocm.in 中移除已在 common.txt 中声明的包 (einops、opencv-python-headless、numpy、pydantic), 避免重复声明。
4. 生成锁定文件: 重新运行 pip-compile 更新 cuda.txt、rocm.txt、xpu.txt, 将 common.txt 中的依赖 (如 anthropic、compressed-tensors、depyf 等) 及其子依赖加入, 确保所有测试依赖版本固定。
5. Docker 适配: 在 docker/Dockerfile.cpu 中对应调整, 使 CPU 镜像也能使用更新后的依赖管理。

关键文件:

- tests/tool_parsers/test_structural_tag_registry.py (模块 测试; 类别 test; 类型 test-coverage): 核心测试改动, 将精确字典比较改为语义化断言, 兼容 xgrammar 0.2.3

- requirements/test/cuda.in (模块 依赖配置; 类别 test; 类型 test-coverage) : 添加 common.txt 引用并剔除重复包, 是依赖固定的关键配置变更
- requirements/test/rocm.in (模块 依赖配置; 类别 test; 类型 test-coverage) : 剔除重复依赖, 保持与 CUDA 一致的依赖管理
- requirements/test/xpu.in (模块 依赖配置; 类别 test; 类型 test-coverage) : 添加 common.txt 引用, 实现对 XPU 平台的依赖固定
- requirements/test/cuda.txt (模块 依赖锁定; 类别 docs; 类型 documentation) : 更新后的锁定文件, 包含 common.txt 的所有依赖, 保证可重现性
- requirements/test/xpu.txt (模块 依赖锁定; 类别 docs; 类型 documentation) : 更新后的锁定文件, 包含 common.txt 的所有依赖, 保证可重现性
- requirements/test/rocm.txt (模块 依赖锁定; 类别 docs; 类型 documentation) : 更新后的锁定文件, 与 common.txt 同步
- docker/Dockerfile.cpu (模块 容器配置; 类别 infra; 类型 infrastructure) : 适配新依赖管理, 确保 CPU 镜像构建正确

关键符号: test_get_model_structural_tag_supports_vllm_hermes,
test_hermes_required_tool_calls_use_empty_separator

评论区精华

本次 PR 未触发实质技术讨论。[chaunceyjiang](#) 与 [mgoin](#) 直接 Approved, [claude\[bot\]](#) 自动评论由于来自 fork 未执行审查。作者在 CI 评论中说明 Hybrid SSM 失败是 main 分支已有的问题, 与本次改动无关。

- 暂无高价值评论线程

风险与影响

- 风险: 风险极低: 测试断言改为语义比较, 覆盖了与之前相同的验证点; 依赖固定通过 pip-compile 生成锁定文件, 遵循推荐实践。唯一需留意的是 .txt 锁文件的更新可能引入其他包的版本变更, 但作者已运行完整 CI 确认无回归。
- 影响: 对开发者: CI 环境完全可重现, pip install -r requirements/test/<platform>.txt 将安装与 CI 一致的依赖, 避免本地测试因依赖版本差异而与 CI 不符。对用户: 无直接影响, 所有变更限于测试基础设施和测试用例。对系统: 降低因临时依赖问题导致的 CI 失败概率, 提升测试可靠性。
- 风险标记: 低风险, CI 依赖管理改进, 测试健壮性提升

关联脉络

- 暂无明显关联 PR