

PR #36465 完整报告

sgl-project/sglang

[Kernel] Declare the PyTorch ABI dependency in sglang-kernel wheels

合并时间: 2026-08-27 06:24

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/36465>

PR 分析: sglang-kernel 声明 PyTorch ABI 依赖

执行摘要

此 PR 通过修改 `python/sglang/kernels/aot/pyproject.toml`, 为 sglang-kernel 的 CUDA wheel 显式声明 PyTorch 2.13.0 依赖, 解决了因缺少 ABI 约束而导致的安装不兼容问题。变更简单、聚焦, 修复了 issue #36462, 对用户安装和依赖解析有积极影响。

功能与动机

`sglang-kernel` 包含针对特定 PyTorch C++ ABI 编译的原生扩展, 但 wheel 元数据未声明对 `torch` 的运行时依赖, 导致 pip 可解析安装与编译版本不匹配的 PyTorch, 导入即崩溃。issue #36462 详细说明了该问题和 ABI 映射, 当前 `main` 文档与构建均使用 Torch 2.13.0, 因此需在元数据中固化该版本。

实现拆解

1. 构建系统依赖: 在 `[build-system] requires` 中将 `torch>=2.8.0` 改为 `torch==2.13.0`, 确保隔离构建环境使用与发布工作流一致的 PyTorch ABI。
2. 运行时依赖: 在 `[project] dependencies` 中添加 `torch==2.13.0`, 使生成的 wheel 元数据声明运行时依赖, 阻绝不兼容组合。
3. 作用范围: 此变更仅针对 CUDA 包; CPU、ROCm、MUSA 使用独立配置, 需各自声明 ABI 要求。
4. 测试: 无运行时影响, 未添加测试; 已通过 PEP 440 校验和元数据检查。

`python/sglang/kernels/aot/pyproject.toml`

核心变更文件, 声明 sglang-kernel 对 PyTorch 的 ABI 依赖, 修复安装兼容性问题。

```
[build-system]
requires = [
    "scikit-build-core>=0.10",
    "torch==2.13.0", # 固定构建隔离环境的 PyTorch ABI, 避免解析器选择不同版本
    "wheel",
]
build-backend = "scikit_build_core.build"

[project]
name = "sglang-kernel"
```

```
version = "0.4.6.post1"
dependencies = ["torch==2.13.0"] # 声明运行时依赖，使安装元数据与构建 ABI 一致
```

评论区精华

无讨论，审阅者 Fridge003 直接批准。

风险与影响

风险：固定 `torch==2.13.0` 可能导致解析器在一个已有其他 torch 版本的环境中选择回溯或失败，但这是预期的兼容性约束。未进行完整 CUDA wheel 构建验证，但元数据和 PEP 440 校验已通过。影响：仅影响 `sglang-kernel` 的 CUDA wheel 安装元数据，用户将从更可靠的依赖解析中受益，避免 ABI 不匹配的导入错误；对运行代码和性能无影响。

关联脉络

此 PR 直接修复 issue #36462，该 issue 详细描述了元数据缺失问题及 ABI 映射表。同时，这与近期多个依赖管理相关的 PR（如 #36493 升级 mooncake）体现项目对依赖和元数据声明一致性的重视。未来可能需要为 CPU/ROCm/MUSA 包添加类似的精确依赖声明，以全面解决跨平台 ABI 兼容问题。