

PR #50144 完整报告

vllm-project/vllm

[CPU] Fix s390x builds and update torch version in dockerfile

合并时间: 2026-07-29 12:52

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/50144>

执行摘要

- 一句话: 修复 s390x 构建并升级 Docker 中 PyTorch 版本
- 推荐动作: 该 PR 是针对特定平台 (s390x) 的修复和依赖升级, 内容明确, 风险低。建议合并。

功能与动机

IBM s390x 平台上 VLLM 构建失败, 原因是 `FP32Vec4`、`FP32Vec8` 和 `FP32Vec16` 的拷贝构造函数被标记为 `explicit`, 导致无法进行隐式类型转换, 编译器报错。同时需要升级 PyTorch 版本以支持新特性和修复。

实现拆解

1. 修正拷贝构造函数声明: 在 `csrc/cpu/cpu_types_vxe.hpp` 文件中, 将 `FP32Vec4`、`FP32Vec8` 和 `FP32Vec16` 的拷贝构造函数从 `explicit` 改为非 `explicit`, 以允许隐式类型转换, 修复 s390x 上的编译错误。
2. 升级 Dockerfile 中的依赖版本: 在 `docker/Dockerfile.s390x` 中, 将 PyTorch 版本从 2.11.0 升级到 2.13.0, 并将 torchvision 版本从 v0.26.0 升级到 v0.28.0, 以匹配新 PyTorch 版本。

关键文件:

- `csrc/cpu/cpu_types_vxe.hpp` (模块 CPU 内核; 类别 source; 类型 core-logic; 符号 `FP32Vec4::FP32Vec4`, `FP32Vec8::FP32Vec8`, `FP32Vec16::FP32Vec16`): 修复核心: 将三个 `FP32Vec` 类的拷贝构造函数从 `explicit` 改为非 `explicit`, 解决 s390x 编译错误。
- `docker/Dockerfile.s390x` (模块 Docker; 类别 infra; 类型 infrastructure): 升级 PyTorch 和 torchvision 版本, 确保 s390x 构建环境兼容性。

关键符号: `FP32Vec4::FP32Vec4`, `FP32Vec8::FP32Vec8`, `FP32Vec16::FP32Vec16`

关键源码片段

`csrc/cpu/cpu_types_vxe.hpp`

修复核心: 将三个 `FP32Vec` 类的拷贝构造函数从 `explicit` 改为非 `explicit`, 解决 s390x 编译错误。

```
// FP32Vec4: 移除 explicit 以允许隐式拷贝
```

```

struct FP32Vec4 : public Vec<FP32Vec4> {
    // ...
    FP32Vec4(const FP32Vec4& data) : reg(data.reg) {} // 不再 explicit
};

// FP32Vec8: 移除 explicit
struct FP32Vec8 : public Vec<FP32Vec8> {
    // ...
    FP32Vec8(const FP32Vec8& data) { // 不再 explicit
        reg.val[0] = data.reg.val[0];
        reg.val[1] = data.reg.val[1];
    }
    // ...
};

// FP32Vec16: 移除 explicit
struct FP32Vec16 : public Vec<FP32Vec16> {
    // ...
    FP32Vec16(const FP32Vec16& data) { // 不再 explicit
        reg.val[0] = data.reg.val[0];
        reg.val[1] = data.reg.val[1];
        reg.val[2] = data.reg.val[2];
        reg.val[3] = data.reg.val[3];
    }
    // ...
};

```

docker/Dockerfile.s390x

升级 PyTorch 和 torchvision 版本，确保 s390x 构建环境兼容性。

```

# 安装 torchvision
ARG TORCH_VISION_VERSION=v0.28.0 # 从 v0.26.0 升级
WORKDIR /tmp
RUN --mount=type=cache,target=/root/.cache/uv \
    git clone https://github.com/pytorch/vision.git && \
    cd vision && \
    git checkout $TORCH_VISION_VERSION && \
    uv pip install torch==2.13.0 --index-url https://download.pytorch.org/whl/cpu && \ # 从 2.11.0 升级
    python setup.py bdist_wheel

```

评论区精华

PR 创建者请求审查，维护者 bigPYJ1151 通过 [/ci run](#) 触发 CI，并最终批准。无实质性审查讨论。

- 暂无高价值评论线程

风险与影响

- 风险：风险较低。拷贝构造函数从 explicit 改为非 explicit 可能会引入隐式转换，但这些类主要用于内部向量运算，隐式拷贝在实际使用中很常见且安全。PyTorch 版本升级可能引入 API 不兼容，但跳过一个版本且已有其他架构验证，风险可控。
- 影响：影响范围较小，仅涉及 s390x 架构的构建流程。对于使用该架构的用户，构建将恢复正常。PyTorch 版本升级会带来性能优化和新特性。
- 风险标记：隐式类型转换，依赖版本升级

关联脉络

- 暂无明显关联 PR