

PR #43881 完整报告

vllm-project/vllm

[ROCm] cmake: support PYTORCH_FOUND_HIP for torch 2.13 native HIP language support

合并时间: 2026-05-30 13:16

原文链接: <http://prhub.com.cn/vllm-project/vllm/pull/43881>

执行摘要

- 一句话: 修复 PyTorch 2.13 下 ROCm 构建失败
- 推荐动作: PR 变更明确且经过审批, 可安全合并。内部无需精读, 但涉及 CMake 配置维护的工程师可了解此兼容性处理模式。

功能与动机

PyTorch 2.13 switched from FindHIP.cmake to CMake native enable_language(HIP) in pytorch/pytorch@d921fd0, setting PYTORCH_FOUND_HIP but never HIP_FOUND, causing vLLM's CMakeLists.txt to fall through to the fatal error path.

实现拆解

1. 定位问题: 在 CMakeLists.txt 中, HIP 检测仅依赖 HIP_FOUND, 而 PyTorch 2.13 不再设置该变量。
2. 修改 CUDA 分支: 在 if (NOT HIP_FOUND AND CUDA_FOUND) 前增加 NOT PYTORCH_FOUND_HIP, 避免在 HIP 已安装时进入 CUDA 分支。
3. 修改 HIP 分支: 将 elseif(HIP_FOUND) 改为 elseif(HIP_FOUND OR PYTORCH_FOUND_HIP), 使系统能够正确识别 HIP 环境。仅修改了 CMakeLists.txt 中的两个条件行 (+2/-2), 无其他文件变动。

关键文件:

- CMakeLists.txt (模块 构建配置; 类别 config; 类型 configuration): 唯一变更文件, 修改了 HIP 检测条件, 修复 PyTorch 2.13 兼容性。

关键符号: 未识别

关键源码片段

CMakeLists.txt

唯一变更文件, 修改了 HIP 检测条件, 修复 PyTorch 2.13 兼容性。

```
# Set up GPU language and check the torch version and warn if it isn't
# what is expected.
if (NOT HIP_FOUND AND NOT PYTORCH_FOUND_HIP AND CUDA_FOUND)
  set(VLLM_GPU_LANG "CUDA")
```

```
if (NOT Torch_VERSION VERSION_EQUAL ${TORCH_SUPPORTED_VERSION_CUDA})
  message(WARNING "Pytorch version ${TORCH_SUPPORTED_VERSION_CUDA} "
    "expected for CUDA build, saw ${Torch_VERSION} instead.")
endif()
elseif(HIP_FOUND OR PYTORCH_FOUND_HIP)
  set(VLLM_GPU_LANG "HIP")
  # ...
endif()
```

评论区精华

讨论极少：Harry-Chen 向 tjtanaa 发起了评审请求，tjtanaa 回复“Looks fine to me that we can compile the code for AMD CI”。作者提到 CI 可能卡在了无关测试上。

- 评审请求 (other): tjtanaa 随后回复并批准。
- CI 状态 (other): 最终 CI 通过并合并。

风险与影响

- 风险：风险极低。改动仅涉及 CMake 中两个条件判断，且逻辑与旧版 PyTorch 完全兼容（PYTORCH_FOUND_HIP 在 torch \leq 2.12 中不存在，条件短路为假）。不影响 CUDA 路径或其他平台。
- 影响：影响范围限于使用 ROCm 和 PyTorch \geq 2.13 的用户，修复了此前无法编译的问题。对已有 CUDA 构建无影响。团队需注意未来 PyTorch 内部 CMake 变量的进一步变化。
- 风险标记：构建修复，条件判断变更，低风险

关联脉络

- 暂无明显关联 PR