

PR #2222 完整报告

radixark/miles

[AMD] Remove PR#301 FileSystemWriterAsync swap — superseded by Megatron-LM#74

合并时间: 2026-08-06 14:33

原文链接: <http://prhub.com.cn/radixark/miles/pull/2222>

执行摘要

- 一句话: 移除 ROCm 检查点写入器替换, 依赖上游修复
- 推荐动作: 值得快速浏览, 重点确认两点: 一是运行集群的 Megatron-LM 版本已包含 #74 修复; 二是 ROCm CI 对 checkpoint 保存路径有覆盖。设计决策上, 将平台 workaround 上推至上游并在本地及时清理, 是值得借鉴的做法。

功能与动机

PR body 明确说明: PR#301 引入 ROCmFileSystemWriterAsync 以修复 ROCm pinned-memory segfault, 通过子类化 FileSystemWriterAsync 并重写 preload_tensors; Megatron-LM#74 在同一基类上修复了该 bug, 使本地子类替换不再必要。关联 Issue Megatron-LM#74 标题为 'Tiny fix fsdp distributed update weight error', 属于上游修复, 因此本地 workaround 应清除。

实现拆解

1. 删除 miles/utils/rocm_checkpoint_writer.py: 整个文件 (27 行) 被移除, 包含 ROCmFileSystemWriterAsync 类与 preload_tensors 静态方法的实现。该实现通过将 non_blocking 强制置 False 避免 HIP 下张量进入 pinned memory 后 fork 崩溃。
2. 在 miles/backends/megatron_utils/model.py 的 initialize_model_and_optimizer() 中删除 HIP 检测块 (8 行): 原先在模型初始化前替换 filesystem_async 模块的 FileSystemWriterAsync 类, 现在直接使用上游基类。
3. 在 tools/convert_hf_to_torch_dist.py 的 main() 中删除同类注入块 (10 行): 原先同时替换 filesystem_async 与 torch 两个策略模块的 FileSystemWriterAsync, 现一并移除。
4. 配套说明: 无测试、配置或文档变更; 行为正确性完全依赖 Megatron-LM 版本中是否包含 #74 修复。

关键文件:

- miles/utils/rocm_checkpoint_writer.py (模块 工具层; 类别 source; 类型 deletion; 符号 ROCmFileSystemWriterAsync, preload_tensors): PR 核心改动: 整个文件被删除, 定义的 ROCmFileSystemWriterAsync 及 preload_tensors 是此前 ROCm 检查点 workaround 的载体。
- miles/backends/megatron_utils/model.py (模块 后端; 类别 source; 类型 core-logic; 符号 initialize_model_and_optimizer): ROCm workaround 的主要注入点之一:

initialize_model_and_optimizer() 原先在 HIP 分支替换 FileSystemWriterAsync 类，删除后模型初始化不再依赖本地补丁。

- tools/convert_hf_to_torch_dist.py (模块 转换工具；类别 source；类型 refactor；符号 main)：权重转换工具的注入点：main() 原先在 HIP 下同时替换 filesystem_async 与 torch 策略模块的 FileSystemWriterAsync，一并移除。

关键符号：ROCmFileSystemWriterAsync.preload_tensors,
initialize_model_and_optimizer, tools.convert_hf_to_torch_dist.main

关键源码片段

miles/utils/rocm_checkpoint_writer.py

PR 核心改动：整个文件被删除，定义的 ROCmFileSystemWriterAsync 及 preload_tensors 是此前 ROCm 检查点 workaround 的载体。

- # 文件 miles/utils/rocm_checkpoint_writer.py (本 PR 已整体删除)
- # 背景：PR#301 曾以子类化 FileSystemWriterAsync 的方式规避 ROCm/HIP 下
- # pinned-memory 导致的 fork 段错误；该 workaround 已被 Megatron-LM#74 上游修复取代。

```
import torch
from megatron.core.dist_checkpointing.strategies.filesystem_async import
FileSystemWriterAsync

class ROCmFileSystemWriterAsync(FileSystemWriterAsync):
    """ROCm 兼容的 FileSystemWriterAsync 包装类，本 PR 中已删除。"""

    @staticmethod
    def preload_tensors(*args, **kwargs):
        # HIP 平台上将 non_blocking 强制改为 False,
        # 避免张量进入 pinned memory 后触发 fork 段错误。
        if torch.version.hip:
            print("HIP/ROCm detected: setting non_blocking=False in preload_tensors")
            if "non_blocking" in kwargs:
                kwargs["non_blocking"] = False
            elif len(args) > 1 and isinstance(args[-1], bool):
                # non_blocking 通常是最后一个位置参数
                args = args[:-1] + (False,)
            return FileSystemWriterAsync.preload_tensors(*args, **kwargs)
```

评论区精华

本 PR 无 review 评论。评审人 guapisolo 直接批准 (APPROVED, body 为空)；关联 Issue Megatron-LM#74 上有 XinyuJiangCMU 的 'LGTM'。整体属于低争论的清理型变更。

- 整体评审 (other): 无需进一步修改，直接合并。

风险与影响

- 风险：主要风险来自对上游修复的依赖：若运行环境中 Megatron-LM 版本未包含 #74 修复，ROCm/HIP 训练与 checkpoint 保存路径可能重新出现 pinned-memory 导致的 fork 段错误。受影响文件为 miles/backends/megatron_utils/model.py 与 tools/convert_hf_to_torch_dist.py，二者原先在 HIP 分支注入 workaround，删除后无本地兜底。此外，本 PR 未附带测试，ROCm 路径的回归只能依赖现有 ROCm CI（PR#1606 引入）覆盖。由于改动为纯删除，冲突风险较低，但需确认依赖版本。
- 影响：用户影响：AMD/ROCm 用户的行为不变（上游修复后行为一致），但若上游版本未更新，可能回归段错误；系统影响：减少运行时对本地模块的 monkey-patch，简化维护；团队影响：消除了与上游并行维护的补丁代码，后续升级 Megatron-LM 时不再需要同步该 workaround。影响程度中等偏小，集中在 ROCm 训练与权重转换两条路径。
- 风险标记：依赖上游修复版本，ROCm 路径回归风险，缺少直接测试覆盖

关联脉络

- PR #1606 ci(rocm): add ROCm CI workflow for MI300X self-hosted runners: 同为 ROCm/AMD 支持线；本 PR 删除的 workaround 对 ROCm 训练路径的回归需由该 CI 覆盖验证。
- PR #2215 fix(mtp): double-shift GPT-path MTP labels: 与本次均修改 miles/backends/megatron_utils/model.py，同一文件上的持续演进，需注意合并顺序与冲突。