PR #7126 完整报告

PaddlePaddle/FastDeploy

[Iluvatar] Fix cuda graph error for tp > 1 in ernie models

合并时间：2026-04-01 19:13

原文链接：http://prhub.com.cn/PaddlePaddle/FastDeploy/pull/7126

执行摘要

- 一句话：修复 Iluvatar 平台上 ERNIE 模型在 tensor parallel 大于 1 时的 cuda graph 错误。
- 推荐动作：该 PR 值得精读以理解 Iluvatar 平台上的 cuda graph 处理策略和分布式通信优化。重点关注 tensor_model_parallel_all_reduce 函数中的平台分支逻辑设计，以及模型运行器中的条件禁用机制，这些是适配异构平台的关键技术点。

功能与动机

根据 PR body，动机是 ' 修复 ernie 模型当 tp>1 时，cuda graph 报错 '，直接针对 Iluvatar 平台上的特定错误，确保 ERNIE 模型在分布式并行环境下能正常使用 cuda graph 功能。

实现拆解

实现方案分为四个关键部分：1) 在 fastdeploy/distributed/communication.py 中修改 tensor_model_parallel_all_reduce 函数，为 Iluvatar 平台添加特殊路径，使用 stream.all_reduce 并设置 use_calc_stream=True，以避免 cuda graph 错误；2) 在模型加载器文件 default_loader.py 和 default_loader_v1.py 中扩展 clean_memory_fragments 方法，将内存清理逻辑应用到 Iluvatar 平台；3) 在 fastdeploy/worker/iluvatar_model_runner.py 的 __init__ 方法中添加条件判断，对于 ERNIE-VL 模型在 tp>1 时禁用 cuda graph，并优化日志记录；4) 更新 CI 测试脚本 scripts/run_ci_iluvatar.sh，添加 tp=2 的测试用例以验证修复。

关键文件：

- fastdeploy/distributed/communication.py（模块 Distributed Communication）：修改了核心通信函数 tensor_model_parallel_all_reduce，为 Iluvatar 平台添加特殊处理以避免 cuda graph 错误，是修复的关键部分。
- fastdeploy/worker/iluvatar_model_runner.py（模块 Worker）：添加了条件逻辑以在 ERNIE-VL 模型 tp>1 时禁用 cuda graph，并优化日志记录，直接影响模型运行行为。
- scripts/run_ci_iluvatar.sh（模块 CI）：扩展了 CI 测试脚本，添加 tp=2 的测试用例，确保修复被验证，增强了测试覆盖。

关键符号：tensor_model_parallel_all_reduce, clean_memory_fragments, IluvatarModelRunner.init

评论区精华

review 讨论非常有限，只有一次来自 EmmonsCurse 的批准评论 'LGTM～ /skip-ci coverage'，表示快速批准，没有深入的技术争议或设计权衡讨论。

- Review Approval (other): Review 通过，没有争议。

风险与影响

- 风险：技术风险包括：1) 修改 fastdeploy/distributed/communication.py 中的 tensor_model_parallel_all_reduce 函数引入了平台特定逻辑（Iluvatar 分支），可能影响其他平台的兼容性或性能，例如如果其他平台误用该逻辑；2) 在 iluvatar_model_runner.py 中禁用 cuda graph 可能导致 ERNIE-VL 模型在 tp>1 时性能下降；3) codecov 评论指出 patch coverage 为 42.86%（8 行缺失覆盖），表明测试覆盖不足，可能有未覆盖的边缘情况，如其他模型类型或配置组合。
- 影响：影响范围主要限于使用 Iluvatar 平台运行 ERNIE 模型的用户，特别是当配置 tensor parallel 大于 1 时。修复后，这些场景将不再出现 cuda graph 错误，提升了系统稳定性和用户体验。对团队的影响是增加了 Iluvatar 特定逻辑的维护负担，需确保未来变更不破坏其他平台功能。
- 风险标记：平台特定变更，测试覆盖率不足，性能潜在影响

关联脉络

- 暂无明显关联 PR