

PR #6731 完整报告

PaddlePaddle/FastDeploy

[CI] 【Hackathon 10th Spring No.34】 async_expert_loader 单测补充

合并时间: 2026-03-31 15:29

原文链接: <http://prhub.com.cn/PaddlePaddle/FastDeploy/pull/6731>

执行摘要

- 一句话: 为 `async_expert_loader` 模块补充单元测试, 覆盖率提升至 100%。
- 推荐动作: 该 PR 主要对测试工程师和模块维护者有参考价值:
 - 值得阅读以学习如何在 FastDeploy 中编写高质量单元测试, 特别是减少 mock、模拟外部依赖的技巧。
 - 对于关注 EPLB 或 Loader 模块的开发者, 可了解测试覆盖的分支和异常情况。
 - 由于不涉及生产代码变更, 无需深入分析核心算法。

功能与动机

根据 PR body, 动机是 'No.34 功能模块 fastdeploy/eplb/async_expert_loader.py 单测补充', 作为 PaddlePaddle Hackathon 任务, 目标是提高该模块的测试覆盖率以确保代码质量和稳定性。

实现拆解

实现集中在单个文件 `tests/eplb/test_async_expert_loader.py` 上:

1. 重构测试类为 `TestAsyncExpertLoader(unittest.TestCase)`, 以匹配同目录其他测试文件格式。
2. 添加辅助函数如 `_shm_buffer`、`_eplb_config` 和 `stub` 类 (如 `_StubSafeFile`、`_StubCudart`), 以模拟外部依赖 (CUDA、OS 层) 而减少 `MagicMock` 使用。
3. 扩展测试用例, 覆盖检查点解析、张量保存 / 加载往返、不同数据类型分支 (如 `bf16`)、加载器生命周期和异常处理路径。
4. 在 CI 环境中强制使用 CPU 设备, 避免 GPU 相关段错误。

关键文件:

- `tests/eplb/test_async_expert_loader.py` (模块 EPLB): 唯一被修改的文件, 包含所有单元测试的补充和重构, 覆盖 `async_expert_loader` 模块的各种分支和异常情况。

关键符号: `_shm_buffer`, `_eplb_config`, `TestAsyncExpertLoader.test_create_mmap`, `TestAsyncExpertLoader.test_load_ep_checkpoint`, `TestAsyncExpertLoader.test_save_tensor_to_shm_mem`

评论区精华

核心讨论围绕测试代码风格和 mock 使用：

- 审核者 CSWYF3634076 要求 ' 把单测都保留在 class TestAsyncExpertLoader(unittest.TestCase) 中，尽最大可能减少 mock 的内容，使用真实的逻辑 '。
- 作者 cloudforge1 响应并重构代码，参照同目录 test_eplb.py 等文件风格，将 bf16 等路径改为真实逻辑执行，仅保留必要 stub 用于 CUDA/OS 层。
- 讨论结论是 PR 被批准，表明更改符合要求。
- 测试代码风格和 mock 使用 (testing): 作者 cloudforge1 重构代码以满足要求，PR 被批准。

风险与影响

- 风险：技术风险较低：
 1. 测试代码不修改生产逻辑，无回归风险。
 2. 移除 MagicMock 并改用真实逻辑可能增加测试对环境的依赖性，例如在 CI 中强制使用 CPU 设备以避免 GPU 段错误（提交中已修复）。
 3. 新增的 stub 类需要维护，但仅用于测试，影响有限。
- 影响：影响范围局限于测试模块：
 1. 对用户无直接影响，因不改变功能或 API。
 2. 对系统：提高 async_expert_loader 模块的测试覆盖率至 100%，增强代码质量和未来变更的信心。
 3. 对团队：提供了测试编写的良好实践示例（减少 mock、使用真实逻辑），可能促进类似测试的改进。
- 风险标记：仅测试代码变更，CI 环境配置敏感

关联脉络

- PR #6730 [CI] 【Hackathon 10th Spring No.33】 config 单测补充：同属 Hackathon 任务，补充单元测试以提高覆盖率。
- PR #6771 [CI] 【Hackathon 10th Spring No.29】 engine unit test: 类似地，为引擎模块添加全面单元测试，覆盖率达到 92%。