

PR #2055 完整报告

THUDM/slime

[ci] clean up ci

合并时间: 2026-06-11 16:24

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2055>

执行摘要

- 一句话: 移除 e2e-test-short, 扩展 Megatron 测试矩阵
- 推荐动作: 值得关注:
 1. 确认被删除的测试是否被其他测试充分覆盖, 必要时添加缺失用例。
 2. 该 PR 展示了将临时 / 短测试合并到统一矩阵的模式, 可评估是否对新 PR 的测试组织有参考价值。

功能与动机

根据 commit message 和变更内容, 此 PR 旨在清理 CI 配置, 简化维护: 将分散的短测试合并到核心 Megatron 测试矩阵, 减少重复的 job 定义, 并移除不再推荐的测试路径。

实现拆解

1. 修改 Jinja2 模板文件 .github/workflows/pr-test.yml.j2: 移除 e2e-test-short job 定义, 在 megatron_tests 列表新增大量测试用例 (如 test_full_disk_weight_update.py、test_quick_start_glm4_9B.py 等), 并调整 e2e-test-image 使其复用 megatron_tests 矩阵。
2. 重新生成 .github/workflows/pr-test.yml: 移除了整个 e2e-test-short job (85 行删除), 只保留必要的结构定义 (3 行新增)。
3. 删除两个不再属于独立 job 的测试文件: tests/test_delta_weight_update.py (143 行) 和 tests/test_qwen2.5_0.5B_ppo_critic_only_short.py (129 行)。
4. 更新中英文 CI 文档 (docs/en/developer_guide/ci.md 和 docs/zh/developer_guide/ci.md): 移除关于 run-ci-short 的描述, 修改 run-ci-image 的描述以反映其复用 megatron 矩阵的变化。

关键文件:

- tests/test_delta_weight_update.py (模块测试; 类别 test; 类型 deletion; 符号 prepare, execute): 被删除的 E2E smoke 测试文件, 直接导致对应功能失去独立 CI 覆盖
- tests/test_qwen2.5_0.5B_ppo_critic_only_short.py (模块测试; 类别 test; 类型 deletion; 符号 prepare, execute): 被删除的 PPO critic-only 短测试, 对应功能失去独立 CI 覆盖

- .github/workflows/pr-test.yml (模块 CI 工作流; 类别 infra; 类型 infrastructure) : 生成的 CI 工作流文件, 删除了 e2e-test-short job (减少 85 行)
- .github/workflows/pr-test.yml.j2 (模块 CI 工作流; 类别 infra; 类型 infrastructure) : Jinja2 模板核心变更, 移除 e2e-test-short 并扩展 megatron_tests, 控制整个 CI 生成
- docs/en/developer_guide/ci.md (模块 开发者文档; 类别 docs; 类型 documentation) : 更新英文 CI 文档, 移除 run-ci-short 描述, 调整 run-ci-image 描述
- docs/zh/developer_guide/ci.md (模块 开发者文档; 类别 docs; 类型 documentation) : 更新中文 CI 文档, 与英文文档同步变更

关键符号: prepare, execute

关键源码片段

tests/test_delta_weight_update.py

被删除的 E2E smoke 测试文件, 直接导致对应功能失去独立 CI 覆盖

```
# 该文件已被删除, 之前用于测试 disk-backed delta weight updates。
# 它通过 prepare() 准备数据, execute() 运行训练并验证 delta 文件生成。
# 下方是被删除的完整源码片段:
```

```
"""E2E smoke test for disk-backed delta weight updates.
```

```
Runs a tiny Qwen3.5-0.8B job so the first weight update seeds the delta
snapshot and the post-train update publishes sparse delta files through
``update_weights_from_disk(load_format="delta", files=...)``.
"""
```

```
import os
import tempfile
from pathlib import Path
```

```
import slime.utils.external_utils.command_utils as U
```

```
MODEL_NAME = "Qwen3.5-0.8B"
MODEL_TYPE = "qwen3.5-0.8B"
NUM_GPUS = 4
TORCH_DIST_CKPT = f"/dev/shm/{MODEL_NAME}_torch_dist"
```

```
def prepare():
    U.exec_command("mkdir -p /root/models /root/datasets")
    U.exec_command(f"hf download Qwen/{MODEL_NAME} --local-dir /root/models/{MODEL_NAME}")
    U.hf_download_dataset("zhuzilin/gsm8k")
    U.convert_checkpoint(
        model_name=MODEL_NAME,
        megatron_model_type=MODEL_TYPE,
```

```

    num_gpus_per_node=NUM_GPUS,
    dir_dst="/dev/shm",
)

def execute():
    with tempfile.TemporaryDirectory(prefix="slime_delta_weight_update_") as delta_dir:
        ckpt_args = f"--hf-checkpoint /root/models/{MODEL_NAME}/ " f"--ref-load {TORCH_DIST_
        CKPT} "
        # ... (中间 args 省略, 详见上下文)
        pass

if __name__ == "__main__":
    prepare()
    for proxy_var in ("http_proxy", "https_proxy", "HTTP_PROXY", "HTTPS_PROXY"):
        os.environ.pop(proxy_var, None)
    execute()

```

[.github/workflows/pr-test.yml.j2](#)

Jinja2 模板核心变更, 移除 e2e-test-short 并扩展 megatron_tests, 控制整个 CI 生成

```

{# 以下是模板文件的部分片段, 展示 megatron_tests 定义和 jobs 调整 #}
<% set megatron_tests = [
    {'test_file': 'test_full_disk_weight_update.py', 'num_gpus': 4},
    {'test_file': 'test_quick_start_glm4_9B.py', 'num_gpus': 8, 'enable_eval': '0'},
    {'test_file': 'test_glm4.7_30B_A3B_pd_mooncake.py', 'num_gpus': 8},
    # ... 更多测试已添加, 覆盖多种模型和场景
] %>

<% set jobs = {
    # 'e2e-test-short' 已被移除
    'e2e-test-sglang-config': {
        'label': 'run-ci-sglang-config',
        'tests': [ ... ]
    },
    'e2e-test-megatron': {
        'label': 'run-ci-megatron',
        'tests': megatron_tests # 现在 megatron 矩阵包含更多测试
    },
    # ... 其他 jobs
} %>

```

评论区精华

本次 PR 无 review 评论和讨论。

- 暂无高价值评论线程

风险与影响

- 风险：主要风险在于测试覆盖的丢失：直接删除的两个测试文件（delta weight update 和 PPO critic-only）并未显式包含在新 megatron_tests 列表中，可能导致对应功能缺乏 CI 验证。虽然部分场景可能被其他 megatron 测试间接覆盖，但需要确认。另外，CI 工作流依赖生成的 YAML，模板变更需确保生成正确。
- 影响：影响主要限于开发者和 CI 维护者：e2e-test-short job 不再可用，需要使用 run-ci-megatron 或 run-ci-changed 触发相关测试。对最终用户无影响。团队需要更新 PR 标签使用习惯（不再使用 run-ci-short）。
- 风险标记：测试覆盖遗漏，CI 配置变更

关联脉络

- PR #2021 Support update_from_disk: 被删除的 test_delta_weight_update.py 直接测试了 update_from_disk 功能
- PR #2030 [examples]: add qwen3.5-9b model config and fully_async example: 被删除的 test_qwen2.5_0.5B_ppo_critic_only_short.py 与 qwen2.5 系列模型测试相关