

PR #2269 完整报告

radixark/miles

[fix] drop --disable-weights-backuper from the nemotron CI test

合并时间: 2026-08-08 11:01

原文链接: <http://prhub.com.cn/radixark/miles/pull/2269>

执行摘要

- 一句话: 修复 Nemotron CI 测试的废弃参数残留
- 推荐动作: 不值得精读, 但值得快速浏览作为一个跨 PR 清理的范例: 删除 CLI 参数时应当全局 grep 所有引用 (包括测试 extra_args、shell 脚本、示例配置), 避免遗漏导致 CI 失效。可顺带关注 PR #2223 引入的默认 rematerialize 行为在后续测试中的表现。

功能与动机

PR #2223 (commit 52ac174d9) 从 miles/utils/arguments.py 中移除了 --disable-weights-backuper, 但漏掉了 tests/e2e/megatron/model_scripts/test_nemotron_3_ultra_4layer_ci.py 中仍然传入该参数的注册。结果是每次 megatron 域 CI 运行都在 argparse 阶段直接失败 (train.py: error: unrecognized arguments: --disable-weights-backuper, 约 93 秒后且未进入任何 GPU 计算), 首次失败出现在 run 31223739466。PR body 指出 git grep disable-weights-backuper 在修复后已无结果。

实现拆解

解决过程分四步:

1. 定位残留引用: 在 tests/e2e/megatron/model_scripts/test_nemotron_3_ultra_4layer_ci.py 的 _args() 中, extra_args 仍以 ("--ci-test " "--ci-disable-logprobs-checker " " --disable-weights-backuper ") 拼接参数, 而对应 CLI 参数已在 miles/utils/arguments.py 中删除。
2. 删除残留参数: 将 extra_args 改为 ("--ci-test " "--ci-disable-logprobs-checker "), 保留 --ci-test 以确保 Megatron 到 SGLang 权重等价性检查继续执行; 该检查正是此测试的核心守护目标。
3. 选择最小验证集: PR body 说明选用 run-ci-model-scripts 标签 (17 个测试) 而非 megatron 标签 (43+ 个测试) 来验证修复, 以最小的 GPU 开销确认问题解决。
4. 配套动作: 无。该变更本身就在测试配置内, 不涉及参数 schema、部署或文档修改。

关键文件:

- tests/e2e/megatron/model_scripts/test_nemotron_3_ultra_4layer_ci.py (模块 模型脚本; 类别 test; 类型 test-coverage): 唯一变更文件: 在 _args() 的 extra_args 中删除已废弃的 --disable-weights-backuper 参数注册, 保留 --ci-test 与 --ci-disable-logprobs-checker, 修复 PR #2223 遗留的 argparse 失败。

关键符号: `_args`

关键源码片段

[tests/e2e/megatron/model_scripts/test_nemotron_3_ultra_4layer_ci.py](#)

唯一变更文件: 在 `_args()` 的 `extra_args` 中删除已废弃的 `--disable-weights-backuper` 参数注册, 保留 `--ci-test` 与 `--ci-disable-logprobs-checker`, 修复 PR #2223 遗留的 `argparse` 失败。

```
def _args() -> ScriptArgs:
    return ScriptArgs(
        model_org="CharyZeng",
        model_name="NVIDIA-Nemotron-3-Ultra-550B-A55B-BF16-4layer",
        mode="debug_minimal",
        num_nodes=1,
        num_gpus_per_node=8,
        num_rollout=2,
        rollout_batch_size=8,
        n_samples_per_prompt=2,
        global_batch_size=16,
        skip_saving=True,
        # PR #2223 (commit 52ac174d9) 已将 --disable-weights-backuper 从 arguments.py
        移除,
        # 此测试里残留的参数注册会让 CI 在 argparse 阶段直接失败。
        # 因此这里必须同步删除, 并保留 --ci-test 以便继续守护
        # Megatron 到 SGLang 的权重等价性检查 (该检查由 --ci-test 开启)。
        extra_args=("--ci-test " "--ci-disable-logprobs-checker "),
    )
```

评论区精华

本 PR 没有任何 review 评论, 审核者 Zhichenzzz 直接批准。关键决策说明都在 PR body 和 commit message 中: 参数删除遗漏是问题根源; 验证时选择 `run-ci-model-scripts` 标签 (17 个测试) 而不是 `megatron` (43+ 个测试), 是性价比的回归验证选择。

- 暂无高价值评论线程

风险与影响

- 风险: 风险很低: 变更只删除一个已不存在的参数, 不改变训练逻辑; 该测试的 CI 门禁指标 (`train/grad_norm`、`train/ppo_kl`、`train/train_rollout_logprob_abs_diff`、`train/train_rollout_kl`、`rollout/raw_reward`) 保持不变。影响限于 CI 层面: 修复 `megatron` 域 `suite` 的持续早期失败, 避免每次约 93 秒的无效运行。对用户无影响。需要注意的是, 这个修复依赖 PR #2223 的默认 `rematerialize` 行为; 如果将来该参数被重新引入, 需要重新评估此测试的覆盖。
- 影响: 影响范围集中在 CI 测试链路: 恢复 `tests/e2e/megatron/model_scripts/test_nemotron_3_ultra_4layer_ci.py` 在 `stage-c-8-gpu-h200 suite` 上的可用性, 消除 `megatron` 领域每个 PR 运行时的 `argparse` 早期失败。对最终用户没有任何行为影响; 对团队而言, 减

少了 CI 无效占用的 GPU 时间和失败噪音，并让 Nemotron-3-Ultra 4-layer 的权重一致性守护得以恢复。

- 风险标记：CI 回归修复，仅测试配套改动，上游参数移除遗漏

关联脉络

- PR #2223 Remove --disable-weights-backuper; default eligible colocate launchers to rematerialize: 直接上游 PR: 从 miles/utils/arguments.py 移除 --disable-weights-backuper 并默认启用 rematerialize，但漏掉了本测试中的参数注册，PR #2269 完成收尾。
- PR #2257 test(ci): disable router-equivalence and MiniMax-M2.7 E2E: 同属 CI 测试套件维护方向：禁用过时用例、释放 GPU 容量，与本 PR 一起体现 CI 套件持续清理的趋势。
- PR #2230 fix(ci): disable MI300X runner jobs: 同为 CI 测试配置调整：管理自托管 runner 与用例状态，与本 PR 同属 CI 可靠性维护范畴。