

PR #7466 完整报告

verl-project/verl

[cfg, megatron, doc] fix: drop unused actor.router_replay in favor of engine config

合并时间: 2026-08-24 10:43

原文链接: <http://prhub.com.cn/verl-project/verl/pull/7466>

执行摘要

- 一句话: 移除 actor.router_replay 死键, 路由回放统一走引擎配置
- 推荐动作: 值得快速精读。这是一个典型的“配置死键静默失效”修复案例, 核心看点是 review 中 wuxibin89 的“直接删除优于兼容”决策如何改变实现方向, 以及用契约测试锁定配置 schema 的写法。对后续任何“文档与代码不一致、双份配置入口”问题都有参考价值。

功能与动机

大 MoE 训练中训练引擎与推理引擎的路由计算差异会导致专家选择不一致、引入训练噪声, 路由回放 (R2/R3) 通过锁定专家路由路径来稳定训练 (见 #3762 引用的 GSPO 论文 arXiv:2507.18071)。#7463 定位到两套独立 schema: 顶层 actor.router_replay (死键, 无人消费) 与引擎 actor.{megatron,veomni}.router_replay (实键), worker 只读引擎副本, 官方 NPU 指南却指向死键。Issue 原文: "Following the NPU guide (or actor.yaml) silently leaves routing replay disabled. For large MoE this is train/infer route mismatch with no error." 同时 experimental/separation/ray_trainer.py 用顶层死键决定 R2/R3 冲突处理, 导致正确设置 actor.megatron.router_replay.mode=R2 的用户仍走 R3 分支。PR body 明确决策: "drops the unused top-level key instead of aliasing it"——删除而不是兼容。

实现拆解

1. 定位双份 schema 与消费差异: 依据 #7463 复现脚本确认
verl/workers/engine_workers.py 仅读取 self.config.actor.{megatron,veomni}.router_replay, 而 ActorConfig / actor.yaml 暴露的顶层 router_replay 在组装 TrainingWorkerConfig 时被丢弃, 是纯 no-op。
2. Schema 收敛: verl/workers/config/actor.py 删除 ActorConfig.router_replay: RouterReplayConfig 字段及其 Args 文档; verl/trainer/config/actor/actor.yaml 删除 19 行 router_replay 配置块; 同步清理 4 份生成 YAML (_generated_ppo_trainer.yaml、_generated_ppo_megatron_trainer.yaml、_generated_ppo_veomni_trainer.yaml、_generated_ppo_torchTitan_trainer.yaml), 保证 Hydra struct 模式实例化不报错。
3. 修正分离式训练器: verl/experimental/separation/ray_trainer.py 的 _fit_compute_log_prob 中, 路由冲突处理从读顶层死键改为根据 actor.strategy 定位 actor.megatron / actor.veomni 子配置再取 router_replay.mode, 非引擎策略回退 disabled (与旧默认行为等价)。

4. 文档与测试配套: transfer_to_npu_guide.md 的 R2/R3 示例改为引擎键并注明“顶层 actor.router_replay 已移除”; parameter_and_metrics.md 参数表同步修正。新增 tests/workers/config/test_actor_router_replay_sync_on_cpu.py, 3 个 CPU 测试分别断言“顶层无该字段”“megatron 引擎可承载 R3”“veomni 引擎可承载 R2”。
5. 实现演进: 首个 commit 按 issue 建议实现“honor or fail” (把顶层值复制到引擎键、冲突时抛错); wuxibin89 review 提出“直接删除”后, 第二个 commit 改为删除方案并同步更新全部文档与生成配置。

关键文件:

- verl/experimental/separation/ray_trainer.py (模块 分离训练; 类别 source; 类型 core-logic; 符号 _fit_compute_log_prob): 核心逻辑修复点: R2/R3 冲突处理从读顶层死键改为按 strategy 读引擎子配置, 消除与主 worker 的口径不一致。
- verl/workers/config/actor.py (模块 配置模型; 类别 source; 类型 configuration; 符号 ActorConfig): schema 源头修改: 从 ActorConfig 删除顶层 router_replay 字段, 使 Hydra struct 模式对死键直接报错。
- verl/trainer/config/actor/actor.yaml (模块 配置文件; 类别 config; 类型 configuration): 配置源头: 删除 19 行 router_replay 块, 官方 actor 配置与 dataclass 契约保持一致。
- tests/workers/config/test_actor_router_replay_sync_on_cpu.py (模块 单元测试; 类别 test; 类型 test-coverage; 符号 test_actor_config_has_no_top_level_router_replay, test_mcore_router_replay_lives_on_engine, test_veomni_router_replay_lives_on_engine): 新增契约测试: 断言 ActorConfig 无顶层 router_replay, 且 megatron / veomni 引擎子配置可正常承载 R2 / R3。
- verl/trainer/config/_generated_ppo_megatron_trainer.yaml (模块 生成配置; 类别 config; 类型 configuration): 生成配置同步移除顶层 router_replay 块, 避免 Hydra struct 模式下实例化报错。
- verl/trainer/config/_generated_ppo_torchitan_trainer.yaml (模块 生成配置; 类别 config; 类型 configuration): 生成配置同步移除顶层 router_replay 块, 保证与 actor.yaml 契约一致。
- verl/trainer/config/_generated_ppo_trainer.yaml (模块 生成配置; 类别 config; 类型 configuration): 生成配置同步移除顶层 router_replay 块, 保证与 actor.yaml 契约一致。
- verl/trainer/config/_generated_ppo_veomni_trainer.yaml (模块 生成配置; 类别 config; 类型 configuration): 生成配置同步移除顶层 router_replay 块, 保证与 actor.yaml 契约一致。
- docs/ascend_tutorial/dev_guide/model_dev/transfer_to_npu_guide.md (模块 NPU 文档; 类别 docs; 类型 documentation): 官方 Ascend NPU 指南曾指向死键, 修复后改为引擎键并声明顶层键已移除。
- docs/ascend_tutorial/dev_guide/model_dev/parameter_and_metrics.md (模块 NPU 文档; 类别 docs; 类型 documentation): 参数表同步修正 router_replay 键路径, 避免再次引入死键。

关键符号: _fit_compute_log_prob, test_actor_config_has_no_top_level_router_replay, test_mcore_router_replay_lives_on_engine, test_veomni_router_replay_lives_on_engine

关键源码片段

verl/experimental/separation/ray_trainer.py

核心逻辑修复点：R2/R3 冲突处理从读顶层死键改为按 strategy 读引擎子配置，消除与主 worker 的口径不一致。

```
# verl/experimental/separation/ray_trainer.py
# 分离式训练中 R2 / R3 的路由冲突处理，统一读取引擎侧配置。
# 修复前读取顶层死键，导致 `actor.megatron.router_replay.mode=R2`
# 时仍误走 R3 分支。
if "routed_experts" in batch.batch and "routed_experts" in old_log_prob.batch:
    actor_cfg = self.config.actor_rollout_ref.actor
    # 按 training strategy 定位对应的引擎子配置块
    if getattr(actor_cfg, "strategy", None) == "megatron":
        engine_cfg = getattr(actor_cfg, "megatron", None)
    elif getattr(actor_cfg, "strategy", None) == "veomni":
        engine_cfg = getattr(actor_cfg, "veomni", None)
    else:
        engine_cfg = None
    # 引擎配置缺失时回退 disabled，与旧默认行为等价
    router_mode = getattr(getattr(engine_cfg, "router_replay", None), "mode", "disabled")
    if router_mode == "R2":
        # R2 回放旧路由：丢弃当前前向计算出的 routed_experts
        batch.batch.pop("routed_experts")
    else:
        # R3 沿用新路由：丢弃 old_log_prob 中的 routed_experts
        old_log_prob.batch.pop("routed_experts")
```

tests/workers/config/test_actor_router_replay_sync_on_cpu.py

新增契约测试：断言 ActorConfig 无顶层 router_replay，且 megatron / veomni 引擎子配置可正常承载 R2 / R3。

```
# tests/workers/config/test_actor_router_replay_sync_on_cpu.py
from dataclasses import fields

from verl.workers.config.actor import ActorConfig, McoreActorConfig, VeOmniActorConfig
from verl.workers.config.engine import EngineRouterReplayConfig, McoreEngineConfig,
VeOmniEngineConfig
from verl.workers.config.optimizer import OptimizerConfig

def test_actor_config_has_no_top_level_router_replay():
    # 契约断言：ActorConfig 不得再暴露顶层 router_replay 字段
    assert "router_replay" not in {f.name for f in fields(ActorConfig)}

def test_mcore_router_replay_lives_on_engine():
    # Megatron 路由回放只挂在 actor.megatron.router_replay 上
    cfg = McoreActorConfig(
```

```

rollout_n=1,
ppo_micro_batch_size_per_gpu=1,
megatron=McoreEngineConfig(router_replay=EngineRouterReplayConfig(mode="R3")),
optim=OptimizerConfig(lr=1e-6),
)
assert not hasattr(cfg, "router_replay")
assert cfg.megatron.router_replay.mode == "R3"

def test_veomni_router_replay_lives_on_engine():
    # VeOmni 引擎同理：R2 模式挂在 actor.veomni.router_replay 上
    cfg = VeOmniActorConfig(
        rollout_n=1,
        ppo_micro_batch_size_per_gpu=1,
        use_remove_padding=True,
        veomni=VeOmniEngineConfig(router_replay=EngineRouterReplayConfig(mode="R2")),
        optim=OptimizerConfig(lr=1e-6),
    )
    assert not hasattr(cfg, "router_replay")
    assert cfg.veomni.router_replay.mode == "R2"

```

评论区精华

wuxibin89 在 [verl/trainer/config/actor/actor.yaml](#) 上提出关键意见："I think we can drop [actor.router_replay](#) since it have been move to [actor.{megatron,veomni}.router_replay](#)"——与其在 issue 建议的“复制 / 冲突报错”方案上做兼容层，不如直接删除死键，让配置模型只保留一个权威入口。作者回复确认删除方案落地："Dropped [actor.router_replay](#) from [actor.yaml](#) and [ActorConfig](#). Routing replay is now only [actor.{megatron,veomni}.router_replay](#), matching this review. Docs and generated trainer YAMLS are updated; [actor_rollout_ref.actor.router_replay.mode=...](#) will no longer compose." 即错误配置会立即失败而不是静默忽略。

- 是否直接删除 [actor.router_replay](#) 而不是做兼容 (design): 采用删除方案：保留引擎侧唯一入口，顶层键在 Hydra struct 模式下直接报错；作者第二个 commit 落实并同步文档与生成配置。
- 删除后的配置行为与文档同步确认 (other): 双方对齐：死键不再 compose，错误配置立即失败而不是静默忽略。

风险与影响

- 风险：兼容性（有意的破坏）：使用旧键 [actor_rollout_ref.actor.router_replay.mode=...](#) 的存量脚本会在 Hydra struct 模式下启动报错。由于旧键本就是 no-op，报错比静默失效更安全，但需在 release note 中给出迁移指引。行为修正：
experimental/separation/ray_trainer.py 的 R2/R3 判定改读引擎键后，若用户只设置了死键 R2 而引擎键保持 disabled，将从误走 R2 分支变为回退 disabled——这是 bugfix 的预期行为变化。残留不一致：PR body 明确 [ref.yaml](#) 的 [ref.router_replay](#) 未在本 PR 处理，属于已知遗留。测试覆盖局限：新测试只锁定 dataclass 级配置契约，未直接覆盖

_fit_compute_log_prob 的 R2/R3 分支逻辑，分离式训练器改动缺少针对性单测。生成配置漂移：4 份 _generated_*.yaml 与 actor.yaml 的一致性依赖生成脚本，后续重新生成配置时需防止死键复活。

- 影响：对用户：Ascend NPU 大 MoE 用户按文档配置 R3/R2 将真正生效；错误配置立即报错而非静默失效，避免 train/infer 路由不一致带来的训练不稳定。对系统：消除“第二套真相”，配置模型单源化；experimental/separation 与主 worker 的路由回放判定口径一致。对团队：新增契约测试降低回归风险；后续改动 engine 配置时需同步生成 YAML。影响范围主要集中在 Megatron / VeOmni + 路由回放用户，其余策略（fsdp / torchtitan 等）因 router_mode 回退 disabled 而行为不变。
- 风险标记：配置契约变更，存量脚本兼容，separation 行为修正，生成配置漂移，测试覆盖局限

关联脉络

- PR #7536 [cfg] fix: drop unused ref router replay config: 与本 PR 对称的清理：删除 ref.yaml 中未使用的 ref.router_replay，共同完成 router_replay 配置入口的全面收敛。
- PR #7407 [megatron,veomni] feat: use torch.int16 for routed_experts: 同属 MoE 路由回放功能线：路由索引改 int16、回放缓冲重构，与本 PR 的引擎侧 router_replay 配置紧密相关。