

PR #6296 完整报告

verl-project/verl

[rollout] fix: skip_tokenizer_init=True for OPD teacher

合并时间: 2026-05-09 16:54

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6296>

执行摘要

- 一句话: OPD teacher 跳过 tokenizer 初始化
- 推荐动作: 建议精读该 PR 用于了解 OPD teacher 的配置要求。值得关注的设计决策是将 skip_tokenizer_init 设为硬编码默认值而非从外部配置传递, 简化用户使用。

功能与动机

OPD teacher 需要跳过 tokenizer 初始化, 以启用 Token-In-Token-Out 模式。参见 PR body: "OPD teacher should skip tokenizer initialization to enable **Token-In-Token-Out**。"

实现拆解

1. 在 verl/trainer/config/distillation/distillation.yaml 中, 于 teacher_models.teacher_model.inference 下将 skip_tokenizer_init 从 false 改为 true。
2. 同步更新所有生成的 PPO trainer 配置文件, 确保一致性:
 - verl/trainer/config/_generated_ppo_megatron_trainer.yaml
 - verl/trainer/config/_generated_ppo_torchTitan_trainer.yaml
 - verl/trainer/config/_generated_ppo_trainer.yaml
 - verl/trainer/config/_generated_ppo_veomni_trainer.yaml 每个文件只有一行变更, 位于 distillation 段落的 teacher_models 下。

关键文件:

- verl/trainer/config/distillation/distillation.yaml (模块 蒸馏配置; 类别 config; 类型 configuration): 蒸馏主配置, 直接影响 OPD teacher 行为, 是变更的核心文件。
- verl/trainer/config/_generated_ppo_megatron_trainer.yaml (模块 生成配置; 类别 config; 类型 configuration): 生成的 Megatron 后端 PPO 配置, 包含蒸馏配置节, 同步修改以保持一致性。
- verl/trainer/config/_generated_ppo_torchTitan_trainer.yaml (模块 生成配置; 类别 config; 类型 configuration): 生成的 TorchTitan 后端 PPO 配置, 同步修改。
- verl/trainer/config/_generated_ppo_trainer.yaml (模块 生成配置; 类别 config; 类型 configuration): 生成的通用 PPO 配置, 同步修改。
- verl/trainer/config/_generated_ppo_veomni_trainer.yaml (模块 生成配置; 类别 config; 类型 configuration): 生成的 VeOmni 后端 PPO 配置, 同步修改。

关键符号：未识别

关键源码片段

verl/trainer/config/distillation/distillation.yaml

蒸馏主配置，直接影响 OPD teacher 行为，是变更的核心文件。

```
# verl/trainer/config/distillation/distillation.yaml ( 修改 )
teacher_models:
  teacher_model:
    _target_: verl.workers.config.DistillationTeacherModelConfig
    key: null
    model_path: null
    num_replicas: 0
    inference:
      _target_: verl.workers.config.RolloutConfig
      name: ${oc.select:actor_rollout_ref.rollout.name}
      dtype: ${oc.select:actor_rollout_ref.rollout.dtype}
      # ... 其他 rollout 参数 ...
      # 关键变更：设为 true 以启用 Token-In-Token-Out 模式
      skip_tokenizer_init: true
      prompt_length: ${oc.select:actor_rollout_ref.rollout.prompt_length}
      response_length: ${oc.select:actor_rollout_ref.rollout.response_length}
      temperature: ${oc.select:actor_rollout_ref.rollout.temperature}
```

评论区精华

无代码 review 讨论。仅有一条 gemini-code-assist 的自动评论表示无反馈。该 PR 未经人工 review，作者直接合并。

- 暂无高价值评论线程

风险与影响

- 风险：风险较低。变更仅涉及配置开关，逻辑简单。但仍需确认：如果某些 teacher 模型依赖 tokenizer 初始化（如需要词汇扩展），skip_tokenizer_init=True 可能导致推理行为异常。建议确保所有 OPD teacher 用例都支持 raw token 输入。
- 影响：影响范围：所有启用蒸馏训练的配置，包括 Megatron、TorchTitan、VeOmni 等后端。影响程度：对使用 OPD teacher 的蒸馏流程是必要修复，否则 teacher 无法正确运行。对其他不使用蒸馏的场景无影响。
- 风险标记：配置变更未验证依赖，缺少测试覆盖

关联脉络

- PR #6265 [reward, cfg] fix: correctly use RewardModelConfig in reward config: 同属配置修复类型，涉及生成配置与事实配置的一致性维护。