

PR #6502 完整报告

verl-project/verl

[fsdp] fix: add sp and use_remove_padding validation for SFT and RL in fsdp engine

合并时间: 2026-05-27 20:05

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6502>

执行摘要

- 一句话: 集中 SP 与 use_remove_padding 校验至 FSDP 引擎
- 推荐动作: 值得快速合并。这是一个低风险、高收益的 bugfix, 消除了 SFT 场景的校验盲区。
关注点: 需确认 SFT 相关文档和示例配置中是否已默认设置 use_remove_padding=True 以配合序列并行。

功能与动机

在 FSDP 引擎中, 当启用 Ulysses 序列并行 (ulysses_sequence_parallel_size > 1) 时必须同时启用 use_remove_padding。此前该限制校验只在 FSDPActorConfig.validate() 和 FSDPCriticConfig.post_init() 中存在, 而这两个类属于 RL 专用配置包装器; SFT 训练器直接使用 FSDPEngineConfig, 完全绕过了这些包装器, 导致 SFT 场景完全缺失校验, 错误配置下会静默失败或出现难以定位的运行时异常。

实现拆解

1. 在 verl/workers/engine/fsdp/transformer_impl.py 的 FSDPEngine.init() 中, 设置了 self.use_remove_padding 之后立即检查: 若 ulysses_sequence_parallel_size > 1 且 use_remove_padding 未开启, 则抛出 ValueError。
2. 从 verl/workers/config/actor.py 的 FSDPActorConfig.validate() 中删除了之前相同的校验逻辑。
3. 从 verl/workers/config/critic.py 的 FSDPCriticConfig.post_init() 中删除了之前相同的校验逻辑。
4. 所有改动仅涉及校验逻辑的迁移, 无新增依赖或功能变化。

关键文件:

- verl/workers/engine/fsdp/transformer_impl.py (模块 FSDP 引擎; 类别 source; 类型 core-logic; 符号 FSDPEngine.init) : 添加核心校验逻辑, 确保 SFT 和 RL 路径均被覆盖。
- verl/workers/config/actor.py (模块 配置; 类别 source; 类型 core-logic; 符号 FSDPActorConfig.validate) : 删除 FSDPActorConfig.validate() 中重复的校验逻辑。
- verl/workers/config/critic.py (模块 配置; 类别 source; 类型 core-logic; 符号 FSDPCriticConfig.post_init) : 删除 FSDPCriticConfig.post_init() 中重复的校验逻辑。

关键符号: FSDPEngine.init, FSDPActorConfig.validate, FSDPCriticConfig.post_init

关键源码片段

verl/workers/engine/fsdp/transformer_impl.py

添加核心校验逻辑，确保 SFT 和 RL 路径均被覆盖。

```
# verl/workers/engine/fsdp/transformer_impl.py (FSDPEngine.__init__)

self.use_remove_padding = self.model_config.use_remove_padding

# 校验：使用 Ulysses 序列并行时必须启用 use_remove_padding
# 该校验之前仅在 RL 配置类中存在，SFT 路径缺失，导致静默失败
if self.engine_config.ulysses_sequence_parallel_size > 1 and not self.use_remove_padding:
    raise ValueError(
        "When using sequence parallelism (ulysses_sequence_parallel_size > 1), "
        "you must enable `use_remove_padding`."
    )
```

评论区精华

无实质性讨论。仅包含 gemini-code-assist[bot] 的自动摘要评论，wuxibin89 直接批准。

- 暂无高价值评论线程

风险与影响

- 风险：风险较低。校验逻辑从配置层迁移至引擎层，对 RL 路径无影响（之前已覆盖），对 SFT 路径新增了必要校验。潜在风险：若存在某些合法场景需要 `ulysses_sequence_parallel_size > 1` 但 `use_remove_padding=False`（理论上不兼容），现在会被拦截；但根据设计文档，该组合不应存在。
- 影响：影响范围：所有使用 FSDP 引擎的 SFT 训练任务。若之前 SFT 配置中意外设置 `ulysses_sequence_parallel_size > 1` 但未设置 `use_remove_padding=True`，训练将不再静默失败，而是抛出明确的 `ValueError`。RL 训练行为不变。
- 风险标记：缺少测试覆盖

关联脉络

- 暂无明显关联 PR