PR #7577 完整报告

verl-project/verl

[veomni] feat: DeepSeek V4 QAT bf16 fake quant training

合并时间：2026-08-28 10:37

原文链接：http://prhub.com.cn/verl-project/verl/pull/7577

执行摘要

- 一句话：DeepSeek V4 支持 QAT bf16 伪量化训练
- 推荐动作：该 PR 值得关注，因为它引入了新的训练模式（QAT）并增强了 GSPO 损失的灵活性。建议补充单元测试覆盖新配置项和损失聚合逻辑，并确认与 VeOmni 侧 PR 的兼容性。

功能与动机

PR 标题和 body 表明目标是支持 DeepSeek V4 的 QAT bf16 伪量化训练，并关联 VeOmni 侧的 PR#1089。通过引入 qat_implementation 配置项，允许用户选择量化实现（如 fp8_blockwise），同时调整 GSPO 损失聚合方式，使其更灵活。

实现拆解

1. 新增 QAT 配置项：在 verl/workers/config/engine.py 的 VeOmniEngineConfig 中添加 qat_implementation 字段，默认值为 "none"，并同步更新 verl/trainer/config/engine/veomni.yaml 和 verl/trainer/config/_generated_ppo_veomni_trainer.yaml，为 VeOmni 引擎提供量化实现选择。
2. 调整 GSPO 损失聚合方式：修改 verl/trainer/ppo/core_algos.py 中 compute_policy_loss_gspo 函数，将原本硬编码的 loss_agg_mode="seq-mean-token-mean" 改为使用传入的参数 loss_agg_mode，默认仍为 "seq-mean-token-mean"，但支持其他模式。
3. 更新训练示例脚本：在 examples/grpo_trainer/run_deepseek_v4_veomni.sh 中，将模型路径改为 DeepSeek-V4-Flash-Base，设置 qat_implementation=fp8_blockwise，并添加 max_model_len 配置，调整实验名称和响应长度等参数。
4. 无测试变更：本次改动没有对应的测试文件更新，但配置和代码的改动较小，风险可控。

关键文件：

- verl/trainer/ppo/core_algos.py（模块 训练算法；类别 source；类型 core-logic；符号 compute_policy_loss_gspo）：修改 GSPO 策略损失聚合方式，从硬编码改为可配置，影响训练核心逻辑。
- verl/workers/config/engine.py（模块 引擎配置；类别 source；类型 configuration；符号 VeOmniEngineConfig）：新增 qat_implementation 配置项，是 QAT 功能的核心配置入口。
- verl/trainer/config/_generated_ppo_veomni_trainer.yaml（模块 配置；类别 config；类型 configuration）：生成的 VeOmni 训练配置文件中同步增加 qat_implementation 字段。

- verl/trainer/config/engine/veomni.yaml（模块 配置；类别 config；类型 configuration）：VeOmni 引擎配置模板中增加 qat_implementation 字段。
- examples/grpo_trainer/run_deepseek_v4_veomni.sh（模块 示例脚本；类别 other；类型 configuration）：示例脚本演示 QAT 配置，切换到 Base 模型并设置 fp8_blockwise。

关键符号：compute_policy_loss_gspo

关键源码片段

verl/trainer/ppo/core_algos.py

修改 GSPO 策略损失聚合方式，从硬编码改为可配置，影响训练核心逻辑。

```
# verl/trainer/ppo/core_algos.py
def compute_policy_loss_gspo(...):
    # ...
    # 原实现硬编码为 seq-mean-token-mean，现改为支持传入的 loss_agg_mode
    pg_loss = agg_loss(
        loss_mat=pg_losses,
        loss_mask=response_mask,
        loss_agg_mode=loss_agg_mode, # 默认为 "seq-mean-token-mean"
        **config.global_batch_info
    )
    # ...
```

verl/workers/config/engine.py

新增 qat_implementation 配置项，是 QAT 功能的核心配置入口。

```
# verl/workers/config/engine.py
@dataclass
class VeOmniEngineConfig(EngineConfig):
    # ...
    qat_implementation: str = "none" # 新增 QAT 实现选择，默认 none（关闭）
    # ...
```

评论区精华

该 PR 没有 Review 审核和评论，因此没有公开讨论内容。

- 暂无高价值评论线程

风险与影响

- 风险：
  1. 测试缺失：没有为新增的 qat_implementation 配置项和 GSPO 损失聚合方式的改动添加单元测试，存在回归风险。
  2. 配置一致性：qat_implementation 是新增字段，若 VeOmni 侧未同步处理，可能导致配置被忽略或报错，但默认值为 none 保证了向后兼容。

3. GSPO 损失聚合方式变更：修改了 compute_policy_loss_gspo 的默认行为，虽然默认值仍是 seq-mean-token-mean，但调用方必须传入正确的 loss_agg_mode，否则可能改变训练行为，需确认现有调用是否兼容。

- 影响：
  1. 用户影响：支持 DeepSeek V4 的 QAT 训练，用户可通过配置 qat_implementation 启用伪量化，提升训练效率或降低显存占用。
  2. 系统影响：VeOmni 引擎相关的配置结构扩展，需保持与 VeOmni 项目的同步。
  3. 团队影响：为后续 QAT 相关工作提供基础，但需补充测试和文档。 - 风险标记：缺少测试覆盖，核心路径变更，配置一致性风险

关联脉络

- PR #7466 [cfg, megatron, doc] fix: drop unused actor.router_replay in favor of engine config: 都涉及 VeOmni 引擎配置的调整，可能影响配置结构的一致性。
- PR #7536 [cfg] fix: drop unused ref router replay config: 同样涉及 VeOmni 配置清理和生成配置的更新，存在关联。
- PR #7513 [trainer, ckpt, cfg] feat: add config-driven checkpoint callback hook: 都涉及训练配置和生成配置文件的更新，可能共享配置生成流程。