

PR #2030 完整报告

THUDM/slime

[examples]: add qwen3.5-9b model config and fully_async example

合并时间: 2026-06-08 09:36

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2030>

执行摘要

- 一句话: 新增 Qwen3.5-9B 模型配置与全异步示例
- 推荐动作: 该 PR 适合希望快速上手 Qwen3.5-9B 全异步训练的用户参考。对于核心开发者, 可快速阅读以了解模型配置格式, 无深度设计决策需要关注。

功能与动机

9B 参数量的模型在社区中广泛使用, 但 slime 仓库中此前缺少相应的模型配置和启动示例。本 PR 填补了这一空缺, 使开发者能够直接使用 Qwen3.5-9B 进行全异步训练实验。

实现拆解

步骤 1: 创建模型配置文件

在 `scripts/models/qwen3.5-9B.sh` 中定义了 Qwen3.5-9B 的结构参数 (层数、注意力头数、隐藏大小等), 并通过 `slime_plugins.models.qwen3_5` 的 `spec` 函数引入模型实现。特别包含了 `--attention-output-gate` 等 Qwen3.5 特有参数。

步骤 2: 创建全异步运行脚本

`examples/fully_async/run-qwen3.5-9B-fully_async.sh` 整合了模型配置、数据路径、Rollout 参数 (使用 `fully_async` 模式)、模型并行设置 (TP=2)、GRPO 超参以及 SGLang 引擎参数。脚本还包含了节点 NVLink 检测和 Ray 集群启动逻辑, 确保一键运行。

步骤 3: 更新 README

在 `examples/fully_async/README.md` 的文件列表中添加了该示例的一行说明, 方便用户浏览。

关键文件:

- `examples/fully_async/run-qwen3.5-9B-fully_async.sh` (模块 全异步示例; 类别 `infra`; 类型 `example`): 核心运行脚本, 演示如何使用 Qwen3.5-9B 进行全异步 GRPO 训练, 涵盖所有参数设置和环境准备。
- `scripts/models/qwen3.5-9B.sh` (模块 模型配置; 类别 `config`; 类型 `model-config`): 定义了 Qwen3.5-9B 模型的结构参数, 供训练脚本引用。
- `examples/fully_async/README.md` (模块 文档; 类别 `docs`; 类型 `documentation`): 在文档中列出新示例, 方便用户发现。

关键符号：未识别

关键源码片段

examples/fully_async/run-qwen3.5-9B-fully_async.sh

核心运行脚本，演示如何使用 Qwen3.5-9B 进行全异步 GRPO 训练，涵盖所有参数设置和环境准备。

```
# 源自 examples/fully_async/run-qwen3.5-9B-fully_async.sh
# 环境清理和 NVLink 检测已省略

SCRIPT_DIR="$(cd -- "$(dirname -- "${BASH_SOURCE[0]}")" &>/dev/null && pwd)"
source "${SCRIPT_DIR}/../scripts/models/qwen3.5-9B.sh"

MODEL_DIR=${MODEL_DIR:-/root/models/Qwen3.5-9B}
DATA_PATH=${DATA_PATH:-/root/datasets/dapo-math-17k/dapo-math-17k.jsonl}

# Rollout 使用 fully_async 模式，通过指定 rollout-function-path
ROLLOUT_ARGS=(
    --rollout-function-path slime.rollout.fully_async_rollout.generate_rollout_fully_async
    --prompt-data "${DATA_PATH}"
    --input-key prompt
    --label-key label
    --apply-chat-template
    --rollout-shuffle
    --rm-type deepscaler
    --num-rollout 3
    --rollout-batch-size 8
    --n-samples-per-prompt 4
    --rollout-max-response-len 2048
    --rollout-temperature 1
    --global-batch-size 32
    --balance-data
)

# GRPO 超参数，包括 advantage-estimator 和 KL 惩罚等
GRPO_ARGS=(
    --advantage-estimator grpo
    --use-kl-loss
    --kl-loss-coef 0.00
    --kl-loss-type low_var_kl
    --entropy-coef 0.00
    --eps-clip 0.2
    --eps-clip-high 0.28
)

# ... 更多参数（模型并行、优化器、SGLang 等）
```

scripts/models/qwen3.5-9B.sh

定义了 Qwen3.5-9B 模型的结构参数，供训练脚本引用。

```
# 源自 scripts/models/qwen3.5-9B.sh
# 模型规格: Qwen3.5-9B (9B 参数级别)
MODEL_ARGS=(
  --spec "slime_plugins.models.qwen3_5" "get_qwen3_5_spec"
  --disable-bias-linear
  --qk-layernorm
  --group-query-attention
  --num-attention-heads 16
  --num-query-groups 4
  --kv-channels 256
  --num-layers 32
  --hidden-size 4096
  --ffn-hidden-size 12288
  --use-gated-attention
  --normalization RMSNorm
  --apply-layernorm-1p
  --position-embedding-type rope
  --norm-epsilon 1e-6
  --rotary-percent 0.25
  --swiglu
  --untie-embeddings-and-output-weights
  --vocab-size 248320
  --rotary-base 10000000
  # Qwen3.5 特有: 在注意力输出后增加门控
  --attention-output-gate
)
```

评论区精华

本 PR 无 review 讨论。

- 暂无高价值评论线程

风险与影响

- 风险: 风险极低。新增的文件均为示例和配置，不修改任何核心框架代码。潜在风险包括:
 - 用户需要确保模型权重路径存在（默认路径为 /root/models/Qwen3.5-9B），否则脚本会失败。
 - 全异步模式依赖特定的环境配置（如 SGLang 补丁），若环境不一致可能无法运行。
 - 影响: 对用户: 提供开箱即用的 Qwen3.5-9B 全异步训练 demo，降低使用门槛。对系统: 无影响。对团队: 增加维护示例的成本，但贡献有限。
- 风险标记: 低风险，新增示例和配置

关联脉络

- 暂无明显关联 PR