PR #21204 完整报告

sgl-project/sglang

[Diffusion] Revamp Rollout Log-Prob Support with SDE/CPS for RL Post-Training

合并时间：2026-04-09 09:00

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/21204

执行摘要

- 一句话：为扩散模型 RL 后训练新增模块化 Rollout Log-Prob 引擎，支持 SDE/CPS/ODE 策略。
- 推荐动作：建议技术管理者和扩散模型开发者精读此 PR，关注其模块化设计、混合模式集成以及序列并行兼容性的实现细节，为类似功能扩展提供参考。

功能与动机

根据 PR body，RL-based post-training of diffusion models (e.g., FlowGRPO) requires computing per-step log-probabilities along the denoising trajectory。因此，需要为扩散模型添加计算每一步 log-probabilities 的能力，以支持如 FlowGRPO 等强化学习后训练算法。

实现拆解

实现方案包括：1) 新增 post_training 目录，包含 rl_dataclasses.py 定义数据结构、scheduler_rl_mixin.py 实现核心 log-prob 引擎、scheduler_rl_debug_mixin.py 处理调试张量；2) 在 SamplingParams 中添加 rollout 相关参数，并集成验证逻辑；3) 修改 FlowMatchEulerDiscreteScheduler，通过混合模式集成 Rollout 功能；4) 更新 denoising pipeline，添加 _prepare_rollout 和 _collect_rollout_log_probs 生命周期钩子；5) 添加单元测试验证 ODE、SDE、CPS 模式的对齐和正确性。

关键文件：

- python/sglang/multimodal_gen/runtime/post_training/scheduler_rl_mixin.py（模块 diffusion/rl）：核心 Rollout Log-Prob 引擎，实现 SDE/CPS/ODE 采样和 log-prob 计算逻辑
- python/sglang/multimodal_gen/configs/post_training/rl_rollout.py（模块 config）：定义 RL Rollout 参数配置、验证和 CLI 接口
- python/sglang/multimodal_gen/runtime/models/schedulers/scheduling_flow_match_euler_discrete.py（模块 scheduler）：集成 Rollout 到现有调度器 step 方法，确保向后兼容
- python/sglang/multimodal_gen/runtime/pipelines_core/stages/denoising.py（模块 pipeline）：在降噪循环中添加 Rollout 生命周期管理钩子
- python/sglang/multimodal_gen/test/unit/test_scheduler_rollout_unit.py（模块 test）：单元测试验证 Rollout 引擎的正确性和对齐

关键符号：SchedulerRLMixin.flow_sde_sampling, SchedulerRLMixin.prepare_rollout, SchedulerRLMixin.collect_rollout_log_probs, RLRolloutArgs.validate, DenoisingStage._maybe_prepare_rollout

评论区精华

Review 中主要讨论点包括：1) gemini-code-assist[bot] 指出在 scheduler_rl_mixin.py 中潜在除零错误，当 rollout_noise_level 为 0 且 log_prob_no_const 为 False 时；2) mickqian 建议为 RL 参数使用专用解析器以提升代码清晰度，Rockdu 表示同意；3) 关于调度器接口修改，Rockdu 讨论将 batch 参数吸收到 mixin 中以避免侵入式改动；4) 单元测试的注册和清理问题也被提及。

- 除法零错误风险 (correctness): 未在 PR 中解决，建议添加验证防止零噪声级别
- 参数解析器设计 (design): Rockdu 同意，但作为后续优化，当前 PR 未实现
- 调度器接口修改 (design): 最终修改了 step 接口添加 batch 参数，但通过 mixin 管理状态
- 单元测试注册 (testing): 已修复，移除重复注册

风险与影响

- 风险：技术风险包括：1) 除零错误：当 rollout_noise_level 设置为 0 且 rollout_log_prob_no_const 为 False 时，log-prob 计算可能失败，需添加验证；2) 性能影响：新增 log-prob 计算可能增加计算开销，尤其是在调试模式下收集张量；3) 兼容性：修改了调度器 step 接口，添加 batch 参数，可能影响现有代码调用；4) 序列并行集成：需要确保在分布式环境下 log-prob 的归并正确，测试覆盖需充分。
- 影响：对用户影响：为扩散模型 RL 后训练提供了标准化的 log-prob 计算 API，支持多种策略，便于集成强化学习算法；对系统影响：增加了代码复杂性和维护点，但通过模块化设计最小化侵入，确保无 rollout 时无额外开销；对团队影响：引入了混合模式和模块化设计，为后续功能扩展提供参考，但需团队成员熟悉新架构。
- 风险标记：潜在除零错误，接口变更风险，性能开销，序列并行兼容性

关联脉络

- 暂无明显关联 PR