

PR #2118 完整报告

THUDM/slime

sync from internal

合并时间: 2026-06-22 21:05

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2118>

执行摘要

- 一句话: 同步内部代码, 新增 Megatron Teacher Server 及无状态 Adam
- 推荐动作: 建议重点关注 Teacher Server 的架构设计 (尤其是高效采样算法和无状态 Adam), 可为后续类似组件提供参考。对于使用 teacher-forcing 的场景, 此 PR 提供了一套完整的参考实现。

功能与动机

同步内部最新代码库, 使公共仓库对齐内部开发进展。引入 Teacher Server 是为了支持独立于训练流程的预填充和 logprob 计算, 用于 RL 中的 teacher 模型或离线数据生成; 无状态 Adam 满足每步重置动量等特殊优化需求。

实现拆解

步骤 1: 新增 Teacher Server 核心模块在 `slime/backends/megatron_utils/server/` 下创建 `megatron_server.py`, 实现基于 `aiohttp` 的 HTTP 服务器, 包含 Ray remote actor `SampleManager` 管理请求队列与分发, 支持 `/submit/`、`/update` 等端点。步骤 2: 实现高效 TP 分片采样与 logprob 计算在 `logprob_utils.py` 中实现 `sample_from_vocab_parallel_logits_without_full_gather`, 通过两步采样避免全词汇表 `gather`, 降低通信开销; `get_label_token_log_probs_from_vocab_parallel_logits` 用于获取标签 token 的 logprob。步骤 3: 新增 Teacher Server 配置在 `arguments.py` 中定义所有服务器参数, 并通过 `configure_megatron_server_args` 强制设置 teacher-only 模式 (关闭 critic、KL 等), `validate_megatron_server_args` 校验参数合法性。步骤 4: 引入无状态 Adam 优化器在 `stateless_adam.py` 中实现 `StatelessAdam`, 每次 step 重置动量为零, 通过 `bias_correction` 选项适配, 支持 `decoupled_grad`、`master_weights` 等选项。步骤 5: 修改模型层以适配 Teacher Server 在 `model.py` 中添加从外部文件加载 checkpoint 路径的回退逻辑、重新初始化 critic 输出层时使用 `init_method_std`、禁用分布式优化器状态初始化、注入无状态 Adam 等。步骤 6: 增强 trace 监控在 `trace_utils.py` 中重构 `SGLANG_TRACE_META_KEYS` 定义, 引入 PD 阶段细分分段事件, 支持 `_trace_children` 属性实现 span 嵌套记录。步骤 7: 调整 Rollout 与 Ray 组件修改 `slime/rollout/sglang_rollout.py`、`slime/ray/placement_group.py`、`slime/ray/actor_group.py` 以支持新的服务器通信和配置选项。步骤 8: 添加测试 新增 `tests/utils/test_megatron_server_arguments.py`, 覆盖参数添加、配置强制 teacher-only、验证失败场景。

关键文件：

- slime/backends/megatron_utils/server/megatron_server.py (模块 后端；类别 source；类型 core-logic；符号 `_find_project_root`, `_parse_sample_n`, `_normalize_label_token_ids`, `_get_max_request_length`) : Teacher Server 核心实现，包含请求队列、SampleManager、HTTP 端点
- slime/backends/megatron_utils/server/logprob_utils.py (模块 后端；类别 source；类型 core-logic；符号 `sample_from_vocab_parallel_logits_without_full_gather`, `get_label_token_log_probs_from_vocab_parallel_logits`, `_to_cuda_tensors`, `_prepare_rollout_data`) : 实现高效 TP 分片采样与 label token logprob 计算，避免全词汇总表 gather
- slime/backends/megatron_utils/server/arguments.py (模块 配置；类别 source；类型 configuration；符号 `_env_bool`, `_non_negative_int`, `_positive_int`, `_positive_float`) : Teacher Server 全部配置参数定义与验证逻辑
- slime/backends/megatron_utils/stateless_adam.py (模块 优化器；类别 source；类型 core-logic；符号 `StatelessAdam`, `init`, `load_state_dict`, `step`) : 无状态 Adam 优化器，满足每步重置动量的特殊需求
- slime/backends/megatron_utils/model.py (模块 模型；类别 source；类型 core-logic；符号 `get_load_checkpoint_path_by_args`, `_reinitialize_critic_output_layer`, `_noop_init_state_fn`, `_disable_distributed_optimizer_state_initialization`) : 模型层重大重构：checkpoint 加载回退、critic 输出层重新初始化、禁用分布式优化器状态、注入无状态 Adam

关键符号：`sample_from_vocab_parallel_logits_without_full_gather`, `get_label_token_log_probs_from_vocab_parallel_logits`, `SampleManager.submit`, `StatelessAdam.step`, `get_load_checkpoint_path_by_args`, `_patch_megatron_adam`, `_build_sglang_pd_trace_children`, `_append_trace_children`

评论区精华

此 PR 无 review 评论，为内部同步操作，未公开讨论。

- 无 review 讨论 (other): 无结论，已合并。

风险与影响

- 风险：
 - 新引入的 Teacher Server 独立部署，可能增加运维复杂度；若配置不当可能导致端口冲突或超时。
 - 无状态 Adam 改变优化器行为，仅在特定场景（如 teacher 模型无需动量）下使用，若误用可能影响训练收敛。
 - `trace_utils.py` 中 `trace` 结构变更（新增 `span_records`、`start_ts` 等）可能影响已有监控脚本。

- model.py 的 checkpoint 加载逻辑重写 (fallback 路径) 可能与现有 Megatron 版本不兼容。
- 影响：用户影响：提供新的 teacher-forcing 训练路径，用户可通过配置启用独立 Teacher Server；性能优化：TP 分片采样避免全 gather 减少通信开销；系统影响：新增一组独立配置参数和服务器进程，集成需谨慎；团队影响：需要熟悉新模块的运维与调试。
- 风险标记：新增独立 Teacher Server，无状态 Adam 特殊行为，trace 数据结构变更，checkpoint 加载逻辑重写

关联脉络

- PR #2102 Support top_p mask: 同时修改 model.py 和 sglang_rollout.py，同属 Megatron 后端训练优化，共享部分配置与模型改动。
- PR #2088 Add rollout_data_transport nixl: 同时修改 placement_group.py 和 actor_group.py，同属 Ray 后端架构调整，本 PR 进一步扩展了 Ray actor 的配置。