

slime 仓库 2026 第 25 周 (06-15~06-21) 周报

THUDM/slme

周期: 2026-06-15 至 2026-06-21

来源 PR: 17 • 重点 PR: 17 • 自动生成

原文链接: <http://prhub.com.cn/THUDM/slme/reports/2026-06-15-to-2026-06-21>

执行摘要

2026 年第 25 周 (06-15 至 06-21), slime 仓库以 rollout 数据处理链路重构为核心, 配合多项算法增强与性能优化, 共合并 17 个 PR。核心成员 zhuzilin 贡献了大部分变更。

`append_response_tokens` 被统一收敛至 `Sample` 类并加强校验, `top-p mask` 支持强化了采样与训练一致性, 过时的 `BShd` 布局被完全移除。同时新增了 CISPO 损失函数、NIXL 传输管线和 GLM-5.2 模型支持。整体上, 本周有力提升了训练管道的稳健性与扩展性。

本周重点变化

- Rollout 数据处理统一化: PR #2108 与 #2110 将零散在多个 rollout 实现中的响应令牌追加逻辑提取到 `Sample` 类的新方法 `append_response_tokens`, 并强制要求 `log_probs` 传入规则, 新增 `_validate_response_metadata_lengths` 确保长度一致。配合 #2102 的 `top-p mask` 支持, 完全消除了静默数据不一致的风险。
- 数据布局精简: PR #2100 移除所有 `BShd` 相关代码与 `--qkv-format` 参数, Megatron 后端全面采用 THD 格式, 简化了数据加载与损失计算。
- 算法创新: PR #2067 新增 CISPO 优势估计器, 通过 `stop-gradient` 裁剪 IS 权重而非直接裁剪损失, 确保梯度始终流经 `log_probs`, 为 RL 训练提供了新选项。
- 性能优化: PR #2088 增加 NIXL 传输选项加速 rollout 数据传输, 并移除冗余预处理; PR #2082 通过异步初始化使数据加载与引擎启动重叠, 缩短了训练启动时间。
- 模型与新功能: PR #2093 增加 GLM-5.2 744B-A40B 训练支持, PR #2081 同步清理并扩展了 `Sample` 字段 (`apply_chat_template_kwargs`, `custom_rm_path`), PR #2080 移除了废弃的指标地址方法。

模块与主题趋势

改动集中于 rollout 与训练数据管道 (`slime/ray/rollout.py`, `slime/utils/types.py` 等)。标签分布显示 `architecture` 和 `refactor` 占主导, `performance` 与 `feature` 次之。值得一提的是, `test` 标签仅出现 4 次, 核心变更如 `top-p mask` 尚缺伴随测试。作者方面, zhuzilin 提交 12 个 PR, 主导所有架构级改动; EazyReal 贡献了 CISPO 算法; jingshenghang 的 `coding-agent RL` 重构 (PR #2005) 蕴含丰富的 review 讨论, 值得团队研读。

风险观察

1. 核心路径回归风险: PR #2108/#2110/#2102 大幅改动 rollout 数据流, 需补充端到端集成测试覆盖多模态、streaming 等场景。

- 2. 测试缺口：top-p mask (#2102) 和 NIXL (#2088) 等关键特性缺少单元测试。
- 3. 废弃参数迁移：--qkv-format 和 --use-slime-router 已移除，下游训练脚本需更新配置，VLM 示例也需适配。
- 4. CISPO 配置警告：eps_clip < 1.0 时触发警告，用户应调整至 canonical 单边设置（如 --eps-clip 1.0 --eps-clip-high 4.0），避免语义歧义。

重点 PR 速览

PR #	标题	作者	重要性	关键内容
#2108	Extract append_response_tokens to Sample	zhuzilin	9.2	统一 rollout 数据追加方法，所有实现共用 <code>Sample.append_response_tokens</code> 。
#2102	Support top_p mask	zhuzilin	9.1	引入 top-p 候选项掩码，保证采样与训练 token 集一致。
#2100	Remove bshd support	zhuzilin	8.1	彻底移除 BSHD 布局，简化 Megatron 后端。
#2067	Add CISPO advantage estimator	EazyReal	7.2	新增 stop-gradient 裁剪损失函数。
#2088	Add rollout_data_transport nixl	zhuzilin	7.7	基于 NIXL 的高效 rollout 数据传输。
#2093	Support GLM-5.2	zhuzilin	6.3	GLM-5.2 744B-A40B 训练支持。

后续建议

- 1. 补全测试：针对本周重构的核心链路编写统一集成测试，覆盖所有 rollout 变体与新增特性（top-p mask、NIXL）。
- 2. 配置迁移：全面清除现有脚本中的 --qkv-format 和 --use-slime-router，并更新对应文档。
- 3. 算法验证：将 CISPO 应用于实际训练任务，与传统 PPO 进行对比，积累最佳实践。
- 4. 代码审查深化：借鉴 PR #2005 的讨论方式，鼓励关键 PR 内开展深入的设计 Review，提升团队整体代码质量。