

PR #1737 完整报告

THUDM/slime

fix: propagate moe_token_dispatcher_type in bridge model provider

合并时间: 2026-03-22 16:25

原文链接: <http://prhub.com.cn/THUDM/slime/pull/1737>

执行摘要

修复 bridge 模式中因 moe_token_dispatcher_type 参数传播缺失导致的训练失败错误，通过添加兼容性传播提升系统稳定性，影响所有 bridge 模式模型。

功能与动机

此 PR 旨在解决 Issue #1725 描述的 bug: 当使用 bridge 模式时，由于模型提供者未传播 moe_token_dispatcher_type 参数，而 slime 始终设置 variable_seq_lengths=True，导致 Megatron-Core 验证失败，finalize() 抛出 ValueError: Token dispatcher type: allgather does not support variable sequence length。这影响了所有 bridge 模式训练，需传播参数以兼容 Megatron 版本并纠正错误。

实现拆解

在 slime/backends/megatron_utils/model_provider.py 的 wrapped_model_provider 函数中，添加了以下代码：

```
if hasattr(args, "moe_token_dispatcher_type"):
    provider.moe_token_dispatcher_type = args.moe_token_dispatcher_type
```

这确保了 args 中的 moe_token_dispatcher_type 值（在 arguments.py 中已修正为 alltoall）传播到 provider，并使用 hasattr 检查处理属性可能不存在的情况，以兼容较新的 Megatron 版本。变更仅涉及两行添加，逻辑简单直接。

评论区精华

Review 评论为空，无讨论记录。

风险与影响

- 风险：变更风险低，hasattr 检查确保了向后兼容性，但需依赖 args.moe_token_dispatcher_type 已正确设置；无测试覆盖，但逻辑简单，回归可能性小。
- 影响：修复了所有 bridge 模式训练中的 bug，提升系统稳定性；对用户透明，无需额外操作；对团队，减少了错误排查成本并示范了兼容性处理模式。

关联脉络

此 PR 与 Issue #1725 直接关联，修复了其中描述的 bug。从 Issue 可追溯到 PR #1715，后者添加了 `variable_seq_lengths` 传播但遗漏了 `moe_token_dispatcher_type`，导致此问题。整体来看，这是 bridge 模式参数传播机制的完善，反映了在集成 Megatron 时对参数兼容性的持续关注。