

PR #2297 完整报告

THUDM/slime

fix: reject non-positive rollout temperature at parse time

合并时间: 2026-08-21 11:11

原文链接: <http://prhub.com.cn/THUDM/slime/pull/2297>

执行摘要

- 一句话: 参数解析阶段拒绝非正 rollout 温度, 拦截非法 RL 采样配置
- 推荐动作: 值得快速浏览。改动虽小, 但体现了“非法配置在参数解析阶段拒绝, 而不是在训练逻辑里隐式兜底”的工程决策, PR body 对 greedy 温度为什么不是合法 RL 策略的解释清晰; 配合 #2236、#2276 可看到 slime 在参数校验层持续加固的趋势。

功能与动机

PR body 明确指出: `--rollout-temperature 0` is greedy decoding, that is not a usable on-policy RL policy: the sampling distribution is a delta, while log-probs (and PPO ratios) still come from a softmax. Greedy RL should be rejected at config parse, not assumed away in loss code. 即温度 0 下采样分布与概率估计不一致, 会让 PPO/on-policy 更新失真, 应在解析阶段拒绝而不是在 loss 代码里特殊处理。

实现拆解

1. 更新参数帮助文案: slime/utils/arguments.py 的 add_rollout_arguments 中, 将 `--rollout-temperature` 的 help 从“the temperature for the inference engine during rollout.”改为“... Must be > 0.”, 让用户在使用 `--help` 时即可感知约束。
2. 在统一校验入口增加拦截: slime_validate_args 是训练启动时对全量参数做合法性校验的总入口。本次在函数开头 (args.eval_datasets = _resolve_eval_datasets(args) 之后、KL/ref 校验之前) 新增 if args.rollout_temperature <= 0: raise ValueError(...)。该位置意味着只要显式传入非正温度, 启动即失败并给出可读错误, 不会进入后续 rollout 或 loss 计算。
3. 测试配套: tests/test_megatron_argument_validation.py 中先为构造器 make_slime_validate_args 补充 rollout_temperature=1.0 默认值 (避免既有用例因缺少该字段而触发 AttributeError), 再新增参数化测试 test_slime_validate_args_rejects_non_positive_rollout_temperature, 用 0.0 与 -0.1 两个取值验证 ValueError 抛出; 默认温度下的既有用例仍全部通过。

关键文件:

- slime/utils/arguments.py (模块 参数解析; 类别 source; 类型 core-logic; 符号 slime_validate_args, add_rollout_arguments): 统一参数校验入口 slime_validate_args 所在文件, 新增 rollout_temperature <= 0 拦截, 并同步更新 `--rollout-temperature` 的

help 文案，是本次变更的唯一核心逻辑位置。

- tests/test_megatron_argument_validation.py (模块 参数校验; 类别 test; 类型 test-coverage; 符号 make_slime_validate_args, test_slime_validate_args_rejects_non_positive_rollout_temperature) : 为 make_slime_validate_args 补充默认 rollout_temperature=1.0, 新增参数化测试覆盖 0.0 与 -0.1, 保证新校验可回归、既有用例不受影响。

关键符号: slime_validate_args, add_rollout_arguments, make_slime_validate_args, test_slime_validate_args_rejects_non_positive_rollout_temperature

关键源码片段

slime/utils/arguments.py

统一参数校验入口 slime_validate_args 所在文件, 新增 rollout_temperature <= 0 拦截, 并同步更新 --rollout-temperature 的 help 文案, 是本次变更的唯一核心逻辑位置。

```
def slime_validate_args(args):
    args.eval_datasets = _resolve_eval_datasets(args)

    # 温度 <= 0 (含 0) 表示 greedy decoding: 采样分布退化为 delta 分布,
    # 但 log-probs 与 PPO ratio 仍按 softmax 计算, 二者不一致会静默污染
    # on-policy 更新, 因此在这里直接拒绝, 而不是在 loss 代码中兜底。
    if args.rollout_temperature <= 0:
        raise ValueError(
            "--rollout-temperature must be > 0; temperature 0 is greedy decoding "
            "and is not a valid RL policy."
        )

    # 后续的 KL / ref 模型校验逻辑保持不变
    if args.kl_coef != 0 or args.use_kl_loss:
        if not os.path.exists(args.ref_load):
            raise FileNotFoundError(
                f"ref_load {args.ref_load} does not exist, please check the path."
            )
```

tests/test_megatron_argument_validation.py

为 make_slime_validate_args 补充默认 rollout_temperature=1.0, 新增参数化测试覆盖 0.0 与 -0.1, 保证新校验可回归、既有用例不受影响。

```
@pytest.mark.unit
@pytest.mark.parametrize("temperature", [0.0, -0.1])
def test_slime_validate_args_rejects_non_positive_rollout_temperature(monkeypatch,
    temperature):
    module = load_slime_arguments_module(monkeypatch)
    # 测试构造器默认温度为 1.0, 这里通过 overrides 显式传入 0.0 / -0.1,
    # 验证非正温度都会在参数校验阶段被拒绝。
    args = make_slime_validate_args(rollout_temperature=temperature)

    with pytest.raises(ValueError, match="--rollout-temperature must be > 0"):
```

module.slime_validate_args(args)

评论区精华

该 PR 没有任何 review comment 或 discussion thread，变更由单一 commit 直接合入（merged_by: zhuzilin）。设计讨论实际沉淀在 PR body 中：作者论证了“为什么要在解析期拒绝而非在 loss 代码兜底”，以及“为什么 eval 温度 0 仍然合法”，这两点构成本次变更的核心决策依据。

- 暂无高价值评论线程

风险与影响

- 风险：
 1. 启动路径破坏性（有意为之）：slime_validate_args 是所有训练 /rollout 启动必经的校验入口，任何显式传入 --rollout-temperature 0 或负值的既有脚本会在启动阶段立即失败。这是设计意图，但对依赖旧行为的实验脚本属于显式 breaking change。
 2. 覆盖盲区：新增测试只覆盖 slime_validate_args 函数本身，未覆盖 argparse 真实解析路径。由于该校验直接读取 args.rollout_temperature，且该属性在解析阶段由 type=float 保证，风险较低。
 3. 语义隔离：校验仅限 rollout 路径，--eval-temperature 0 保持合法，不会误伤 greedy 评测场景。
 - 影响：对用户：显式配置 --rollout-temperature 0 或负值的命令会得到清晰报错，避免训练若干步后才暴露采样与 log-probs 不一致的问题；默认配置（1.0）与正温度配置完全不受影响。对系统：校验为常数时间，无性能开销，运行时路径零改动。对团队：与 #2236、#2276 延续“启动前参数校验加固”的实践，为后续同类参数约束提供了可复用的测试模式。
 - 风险标记：核心启动路径变更，潜在 breaking change 风险，缺少 argparse 层测试覆盖

关联脉络

- PR #2236 fix: don't overwrite an explicitly set --start-rollout-id: 与本次 PR 共享 slime/utils/arguments.py 和 tests/test_megatron_argument_validation.py，同为启动前参数校验加固。
- PR #2276 Add args check for --save-debug-train-data: 同文件、同测试文件，同为冲突参数在启动阶段拦截，属于同一演进线。