

PR #2741 完整报告

radixark/miles

Add Terminus 2 compaction training example

合并时间: 2026-08-26 03:44

原文链接: <http://prhub.com.cn/radixark/miles/pull/2741>

执行摘要

- 一句话: 新增 Terminus 2 compaction 训练示例
- 推荐动作: 值得精读, 尤其是 `run.py` 中 `_build_train_args` 对 session server v2 与 TITO、GRPO 的串接方式, 以及 README 中对 compaction 心理模型的解释。对需要设计训练配方或复现 RL 实验结果的同学, 这是一份很好的参考模板; 对核心代码维护者, 仅需关注新增测试是否与 `py_harness` 的兼容性。

功能与动机

PR body 指出 Terminus 2 可以总结长 episode 并暴露多条线性轨迹分段, 而 Miles session server v2 保留这些分段、掩码共享策略 token 使其只训练一次, 并将分段归为一次 rollout 供 GRPO 归一化。为了让这一完整设置可复现, 而不需要复制私有主机名、路径或凭据, 作者添加了该示例。默认配方在单机 8 GPU 上运行 100 步, 覆盖 89 行 Terminal-Bench 2 prompt 数据集。

实现拆解

1. 新增启动脚本 `examples/experimental/terminus-compaction/run.py`: 定义继承 `U.ExecuteTrainConfig` 的 `ScriptArgs` 数据类, 集中管理模型路径、rollout 批量、Agent 网络等参数; 提供 `_hf_checkpoint`、`_ref_load`、`_save_dir`、`_save_traces_dir` 等路径推导辅助函数, 以及 `_checkpoint_args`、`_rollout_args`、`_performance_args`、`_grpo_args` 等参数拼装函数, 最终由 `_build_train_args` 生成完整训练命令, 默认启用 `--use-session-server v2`、`--tito-model glm47` 与 `--use-miles-dashboard`。
2. 新增快速测试 `tests/fast/examples/experimental/terminus_compaction/test_run.py`: 通过 `py_harness.import_launch_script` 导入脚本, 用 `shlex` 拆解 `_build_train_args` 生成的 `argv`, 断言 session server v2、TITO 模型、batch 大小、GPU 数、dashboard 与 trace 开关等关键配置; 同时用 `_extra_env_vars` 验证 Harbor 适配器环境变量复用与可选主机覆盖的省略行为。
3. 新增文档 `examples/experimental/terminus-compaction/README.md`: 先说明 compaction 的 mental model (summarization + linear history -> 轨迹树 -> 每叶子一个 Sample), 再给出 Harbor agent server 启动步骤 (依赖 harbor-miles-v0.20.0 分支) 与 `HARBOR_*` 环境变量, 最后覆盖完整数据集转换、dashboard、trace 和验证方法。
4. 配套改动: 无配置、schema 或部署文件改动, 仅新增示例、测试与文档。

关键文件：

- `examples/experimental/terminus-compaction/run.py`（模块 示例脚本；类别 `source`；类型 `core-logic`；符号 `ScriptArgs`, `_hf_checkpoint`, `_ref_load`, `_save_dir`）：核心启动脚本：定义全部训练参数、参数拼装逻辑与 Harbor 环境变量传递，是本 PR 的实际业务逻辑所在。
- `tests/fast/examples/experimental/terminus_compaction/test_run.py`（模块 示例测试；类别 `test`；类型 `test-coverage`；符号 `_value`, `args`, `test_recipe_enables_compaction_aware_session_training`, `test_recipe_records_dashboard_traces_and_honors_gpu_count`）：快速测试：覆盖关键参数拼装、`dashboard/traces` 开关、可选网络覆盖与 Harbor 环境变量继承，防止脚本配置漂移。
- `examples/experimental/terminus-compaction/README.md`（模块 示例文档；类别 `docs`；类型 `documentation`）：文档：说明 `compaction` 心理模型、Harbor agent server 启动步骤、环境变量与验证方式，是示例可复现性的关键支撑。

关键符号：`_hf_checkpoint`, `_ref_load`, `_save_dir`, `_save_traces_dir`, `_checkpoint_args`, `_rollout_args`, `_performance_args`, `_grpo_args`, `_build_train_args`, `_extra_env_vars`

关键源码片段

`examples/experimental/terminus-compaction/run.py`

核心启动脚本：定义全部训练参数、参数拼装逻辑与 Harbor 环境变量传递，是本 PR 的实际业务逻辑所在。

```
# 训练配置的数据类：继承 ExecuteTrainConfig 以复用 Miles 通用启动参数。
# 路径与网络字段均提供默认值，并支持通过环境变量覆盖，
# 便于在不泄露内部主机名和凭据的前提下复现。
@dataclass
class ScriptArgs(U.ExecuteTrainConfig):
    mode: Literal["normal", "debug_rollout_only"] = "normal"
    run_id: str = field(default_factory=U.create_run_id)
    megatron_model_type: str = "glm4.7-flash"
    megatron_path: str = "/root/Megatron-LM"
    num_gpus_per_node: int = 8

    skip_prepare: bool = False
    model_dir: str = "/root/models"
    model_name: str = "GLM-4.7-Flash"
    hf_checkpoint: str = "" # 留空时自动推导为 model_dir/model_name
    ref_load: str = "" # 留空时自动推导为 model_dir/{model_name}_torch_dist
    save_dir: str = "" # 留空时按 output_dir/run_id/checkpoints 生成
    save_traces_dir: str = "" # 留空时按 output_dir/run_id/details 生成
    prompt_data: str = "/root/tb2_train_89.jsonl" # 完整 89 个任务的 TB2 训练集

    max_seq_len: int = 32768
    rollout_max_response_len: int = 8192
    num_rollout: int = 100
```

```

rollout_batch_size: int = 4
n_samples_per_prompt: int = 8
global_batch_size: int = 32
save_interval: int = 100

# 以下字段默认从环境变量读取，未设置时使用示例默认值。
# 这样既能在示例中直接运行，也允许真实环境通过 env 注入。
agent_server_url: str = field(default_factory=lambda: os.environ.get("AGENT_SERVER_URL",
"http://127.0.0.1:11000"))
agent_model_name: str = field(default_factory=lambda: os.environ.get("AGENT_MODEL_
NAME", "model"))
agent_trial_timeout: int = 7200
router_external_host: str = field(default_factory=lambda: os.environ.get("MILES_ROUTER_
EXTERNAL_HOST", ""))
miles_host_ip: str = field(default_factory=lambda: os.environ.get("MILES_HOST_IP", ""))
session_server_ip: str = field(default_factory=lambda: os.environ.get("MILES_SESSION_
SERVER_IP", ""))

use_prometheus: bool = True
prometheus_port: int = 9090

```

路径推导辅助函数：把所有产出路径集中在这里，保证 checkpoints 与
traces 目录始终随 run_id 变化，避免多次运行相互覆盖。

```

def _hf_checkpoint(args: ScriptArgs) -> str:
    return args.hf_checkpoint or str(Path(args.model_dir) / args.model_name)

def _ref_load(args: ScriptArgs) -> str:
    return args.ref_load or str(Path(args.model_dir) / f"{args.model_name}_torch_dist")

def _save_dir(args: ScriptArgs) -> str:
    return args.save_dir or str(Path(args.output_dir) / args.run_id / "checkpoints")

def _save_traces_dir(args: ScriptArgs) -> str:
    return args.save_traces_dir or str(Path(args.output_dir) / args.run_id / "details")

```

tests/fast/examples/experimental/terminus_compaction/test_run.py

快速测试：覆盖关键参数拼装、dashboard/traces 开关、可选网络覆盖与 Harbor 环境变量继承，防止脚本配置漂移。

```

def test_recipe_enables_compaction_aware_session_training(args, monkeypatch):
    # 移除 WANDB key，确保测试不依赖外部环境变量
    monkeypatch.delenv("WANDB_API_KEY", raising=False)
    argv = shlex.split(run._build_train_args(args))

    # 断言 compaction 感知的会话训练配置：

```

```
# 使用 session server v2 保留轨迹分支，并启用 TITO (glm47) 对共享 token 掩码
assert _value(argv, "--use-session-server") == "v2"
assert _value(argv, "--session-server-ip") == "0.0.0.0"
assert _value(argv, "--tito-model") == "glm47"
# 默认配方：每步 4 prompt × 8 rollout，总 batch 32，100 步
assert _value(argv, "--rollout-batch-size") == "4"
assert _value(argv, "--n-samples-per-prompt") == "8"
assert _value(argv, "--global-batch-size") == "32"
assert _value(argv, "--num-rollout") == "100"
assert _value(argv, "--rollout-max-response-len") == "8192"
assert _value(argv, "--prompt-data") == "/root/tb2_train_89.jsonl"
```

评论区精华

PR 没有实质技术讨论。[claude\[bot\]](#) 自动提示该仓库配置为手动 code review，可评论 [@claude review](#) 触发一次性或持续审核；审核者 [Zhichenzzz](#) 直接批准，未留下文字说明或具体修改建议。

- PR 无实质技术讨论 (other): 无设计争议或未解决问题，PR 作为实验性示例被合并。

风险与影响

- 风险：该 PR 仅新增实验性示例与测试，不触及核心训练、rollout 或后端代码，回归风险低。主要风险集中在：1) 示例依赖外部 Harbor 分支 harbor-miles-v0.20.0 及 HARBOR_TERMINUS_2_ENABLE_SUMMARIZE、HARBOR_TERMINUS_2_LINEAR_HISTORY 等环境变量，分支演进或 API 变动可能导致示例失效；2) 脚本默认路径 (/root 系) 和模型名仅适配容器化环境，在其它部署形态下需要手动覆盖；3) 测试只验证了参数拼装与 env 构成，未实际跑通训练，_build_train_args 与 _extra_env_vars 之外的运行时行为（如 session server v2 的轨迹掩码效果）没有自动化保障。
- 影响：对用户而言，获得了一个可直接运行的 Terminus 2 compaction 训练参考配方，不必再自行拼装参数；对系统而言，不影响既有训练链路，仅新增 examples/experimental/terminus-compaction 目录与 fast 测试；对团队而言，随附 README 和测试降低了复现门槛，便于后续以相似结构扩展其它模型或任务集的 compaction 示例。影响范围限定在实验性示例层，影响程度较小但示范价值明确。
- 风险标记：依赖外部 Harbor 分支版本，测试仅覆盖参数拼装，默认路径为容器内 /root

关联脉络

- PR #2710 Log compaction-aware rollout metrics: 同为 compaction 相关功能：该 PR 添加 compaction 感知的 rollout 指标，本 PR 的示例依赖 session server v2 的轨迹分割与掩码语义，二者共同支撑 compaction 训练链路。
- PR #2200 [RL] Add sampling-support log-prob primitives: 为 on-policy 归一化提供 log-prob 原语，是 GRPO 流程的基础设施，本 PR 的 GRPO 配方在该能力之上构建。
- PR #2535 docs: add Mooncake rollout data transfer guide: 同为实验性示例 / 文档类 PR，展示在 examples 与 docs 层面沉淀可复现配置的模式。