

【OpenLLM Talk】番外 02 : OpenRLHF

框架、OpenLLM Talk 和一些想法

hello, 大家好, 我是羡鱼, 好久不见。

本文属于闲聊。先不谈太远的未来, 只谈谈 OpenLLMAI 已经做了什么、接下来一段时间内想做什么和我个人的一些想法和私货。

P0 级的事项

1.OpenRLHF 的第一个正式版本及正式文档

<https://github.com/OpenLLMAI/OpenRLHF>

本项目一开始的目标是实现一个复现 llama2 的基于 Ray 的工业级分布式对齐框架, 其实目前实现的功能已经超出了 llama2 的内容, 应该在目前的开源框架里面处于一个非常领先的位置。当然了, 离真正的工业级框架可能还有不小的距离。

我们目前正在进行第一个正式版本的扫尾和测试工作 (可能不会特别快, 因为年底大家都很忙), 这里先不多说, 仅做预热。敬请期待我们的正式版本, 也欢迎大家参与贡献, **框架要开源才有生命力**, 尤其欢迎性能方面的测试或者算力方面的支持。

如有意愿贡献或者赞助, 可以在 git 上直接进行或者联系我和各个模块的负责人:

- RL : [初七](#)
- Ray : [Joel](#)
- NLP : 羡鱼 (可以通过知乎、git [Xianyu](#) 或者 OpenLLMAI official email xianyuai@openllmai.top 联系我)

Features

- Compatible with HuggingFace Transformers model.
- Distributed [PPO based on Ray](#).
- Support Multiple Reward models.
- Support [Rejection Sampling](#).
- Support [DPO \(direct-preference-optimization\)/IPO/cDPO](#).
- Support [Conditional Alignment](#) (<https://arxiv.org/abs/2308.12050>).
- Support [top chinese models](#).
- Multi-nodes [training scripts](#) for Slurm.
- Support Wandb log (--wandb).
- Support FlashAttention2 (--flash_attn).
- Support [GPT4 evaluation](#) & PPO vs SFT [examples](#)
- Pre-trained 7B/13B llama2 [checkpoints](#)

TODO

- **RLHF compatible with models larger than 100B using vLLM**
- Allows saving and loading training checkpoints.
- Integrates with the QLora.

Features

- Compatible with HuggingFace Transformers model.
- Distributed [PPO based on Ray](#).
- Support Multiple Reward models.
- Support [Rejection Sampling](#).
- Support [DPO \(direct-preference-optimization\)/IPO/cDPO](#).
- Support [Conditional Alignment](https://arxiv.org/abs/2308.12050) (<https://arxiv.org/abs/2308.12050>).
- Support [top chinese models](#).
- Multi-nodes [training scripts](#) for Slurm.
- Support Wandb log (--wandb).
- Support FlashAttention2 (--flash_attn).
- Support [GPT4 evaluation](#) & PPO vs SFT [examples](#)
- Pre-trained 7B/13B llama2 [checkpoints](#)

TODO

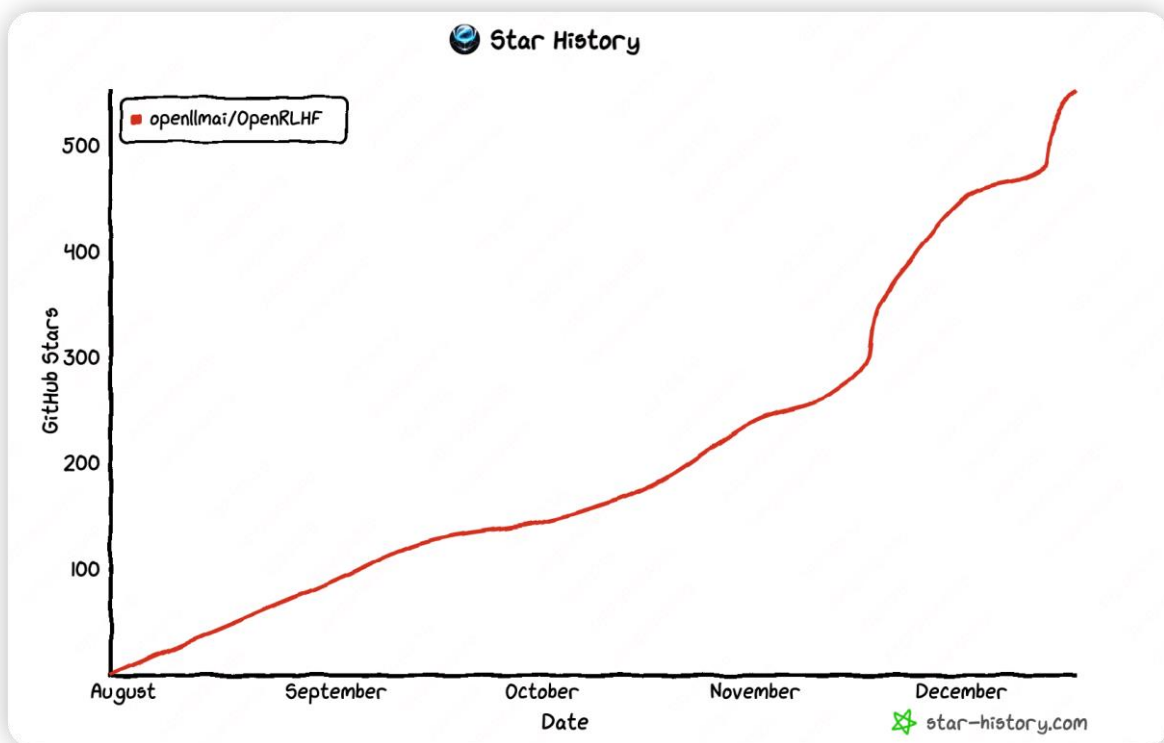
- **RLHF compatible with models larger than 100B using vLLM**
- Allows saving and loading training checkpoints.
- Integrates with the QLora.

Support Matrix

	PPO-max & Best Hyperparameters	Ray	34B Full Tuning with 4 A100	7B Full Tuning with 1 A100 (80G)	7B Full Tuning with 4 RTX4090
OpenRLHF	✓	✓	✓	✓	✓
DeepSpeedChat	✗	✗	✗	✗	✗
ColossalAIChat	✗	✗	✗	✗	✗
TRL	✗	✗	✗	✗	✗

Performance

	7B llama2 RLHF	13B llama2 RLHF (50k samples)
OpenRLHF	-	22 hours with 8 A100
DeepSpeedChat	-	48 hours with 16 A100



Contributors

A big thank you to all our contributors! If you want to contribute, feel free to make a pull request or create an issue.



Citation

```
@misc{hu23openrlhf,  
  author = {Jian Hu and Xibin Wu and Xianyu and Chen Su and Leon Qiu and Daoning Jiang and Qi  
  title = {OpenRLHF: A Ray-based High-performance RLHF framework},  
  year={2023},  
  publisher = {GitHub},  
  journal = {GitHub repository},  
  howpublished = {\url{https://github.com/OpenLLMAI/OpenRLHF}}  
}
```

正式的文档会有中英文版，大概还需要等一等。目前有一些我们组织成员的写的文档或者其他同学的自来水了，可以先看看，希望大家多多支持。

一些预热/旧/自来水 文档：

OpenRLHF 单卡 7B RLHF 全量微调 -- 新增基于 Ray 的 13B/34B RLHF 全量微调 -- RTX

4090 7B 微调 - 蜗牛在花园跑酷的文章 - 知乎

<https://zhuanlan.zhihu.com/p/650758507>

OpenRLHF——新时代的 RLHF 框架 - 王小惟 Weixun 的文章 - 知乎

<https://zhuanlan.zhihu.com/p/672036545>

OpenRLHF 框架详解 - OpenDeployment 的文章 - 知乎

<https://zhuanlan.zhihu.com/p/662933454>

2.重启 OpenLLM Talk，并作为 OpenLLMWiki 的一部分

<https://github.com/catqaq/OpenLLMWiki>

OpenLLM Talk 由于个人工作太忙的原因而暂停了两个月了，近期有不少朋友在问，或许未来一段时间内会重启吧，但同样的，不会太快。

目前将原来的 OpenLLMDoc 项目升级为了 OpenLLMWiki，后续 talk 及 OpenLLMAI 的其他文档、资源都会统一放到里面，也欢迎大家积极参与贡献。

另外，大家不要光顾着催我们，也要多多贡献呀哈哈，这个事情其实非常耗费精力，单靠我和另外几位贡献者很难长久维持（大概取决于我们的工作或者科研状态。

最后，我比较忙，关于加群或者进度之类的我一般只能在晚上统一处理，你可以催，但是请礼貌一点，如果没有及时回，我只能说比较抱歉。不要反复催，可以联系其他的管理员同学或者隔天晚上再催，我不想啰嗦这些的，架不住有些人真的有点儿奇葩。我们是在做开源，不是欠了你的钱！

P1 级的事项

1.在 OpenLLMAI 的项目中正式再启动 1-2 个高优的项目

想做的事情蛮多的，其实有些项目我们内部之前聊过几次了，但是实在太忙了，慢慢来吧。

OpenRLHF Public
A Ray-based High-performance RLHF framework (for large models)
● Python ☆ 553 📄 Apache-2.0 🍴 54 🔄 17 📄 1 Updated 3 hours ago

OpenLLMDE Public
OpenLLMDE: An open source data engineering framework for LLMs
● Python ☆ 12 📄 Apache-2.0 🍴 3 🔄 9 📄 1 Updated on Sep 9

OpenCLLM Public
Chinese LLaMA2 - Enhanced LLaMA2 with PT/SFT/RLHF
☆ 8 🍴 1 🔄 3 📄 0 Updated on Jul 31

OpenSE Public
OpenSE: Open source efficient sentence embedding project for LLMs + Search
☆ 1 📄 Apache-2.0 🍴 0 🔄 0 📄 0 Updated on Jul 9

2. 关于技术分享及讨论

后续也会统一放到 OpenLLMWiki

个人部分，会比较随缘，而且只谈开源的技术，不会涉及任何工作方面的经验：

1. LLM 评测综述：草稿及正式版
2. LLM 幻觉综述：草稿及正式版
3. MOE 综述：简单高效，也许是正确的路，草稿及正式版

专家分享，同样的，比较随缘：

其他主题，可能会邀请相关同学来进行讨论和分享，至于具体的 topic 或者人选，大家可以提名或者报名。

一些想法和私货

1. 我们做开源的虽然很多时候是在用爱发电，但还是希望获得大家的支持的，借用 B 站的话讲，还请大家一键三连！
我们可以白嫖党和伸手党，但请礼貌一点。
2. 关于贡献文档，不要有顾虑（老实说，老问我这种问题，我也没法儿回答），请大胆的走出第一步，技术分享这事儿可能比你想象的要更有意义
3. 关于吊打 G4 的小模型：

作为打工人，我表示理解；作为 NLPer，我表示无奈，不仅国内很多小模型在过拟合榜单，就连 Google 的 gemini 都在文档和视频中搞小动作了，就怎么说呢，混口饭吃还是很难的。

但是，请不要吹的太离谱，骗哥么儿可以，不要连自己都信了。

个人尊重各路榜单，也希望大家可以有自己的判断

4. 关于 MOE：

G4-MOE 的传言，含金量正在上升，至少我个人是比较相信的，所以才有了前面提到的 MOE 综述。这不是个新东西，其实从旧 NLP 时代第一次接触到这玩意儿就感觉是个简单且高效的想法，混合专家，听着真的很合理好吧，什么叫专业的人干专业的事儿，这就叫专业！

广义上来讲，dense model 其实可以看做一种 soft routing 的 MOE model，但实际上没必要每次都都需要所有神经元的参与，路由的粒度也没必要到每个神经元，到一群神经元就行。虽然我们可能事先无法明确定义每一群神经元具体在干什么，但是训练过程中他们自己会学到的，

关于 MOE，先不多说，以后有空再讲。

算了，有空再说，不知不觉又很晚了。