

ChatGLM-Efficient-Tuning

项目总结

[日期]

2023/06/27 -- 2023/06/29

[简介]

(1) 项目背景:

ChatGLM-6B 是一个开源的、支持中英双语的对话语言模型，基于 General Language Model (GLM) 架构，用户可以在消费级的显卡上进行本地部署，ChatGLM-6B 使用了和 ChatGPT 相似的技术，针对中文问答和对话进行了优化。

ChatGLM2-6B 是开源中英双语对话模型 ChatGLM-6B 的二代版本，在保留了初代模型对话流畅、部署门槛较低等众多优秀特性的基础之上引入了新特性。

(2) 项目目标:

1. 让通用模型 chatglm2 不胡编，不提供无依据的法律解释
2. 让 chatglm 熟悉法律条款，明白用户意图，提供专业的解答
3. 使用法律条款及庭审案例训练后，能提供相关索引

[关键成果]

1. 将该微调模型虚拟环境依赖部署成功，调用 GPU 运行 python src/cli_demo.py 和 python src/web_demo.py 脚本实现在命令行和浏览器的成功测试

2. 熟悉该微调模型的数据集接入方式和使用规范，对此板块的代码内容做过完整分析

3. 对后续将引入的法律项目文件 Laws 进行初步修改与调试，将项目虚拟环境依赖添加成功后实现在 scripts 目录下执行 python requessts.py 和 python database.py 脚本后自动处理，将新增法律加入到列表合适的位置以及自动提交并更新 sqlite 的内容

[项目进展]

进展条目	完成情况	问题调挑战
1. 环境搭建	已完成	须注意 cuda 与 torch 版本兼容问题
2. 模型初测	已完成	须调用强大 GPU 测试，要使用虚拟环境

3. 数据接入	进行中	须注意接入格式，思考是否将格式统一为 json
4. 微调训练	未启动	/
5. 指标评估	未启动	/
6. 导出模型	未启动	/
7. 模型引用	未启动	/

[下一步行动]

1. 在新设备上搭建新虚拟环境，调用更强调 GPU 来初测项目，对比分析
2. 参照 dataset_info.json ， config.py 和 Hugging Face 提供的单个数据集的文件格式接入法律项目文件数据集，用作数据准备
3. 根据项目进展依次进行模型训练，指标评估，导出模型，模型引入

[感悟]

在这短短的两三天的项目工作中，我对 ChatGLM2-6B 模型和法律数据集的接入与处理有了初步的了解。通过搭建微调模型的虚拟环境并进行测试，我能够调用 GPU 来提高模型的性能，并在命令行和浏览器中成功测试模型的功能。同时，我也熟悉了微调模型数据集的使用规范，对数据接入和修改进行了实践。这段时间的工作让我深入了解了模型训练和数据处理的重要性，以及如何逐步推进项目的不同阶段。

在未来的工作中，我将继续在新设备上搭建虚拟环境，并利用更强大的 GPU 进行进一步的测试和分析。我将参考相关文档和示例，完善对法律项目文件数据集的接入，并思考如何统一数据格式以提高数据处理的效率。同时，我将启动微调训练、指标评估、模型导出和模型引用等后续工作，以不断提升模型的性能和可用性。

这段时间的项目工作让我意识到在进行复杂任务时，合理规划和分阶段推进工作是至关重要的。同时，与团队成员和相关文档的密切合作也是项目成功的关键。我深感项目工作需要持续的学习和钻研，以不断提升自己的技能和知识储备。我期待未来继续参与并推进这个项目，为公司构建专业的法律问答产品做出贡献。

[参考资料]

描述	链接地址
微调开源项目	https://github.com/hiyouga/ChatGLM-Efficient-Tuning/tree/main
微调思路及原理	https://drive.weixin.qq.com/s?k=ALoAjgdwAA8aQuF3flAKAA8gZYAKo#/preview?fileId=i.1970325096009402.1688852718219936_f.687776344RkEp
法律文件素材	https://github.com/LawRefBook/Laws
微调经验分享	https://www.bilibili.com/video/BV1eL41127uC/?spm_id_from=333.337.search-card.all.click&vd_source=11596aa265fd55c77199ac0f0bb7dd50
官方微调教程	https://www.bilibili.com/video/BV1fd4y1Z7Y5/?vd_source=

	11596aa265fd55c77199ac0f0bb7dd50
ChatGLM 相关产品	https://chatglm.frp.isecsp.com/
微信交流群	LLM 微调应用交流群 2, langchain-chatglm 交流群 37 等

时间: 2023/06/29
记录人: 刘楷