

基于文档驱动自适应编码大模型微调框架

毕业设计答辩

姓名：XXX
学号：XXX
指导教师：XXX

广州大学

2025 年 5 月 26 日

目录

- 1 研究背景
- 2 技术介绍
- 3 系统设计
- 4 实验结果
- 5 总结与展望

- ✔ 研究领域现状：大语言模型在公开数据集与开源项目中表现优异，但在处理企业私有库时存在局限性
- ⚠ 存在的问题：缺乏对私有库专有函数、类及其交互细节的深度理解，无法精准引用库中的类、方法或属性
- 🔧 研究意义：通过深度解析私有库文档，对大语言模型进行微调与优化，提升生成代码的准确性和实用性

前端技术栈

- Gradio - Python 机器学习前端框架

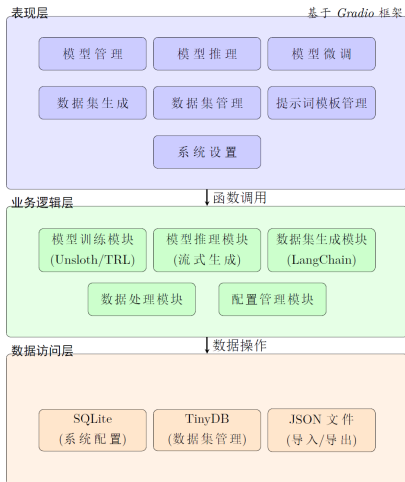
后端技术栈

- Python - 主要开发语言
- SQLite - 轻量级关系型数据库
- TinyDB - 轻量级非关系型数据库
- LangChain - 大语言模型应用框架

人工智能技术

- Qwen2.5 - 开源大语言模型
- Unsloth - 开源微调工具
- QLoRA - 低秩适应微调技术

系统架构



实验条件介绍

硬件配置

- 笔记本型号: Lenovo Legion R7000P 2021H
- CPU: AMD Ryzen 7 5800H
- GPU: NVIDIA GeForce RTX 3060 Laptop GPU (6GB 显存)
- 内存: 16GB DDR4

软件环境

- 操作系统: Ubuntu 22.04 LTS (通过 WSL2 运行)
- Python 版本: 3.11.6
- 深度学习框架: PyTorch 2.1.0+cu121

 微调模型: unsloth 官方 qwen-2.5-3b-4bit 量化模型
(<https://huggingface.co/unsloth/Qwen2.5-3B-Instruct-bnb-4bit>)

 仓库文档: unsloth 官方仓库文档 (<https://docs.unsloth.ai/>)

训练指标变化

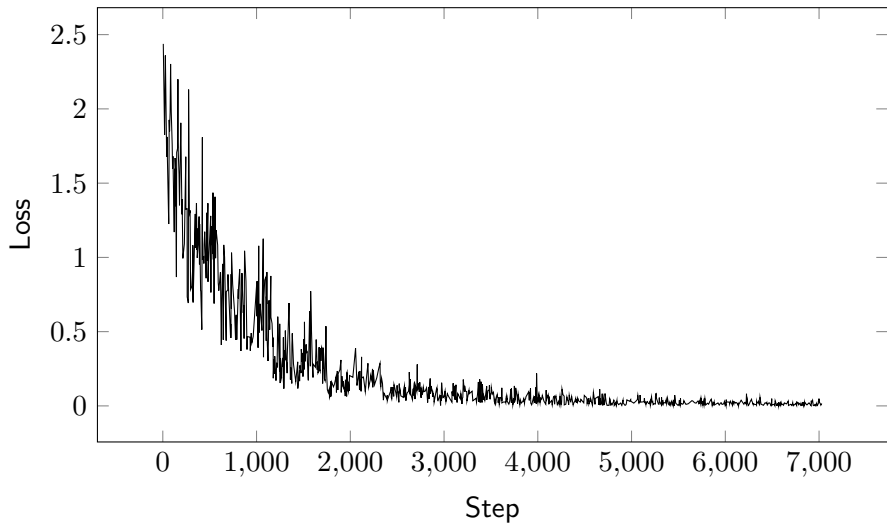


图: 训练损失变化曲线

训练指标变化

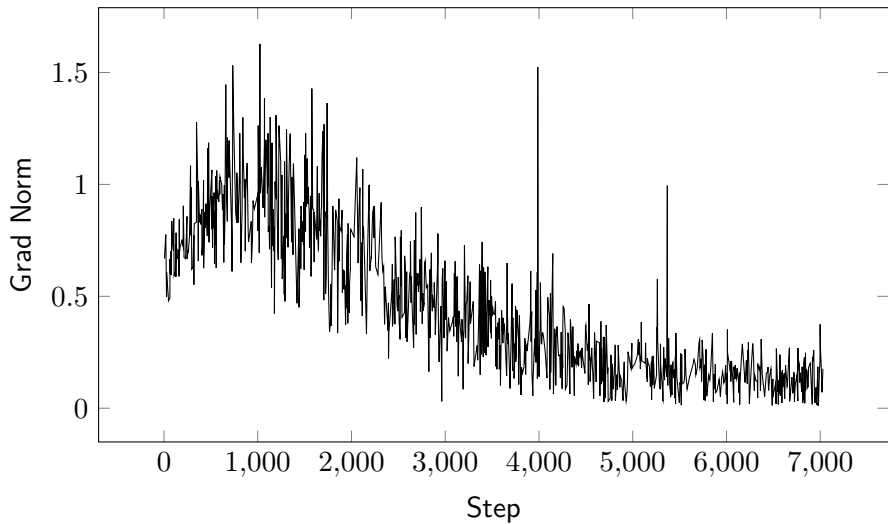


图: 梯度范数变化曲线

训练指标变化

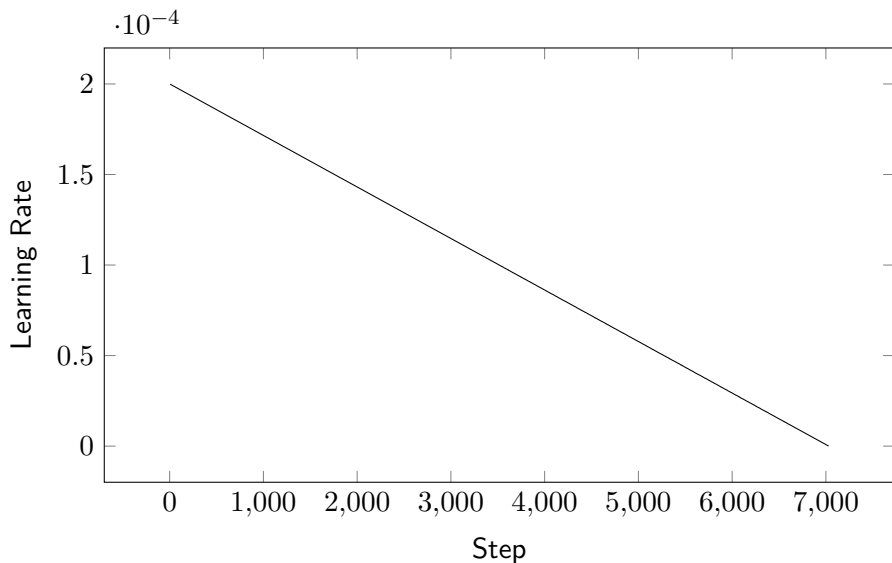


图: 学习率变化曲线

基于文档驱动的自然语言生成大模型微调框架

模型管理

模型训练

模型部署

数据预处理

数据清洗

提示词模板管理

设置

对话

聊天机器人

你知道什么是unsloth框架吗

对不起，我对“unsloth框架”没有听说过，可能这个框架比较小众或者比较新的技术实现方式。如果是在特定领域的应用，可能需要更加具体的信息来给出准确的描述。你可以提供更多的上下文信息让我来尝试回答你的问题。

输入消息

最大生成长度

1024

温度 (Temperature)

0.8

Top-p 采样

0.95

重复惩罚

1.1

清除对话

通过 API 使用 使用 Gradio 界面 设置

基于文档驱动的自然语言生成大模型微调框架

模型管理

模型训练

模型部署

数据预处理

数据清洗

提示词模板管理

设置

对话

聊天机器人

你知道什么是unsloth框架吗

是的，unsloth是一个用于微调LLM的Colab笔记本，它提供了在Google Colab环境中运行和保存微调模型的教程。

输入消息

最大生成长度

1024

温度 (Temperature)

0.8

Top-p 采样

0.95

重复惩罚

1.1

清除对话

通过 API 使用 使用 Gradio 界面 设置

图: 训练前后对比

☰ 主要工作总结

- 提出了文档驱动的自适应编码大模型微调框架
- 实现了三层架构设计（表现层、业务逻辑层和数据访问层）
- 实现了大语言模型的微调

💡 创新点总结

- 开发了基于栈结构的 Markdown 文档解析器
- 采用了 QLoRA 参数高效微调方法
- 构建了基于提示工程的数据集生成工具

🔮 未来工作展望

- 边缘计算部署
- 更多的格式输入支持
- CICD 集成

🐙 软件仓库地址：<https://gitea.carry.fit/carry/gzhu-biyesheji>

感谢聆听!

欢迎各位老师提出宝贵意见