

邓晓飞

男 | 汉族

18794227203 | dengxfabc@163.com

开发技能

1. 熟悉常见的机器学习及深度学习模型和方法,比如逻辑回归,神经网络CNN,RNN,LSTM,GAN
2. 熟悉机器学习,深度学习开发框架的使用,包括pytorch,tensorflow,XGBoost,sklearn
3. 熟悉LLM模型LLaMA,BLOOM,Langchain,ChatGLM预训练,SFT,RLHF等阶段,及分布式训练
4. 熟悉微调模型,自监督学习,注意力机制,Transformer
5. 熟悉常见的NLP研究和开发,包括文本处理,文本分析等,了解prompt工程原理,基于场景的微调数据集构造生产
6. 熟练使用python,java语言及开发框架,Django,FastApi,Springboot和开发工具,了解Linux系统常用命令和shell脚本编写
7. 熟悉常用数据库MySQL,PostgreSQL,Doris,StarRocks及dbeaver工具的使用
8. 熟悉知识图谱的构建流程,熟悉信息抽取、融合和加工处理
9. 熟悉使用python和py2neo库对neo4j数据库操作,数据的增删改查
- 10.了解大数据开发平台EMR集群,Maxcomputer,dataworks,了解使用HiveSql数仓开发和数据提取
- 11.了解使用docker容器化项目部署

项目经历

项目一 亚马逊电商平台在线评论分析

开发工具: python+pycharm

相关技术: Bert+Transformer+pytorch+pandas+nlTK+sklearn+DL

项目描述: 电商评论是重要的在线消费活动。分析在线评论质量,可以有效提高用户体验和消费决策倾向。

责任描述: 负责亚马逊在线评论数据清洗和整理,确保数据的可用性和数据质量;负责评论分析模型的设计,构建,训练,和评估,确保评论分析算法的性能,负责评论推荐方法的设计,训练及评估。

技术描述:

1. 使用LSTM与GANs构建在线评论分析核心、通过Transformer、BERT等实现文本到编码的转换
2. 基于AE和GANs构建生成式匹配推荐模型
3. 使用注意力机制融合文本数据,增强数据特征表示
4. 使用常见指标precesion,recall,F1等对模型进行评估
5. 使用可视化方法对模型评估指标进行可视化记录
6. 项目过程整理,形成文本记录

项目二 基于LLM的中英文多轮对话系统

开发工具: python+xshell+pycharm+AutoDL

相关技术: Transformer+pytorch+pandas+qlora+SFT+RLHF

项目描述: Firefly是一个开源的大模型训练项目,支持对主流的大模型进行预训练和指令微调。中英文多轮对话系统以复旦大学MOSS团队开源的中英文多轮对话数据集和baichuan-7B微调的大语言模型。

责任描述: 负责数据处理和模型微调;

技术描述:

1. 使用数据处理工具将MOSS数据为项目所需的数据格式
2. 使用QLoRa4位量化、双重量化
3. 使用SFT监督微调,训练模型领域认知
4. 使用RLHF微调优化模型
5. 使用并行多卡训练,优化计算资源消耗

项目三 医疗知识图谱构建

开发工具: python+Neo4j+pycharm

相关技术: py2neo+MongoDB

项目描述: 本项目立足医疗领域,以垂直型医药网站为数据来源,以疾病为核心,构建起一个包含7类规模为4.4万的知识实体,11类规模约为30万实体关系的知识图谱

责任描述: 基于垂直网站数据的医疗知识图谱构建,基于医疗知识图谱的自动问答

技术描述:

1. 使用Neo4j图数据库存储数据
2. 使用Cypher查询语句作为问答搜索SQL,支持问答服务
3. 基于结构化数据设计知识schema,构建医疗知识图谱
4. 使用KBQA基于问题模板回答用户用自然语言提出的问题

自我评价

1. 工作认真负责,积极主动,能吃苦耐劳,工作脚踏实地,有较强的责任心
2. 上进心强,勤于学习能不断提高自身能力与综合素质,具有较强的独立工作能力
3. 具有较强的适应能力,纪律性强,意志坚定,具有较强的无私奉献精神
4. 具有良好的身体素质,健康的体魄和乐观的精神

教育背景

2016.09 到 2020.06	西北师范大学	物联网工程	本科
2021.09 到 2024.06	西北师范大学	智能计算与软件	硕士