

基本信息

姓名: 李康华
生日: 1994-10-09

学历: 本科
手机: 13308786435

性别: 男
邮箱: 906695844@qq.com

相关工作经历

- 深兰科技有限公司 (2019年5月 - 2023年4月) NLP 算法工程师
- 新晨科技有限公司 (2023年4月 - 2024年8月) 大模型算法工程师
- 上海锐天投资管理有限公司 (2024年8月 - 2025年3月) 大模型研发工程师

项目经历

一、理财经理智能助手 (RAG) (2023.07-2024.03)

项目背景: 构建金融知识库, 为理财经理提供快速、精准的信息检索和数据分析, 同时也支持结构化查询, 提升理财经理工作效率。

项目职责:

- 结构化查询: 通过规则、槽位识别和实体对齐等传统 NLP 方法, 构建实体-意图分类树。将自然语言转化为SQL语句, 实现结构化查询。训练二分类模型来决定query是否进入结构化查询。
- query 重写: 使用多轮、least to most、step back 等改写方式对4000条数据重新改写, 微调chatglm3-6b, 解决口语化、多轮表达、复杂问题拆解、知识映射等问题。知识召回率较baseline提高18%。(baseline召回率68%)
- 知识增强: 利用大模型根据知识文档生成问答对, 并对其评估与分类, 增强生成问答对的可靠性与实用性, 提高细致问题与概括问题的回答准确性。
- 召回: 采用query-文档的embedding召回、query-answer的embedding 召回、query-文档的字符检索和 quey-query的字符检索等4路召回, 增加知识召回的多样性, 召回率比单一文档向量召回率提升9%。
- 重排: 通过对历史对话数据的badcase分析、大模型增广等方法, 挖掘难负例, 构建 3000 余条正负样例对(正负比例1:10)。微调重排模型, 知识命中率较微调前提升16%。
- 模型部署: 采用 fastapi+gunicorn+nginx+supervisor 部署, 平均首响时间为2.8s。

二、银行智能客服机器人 (CPT+SFT+DPO) (2024.02-2024.08)

项目背景：构建问答机器人，为客户提供业务办理、理财产品、贷款、反诈和其他金融知识的咨询服务，提升客户服务体验与满意度。

项目职责：

- 继续预训练：通过JS逆向爬虫、第三方渠道购买等收集专业书籍、财经新闻、各类金融网站和其他通用数据（共34B），采用规则、minhash等方法对数据清洗与去重（23B），在qwen1.5-1.8b上使用小批量数据做混比实验；基于lora微调qwen1.5-14b，测评达标率较base-model提升9.1%。
- 监督微调：人工标注单、多轮种子数据5000余条，大模型增广4万余条，人工筛选后混合开源通用数据 6.5万条，基于继续预训练模型全参微调。人工测评3.6 分。（满分 5 分）
- DPO：收集金融伦理道德与法律法规和幻觉偏好数据2万对，训练DPO模型，减少模型幻觉和不安全回复，不安全回复率由8.6%降低至2.4%。
- 模型测评：利用自动测试、人工测试的组合，对训练结果进行全方位评估，根据结果反馈进行相应训练数据的调整。
- 模型部署：采用docker+vllm+nginx部署,平均首响时间为1.2s。

三、投资决策智能辅助系统 (Agent) (2024.11-2025.02)

项目背景：构建Multi-Agent投资决策系统，帮助投研人员提供实时的市场分析、投资建议和风险评估。

项目职责：

- 数据采集：集成各大平台网络爬虫工具，实时获取市场数据、新闻资讯和社交媒体情绪。
- 智能体设计：构建个股测评、市场分析、风险评估等10余类agent。采用多智能体协作机制，确保各智能体之间的信息共享和决策一致性。
- 性能调优：调整prompt，全自动完成多步分析，实现复杂任务的全流程实现。

四、搭建企业知识图谱 (传统 NLP) (2022.07-2022.11)

项目背景：从新闻文本中抽取公司1、公司2、公司之间的关系等三要素存入知识图谱。

项目职责：

- 数据预处理：从mysql数据库中提取新闻文本、公司名数据，并清洗。
- 构建实体库：设计公司名简称提取算法，尽可能多去匹配公司名，并做好对应关系，解决实体统一的问题。

- 数据标注：采用 AC 自动机标注新闻文本中的公司名实体，结合 SentenceTransformer、jieba词性分析、构建停用词表来解决实体消歧，设计实体序列排序算法，解决实体嵌套问题，再加以人工校验。
- NER 模型构建：采用 BERT+BiLSTM 进行实体识别，设计实体替换等动态加载训练策略。模型在 人工标注的测试集上 F1 值为 97%。
- 关系抽取：采用 BERT+Casrel 的 Pipeline 模型抽取出控股、供应商、子公司、竞争、客户等关系。训练时对 loss 与标签加权，减小标签不均匀的影响。模型在人工标注的测试集上 F1 值为 86%，比 Joint Model 高出 13%。将公司实体与关系导入到 neo4j 图数据库。

专业技能

- 熟练掌握预训练、SFT、Lora、RLHF、RAG、Agent等大模型任务。
- 熟练掌握文本分类、命名实体识别、关系抽取等传统NLP任务。
- 熟练掌握python、javascript等编程语言。
- 熟练使用pytorch、deepspeed、langchain、swarm、autogen、ollama等框架。
- 熟悉fastapi、flask、gunicorn、docker、nginx等后端部署框架与技术。熟练操作linux。
- 了解GraphRAG、Dify、MCP等技术。
- 熟练掌握Pgsql、Elasticsearch、Milvus、Faiss、MongoDB、Neo4j、Redis等数据库的使用。
- 熟悉JS逆向等复杂场景爬虫技术。

教育经历

云南师范大学	金融学	2013.09-2017.07
--------	-----	-----------------

自我评价

扎实的代码工程能力，丰富的内网开发与部署的经验。极强的学习能力，抗压能力强。对数据敏感，善于思考，具有发现问题并解决的能力。