



杨旭

电话: 18428365776 | 邮箱: 1290682268@qq.com | 微信号: 18428365776 | 工作年限: 4 年

教育背景

上海交通大学	凝聚态物理	硕士	2017.09 - 2021.07
四川大学	物理学类	本科	2013.09 - 2017.06

工作经历

上海蜜度科技股份有限公司 (LLM 算法工程师)	2023.05 - 至今
上汽集团人工智能实验室 (知识图谱高级工程师)	2021.10 - 2023.04
北京大学信息技术高等研究院 (NLP 算法实习生)	2021.03 - 2021.06

项目经历

RAG 相关2024.04 - 至今

- 参与政府某知识问答项目，负责搭建整个知识问答项目的 pipeline 和后续的迭代优化工作。该项目的重难点在于表格类型数据较多，问答涉及跨表格问答。针对此，在文档处理（细粒度文档解析）、检索（chunk 增强、混合检索+精排）以及模型侧（构造带引用的 sft 语料，微调大模型）分别做了优化工作，最终在评测集上问答准确率超过 90%，该问答系统的效果得到客户认可，项目已正式上线运行。
- 负责公司 RAG 系统中表格问答基础能力的提升工作。先后在文档解析侧（加入版面分析+表格提取模型，提高对表格解析能力）、模型侧（参与构建表格问答效果评测集、参与构建 sft 训练语料（QA 对），模型训练）等做了优化工作。从效果上来看，整体问答准确率提升 10 个百分点以上。
- 其它：参与某英文文档问答项目 pipeline 搭建和迭代优化工作。特别地对该问答项目中出现的跨表格和跨多文档问答等传统 RAG 的难点问题，在多个方面进行过优化工作（如对 query 进行实体抽取，优化大表格拆分方式、复杂表格转文本描述、优化检索策略等），并取得了较为明显的效果提升。

某终结报告书智能写作项目2024.01 - 2024.04

- 该项目主要任务是利用大模型，基于案件笔录、案件调查记录等多种信息撰写案件调查终结报告。该项目的重难点在于文档数量众多且命名不规范、要求撰写的报告内容具有相当的专业性。针对此，我们规范了文档命名规则，并基于此过滤无效文档，对重要类型的文档建立基于规则的清洗流程；模型侧，利用了 prompt 工程、few-shot learning 等技术，最终实现高效、专业的案件报告生成，该撰写效果得到客户认可，并已正式上线试运行。

- 该项目主要任务是利用大模型基于大量的市民工单文本生成热点分析报告。该分析报告的主要内容包括热点主题以及该主题下的摘要总结。在项目中，主要利用了文本聚类、指令微调（微调后的大模型可以生成专业的主题以及摘要，从而满足专业化场景需求）等技术，实现了全自动、高效的分析报告生成。该项目目前已上线使用。

其它工作

- 大模型微调相关：单文档、多文档的标题、摘要生成、文本分类任务
- 知识图谱相关：先后参与了医疗领域知识图谱、故障诊断等知识图谱的构建工作、以及上层应用的算法开发工作，主要工作内容包括：参与前期知识图谱构建的数据清洗与整理工作、训练关系抽取模型；应用方面，基于 noisy-or 贝叶斯模型实现了智能辅助问诊功能、基于知识图谱的故障诊断算法开发。

其它

- **技能**：熟悉 python 语言，掌握 mysql、mongodb、redis 等主流数据库的使用方式，具有良好的编程能力；熟悉 pytorch 等常用深度学习框架；熟悉大模型的常用训练方法、调优及模型部署（vllm、lmdeploy 等）；
- **语言**：英语六级，具备良好英文文献阅读能力；
- **个人评价**：具有良好的沟通能力和逻辑思维能力，自驱力强，工作态度认真踏实。