

张琰

18734299161 (微信同号) | lz2912721@gmail.com | 算法工程师 (2年) | 2000.12



个人优势

- 熟练使用Python及NumPy、Pandas等库，高效完成数据处理任务。
- 熟悉掌握KNN、线性回归、逻辑回归、SVM、决策树等机器学习算法。
- 熟悉NLP相关算法TF-IDF、CNN、RNN、Transformer 和 BERT。
- 熟悉大模型相关算法及应用：Lora、混合精度、VLLM、RAG。

工作经历

山西华翔集团

算法工程师

2023年7月-2025年6月

- 负责数据的收集与处理。
- 负责机械制造故障诊断与处理系统的开发与优化，专注于提升设备维修效率和降低停机时间。
- 参与系统需求分析、设计和开发，确保系统功能满足企业实际需求。
- 维护和优化系统性能，确保系统的稳定运行和高效响应

项目经历

基于RAG的设备故障维修指导助手

2024年7月-2025年6月

项目背景：企业面临设备复杂度增加、熟练技师短缺、维修知识分散及故障诊断效率低等问题，急需一套智能系统提升设备维护效率和维修成功率。

项目功能：故障诊断-方案推荐，构建“故障症状输入-原因分析-维修方案生成”智能决策链，并融合设备安全操作规则，生成解决方案。

项目流程：

- 通过需求收集数据：**明确功能需求，收集设备手册、维修记录，历史故障库数据。
- 文档解析与分块：**使用marker-pdf工具+ Qwen2-VL-7B解析pdf文档，采用‘基本块-子块’的分割方式。
- 数据预处理与标准化：**清洗数据，建立标准话术表。
- 知识库与向量数据库构建：**利用Elasticsearch构建知识库与向量数据库，并通过bge-large-zh-v1.5嵌入模型实现元数据的高效关联与检索。
- 检索与生成：**设计混合检索策略并进行结果重排，以及reranker模型的对比选择。
- 模型开发：**通过精心设计的提示词工程，利用DeepSeek-R1-14B模型生成高质量的回复。
- 本地部署与测试：**用Vllm在本地部署模型，并大模型大模型进行性能评估。

项目成果：

- 提升诊断效率：**平均故障诊断时间缩短至5分钟以内。

- 提高诊断准确率：**故障根因分析准确率达95%。
- 增强检索能力：**知识库检索召回率达94%，精确率达96%。
- 降低维修成本：**维修人工成本降低35%。
- 知识库建设：**构建覆盖20000+故障案例的结构化知识库，每年新增3000+案例。

设备故障维修领域命名实体识别

2024年11月-2024年12月

项目背景：现有RAG系统在处理非结构化或半结构化文本时，面临对文本中关键实体识别和理解局限性的问题，需要引入NER技术解决这些问题。

项目功能：自动识别并从文本中提取关键信息，包括设备部件、系统、型号、故障现象与症状、维修操作与步骤，以及原因与解决方案。

项目流程：

- 数据准备与标注：**明确需识别的实体类型，收集并人工标注大量维修文本数据。
- 模型开发与训练：**选择BERT预训练模型并进行领域适应性预训练，随后基于标注数据进行全量微调训练，优化模型性能。
- 评估与优化：**采用F1-Score评价指标来评估模型。
- 部署与应用：**将训练好的NER模型部署为API服务，集成到业务流程中。

项目成果：项目通过自动、准确地识别和提取文本中的关键实体，成功解决了现有RAG系统在处理数据时的痛点。与项目初期未引入NER技术前相比，这使得RAG系统在知识抽取召回率上提升了20%，并在问答精确度上提升了15%。

自我评价

- RAG系统构建与优化：**熟悉RAG系统从文档处理到大型语言模型应用的完整流程，能够进行多模态文档解析、向量数据库构建，并设计有效的检索策略。
- NER技术应用经验：**具备NER模型的开发和部署经验，包括数据准备、模型训练，能解决文本中的实体识别挑战。
- 技术融合能力：**能够将命名实体识别（NER）等技术融入RAG系统，以提升信息抽取效率，从而改进系统整体性能。
- 解决实际工程问题：**致力于将AI技术应用于实际业务场景，解决具体痛点，在系统落地和工程化方面积累了一定经验。
- 注重团队协作与沟通：**在项目中积极与团队成员沟通协作，确保技术方案有效落地，并共同推动项目按计划完成。

教育背景