

盛星达

性别：男

电话：15645103211

工作经验：7 年

邮箱：306878447@qq.com

教育经历

对外经济贸易大学 | 统招本科 2014.09-2018.06

相关技能

- 大模型工程化**：熟悉 LangChain 框架，熟悉 coze dify 平台的开发和应用，具备从大模型预训练到蒸馏部署的全链路经验
- 数学和统计学基础**：高等数学：理解微积分、线性代数、矩阵运算等概念。概率论与数理统计：熟悉概率分布、假设检验、贝叶斯定理等。优化理论：了解梯度下降、随机梯度下降、牛顿法等优化算法；
- 编程技能**：熟练掌握 Python、C，PyTorch 框架；
- 机器学习专业知识**：理解各种机器学习算法的原理和应用，如有监督学习（如线性回归、逻辑回归、决策树、随机森林、神经网络等）、无监督学习（如聚类、降维等）。熟悉模型评估方法，如交叉验证、混淆矩阵、准确率、召回率、F1 分数等。了解超参数调优技术；
- 深度学习专业知识**：熟悉各种深度神经网络架构，如 CNN、RNN、LSTM、BERT、T5、Transformer 等；
- 数据预处理和特征工程**：能够有效地清洗和预处理数据，包括缺失值处理、异常值处理、数据标准化或归一化等。熟悉特征选择和特征降维技术，以提高模型的性能和减少过拟合；
- 问题解决和沟通能力**：能够独立地分析和解决复杂的技术问题。能够与团队成员、产品经理和其他利益相关者有效地沟通，解释技术解决方案和限制；
- 持续学习和研究能力**：愿意不断学习和掌握新的技术和方法，以应对快速发展的机器学习领域。能够阅读和理解最新的学术论文和开源项目，以拓展自己的视野和知识；

工作经历

胥瓷信息科技有限公司 | 大模型应用算法工程师 2024.12-至今

大模型算法开发与利用、产品配套组件的开发部署、模型评估与验证、数据预处理与特征工程

- 政府及国企行业大模型融合落地、大模型组件开发与调优
- 主导政务与国企领域大模型应用场景设计，完成从技术方案到落地的全流程实施，推动大模型在公文处理、决策支持等场景的深度应用
- 开发企业级大模型核心组件库，优化模型推理效率 30%，支持高并发政务场景需求
- 构建私有化模型微调框架，结合传统 CV/NLP 模型实现多模态任务性能提升 15%-20%
- 设计政务知识图谱动态更新机制，通过向量化引擎实现亿级数据实时检索响应时间 $\leq 200\text{ms}$
- 牵头智能辅审系统开发，整合造价模型与规则引擎，实现工程审计效率提升 40%

盟识（上海）科技有限公司 | 算法工程师 2020.7-2024.9

算法研究与开发、数据处理与分析、模型评估与验证、数据预处理与特征工程

- 研究和评估各种机器学习算法，包括监督学习、非监督学习、强化学习等。开发新的机器学习算法或改进现有算法的性能。实现和测试算法在模拟环境或实际数据集上的性能。
- 清洗和预处理原始数据，使其适合机器学习模型的训练和测试。进行数据探索性分析和可视化，以发现数据中的模式和趋势。构建特征工程流程，提取和选择对模型性能影响最大的特征。
- 使用适当的机器学习框架（PyTorch）来训练模型。调整模型的超参数以优化性能，如学习率、批量大小、网络结构等。使用交叉验证等技术来进行超参数调优过程。
- 设计和实施模型评估策略，包括准确率、召回率、F1 分数、AUC 等指标。使用测试集或验证集来评估模型的性能，并确保模型没有过度拟合训练数据。进行模型比较和选择，以确定最佳模型或模型组合。
- 跟踪机器学习领域的最新研究成果和技术趋势。参加学术会议、研讨会和在线课程，不断提升自己的专业技能和知识水平。

北京燃数科技有限公司 | 算法工程师 2018.09-2021.02

需求分析、爬虫设计：爬虫开发：反爬虫：数据清洗和存储：性能优化：数据清洗和整合：

- 与业务团队、产品经理或数据科学家沟通，了解他们需要从哪些网站或 API 获取数据。
- 分析和理解目标网站的结构、规则、反爬虫机制等。
- 设计爬虫的总体架构，包括爬取策略、数据存储方案、错误处理等。选择或开发合适的爬虫框架（Scrapy、BeautifulSoup、Selenium 等）。
- 编写代码实现爬虫的各个模块，如 URL 管理器、网页下载器、网页解析器、数据输出器等。
- 使用 HTTP 请求库（requests、urllib 等）发送请求，获取网页内容。解析网页内容，提取所需数据（使用 XPath、CSS 选择器、正则表达式等）。
- 研究和应对目标网站的反爬虫策略，如 IP 封锁、验证码、动态加载内容等。
- 使用代理 IP、设置请求头、模拟用户行为等方式绕过反爬虫机制。
- 对提取到的原始数据进行清洗、去重、格式化等处理。将处理后的数据存储到数据库（MySQL、MongoDB、Redis 等）或文件系统中。

项目经历

2024.12-至今

政务大模型融合应用平台

项目背景：针对政府公文处理效率低、多部门协作复杂等问题，开发基于大模型的智能政务平台，整合文档解析、知识库管理、智能审批等核心功能。

项目内容：针对政府公文处理效率低、多部门协作复杂等问题，开发基于大模型的智能政务平台，整合文档解析、知识库管理、智能审批等核心功能。

项目成果：落地 3 个省级政务平台，日均处理公文超千份，人工审核工作量减少 60%；知识库检索响应速度提升 50%，关

联推荐准确率提升至 80%；数字人系统通过工信部认证，意图识别准确率达 93.2%，入选 2025 年智慧城市示范项目

2024.12-至今

基建工程造价大模型

项目背景：为国企基建项目开发智能造价评估系统，解决传统造价模型依赖人工经验、调整周期长等问题。

技术方案：大模型基座融合 Transformer 与 GNN 架构，构建支持工程量清单自动分解的多任务模型；开发基于 LoRA 的轻量化微调方案，模型迭代周期从 2 周缩短至 3 天；设计可视化决策看板，集成成本预测、异常检测、趋势分析模块

项目成果：在某高铁项目中实现造价估算误差率 $\leq 2.5\%$ （行业平均 5%-8%）；通过知识图谱关联 100+国家标准规范，自动合规检查覆盖率提升至 90%

2023.10-2024.09

大模型微调以及提示词工程和 RAG 搭建

项目背景：大模型（如 GPT、BERT 等）的出现，使得机器能够在文本生成、问答系统和信息检索等任务中表现出色。然而，现有的大模型往往面临着计算资源消耗大、训练周期长以及适应性差等挑战。为了解决这些问题，我们启动了自研大模型项目，结合提示词工程（Prompt Engineering）与 RAG（Retrieval-Augmented Generation）技术，旨在提升模型的实用性和灵活性。

项目内容：

- 自研大模型：**
 - 模型架构：基于 Transformer 架构设计，采用自回归和自编码技术相结合，构建具有强大生成能力的语言模型。
 - 数据收集与预处理：从多种来源（如互联网文本、专业文献、社交媒体等）收集大规模文本数据，进行数据清洗和标注，确保数据的多样性和质量。
 - 模型训练：采用分布式训练策略，提高训练效率，并通过模型压缩技术（如剪枝、量化等）降低模型的计算资源需求。
- 提示词工程（Prompt Engineering）：**
 - 提示词设计：根据不同应用场景，设计高效的提示词以引导模型生成更精准的输出。例如，在问答任务中设计特定问题格式，以提高模型的回答质量。
 - 自适应提示：通过用户反馈和在线学习机制，实时调整提示词，提高模型的适应性和准确性。
- RAG（Retrieval-Augmented Generation）：**
 - 信息检索模块：结合信息检索技术，构建检索模块，根据用户输入实时检索相关文档或知识库，以增强生成内容的准确性和丰富性。
 - 生成模块：在检索到的信息基础上，利用自研大模型生成自然语言响应，实现信息与生成的有机结合。

项目成果：

性能提升：

自研大模型在文本生成和问答任务上表现优越，较现有模型在多个基准数据集上提高了 10-20% 的准确率。

通过提示词工程和 RAG 技术，生成的内容更加相关且具有上下文连贯性，显著提升了用户体验。

应用场景扩展：

成功应用于客户服务、智能助手、内容生成等多个领域，帮助企业提升了工作效率和服务质量。

模型的灵活性和可扩展性使其能够快速适应不同业务场景的需求。

用户反馈与迭代：

根据用户反馈，不断优化提示词和模型表现，形成闭环改进机制，使模型在实际应用中不断迭代和进步。

2023.12-2024.6

围绕 LangChain 开发对话型 AI Agent

项目背景：企业对高效、智能的对话系统需求日益增加。对话型 AI Agent 能够通过自然语言与用户进行互动，提供即时反馈和解决方案，极大提升用户体验与满意度。LangChain 作为一种强大的框架，能够简化对话系统的构建过程，并支持多种功能的集成，成为开发对话型 AI Agent 的理想选择。

项目负责的内容：

- **LangChain 框架分析与设计：**
 - 框架优势：深入分析 LangChain 的功能，了解其在对话管理、上下文处理和多轮对话等方面的强大能力。
 - 系统架构设计：设计基于 LangChain 的对话型 AI Agent 系统架构，包括模块划分、数据流和交互方式。；
- **对话型 AI Agent 功能开发：**
 - 自然语言理解（NLU）模块：利用 LangChain 提供的模型，开发能够理解用户意图和提取关键信息的 NLU 模块。
 - 对话管理模块：实现对话状态跟踪和上下文管理，确保 AI Agent 能够根据历史对话内容做出智能响应。
 - 多模态交互：支持文本、语音等多种输入方式，使用户能够通过不同的交互方式与 AI Agent 进行沟通。
- **智能响应生成：**
 - 文本生成模型集成：将现有的文本生成模型（如 GPT 系列）与 LangChain 结合，生成自然流畅的对话响应。
 - 个性化推荐：根据用户的历史行为和偏好，AI Agent 可以提供个性化的建议和服务。
- **系统测试与优化：**
 - 用户测试：开展用户体验测试，收集反馈并根据用户需求不断优化 AI Agent 的对话能力和响应质量。
 - 性能优化：在不同场景下测试系统性能，优化响应速度和处理能力，以确保 AI Agent 在高并发情况下仍能稳定运行。
- **部署与应用：**
 - 云端与本地部署选项：根据客户需求，提供灵活的部署方案，确保系统在不同环境下的兼容性。
 - 集成现有系统：与企业现有的客服系统、知识库等进行集成，提升系统的整体效能。
 - 通过需求分析，环境搭建，爬虫设计，爬虫开发，反反爬虫策略，测试与调试，部署与监控，数据可视化与导出等一系列的流程实现产品落地；

项目成果：

- 用户交互效率提升：对话型 AI Agent 在多个测试环境中运行后，用户交互效率提升了 30%，大幅减少了用户查询问题的时间。
- 客户满意度提高：用户反馈调查显示，客户对 AI Agent 的满意度达到 85%，其中 70% 的用户认为 AI Agent 能有效解决他们的问题。
- 多场景应用成功：AI Agent 在客服、医疗咨询和教育培训等多个领域成功部署，显示出良好的适应性与灵活性。
- 数据处理与分析能力增强：通过与企业数据系统的集成，AI Agent 能够实时分析用户数据，为决策提供支持，帮助企业制定更精准的市场策略。该系统能够定时爬取目标电商网站上的产品信息，并将数据存储到数据库中供后续分析和使用；

2023.6-2023.12

信用卡欺诈检测模型

项目背景：银行和金融机构面临着巨大的欺诈风险，需要借助先进的技术手段来识别和预防欺诈行为。构建一个信用卡欺诈检测模型，以提高欺诈检测的准确性和效率，降低误报率和漏报率；

项目内容：

数据收集与预处理：

- 收集信用卡交易数据，包括交易金额、交易时间、交易地点、商户类型等信息。
- 对数据进行清洗和预处理，包括处理缺失值、异常值、重复数据等。将数据进行标签化，区分欺诈交易和非欺诈交易；

特征工程：

- 提取与欺诈行为相关的特征，如交易金额、交易频率、交易地点变化等。
- 对特征进行转换和缩放，以满足机器学习模型的要求。使用特征选择技术，筛选出对模型性能影响最大的特征。

模型选择与训练：

- 选择适合欺诈检测的机器学习算法，逻辑回归、随机森林、SVM。
- 使用训练数据集对模型进行训练，调整模型的超参数以优化性能。
- 使用验证数据集对模型进行评估，比较不同模型的性能并选择最佳模型。

模型评估与优化：

- 用测试数据集对最终选定的模型进行评估，计算准确率、召回率、F1 分数等指标。
- 分析模型的错误案例，识别可能的改进方向。
- 使用集成学习等技术对模型进行优化，提高性能。

模型部署与监控：

- 将训练好的模型部署到生产环境，实现实时欺诈检测。
- 构建监控系统，对模型的性能进行实时监控和告警。
- 定期收集新的交易数据，对模型进行再训练和更新，以适应新的欺诈模式。

项目成果：

- 成功构建了一个高准确率的信用卡欺诈检测模型，有效降低了误报率和漏报率。
- 通过实时欺诈检测，银行能够及时发现并阻止欺诈行为，降低了欺诈损失。
- 模型提高了交易审批的自动化程度，提高了客户体验。

2021.8-2022.10

皮革纹理识别

项目描述：与皮革厂进行合作，对三千份皮革纹理进行拍照，针对用户给出的皮革纹理进行相似度前五匹配。针对现实中购买的需求以及不断迭代的皮革纹理的标记需求设计一款落地的产品。

项目负责的内容：

需求分析：

- 与项目团队（包括产品经理、设计师、皮革专家）紧密合作，深入理解项目的目标和需求。
- 包括了解各种皮革纹理的特点、分类方式、应用场景等，以及明确系统需要识别的纹理类型、识别精度等要求。
- 参与前期数据收集和训练标准的设定，调整拍照焦距清晰度等一系列指标以达到模型训练与实际落地的情况保持一致。

算法实现：

- 选择 VGG 以及 GoogleNet，实现皮革纹理识别的核心算法。
- 包括图像预处理、特征提取、分类器训练等步骤。
- 关注算法的性能和效率，确保系统能够满足实时性和准确性的要求。
- 通过一系列数据增强来提高模型的准确性。

项目成果：

- 成功开发了一套基于深度学习的皮革纹理识别算法。
- 该算法通过 GoogleNet 对皮革纹理图像进行特征提取和分类，能够准确识别出不同的皮革纹理类型。
- 算法在大量标注的皮革纹理图像数据集上进行了训练和验证，取得了优异的识别准确率和泛化能力。
- 针对皮革纹理的复杂性和多样性，采用了多种模型优化策略，如数据增强、迁移学习、模型融合等，进一步提高了模型的识别性能和鲁棒性。

- 通过对不同纹理类型的识别结果进行分析和评估，不断优化和调整模型参数，确保模型在各种场景下都能保持稳定的性能。
- 将皮革纹理识别系统应用于皮革制品的质量控制和分类管理中，实现了对大量皮革制品的快速、准确识别和分类。

2021.12-2022.10

医疗命名实体识别

项目背景：在医疗领域，处理和分析大量的病历文本数据对于疾病诊断、治疗方案制定和药物研发等至关重要。然而，这些文本数据往往包含大量的医学术语和专有名词，给信息的自动提取和分析带来了挑战。因此开发一个高效、准确的医疗命名实体识别系统，以支持医疗领域的文本挖掘和知识图谱构建。

项目内容：

需求分析：

- 与医疗专家、产品经理或项目团队紧密合作，深入理解医疗命名实体识别的具体需求。
- 这包括明确需要识别的医疗实体类型（如疾病、药物、症状、治疗方法等），以及识别的准确性和效率要求。

数据收集与预处理：

- 负责收集相关的医疗文本数据，并进行预处理工作。
- 包括文本清洗、去重、分词、词性标注等步骤，以确保数据的质量和一致性。

算法设计与实现：选择 BERT，并对其训练和优化，以提高识别的准确性和效率。

模型评估与调优：

- 在算法实现后，对模型进行评估，以验证其性能和效果。
- 通过选择合适的评估指标（准确率、召回率、F1 值等），并使用测试集对模型进行测试。根据评估结果，对模型进行调优，以提高其性能。

文档编写与培训：在项目完成后，编写相关的技术文档和用户手册，为医疗专家或用户提供必要的培训和支持。

项目成果：

- 开发了一个基于深度学习的医疗命名实体识别模型，该模型能够准确识别病历文本中的医学命名实体，包括疾病名称、药物名称、手术名称、解剖部位等。
- 在标准的医疗命名实体识别数据集上进行测试，模型的准确率、召回率和 F1 分数均达到了业界领先水平。
- 将医疗命名实体识别系统应用于实际的病历文本数据分析中，成功从大量病历文本中提取出了丰富的医学知识。
- 通过与医疗知识图谱的结合，实现了对病历文本中实体关系的自动抽取和展示，为医生提供了更加直观、便捷的知识查询和分析工具。