

宇中强

13622084038, zhongqiang_yu@sina.com
34 岁, 男, 天津, 求职目标: 机器学习、后端开发



教育背景

2014.09-2017.07	河北工业大学	仪器科学与技术 硕士
2009.09-2013.07	河北工业大学	测控技术与仪器 本科

优势亮点

- 熟悉深度学习及其相关神经网络，包括 CNN、RNN、Transformer 等。
- 熟悉大模型相关训练方法，包括 DeepSpeed、lora 等。相关大模型，包括 llama、Qwen、MiniCPM、Phi3 等；使用 xtuner、unsloth 进行模型微调；RAG 检索、Agent 智能体等相关内容。
- 了解语音、文本、图像方向常用框架，对大模型、多模态和元学习的 paper 算法有一定的跟进学习。熟悉 Pytorch、Tensorflow2.x 框架及端侧部署 ONNX、TensorRT 等。fastText、SentenceBert。
- 了解机器学习分布式训练，主从结构参数服务器的数据通信、GRPC。了解强化学习相关算法 PPO，对奖励函数有一定尝试经验。
- 熟悉 Python 和 JavaSE，了解常用的 Python 后端框架，Linux Shell 使用和 Docker 使用。
- 对算法调优有一定经验。
- 熟悉机器学习相关算法及原理。
- 了解基础算法和数据结构。

自我评价

- 工作背景：7 年+人工智能项目工程师，任算法岗工程师，做大模型相关任务。累计完成 10+个项目，曾从 0 至 1 主导完成语音距离检测、传感器阵列电子鼻和供热优化等项目的需求分析、解决方案制定和工程实现。语音距离检测准确率达 94%，传感器阵列电子鼻问题准确率达 90%以上。为这些项目在公司其他产品的开展奠定基础。
- 工程实现：人工智能工程编程语言使用 Python，JavaSE 语言等。
- 综合能力：拥有较强的自我管理意识，对待工作认真负责，同时具有较强的思维逻辑及问题分析能力、解决能力，不断学习行业知识。

工作经历

2023.11-2024.10	软通动力信息技术有限公司	北京
人工智能工程师 部门: 搜索推荐 汇报上级: 部门负责人 2023.11-至今		
工作概述: 负责驻场团队搜索推荐部门做一些大数据处理工作和检索类工作。内容有，对大数据的采样方法进行调整，使用不同时间段数据的对比实验，以提高模型的推荐效果。对图像进行审美标准评分模型。使用 RAG 方法对错误的新闻文章进行检测。		
2022.09-2023.08	天津亚虎科技有限公司	天津
人工智能工程师 部门: 算法部 汇报上级: 部门负责人 2022.09-2023.08		
工作概述: 负责亚虎科技检验报告单 OCR 识别算法的选择、代码编写以及部署工作。使用 PaddleOCR 算法分别在服务端和移动端进行测试和部署。		
2021.12-2022.09	国防创新研究院	天津
人工智能工程师 部门: 算法部 汇报上级: 部门负责人 2021.12-2022.09		
工作概述: 负责国防创新技术研究院，强化学习小组的强化学习算法。任职期间参与完成强化学习相关的 pytorch 代码编写、半精度实验、强化学习算法更新，分布式强化学习工程实现。其中，主要使用 python 代码，完成强化学习仿真实验的验证。		
2020.05-2021.12	中广智诚有限公司	天津

人工智能工程师 | 部门: 算法部 | 汇报上级: 部门负责人 2021.05-2021.12

工作概述: 负责中广智诚算法部开源算法落地。任职期间完成的项目有台标的目标检测、语音分离项目算法实现及 fastapi 后端代码。由于项目需进行快速上线, 主要使用开源代码进行迁移学习并完成上线。其中, 目标检测是用甲方提供的图片根据 yolov5 进行算法迁移, 语音分离使用 spleeter 开源模型。最后把算法融入 fastapi 框架。

2019.08-2020.05 德尔塔科技有限公司-天津分公司 天津

自然语言处理工程师 | 部门: 算法研发部 | 汇报上级: 部门负责人 2019.08-2020.05

工作概述: 负责公司自然语言处理相关的数据分析、算法设计、结果分析和内容重构等工作, 保证项目及时交付, 避免项目扣款。

- **算法整合:** 负责算法部门数据处理, 保证算法正常运行。根据客户不同需求和场景, 搭建产品模型, 选择不同算法进行工程化, 比较结果。累计完成 4 个+产品模型的搭建。目的是提升 AI 准确率。
- **内容重构:** 根据公司之前的产品代码框架, 在新框架下进行算法二次重构。在一个月时间完成 10 个+原有自然语言处理代码重构。
- **解决方案供给:** 根据系统运行情况, 契合客户真实需求, 配合测试同事, 一起进行为客户提供专业的解决业务方案, 包括压力测试、性能测试等, 确保客户工作顺利开展, 提升客户满意度。

2017.08-2019.08 汉威电子股份有限公司 郑州

机器学习算法工程师 | 部门: 研发中心 | 汇报上级: 部门领导 2017.08-2019.08

工作概述: 公司专门做气体传感器。为扩大传感器销量, 制定传感器阵列项目, 即, 使用多种传感器对混合气体识别, 方法是机器学习。机器学习需要了解项目, 了解数据本身规律。所以本项从数据采集、数据预处理、算法选择、工程实现大部分有本人完成。

- **需求调研:** 向资深气体传感器同事请教, 传感器相关原理、特性, 从而对数据采集、算法选择提供了前提。
- **数据搜集:** 请教实验员, 根据不同气体, 不同浓度, 使用集气箱进行气体采集。
- **项目管理:** 在了解需求和传感器信息之后, 选择相关的机器学习方法进行试验。在和传感器专家的沟通之后, 制定从单一传感器-单一气体到传感器阵列-混合气体从简入繁的进行试验。

项目经历

2023.11-至今 软通动力信息技术有限公司 大数据算法工程师

项目描述: 辅助对推荐算法的数据进行采集(数据集采样、数据集划分处理), 图像审美/美学检测等任务。实现和优化产品 RAG 开发, 结合检索模块和生成模块, 提高模型的泛化能力和准确性。定期对模型进行评估和调优, 确保模型的稳定性和可靠性。使用 xtuner 微调大模型。利用 LangChain 搭建 Agent 智能体。

项目职责:

- ◇ 工程化实现

2022.09-2023.08 天津亚虎科技有限公司 深度学习算法工程师

项目描述: 负责算法的选定, 模型测试、部署方案。对验报告单的内容进行 OCR 识别。使用 paddleOCR 进行算法推理, 测试在服务端(Nvidia T4)和移动端(Android)模型的区别, 以及在不适用版面分析的情况下, 使识别的内容尽可能达到识别需求。

项目职责:

- ◇ 工程化实现

项目业绩: 服务端 Python 代码每张报告单推理时常 0.6s, 基本满足使用要求。

2021.12-2022.09 国防创新研究院 深度学习算法工程师

项目描述: 负责创新研究院强化学习算法及其相关的任务。期间, 把强化学习 tensorflow 代码转为 pytorch, 使用半精度进行训练。代码环境放到 Nvidia Jetson, 并可以实现推理。使用近期论文里的优化点对原有强化学习算法更新。使用主从参数服务器进行分布式训练。

项目职责:

- ◇ 算法更新和工程化实现

项目业绩: 对强化学习算法进行更新, 放到 Nvidia 开发版推理和分布式训练。

2020.05-2021.12

中广智诚有限公司

深度学习算法工程师

项目描述: 负责中广智诚算法部开源算法落地。任职期间完成的项目有台标的目标检测、语音分离项目算法实现及 fastapi 框架后端代码。由于项目需进行快速上线, 主要使用开源代码进行迁移学习并完成上线。其中, 目标检测是用甲方提供的图片根据 yolov5 进行算法迁移, 语音分离使用 spleeter 开源模型。最后把算法融入 fastapi 框架。

项目职责:

- ✧ **项目算法:** yolov5 和 spleeter 开源模型

项目业绩: yolov5 可以正式投入生产, spleeter 应用实际数据集效果比官方预训练模型好。

2019.08-2020.05

自然语言处理平台

自然语言处理工程师

项目描述: toB 的项目。基于旧德尔塔自然语言处理项目进行重构改造, 同时在平台原有算法基础上, 新增了命名体识别、文本摘要等项目。

项目职责:

- ✧ **项目重构:** 在其他同事协助下, 由本人完成。使用 java 语言对旧架构内的算法和代码逻辑进行移植, 目的是原有架构算法结果和新架构算法结果一致。
- ✧ **项目算法:** 命名体识别: 在新框架下, 使用传统基于字典方法进行命名体识别算法, 在任职期间, 使用基于深度学习模型进行算法改进。文本摘要: 使用抽取式摘要 (textRank 算法) 对 1000 字内的中文段落进行摘要。
- ✧ **进度把控:** 通过项目规划, 有效规避项目延期风险, 确保按时保质上线。

项目业绩: 重构的项目按时交付, 避免项目延期扣款, 同时确保客户工作顺利开展。

2018.11-2019.08

语音距离检测

智能语音工程师

项目描述: 该项目跟政法大学合作, 基于校方模拟法庭语音转文字项目其中子项目。法庭记录从笔录改为自动语音转文字系统, 但当两人同时说话, 转录系统会将两人话语整合在一个句子中, 无法阅读。解决方案为通过麦克风和说话人距离进行语音区分, 再转录。由此该算法根据距麦克风不同距离的声音特征提取进行区分。算法为论文复现。

项目职责:

- ✧ **数据分析:** 找到 10 个男/女同事, 根据 0.5 米为界限, 收录两组语音数据。进行语音数据预处理, 提取特征值, 以备算法使用。
- ✧ **项目算法:** 将上述提取的特征值, 进行响度、峰度等六维特征进行变换, 以六维特征值放入支持向量机 svm 进行二分类。通过网格搜索进行算法参数选取, 对训练集/测试集进行交叉验证, 验证最后的算法效果。
- ✧ **算法效果:** 算法最终结果, 二分类的分类精度为 94%。用到语音转文字项目中, 取得了良好的效果。

项目业绩: 该项目为模拟法庭语音转录系统的成功奠定了基础。在转录系统中, 可以把两人同时说话的转录问题解决了, 能做到会其意, 但流畅度有待提升。

2018.11-2019.08

唤醒词检测(KWS)

智能语音工程师

项目描述: 语音唤醒词项目类似 “Hey Siri” 功能。属于语音方向检测的子项目。使用 librispeech 开源语料库进行训练, 在使用 kald 进行语音对齐之后, 使用三层 lstm 进行语音特征提取, 使用最后一层 lstm 作为特征标识模板。然后, 需要用户说三遍唤醒词进行再训练, 得到最终唤醒词向量模板。实际使用时, 用户说出唤醒词, 程序根据算法计算出唤醒词向量, 与模板向量进行余弦距离比对, 根据相似性进行结果判断。算法为论文复现。

项目职责:

- ✧ **项目算法:** 使用深度学习 tensorflow、keras 框架, sklearn 第三方库进行工程化实现, 开源语料库进行特征提取。
- ✧ **算法效果:** 算法实际使用时, 尝试若干英文词汇和中文词汇作为唤醒词进行识别, 英文识别正确性在 70%以上, 中文识别正确性在 60%以上。

项目业绩: 将唤醒词作为语音方向检测的子模块参加竞赛。

2017.08-2018.11

电子鼻阵列

机器学习工程师

项目描述: 不同气体传感器对不同气体的敏感程度不同, 筛选三种传感器制成电子鼻阵列, 对混合气体种类进行识别区分。

项目职责:

- ✧ **项目规划:** 在传感器专家指导下, 从 0 至 1 主导完成电子鼻的项目落地。包括, 数据采集, 数据处理, 算法建模, 结果显示等环节。
- ✧ **数据采集:** 从在集气箱内单一传感器通入单一气体, 到集气箱内电子鼻阵列通入混合气体, 采集 600 多组不同传感器, 不同气体的数据。
- ✧ **项目算法:** 由于是分类模型, 使用逻辑回归作为主要算法, 从线性拟合、协方差等方法入手, 研究传感器对气体敏感程度。由于传感器本身温漂、时漂的问题, 数据会有一定不稳定情况, 导致混合气体分类精度在 70%左右。

右。

项目业绩：在职期间，电子鼻产品试用阶段通过，在公司内部作为新产品展示。