

周戩

求职意向：AI 算法工程师

热爱算法，喜欢理论加实践，坚信 AI+Biotechnology 造福人类

27 岁 | 18475487268 | 浙江杭州 | 4 年经验 | beautifulsumday@gmail.com



工作经历

杭州诺辉健康科技有限公司

算法工程师

2022.06--至今

项目名：NGS 肝癌早筛产品研发 (IVD) **算法负责人**

项目主要负责部分：产品阳性判断值的确定、算法模型的搭建、算法云平台建立。

主要工作内容：根据 miRNA 和蛋白数据，建立数据分析流程及机器学习模型，筛选 miRNA 并确定模型，最终实现确定灵敏度和特异性双 90 以上的模型。

其他工作内容：算法云平台的搭建。通过 python web 框架配合 IT 搭建算法云平台、实现可视化操作算法分析数据。

项目主要设计到的技术点：特征工程、机器学习算法（主要随机森林、逻辑回归、xgboost、SVM 等）、flask、docker 等

深圳华大基因科技服务有限公司

AI 算法工程师

2019.03--2022.04

项目名：遗传疾病的自动化诊断 (LDT) **算法负责人**

项目主要负责部分：对 wes/wgs 产品，建立自动化疾病诊断过程。

主要工作内容：通过自然语言处理算法，根据医学文献和公有数据，自动化寻找疾病诊断过程中的证据，结合临床特征和变异位点信息，最终判断患者的遗传性疾病。

变异位点知识库的建立：整合 NCBI、OMIM 等平台上的基因组学数据、蛋白组学数据和表型数据等，设计数据库结构，建立公有知识库，得到变异位点的特征信息：人群频率、遗传模式、功能危害性（基因产物有害影响）等。

文献知识库的建立：通过实体提取和关系提取，从医学文献中提取出基因、表型和疾病之间的关联关系，根据 ACMG 指南，建立文献知识库。

变异位点数据预处理：根据公有知识库和文献知识库，对变异位点进行特征查找，并处理噪声数据。

算法的设计：通过变异位点特征，结合 ACMG 规则，并利用机器学习算法 xgboost 对不同的特征进行综合分类，最终得到变异位点的致病性判定。

微服务的搭建：利用 docker，python 的 web 框架，搭建应用服务。

项目名：基因排序推荐

算法负责人

项目内容简介：通过机器学习算法，根据病人的表型信息及变异位点信息，推荐出最有可能的致病基因。

数据收集：调查和收集生物行业的共有数据库，包括人类和各种生物的表型、基因组学数据、蛋白组学数据，疾病等数据库。

数据的清洗及对齐合并：对各数据库数据进行标准化，合并各数据库。最终提取出表型与基因关联数据库、疾病与关联数据库、基因与基因关联数据库。

算法的设计：利用 tf-idf 算法对数据库进行计算得到表型与基因关联稀疏矩阵和基因与基因关联稀疏矩阵，然后利用 svd 对矩阵解析。

算法微调：根据稀有的 1500 例标注样本，利用 word2vec 的负采样算法对矩阵进行微调，最终得到表型与基因之间

的关联关系。

项目结果：top20 的准确率达到了 91%。

项目名：临床特征的抽取及标准化 **算法负责人**

项目内容简介：通过改进原有的机器学习算法，从病人的临床特征文本中提取出临床特征词条，并标准化到 1.8w 种标准表型词条，提取算法有完全匹配、模糊匹配和历史匹配。

数据预处理：对临床特征文本进行去停用词等处理，对人类表型词条进行 1-gram、2-gram 处理。

自研算法词条提取：根据机器翻译自动评估算法 bleu 思想进行改进，计算临床特征文本中某个片段与人类标准表型词条的相似性，利用某片段的离散程度作为惩罚项对相似性就行修正，从而提取标准人类表型词条，做到完全匹配和模糊匹配。

设计平台：用 python 的 web 框架设计 http 接口，根据客户针对提取的结果的反馈信息，设计线上学习机制和历史匹配。

项目结果：项目上线后半年左右，准确率和召回率均达到 99%以上。

深圳市脑洞科技有限责任公司 AI 算法工程师 2018.03--2019.03

项目名：智能语言面试评分系统 **主要负责语音部分和少数文本部分**

项目内容简介：根据考公务员学生的面试语音，对面试内容进行评分

语音文件预处理：对语音文件的编码格式进行统一化、利用 RNN 对语音进行降噪处理，利用 GMM 模型对语音进行语音激活检测（有无人声声音）。

语音特征的提取：利用 librosa 的 python 包对语音进行处理，提取情绪浮动、时间利用率等特征。

语音识别：利用百度的语言识别 api 对语音转文本识别，即文本的预处理。

文本特征提取：搭建 bert 服务计算文本的相似度和利用 hanlp 提取关键词等，从而提取中心思想、流畅性等特征。

计算总分：根据提取的特征，对特征进行 0-10 的标准化，从而计算总分。

专业能力

算法能力：对传统机器学习（xgboost、hmm、crf、rf 等）和深度学习算法（rnn、transformer 等）具有深刻的理解，并能根据业务现状对算法进行创新改进，使得项目能很好的落地。

专业知识：熟悉基因组学、蛋白组学、分子生物学等。获得华大基因认证《生物信息分析员》证书。

科研能力：时刻关注并实践前沿的 AI 算法和 AI 在生物领域的创新，与时俱进。

落地能力：熟练 python、pytorch、sklearn 等工具，熟悉数据库 mongodb，熟练 docker 生态。

教育及专利

江西理工大学 本科：机械设计制造及其自动化 2014.09--2018.07

专利：《一种人类表型标准用语确定方法及装置》、《一种基因关联程度确定方法及装置》。

数学建模：国家二等奖、亚太赛区一等奖。