

OBJECTIVE 机器学习、自然语言处理、数据挖掘等相关职位

EDUCATION 硕士 计算机科学与技术 （免试）
郑州大学
2014 ~ 2017

学士 计算机科学与技术
郑州大学
2010 ~ 2014

EXPERIENCE 2016.04~2016.06 基于加权KNN算法的自动诊断模块设计：此项目将处理后的半结构化神经内科电子病历作为知识库（896例），通过加权KNN算法自动决策新患者所属的疾病类别（主要类别有癫痫、脑出血、脑梗死、帕金森、格林巴利等）。
担任职责：病例结构化，加权KNN分类模型的实现与优化
合作单位：南京富士通，郑州大学第一附属医院

2016.03~2016.06 基于动态最高特征对算法的基因表达分类：此工作是在传统最高特征对算法的基础上，提出的一种适用于时间序列基因表达分类的算法。
担任职责：DTSP算法的实现
论文：《DTSP-V:A Trend-based Top Scoring Pairs Method for Classification of Time Series Gene Expression Data》，BIBM 2016审稿中

2016.03~2016.05 胎儿体重预测模型：此项目从妇产科电子病例中，使用手工规则自动提取所有与胎儿体重相关的生理参数（2015年全年共计14844例），并利用特定结构的多层感知器构建胎儿体重预测模型。
担任职责：数据清洗，模型的选择、设计与实现
论文：《基于深度神经网络的胎儿体重预测》，已被《计算机科学》杂志录用

2015.10~2016.01 基于循环神经网络的电子病例匿名化框架TS-RNN：此工作提出一种适用于自动识别电子病例中隐私信息的循环神经网络结构。TS-RNN框架在一个中文数据集与两个英文数据集上均有优秀的性能，中文数据集是实验室与河南省妇幼保健院合作构建的数据集（485份，内含48072隐私实体）。
担任职责：病例文本预处理，TS-RNN网络结构的设计、实现
论文：《Learning to Recognize Protected Health Information in Electronic Health Records with Recurrent Neural Network》，已被国内NLP顶会NLPCC 2016录用

2015.09~2016.06 电子病例匿名化与信息抽取软件：通过手工制定的规则识别电子病例中的隐私信息、患者参数等信息，随后将隐私信息从电子病例中移除，将各项参数提取为格式化数据并存储为EXCEL文件。
担任职责：手工规则的制订，抽取模块开发及Windows平台界面开发
软件著作权：《电子病历处理软件V1.0》
合作单位：河南省妇幼保健院

2015.04~2015.07 开放领域的中文问答系统：此系统将百度百科、维基百科、百度知道、必应网典等作为在线知识库，利用中文分词、文本分类、语义角色标注等多种自然语言处理技术搭建开放域问答系统。系统主要针对的问题类型为人物、地点、时间等事实类问题。此系统参与NLPCC 2015 Chinese Open QA评测任务获得第二。
担任职责：QA System框架设计、问题分类模块的实现
软件著作权：《中文问答系统V1.0》

2014.04~2014.08 道路交通状况分析系统：系统通过抓取郑州市公交实时GPS候车信息，获取公交车在道路上的运行情况，进而估计特定路段的交通拥堵状况。此系统分为服务器端与Android客户端，服务器端部署于百度应用引擎（BAE）。

担任职责：公交候车信息的抓取与分析、系统在BAE的部署

实习单位：河南省信息网络重点开放实验室

2013.10~2014.05 校园新闻的分类与信息抽取系统：系统在每日定时抓取学校官方网站的新闻公告后，使用SVM对新闻进行分类，随后通过规则模板抽取新闻的主要信息。

担任职责：新闻分类模块与抽取模块的设计、开发

学士学位论文：《文本分类与信息抽取系统的设计与实现》 校级优秀毕业论文

SKILLS

- 学习模型：SVM, LSTM, KNN
- 英语：CET6（482分）
- 平台：了解Linux，熟悉Keras，熟悉MySQL
- 编程语言：JAVA > Python > Matlab > C Sharp

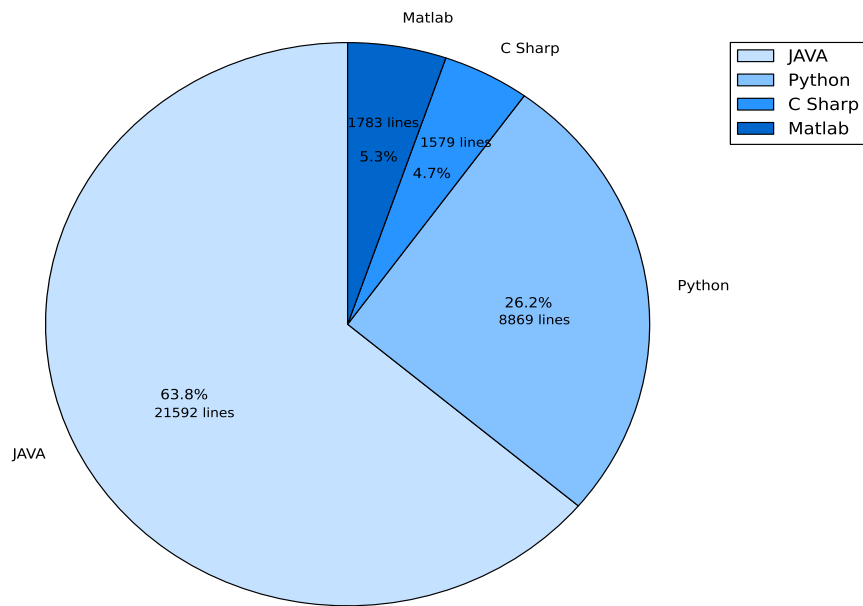


图1. 2014-2016年编程语言使用情况

LINKS

CSDN博客：<http://blog.csdn.net/zzukun>

个人主页：<http://zzukun.com>

GitHub：<https://github.com/zzukun>