

技能简介

- 编程语言: Python、C/C++、Java、html、css、js、Sql、Php
- 具有爬虫、机器学习、数据挖掘、数据分析、软件开发、网页开发、NLP 等相关经验

教育背景

- 2013.09 – 2017.07 汕头大学 计算机科学与技术 本科

在校事迹

- 2016.04 – 2016.05 获第四届“泰迪杯”全国数据挖掘挑战赛 全国三等奖
- 2016.08 – 2016.10 获阿里云安全算法挑战赛 季军

工作履历

- 2017.07 – 2018.03 深信服科技股份有限公司 算法工程师

项目/工作经验

2016.12-2017.05	毕业设计：面向招聘网的文本挖掘与推荐系统设计与实现
项目简介	<p>A 文本挖掘：</p> <ol style="list-style-type: none">1. 使用 python 爬取招聘网站的职位信息，建立数据仓库。2. 对数据进行清洗、分词、建立关键字字典、分类。3. 将处理后的数据按时间绘制图表得出一些岗位需求在职位类型上和地域上的分布信息。4. 采用 ARIMA 数学模型进行建模，得到未来岗位需求的预测区间。 <p>B 推荐系统：</p> <ol style="list-style-type: none">1. 设计系统架构。2. 数据仓库建立，用于存放用户信息、职位信息、用户行为等数据。3. 实现基于用户行为的协同过滤推荐算法、基于特征与流行度的推荐算法，并对推荐结果进行数据匹配验证。 <p>备注：本毕业设计的内容丰富性、创新性、挑战性得到评委导师们的认可，取得了不错的成果。</p>
2017.07-2018.03	基于人工智能的网页篡改检测
项目简介	<ol style="list-style-type: none">1. 数据爬取：为项目提供数据支撑，整站爬取各种网页、自动判断并爬取恶意网页、模拟浏览器自动判断并爬取 js 跳转网页。2. 关键词库自动扩充及半自动扩充：综合运用 word2vec、模拟浏览器搜索、TF-IDF 实现关键词库快速建立。3. 传统机器学习方法效果验证：利用词袋建立及 TF-IDf 构造特征，实现并验证 RF、LR、贝叶斯等传统机器学习方法的检测效果。4. lsi+xcgboost 实现：使用 gensim 库实现 lsi 主题抽取建立特征，并应用 xcgboost 分类器实现篡改检测。 <p>备注：除了上述技术性工作外，还负责网页篡改运营，收集整理运营数据并分析检测效果与误报漏报原因，该模块在误报率与漏报率上，效果均比原有模块好一倍。</p>

