

段凌霄

算法工程师

15712974785

@ duanlingxiao1219@sina.com

北京

西城区黄寺大街 23 号

askwitionary.github.io



教育背景

圣路易斯华盛顿大学

计算机科学与技术 硕士

2015.08 – 2017.06

密苏里, 美国

- 数据挖掘与机器学习方向
- 对学校医学院数据库中 Alzheimer's Disease 病人的 DNA 数据进行挖掘

惠提尔学院

数学与应用数学 学士

2010.09 – 2014.05

加利福尼亚, 美国

- 2011 曾获 Dean's List 提名
- 本科连续四年获得半额奖学金 (每年 22500 美金)

工作经历

算法工程师

北京三点一刻科技有限公司

2018.05 – 2018.09

北京

- 研发并调试平台里项目、甲乙方的多维征信、匹配、推荐逻辑与算法
- 以 Tensorflow 的 CNN 为基础, 搭建图像识别模型, 帮助爬虫与自动化测试同事解决自动识别图形验证码的问题
- 图片文字转文本的优化、聊天记录截屏中 emoji 识别
- 探索、学习各个相关开源项目的代码和文档并基于可能用到的场景做 demo, 如 LSTM 聊天交互、自动写稿、趋势预测; 决策树、随机森林分类; CNN 文本、图像分类; NLP 相关: 分词、句法分析、词向量、关键词提取等

数据库开发、管理员

comScore, Inc.

2017.07 – 2017.08

弗吉尼亚, 美国

- 根据客户需求为客户创建数据库逻辑与框架设计, 范围包括 MySQL, PostgreSQL 等
- 为客户提供数据库维护服务

数据分析师

AIE International Inc.

2014.08 – 2015.06

新泽西、加利福尼亚, 美国

- 为客户提供多方位咨询服务
- 协助创建、维护客户信息库并提供数据分析

个人期望



期望行业

医疗大数据行业



期望职位

机器学习、人工智能、数据挖掘 (大数据相关)

算法工程师、研究员

编程及相关语言

Python

JAVA

LaTeX

Linux/Shell

MediaWiki

SQL

常用软件及平台

MS Office

Mathematica

Maple

MATLAB

IDL

Blender

Photoshop

Unity

自然语言

中文

英文

西班牙语

数据挖掘与机器学习常用工具

Tensorflow

SK-Learn

Caffe

决策树

随机森林

LSTM

卷积神经网络

GBDT

线性、非线性规划

自然语言处理相关

基于深度学习的图像、视频物体识别与追踪

个人项目

📅 2018.01 – 2018.04 📍 加利福尼亚, 美国

- 基于 Google 开源项目，对图片及视频中出现的常见物体进行识别及追踪
- (视频) 通过连续位置追踪确定目标并对其做简单行为判定 (对视频质量要求高, 运算成本大)

对 Alzheimer’s Disease 病人基因与进行数据挖掘

研究生个人项目

📅 2016.01 – 2016.12 📍 密苏里, 美国

- 使用华盛顿圣路易斯大学医学院老年痴呆症病人以及对照组正常人的基因片段进行数据挖掘, 对病人与非病人的基因进行分类
- 对较大规模数据进行清洗与标准化
- 对 8000 多个基因维度进行维度分析, 缩小可能为致病基因的范围
- 学习、练习了数据挖掘相关的基础方法, 包括各种回归分析、决策树、临近算法、SVM 等

生物医学图像处理几何运算

研究生个人项目

📅 2015.08 – 2015.12 📍 密苏里, 美国

- 使用 Mathematica 对学校医学院的图片素材进行分析与处理, 主要达到的功能包括精确寻找 CT 骨骼结构和寻找 MRI 肿瘤位置与轮廓等
- 使用 Mathematica 对骨骼、器官进行 3D 建模并进行以下处理: 寻找主体、体积计算、表面积计算、模型锐化与平滑化、模型化简等

网站爬虫抢购插件

个人项目

📅 2017.12 – 2018.4 📍 加利福尼亚, 美国

- 鉴于如 Supreme 等潮牌, 尤其联名款的商品常常被一抢而空的情况, 考虑到需求远远大于市场, 利用 Python 写了一个定时抢单脚
- 本。脚本特点如下:
- 使用代理服务器实现高并发防止 IP 被封
 - 可个性化修改需要抢的数量及商品
 - 直接对 html 页面进行解析, 解决部分改版后不能用的问题

图片验证码识别

北京三点一刻有限公司 算法工程师

📅 2018.06 – 2018.07 📍 北京

- 根据各个以图片为验证码的网站, 包括数字、大小写字母识别; 鼠标滑块; 按顺序选字等图像人机验证的生成逻辑, 按照下列步骤实现对人机验证的突破
- 识别机制
 - 分析验证码生成逻辑
 - 仿制验证码生成器从而提供无限量训练数据
 - 根据不同验证码特点设定突破步骤
 - 搭建模型并进行训练
 - 根据训练过程和测试结果调参重复训练
 - 将正确率合格的模型封装上线
 - 成果
 - 对本平台网站图片识别率达 99
 - 对外部平台的图片验证码识别率根据复杂程度从 50% ~ 85% 不等

数据分析与征信

北京三点一刻有限公司 算法工程师

📅 2018.05 – 2018.08 📍 弗吉尼亚, 美国

- 根据平台数据以及爬取的外部平台数据实现对营销需求、甲方、乙方的多维度评分。评分系统通过比较分类中其他项目归一化, 并实时更新, 基本不用维护。
- 评论文本打标签、提取句子主干、分类
- 需求统一化分类
- 分析各分类下需求描述的关键信息点, 并且使用 TF-IDF 为基础模型对需求描述中非结构化信息点进行分析评价, 并使用以上方法评价描述信息完整度
- 对每个用户 (公司/个人) 根据百度精确搜索获得的条目数确定其品牌、影响力
- 根据相关信息, 使用统计学方法获取其他维度 (共 13 个) 评分

数据分析师

AIE International Inc.

📅 2014.08 – 2015.06 📍 新泽西、加利福尼亚, 美国

- 为客户提供多方位咨询服务
- 协助创建、维护客户信息库并提供数据分析