

彭聪

上海市锦绣路 3088 弄 54 号 102 室 | (+86)150-2180-8015 | pengcong@ucla.edu**专业技能:****编程语言:** Java, Python, JavaScript(React), C++, PHP, HTML/CSS, MATLAB**办公技能:** Git, Linux/Windows, SQL(MySQL/Mongo), Selenium**教育背景:****加州大学洛杉矶分校 (UCLA)**

预计毕业时间: 2017.12

- 电子工程, 信号与系统方向 (GPA: 3.6/4)

北京邮电大学 (BUPT)

2012.08 - 2016.06

- 通信工程 (GPA : 88/100)

里斯本高等理工学院 (Instituto Superior Técnico, Lisboa) 电子与计算机工程

2015.02 - 2015.06

专业经历: [课程编号/经历性质: 学科名/研究课题 (所用工具)]**[实习] 数据挖掘实习生, 平安科技, 上海 (Python, Mongo dB, Selenium & Hive)**

2017.07 – present

基于 Scrapy 的大型网络爬虫平台, 该爬虫支持 (1) 微博 (2) 百度热点 (3) 搜狗微信公众号热搜 (4) 百度指数

- 使用 BeautifulSoup, Xpath 解析, 通过维护动态 IP、Cookie 池突破反爬机制。
- 使用 Selenium 模拟登陆处理验证码、手势码以及异常操作等问题。
- (1) 的结果作为基于兴趣分类的平安保险推荐系统的输入, (2, 3 & 4) 的结果接入大数据平台作为实时爬取接口。

[COM SCI 230] 软件工程, Dynamic Universal Software System Visualization Kit (C++, Python, & JS)

- 静态解析器: 解析代码库的方法以及各参数的属性, 通过 JSON 包输入 IR (Intermediate Representation) 模块。
- IR: 将存储的 JSON 文件上传到 Django 上, 并且使用 PageRank 算法对类中的方法根据调用情况进行排序, 输出结果传递给 AE (Attributes Extractor)。
- AE: 解析 JSON 包, 统计数据得到可视化绘制标准, 分别有 LOC、方法数目以及属性数, 将统计数据交给画图器。
- 用 D3.js 绘制出软件可视化图。(Hierarchy Tree, Sunburst, Edge Bundling, Circle Packing & PageRank)

[COM SCI 130] 软件工程, iOS React-Native Mobile application Development – Front End (JavaScript, React)

- 在现有的用户界面上进行修改, 并且加上一些相同风格的图形界面丰富应用的实用性。(Google Map, Yelp APIs)
- 和后端进行交互从新搭建的数据库获取数据, 并在应用上合理的显示。(Mongo dB)
- 针对用户需要, 对应用所提供的服务进行扩展, 使用户使用更加方便。(对菜单栏, 登录界面的 Components 优化等)

[EL ENGR 219] Large-Scale Data Mining: Models and Algorithms (Python & MATLAB)

2017.01 - 2017.03

- 回归分析: 根据不同的回归算法 (线性回归, 脊回归, 逻辑回归, 多项式回归等) 分别对网络备份以及房价数据进行回归分析, 对于数据的欠拟合以及过拟合情况分别进行处理。
- 分类、聚类分析: 对于新闻文本数据进行分类, 从中提取特征值, 通过不同的特征提取策略 (TFxIDF, LSI, PCA 等), 对不同的分类算法 (线性 SVM, soft margin SVM, 朴素贝叶斯等) 进行性能分析; 同时利用 K-means 算法对文本进行非监督学习, 再与之前的标签进行对比。
- 协同过滤分析: 通过 NMF 对评分矩阵进行降维, 再利用 ALS 协同过滤, 实现了一个基于网友评分的电影推荐系统。

[EL ENGR 210A] Adaptation & Learning (MATLAB)

2017.01 - 2017.03

- 从底层设计了一种不仅限于图片分类的卷积神经网络系统, 从基础结构, 前馈, 反馈算法上进行设计。
- 在设计 CNN 网络过程中成功地处理 padding, partition, pooling, permutation 等问题, 并且在设计过程中考虑节约内存以及计算效率的问题, 并且提出了一种新的生成卷积层的方法。
- 对 MNIST 以及 ImageNet 上的数据集进行训练, 数据集大小分别为 60000 张 28*28 的数字图像 (10 类) 以及 1000 张 256*256 的动物照片 (4 类), 测试的准确度分别为 94.11% 和 45.24%。

[本科毕业设计] 基于自动化路径识别的 Wi-Fi 定位算法优化分析 (Java, Python & SQL) 清华大学 2015.11 - 2016.05

- 通过大量分析用户在给定区域的历史路径, 总结出三种运动特性, 分别是连续变向概率、转弯角度范围以及整体转弯频度, 并通过对室内环境进行两个层面的建模 (连通矩阵以及辅助点), 由此提高算法的定位精度,
- 提出一种在采集过程进行单次测量就能够保证定位精度的普适算法, 通过获取用户在室内的移动路径, 以聚类算法将用户的运动特性进行提取, 作为判断路径的参考, 最后在进行定位的过程利用分析得到的运动特性对路径进行修正, 实验结果表明本算法能够在保证能耗一定的情况下达到令人满意的定位精度。