

陈曦

xchen595@163.com | +86 177-0642-7612 | www.linkedin.com/in/xichen612

资格能力概述/SUMMARY OF QUALIFICATIONS

- 有上进心，有创造力，学习能力强，具备优秀的团队合作精神。
- 对大数据和数据科学富有激情和热爱。
- 在机器学习，深度学习和数据挖掘上有很强的专业技能和学习能力。
- 熟练 Python, Tensorflow, R, 熟悉 Unix Shell, Spark。

教育背景/EDUCATION

统计系, 美国加州大学戴维斯分校 (University of California, Davis)

Davis, CA

统计学 硕士(M.S)

2014年9月 - 2016年4月

数学与统计学院, 山东大学 (Shandong University)

Shandong, CHINA

数学与应用数学/金融学 学士(B.S)

2010年9月 - 2014年6月

研究经历/SELECTED RESEARCH EXPERIENCE

项目, 美国加州大学戴维斯分校 (UC-Davis)

Davis, CA

纵向数据的过度分散化(overdispersion)

2016年3月 - 2016年4月

- 应用泊松混合模型和负二项模型解决癫痫病发病量数据的过度分散化问题。
- 通过 Generalized Estimating Equations (GEE) 和 Monte-Carlo EM 算法估计方差分量。
- 用 Jackknife and Bootstrap 方法得到模型系数的标准差。
- 比较协方差矩阵模型和之前两个模型结果的利弊。

Mashable News 热点和普及度分析(popularity analysis)

2015年5月 - 2015年6月

- 从时间, 频道等对网站上各个文章的转载数进行分析探索。
- 根据抓取的关键词画出 2013 和 2014 两年的词频云图。
- 建立 Random Forest 模型对日分享量进行预测并得到最重要的几个性质。

相关工作经历/RELATED WORK EXPERIENCE

大数据算法工程师, 海康威视研究院 (HRI)

Hangzhou, CHINA

犯罪嫌疑人分析推荐

2017年4月 - now

- 搭建基于长短期记忆网络(LSTM)的深度学习框架对目标对象的交互行为进行学习, 在隐性关系进行识别任务上, 比传统机器学习方法在精确度上高 5%。
- 充分利用安防行业数据中的关系图, 时间及空间信息, 研发由 CANE 算法改进时空图挖掘算法在犯罪人员关系图谱中提取 graph embedding, 落地效果比传统 GCN 如 GraphSAGE 在 top50 上高出 10%。

科研助理, 美国劳伦斯伯克利国家实验室 (LBNL)

Berkeley, CA

城市居民用电量分析

2016年1月 - 2017年3月

- 应用 ETL 对 100GB 的原始数据进行清洗得到基于小时的居民用电量数据。
- 建立广义线性混合模型对居民每天高峰时期用电基本线进行估计。
- 通过模型不确定性参数对 8 个节能项目的稳定性进行分析。
- 对之前的模型进行贝叶斯序贯学习研究并寻找对于不同数据分割的收敛点。

数据科学实习, 美国宾州州立大学(PSU)

State College, PA

基于土壤物理和化学性质的量化分类

2015年6月 - 2015年8月

- 探索全球土壤和美国土壤物化性质分布, 制成描述性统计量界面。
- 对土壤分层差异做 Tukey's 配对比较检验。
- 在主成分维度对土壤主次层级进行 K-means 聚类可视化。
- 应用 random Forest, SVM 等机器学习方法对土壤层级进行分类预测。

个别荣誉奖项/SELECTED HONORS & AWARDS

- 美国数学建模大赛二等奖, USA 2012
- 全国数学建模大赛国家二等奖, CHINA 2012
- 三等奖学金, 山东大学 2010 - 2012

其它信息/ADDITIONAL INFORMATION

- 编程: Python, R, MATLAB; 数据库: SQL; 分布式计算: Spark; 深度学习: TensorFlow, Keras;
- 爱好特长: 篮球, 羽毛球