

工作经历

- 2020.08 -

英国巴克莱银行，毕业生数据科学家, 数据科学团队（第二次轮岗）
- 2019.08 - 2020.07

英国巴克莱银行，毕业生后端开发, 卡平台团队（第一次轮岗）
- 技术栈:

Jenkins, Jira, Confluence, BitBucket, Openshift, Docker, GridGain, Maven, Gradle, Wiremock, Mockito, Spring Boot, AWS, SonarQube, Karate, AppDynamics

• 端到端功能开发，测试（单元，功能，性能），部署（CD）

• 为API添加缓存层，以减少重复数据获取的延迟

• 使用内部的Spring Boot模板迁移遗留代码，并对遗留代码进行重构以提高其可读性和性能

• 从头构建内部使用的工具（比如 git hooks）和 脚本（python / bash）来自动化软件开发工程，并减少潜在的人类错误

• 参与了NLP项目（比如基于BERT的地址命名实体识别），同时跟进NLP领域的最新进展，特别是在迁移学习，模型压缩（低资源推断）和多模态融合方向

• 在新冠疫情期间远程工作的半年中，和团队一起仍然保持了较高的沟通和交付效率

教育背景

- 2018 - 2019

伦敦大学学院，网络科学和大数据分析（硕士），Distinction

核心课程: 概率图模型; 复杂网络; 情感计算; NLP; 信息检索; 多智能体AI; 应用机器学习; 深度学习导论
- 2016 - 2018

英国利物浦大学，互联网计算，一等荣誉学士

核心课程: 软件工程; 数据库开发; 网络原理（OSI导论）; 面向对象编程; 分布式系统原理; 软件开发工具（主要关于测试）; C语言和内存管理; iOS编程（Swift）; 知识表达和推理（Modal Logic 和 Descriptive Logic）; 多代理系统（MARs）; E-commerce（拍卖机制, RSA, DH密钥交换, 椭圆曲线加密）
- 2014 - 2016

西交利物浦大学，信息与计算科学

核心课程: 计算机系统; 数据库导论; Java编程导论; 算法基础和问题求解; 数据结构; 操作系统概念; 微积分; 离散数学导论

项目

- 2019.06 - 2019.09

项目实习（研究生学位论文）@ **Astroscreen**

技术栈: Python, Keras, Tensorflow, MulticoreTSNE, Matlabplot

社交网络语言的来源识别（推特和Gab） 完成了一个爬虫从Gab.com爬取语言（帖子），使用正则表达式预处理数据，fine turning BERT 和 XLNet来分类文本来源，并将输出使用 t-SNE可视化，使用准确率，F1 Score, 困惑矩阵，马修斯相关系数评价模型。
- 2019.03 - 2019.04

信息检索课程项目

使用Fact Extraction and Verification (FEVER)数据集进行了多项练习。

包括词频统计并验证zipf's law;实现向量空间文档索引（TF-IDF）;实现查询似然（Query likelihood）文档索引（并分别应用Laplace平滑，Jelinek-Mercer平滑和Dirichlet平滑），实现逻辑回归比较句子相似性;实现Precious, Recall和F score函数;使用神经网络检验文档Truthfulness。
- 2019.02 - 2019.03

Integrating BERT and Embeddings into CommonsenseQA Challenge

我们在CommonsenseQA 1.0数据集（3选项）上fine-turning了Google BERT并尝试整合了Conceptnet Numberbatch and ELMo 词嵌入来尝试提高模型性能。这个数据集包含一组需要人类常识来回答的多选题。

使用BERT+ELMo的组合我们在验证集上取得了68.79%的准确率（BERT: 67.47%; BERT + Numberbatch: 67.68%）。多次实验取最佳结果
- 2019.02 - 2019.02

"识别由食物引起的情感" 项目标注平台

技术栈: Bootstrap, JQuery, Recordrtc.js, Java, Play! framework 2, PostgreSQL, Cloud Foundry

一个用来收集数据（食物味觉情感）的简单的自适应web app。这个网站曾有个demo运行在<https://affective-computing-data-collection-dist.cfapps.io/sessions>（当前因云服务费用用尽已下线） 本项目是我在一个周末完成的，是这个项目的草稿版本（后续团队中其他人移除了个人信息并创建了一个本地版本）
- 2019.02 - 2019.03

通过CTR预测和拍卖函数选择以最大化广告点击数

在一个实时广告需求方平台拍卖的历史数据库中预测是否用户想要点击在线广告。预测结果作为一个拍卖竞价函数的输入去预测出价价格。同时总出价会有一个最大限制。数据集是非平衡数据集（超过300000条拍卖纪录中约有3000个正样本（点击数））。我们尝试了许多不同的模型（XGBoosting, 浅层nn, 逻辑回归）和一些竞价函数。我们也在这个项目应用了 undersampling 和重校正技巧。我们与其它学生在leaderboard进行了比赛（共30组学生），取得了第3名。

（我们取得了185个点击，前两名均取得了186个点击）。
- 2018.12 - 2019.01

上传越多的up主越受欢迎吗？一个对bilibili基于网络的分析

技术栈: Python, networkx, graph-tool, MySQL

该项目检验了一些b站用户（up主）关注网络的属性（度分布和assortative系数），并通过网络可视化检查了是否上传视频数量（反映活跃度）和节点入度（反映受欢迎程度）是相关的。过程中写了一个爬虫来通过b站的RESTful API抓取数据。
- 2017.09 - 2018.05

2维情况下对集群协议(Population Protocol)和网络构造协议(Network Constructor)的模拟，可视化和实验分析

集群协议（Population protocol）是一个分布式计算理论模型。该模型包含一组不可区分的agent。Network constructor 和 Terminating grid network constructor 是对集群协议的扩展，不同之处在于其目标在于构建不同拓扑的网络而非计算函数。
- 2017.04 - 2017.05

bookswop.me (https://github.com/billweasley/Distributed_library_Release)

技术栈: Play! framework 2 (以Java为主), MySQL, Nginx, AWS

针对于英国大学学生和教师的图书交换平台。此为大三小组项目（小组9人）。
- 2016.08 - 2016.09

西浦ICE学习平台爬虫 (<http://shellcottage.me/firstSpider>)

技术栈: JSoup, Java Swing

一个可以自动备份西浦ICE学习平台资料的爬虫。

奖励和荣誉

- 2015年 10月

国家励志奖学金
- 2015年 7月

西浦学业成就奖（升学奖学金）