

徐彩旭

✉ csxucaixu@gmail.com · ☎ 176-1079-3303 · 🌐 <https://calxu.github.io/about/>

🔑 基本信息

性 别: 男

学 历: 硕士

政治面貌: 中共党员

年 龄: 27

籍 贯: 江苏

求职意向: 大数据方向

🎓 教育/工作背景

度小满金融（前百度金融），北京 201807-至今

数据挖掘工程师 涉及模型/策略/分析

苏州大学, 江苏, 苏州 2015-2018

硕士 计算机技术

专业排名前 15%

东南大学成贤学院, 江苏, 南京 2011-2015

学士 计算机科学与技术

专业排名前 3%

👨‍💻 项目经历

风控挖掘子模型 2019.02-至今

项目属性: 线上风控子模型分

项目角色: 项目主要贡献者

主要工具: XGBoost, Liblinear

挖掘强相关的子模型分供模型和策略使用, 涉及技术主要有 ETL、RFM 挖掘方法、评估;

- 平台环境搭建 (Hadoop/MapReduce/Hive/Spark)
- 数据挖掘: RFM 动态窗口滑动子模型、底层变量自动化 ETL 与例行;
- 效果评估: 单变量相关性分析 (梯度/WOE/IV) 的评估;

黑名单系统 2018.07-至今

项目属性: 风控基础设施

项目角色: 项目主要贡献者

主要工具: Python, Linux, Shell, MySQL, Hadoop, MapReduce, Hive, XGBoost, Liblinear

黑名单系统重构与升级, 涉及技术有 ETL、分布式爬虫、图关联、RFM 挖掘子模型、自动化评估;

- 数据挖掘: 分布式爬虫、图关联网络关联、RFM 动态窗口滑动子模型;
- 数仓管理: 数仓自动化 ETL、数据库自动化管理监控;
- 效果评估: 重要指标 (Precision/命中数/命中率/Lift) 的自动化评估、自动化展示;

基于位置社交网络的用户关系强度预测 2016.01-2017.02

项目属性: 硕士毕业设计 (机器学习项目)

项目角色: 独立研发完成

主要工具: Python3, Linux, C++, MySql, Scikit-Learn, XGBoost

利用斯坦福大学开源 SNAP 数据 (近亿条签到数据, 几百万用户) 构建分类模型, 模型预测任意用户对关系强度。在召回率相同的情况下, 精确度比最好的模型提高 10%。

- 平台搭建 (Linux 系统、Python3、XGBoost);
- 数据 ETL 处理与管理, 多维度抽取特征、构建多视角分类器模型、模型调测;

轨迹相似度计算系统

2016.03-2016.08

项目属性: 比赛项目 (第五届中国软件杯决赛三等奖)

项目角色: 项目主要负责人

主要工具: Java, Python3, JS, UML, Linux

利用全国近三千万轨迹数据搭建相似度计算平台, 平台可以快速响应请求: 计算任意两条轨迹的相似度; 任意画出一条轨迹, 从数据库中快速检索出与当前轨迹最相似的轨迹, 并以地图显示的形式反馈给用户。

- 数据的聚类处理与存储;
- 该项目采用 BS 模式, 前台采用 JSP+CSS+DIV;
- 后台 Java 实现, 构建基于聚类簇的数据结构 (以响应快速读取操作); 定义相似度并设计高效算法;

公司信誉度评估分类系统

2015.06-2015.11

项目属性: 实习项目 (苏州汇誉通数据科技有限公司)

项目角色: 项目贡献者

主要工具: Python3, Linux, Scikit-Learn

到各大信息公开网站采集公司的异常记录、公司基本属性等数据; 根据数据构造特征; 对部分公司进行聚类标记, 构建随机森林回归预测模型, 模型输出公司所属类别。

- 数据的采集与清洗;
- 数据存储与管理;
- 特征工程与模型调测;

扫雷游戏设计

2014.12-2015.04

项目属性: 课程毕业设计

项目角色: 独立设计开发

主要工具: Java, UML, Swing 包, MySql

分析扫雷游戏需求并用 UML 建模 (包含初级、中级和高级三个级别), 记录进行游戏所花费的时间, 完成游戏则记录最快玩家所花费的时间, 并将这些信息写入数据库。

- 利用 UML 建立用例图和活动图; 游戏的图形展示界面利用 swing 包搭建;
- 利用 UML 建立详细类图、类关联图及交互图; 并设计递归生成雷区的算法;
- 设置完整的测试用例对游戏进行功能、边界、负面三个层面的测试;

居民信息管理系统

2013.06-2013.09

项目属性: 暑假实习项目 (盐城思源网络科技有限公司)

项目角色: 项目贡献者

主要工具: C#, ASP.Net, Html, CSS, JS

对公司原始的客户信息系统进行升级, 本人负责整改与添加新功能, 整改用户的 UI 界面, 整改信息筛选模块 (改写 Sql 语句), 添加地图显示小区位置模块。

- 整体采用 ASP.Net 框架, 前台使用 html、CSS、JS 技术;
- 后台使用的是 C# 语言, 数据库使用 MySql 数据库;
- 改版后的客户信息管理系统交互性极大提高, 界面更加简洁大方;

语言水平

- 中文 (普通话): 母语
- 英语: 听、说良好, 读、写流利, CET-4/CET-6

🔧 IT 技能

- 编程语言: Python = Shell = SQL > awk > JAVA > C/C++ > C#;
- 大数据工具: 熟悉 Linux、Hadoop、MapReduce、Hive; 了解 Spark;
- 机器学习框架: 熟悉 XGBoost、Liblinear、Word2vec、LDA;
- 数据库: 熟悉 MySQL、Oracle、PostgreSQL、Hive、KV 存储的数据库;
- 软件工程: Git/Github/Gitlab、UML、OOA/OOD;

📍 社会活动

计算机科学与技术学院学工助理, 东南大学成贤学院	2013
院 DV 工作室技术部副部长	2012
院科学与技术协会副部长	2012

★ 获奖情况

基于位置社交网络的用户关系强度预测方法、装置及设备 (专利)	2018.05
轨迹相似度分析与预测系统 (计算机软件著作权)	2018.01
基于时空数据的用户社交关系强度预测软件 (计算机软件著作权)	2016.05
苏州大学三等奖学金	2015-2016
东南大学成贤学院优秀毕业生	2015.06
东南大学成贤学院三好学生	2013-2014
东南大学成贤学院综合一等奖学金	2013-2014
全国计算机二级 C 语言能力等级证书	2013
江苏省高等学校计算机二级 VFP 能力等级证书	2013
院级二等奖学金	2012-2013
院级三等奖学金	2012-2013
院级三等奖学金	2011-2012

☆ 学科竞赛

DataCastle: 交通线路通达时间预测, 26 名	2017.01
第五届“中国软件杯”大学生软件设计大赛决赛三等奖	2016.08
东南大学第五届程序设计竞赛二等奖	2014.05
院级网页设计制作大赛一等奖	2011.12

📄 论文发表

- Inferring Social Ties from Multi-view Spatiotemporal Co-occurrence (第一作者, CCF C 类)
- Context Co-occurrence Based Relationship Prediction in Spatiotemporal Data (第一作者, EI 检索)

💡 自我评价

- 热爱代码, 代码洁癖, 注重代码的可读性;
- 具备数据思维; 对从事大数据方向的工作有很强热情;