

徐彩旭

✉ csxucaixu@gmail.com · ☎ 176-1079-3303 · 🌐 www.xucaixu.com

🏠 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: 数据挖掘方向

👤 个人背景

京东-京东数科, 北京大兴区

2020.03-至今

💼 数据挖掘工程师, 涉及企业图谱/企业指标体系

百度-百度金融 (度小满金融), 北京海淀区

2018.07-2020.01

💼 数据挖掘工程师 (实习/正式), 涉及风控模型/策略/分析

苏州大学, 江苏苏州

2015.09-2018.07

🎓 硕士, 计算机技术 (专业排名 *Top 15%*)

东南大学成贤学院, 江苏南京

2011.09-2015.07

🎓 学士, 计算机科学与技术 (专业排名 *Top 3%*)

⚙️ 个人技能

- 编程语言: Python = SQL = Shell > awk > JAVA > C/C++ > C#;
- 大数据工具: Linux、Hadoop、MapReduce、Hive; 了解 Spark;
- 数据库: MySQL、PostgreSQL、Hive、KV 键值对数据库;
- 机器学习框架: XGBoost、Liblinear、Word2Vec、LDA;
- 软件工程: Git/Gitlab/GitHub、UML、OOA/OOD;
- 语言水平: 英语读写流利, CET-4/CET-6;
- 其它: 数据结构、Neo4j、LaTeX、Markdown、Office 系列、数据可视化、静态 Web 设计;

🏆 获奖情况

- 专利 × 1 2015.09-2018.07
- 计算机 EI 检索会议论文 × 2 2015.09-2018.07
- 计算机中文核心期刊论文 × 2 2015.09-2018.07
- 计算机软件著作权 × 2 2015.09-2018.07
- 全国软件设计大赛决赛三等奖 × 1 2015.09-2018.07
- 计算机科学与技术学院奖学金 × 6 2011.09-2018.07
- 计算机能力等级证书 × 2 2011.09-2015.07
- 本科优秀毕业生兼三好学生 × 1 2011.09-2015.07
- 本科软件设计大赛三等奖 × 2 2011.09-2015.07

📄 论文发表

- **Caixu Xu** and Ruirui Bai. Inferring Social Ties from Multi-view Spatiotemporal Co-occurrence (APWeb-WAIM 2018, 第一作者, CCF C 类)
- **Caixu Xu**, JianFeng Yan and etc. Context Co-occurrence Based Relationship Prediction in Spatiotemporal Data (CMSA 2018, 第一作者, EI 检索)

👤 项目经历

企业征信图谱

2019.03-至今

- **项目属性:** 京东数科 B 端解决方案产品-企业图谱
- **项目角色:** 项目主要贡献者
- **主要工具:** SQL, python, Linux, Neo4j, 图存储与计算

构建企业图谱基础能力层和应用层, 为业务方提供图谱产品输出。主要涉及技术点如下: (a) 基础能力层: 十亿量级顶点与边基础能力构建 (图谱 1/2 度能力、连通子图的多层穿透); (b) 图谱应用层: API 接口 (股权穿透、司法关联), 图谱可视化 (图的存储与设计), 企业指标体系 (近 500 个)。

风控挖掘子模型

2019.02-2019.11

- **项目属性:** 度小满金融信贷风控-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, 主要有 ETL、RFM 挖掘方法、效果评估。主要涉及技术点如下: (a) 平台环境搭建 (Hadoop/MapReduce/Hive/Spark); (b) 数据挖掘: RFM 动态窗口滑动子模型、底层特征自动化 ETL 与例行; (c) 效果评估: 单变量相关性分析 (梯度/WOE/IV) 的评估。

黑名单系统

2018.07-2019.08

- **项目属性:** 度小满金融信贷风控-线上风控基础设施
- **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, 主要有 ETL、分布式爬虫、图关联、RFM 挖掘子模型、自动化评估。主要涉及技术点如下: (a) 数据挖掘: 分布式爬虫、图关联风险传播、RFM 动态窗口滑动特征; (b) 数仓管理: 数仓自动化 ETL、数据库自动化管理监控; (c) 效果评估: 重要指标 (Precision/命中数/命中率/Lift) 的自动化评估、自动化监控。

基于时空上下文共现的用户关系强度预测

2016.01-2017.02

- **项目属性:** 硕士毕业设计
- **项目角色:** 独立研发
- **主要工具:** Python3, Linux, C++, MySql, Scikit-Learn, XGBoost

利用开源数据 (近亿条签到数据) 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%。主要涉及技术点如下: (a) 平台搭建 (Linux 系统、Python3、XGBoost); (b) 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 (第五届中国软件杯决赛三等奖)

2016.03-2016.08

利用全国轨迹数据搭建计算平台, 快速响应计算相似度的请求, 并从数据库中秒级检索出最相似的轨迹。主要涉及技术点如下: (a) 数据的聚类处理与存储 (Python3、Linux); (b) 采用 BS 模式, 前端采用 JSP+CSS+DIV; 后端用 Java 实现构建基于聚类簇的数据结构 (UML), 秒级响应并计算相似度。

公司信誉度评估分类系统

2015.06-2015.11

到各大信息公开网站采集公司的异常记录、基本属性等数据, 构造特征并对部分公司进行聚类标记, 从而构建随机森林分类预测模型, 预测公司所属类别。主要涉及的技术点如下: (a) 数据的采集与清洗 (Python3、Linux); (b) 数据存储与管理 (MySQL); (c) 特征工程与模型调测 (Scikit-Learn)。

扫雷游戏设计

2014.12-2015.04

扫雷游戏 UML 建模并 Java 实现其功能, 玩家信息与 MySQL 数据库对接。主要涉及技术点如下: (a) 建立 UML 用例图和活动图, 游戏的图形界面展示; (b) 建立 UML 详细类图、类关联图及交互图, 并设计生成雷区的算法; (c) 设置测试用例对游戏进行功能、边界、负面三个层面测试。