

# 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

滴滴, 北京海淀区	2020.12-至今
💡 数据挖掘工程师, 涉及运筹算法/机器学习	
京东-京东数科, 北京大兴区	2020.03-2020.12
💡 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, 北京海淀区	2018.04-2020.03
💡 数据挖掘工程师, 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 (个人排名 <i>Top 15%</i> )	
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 (个人排名 <i>Top 2%</i> )	

## ⚙️ 个人优势

- **个人经验:** 3.5 年互联网大厂的项目经验, 做过很多数据分析相关的工作, 从 0 到 1 负责开发过 5 个工业级项目, 重构过 2 个工业级项目;
- **落地经验:** 熟悉工业界常用的建模方法论和流程, 可熟练使用常用数挖方向的语言和框架;
- **合作经验:** 能很好地与产品、运营、工程团队合作把项目做大;
- **其它:** 具备对业务问题抽象, 抓住业务问题本质和建模的能力; 具备快速学习、快速迭代和一定的抗压能力;

## ⚙️ 个人技能

- **编程语言:** Python = SQL = Shell > Scala > 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;

## 👥 项目经历

运筹定价	2021.04-2021.09
• <b>项目属性:</b> 配送司机定价	
• <b>项目角色:</b> 项目主要负责人	
• <b>主要工具:</b> SQL, Jupyter, python, Linux	

协助业务方降本，分析业务问题痛点并进行抽象建模；利用运筹算法解决司机定价问题；最终该定价方案已经全国推广，最终收益：保证了司机收入的公平性同时实现 3% 的降本；主要涉及技术点如下：a) 对司机进行多因素定价，包括：起步价、团点数、里程、重量、逆向费用；b) 抽象业务问题的优化目标，总定价最小化的同时方差最小；c) 抽象业务问题中的约束，包括：总成本 5% 的波动、仓库供应商总成本满足在区间内；d) 监控指标层面的搭建，包括：司机收入分层、司机收入变化、司机留存率；

## 履约成本拆分

2021.05-2021.11

- **项目属性：** 滴滴橙心优选履约成本基础能力
- **项目角色：** 项目主要负责人
- **主要工具：** SQL, python, Linux, Shell

协助业务方看清履约成本的结构，抽象业务需求并对各项成本拆分，构建橙心履约体系基础成本能力；该成本拆分项目已经在全公司推广，最终收益：橙心的基础能力，服务 9 大业务场景，滴滴公司内部表的价值排行 Top 0.26% (699/264301)，服务下游 4048 个依赖表；主要涉及点如下：a) 与业务拉齐目标并抽象出拆分方案，多次核对并拉齐；b) 对仓租、水电、人力、物资、冷链、配、管理人员，7 项总成本按一定的业务逻辑进行拆分；c) 配送成本按网约车拼车逻辑算法进行边际成本的拆分；d) 监控指标层面的搭建，包括：区域成本看板、超远团看板、团长商分、低效团看板；

## 司机配送规划

2020.12-2021.05

- **项目属性：** 司机配送路径规划
- **项目角色：** 项目主要负责人
- **主要工具：** SQL, python, Linux, Jupyter, Shell

分析司机的配送行为，规划推荐出司机配送顺序；最终收益：目前已经推广到全国司机，平均节省距离 6km (节省 14.8%)，平均节省时间 17min (节省 20.9%)；主要涉及点如下：a) 分析司机的配送行为/配送特点，并找到相关可优化的点；b) 利用基于区块的贪心搜索方式快速上线一版比司机人工好的策略；c) 从产品层面优化基于下一个配送点的位置推荐策略算法，50% 司机反馈良好；

## 企业征信图谱

2020.03-2020.08

- **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 (投资任职、股权穿透、最终受益人、司法关联)；图谱可视化 (图存储与设计)；企业指标体系 (近 500 个)；最短路径计算。

## 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。

## 🏆 学校奖项

- |                     |                 |
|---------------------|-----------------|
| • 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 |

## 📖 学校论文

- **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)