

个人简介

两年半工作经验，有基于 C++ 的大型分布式 NoSQL 数据库项目开发经验。熟悉使用 Python 开发后台服务、自动化系统，进行数据分析、建模与可视化，有基于 React 的前端开发经验。作为主力工程师参与设计、开发风控决策规则分析系统、自动化风控解决方案生成平台、实时决策数据监控报警平台。

工作经历

PayPal · 数据分析与决策管理部门

数据工程师

2017 年 8 月 - 至今

- 参与风控数据分析 (SQL, Pandas)，并构建了风控决策规则全方位分析系统 (Flask, React)，包括规则解析引擎、近实时决策监控、决策规则性能指标评估、决策因子引用分析等模块，节省约 80% 的分析成本
- 参与设计、构建从损失报警到生成解决方案的自动化风控平台 HyperNet，自动生成的风控方案有效性不低于人工解决方案，缩短 2-3 周生成风控方案的周期，节省约 70% 人力成本，每月止损收益约 \$80k

HyperNet 自动化风控平台 - 前后端设计与开发

- 构建统一数据预处理平台，对不同源数据 (HDFS, Teradata) 进行 Spark ETL，进行特征统计分析与特征编码，整合生成训练与测试数据集
- 使用 XGBoost 构建决策树模型，按风控经验指标过滤决策规则生成风控方案
- 使用 React 构建前端损失报警与自动风控方案监控管理平台

Spider 实时决策数据监控报警平台 - 后端开发

- 开发报警任务执行系统，消费风险决策引擎产生的警报 (Apscheduler)，并调度执行事后聚合分析 (Pandas, Spark)、决策反馈、可视化 (Matplotlib)、邮件报警等任务。
- 基于 Celery, RabbitMQ 实现任务链和异步优先级任务调度，提供更高效稳定的报警服务

百度 · 基础架构部

研发工程师 · Mola 大型分布式 NoSQL 数据库

2017 年 3 月 - 2017 年 7 月

- 在 Proxyserver 中为百度云公网访问 MolaDB 服务添加标准访问日志 (C++)，添加请求 key 记录，有效缩短后台故障排查时间
- 解决 Node 中分片恢复过程 Task 修改异步引发恢复错误问题，有效降低了 Mola 服务的故障率

百度 · 生态数据产品部

实习数据工程师 · 精益-广告投放数据分析

2016 年 7 月 - 2016 年 10 月

- 精益对凤巢、Holmes 等广告数据进行监测与评估，建立漏斗模型 (人群转化、流动、归因分析) 分析广告营销活动效果
- 对上游广告源数据进行 ETL(Hive) 及数据拼接、编写 Spark UDF，优化 ETL 速率，并线上验证数据分析准确性

荣石投资管理有限公司

实习算法工程师 · 量化投资策略分析 (毕业设计)

2016 年 2 月 - 2016 年 6 月

- 基于 python 实现 Hurst 指数、SMA 双均线突破、分类预测择时策略
- 提出基于最大信息熵的股票特征选取方法，对比主成分分析、奇异值分解等方法高出约 4.5% 的预测精确度
- 提出 SRA-Voting 混合二元分类模型，对比 SVM、随机森林、AdaBoost 高出约 3.125% 的预测精确度

人行道网络信息技术有限公司

实习软件工程师 · 东方名城社区系统

2015 年 6 月 - 2015 年 12 月

- 参与开发以五角场镇为试点的社区治理系统 (社区治理服务、邻里社交、社区公共服务)，主要负责系统需求分析、移动端原型设计、Android 应用开发、部分 Java Web 后台服务 (SSM) 开发

技能

编程语言 Python, C++, Java, SQL, Shell, JavaScript

框架技术 Pandas, Flask, Celery, React, Redux, Spark

工具 Teradata, MySQL, MongoDB, Emacs, Git, Jenkins, Docker, Excel, Tableau

英文 CET6, GRE, 良好听说读写能力