

罗嘉濠 (Edward)

基本资料

年龄: 33 岁

邮箱: jiahao.luo2014@gmail.com

期望职位: 机器学习高级工程师

期望年薪: 面议

教育经历

Georgia Institute of Technology (Atlanta, USA)

08/2015 – 05/2017

- 硕士 / 计算机科学与工程 GPA: 3.4/4.0
- 核心课程: Machine Learning, Artificial Intelligence, Computational Data Analysis and Modeling and Simulation.

华南理工大学

09/2010 – 07/2014

- 优秀毕业论文 (典型城市风气候分析及其在建筑抗风中的应用) / 国家励志奖学金 / 利海奖学金

主要技能

- 编程语言: Python / Scala / Java / SQL
- 深度学习框架: 熟悉 PyTorch / Tensorflow
- 大数据处理框架: 掌握 Apache Spark / Hive
- 其他: Git / Tableau / Linux / Docker

工作经验

机器学习高级工程师 – 顺丰科技有限公司

07/2017 – 至今

- 对海量的物流数据进行数据提取、清洗和转换。
- 根据不同的业务应用场景需求, 深入分析与洞察, 提炼和发现业务规律, 指导模型特征构建, 定位产品相关数据问题及分析优化。
- 利用机器学习等技术挖掘海量物流运单数据, 开展理赔特征分析与洞察, 构建理赔风控预警模型。
- 开展理赔成因评估, 构建和优化客户画像, 物流网点画像以及快递小哥画像, 并进行检验评价。
- 面向探索性项目, 采用业界不同的流行算法进行探索和研究。

解决方案架构师 – (SSP) Southeastern Security Professionals, LLC

12/2014 – 07/2015

- 负责安防网络与数据库的智能化管理, 通过机器学习算法分析监控视频流、设备日志等数据, 优化网络传输效率与数据存储安全性。
- 运用数据挖掘技术, 深度分析产品使用痛点、安全隐患及用户潜在需求, 输出数据驱动的产品迭代策略, 助力安防解决方案优化升级。
- 运用空间数据分析与建模技术, 结合场地监控覆盖范围、设备性能参数等数据, 优化安防设备布局与预警算法, 提升区域安全管理效率与风险响应能力。

比赛经验

技术嘉年华算法挑战赛 (冠军 1/53) – 顺丰科技有限公司

10/2019 – 12/2019

- 顺丰产品推荐: 基于顺丰亿级运单信息, 根据历史记录判断并推荐给客户应采用的顺丰速运产品。主要负责特征工程, 采用了 LightGBM 模型训练, 以及相关的样本平衡方法, 最后部署 docker。
- 顺丰客服对话意图识别: 根据客服对话 (文字), 通过自然语言处理技术, 识别出客户来电意图。主要负责对话文本多方式增强, 采用了 BERT-base 模型训练, 以及多模型融合, 最后部署接收 API 请求的 client 端。
- 除了作为比赛的主要参与者, 还承担了团队 team leader 角色, 组建队伍, 任务安排, 制定时间轴等。

项目经历

物流理赔风险预警项目 (Apache Spark - Scala / Hive SQL / XGBoost – Random Forest)

07/2017 – 至今

项目背景

由于导致快递件发生理赔的原因可能出现在快递件运输的各个阶段，为此，基于快递件理赔大数据，对快递在揽件阶段、运输阶段以及理赔阶段整个流程进行分析，并对各阶段的主要实体进行画像构建。（收寄件客户画像、托寄物画像、运输中转网点画像、快递员画像以及理赔专员画像）

项目描述

- 基于快递件理赔大数据，对快递在揽件阶段进行分析，并以收寄件客户为实体进行画像构建。根据业务方需求，在理赔阶段的快递件采取关键字规则匹配，建立高风险客户黑名单。
- 建模阶段，针对托寄物的品类对个别地区采取分地区建模，如生鲜水果大省海南、山东等，大闸蟹产地江苏苏州、江苏泰州等，分地区建模能有效提高模型的效果。
- 根据模型的理赔风险预警值，在揽件阶段进行开箱拍照、身份证校验等操作，从而降低理赔风险，减少理赔成本。

项目职责

- Hive 大数据平台上进行数据提取、分析、准备与清理，提取基本特征。
- 基于 Spark/Scala 对预处理后的数据进行特征工程，丰富特征。主要以时间维度和空间维度进行特征扩充，时间维度包括按不同时间窗口对目标特征进行统计，空间维度包括对客户寄收件地址变换统计等，并利用 XGBoost 和随机森林算法进行建模，得出理赔风险预警值。
- 结合业务需求，对理赔风险预警值进行合理划分，平衡召回率和精确率，后续模型上线后，特征不断优化和改进。

项目成果

- 目前项目快递员揽件阶段以及理赔员理赔阶段建模已上线运行，并发表了以下两项专利。
- 专利号 201810067080.0 《一种基于 XGBOOST 算法的快件理赔预警技术》，专利号 201810067068.X 《基于随机森林算法的理赔预警模型》

物流网点建模项目 (Apache Spark - Scala / Hive SQL / XGBoost – Random Forest)

07/2018 – 至今

项目描述

- 基于快递件理赔大数据，对快递在运输阶段进行分析，并以运输中转网点为主要实体进行画像构建。
- 建模阶段，针对托寄物的品类对个别地区进行分地区建模，如生鲜水果大省海南、山东等，大闸蟹产地江苏苏州、江苏泰州等，分地区建模能有效提高模型的效果。
- 根据模型得出的理赔预警值，在运输阶段进行暴力分拣检测操作、中转网点评分制等措施，从而定位理赔原因，控制理赔风险，降低成本。

项目职责

- Hive 大数据平台上进行数据提取、分析、准备与清理，提取基本特征。
- 基于 Spark/Scala 对预处理后的数据进行特征工程，丰富特征。主要以网点维度和运输路线维度进行特征扩充，中转网点维度指的是对不同网点代码的目标特征进行统计，衡量不同网点的优劣，运输路线维度指的是对全网路线的目标特征进行统计，从而发现出理赔风险较高路线。最后利用 XGBoost 和随机森林算法进行建模，得出理赔风险预警值。
- 结合业务需求，对理赔风险预警值进行合理划分，后续模型上线后，特征的不断优化和改进。

项目进度

- 目前已经完成物流网点基础画像刻画，并进行模型构建中。

大数据物流图谱 (Python / Apache Spark – GraphX / JavaScript – D3.js)

03/2018 – 12/2018

项目描述

- 基于历年物流数据，以不同的寄件和收件方为节点，寄收件关系为边，寄收件数量、占比、金额等为权重，构建物流大数据图谱，研究探索不同公司客户之间的物流关系（供应链关系、金融关系）。

项目职责

- 基于 NetworkX Python 库对提取数据进行物流图谱构建，计算图谱相关结构特性，如节点的 PageRank 值、经过每个节点的三角形数目等。
- 基于 SIR 理论模型，建立图谱节点间的风险传播模型，模拟随机风险对图谱中其余节点的影响，探索不同公司客户间风险传播的影响及抗风险能力。
- 利用 D3.js，对提取建立的物流图谱进行可视化。