

谢 坤

Room 5569, Academic Building, HKUST, Clear Water Bay, KLN, HK
<https://cnxiekun.github.io>

Tel: (852) 9587-5869
kxie@connect.ust.hk

教育背景

- 香港科技大学 香港
 - 工业工程与决策分析学博士 9/2014 – 10/2018 (预计)
 - 相关课程: Machine Learning, Algorithms, Deterministic Models in OR, Stochastic Models in OR, Advanced Production Planning and Control, Statistics, Stochastic Calculus, Convex Optimization 等
 - 课程助教: Machine Learning, Engineering Probability and Statistics, Service Engineering and Management, Global Supply Chain Management, Financial Engineering and Risk Management 等
- 南京大学 江苏省南京市
 - 数学系理论数学学士 9/2010 – 6/2014
 - 相关课程: 实变函数, 数学分析, 高等代数, 概率论, 随机过程, C++, 数值计算, 数值分析

研究兴趣

- 机器学习, 算法, 运筹优化与随机模型, 数据挖掘与分析

科研及项目经历

- 无穷维状态空间的连续马尔可夫链在股票市场的应用 香港
 - 博士毕业论文课题, 指导老师: Jiheng Zhang 教授, 工业工程与决策分析学系 1/2015 – 10/2018 (预计)
 - 观察真实数据, 用泊松过程对股票市场的买家卖家的各类决定建立数学模型
 - 利用排队论, 概率论尤其是大数定理, 中心极限定理的知识求得在极限 (平稳) 状态下买家卖家的最优价格的分布
 - 与导师关于该课题的研究尚未完成, 仍在进行中
- 供应链排产 香港/广东省深圳市
 - 华为技术有限公司合作项目, 项目开发人员 11/2017 – 10/2018 (预计)
 - 华为公司是全球主要的通信设备与通信技术解决方案供应商之一, 其产品包含 8 万种有复杂 BOM (Bill of Material) 树关系的半成品, 需要在考虑产能、运输等约束下实现每天的订单在数十个加工地点、数百条产线上的多目标优化排产
 - 基于各类约束条件以及目标函数, 建立合理的数学模型; 受限于问题规模, 常用的 Solver (如 Gurobi) 无法在合理的时间内求解; 利用分解思想, 将原问题的求解分为四步, 其中每步都能在短时间内建模求解
 - 算法还在开发以及完善中, 预计最终会在华为内部申请潜在高价值专利, 同时投稿 IEEE
- 急单预测 香港/广东省深圳市
 - 华为技术有限公司合作项目, 项目开发人员 11/2017 – 10/2018 (预计)
 - 华为 60% 以上订单的交货周期小于整个产品的生产和组装周期, 生产计划需要根据对未来的预测, 设计合理的生产提前量以较高的半成品库存周转率保证订单满足率
 - 基于历史订单量的随机序列, 建立机器学习模型并提取重要特征; 受集成算法思想 (如随机森林) 的启发, 设计了一个基于数据集成的通用算法, 不受限于模型的选择, 且极大地提高了预测的准确性
 - 与华为原有算法相比, 无论是“绝对误差”、“相对误差”或是“均方差”都至少提高了 30%
 - 算法的思想步骤已在华为内部申请了潜在高价值专利; 算法的效果也有理论分析支撑, 预计投稿 IEEE

- 销量预测** 香港/广东省深圳市
 - 顺丰科技有限公司合作项目，项目开发人员 4/2017 – 10/2017
 - 顺丰拥有各个品类的海量运单数据，需对各品类在各城市未来一段时间的件量进行精准预测
 - 利用 Hive 以及 MongoDB 预处理及存储数据，随后人为选取及添加一些重要特征，利用 Scikit-Learn 中的 RandomForestRegressor 建立模型，并进行训练以及交叉验证
 - 其中特征的选取以及模型参数非常重要，不能一味地将 Scikit-Learn 当成一个黑盒使用
 - 最终城市覆盖率的预测精度从 62.8% 提高到了 90.7%，件量预测值的平均绝对误差从 15.2% 降低到了 3.7%，目前该算法已在顺丰内部上线使用
- 商圈潜客** 香港/广东省深圳市
 - 顺丰科技有限公司合作项目，项目开发人员 4/2017 – 10/2017
 - 顺丰拥有众多商户，同时也积累了海量的线下数据，希望以此帮助这些商户识别其已有客户的特征及商业价值，并为其挖掘潜在客户，最后中国地图上绘制人群分布图
 - 先利用 KMeans 将商户已有的客户聚类为忠实以及摇摆两类客户群，并将他们在各品类的消费习惯作为特征向量，随后利用 RandomForestClassifier 去顺丰余下的海量数据中匹配出消费习惯相似的两类人群
 - 利用百度开源的纯 JavaScript 图表库 ECharts 及全国各区/县的经纬度信息绘制各类人群的地域分布图
 - 基于顺丰的具体业务要求，数据的预处理、特征多少的选取以及模型参数的调整显得尤为重要
- 风险客户识别** 香港/广东省深圳市
 - 顺丰科技有限公司合作项目，项目开发人员 4/2017 – 10/2017
 - 顺丰的保价理赔流程相对简易，部分客户存在恶意索赔行为，需根据诈保历史数据，利用大数据分析识别出风险客户，为理赔人员在理赔生成初期提供预警
 - 利用 Hive 数据库统计分析出理赔的钱主要出自哪里，以此为基础确定之后的特征选取标准
 - 正样本的缺失以及数据的极度不平衡性导致传统机器学习的算法准确率低，同时顺丰运营方面也难以执行，最终采用 RandomForest、Undersampling 以及 Mini-Batching 的方法训练并测试模型
 - 为顺丰提供多种解决方案以减少其每年的理赔金额，目前算法已在顺丰内部上线使用
- 顾客具有参考效应的排队竞争模型** 江苏省南京市
 - 本科毕业论文课题，指导老师: 沈厚才，吴婷教授，工程管理学系 9/2013 – 6/2014
 - 两家公司在市场标准存在的情况下对承诺给顾客的平均等待时间进行博弈，其中市场标准是内生的，取决于两家公司的决定
 - 其中一家公司是领先者 (leader, 先决定其顾客的平均等待时间)，另一家是跟随者 (follower)，公司的顾客率取决于该公司的等待时间与市场标准的差距，公司的最终目标是最大化其收益
 - 利用排队论及博弈论的知识，我们证明了当市场标准等于两家公司给出的较短的等待时间时 Stackelberg Equilibrium 的存在唯一性，并给出了其精确解
- 网上选课及数据库在线查询** 江苏省南京市
 - 数据库课程项目，指导老师: 吴朝阳教授，数学系 10/2011 – 12/2011
 - 使用 SQL 和 Python 语言编写了一个在线查询课程信息的数据库
 - 了解系统的数据结构，并从学校官方网站和数据库中提取所需数据

技能

- **编程语言:** 熟悉 Python (个人项目主力语言); 了解 Java, JavaScript, MatLab, SQL, JSON, CSS, HTML
- **平台:** Scikit-Learn, TensorFlow, Gurobi, MySQL, MongoDB, Apache Hive, Gaode Map API, GitHub
- **语言:** 普通话 (母语), 英语 (流利)

荣誉奖章

香港科技大学博士生全额奖学金	2014 – 2018
优秀毕业生 (南京大学颁发)	2014
二等人民奖学金 (前 5%，中国教育部颁发)	2012, 2013
数学系拔尖计划奖学金 (南京大学数学系颁发)	2011, 2012, 2013
三等人民奖学金 (前 15%，中国教育部颁发)	2011
优秀学生干部 (南京大学颁发)	2011
江苏省数学竞赛一等奖 (随后被南京大学数学基地班提前录取)	2009

课外活动

数学学术交流，北京大学	8/2012
四校联合暑期数学夏令营，厦门大学	8/2011
首批南京大学拔尖计划学术共同体委员会成员	2/2011
南京大学数学系乒乓球队核心队员	9/2010 – 6/2014
南京大学数学系年级长	9/2010 – 6/2014