



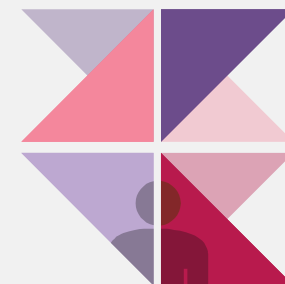
Profiling The World

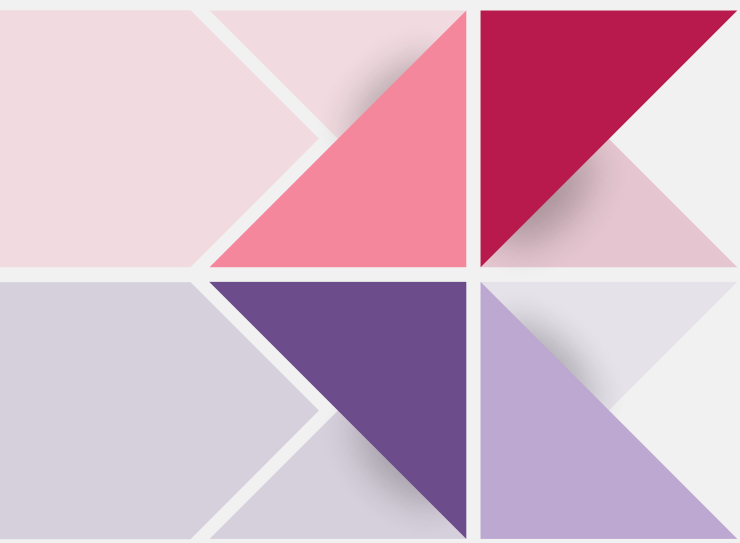
AI引擎研究 团队介绍

牛琨 博士/副教授

北京邮电大学软件学院移动计算与大数据分析挖掘技术研究中心

niukun@bupt.edu.cn





01

我们是谁

来自北京邮电大学的AI引擎研究团队

团队负责人：牛琨 博士

北京邮电大学软件学院 副教授 硕士生导师



个人介绍

- 北京市科技进步二等奖获得者；
- 第一批北京市高校青年英才；
- 北京邮电大学“周炯槃”优秀青年教师励志奖获得者；
- 发改委、财政部等项目评议专家，多家学术期刊/国际会议审稿人；
- 北京邮电大学大学生创新性实验计划优秀指导教师，教学观摩评比优秀奖获得者；
- 北京邮电大学师德标兵，北京邮电大学“我心中的好导师”。



科研概况

- **研究方向**：大数据分析挖掘、智能信息处理、大数据运维及其行业应用；
- **学术研究**：2005年至今以第一作者在国内外期刊会议上发表中英文学术论文50余篇，包括SCI检索6篇，EI检索29篇，授权中国发明专利3项，出版专著1部；
- **科研项目**：以项目负责人或项目主研人主持及参与科研项目共计30余项，项目金额累计达2127万元。

研究方向：AI与大数据

人工智能 & 大数据



我们擅长什么

AI系统、算法建模、数据分析

数据分析

03

数据科学的基础是对数据的系统分析，包括机器学习(含深度学习、强化学习)、统计分析、预处理等

AI引擎研发

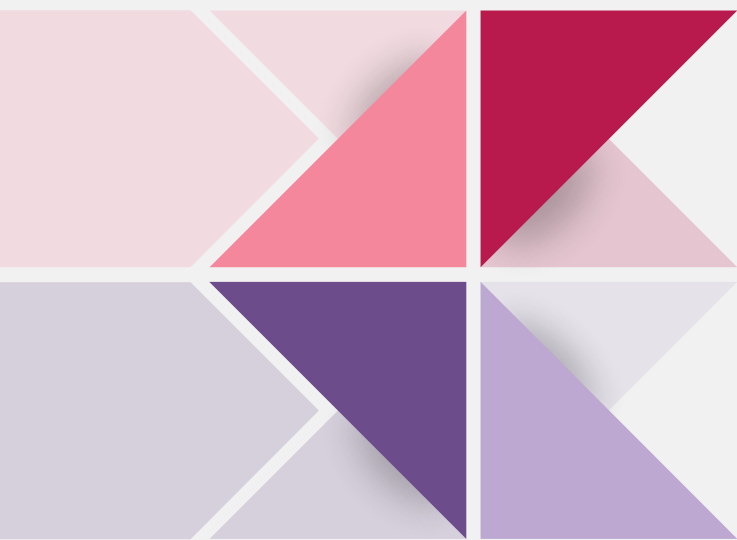
AI引擎一般包括用户画像、情景识别和决策系统三个主要部分

01

02

算法建模

基于定制化场景的建模是数据科学的桂冠，也是行业应用的核心竞争力



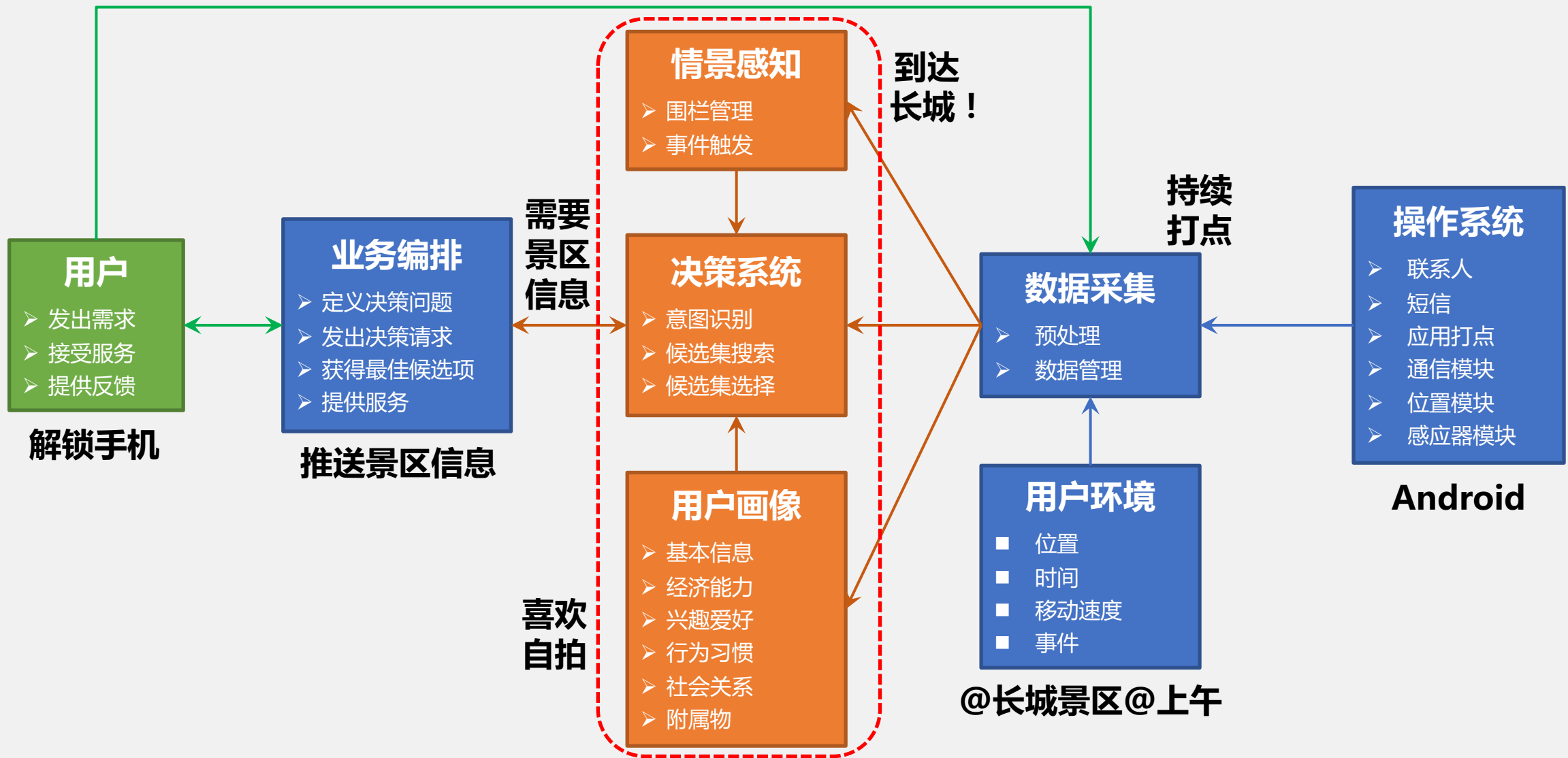
02

我们能做什么

AI引擎研发、大数据建模、行业应用

A引擎研发

用户画像、情景感知、决策系统



大数据建模

学习历史，判断未来

模型训练

通过构造模型，把历史数据中的规律知识进行结构化表述

例如：对已知产生离网/续约/订购/欺诈等行为的客户，通过模型建立聚类规则、分类器等模型，有效记录用于未来的知识

模型应用

运用模型，把规律用在行业应用上，向数据要效益

例如：利用模型预测得出未来产生某种标定行为的用户清单，按照概率排序执行业务动作，精准投放资源

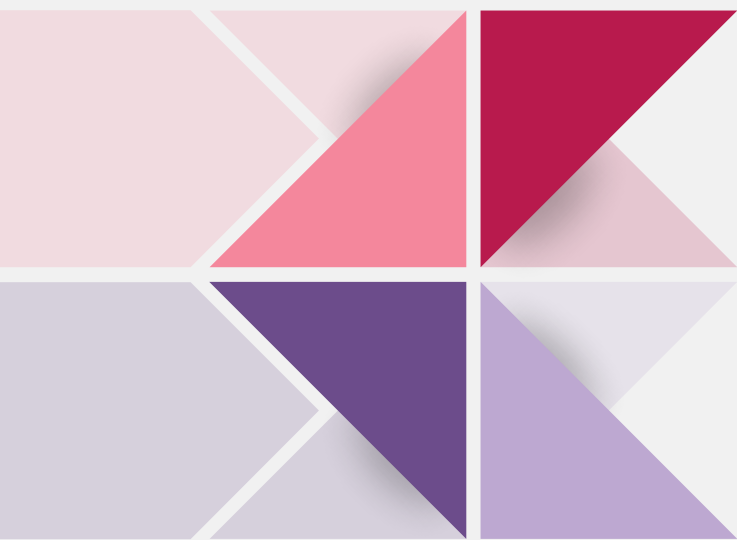
行业应用

用户洞察、用户维系、异网策反



- **聚类模型**：自然分组，物以类聚
- **分类模型**：预测目标类行为概率
- **预测/回归模型**：周期性预测时序
- **关联模型**：生成组合，判断元素关系

- **用户分群**：分而治之，因势利导
- **行为预测**：负向预警，正向激励
- **时序预测**：战略依据，战术素材
- **关系分析**：重点研判，牵引提升



03

我们做过什么

学术研究/应用/培训@电信/电商/金融/保险大数据

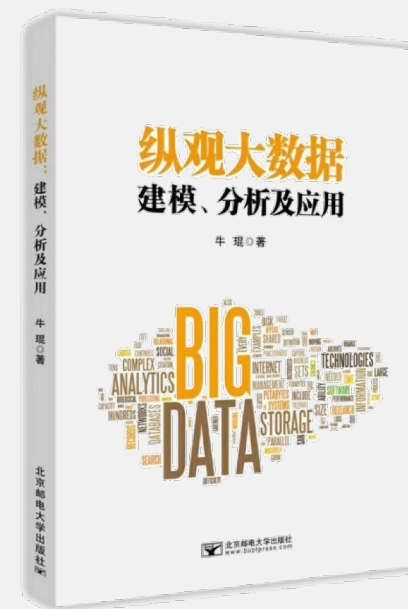
学术：发表论文50余篇

第一作者论文SCI检索6篇、EI检索29篇，授权3项中国发明专利，出版专著1部



学术研究主要成果

- **BTP: A Bedtime Predicting Algorithm via Smartphone Screen Status**[Wireless Communications and Mobile Computing]
- **Real-Time Taxi-Passenger Prediction with L-CNN**[IEEE Transactions on Vehicular Technology]
- **Predicting Ride-hailing Service Demand via RPA-LSTM**[IEEE Transactions on Vehicular Technology]
- **Subspace clustering for vector clusters**[Journal of Internet Technology]
- **A Novel Learning Approach to Improve Mobile Application Recommendation Diversity**[2018 IEEE International Conference on Data Mining Workshops]
- **A developed apriori algorithm based on frequent matrix**[ACM International Conference Proceeding Series]
- **K-means+: A developed clustering algorithm for big data**[International Conference on Cloud Computing and Intelligence Systems]
- 专利：**Web服务质量的分布式测量系统和方法**
- 专利：**分布式数据的一致性维护系统和方法**
- 专著：《**纵观大数据-建模、分析与应用**》



应用：聚焦大数据建模

涵盖科技部/市教委、电信运营商、互联网公司、创业公司



主研项目

- **【跨境电子商务大数据平台联合创新研发合作】** 国家重点研发计划
- **【自主网络品牌电子商务服务技术研发与应用示范】** 国家科技支撑计划
- **【物联网环境下多源上下文信息的协同认知理论和方法研究】** 国家自然科学基金
- **【物联网总体架构及关键技术研究】** 科技部重大专项
- **【云服务开发平台示范项目(二期)】** 百度在线网络技术(北京)有限公司
- **【基于电商云平台的大数据知识发现及应用系统】** 北京京东尚科信息技术有限公司
- **【物联网大数据处理及分析方法研究】** 北京市教委高校“青年英才计划”
- **【用户信用度评估模型课题咨询服务】** 中国移动(深圳)有限公司
- **【漫游用户行为分析研究】** 中国移动(深圳)有限公司
- **【数据监控分析与广告推荐方法研究】** 北京赛弗网络科技有限责任公司

培训：深耕电信行业

涵盖电信运营商、电商公司、咨询公司



培训经历

- **中国电信**集团公司、北京研究院、甘肃、宁夏、湖南、广东、陕西、山东、广西、内蒙古、新疆省公司
- **中国移动**研究院、卓望科技、四川、贵州、云南、广东、陕西、山东、江苏、湖北、浙江、湖南、江西、山西省公司
- **中国联通**集团公司、研究院、广东、安徽、山东、河南、辽宁、山西、河北省公司
- 北京益普索市场咨询有限公司
- 新疆广电网络有限责任公司
- 中证资本市场运行统计监测中心
- 北京市国家安全局
- 北京友谊医院
- 清华大学附属小学



案例：跨境电子商务大数据平台联合创新研发合作

国家重点研发计划[2016YFE0204500]

本课题预期解决跨境电商中因宗教、语言、反恐等差异带来的问题，如商品展示中涉及的语言禁忌、反宗教图片识别和过滤等；解决跨境电商中遇到的风控问题，如反欺诈、商家、个人的信用体系评价等；解决跨境电商中因用户群体差异、技术手段不足带来的精准营销等问题。



实时数据采集及计算

突破基于数据库日志实时解析的实时数据采集技术，实现多源异构数据库间的数据实时采集，并建设实时数据计算平台。



建设机器学习平台

突破特征工程、算法任务监控、算法任务管理、算法任务执行引擎等机器学习算法应用中的技术问题，构建一站式的机器学习平台，提供机器学习算法的统一管理及应用服务。



构建一站式的数据服务体系

实现数据查询引擎、数据可视化产品体系，提供快速的数据查询服务和丰富的、定制化的数据产品和工具。

案例：自主网络品牌电子商务服务技术研发与应用示范

国家科技支撑计划[2015BAH03F00]

该课题针对电商平台中的大数据应用展开研究，主要包括电商大数据平台架构设计及部分面向电商特定主题的数据分析算法研究。

任务 01

软件服务架构

开发适用于电子商务行业应用的，基于云计算技术的大数据应用、分析和开发软件体系架构

任务 02

设计与测试平台

建立起覆盖大数据分析全生命周期的设计、开发、测试和部署平台

任务 03

服务能力要求

支持大数据多维分析、大数据挖掘、大数据服务开发

任务 04

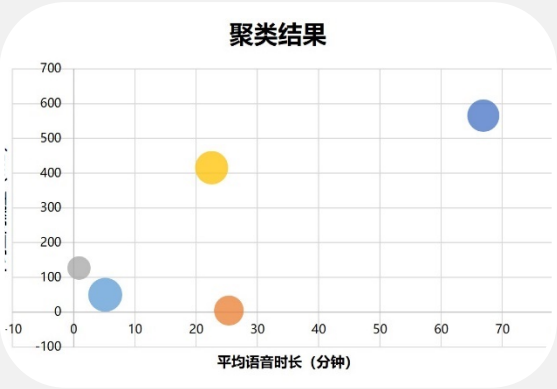
应用验证

通过京东商城在电子商务上的典型应用，验证平台的大数据存储、管理、计算、建模、多维分析、挖掘、应用扩展以及服务化应用能力

案例：漫游用户行为分析研究

中国移动(深圳)有限公司

项目拟通过分析客户运营商的用户的漫游清算数据，获得其漫游用户的漫游通信行为特征，识别出漫游领域潜在的高价值用户和通信轨迹异常用户，优化用户与业务匹配性，从而提高漫游业务量，达到**为客户运营商提升收入，为清算中心提升清算服务收入**的目的。



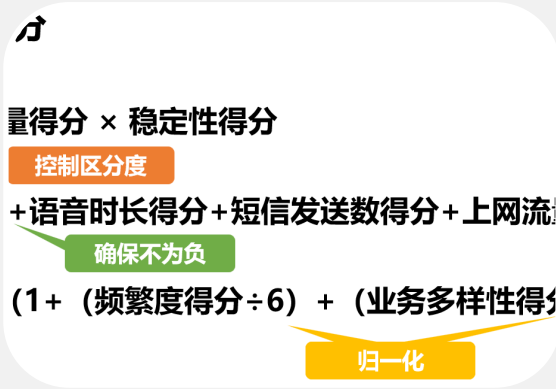
出访用户细分模型

- 对某国际运营商的100多万出访型国际漫游用户通过聚类得出了细分模型



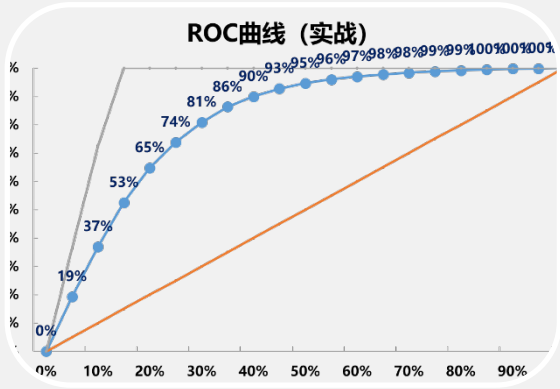
客户群特征总结

- 对各客户群进行特征总结
- 最终输出营销策划建议



价值评分模型

- 对每个用户进行科学的价值评分，综合考虑业务量和稳定性
- 结合细分模型定位各客户群价值



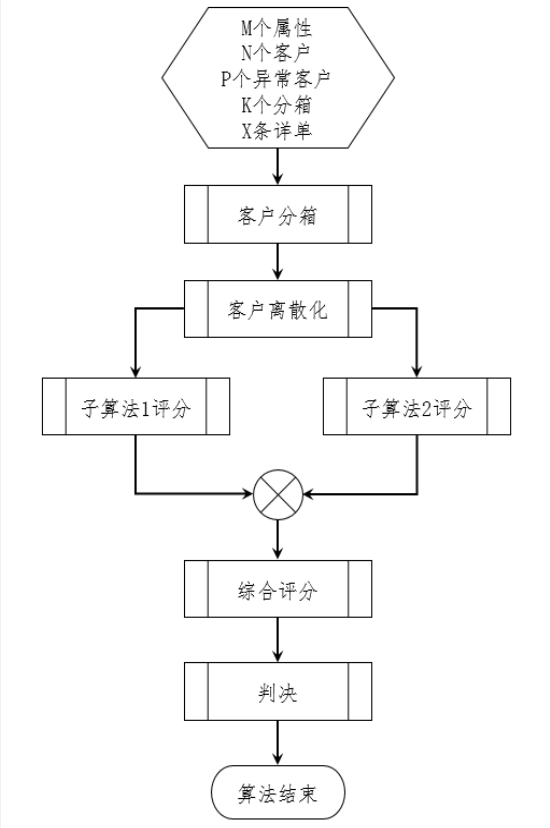
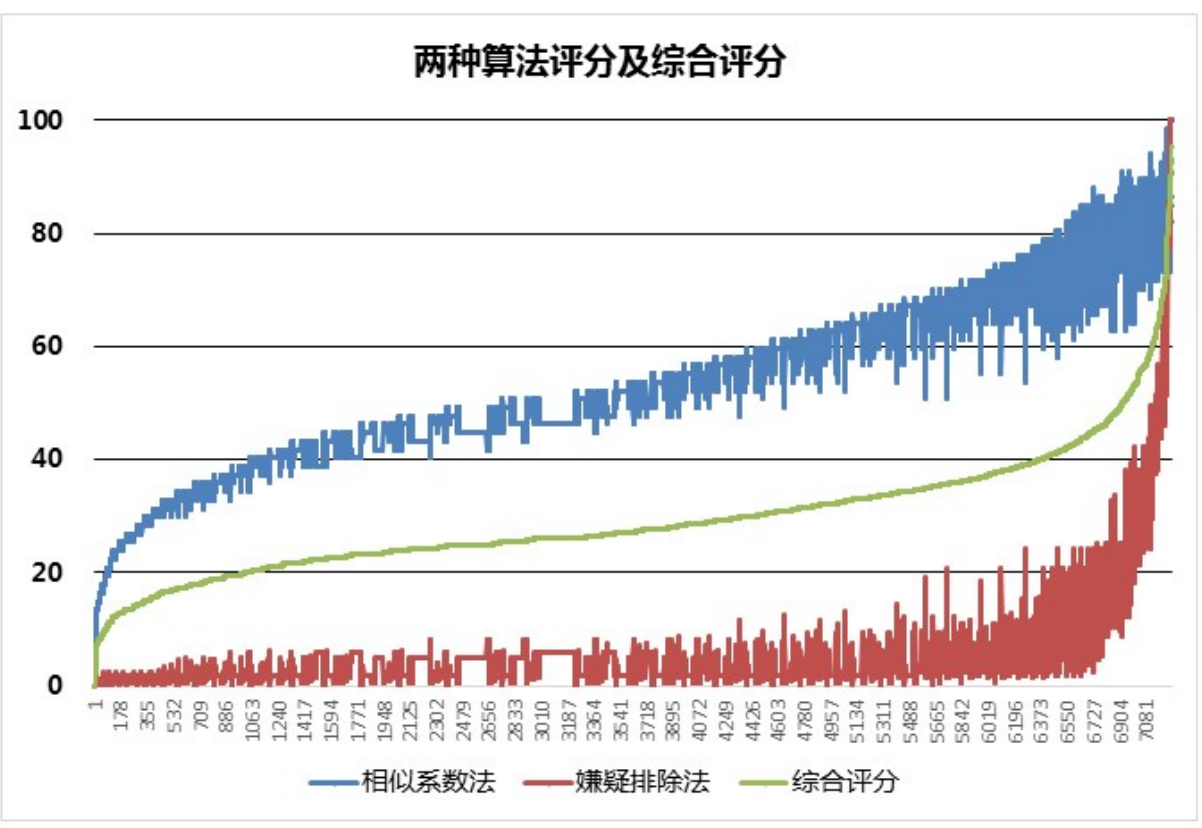
异常用户预警模型

- 建立预警模型，对下个月产生的异常行为生成名单并验证
- 5%处敏感度是19%，提升到3.8倍，准确率54.9%

案例：用户信用度评估模型课题咨询服务

中国移动(深圳)有限公司

建立一套用户风险信用模型，通过该模型，评估用户当前欺诈欠费风险，并对用户欺诈欠费风险进行量化和评级，以便于根据风险评估结果采取差异化处理流程。



核心价值

- 传统预警算法局限性：只能依托现有欺诈模式，不能兼顾敏感性和准确性
- 本算法优势在于综合采用了基于加法和基于乘法的两种子算法，表现更均衡
- 采用分箱和离散化设计把异常行为特征化，能直接识别新的欺诈模式，只要该行为与总体不够同步

案例：物联网大数据处理与分析方法研究

北京高校“青年英才计划”[YETP0453]

本课题的预期目标是聚焦物联网大数据计算，进行数据预处理、低密度关键信息挖掘两项内容的研究，提出行之有效的数据预处理与知识发现方法。

01. 大数据预处理方法研究

1. 系统学习相关概念和基础技术理论；搭建本地数据库；
2. 基于Java开发平台开发数据提取和装载程序，并测试可用性和容量；
3. 对物联网大数据集进行探索性分析，研究非结构化或半结构化数据的预处理方法；
4. 基于该方法，开发相关程序，测试该预处理方法的有效性和复杂度

02. 低密度数据重要信息挖掘研究

1. 采用简单的搜索和查找算法探索性研究物联网大数据；
2. 针对关键信息的分布特征提出对关键信息的挖掘方法；
3. 对物联网大数据集的切片测试关键信息挖掘方法有效性；
4. 在实验基础上确定针对物联网大数据的关键信息挖掘算法

03. 原型仿真系统开发

1. 根据软件工程理论进行设计；
2. 实现物联网大数据的映射化简处理能力及其它功能，同时具备对低密度存在的指定关键信息进行快速挖掘能力



Thank you

@牛琨 : 18911815860 , niukun@bupt.edu.cn