



用户画像介绍和在金融行业的应用

汇付天下 裔隽

2016.11 v1.1.1

目录

01. 用户画像概况

02. 用户画像体系构建

03. 行业应用案例介绍

04. 汇付天下用户画像应用

01. 用戶画像概況

■ 人格的研究



- 人格

指人类心理特征的整合和统一体，是一个相对稳定的结构组织。并在不同时间和地域下影响着人的内隐和外显的心理特征和行为模式 (Wikipedia)

- 人格心理学

心理学的分支之一，系统解释或说明各种人格现象，研究内容包括人格结构、人格动力、人格发展和人格适应等

弗洛伊德：精神分析论

班杜拉：社会认知论

马斯洛：人本主义

埃森克：生物学派

■ 人格心理学



• 人格心理学主要理论

- 体液学说：古希腊 希波克拉底 提出，将人的气质分为多血质、粘液质、胆汁质和抑郁质
- 人格分析理论：1913 年由瑞士的 荣格 提出，将人格划分为内倾型和外倾型，并结合感觉、思维、情感和直觉四种基本机能，得出了性格的八种机能类型
- 大五模型 (Big Five)：从外向型、神经质、和善性和严谨自律性对人格进行分类
- 星座、生肖和血型等分类方法

■ 人格心理学与用户画像



人格心理学

研究现实的人，探索、描述和揭示个人思想、情绪及行为的独特模式，并综合诸多足以影响个人的各种环境交互作用的过程，进而对现实社会中的个人作整体性介绍。

用户画像

通过用户调研去了解用户，根据他们的目标、行为和观点的差异，将他们区分为不同的类型，然后每种类型中提取出典型特征，进而为企业应用提供数据基础。

目前用户画像的应用，正是依赖于假设人们会因为接近的地域、教育和环境影响等等因素形成类似的事物偏好和消费习惯等。

■ 用户画像的定义

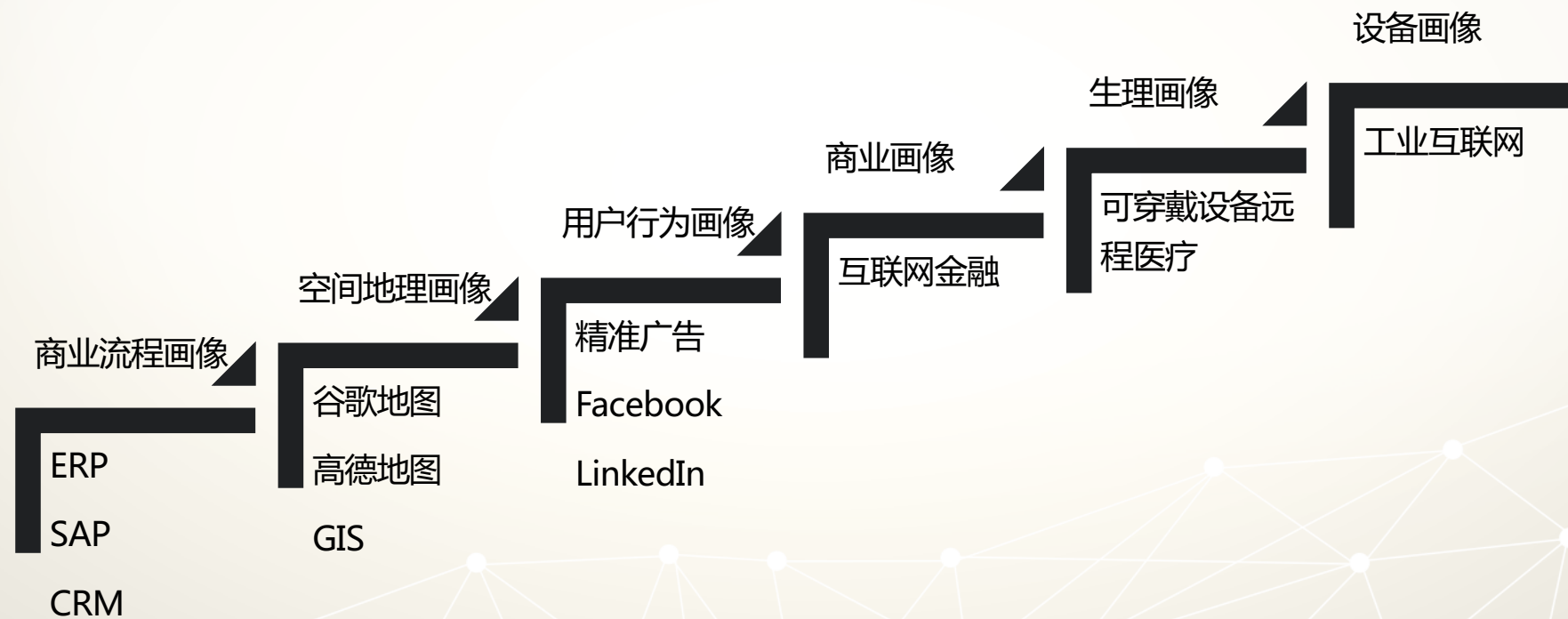


- 用户画像

真实用户的虚拟代表，是建立在一系列真实数据之上的目标用户模型。(交互设计之父 Alan Cooper)

■ 用户画像的发展

随着互联网金融的迅速发展，大数据技术得到日益广泛的应用，大数据时代需要数据化的画像。



■ 大数据时代的数据特征

传统数据技术

大数据技术

以结构化为主

数据
类型

非结构化，结构化

样本模式

分析
范围

全数据模式

批量式分析，事后见效

分析
速度

实时分析，立竿见影

反映整体情况，描述现状

应用
特点

支持个体差异分析，预测未来

■ 大数据时代用户画像的特征

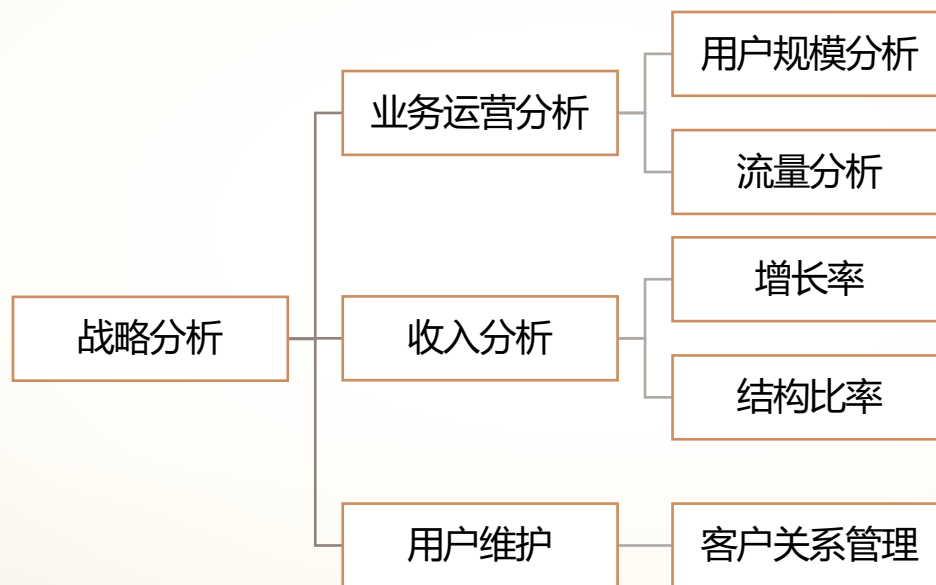
- 整合多种类型数据，用户画像更全面和更细分
 - 结构化数据
 - 非结构化数据
- 依托大数据技术，用户画像实现实时更新和快速查询
 - Hadoop 生态圈
- 结合丰富的业务场景，用户画像应用更创新
 - 精准营销
 - 在线实时推荐



■ 大数据时代用户画像的意义

- 战略层面

帮助企业分析其各业务的经营状况，收入的用户构成情况，从而加强与用户关系



■ 大数据时代用户画像的意义

• 产品层面

优化产品功能，改善用户体验，实现精细化运营等

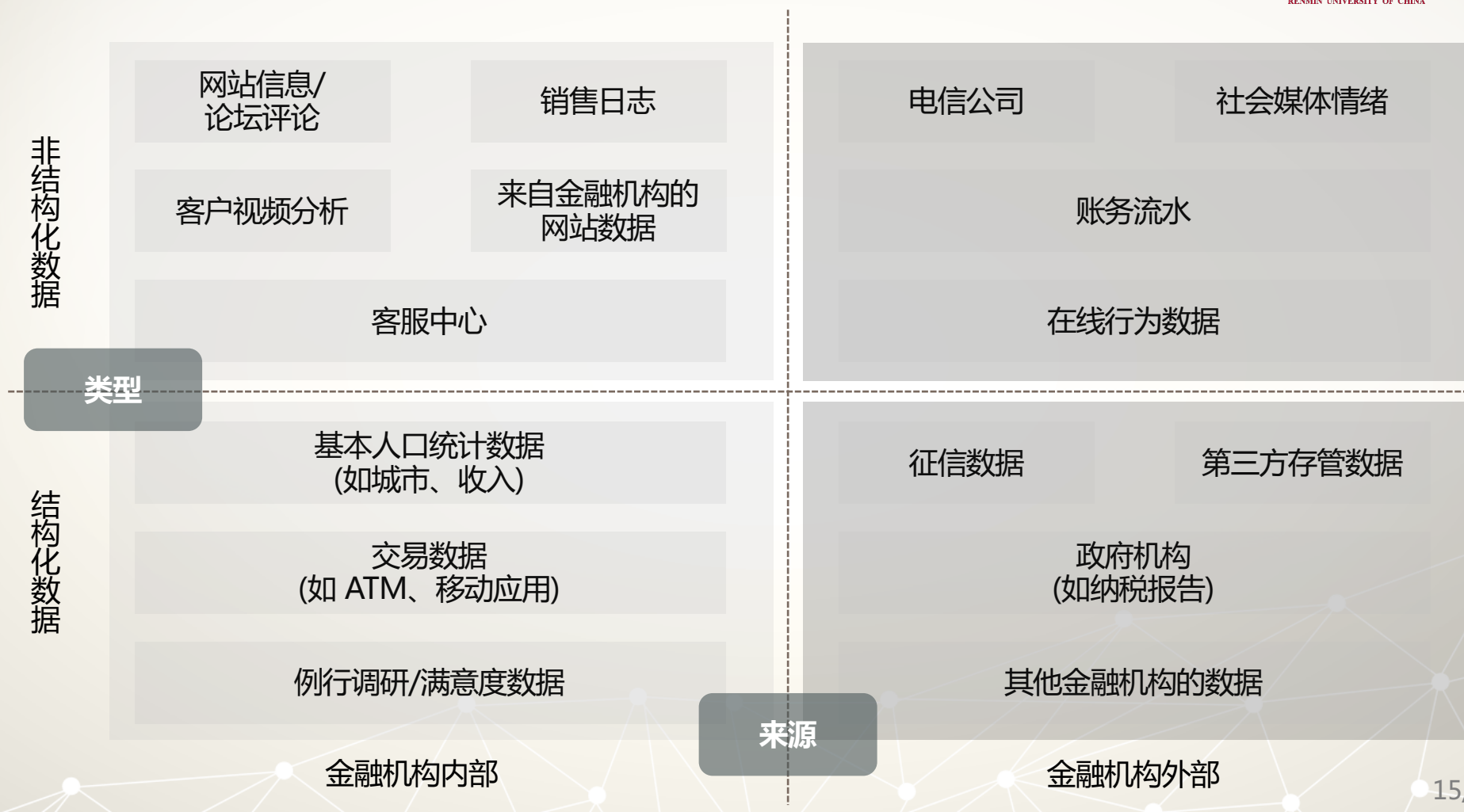


02. 用戶画像體系构建

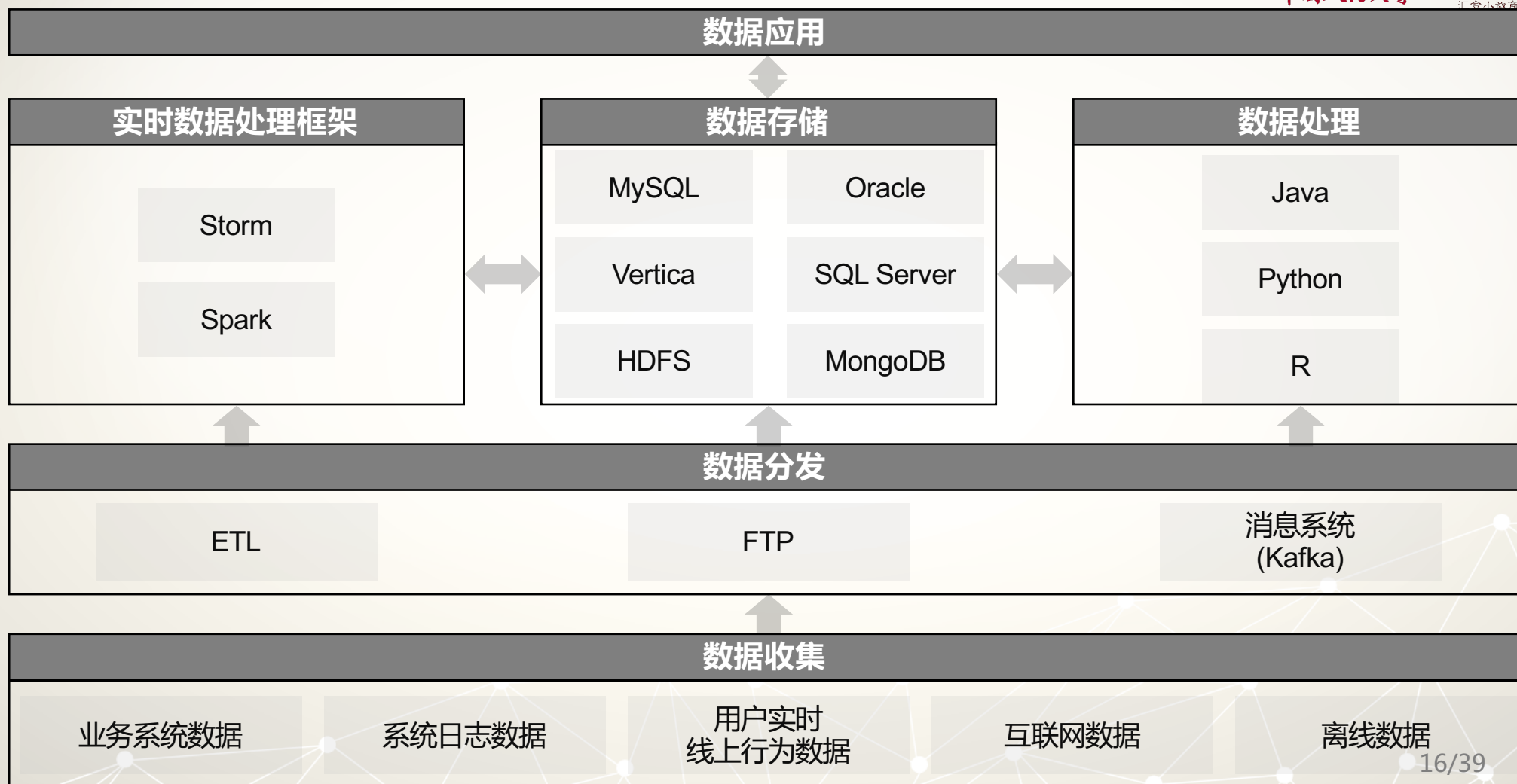
■ 用户画像体系构建—构建流程



■ 用户画像的构建流程—基础数据收集



■ 用户画像的构建流程—数据集市



■ 用户画像的构建流程—行为建模



- 监督学习

利用一组已知类别的样本调整分类器的参数，使其达到所要求性能的过程，也称为监督训练或有教师学习。常见的监督学习算法包括回归分析和统计分类

- 无监督学习

与监督学习相比，训练集没有人为标注的结果。常见的无监督学习算法有聚类

- 半监督学习

利用少量的标注样本和大量的未标注样本进行训练和分类的问题。常见的半监督学习算法有自训练算法和生成模型

- 增强学习

通过观察来学习做成如何的动作。每个动作都会对环境有所影响，学习对象根据观察到的周围环境的反馈来做出判断

■ 用户画像体系构建—标签体系

事实标签

人口属性统计

账户历史趋势

产品购买时间

产品购买次数

渠道使用频率

...

模型标签

产品购买偏好

用户活跃度

用户风险评分

渠道使用偏好

用户关联关系

...

预测标签

消费能力

违约概率

用户流失概率

人群属性

用户近期需求

...

■ 用户画像体系构建—标签体系举例

姓名：洪某某

性别：男

年龄：32 岁

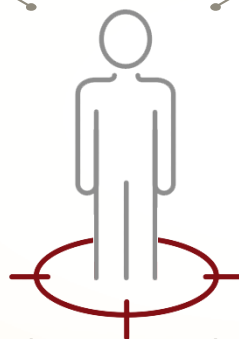
地区：江苏镇江

投标金额：约 360 万元

注册平台个数：10 个

基金投资金额：约 20 万元

贡献手续费：约 3 千元



用户类型：重要发展用户

活跃等级：6

偏好产品收益率：10%-15%

基金偏好产品：股票型基金

身份证是否实名验证：是

绑定银行卡数量：1 张

是否通过人脸验证：否

■ 用户画像应用场景



- 行业报告与用户研究

通过对用户画像的分析可以了解行业动态，进而可以指导平台更好的运营和把握大方向，同时也能给相关公司提供细分领域的深入洞察

- 优化产品设计

改变原有的“先设计再销售”的传统模式，在研发新产品前，先基于产品期望定位，在用户画像平台中分析该用户群体的偏好，有针对性的设计产品，从而改变原先新产品高失败率的窘境，增强销售表现

- 精准营销

针对已有产品，寻找所偏好的精准人群分类，以及这些人群在信息渠道和购买渠道上的分布比例，来决定广告投放和活动开展的位置和内容等，实现精准营销

03. 行业应用案例介绍

■ Netflix 用户画像应用案例



- Netflix 大赛

2006 年 Netflix 设立一项大赛，公开征集电影推荐系统的最佳电脑算法，第一个能把现有推荐系统的准确率提高 10% 的参赛者将获得一百万美元的奖金

基于推荐算法，以历史评价超过 50 个的观众作为样本，准确预测他们的口味

从获胜的团队中得到了现有算法 10.06% 的提升

第二个百万大赛，为不经常做影片评级或无评级数据的客户推荐影片

数据集有一亿条，包括评级数据、顾客年龄、性别、居住地区邮编和以前观看过的影片等信息。要求使用一些隐藏着观众口味的地理数据和行为数据进行预测

因为私隐问题，Netflix 在 2010 年 3 月宣布取消了第二个百万美金大奖赛

■ 百分点用户画像应用案例—标签体系



• 百分点公司介绍

百分点拥有全国最大的第三方消费偏好数据平台

百分点推荐引擎是一款提升电子商务零售网站整体营销性能的个性化推荐工具

核心产品包括技术层的大数据操作系统，管理层的用户画像标签工场，以及应用层的推荐引擎、分析引擎和营销引擎

• 百分点标签体系



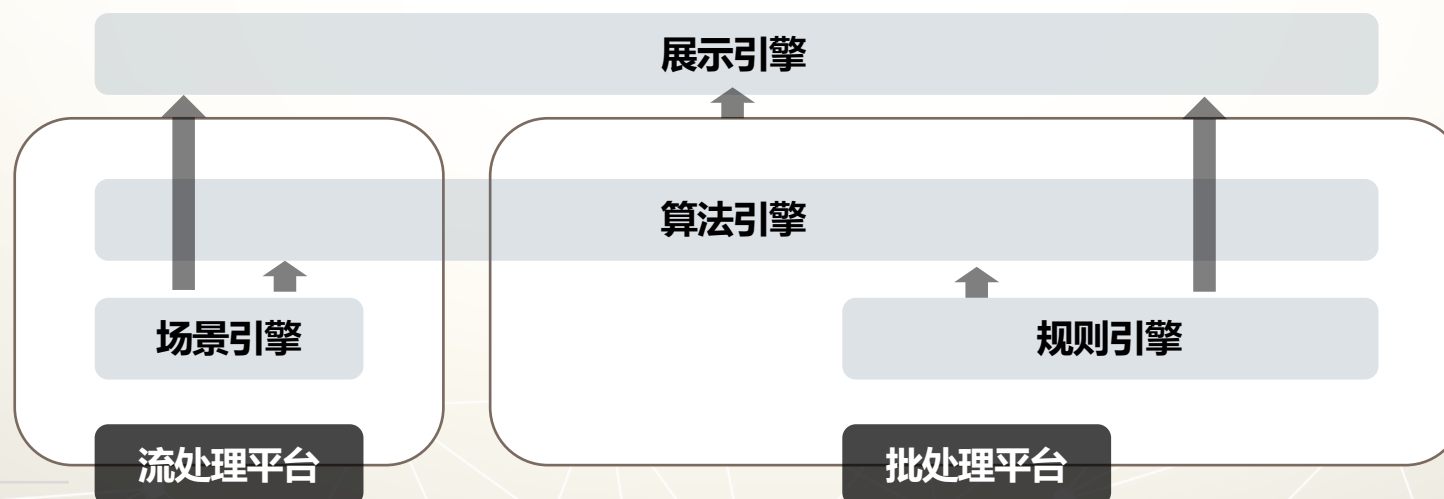
■ 百分点用户画像应用案例—推荐系统

场景引擎：个性化的核心，判断用户处于哪个购物环节，有什么样的购物目标

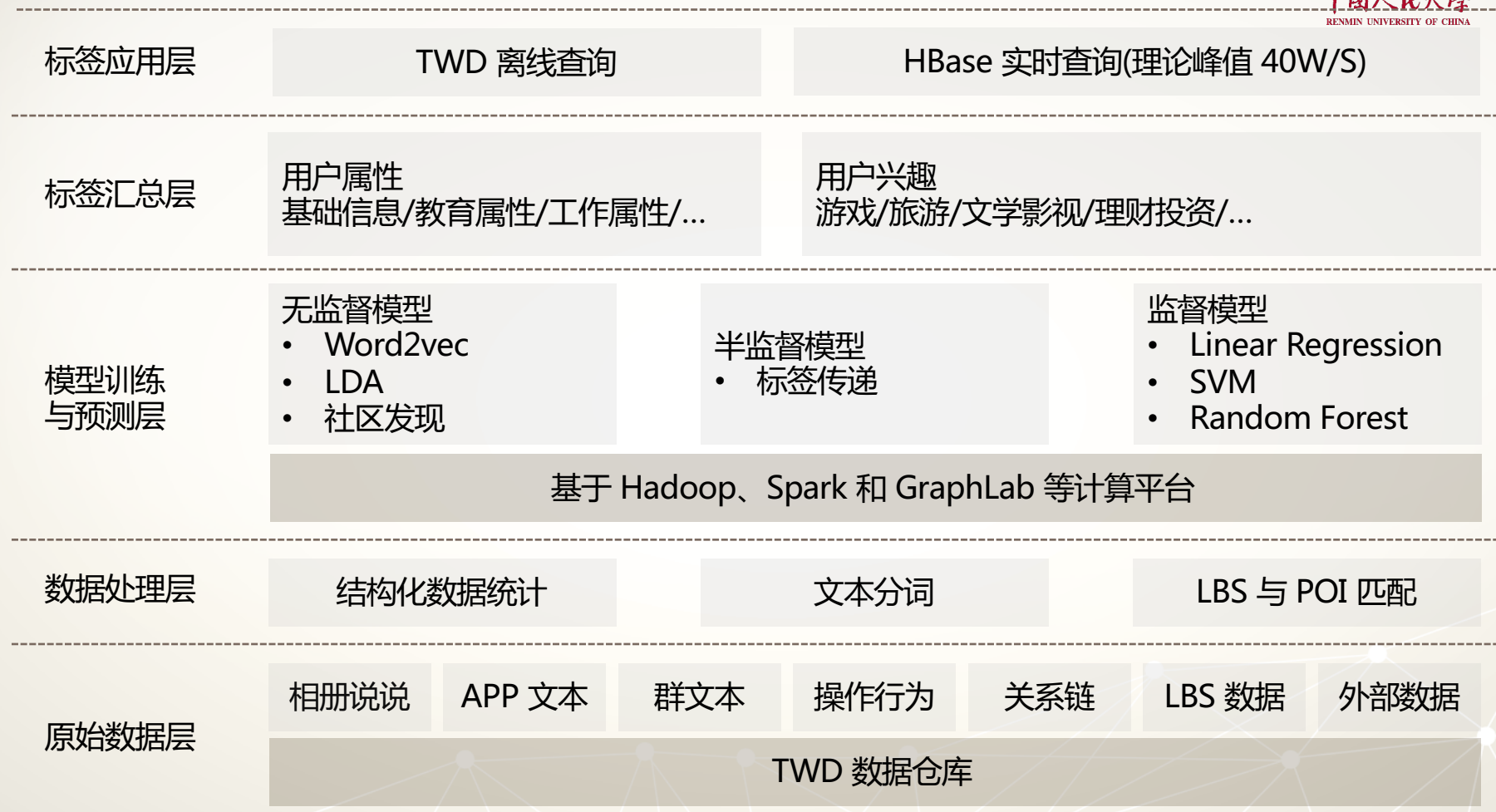
规则引擎：业务的核心，结合用户、场景、算法输出数据和业务 KPI，决定为用户推荐哪些内容

算法引擎：计算用户之间的相似度、商品之间的相似度、用户对商品的评分、用户分群和热门排行

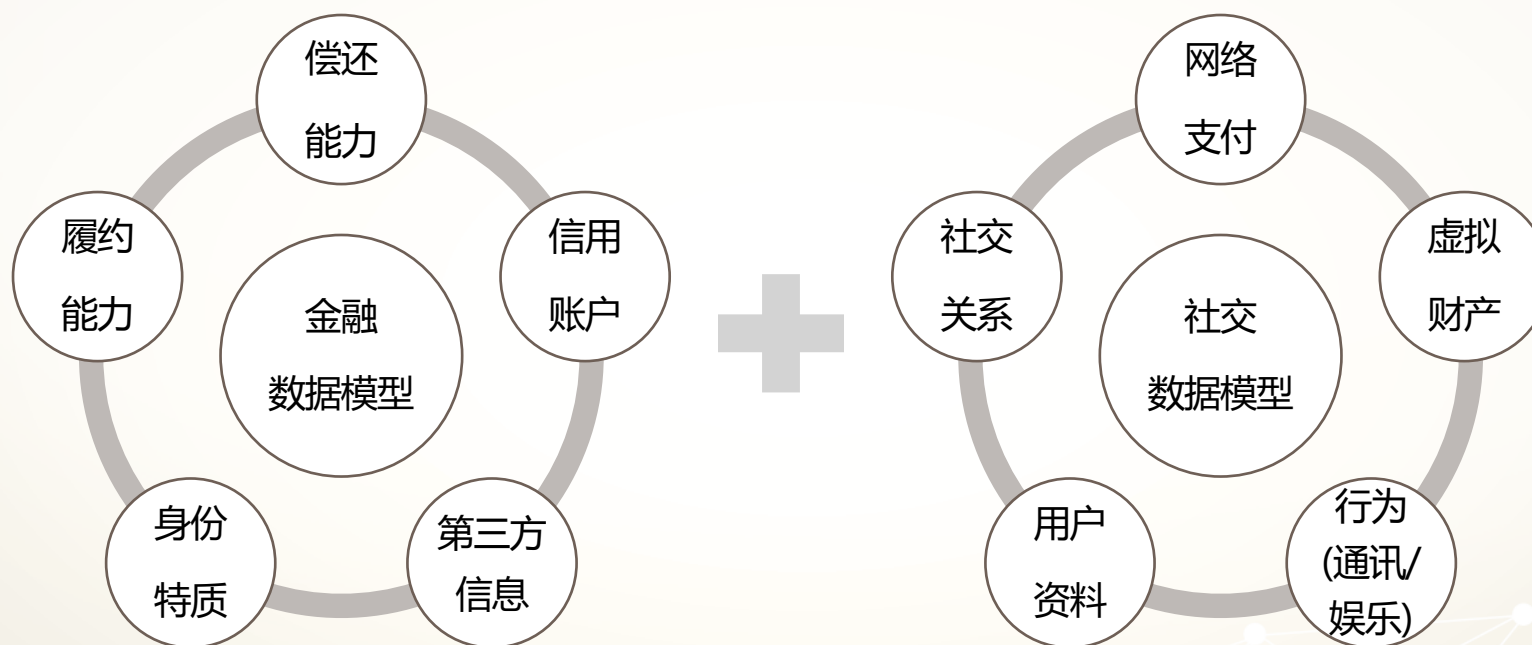
展示引擎：将推荐内容以最佳的展示方式呈现在用户面前



■ 腾讯用户画像应用案例—系统部署

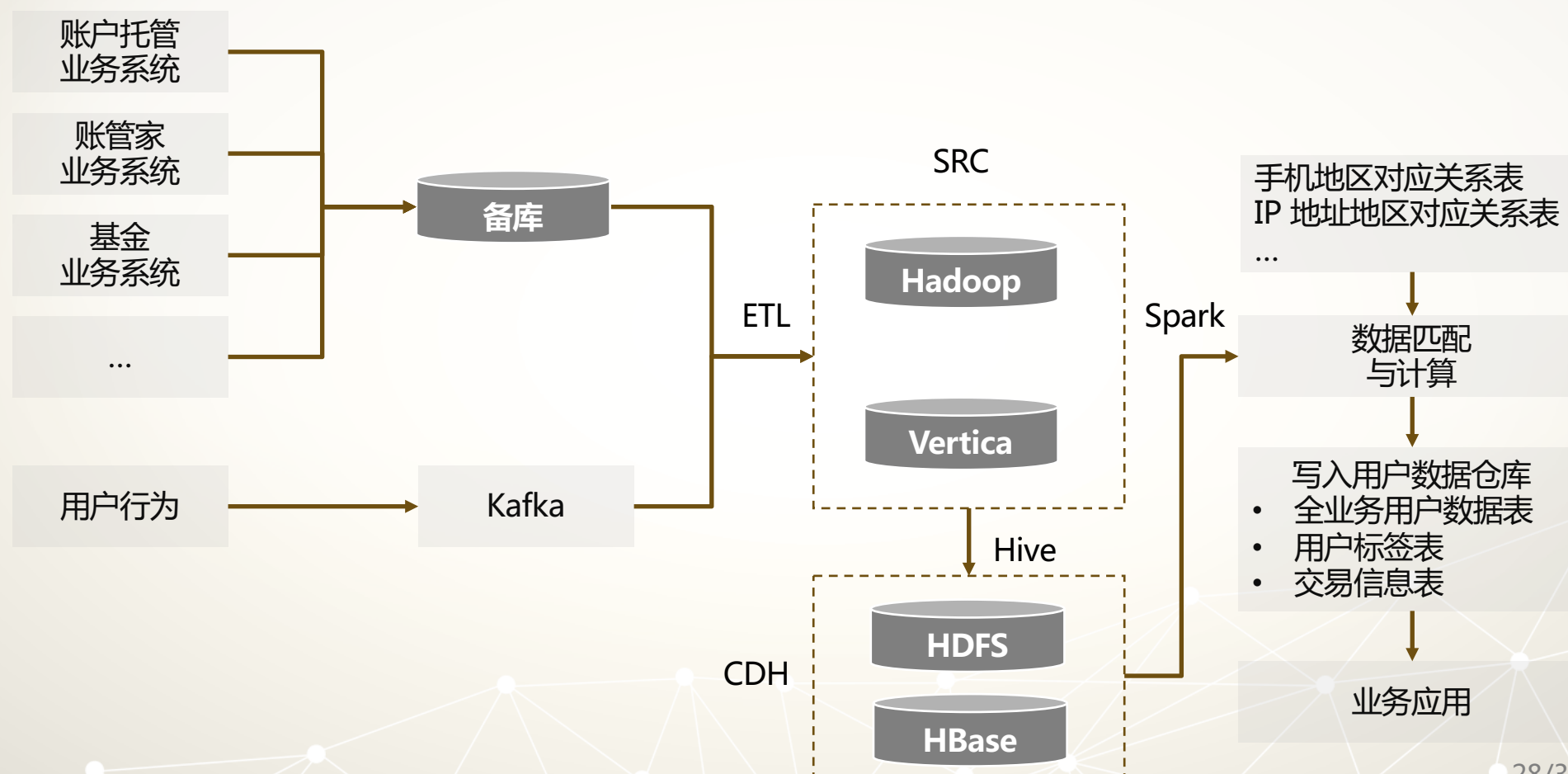


■ 腾讯用户画像应用案例—征信



04. 汇付天下用户画像应用

■ 汇付天下用户数据仓库



■ 汇付天下用户画像标签体系



■ 汇付天下用户画像应用举例



用户研究分析报告

监控业务运营，跟踪用户行为，帮助企业战略规划

POS 经营性套现
商户识别

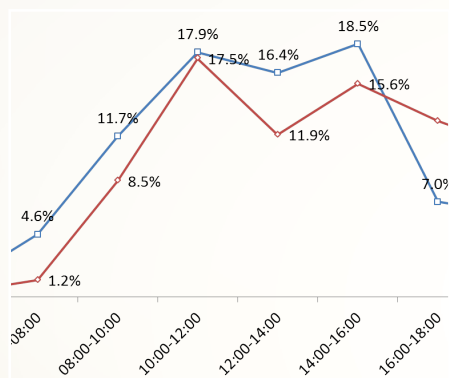
识别业务风险，有效提高风控能力

客户关系管理

吸引新客户、保留老客户以及提升客户粘性

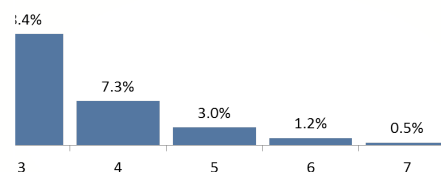
■ 用户研究分析报告

用户行为跟踪



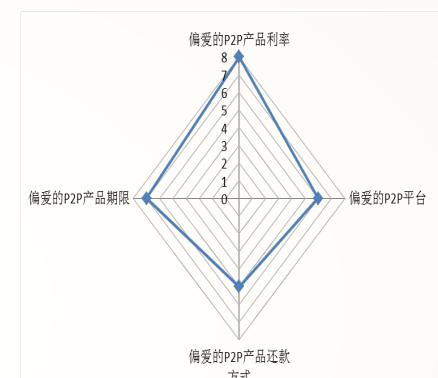
了解用户日常各类业务操作时段，为广告推送和产品推荐提供时间维度的参考。

用户活跃等级



监测用户操作，评定活跃等级，利用这一属性，挖掘潜在用户，推动交易量的增长。

产品偏好



探究用户对各类产品偏好，匹配相似产品，推荐用户购买，提高交易量。

■ POS 经营性套现商户识别—概况



- 背景

POS 商户利用终端机具，帮助贷记卡持卡人进行套现，并从中获取一定收益

- 目标

借助聚类算法对商户进行无监督的分类，探究各类商户的行为特征，在此基础上，识别出具有较高套现嫌疑的风险商户，进而方便业务部门采取相应的措施

■ POS 经营性套现商户识别—流程

前期准备

需求方
沟通

收集业务
相关信息

初定方案
与需求方
确认

分析建模

模型建立

挑选行为特征标签
数据预处理

方法评估

- K-means 算法
- ...

模型评估

否

是

模型确认

实施及跟踪

数据计算

ETL

调度配置

Spark+Python

结果输出

风控系统
监控
与处置

记录模型表现

模型跟踪

是

模型结果是否稳定？
模型排序是否有效？
模型是否无偏？

风控需求

■ POS 经营性套现商户识别—结论

行为特征描述	类别鉴定
单笔金额：高 单卡平均笔数：一般 有交易天数占比：高 卡数：极多 回头客比例：较高 特点：单笔金额高，卡数极多，有交易天数占比高，疑似经营性套现	疑似经营性套现商户
单笔金额：一般 单卡平均笔数：极高 有交易天数占比：一般 卡数：一般 回头客比例：极高 特点：单卡平均笔数极高，回头客占比极高，但卡数很低	正常商户
单笔金额：极低 单卡平均笔数：一般 有交易天数占比：极高 卡数：极少 回头客比例：极高 特点：回头客比例很高，平均交易时长短	新尝试商户
...	...

■ 客户关系管理—概况



- 背景

分析投资用户在销售、营销和服务上的交互，从而提升其管理方式，向用户提供创新式的、个性化的交互和服务的过程

- 目标

帮助投资平台吸引新客户、保留老客户以及将已有客户转为忠实客户，增加市场份额

■ 客户关系管理—算法介绍

- RFM 模型是客户关系管理的重要分析模式，是衡量客户价值的重要工具

- RFM 指标定义：

1. 最近购买时间 (R)

顾客最近一次的购买时间和分析时间点间隔的天数

2. 购买的频率 (F)

计算期内顾客购买产品或服务的次数

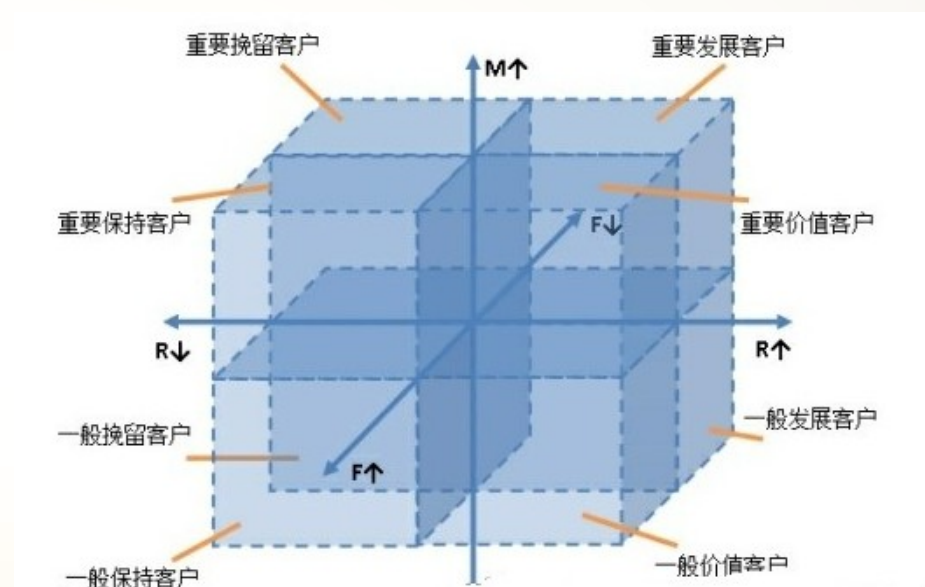
3. 购买金额 (M)

计算期内顾客购买的总金额

- RFM 建模

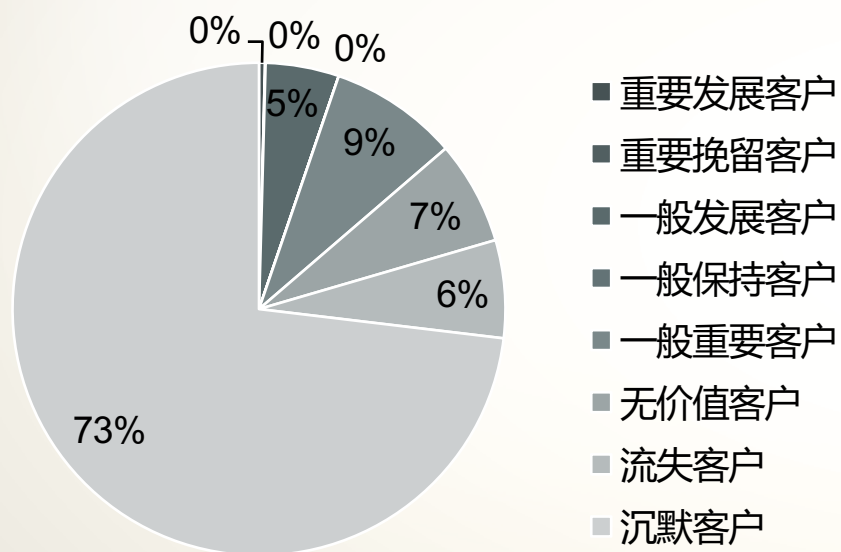
1. 根据 (R、F、M) 指标进行聚类

2. 聚类中心的 (R、F、M) 与样本总体的 (R、F、M) 比较，确定客户类型



■ 客户关系管理—结论

基于用户忠诚度分析，选取用户的投资近度、频率及金额标签，利用 Spark 的高效大数据分布式计算能力，通过标签聚类将用户分为多种类型。



• 模型结果

重要发展客户

最近一次交易间隔时间短、单笔购买金额大且投资频率较低

重要挽留客户

单笔投标金额较大，交易具有偶然性

一般发展客户

交易间隔时间短，投标频率和单笔投标金额的较低

... ..

■ 总结

- 主要挑战

如何充分利用各数据源并建立之间的联系

如何使得用户画像灵活应对各种业务场景

如何高效处理海量用户数据

- 解决方案

形成底层事实类标签，基于底层标签抽象更上层用户标签

建立完善的用户画像标签体系，支持从不同维度、颗粒度进行用户信息挖掘

搭建大规模和机器学习计算平台，结合跑批处理和实时计算满足高性能数据处理



THANK YOU

