

移动用户画像在挖掘与服务高净值客户的应用

沈吉梅

(东北大学 工商管理学院, 辽宁沈阳, 110169)

摘要

随着互联网金融的深入发展, 移动金融应用也日益普及。面对数以亿计的用户及海量的数据, 金融企业要服务好用户, 必须对用户进行画像。通过用户画像, 金融企业能够为用户提供个性化的服务, 同时促进金融企业健康稳健发展。高净值客户逐渐成为财富管理机构尤其是证券企业的重要服务对象。因为现在高净值客户市场体量日益增大, 从证券企业开展财富管理业务显得日重要, 利用用户画像技术对高净值客户客户关系管理进行全面而深入的研究。

关键词: 移动用户画像 高净值客户 数据挖掘

1 引言

客户关系管理(CRM)是指公司与客户, 尤其是公司与客户关系的价值管理。管理, 这既是一门科学也是一门艺术。由于金融业存在二八原则 [1], 也就是说只占 20% 证券行业的客户创造了高达 80% 的利润这个原则。对于证券公司, 它为获得更高的市场回报而对高端客户的特殊的管理就显得尤为重要。因此, 证券公司必须标注客户进行分类管理, 以确保经营活动的有效性 [2]。挖掘不同用户不同的偏好, 提供个性化的服务已逐步成为证券公司的核心理念。

随着智能手机的普及, 人们的生活行为与移动智能设备链接更为紧密, 人们可以随时随地通过移动终端设备取得服务, 各种与移动终端相结新兴的商业模式正在增长。在这种背景下, 移动互联网金融产品和服务已经逐渐渗透进入了人们的生活。互联网金融业务数据和用户数据正在飞速增加, 用户与用户之间、企业与企业之间的关系越来越强。海量用户信息和交易数据, 使得证券公司进行信用评价, 挖掘潜在高净值客户; 分析其投资偏好, 个性化的高净值客户的产品和服务; 对于高价值客户有针对性的风险管理是可能的。

从理论方面上看, CRM 发源于 1980 年代的欧美, 而传入我国的时间就更晚。CRM 的研究主要从战略和技术方面进行研究, 关于高净值客户的 CRM 的理论研究很少, 而应用用户画像进行高净值客户挖掘与个性化服务的研究几乎没有。因而, 本文将新型技术与传统 CRM 结合, 具有较强的理论意义。

从实用的角度来看, 当今世界经济, 尤其是中国的经济, 正处于高速发展中, 催生出一批拥有巨额财富的潜在用户。本文旨在应用 MPP+Hadoop 大数据技术框架的用户画像对投资者进行精细划分, 了解这些高净值客户的需求并深入挖掘高净值客户者投资、融资操作中的内涵, 分析出其投资方向、风险水平等, 为证券公司资产挖掘潜在客户及投资策略构建提供借鉴, 具有重要的现实意义。

2 移动用户画像与高净值客户

本小节主要对相关概念与理论进行概述，首先介绍用户画像的定义与其在移动金融领域的应用，然后阐述高净值客户的概念与相关财富管理业务理论，最后对国内外相关文献进行梳理，为后文提出的技术框架提供理论依据。

2.1 相关概念与理论概述

2.1.1 用户画像与其在移动金融的应用

用户画像是通过收集分析用户的信息标签集合（如基本属性、社交关系、个人偏好等主要信息），进行分析而描绘出的数字个体 [3]。在大数据技术的帮助下，可以了解收集用户更多的信息，对于证券公司来说，用户画像本质就是从公司业务角度出发对投资者进行分析，了解投资者的各种功能（如基本属性，购买力，投资行为特征，兴趣爱好，心理特点，社交网络等），发现用户需求，寻找信用等级高、风险低、给企业创造可观利润的客户。



图 2-1 金融客户标签构建图

由于证券公司拥有大量用户信息和交易数据，使得对各类投资者信息进行数据挖掘、特征分析、抽象归类，从中洞察客户行为偏好，全方位了解客户特征成为可能，为证券公司进行潜在客户挖掘、风险管理提供工具支持、数据支持。但是光靠证券公司的数据还远远不够，还要广泛依靠银行、社交媒体、第三方公司等外部数据，以丰富客户标签，使用户画像有更高的“像素”。

金融客户标签依托大数据技术进行，如图 2-1 所示，首先时数据选取与加工，包括：采集证券公司的行内外数据源，行内数据来自比如证券公司的交易系统、专家系统等来源，行外数据来自政府部门或者银行等，对这些数据使用数据挖掘、自然语言处理、知识图谱等手段对数据进行分类加工，这个过程用来发现用户行为特征。客户标签存储层主要通过大数据技术，从而实现特征识别，进一步构建客户标签库。通过客户标签库可以对其进行用户画像构建，从而进行客户群体分割与投资者风险监控等。

2.1.2 高净值客户

高净值个人（HNWI）是一个广泛的概念，对其定义学术界和实务界并未达成完全一致。Credit Suisse [4] 和 Capgemini [5] 发布的《2018 年世界财富报告》中，富裕人士是指个人、股票、债券、共同基金，银行账户以及其他流动资产在 100 万美元以上的人（需排除自住和自用的房产）。在由中国招商银行和贝恩公司发行的《2017 年中国私人财富报告》中 [6]，高净值客户被定义为具有超过 1000 万元可投资资产的个人；而超高净值客户指的是可投资资产超过 1 亿元，可投资资产指的是个人的金融资产和投资性房产综合。由此可以看出，富裕人士的定义在国外与国内基本相同，因此可以选取可投资资产在 100 万美元作为划分标准。

与 HNWI 紧密相关的一个概念是私人财富管理。私人财富管理主要指的是金融机构为富裕个人提供全方位、多层次、专业定制的金融产品和金融服务。下面选取具有代表性的主流理论（现代资产组合理论、生命周期理论、流程再造理论以及市场细分理论）进行简单介绍。

（1）现代资产组合理论

现代投资组合理论在 1952 年由马科维茨提出，即理性的投资者不仅要高回报，与此同时还需要风险尽可能低。因此，他构造了方差收益模型，衡量风险与收益的关系，找到理性投资边界与合理的资产组合。[7]。

现代资产组合理论是典型的新古典金融学研究领域内容，即把资产的风险分割为系统性风险和非系统性风险，通过资产组合在投资组合中的资产个数区域无穷大时将整个投资组合的非系统性风险降低到零，即只存在系统性风险，使得投资组合风险尽量控制在不可避免的风险上。其基本假设就是理性人假设，即投资者希望利益最大化。

对于证券公司而言，通过对客户的交易行为等进行分析，获取客户的风险偏好水平，为其定制个性化的投资组合。根据马科维茨理论，进行不同风险接受程度和厌恶程度的用户画像特征分析后，不同种类的投资者的投资组合见表 2-1。

表 2-1 不同风险厌恶水平下投资者的投资组合

风险水平	投资组合	适合的投资者
收益型	100%债券市场	不能容忍本金存在任何损失，期望比定期存款收益率略高一点
保守型	10%-20%股票市场 80%-90%债券市场	期望收益比较稳定，比定期存款收益高一些，可以承担较小的风险损失
均衡型	50%股票市场 50%债券市场	期望投资兼具成长与稳定性，可以承担风险损失
成长型	70%-80%股票市场 20%-30%债券市场	期望投资取得较高的收益，并愿意承为之担较高的风险损失
激进型	100%股票市场	期望取得极高的收益，并愿意为之承担本金损失甚至更多风险

（2）生命周期理论

马科维茨提出的资产组合理论为不同的风险承受水平的投资者提供了不同的资产组合方案，但是人的风险承受水平不是一成不变的，其需求也不是一成不变的，金融机构必须根据用户的生活金融产品和不同的阶段提供金融服务，构建理财目标要考虑用户的生命周期。生命周期理论对人生不同阶段进行了分析，这就为证券公司开展财富管理提供了相应的指引。

生命周期理论最早在 1959 年由 Modigliani 等人于提出，Modigliani 从人的不同阶段阐述了人的不同需求，并根据不同的需求由此引出了不同理财目标的选取。该理论认为，一个人可以被分为青年、中年和老年，其划分的主要依据是通过劳动获得收入的能力 [8]。

在青年阶段，人处于人力资本的积累时期，大部分时间在学习，个人收入偏低，消费有可能超过收入；在中年阶段，青年期积累的人力资本逐渐转化为现实的劳动力，通过不断的劳动使得收入逐渐提高，收入会大于消费，这一阶段所积累的财富不仅要用以偿还上一阶段父母的养育，还要为下一阶段自己的老年做准备；在老年阶段，人力资本逐渐消耗完，人的收入水平逐渐下降，消费又超过了收入。总而言之，人生的第一阶段即青年时期为人力资本积累和财富的负债过程，第二阶段即中年时期为人力资本的小号及财富的偿还与积累时期，第三阶段即为老年时期则为人力资本和财富积累的消耗阶段，对于不同生命阶段的收入和支出来看，财富管理有不同的目标。

在对高净值客户进行财富管理时，有必要充分考虑客户所处人生的哪个时期，整合用户的风险偏好，并定制不同的财务目标，为客户收入的能力。表 2-2 显示客户在人生不同阶段的财务目标。

表 2-2 人生不断阶段客户的理财目标

生命周期的阶段	收支情况和风险承担能力	理财目标
积累阶段	收入一般小于支出 风险承担能力较大	储蓄，教育投入，为购房做准备等
巩固阶段	收入稳定增长，收入大于支出，承受中等风险	子女教育投入，为年老消费积蓄资金，分散风险等
消耗阶段	收入不断下降，支出大于收入，选择风险水平极低的投资方式	医疗保障、养老金准备、旅游休闲等

（3）业务流程再造理论

业务流程再造理论是于 1991 年由 Hammer 首先提出，是对企业业务流程进行本世的思考和再次设计，这使公司能够做出的个体经营绩效的重要指标，包括成本，质量，服务和速度显著的进展。Hammer 认为，信息技术是 BPR 的助推器。Venkatraman 指出信息技术是 BPR 的赋能机制 [10]。

当证券公司在对高净值客户开展财富管理时，能否进行业务流程关系到传统证券公司能否由原先的被动服务转变为主动服务类型，证券公司可以充分利用信息技术，掌握高净值客户信息，并掌握高净值客户的需求，然后设计相应的产品和服务，这关系到证券公司能否提高客户满意度和忠诚度，进而实现证券公司自身的盈利这一目标。

（4）市场细分理论

市场细分理论于 1956 年由 Smith 首先提出，它指的是根据消费者需要、购买习惯、购买能力等之间的差异，将消费者市场细分成不同类型的群体，在同一个群体内消费者有类似的消费倾向，并且不同的群体消费者之间存在一定的差异 [11]。

根据市场细分理论，证券公司在对 HNWI 进行财富管理时，在充分了解客户信息的前提下进行用户画像分析，同时，结合资产规模、投资风格等证券公司自身因素，制定并采取相应的客户关系管理策略，从而优化资源配置，在证券投资管理市场上取得充分的优先权与主动权，从而提高 HNWI 的满意度，实现预期盈利目标。

2.2 相关文献综述

客户关系管理理论为 Berry 等人于 1999 年提出，最初主要指的是“联系管理”，也就是尽可能多的收集客户相关信息 [12]，并逐渐由信息收集衍生为客户关怀，即应用信息技术实现用户个性化服务而不仅仅是停留在信息收集 [13]。此外，随着社会信息技术的不断发展，海量的数据使得数据的收集、数据的分析与处理技术不断迭代更新，不断丰富发展了客户关系管理理论。

目前，国内外相关学者具体针对应用用户画像对高净值客户进行客户关系管理的研究非常少，之前的文献大多是从战略层面、技术层面以及两者相结合的层面作为切入点，从而对客户关系管理进行研究。

2.2.1 战略层面

从战略角度上看，CRM 主要是从能力和资源进行客户管理，主张企业需要提高获取客户关键信息的能力，主要通过获取信息后合理利用资源来提高企业的竞争力和企业效益。

Berry 等人认为客户关系管理是一种关系营销，通过收集、分析、处理客户的信息，主动采取相应营销行动，以其通过自身影响客户的行为 [14]。基于营销的 CRM 理论强调通过关系营销形成与维护稳定的客户关系，并以此服务企业的盈利目标。Ahearne 认为客户关系管理的关键在于如何使用正确的方法，使用正确的工具分析，从而获得正确的结果，从战略角度来看，关键是如何划分客户群体并对不同的群体分开分析 [15]。

Geib 相信是客户关系管理过程中，有必要紧密结合公司的战略目标，结合公司的战略目标，才能为客户提供更好的产品和服务 [16]。Pedron 等人相信客户关系管理是一个系统的宏观过程，并且在这个过程中包含很多小的子流程，并且通过这些流程可以找到与客户的长期合作过程。此外，从战略角度分析这些过程也可以反映客户关系的生命周期特征 [17]。

2.2.2 技术层面

从技术角度上看，主流的 CRM 主要用到数据库技术、信息处理技术、 workflow 管理技术等信息技术。

Brown 首先系统研究了客户信息的获取和客户信息的分析等内容，并对客户关系管理中的数据挖掘、数据分析、模式发现、CRM 绩效评估等技术分析进行了深入的探讨 [18]。Josiassen 研究了信息技术在客户关系管理中的作用，他研究结果表明，利用先进的信息技术可以有效地提高客户关系管理的水平和效率 [19]。Navimipour 通过自己的研究证明了有效利用信息技术对于显著提高公司绩效由明显的作用 [20]。

使用信息技术建立更有价值的客户关系需要利用获取、开发和维护有利可图的客户组合过程中的相关信息和经验。在客户知识管理的概念中，Xu 强调在客户管理过程中，重点是通过数据收集、数据存储、数据共享和使用客户知识发现 [21]。

2.2.3 战略与技术相结合

客户关系管理不仅需要从战略层面进行宏观把控，同时也需要信息技术的支撑，兼顾这两者才能更好的充分调动宏观、微观各项因素，提高社会全要素生产率。

Swift 首先提出 CRM 不仅仅是一个技术问题的更多。它的目标是提升公司的竞争力，通过建立和维护与客户的长期合作关系，实现公司的经济效益 [22]。Jafari 等人提出成功的公司和组织通常通过战略与技术结合，牢固与客户的关系来争取竞争优势 [23]。

目前，如何准确有效地从海量数据中挖掘用户信息成为大数据背景下 CRM 的一大难题 [25]。利用数据处理综合多种因素进行 CRM，对于公司精准定位、提供

个性化服务在现在显得尤为重要 [26]。

从以上分析可以看出，应用信息技术，采用数据挖掘方法有助于客户关系管理绩效的提升与用户体验的提升。因此，本文选取用户画像技术，通过数据分析挖掘客户标签、挖掘高净值客户并提供个性化服务，以期丰富客户关系管理理论研究。

3 主要数据来源与技术框架

3.1 主要数据来源

证券公司可用的客户移动信息包括但不限于：基本信息、交易信息、风险信息、人行征信、网购信息等。

3.1.1 客户信息

客户信息包括移动金融用户的姓名、年龄、性别、教育背景、手机类型、以及由此延伸的手机用户的实际话费、手机号码注册地、身份证居住地址等。通过收集用户的通信记录，可以描绘出该用户的社交网络。社交网络的研究分为两种范式，第一种是以某个人为核心点，重点探讨研究其他各节点和核心点的连接互动关系以及连接强度，第二类网络是以整个社会为整体，研究重点在于网络结构以及信息如何在网络内部扩散 [27]。可以充分利用与该投资者密切联系的社会网络群体中的一些特征，如手机型号的整体价格水平、居住地点平均房价等推测该用户所处的社会环境，以挖掘潜在的 HNWI。

3.1.2 交易信息

用户交易信息来源主要有两个渠道：证券公司自身与其他第三方机构。包括投资者每个月在该证券投资多少、投了哪些产品、用何种方法支付、是否有按揭等。

3.1.3 风险信息

用户的风险信息，可以看到主要从操作风险和信用风险。操作风险主要是指由信息系统或内部控制缺陷意外丢失的风险。对于证券公司，引起操作风险的投资者的原因包括人为错误、电脑系统故障等。对于证券公司角度来说，可以监控该用户是否有操作失误导致巨额损失，或提示，当它即将执行危险操作。信用风险，也称交易对手风险和履约风险是指交易对手不履行到期债务的风险。证券公司可以银行或其他金融机构合作，共享数据，了解该用户是否存在坏账、呆账或其他不履行债务契约的行为。

3.1.4 人行征信信息

2006 年 3 月，中国人民银行建立了中国人民银行征信中心，对企业和个人信用信息系统（即金融信用信息基础数据库）直接负责，也称为公司和个人信用信息库）的建设，运营和维护。信用信息系统被广泛应用于金融机构的信贷风险管理，有效地解决信息不对称问题，提高公共资金的便利，创造更多的融资机会，促进信贷市场的发展。，显着提高了社会信用意识。证券公司可以借助该系统，对用户的行为进行监控。

3.1.5 网购信息

网购信息包括互联网中的电商消费及相关信息。比如蚂蚁花呗积分，用户用移动支付消费了多少钱，购买了那些产品或服务，有多少是使用移动金融平台的线上和线下消费？什么时候用户浏览购物网站、金融网站，浏览这些网页，并有点击行为和购买行为？什么是用户常用的登录地址和送货地址、常去的地方有那些？用户常买的产品有那些，价格是多少，是国外品牌还是国内品牌？等等。

3.1.6 其他信息

如移动用户经常在的坐标地点、在每个对应坐标地点停留时长等信息。

3.2 技术框架

传统的数据处理方式是一种行存储方法，支持一般的增加、删除、修改等基本操作，进行统计分析时效率较低。此外，传统的数据库系统使用关系型数据库，如 Oracle、IBM / DB2、Microsoft/ SQLSERVER 等等，不支持半结构化和非结构化的数据的存储和处理。金融行业有大量的半结构化、非结构化数据存储、查询、分析和处理的需求，传统数据存储方式已无法完成该需求。与此同时，进行用户画像构造还需要数据处理如整理、交叉分析比较和海量数据的深度挖掘。结合上述两个要求，目前可以适用于移动金融用户画像分析的主要有以下三种主流处理架构：

3.2.1 MPP 数据库架构

MPP 是结合一种高效的分布式计算模型和 SharedNothing 体系结构，如图 3-1 所示，该方式能够成为大数据处理的一种可行的方式，这是因为它支持海量数据处理技术，如粗粒智能索引、列存储、高效的压缩等等。结构化数据分析类应用的支撑。

这种数据库集群方式具有高性能和高扩展的特点，能够有效支撑 PB 级的结构化数据处理，广泛应用于行业数据仓库和各类结构化数据分析领域。金融数据中如财报等信息具有高度结构化的特征，因此在结构化数据分析中，MPP 数据库是最佳的选择。

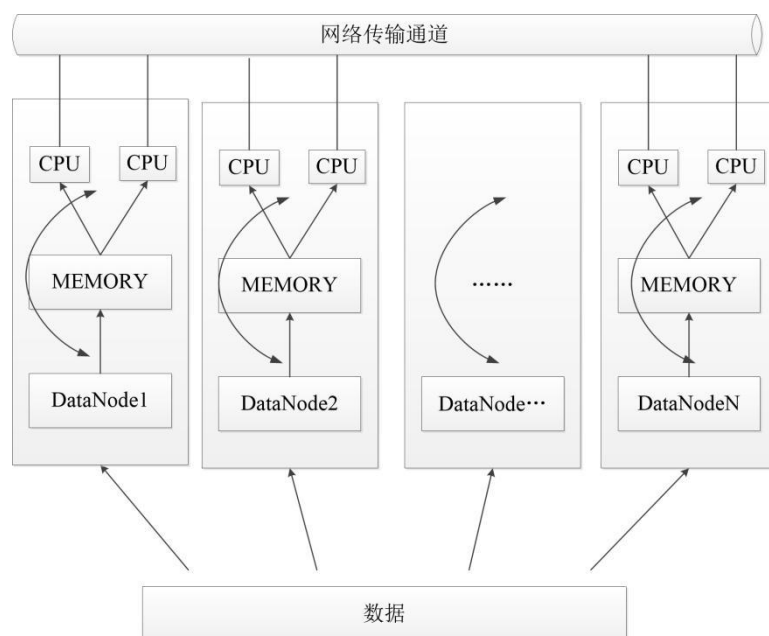


图 3-1 MPP 数据库架构

3.2.2 基于 Hadoop 的技术扩展和封装

Hadoop 的技术扩展和封装依托于开源社区的优势，加速其相关技术不断进步和迭代更新，可有效支撑非结构化、半结构化数据处理、复杂的 ETL 流程、数据挖掘和计算模型。该技术基于开放的 X86 架构服务器部署以及 HDFS 的分布式文件系统，具有高存储、低成本扩容、易扩展性等优势，对于金融从业者来说，其主要缺陷包括对大表复杂关联处理性能、数据更新性能较差、SQL 兼容性较差等。

3.2.3 大数据一体机技术

大数据 All-in-One 的技术是大数据存储、处理、展示全功能、软硬整合，面向程序的产品。它是专为大数据分析和处理而设计的，它的软件和硬件具有很强的针对性。它由集成服务器、存储设备、操作系统、数据库管理系统和为进行数据查询、处理和分析目的进行了优化的软件组成。具体地，为了金融领域进行数据查询和处理分析的目的，这所有功能于一身的机器具有良好的稳定性和可扩展性，为金融分析师快速高效的查询有效数据是进行挖掘潜在客户与合理投资的前提。

3.2.4 MPP+Hadoop 大数据框架

上述处理架构具有各自的优缺点和应用场景，在实际金融场景应用中，基于列存储 MPP 架构的新型数据库，与 Hadoop 生态系统结合混搭使用方案，结合各自优势，实现功能互补，能解决金融大数据存储的诸多复杂需求。

其中，MPP 用来处理 PB 级的，高质量的结构化的数据，同时提供丰富的 SQL 和应用程序事务的支持；HADOOP 是半结构化的、非结构化的数据处理、流处理和大规模批量操作。

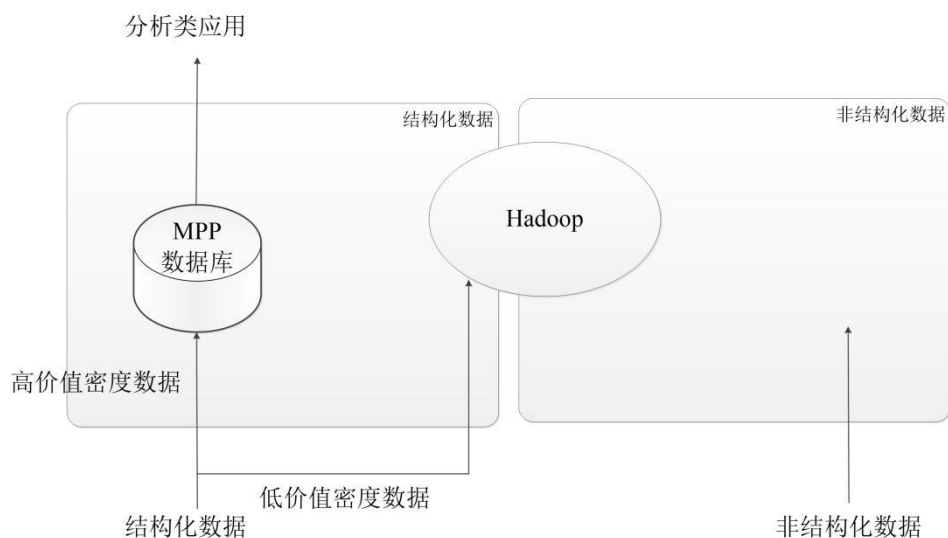


图 3-2 混合方案金融数据处理技术架构

3.3 基于混合方案的高净值客户用户画像构建方法

证券公司在数据存储、查询分析、资源协调与管理方面遇到了类似经济学中“三元悖论”，也就是说，只能满足在同一时间两个目标，而放弃另外一个目的是最大限度地发挥整体效用。类似的，Hadoop 和 MPP 目前只能满足以下 2 项功能，而不能满足全部要求。在此限制下，选取用户画像构建主要需要实现的目标如下。

（1）实时

在实时方面，单节点执行系统（如应用 MPP）具有明显优势，其他方式会在一定程度上弱化实时性能。虽然 Spark 技术提高了实时性能，但实施成本过高，同时并不完全成熟，相关技术有待进一步的完善。

（2）可扩展

可扩展即要按数据量的增加而扩容。MPP 扩容到一定程度后因传输等因素将会受到限制，而 Hadoop 的 Mapreduce 在此方面表现较好

（3）数据复杂查询与复杂分析的处理能力

Hadoop 和 MPP 均可以通过算法实现此功能，但会有难易与熟练程度上的差别。

为了实现上述目标，进行数据汇聚整合、数据整理、数据存储、数据分析、用户画像构建，下面进行详细阐述。

3.3.1 数据汇聚整合

数据聚合和整合需要的数据类型和数据容量，以根据该要求进行扩展根据需要，一种证券公司建立大数据服务体系如图 3-3 所示。

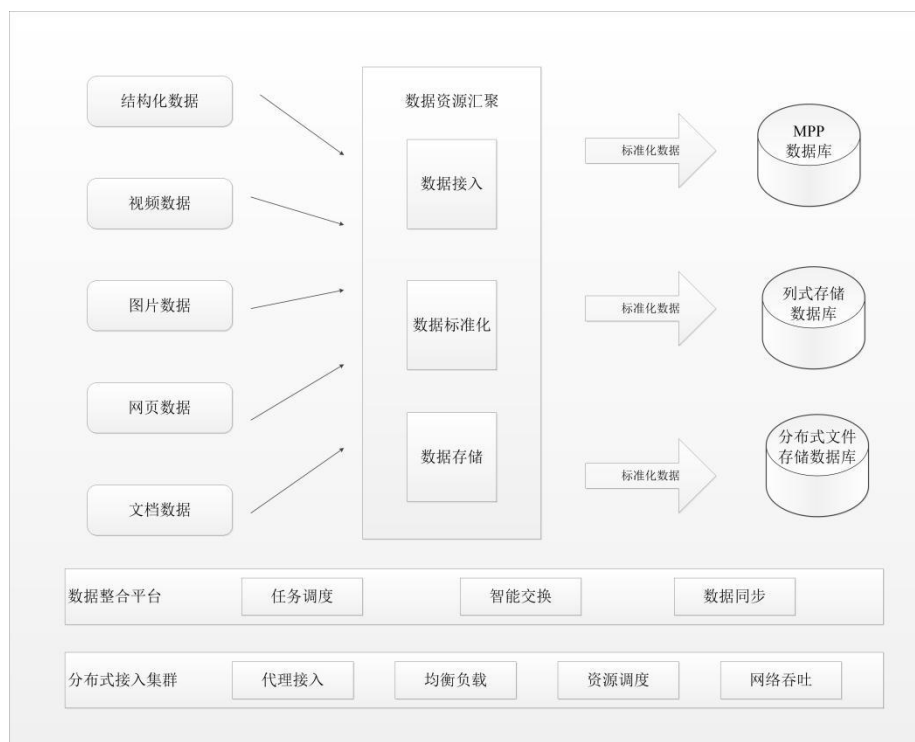


图 3-3 数据汇聚整合功能框架图

金融大数据聚合和集成基于底层分布式访问集群和数据集成平台的功能，通过数据访问和数据实现大量异构数据，如结构化数据、自适应数据、图片数据、网页数据和文档数据标准化。并且数据存储功能，形成标准化数据，并根据数据存储的性质保存到 MPP 关系数据库，列存储数据库和分布式文件存储的数据库。其中，数据集成系统可以分成三个逻辑层次：数据汇聚层、数据组织层、数据服务层，如图 3-4 所示。



图 3-4 系统逻辑分层图

数据聚合功能是指各种类型数据的集合，它可分为模块，例如数据接入适配，数据缓冲，数据同步的功能的基本功能模块是通过企业服务总线技术（ESB）实现，它采用了统一的数据访问总线，以获得不同的业务系统和数据库，如传统的关系型数据，大规模并行关系数据库，MPP 和 Hadoop 数据库来实现。提供对数据访问的安全性。数据聚合层需要 ESB 安全访问客户端和 ESB 安全前端系统，以支持跨网络的安全数据访问。

3.3.2 数据整理

数据整理包括数据清洗、数据格式转换、数据分类及其他功能。

（1）数据清洗

数据清洗采用数据集成工具可用数据提取、数据清理、数据转换和处理异构源数据，并将其加载到多个目标数据库中，为用户标签挖掘与用户画像建立提供数据支持。

（2）数据转化

数据被清洁后，将数据根据转换规则，并在基本的关系的库中的信息模型转换。数据转换可涉及操作以上的数百或数千，所以如何让他们高效，可扩展性和可维护性是非常重要的。数据转换基于一组转换规则，并以开放的方式进行管理。同时，转换算法可以根据个人需要进行修改和添加。如果需要复杂的数据转换，证券公司可以使用转换的结合来实现自己的目标，及时通过算法重新与转化组合提高系统整体效率，降低数据转化时空复杂度。转换流程如图 3-5 所示。

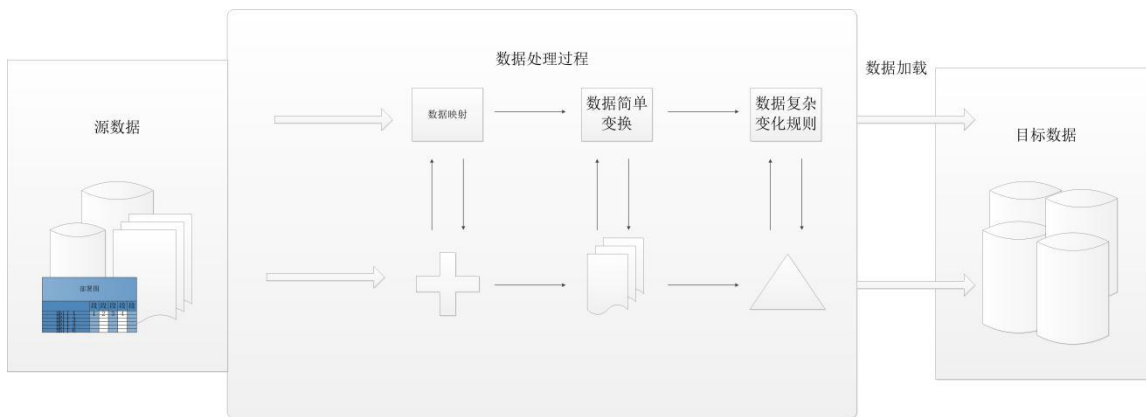


图 3-5 数据转换图

（3）数据加载

提取、清理数据并建立映射关系后，可以加载数据，就可以进行数据的装载工作，同时建立数据的关联关系。

3.3.3 数据存储

该数据被存储在数据采集，数据聚合和整合处理。基于数据类型和数据应用划

分，分别进入基础数据库（客户信息）、特别的库（交易信息）、相关的库（风险信息）、历史库（人行征信信息、网上购物信息）、缓存库、配置库和管理库。其中，基础库，相关联的库，配置库和管理库使用 **MPP** 数据库，以确保数据的更新频率，提高数据存取效率；历史库和缓存库的实时性要求并不高可以使用 **NoSQL** 数据库、**HBase** 的存储；此外，可以利用 **Storm** 或 **Spark** 等对流数据或图像数据进行图计算。

在具体实施方式中，高速缓存库可以由高性能的 **Oracle** 数据库服务器的，用于数据缓存和数据集成过程 **ETL** 处理中间库；基础库包含两个方法：流水数据和全量数据更新，而流水数据和全量数据是同时双向融入 **HBase** 的和 **MPP**，分别满足实时快速的查询要求和离线数据分析要求；集成到 **HBase** 的历史库和问题库，专题库建立了符合实际的大数据应用的需求，完成 **HBase** 的分析和专题数据的模型整合，然后同步到 **MPP**。存储设计结构示于图 3-6。



图 3-6 大数据存储架构

3.3.4 用户画像构建

一般以为，用户画像的样本可以从证券公司已有的记录中进行筛选，但这种对已投资客户情况来进行逻辑判别，实质上会陷入“先验性误导”中，因为很多高净值客户可能并没有进入资本市场，或者交易甚少，那么就需要对用户进行全方位的数据分析，通过多维标签组合，客户的轮廓可以直观地显示出来，形成一个完整的、动态的、三维的客户肖像，或标签可以灵活组合，根据业务场景和管理的目的，以形成满足各个需要专业的肖像的领域；立体画像是全局性的，专业的画像是局部性的。

其中，三维肖像使用数据挖掘和关联分析技术，通过多维度标签相结合，全面描绘顾客的特点，并建立顾客消费、缴费、投资和融资的各个行为标签，集成了客户立体图像。通过客户立体图像的统一管理，挖掘看似孤立的信息背后的内在关联，并提供用于监控模型设计、交叉风险管理和客户精准营销的企业级决策的信息支持。图 3-7 显示了证券公司的高净值用户立体画像的示意图。



图 3-7 高净值客户立体画像图

专业人像。基于大数据分析和标签组合的想法，对画像深入分析。基于统一的平台，证券公司可以根据管理的目的，构建如风险肖像和营销画像等专业的肖像，为用户画像结果广泛应用于风险管理、市场推广和客户关系维护等提供了强有力的支持，并使证券公司准确预测客户的交易行为，迅速了解潜在的风险，并及时把握市场机遇成为可能。图 3-8 展示了某证券公司用户专业画像的示意图。

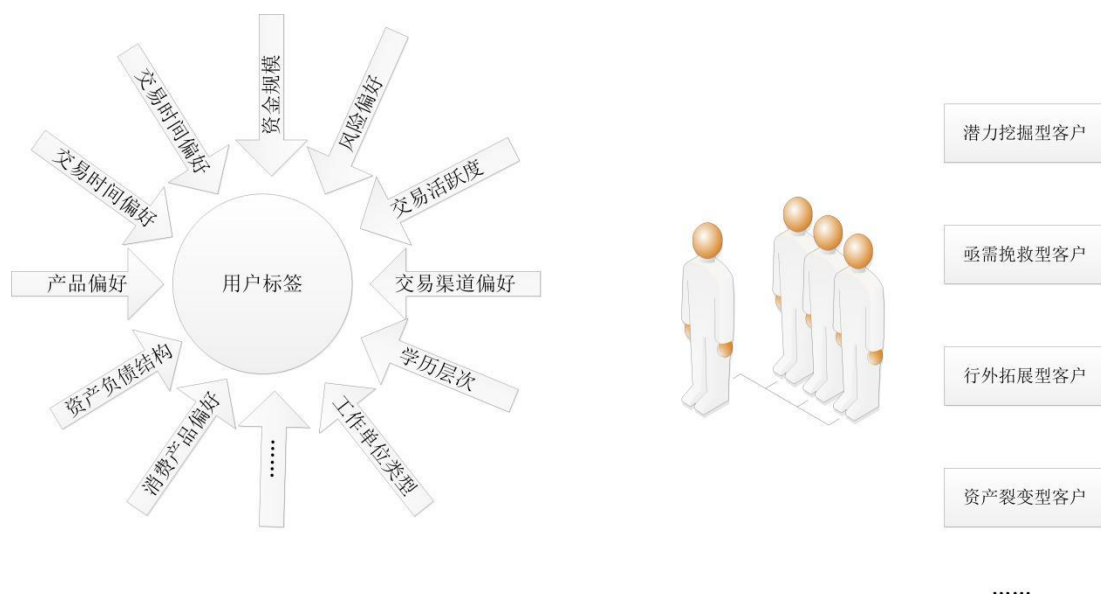


图 3-8 高净值客户专业画像图

标签作为客户特征化的定性表达和量化反映，需要遵循相关性、有效性、适用性、全面性原则，实现客户标签全方位覆盖与多视角展现。标签的合理分类对标签使用至关重要，本文是在静态与动态数据信息的整合基础上，结合证券公司管理需要从公用与专属两个方面将客户标签划分多个维度，每个维度下细分若干个标签。公用类一般是证券公司各业务条线都可能涉及的标签，如客户的收付行为、消费行为、线上行为以及基本属性、联系信息；专属类一般是满足证券公司特定目的的标签，如用于风险管理的操作风险、信用风险等标签，用于市场营销的投资行为、融资行为等标签。根据用户目的不同，也有的将标签划分为一级、二级、三级等，不管如何分类，最关键的是标签应该有自己的内在分类逻辑，并满足“封闭性”管理思路要求，以保证标签的可拓展性。标签分类应充分考虑到各业务条线或各管理领域要求，避免分类的交叉、重复形成资源浪费。

总结

用户画像标记的组合和标签的一个重要应用，但不是唯一的。单一的个人标签难以客观、全面地反映客户的轮廓，只有多维度标签组合才可以体现其价值。事实上，这种组合的过程建模过程（如风险模型和营销模式），并且也是实现挖掘用户的目的的过程。标签模型一般包括以下几个方面：场景分析、战略制定、标签确定、战略执行、效果评估、标签的优化，并通过反复训练和验证，实现精准营销、挖掘价值、寻找商机，实现数据价值的增值的方式。

在“互联网+”的背景下，大数据技术在金融的业务应用领域将不断拓展，而智能手机技术和功能的不断推陈出新，它进一步触发的消费模式的转变，产业链的开放和更广泛的数据融合。在正确处理个人信息安全和隐私保护的前提下，以客户为中心的手机用户画像应用研究，有利于如证券公司等金融机构充分利用现有的数据资源，有效实现精准营销和个性化服务，同时在高净值客户挖掘方面也可实施有益的探索与实践。

本文从用户画像角度出发,提出了一种通过 MPP+Hadoop 结合的方式对证券公司的数据管理方式,为其进行高净值客户挖掘与服务提供了依据,具有指导现实管理的意义。

参考文献

- [1] Pareto V. Cours d'économie politique[M]. Librairie Droz, 1964.
- [2] Kumar V, Reinartz W. Customer relationship management: Concept, strategy, and tools[M]. Springer, 2018.
- [3] Baden R, Bender A, Spring N, et al. Persona: an online social network with user-defined privacy[C]//ACM SIGCOMM Computer Communication Review. ACM, 2009, 39(4): 135-146.
- [4] Credit Suisse. Global Wealth Report 2018 [EB/OL]. <https://www.credit-suisse.com/corporate/en/research/research-institute/global-wealth-report.html>, 2018.
- [5] Capgemini. Global Wealth Report 2018 [EB/OL]. <https://www.capgemini.com/resources/world-wealth-report-2018-infographic/>, 2018.
- [6] 招商银行, 贝恩公司. 2017 年中国私人财富报告. images.cmbchina.com/cmbcms/201706/59d61c7c-8119-4444-92d5-af4ab6d5341f.pdf, 2017.
- [7] Markowitz H. Portfolio selection[J]. The journal of finance, 1952, 7(1): 77-91.
- [8] Modigliani F, Miller M H. The Cost of Capital, Corporation Finance, and the Theory of Investment: Reply[J]. American Economic Review, 1959, 49(4):655-669.
- [9] Hammer M. Reengineering work: don't automate, obliterate[J]. Harvard business review, 1990, 68(4): 104-112.
- [10] Venkatraman M P. The impact of innovativeness and innovation type on adoption[J]. Journal of Retailing, 1991, 67(1):51-67.
- [11] Smith W R. Product Differentiation and Market Segmentation as Alternative Marketing Strategies[J]. Journal of Marketing, 1956, 21(1):3-8.
- [12] Berry M, Linoff G. Mastering data mining: The art and science of customer relationship management[M]. John Wiley & Sons, Inc., 1999.
- [13] Chen I J, Popovich K. Understanding customer relationship management (CRM) People, process and technology[J]. Business process management journal, 2003, 9(5): 672-688.
- [14] Berry L L. Relationship marketing of services — growing interest, emerging perspectives[J]. Journal of the Academy of Marketing Science, 1995, 23(4):236-245.
- [15] Aharne M, Rapp A, Mariadoss B J, et al. Challenges of CRM implementation in business-to-business markets: A contingency perspective[J]. Journal of Personal Selling & Sales Management, 2012, 32(1): 117-129.
- [16] Geib M , Riempp G . Customer Knowledge Management[M]. Springer International Publishing, 2014.
- [17] Pedron C D, Picoto W N, Dhillon G, et al. Value-focused objectives for CRM system adoption[J]. Industrial Management & Data Systems, 2016, 116(3): 526-

- [18] Brown S A, Coopers P W. Customer relationship management: A strategic imperative in the world of e-business[M]. John Wiley & Sons, Inc., 1999.
- [19] Josiassen A, Assaf A G, Cvelbar L K. CRM and the bottom line: Do all CRM dimensions affect firm performance?[J]. International Journal of Hospitality Management, 2014, 36: 130-136.
- [20] Navimipour N J, Soltani Z. The impact of cost, technology acceptance and employees' satisfaction on the effectiveness of the electronic customer relationship management systems[J]. Computers in Human Behavior, 2016, 55: 1052-1066.
- [21] Xu G . Research on Customer Knowledge Management Based on CRM[C]// International Conference on Intelligent Human-machine Systems & Cybernetics. IEEE, 2014.
- [22] Ronald S. Accelerating customer relationships : using CRM and relationship technologies[M]. 2001.
- [23] Jafari Navimipour N , Rahmani A M , Habibizad Navin A , et al. Expert Cloud: A Cloud-based framework to share the knowledge and skills of human resources[J]. Computers in Human Behavior, 2015, 46:57-74.
- [24] Kaul D. Customer Relationship Management (CRM), Customer Satisfaction and Customer Lifetime Value in Retail[J]. Review of Professional Management, 2017, 15(2): 55-60.
- [25] Song M , Zhao X , Haihong E , et al. Statistics-based CRM approach via time series segmenting RFM on large scale data[J]. Knowledge-Based Systems, 2017:282-291.
- [26] Ennaji F Z , Abdelaziz E , Sadgal M , et al. Multi-Agent Framework for Social CRM: Extracting and Analyzing Opinions[C]// Computer Systems & Applications. IEEE, 2017.
- [27] Watts D J , Strogatz S H . Collective dynamics of ‘small-world’ networks[J]. Nature, 1998.