

基于大数据技术的客户全面感知数据分析研究与应用

随着新经济时代的来临，企业的战略中心正向“以客户为中心”转变，客户已经成为企业最重要的资源。电力企业拥有一个庞大的客户数据，通过对从各个渠道获取的客户需求数据获进行深入挖掘，分析客户对电力客户服务的需求具有重要价值。

林鸿¹ 徐伟杰¹ 韩毅平¹ 郑勇锋^{2/1}. 北京国网信通埃森哲信息技术有限公司
2. 北京国电通网络技术有限公司



林鸿/高级工程师

关键词/Keywords

大数据技术 ·
数据分析 ·
客户全面感知 ·
数据挖掘 ·

0 引言

随着新经济时代的来临，企业的战略中心正从“以产品为核心”向“以客户为中心”转变，客户已经成为企业最重要的资源。随着市场经济的发展，市场从生产主导转变为需求主导，从产品的时代转变为服务的时代，尤其是电子商务时代的到来，客户关系管理将成为企业间竞争的最核心的问题。企业要适应竞争的要求，就必须从“以产品为中心”的运营理念与方式转变为“以客户为中心”的运营理念与方式。

在输配网分离的改革背景下，各地区的供电企业甚至是民营企业的竞争也逐步走向正规，更多的大客户尝试直购电试点工作的推广，“优质客户”成为电力公司争取的对象，包括居民客户（通过电商平台、车联网、物联网等实现交叉销售），或者是工业客户（量大稳定、节能产品、客户内部电网运维等）。随着电力市场化改革的深入，供电企业也要适应打破垄断的环境，树立一切以客户利益为先的管理理念，努力提供服务质量。

电力企业拥有一个庞大的客户数据，使用传统的数据处理方式已经很难达到企业发展的需要，随着发展，现今的数据库已经从起初的数据整理、存储及对联系数据处理的支持转向了大数据重组、数据库分析、数据仓库以及与此相关的数据挖掘等更深层次的数据利用阶段。

通过对从各个渠道获取的客户需求数据获进行深入挖掘，分析客户对电力客户服务的需求及对供电企业的发展的引导作用，提升客户的精细化管理，把握日渐复杂的市场脉络，加快新能力的建设步伐，由此迅速扩大规模、捕捉全新机遇、积极开发非传统市场，为电力客服及客户引导性的业务发展提供数据支撑。

在十二五期间，国网营销信息化建设将建成“一个支撑，两大平台”（营销2.0），也即集约化的营销基础数据支撑、模块化的营销业务平台及智能化的客户服务平台。国网客服和营销的平台数据为使用大数据技术对客户全面感知的需求分析提供了良好的数据基础。

1 课题研究重点

课题使用大数据技术对客户全面感知的需求进行分析，项目的研究原理即大数据的洞察及分析挖掘原理。大数据技术使用新一代的技术和构架，以经济的方式，高速地捕获、发现和分析技术，从各种超大规模的数据中提取价值。数据挖掘是从大量的数据中，抽出出潜在的、有价值的知识（模型或规

则)的过程。聚类分析就是将一个数据集合分解成多个有相似特征的类的分析过程,它是一种重要的数据挖掘工具,是一种多变量统计方法,用来研究样本分类的问题。关联分析是数据挖掘领域常用的一类算法,主要用于发现隐藏在大型数据集中有意义的联系,所发现的模式通常用关联规则或频繁项集的形式表示。

本课题研究主要在大数据架构设计研究、用户感知需求分析数据整合和用户感知需求分析研究三个方面,具体包括:

1.1 源数据的收集和存储

目前客户全面感知涉及营销系统达15个,到2013年底,已经建成转运维的应用系统5个,推广阶段的应用系统6个,正在建设的应用系统4个,而且根据部署情况还存在增长的可能。由于涉及数据众多,客户感知所涉及源数据的收集和整理、接口及存储难度较大,移动相关的数据也存在数据格式的转化问题。

针对不同的来源数据,开发数据接口工具及整理转换工具。对目前营销系统所涉及的系统数据类型进行梳理,选取合适的开发平台及存储方式,整理不同类型数据的存储方法及其接口设计的最佳实践,从多个维度对数据不同的存储方法进行评估,选择最适合本项目的数据处理方法。

1.2 用户行为个体的服务痕迹识别和归集

国内电力公司目前对于用电客户基本信息和业务信息通常采用“认表不认人”的管理策略,也就是能够准确记录每一块电表是否有行度、是否有欠费,但是无法识别是哪一位客户(如房东还是租客)欠费、这位客户名下的所有物业是否都存在欠费情况,也无法识别到营业厅或者95598咨询的客户之前是否通过其他渠道有过类似沟通。

需要按照客户服务生命周期梳理电力公司与客户的每一个互动环节及渠道,通过应用视频识别、来电号码识别等技术手段体系化的收集、分类并整理客户跨渠道服务痕迹。对客户在服务过程中每一步骤、每一环节所留下的印迹进行跟踪记录,形成全双工模式的客户服务痕迹及感知管理。试点营业厅和现场作业车辆GPS系统需要进行相应的监控摄像头人脸识别功能识别、环境识

别功能升级;95598和营销系统需要进行配套的客户服务痕迹跟踪记录能力升级。

1.3 用户感知需求辨识研究

用户感知需求辨识,需要进行辨识处理用户需求感知的影响因素众多、关系复杂,需要建立考虑相关影响因素和关系的模型;根据需求辨识的需要,需建立量化的需求分析指标和权重设定体系。模型和指标体系的建立过程要考虑其复杂性和完整性,并根据需求的不同特点设置客户感知指标体系,并选择主观(专家打分法、德尔菲法)和客观(熵权法)相结合的方法科学确定指标和权重,通过指标分析确定客户感知需求的层次及需求紧迫程度,为下一步构建影响因素间的耦合关系提供数据支撑,降低了预测模型的复杂度,同时有效缩小了影响因素考虑范围,减少相应数据处理工作量。

1.4 用户感知需求分析研究

用户感知全面涉及营销客户系统,需要分析的用户感知需求多样,且分析角度均有变化,从不同角度的进行需求分析需要建立对应的分析模型,模型的建立存在多样性和复杂性。

用户感知需求数据之间存在关联性,用户的需求数据中隐藏着用户的需求的需求特性,对这些用电数据进行挖掘并研究用户类型,分析的结果可以帮助供电企业了解用户需求的共性及特性,并为预测模型的选择提供数据支撑。在解决过程中,应遵循以下步骤:一是需要对数据源进行梳理;二是在用户感知辨识的基础上开展功能需求分析,功能需求分析要紧密结合实际客户需求,要以“感知求全面、问题发现为重点、原因分析要关键、举措制定应有效”为原则,展开功能的系统化分析。

2 基于大数据技术的客户全面感知数据分析研究

信息和能源技术的飞速发展和大数据技术的广泛应用正深刻改变着电力行业的客户服务和创新模式。新一代的客户需求呈现移动性(如泛在互联)、虚拟性、个性化(如推荐服务)、社会性(如社交媒体)和极端数据(如富媒体、大数据)等鲜明的新特征。特别是从电力行业在重要

战略技术方面的走向来看，云计算、大数据等新型计算模式以及社会化网络应用等进一步凸显了这些特征的深化。研究在电力客户需求引入大数据技术对客户需求进行分析和挖掘，拟从四个方面开展研究与应用：

1) 从客户需求出发，以客户用电服务生命周期和客户全方位感知两个维度为切入点，利用现有客户服务数据和网络抓取数据进行挖掘，通过客户感知包括安全、可靠、精确、便捷和形象等多个方面的内容，研究建立综合评价基于客户感知的客户需求评价的整体框架。在基于大数据的客户全面感知分析方法的实现则依赖于越来越纯熟的数据挖掘术。数据挖掘按照其功能以及应用来划分，主要有：分类、关联、时间模式和聚类，可以具体应用到分析中的客户群体分类分

析、客户背景分析、客户满意度分析和客户的需求分析等。

2) 客户感知的客户需求分析及建模是以客户细分为基础的，客户细分可以采用分类的方法，也可以采用聚类的方法。比如，可以将客户分为高价值和低价值的客户，然后确定对分类有影响的因素，再将拥有相关属性的客户数据提取出来，选择合适的算法对数据进行处理得到分类规则。使用聚类的方法，则在之前并不知道客户可以分为几类，在将数据聚类后，再对结果数据进行分析，归纳出相似性和共性。每一类别的客户具有相似性的属性，而不同类别客户的属性也不同，从而确定特定消费群体或个体的兴趣、消费习惯、消费倾向和消费需求，进而推断出相应消费群体或个体下一步的消费行为。如图1所示。



图1 客户细分模型图

如图2所示，客户细分所采用的聚类分析方法是分类学发展而来的，聚类（clustering）是将类似的数据归类到一起，形成一个新的类别进行分析。对象的簇（聚类）的形成办法为：使得在一个簇中的对象具有很高的相似性，而与其他簇中的对象很不相似。所形成的每个簇可以看作一个对象

类，由它可以导出规则。随着多变量分析的逐步进入，聚类分析就逐步地独立与数值分类，形成了一个相对独立的分支。在众多的聚类算法中，比较常用的层次算法、划分算法都是研究的热点。近年来研究比较热门的划分聚类算法有K-means算法、模糊C-均值算法、K-调和均值算法以及谱聚类算法等。

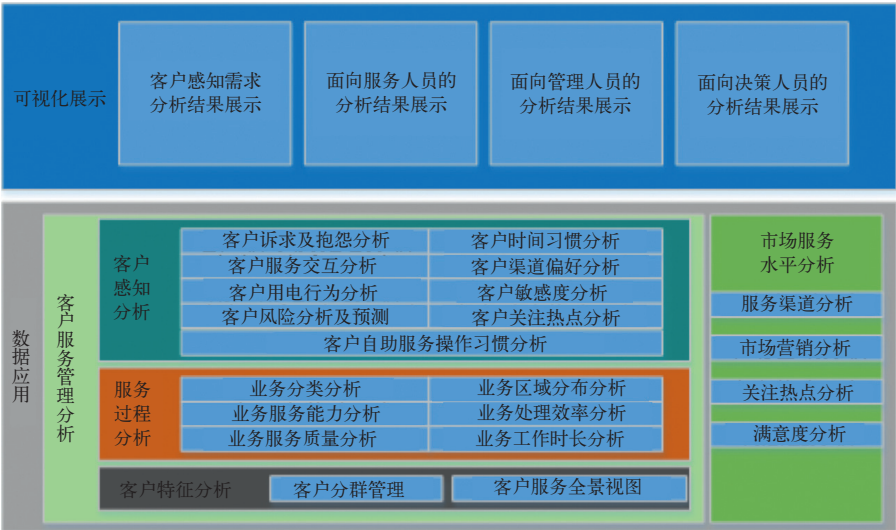


图2 系统总体架构图

客户需求信息是随时间动态变化的,目前通常使用的分析方法都是选用一个或多个因素对客户需求作单一的或综合的定性分析,这种方式不能精确地描述客户需求间的差异,只有综合运用定性和定量分析方法才能对客户需求做出真实量化的反映。客户需求分析是将客户按照客户需求划分、聚类为许多的客户群,分属于同一客户群的客户的需要和欲望极为相似;分属于不同客户群的客户对同一产品的需要和欲望存在着明显的差别。客户需求的多样化体现在客户对产品功能属性组合需求的多样化和不同属性水平组合的多样化两大方面。客户需求可以描述成一系列的产品属性(主要为功能属性),每个属性又分为不同的属性水平,不同属性及属性水平组合形成不同的需求模块,最终根据产品模块规划和模块组合获得某一市场细分区段偏好程度最高的模块化组合产品。

3 基于大数据技术的客户全面感知数据分析应用系统

开发客户全面感知数据分析原型系统,需要数据的准备及相关数据接口工具的开发,完成数据的清洗和分析功能,按照分析模型开发分析所需的算法模块;开发基于客户感知的客户需求分析的应用软件功能,对客户感知需求数据进行分类型、分过程及对不同用户类型进行应用分析,并加入对当前市场服务水平的分析,形成的分析结果进行面向不同操作人员的可视化展示。

3.1 系统业务应用架构

主要的分析类型有:

1) 客户类型细分及客户需求分析。客户不断嬗变,并表现出一系列的核心态度、偏好和行为特征。这些核心特征共同构成了新一代客户的特征。通过这些特征对客户进行分类,并有针对性地对不同类型的客户的需求进行分析。从产品创新、客户互动模式、新型合作关系以及数据在消费者日常生活中不断变化的作用等维度出发,深入探讨供电企业如何来顺应关键的发展趋势。

2) 客户诉求及抱怨分析。通过95598日志记录、社交网络客户诉求文本分析和客户语音识别等,结合生产运行、规划建设等与客户相关的数

据,综合分析客户诉求与抱怨,明确当前业务薄弱环节与问题。

3) 客户服务交互分析。通过追踪与分析同一用户在不同服务渠道上的服务痕迹,掌握用户客户的服务渠道偏好、使用痕迹,寻找有效影响客户的方法和服务接触点,为业务推广、渠道交互性提升和服务渠道引导提供实时和准实时数据支持。

4) 用户用电行为分析。通过海量的用户实时用电曲线、负荷曲线等计量数据的采集与分析,明确用户用电行为特征与模式,为引导开展有序用电及需求侧管理提供决策。

5) 客户风险分析及预测。通过客户的价值、需求与行为等多个维度的属性,开展客户的缴费行为分析和用电习惯分析,构建预测模型评判客户的欠费及用电风险,降低电费回收风险及用电安全风险。

6) 客户时间习惯分析。基于客户档案、来电区域以及客户服务记录,智能分析客户所进行交互服务的时间习惯。通过大数据分析,寻找客户服务的时间习惯。

7) 客户渠道偏好分析。基于客户档案、服务来源以及客户服务记录,智能分析某一特定客户群的渠道偏好行为,并建立客户的渠道偏好标签,如偏好电话服务、偏好微信服务等,开展个性化服务。

8) 客户敏感度分析。基于客户档案以及客户服务记录,智能分析某一特定客户群对于不同事件的响应频次,并建立客户的行为标签,如经常投诉客户、故障停电后反复报修客户等,开展有针对性的服务。

9) 客户关注热点分析。基于客户档案以及客户服务记录,智能分析某一特定客户群关注热点分析,可基于受理信息、录音到文本的通话记录,建立词语模型库,分析词语发生的频率,找出客户群关注的热点以及趋势。

10) 客户自助服务操作习惯。基于客户档案以及客户服务记录,智能分析某一特定客户群在IVR自动语音、网站、APP、短信和微信的操作轨迹,寻找客户操作的规律和客户聚焦内容,进而对自助服务进行优化,并对特定客户群提供个性

化服务。

3.2 数据架构

通过对大数据处理技术将这些数据转化为业务需要的数据模型，发现大数据中的“信息”和

知识，为行业内外提供大量的高附加值的内容增值服务。对数据特性、数据来源、数据获取方式和数据获取频率等不同角度对大数据进行分类、规划、设计及部署。如图3所示。

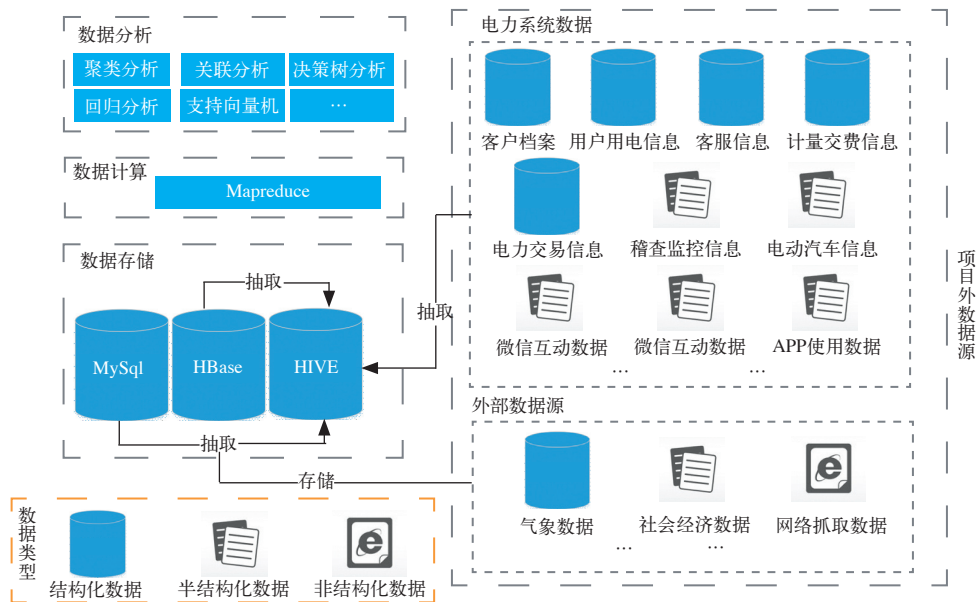


图3 数据架构图

数据处理和分析如图4所示，数据处理包括：

- 1) 大数据抽取。根据预定义的数据过滤校验规则，从相关的业务系统，如95598客服中心及营销客服其他相关应用的全量事实数据信息。
- 2) 分布式数据存储与计算。根据用电量分析业务要求，构建分布式数据存储数据库以及支撑分布式计算的平台。

- 3) 客户需求数据分析。在数据分析过程中，首先对初始数据进行轻度加工，形成分析型数据，然后通过分析模型的应用，得到分析结果数据并形成价值服务数据。
- 4) 客户需求分析结果可视化展示。对大数据应用基于准备数据，依据分析模型，分析统计出结果，需要对分析结果进行可视化展现以及验证。



图4 数据处理和分析示意图

3.3 技术架构

技术架构如图5所示。大数据的关键技术既包括数据分析技术等核心技术，也包括数据管理、数据处理和数据可视化等重要技术：

- 1) 数据抽取技术。对用电信息采集数据库中的海量各类电量数据进行数据抽取，抽取有用数据存入本系统数据仓库中。其数据库建库方式不同于

用电采集及其他联机事务的数据库，数据主要用于分析，数据设计按照面向业务设计原则进行设计。本系统数据来源于用电信息采集数据库，经过数据抽取、数据清洗转换和数据加载三个过程生成用于用户能耗分析。数据是通过数据抽取工具从用电信息采集系统中自动定期抽取，抽取的数据针对不同业务要求进行过滤，过滤掉脏数据和不完整数据形

成可用于用电分析的数据。

2) 数据分析技术。研究基于用户全面感知分析的需求数据库的数据挖掘技术,对大量需求数据进行数据准备、规律寻找和规律表示,挖掘的主要任务有数据关联、聚类、分类、异常分析、特异群组分析和演变分析等,以此实现用户需求准确分析,为服务及市场发展提供技术职称。

3) 数据服务技术。为展示客户端提供数据访问服务,数据服务内嵌内存缓存数据库,用于提高数据访问性能。通过数据服务提供的多种分析数据输出方式接口,为实时展现用电量变化情况提供强大支撑。研究建立与社会公共资源平台信息交互机制,将用户用电信息与节能方案快速推送给用户。

4) 数据展现技术。数据的全方位、多角度展现。借鉴与应用股票软件的优点,开发快速展现客户端应用程序,提出按照用户、地区、行业、产业和时间多种角度进行用电信息的实时动态展示方法,展现K线图、用电趋势图等,客户端具备本地存储与及时计算能力,使数据展现与指标分析更快速。

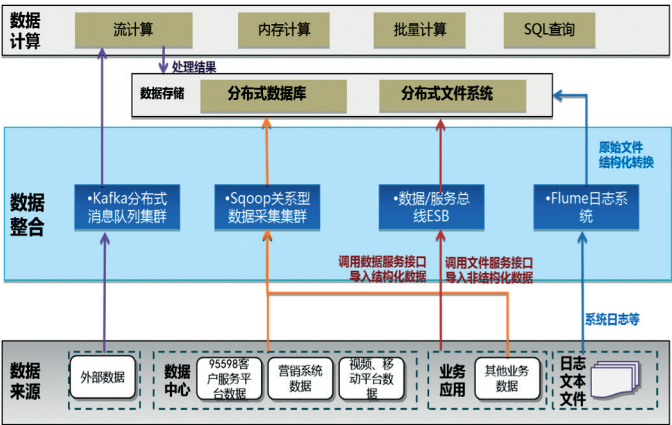


图5 技术架构图

3.4 安全架构

安全架构针对系统的具体业务特点,按照管理信息内、外网分离的安全要求建立起强健的信息安全防范体系,有效保护系统的信息安全,防范黑客和非法入侵者的攻击。系统应用的安全体系可分为技术层面的安全和管理层面的安全两个部分。技术层面的安全设计主要包括应用安全、数据安全、系统安全、网络安全和物理安全等,其中应用安全是业务安全防护体系的核心。管理层面的安全主要包

括安全组织及人员保证、安全管理制度和安全技术规范、安全考核及监督等内容。

4 结束语

本课题利用大数据技术对客户全面感知的需求进行分析,分析用户行为个体的服务痕迹识别和归集、用户感知需求辨识和用户感知需求分析研究,从客户需求出发,以客户用电服务生命周期和客户全方位感知两个维度为切入点,利用现有客户服务数据进行挖掘。研究基于大数据技术的客户全面感知数据分析应用系统,包括系统的业务应用架构、数据架构、技术架构和安全架构。

参考文献

[1] 萨科齐戴安迪·辛格, 尼尔玛拉·辛格. 大数据分析 [C]. 国际会议交流, 信息与计算技术, 孟买, 2012(10): 19-20.

[2] Robert Fonow. In charge of big data - to grasp the pulse of the information age [N]. China Brief, 2012.

[3] 孟小峰, 慈翔. 大数据管理: 概念, 技术与挑战 [J]. 计算机研究与发展, 2013, 50(1): 146-169.

[4] 王秀磊, 刘鹏. 大数据关键技术[J/Online]. 网络出版地址, <http://www.cnki.net/kcms/detail/34.1228.TN.20130627.0859.001.html>.

[5] 关志刚. 福布斯: 当智能电网遇见大数据 [J/Online]. 能源, 网络地址<http://www.ctocio.com/ccnews/6094.html>[2012.May10].

[6] 赵春雷. “大数据”时代的计算机信息处理技术[J]. 世界科学, 2012(2): 30-31.

[7] 陈康, 郑纬民. 云计算: 系统实例与研究现状[J]. 软件学报, 2009, 20(5): 1337-1348.

[8] 窦万春, 江澄, 等. 大数据应用的技术体系及潜在问题[R]. 江苏: 南京大学, 2013.

[9] 余立军. 从企业价值链看电网企业“三集五大”集约化管理[J]. 能源技术经济, 2011(5).

作者简介

林鸿 (1973-), 男, 博士, 高级工程师, 研究方向为电力系统信息通信。

(收稿日期: 2015-12-10) **EA**
(编辑: 崔文静)