

电力客服中心用户行为分析研究与实现

陈俐冰¹, 何容¹, 邱林¹, 王颖¹, 孙天昊²

(1. 国网重庆市电力公司 客户服务中心, 重庆 400017;

2. 重庆大学 计算机学院, 重庆 400044)

摘要:目前电力客服中心已经开展了多元化服务渠道为客户提供全方位的服务,但缺少对各渠道的深度挖掘和分析。针对这种情况,提出基于大数据进行电力客服中心用户行为分析,将为多个电力客户服务渠道建立智能在线监测分析系统。首先进行分布式多线程的数据采集,然后基于大数据从多维度进行用户在线行为分析,包括统计分析和聚类分析等。其中统计分析包括服务渠道指标、服务功能总指标、单个服务功能指标、客户访问时间指标、客户区域分布等,聚类分析采用K-means聚类根据忠诚度、使用频度、贡献度三个指标进行用户细分挖掘分析。实现的系统能够获得各渠道的客户行为特性,为企业全面掌握各服务渠道提供了支撑,有助于提供更为智能便捷的个性化服务。

关键词:大数据;电力大数据;数据挖掘;用户行为分析

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2017)02-0116-04

doi:10.3969/j.issn.1673-629X.2017.02.026

Research and Implementation of User Behaviors Analysis of Electric Power Customer Service Center

CHEN Li-bing¹, HE Rong¹, QIU Lin¹, WANG Ying¹, SUN Tian-hao²

(1. Customer Service Center, State Grid Chongqing Electric Power Corporation, Chongqing 400017, China;

2. College of Computer, Chongqing University, Chongqing 400044, China)

Abstract: At present, the electric power customer service center has already carried out the diversified service channels to provide customers with a full range of services, but the lack of depth mining and analysis for each channel. In view of this, the user behaviors analysis of electric power customer service center is conducted based on big data. An intelligent online monitoring and analysis system is set up for multiple power customer service channels. First it collects data using distributed multi thread, then makes online analysis of user behavior with multiple dimensions based on big data, including statistical analysis and cluster analysis. Statistical analysis contains service channel index, service function total indexes, service function single index, customer access time index, customer area distribution, etc. Cluster analysis uses K-means clustering to analyze user segmentation according to loyalty, frequency and contribution. The system implemented can get the customer behavior characteristics of each channel, provide support for enterprises to fully grasp the various service channels and help to give more intelligent and convenient personalized service.

Key words: big data; power big data; data mining; user behavior analysis

0 引言

随着数字信息化时代的迅猛发展,信息量也呈爆炸性增长态势。近年来,电力行业信息化发展迅速,特别是伴随着下一代智能化电网的全面建设,以物联网和云计算为代表的新一代IT技术在电力行业得到广泛应用,电力数据资源开始急剧增长并形成了一定的规模。

电力大数据是以业务趋势预测、数据价值挖掘为

目标,利用数据集成管理、数据存储、数据计算、分析挖掘等方面的核心关键技术,实现面向典型业务场景的模式创新及应用提升^[1]。电力大数据的应用将推动公司业务发展和管理水平。其中客户服务中心是公司优化整合服务资源,打造“全业务、全天候,服务专业化、管理精益化、发展多元化”的供电服务平台。

目前重庆电力客服中心已经开展了多元化服务渠道为客户提供全方位的服务,包括短信、网站、手机

收稿日期:2016-03-29

修回日期:2016-08-03

网络出版时间:2017-01-10

基金项目:国家自然科学基金资助项目(61472051);重庆市科技计划项目(cstc2013kjrc-qmrc40003)

作者简介:陈俐冰(1979-),女,经济师,研究方向为信息数据分析。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170110.0941.020.html>

APP、微信、有线电视等。

文中将基于电力大数据对电力客服中心的用户行为进行分析研究。用户行为的研究与心理学、社会学、社会心理学、人类学以及一切与网络行为的学科密切相关,它研究网络用户行为的规律性,借以控制并预测网络用户行为,并实现特定的目标。具体来说,网络用户行为研究就是分析网络用户的构成、特点及其行为表现出来的规律^[2]。

对短信、网站、手机 APP、微信、有线电视等多个电力客户服务渠道建立智能在线监测分析应用,进行分布式多线程的数据采集,从多维度进行用户在线行为分析,研究用户行为,各渠道的客户群体特性,为提供更为智能便捷的服务奠定基础。进行全面系统的用户行为分析,有助于掌握各渠道客户实际使用情况,为有针对性地进行各服务渠道的优化提升,进一步提供个性化服务定制奠定基础。

1 相关研究

目前基于大数据的用户行为分析最常用的技术是数据挖掘技术,包括大数据平台的构建、数据整合、数据挖掘、分析展示等方法。

文献[2-3]研究了基于大数据的网络用户行为分析。文献[4]实现了一个基于数据挖掘技术的电网用户行为分析系统进行用电客户细分、客户信用等级评估和欠费高风险客户预测。文献[5]提出了一个基于大数据,云计算和智能网格的关系分析的电力大数据平台的通用框架。电力大数据的关键技术包括大数据管理技术、大数据分析技术、大数据处理技术和大数据可视化技术。

文献[1,6-7]研究了大数据技术在电力行业的具体应用。其中,文献[1]利用大数据技术应用于客户服务中心,可实现面向典型业务场景的模式创新及应用提升,包括服务质量实时监控,热点问题集中处理等。文献[6]根据电力行业特征,分析了电力大数据的生成阶段、大数据的应用方式等。文献[7]讨论了大数据时代下电力企业数据挖掘的过程。

文献[8-11]研究了数据挖掘技术在客户服务中的应用。其中,文献[8]具体研究了数据挖掘技术在电信客户服务的应用。文献[9]提出了一个大数据应用框架来分析客户行为。首先客户相关数据被转化为拓扑数据结构,然后使用拓扑关系生成共生矩阵,并用来为客户行为分析推导马尔可夫链模型。文献[10]提出并实现了一个网络用户行为分析系统,该系统基于 Hadoop 分布式平台获取流量,分析用户行为。文献[11]通过个案研究,大数据分析支持创建,提高和增强各种商业服务,既能显著提高客户体验,又能为企业

创造价值。

2 用户行为分析系统研究与实现

基于目前大数据技术和数据挖掘技术,研究和实现了电力客户服务中心的用户行为分析系统。

2.1 系统流程

系统按照图 1 所示流程进行。先从多个数据源中基于大数据平台进行数据采集,具体数据源包括为客户提供全方位服务的短信、网站、手机 APP、微信、有线电视等多元化服务渠道所产生的客户服务数据,具体包括用电查询、电费缴纳、业务办理、信息订阅、服务开通等。接着对采集到的数据进行建模,包括数据约简和转换等环节,形成标准化的可以使用的数据,如统一标准的服务功能访问记录等。建模好的数据通过 Sqoop 工具进行抽取,整合到 Hadoop 大数据平台的分布式数据库 HBase 中。然后使用数据挖掘技术进行具体的用户行为分析,从多维度(如服务渠道指标、服务功能总指标、单个服务功能指标、客户访问时间指标、客户区域分布等)进行包括用户行为特征、客户群体特性、服务渠道特征等行为分析。最后进行可视化的结果展示,包括柱状图、曲线图、饼图和条形图等,为进一步决策提供支撑。

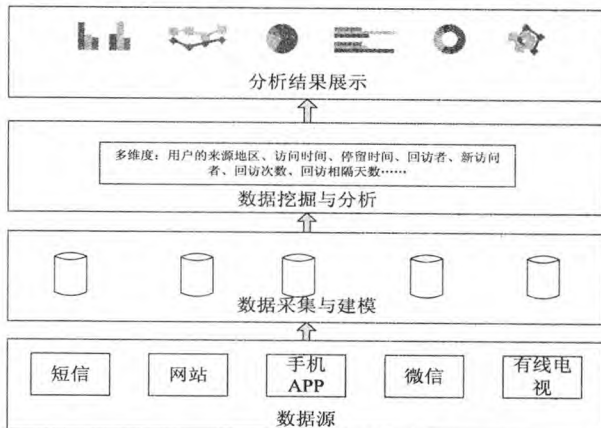


图1 用户行为分析流程

2.2 系统设计

用户行为分析系统的功能包括数据采集与建模、数据挖掘与分析、结果可视化展示等关键内容,如图 2 所示。

(1)数据采集具体包括数据源选择与数据源维护等功能。数据源选择,就是选择需要进行处理的数据源,通过数据接口的方式进行数据采集。数据源维护,就是数据源相关信息(如名称、链接地址、用户名和密码等信息)的修改。

(2)数据建模主要加工在数据采集中所选择的数据源,包括数据清洗、数据集成、数据归约和数据转换四个功能。数据清洗是在确定清洗的数据属性后对所

选择的数据源进行清洗加工。数据简化主要是对数据源中某些数据按月份或年份做汇总处理。数据归约是对所选择的数据源中某些不相关属性利用直接删减法、信息增益法等进行处理。数据转换包括数据标准

化和数据泛化,主要是对数据源中的连续值和离散值做不同的处理。针对数据源中的连续值,采用标准化作为数据转换的方法,而针对数据源中的离散值,采用数据泛化的方法。

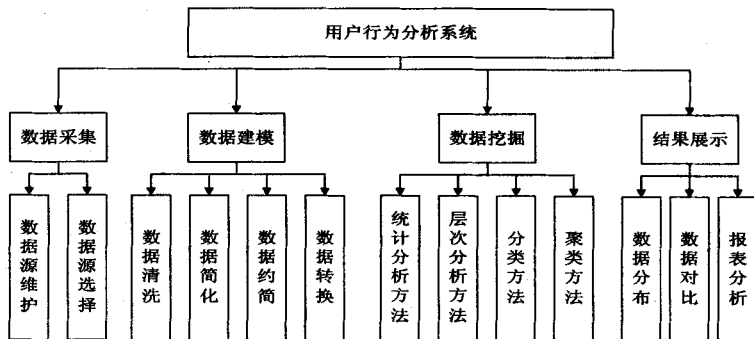


图 2 用户行为分析系统的功能

(3)数据挖掘模块,可选择使用统计分析方法、层次分析法、KNN 分类算法和 K-means 聚类算法等对用户行为进行具体的分析和预测。

统计分析方法是指对目标数据集中的关系属性进行整理归类并进行解释的数据挖掘方法^[12]。比较常用的有:因子分析、相关分析、判别分析、回归分析等。

K-means 是一种基于样本间相似度度量的非监督学习的聚类算法^[13]。可以用来对用户进行分类等。首先随机选取 k 个点作为 k 个初始聚类中心,然后对剩下的点依据对 k 个聚类中心的相似度划归到对应的聚类中,再更新每个聚类的中心,不断迭代,直到收敛。

(4)结果可视化展示对挖掘出的结果进行数据分布、数据对比、报表生成。数据分布可以分别按照客户类别、渠道、功能等不同的分类条件输出相应的统计信息,并以饼图或柱状图的直观化形式呈现。数据对比可以分别按照渠道不同的对比属性输出相应的对比信息,并以柱状图形式输出对比结果。报表生成分为年报表和月报表,按照渠道的不同属性生成相应的报表。

2.3 系统实现

最终实现的系统主要完成了各类指标的统计分析和用户细分。

(1)统计分析。

系统分别从服务渠道指标、服务功能总指标、单个服务功能指标、客户访问时间指标、客户区域分布等几个方面进行统计分析。其中,对电力客服中心的用户行为的统计分析包括以下几个指标:

①服务渠道指标:对每个服务渠道统计出渠道的总用户数、访问次数和访问总时长。

②单功能指标:对每个服务功能统计出其用户数、访问时间和访问时长。

③功能总指标:对每个服务功能统计出各服务功能的总用户数、访问次数和访问总时长。

④用户喜好指标:从访问渠道、访问时长、所访问服务功能三个方面统计用户的喜好。

⑤用户区域分布指标:统计用户所属区域的分布情况,给出分布图及其各占比例。

数据分布和对比包括:

①指标在时间上的分布:包括指标在不同时间粒度(日、周、月、年)上的分布;分布对比(同一指标在不同时间粒度上的对比,不同指标在同一时间粒度上的对比)。

②指标在空间上的分布,包括不同指标在一个地区上的分布;同一指标在不同地区分布的对比。

③时间和空间的结合:指标在同一时间粒度不同地区的分布,指标在不同时间粒度同一地区的分布,分布对比。

用户喜好分析包括渠道、时间、业务等几方面,具体为:

①渠道爱好:以用户在一段时间内对各个渠道的访问次数作为衡量用户对渠道的喜好标准。

②时间爱好:以用户在各个时段访问时长作为衡量用户时段喜好的标准。

③功能爱好:以用户在一段时间内对各个服务功能访问次数作为衡量用户对功能的喜好标准。

系统实现的数据分析界面如图 3 所示。

其中,渠道分析分别向系统用户展示每个渠道的用户数、访问次数、访问频度等,用统计图和表格两种方式显示。服务功能总指标分析以统计图和表格的方式显示不同渠道下各服务的用户数、访问次数、访问总时长。服务历年走势分析中系统用户选择待分析服务和起止年份,选择待分析指标(用户数、访问次数、访问总时长)后生成趋势图。客户访问时间分析中系统用户选择待分析服务,设置起止时间后点击生成分析图。客户区域分布分析统计出客户所属区域的分布情

况并以饼状图和表格两种形式展示结果。

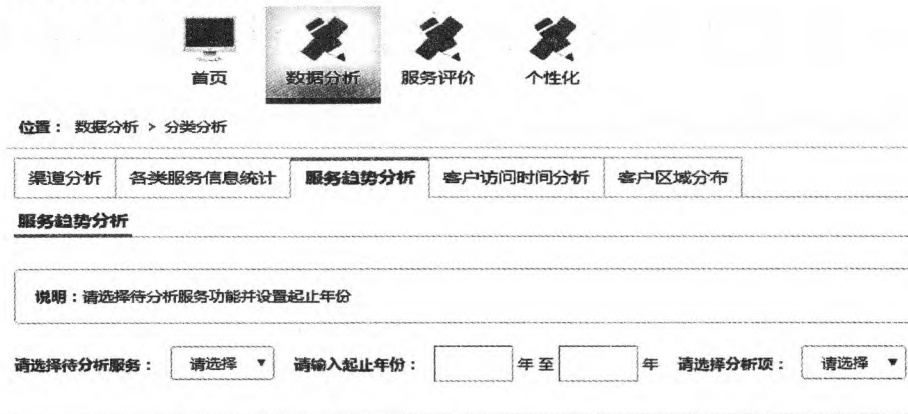


图 3 用户行为分析系统界面

(2) 用户细分。

客户细分是把客户按照一定的标准划分为多个不同的客户群,以便有针对性地实行不同的营销策略^[14]。具体的客户细分指标可以是用户行为的不同、行为习惯的不同或者客户需要的不同等等。对客户进行正确的细分,是电力企业制定营销策略的起点。

目前,电网企业的一些静态分类方式,例如根据行业类别、用电类别、用电方式、电压等级等等,并不能很好地反映出客户的行为习惯和动态需求,而在实际情况中,客户的忠诚度、客户对各渠道服务的使用情况以及客户对企业利润的贡献等指标得到了极大的关注。文中使用 K -means 聚类技术根据这三个指标进行用户细分挖掘分析。

①忠诚度:主要表现在客户从开户至今所经历的时间(以月为单位)。

客户保持时长 = 当前年月 - 开户年月

②使用频度:是以客户对各渠道服务的使用频度情况。使用情况可由用户对服务的月平均点击次数来衡量。

使用情况 = 客户对各服务总的点击次数 / 客户保持时长

③贡献度:主要以客户的月平均用电量来衡量。该指标是电网公司关注的重点指标,因而也是用电客户细分的关键指标之一。

月平均用电量 = 累计用电量 / 客户保持时长
系统实现的客户细分界面如图 4 所示。

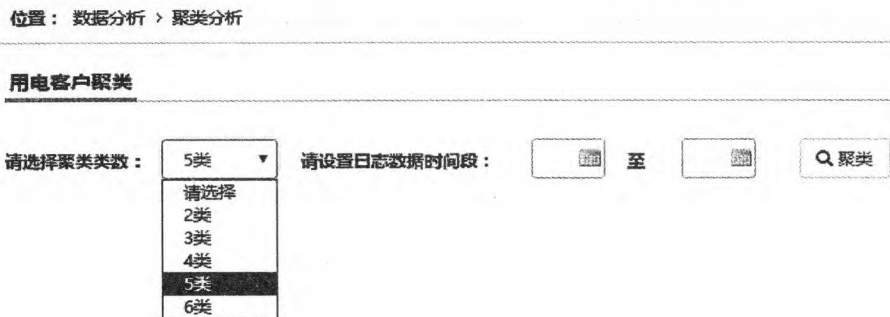


图 4 客户细分界面

在用户设置好聚类的 K 值以及时间段之后,系统使用 K -means 聚类算法训练学习得出用户细分结果。

3 结束语

基于大数据技术和数据挖掘技术,研究和实现了电力客户服务中心的用户行为分析系统。该系统已投入实际运行,能够根据短信、网站、手机 APP、微信、有线电视等多元化服务渠道所产生的客户服务数据进行用户行为分析,包括用户行为特征、客户群体特性、服务渠道特征等。通过用户行为分析,发现用户行为规律,对企业 and 用户都有重要意义。对企业来说,用户行

为分析可以发现用户对服务渠道、业务的使用特征,帮助企业进行合理资源规划,使得资源在最大限度满足用户需求的同时,降低运营成本;帮助企业了解用户对服务渠道、业务的兴趣所在,从而为企业制定个性化服务提供依据。对用户来说,企业根据用户行为进行了服务优化,能够更快更方便地使用网络服务。

参考文献:

[1] 陈超,张顺仕,尚守卫,等. 大数据背景下电力行业数据应用研究[J]. 现代电子技术,2013,36(24):8-11.

(下转第 124 页)

点。但是,利用该理论在确定评价样本矩阵时,由于专家对各指标的理解不同,其评分的主观依赖性较强。因此,还需进一步对评价数据进行差异度分析,并在平台实际应用和推广中不断验证。该研究对定量评价农业信息平台做了一次有益的探索与实践,对其他各类农业信息平台的定量分析评价具有参考借鉴价值。

参考文献:

- [1] 陈关胜. 基于灰色关联度分析法的电子商务网站评价研究[D]. 南昌:江西财经大学,2013.
 - [2] 冯伟,王修来,马宇玲,等. 基于多层次灰色理论的科技资源整合效果评价模型[J]. 技术经济,2009,28(5):16-20.
 - [3] Li Min, Ren Yueying, Zhu Zhu. Usability evaluation of independent-sales B2C fashion website based on consumer's perspective[J]. iBusiness, 2013, 5(3):96-99.
 - [4] 唐飞. 我国新农村建设与多层次灰色理论评价模型研究[D]. 重庆:重庆大学,2009.
 - [5] Li Fangyu, Li Yefei. Usability evaluation of e-commerce on B2C websites in China[J]. Procedia Engineering, 2011, 15(1):5299-5304.
 - [6] 刘虹,孙建军,郑彦宁,等. 网站评价指标体系设计原则评述[J]. 情报科学,2013,31(3):156-160.
 - [7] 周蓓婧,侯伦. 基于层次分析法的团购网站综合评价研究[J]. 管理学家:学术版,2012(2):50-66.
 - [8] 王谨乐. 我国电子商务网站综合评价研究与应用[D]. 合肥:合肥工业大学,2008.
 - [9] 刘雨薇. 我国 B2C 电子商务平台的综合评价研究[D]. 长春:吉林财经大学,2013.
 - [10] 丁志同. 基于多层次灰色理论的高校教师素质绩效评价[J]. 统计与决策,2011(15):67-69.
 - [11] 张健,张再生,李广. 基于多层次灰色理论的乡镇基层公务员胜任力评价[J]. 工业工程,2011,14(2):26-30.
 - [12] 胡容,黄南京. 基于多层次灰色理论的综合评价模型及应用[J]. 四川大学学报:自然科学版,2005,42(5):889-895.
 - [13] Li Hongjun, Suen C Y. A novel non-local means image denoising method based on grey theory[J]. Pattern Recognition, 2016, 49:237-248.
 - [14] 郑运虎,姜峰. 基于多层次灰色评价理论的海洋输油管道安全评估方法[J]. 辽宁石油化工大学学报,2014,34(4):22-26.
 - [15] Liu Sifeng, Forrest J, Yang Yingjie. A brief introduction to grey systems theory[J]. Theory and Application, 2012, 2(2):89-104.
- +++++
- (上接第 119 页)
- [2] 任思颖. 基于大数据的网络用户行为分析[D]. 北京:北京邮电大学,2015.
 - [3] 刘淑鑫. 面向大数据的网络用户行为分析研究[D]. 上海:东华大学,2015.
 - [4] 林嘉晖. 基于数据挖掘的电网用户行为分析系统的设计与实现[D]. 广州:中山大学,2013.
 - [5] Zhan Jie, Huang Jinxin, Niu Lin, et al. Study of the key technologies of electric power big data and its application prospects in smart grid[C]//Proceedings of 6th IEEE PES Asia-Pacific power and energy engineering conference. [s. l.]: IEEE, 2014.
 - [6] 赵云山,刘焕焕. 大数据技术在电力行业的应用研究[J]. 电信科学,2014,30(1):57-62.
 - [7] 卢建昌,樊围国. 大数据时代下数据挖掘技术在电力企业中的应用[J]. 广东电力,2014,27(9):88-94.
 - [8] 陶晓英. 大数据支撑下的客户服务应用探索[J]. 信息通信技术,2014,8(6):14-20.
 - [9] Zin T T, Tin P, Toriu T, et al. A big data application framework for consumer behavior analysis[C]//2013 IEEE 2nd global conference on consumer electronics. [s. l.]: IEEE, 2013:245-246.
 - [10] Guan Jianfeng, Yao Su, Xu Changqiao, et al. Design and implementation of network user behaviors analysis based on Hadoop for big data[J]. Communications in Computer and Information Science, 2014, 490:44-55.
 - [11] Ankur B, Nikita M, Swadesh P, et al. Forecasting consumer behavior with innovative value proposition for organizations using big data analytics[C]//Proceedings of 4th IEEE international conference on computational intelligence and computing research. [s. l.]: IEEE, 2013:1-4.
 - [12] Liu Hong. Discussion on the statistical analysis method[C]//7th international joint conference on computational sciences and optimization. [s. l.]: [s. n.], 2014:383-385.
 - [13] Khandare A, Alvi A S. Survey of improved k-means clustering algorithms; improvements, shortcomings and scope for further enhancement and scalability[M]//Information systems and intelligent application. [s. l.]: [s. n.], 2016:495-503.
 - [14] 耿筱媛,张燕平,闫屹. 改进的 K-means 算法在电信客户细分中的应用[J]. 计算机技术与发展,2008,18(5):163-167.