

DOI: 10.7667/PSPC171615

基于数据挖掘的电动汽车用户细分及价值评价方法

张禄, 李国昌, 陈艳霞, 孙舟, 王伟贤, 田贺平

(国网北京市电力公司电力科学研究院, 北京 100075)

摘要: 用户细分可以掌握不同电动汽车用户充电行为的特征及其之间的差异性, 对充电服务运营企业具有重要意义。基于运营管理系统迅速积累的大量充电服务数据, 对全量数据进行探索性分析, 筛选出细分模型关键变量, 给出了基于数据挖掘技术和 K 均值(K_MEANS)聚类算法的电动汽车用户细分方法, 提出了电动汽车用户价值评价方法。针对北京地区电动汽车用户开展分析并得到用户行为特征及价值评价结果。相关结论可为运维管理机制优化和精准营销策略制定提供数据支撑。

关键词: 电动汽车; 用户细分; 数据挖掘; K_MEANS 算法; 用户价值评价

Customer segmentation and value evaluation method based on data mining for electric vehicles

ZHANG Lu, LI Guochang, CHEN Yanxia, SUN Zhou, WANG Weixian, TIAN Heping

(Electric Power Research Institute, State Grid Beijing Electric Power Company, Beijing 100075, China)

Abstract: Customer segmentation is of great significance for charge service operators to obtain the features of charging behaviors and individual differences between various Electric Vehicle (EV) users. Based on large numbers of charging service data fast accumulated by operations management system, exploratory data analysis is applied to all the historical data in the database. Firstly key variables are screened out to the segmentation model, and then the EV customer segmentation method by data mining technique and K -MEANS algorithm is presented. Secondly the customer value evaluation method is proposed, and charging behavioral features and customer values are analyzed based on Beijing EV customers. At last, conclusions and suggestions are given, which would provide data supports for the improvement of operation and maintenance management and decision-making of the precision marketing.

This work is supported by Science and Technology Project of State Grid Corporation of China (No. 52020116000J).

Key words: electric vehicle; customer segmentation; data mining; K -MEANS algorithm; customer value evaluation

0 引言

近些年, 在大城市汽车保有量迅速增加和实施大气污染治理、降低化石能源消耗的双重压力下, 节能环保的电动汽车获得了政府和车企的大力推广应用, 车辆技术的进步和充电网络的完善, 也使电动汽车得到了更多消费者的认可和青睐^[1-5]。随着电动汽车数量的迅速增长, 用户日常出行不但对公共充电服务的需求不断增加, 而且对充电设施运营单位的服务水平提出了更高要求^[6-7]。充电设施运营单位为了提高精益化管理水平、增加盈利并改善用户充电体验, 有必要对电动汽车用户开展客户细

分研究, 得到精准营销策略, 实现差异化服务。

以客户为导向的细分方法主要包括四类: 基于客户统计学特征、基于客户行为、基于客户生命周期和基于客户价值, 其中行为细分包括 RFM 分析和客户价值矩阵分析法等, 利益细分包括拟合分析、因素分析、聚类分析和人工神经网络等^[8]。文献[9]以 RFM 为分析变量利用层级分析法对“4×4”型的自组织神经网络模型将价值客户划分为 16 类, 跟 8 种类型分类方法更为细致准确。文献[10]给出了考虑权重的 RFM 模型来分析客户行为规律, 是对经典 RFM 方法的改进。文献[11]通过引入总利润属性改进 RFM 模型, 创建 RFP(Recency, Frequency, Profit)模型, 并对比了两种细分模型的聚类效果。上述三类 RFM 及其改进模型的分析和应用技术均基于传

基金项目: 国家电网公司科技项目资助(52020116000J)

统基础变量,可选择变量相对固定且应用不够灵活。文献[12]利用因子分析法提取细分变量,设计了多指标客户细分模型,能够更好地解释客户行为特征。文献[13]通过引入大均值子矩阵双聚类算法,从客户样本和消费属性的两个维度进行双向聚类,能够挖掘高价值客户群体。文献[14]利用最大距离等分策略选取初始聚类中心,并采用评价函数确定聚类数目,降低了算法结果对预设参数的依赖程度。上述三类利益细分方法在分析维度和应用广度方面都具有更大弹性,能够透过客户行为表象挖掘出更深层次的利益本质。文献[15]提出了基于数据挖掘技术的移动通信行业客户细分方法,利用数据挖掘中的误差平方和准则函数的 K 均值(K_MEANS)聚类算法,建立客户细分模型,为企业制定营销策略提供依据。文献[16]基于 ReliefF 算法建立特征数据模型,并通过 K 均值算法进行聚类分析,可有效提升居民用电行为分类的准确率。文献[17]利用模糊 C 均值聚类法对重要企业用户的用电负荷进行聚类,实现需求侧管理的有序用电目标。综上所述,客户细分技术在零售、金融和移动通信等行业已经得到广泛应用,但鲜有在电动汽车充电网络运营服务领域应用的相关报道。

本文针对北京地区公共领域的电动汽车用户开展充电行为细分研究,采用“大数据”分析思路,在对全量充电服务数据进行探索性分析的基础上,确定细分模型关键变量,采用 K_MEANS 聚类算法完成用户分类,并提出电动汽车用户价值评价方法,给出了评价指标规则及相关分析结论与建议。

1 基于数据挖掘的用户聚类分析法

1.1 聚类分析法概述

聚类分析的目标是将所关心的对象按照一定的规则或标准分成不同的类别,包括非模糊聚类和模糊聚类。非模糊聚类的对象与类之间是确定的从属关系,而模糊聚类中的对象与类之间则是一定概率的从属关系。

1.1.1 聚类分析法的类型

常见的聚类分析法包括层次法和划分法。层次法是指先将观测视为一个整体,然后每次都最相似的两个类合并成一个新的类,直至所有观测达到所预定的分类条件为止。划分法是指在开始就指定某几个类中心,再通过计算将每个观测暂时归到距离其最近的类中心所在的类,继续调整类中心直至收敛。两种聚类分析方法的应用适用性和局限性对比如表 1 所示。

表 1 层次法和划分法的应用特点对比

Table 1 Application comparison of analytic hierarchy process and division method

聚类方法	适用性	局限性
层次法	不需要输入类别,适用于观测数比较少的情形	① 不断将观测合二为一,过程中不恰当的合并会一直延续至结束。 ② 算法多,算法间各有优缺点,没有哪一种算法是绝对有效的。 ③ 不适用于观测值较多的数据。
划分法	适用于观测数据比较多的情况	① 需要提前输入类别数 K 。 ② 需要提前制定 K 个类别的中心。

1.1.2 距离度量方法

聚类分析法的一个关键问题是如何度量两个对象间的相似性。通常,采用 $d(x,y)$ 表示观测 x 和 y 之间的距离,距离越小表示观测之间越相似。

一般情况下,定义距离需要满足 4 个条件:

- 对称性: $d(x,y)=d(y,x)$
- 如果 $x \neq y$, 则 $d(x,y) > 0$
- 如果 $x = y$, 则 $d(x,y) = 0$
- 三角不等式: $d(x,y) \leq d(x,z) + d(z,y)$

以上 4 个条件:对称性表示观测 x 到 y 的相似程度等于观测 y 到 x 的相似程度;第 2 个条件表示如果观测 x 和 y 是两个不同对象,那么它们之间的距离大于 0;第 3 个条件表示如果观测 x 和 y 是两个相同的对象,那么它们之间的距离等于 0;三角不等式是从数学的角度把度量限制在欧式空间,即观测 x 和 y 之间的距离是直线距离,不是不规则的曲线。

常用的距离函数包括欧氏距离、街区距离和闵可夫斯基距离等。

欧式距离是勾股定理在多维空间的推广,公式为

$$d(x,y) = \sqrt{\sum_{i=1}^k (y_i - x_i)^2} \quad (1)$$

街区距离是以线段 xy 为斜边的直角三角形的两直角边的距离之和,其中 $x=(x_1, x_2, \dots, x_k)$ 、 $y=(y_1, y_2, \dots, y_k)$,街区距离的公式为

$$d(x,y) = \sum_{i=1}^k |y_i - x_i| \quad (2)$$

闵可夫斯基距离公式为(其中 λ 为正数)

$$d(x,y) = \left(\sum_{i=1}^k (y_i - x_i)^\lambda \right)^{\frac{1}{\lambda}} \quad (3)$$

街区距离和欧式距离都是闵可夫斯基距离的特殊形式,二者分别对应 λ 值为 1 和 2 的情况。这

三种距离均用来处理连续变量。

1.2 基于数据挖掘的用户聚类分析法

数据挖掘是指将冗杂的数据经过科学处理后提取出具有潜在价值的信息的过程。数据挖掘利用数据库、人工智能、机器学习、神经网络、数理统计、模式识别、高性能计算和数据可视化等多领域技术从庞大数据源中寻找出有价值的隐藏信息，加以分析并归纳成结构模式，帮助企业进行科学决策。

基于数据挖掘的用户聚类就是借助数据挖掘的深层次数据分析方法，对数据库中的全量用户行为变量进行精确化细分，能够对用户行为模式与用户价值进行准确判断与分析，从而掌握用户需求和行为习惯等信息。

与传统的用户聚类相比，基于数据挖掘的用户聚类方法能够提高分类的准确度，对用户行为进行量化描述，其优点如下。

(1) 能够客观反映用户群体的内在特征：基于数据挖掘的用户聚类法相对客观，通过剖析数量有限的客户组特点，深刻了解客户群体构成，并针对不同客户组制定营销方案。

(2) 综合反映对客户多维度的认识：基于数据挖掘的用户聚类法基于用户的多方面特性进行多个维度的分类，全面掌握用户特征，增加营销方案的准确性。

(3) 可以更加深入细致地了解用户价值。

(4) 便于对用户的动态跟踪，对于迅速变化的市场需求，实现市场的动态跟踪。

1.3 K _MEANS 聚类法

聚类方法的选择既要考虑分析目标和对象的特点，又要符合实际需求，并满足算法效率和结果合理性要求。 K _MEANS 聚类法是较常用的聚类分析算法，其重要特点是能够使类群内具有较高相似度，而类群之间的相似度较低，且算法收敛的时间与待分析数据的观测数成正比。本文针对北京地区电动汽车公共充电设施用户及其充电行为开展用户细分研究，用户量及其充电数据量较大，这里选择应用较为成熟的 K _MEANS 聚类法，并利用数据挖掘技术完成细分关键变量的筛选和确定。

K _MEANS 聚类方法的步骤如下：

(1) 选定 K 个观测作为 K 类种子。

(2) 读入所有观测，计算每个观测与 K 个种子之间的距离，并将观测暂时归类到与其距离最近的种子所在的类中。

(3) 根据现有类中的观测，重新计算类的中心，即种子。

(4) 重复第(2)、(3)步，直至收敛。至此，所有 K 类种子最终确定。

(5) 再次读入所有观测，将每个观测归类到与其距离最近的种子所在的类，分类结束。

其中 K 值是预估的， K 值的估计方法包括根据背景知识判断、根据分类目标判断和作图辅助判断等。

1.4 K _MEANS 聚类法在 SAS 中的应用

SAS 软件是由美国 SAS 软件研究所研制的一个模块化、集成化的大型应用软件系统，包括数据访问、数据存储、应用开发、图形处理、数据分析、预测等功能。采用 SAS 中的 PROC FASTCLUS 实现 K 均值聚类法，使用欧式距离计算观测之间的距离。对种子的选择可以由分析人员指定，也可以由过程步在原始输入数据集中选择，一般该过程可以自动选出“足够好”的初始种子。

PROC FASTCLUS 过程步在使用中具有以下特点。

(1) 适用于分析较大数据集：当数据集较小时，该过程步对数据集中的观测顺序较为敏感，同一数据集观测顺序的变化会对结果产生较大影响。

(2) 在该过程步中，方差较大的变量对结果的影响也很大，在使用该过程步之前，应考虑先对数据集进行标准化。

2 数据处理与关键变量筛选

2.1 数据预处理

原始数据来源于电动汽车充换电运营管理系统数据库，包括充电站数据、充电桩数据、用户数据和充电交易记录数据。数据预处理涉及到原始数据表的匹配合并、全量数据质量治理、关键变量筛选和数据标准化等过程。

2.1.1 原始数据表合并

系统数据库的原始数据分散存储在各类结构化的数据表格中，根据用户细分需求对数据表格及其数据进行匹配合并处理，主要操作过程包括：

(1) 根据分析目标筛选相关数据表，包括 300 余条充电站档案信息、3 500 条充电桩档案信息、1.5 万余用户信息以及 150 万余条充电交易记录。

(2) 选择充电设施编号、用户充电卡编号等关键数据项，将充电设施和用户等档案信息与充电交易记录等交易数据进行匹配。

(3) 通过数据表的合并处理，得到 145 万余条全量信息数据，并对 5 万余条匹配失败的异常交易数据进行标记。

2.1.2 数据质量治理

系统数据库存储的原始数据受到终端设备采集准确度、通信信道干扰和系统设计缺陷等因素的影响, 需要经过数据筛选和处理环节优化数据质量, 在全量信息中提取出有效的数据项作为分析模型的输入变量。

文献[18]给出了一种基于 Spark 的并行聚类算法辨识不良数据的方法, 而文献[19]详细介绍了电动汽车充电数据质量的治理技术。基于电动汽车充电服务业务数据之间的数理和逻辑关系, 通过在系统中设计与应用数据筛选与处理模块, 能够有效发现并修复异常数据, 保障下阶段数据分析工作的准确性。

2.2 关键变量筛选

为确定用户细分模型的输入变量, 对所有电动汽车交易记录进行探索性数据分析(Exploratory Data Analysis, EDA), 并对原始变量进行加工, 获得充电量总和、平均卡内余额、交直流充电量占比等 58 个衍生变量, 从用户的充电电量、频度、效率和效益、卡内余额、充电“近度”(最后一次充电到现在的时间长度)、交直流偏好等方面描述了用户行为。

为避免分析模型输入变量之间具有相关性, 需要对变量进行相关性检验分析。相关性检验是指检验两变量是否存在相关关系的一种假设检验, 其中原假设 H_0 和备择假设 H_1 分别为

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases} \quad (4)$$

式中, ρ 是相关系数的参数。通常, 当 $P < 0.05$ 时, 表示两个变量之间的线性关系是显著的。但 P 值的大小不能表示相关性的强弱, 且 P 值会受到样本容量的影响。

通过相关性检验分析, 筛选出 9 个本次研究中可用的无相关变量。在与业务专家人员进行沟通交流后, 最终确定 5 个关键变量作为用户细分模型的输入变量, 如表 2 所示。

表 2 关键变量筛选结果

Table 2 Screening results of key variables

名称	缩写	含义描述
充电总电量	Q	用户充电的累计总电量
充电“近度”	R	最后一次充电到现在的时间长度
充电卡余额	M	用户充电卡内的平均余额
持卡时长	T	用户开通充电卡的时长
充电效率	P	用户平均的实际充电时间(充电电流不为零)与充电桩占用时间的比值

2.3 数据标准化

为避免不同单位量纲的输入变量对聚类模型中的距离计算造成影响, 需要对关键变量进行数据标准化处理。用户聚类分析中的变量通常可以分为三种^[20], 分别为:

(1) 增益变量, 即该变量的值越大, 客户价值越大, 包括充电总电量、充电卡余额、持卡时长和充电效率。增益变量的标准化公式为

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

(2) 损益变量, 即该变量的值越小, 客户价值越大, 例如充电“近度”, 即用户最后一次充电到现在的时间长度。损益变量的标准化公式为

$$x' = \frac{x_{\max} - x}{x_{\max} - x_{\min}} \quad (6)$$

(3) 适度变量, 不能用增益和损益来衡量, 通常采用全量的平均值作为最优值。适度变量的标准化公式为

$$x' = \frac{x - x_{\text{mean}}}{x_{\max} - x_{\min}} \quad (7)$$

式中, $x_{\text{mean}} = \frac{1}{n} \sum_{i=1}^n x_i$, x_{mean} 是该变量全体的平均值。

3 用户细分算法及用户价值评价

3.1 基于 K_MEANS 算法的用户细分方法

基于上述关键变量筛选和相关性检验分析结果, 对表 2 所示的 5 个关键变量进行标准化处理, 并通过分析全量数据得到关键变量的分布情况, 如表 3 所示。获得各变量分别在其全量数据中 1%、25%、50%、75%、90%和 99%位置的数值, 有利于了解每个数据的位置分布情况, 并帮助后续的分类结果评价。

表 3 关键变量数值分布

Table 3 Value distribution of key variables

变量	最小	最大	1%	25%	50%	75%	90%	99%
Q	-0.46	54.33	-0.46	-0.44	-0.37	-0.08	1.03	3.57
R	-0.63	4.98	-0.63	-0.63	-0.63	0.25	1.44	3.80
M	-0.84	18.47	-0.76	-0.46	-0.24	0.10	0.74	3.96
T	-0.74	4.70	-0.74	-0.74	-0.45	0.40	1.26	3.27
P	-1.84	3.64	-1.84	-0.61	-0.12	0.47	1.23	2.97

基于 SAS 软件中的 K_MEANS 算法, 对所有电动汽车用户的充电交易记录数据中的关键变量进行聚类分析, 预设用户分为 10 类, 得到的聚类结果如表 4 所示。其中聚类 1 的中心观测为(-0.46, 4.98, 6.97, -0.74, -1.84), 其他 9 类的聚类中心以此类推。

表 4 聚类中心分布

Table 4 Distribution of clustering centers

聚类	<i>Q</i>	<i>R</i>	<i>M</i>	<i>T</i>	<i>P</i>
1	-0.46	4.98	6.97	-0.74	-1.84
2	-0.20	-0.63	18.47	4.41	-0.83
3	54.33	1.44	0.25	2.69	1.33
4	-0.46	-0.63	-0.84	-0.45	-1.76
5	30.87	1.44	-0.25	2.69	0.46
6	5.35	-0.63	0.21	4.70	2.83
7	1.50	-0.63	4.56	4.70	-0.53
8	-0.40	4.98	-0.05	-0.74	1.35
9	-0.33	-0.63	12.12	2.98	-0.96
10	-0.45	3.21	18.38	-0.74	-0.37

为掌握每个聚类结果的特点以及不同类之间的关系,分别统计分析各聚类的用户数量、均值平方根(Root Mean Square, RMS)标准差、从类中心到观测的最大距离、最近的聚类和聚类质心间的距离等 5 个指标,聚类特性信息如表 5 所示。其中聚类 1 中共包含 143 个用户, RMS 的标准差为 1.04, 从该聚类中心到其余观测的最大距离为 5.63, 距离聚类 1 最近的类是聚类 7, 二者之间的距离为聚类质心间的距离, 即为 3.83。

表 5 聚类特性分析

Table 5 Analysis of clustering characteristics

聚类	用户数量	RMS 标准差	从类中心到观测的最大距离	最近的聚类	聚类质心间的距离
1	143	1.04	5.63	7	3.83
2	4	1.09	2.79	10	4.24
3	1	—	0	5	23.48
4	10 109	0.53	3.78	7	1.64
5	1	—	0	3	23.48
6	1 667	1.00	5.03	8	3.35
7	1 418	0.66	3.82	4	1.64
8	2 088	0.82	4.53	4	2.25
9	27	1.09	4.47	1	4.49
10	6	1.06	3.00	2	4.24

3.2 用户价值评价

建立用户价值评价体系有利于理解聚类结果,分析不同类型用户特征,提高聚类结果的可读性和可执行性,从而指导营销策略的制定。根据全量数据 EDA 特征分析及相关业务专家判断,设定用户特征评价指标阈值,并将用户赋予长期用户、短期用户、流失用户、非流失用户、价值型用户、非价值型用户等 6 种属性,用户特征指标评价规则如表 6 所示。

表 6 用户特征评价指标

Table 6 Evaluating indicators of customers characteristics

变量	价值评价阈值	用户特征
<i>T</i>	≥90%	长期用户
	<90%	短长期用户
<i>R</i>	≤1 个月	非流失用户
	>1 个月	流失用户
<i>Q</i> 、 <i>M</i> 、 <i>P</i>	$Q \geq 50\%$ 且 $M \geq 50\%$ 且 $P \geq 50\%$	价值型用户
	否则	非价值型用户

根据评价规则,持卡时长超过全量用户数据的 90%(约7个月)的认定为长期用户,否则为短期用户;最后一次充电行为小于 1 个月的认定为非流失用户,否则为流失用户;同时满足充电总电量、充电卡余额均值、充电效率均大于全量用户数据的 50%的认定为价值型用户,否则为非价值型用户,由此得到用户分类价值特征如表 7 所示。

表 7 用户价值分类结果

Table 7 Segmentation results of customer value

聚类	用户数量	是否长期	是否价值型	是否流失
1	143	否	否	是
2	4	是	否	否
3	1	是	是	是
4	10 109	否	否	否
5	1	是	否	是
6	1 667	是	是	否
7	1 418	是	否	否
8	2 088	否	否	是
9	27	是	否	否
10	6	否	否	是

综合用户聚类结果与用户特征评价指标,将全体电动汽车用户分为 8 类,得到用户分类的价值评价结果如表 8 所示。针对分类结果进行分析,可得到以下结论:

(1) 在全量电卡中,近 80%的电卡均为短期使用(持卡时间不超过7个月),只有约 20%的电卡为长期使用(持卡时间超过7个月),主要受到电动汽车新增用户较多的影响。

(2) 价值型用户完全分布在长期型用户中,说明电动汽车用户的价值需要一定的使用时间积累才能体现出来;而流失用户大部分分布在短期用户中。

(3) 聚类 6 约 11%的电卡持有者同时具备长期使用、价值型和非流失特征,属于优质用户,需要重点关注并维系客户关系。

(4) “非价值型非流失用户”占比最大,约为全体用户的 75%,属于有待价值提升的用户群体;在长期用户(占 9.37%)中,聚类 2 和聚类 9 的充电总电量

和充电效率偏低, 而聚类 7 仅是充电效率偏低, 更易于转化为价值型用户; 在其余的短期用户(占 65.37%)中, 聚类 4 包含 10 109 张电卡, 其充电总电量、充电效率和电卡余额均处于较低水平, 有待进一步观察。

(5) 聚类 7 的“非价值型”长期用户(占 9.17%)的充电总电量和电卡余额水平较好, 仅是充电效率略低, 只要通过合理的服务引导, 降低充电完成后的充电桩占用时间, 就可将该聚类转化为价值型用户。

(6) 近 15%的电卡属于流失用户, 主要原因是锁卡、故障或损坏等问题造成的电卡更换。

(7) 调查表明, 聚类 3 和聚类 5 各涉及的 1 张已流失电卡均为运维测试卡, 在完成测试任务后已不再使用。

表 8 用户价值评价结果

Table 8 Evaluation results of customer value

一级分类	二级分类	对应聚类	用户数量	占比
长期用户 (3 118 个)	价值型流失用户	3	1	0.01%
	价值型非流失用户	6	1 667	10.78%
	非价值型流失用户	5	1	0.01%
	非价值型非流失用户	2、7、9	1 449	9.37%
短期用户 (12 346 个)	价值型流失用户	—	0	0
	价值型非流失用户	—	0	0
	非价值型流失用户	1、8、10	2 237	14.47%
	非价值型非流失用户	4	10 109	65.37%

3.3 措施建议

结合电动汽车用户细分结果, 对充电设施运维管理、充电服务以及营销策略提出以下建议:

(1) 在运维管理水平提升方面, 由于使用过程中出现电卡故障、锁卡、损坏或遗失等现象, 造成对电卡短期持有且已流失的用户数量较多, 因此需要进一步提高电卡质量, 优化充电系统兼容性, 减少用户换卡频次, 加强用户持卡连贯性; 另一方面, 应根据充电设施故障分布加强日常巡检和故障检修, 减少设备故障停运时间, 提高充电服务能力。

(2) 在“非价值型”向“价值型”用户转化方面, 一是根据长期持卡用户“低价值”分析结果, 分别从充电电量和充电效率等方面入手, 采取应对措施, 例如通过升级充电服务系统, 实时向用户推送充电状态和充电完成消息, 提醒用户在充电完成后及时离开充电车位, 以提高充电设施使用效率; 二是关注短期持卡的“非价值”用户的充电行为趋势, 下钻分析充电量变化趋势和充电行为特征, 对潜在优质用户采取营销策略, 将其转化为“价值型”。

(3) 在精准营销策略方面, 一是持续锁定“价值型非流失”用户, 通过优先预约、积分兑换优惠券等方式维持用户忠诚度; 二是对充电电量多、充电

效率高的用户给予积分奖励、优质评级和特惠服务, 鼓励用户良好的充电行为; 三是对“低价值”用户推送充电设施分布、充电优质服务内容和积分奖励规则等宣传消息, 引导用户向“价值型”转化。

4 结语

本文基于北京地区公共充电设施的充电服务数据, 针对持卡充电的电动汽车用户开展充电行为细分研究。采用“大数据”分析思路, 在对全量充电服务数据的筛选、处理以及探索性分析的基础上, 分析得到用于细分模型输入的充电电量、充电“近度”等 5 个关键变量, 并采用 K_MEANS 聚类算法完成用户分类。针对聚类结果, 提出了电动汽车用户价值评价方法, 给出了评价指标规则及分析结果, 得到价值型非流失优质用户以及流失型用户的分布特点等结论, 为提升运维管理水平和制定精准营销策略提供数据基础和技术支撑。

参考文献

- [1] 宋永华, 阳岳希, 胡泽春. 电动汽车电池的现状与发展趋势[J]. 电网技术, 2011, 35(4): 1-7.
SONG Yonghua, YANG Yuexi, HU Zechun. Present status and development trend of batteries for electric vehicles[J]. Power System Technology, 2011, 35(4): 1-7.
- [2] 张晨曦, 文福拴, 薛禹胜, 等. 电动汽车发展的社会综合效益分析[J]. 华北电力大学学报, 2014, 41(3): 55-63.
ZHANG Chenxi, WEN Fushuan, XUE Yusheng, et al. A preliminary investigation on the overall profit assessment of electric vehicle development to the society[J]. Journal of North China Electric Power University, 2014, 41(3): 55-63.
- [3] 黄学杰. 电动汽车与锂离子电池[J]. 物理, 2015, 44(1): 1-7.
HUANG Xuejie. Electric vehicles and Li-ion batteries[J]. Physics, 2015, 44(1): 1-7.
- [4] 肖湘宁, 温剑锋, 陶顺, 等. 电动汽车充电基础设施规划中若干关键问题的研究与建议[J]. 电工技术学报, 2014, 29(8): 1-10.
XIAO Xiangning, WEN Jianfeng, TAO Shun, et al. Study and recommendations of the key issues in planning of electric vehicles' charging facilities[J]. Transactions of China Electrotechnical Society, 2014, 29(8): 1-10.
- [5] 李秋硕, 肖湘宁, 郭静, 等. 电动汽车有序充电方法研究[J]. 电网技术, 2012, 36(12): 32-38.
LI Qiushuo, XIAO Xiangning, GUO Jing, et al. Research on scheme for ordered charging of electric vehicles[J]. Power System Technology, 2012, 36(12): 32-38.
- [6] 曾鸣, 薛松, 刘宏志, 等. 我国区域电动汽车运营模式及其最优并网规模规划模型[J]. 电网技术, 2012, 36(6): 175-181.

- ZENG Ming, XUE Song, LIU Hongzhi, et al. Research on regional vehicle-to-grid operation mode and optimal share planning model integrating it into electric power system[J]. Power System Technology, 2012, 36(6): 175-181.
- [7] 夏露, 刘畅, 李斌, 等. 面向私人电动汽车的城市公共充电网络运营服务能力评估方法与仿真研究[J]. 电网技术, 2015, 39(12): 3544-3548.
XIA Lu, LIU Chang, LI Bin, et al. Evaluation methodology and simulation of service capability of urban public charging network for private electric vehicles[J]. Power System Technology, 2015, 39(12): 3544-3548.
- [8] 黄顺亮, 王琦. 基于三支决策理论的客户细分方法[J]. 计算机应用, 2014, 34(1): 244-248.
HUANG Shunliang, WANG Qi. Method for customer segmentation based on three-way decisions theory[J]. Journal of Computer Applications, 2014, 34(1): 244-248.
- [9] 刘朝华, 梅强, 蔡淑琴. 基于 RFM 的客户分类及价值评价模型[J]. 技术经济与管理研究, 2012, 1(5): 33-36.
LIU Chaohua, MEI Qiang, CAI Shuqin. The model of customer classification and evaluation of customer value based on RFM[J]. Technoeconomics & Management Research, 2012, 1(5): 33-36.
- [10] HU Yahan, HUANG T C K, KAO Yuhua. Knowledge discovery of weighted RFM sequential patterns from customer sequence databases[J]. Journal of Systems and Software, 2013, 86(3): 779-788.
- [11] 徐翔斌, 王佳强, 涂欢, 等. 基于改进 RFM 模型的电子商务客户细分[J]. 计算机应用, 2012, 32(5): 1439-1442.
XU Xiangbin, WANG Jiaqiang, TU Huan, et al. Customer classification of E-commerce based on improved RFM model[J]. Journal of Computer Applications, 2012, 32(5): 1439-1442.
- [12] 曾小青, 徐秦, 张丹, 等. 基于消费数据挖掘的多指标客户细分新方法[J]. 计算机应用研究, 2013, 30(10): 2944-2947.
ZENG Xiaoqing, XU Qin, ZHANG Dan, et al. New multi-indicator customer segmentation method based on consuming data mining[J]. Application Research of Computers, 2013, 30(10): 2944-2947.
- [13] 林勤, 薛云. 双聚类算法在电信高价值客户细分的应用[J]. 计算机应用, 2014, 34(6): 1807-1811.
LIN Qin, XUE Yun. Application of biclustering algorithm in high-value telecommunication customer segmentation[J]. Journal of Computer Applications, 2014, 34(6): 1807-1811.
- [14] 刘芝怡, 陈功. 基于改进 K-means 算法的 RFAT 客户细分研究[J]. 南京理工大学学报, 2014, 38(4): 531-536.
LIU Zhiyi, CHEN Gong. RFAT customer segmentation based on improved K-means algorithm[J]. Journal of Nanjing University of Science and Technology, 2014, 38(4): 531-536.
- [15] 周颖, 吕巍, 井森. 基于数据挖掘技术的移动通信行业客户细分[J]. 上海交通大学学报, 2007, 41(7): 1142-1145.
ZHOU Ying, LÜ Wei, JING Miao. The research of customer segmentation based on data mining in mobile telecommunication industry[J]. Journal of Shanghai Jiao Tong University, 2007, 41(7): 1142-1145.
- [16] 傅军栋, 杨姚, 罗善江. 智能小区居民用电负荷特征权重分析[J]. 电力系统保护与控制, 2016, 44(18): 41-45.
FU Jundong, YANG Yao, LUO Shanjiang. Residential electricity load features weighting analysis in smart community[J]. Power System Protection and Control, 2016, 44(18): 41-45.
- [17] 陈明照, 毛坚, 杜宗林, 等. 基于聚类法的工业用户需求侧管理(DSM)方案分析与研究[J]. 电力系统保护与控制, 2017, 45(7): 84-89.
CHEN Mingzhao, MAO Jian, DU Zonglin, et al. Analysis on demand side management scheme of industrial enterprise based on clustering method[J]. Power System Protection and Control, 2017, 45(7): 84-89.
- [18] 孟建良, 刘德超. 一种基于 Spark 和聚类分析的辨识电力系统不良数据新方法[J]. 电力系统保护与控制, 2016, 44(3): 85-91.
MENG Jianliang, LIU Dechao. A new method for identifying bad data of power system based on Spark and clustering analysis[J]. Power System Protection and Control, 2016, 44(3): 85-91.
- [19] ZHANG Lu, CHEN Yanxia, ZHU Jie, et al. Data quality analysis and improved strategy research on operations management system for electric vehicles[C] // 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, 2015, Changsha, China: 2715-2720.
- [20] 王扶东, 马玉芳. 基于数据挖掘的客户细分方法的研究[J]. 计算机工程与应用, 2011, 47(4): 215-218.
WANG Fudong, MA Yufang. Research of method for customer segment based on data mining[J]. Computer Engineering and Applications, 2011, 47(4): 215-218.

收稿日期: 2017-11-01; 修回日期: 2018-01-12

作者简介:

张 禄(1984—), 男, 博士, 高级工程师, 主要研究方向为电动汽车数据分析与计算; E-mail: zhanglu7@163.com

李国昌(1977—), 男, 高级工程师, 主要研究方向为电力系统及其自动化控制技术;

陈艳霞(1974—), 女, 博士, 高级工程师, 主要研究方向为智能配电网信息与通信技术。

(编辑 张爱琴)