

Electricity Customer Segmentation Study Based on Payment Behavior

Zhonglian Du¹, Hongwei Xie², Xin Fan², Yan Shi², Huaqin Qin³, Chonghui Guo^{1*}

¹Faculty of Management and Economics, Dalian University of Technology, Dalian Liaoning

²State Grid Inner Mongolia East Power Co., Ltd., Chifeng Inner Mongolia

³Beijing Kedong Power Control System Co., Ltd., State Grid Electric Power Research Institute, Beijing

Email: *dlutguo@dlut.edu.cn

Received: May 30th, 2018; accepted: Jun. 14th, 2018; published: Jun. 21st, 2018

Abstract

It is effective to improve electricity companies' efficiency of collecting electricity charge by subdividing the payment behavior of electricity customers and adopting the improvement measures targeted. Based on the traditional RFM model, this paper proposes an index system and customer segmentation method which is suitable for the power customer system. Firstly, the WFM index system based on the electricity customer's payment behavior is established. Then, K-means clustering method is used to obtain the optimal number of cluster with the best clustering goodness. The SOM neural network is used to cluster the customers, and the attributes of each group are compared with the overall attributes. So, the customer segmentation results are obtained. Finally, this method is applied to the detail data of electricity customer in Ningcheng County, Chifeng City, Inner Mongolia. Four different types of customer groups are extracted, and the corresponding marketing strategies of payment channels are put forward respectively. So, the feasibility and validity of this method are verified.

Keywords

Customer Segmentation, Payment Behavior, RFM Model, SOM Neural Network

基于缴费行为的用电客户细分研究

杜忠莲¹, 谢宏伟², 樊新², 石研², 覃华勤³, 郭崇慧^{1*}

¹大连理工大学管理与经济学部, 辽宁 大连

²国网内蒙古东部电力有限公司, 内蒙古 赤峰

³国网电力科学研究院, 北京科东电力控制系统有限责任公司, 北京

Email: *dlutguo@dlut.edu.cn

收稿日期: 2018年5月30日; 录用日期: 2018年6月14日; 发布日期: 2018年6月21日

*通讯作者。

摘要

对用电客户缴费行为进行有针对性的细分研究,可有效提升电力公司电费收缴工作效率。研究用电客户缴费行为特点,在传统RFM模型的基础上,提出适用于电力客户缴费行为的指标体系和客户细分方法。首先,建立了基于用电客户缴费行为细分的WFM指标体系;然后,通过聚类优度分析获取最优聚类簇个数,采用SOM神经网络对用电客户进行聚类,对各类群体的属性与总体属性进行比较分析,得到客户细分结果;最后,将该方法应用于内蒙古赤峰市宁城县用电客户缴费明细数据中,提取出四类不同特征的客户群体,并分别提出相应的缴费渠道营销策略,验证了该方法的可行性和有效性。

关键词

客户细分, 缴费行为, RFM模型, SOM神经网络

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着经济的快速发展,客户用电服务需求呈现出多样化的趋势。为满足用电客户对服务多样化的需求,供电企业不断引入新型的缴费方式,积极引导用电客户采用新型缴费渠道缴费[1]。受内外形势所迫,电网企业充分认识到转型迫在眉睫,正通过在营销环节打造面向市场的“入口+平台”,扩大核心资源能力,拓展产业布局[2]。近年来,客户关系管理受到越来越广泛的重视。任何高效的客户关系管理都将以扎实的客户细分为基础,其理论依据在于顾客需求的异质性和企业需要在有限资源的基础上进行有效市场竞争[3]。赵铭等利用K-均值聚类方法,对商业银行基金理财客户进行研究,并依据分类结果制定出相应的个性化营销方案[4]。褚格林研究了移动数据业务发展背景下电信客户的细分,并针对不同客户群体初步讨论了相应的市场策略[5]。郭崇慧和赵作为应用SOM神经网络对汽车4S店客户进行聚类,提出客户群在群数量及群属性上随时间变化的分析方法和客户个体的分群演变分析方法[6]。蔡淑琴等利用模糊C均值聚类方法,基于在线客户评论对客户进行细分研究[7]。

客户细分根据细分维度的不同,也对应有不同的细分方法,采用传统的人口统计指标来分析客户,该方法操作简便,但是只适合用于了解客户结构,难以进行客户行为分析。近些年客户行为指标如近度、频度和值度,即RFM模型在客户行为分析和客户细分中得到了广泛运用[8][9][10],但是关于用电客户细分的研究较少。宋才华等利用从电网公司用电客户细分现状与应用需求出发,提出三种客户细分模型,并构建了三步客户细分机制[11]。赵莉等建立了家庭用电信息的数据维度模型,利用K-均值算法对家庭用户进行聚类[12]。胡晓雪等采用当月用电量和当前欠费总额等当前价值指标、电量增长贡献率和历史同期电费增长率等潜在价值指标和地域分布指标,构建电力客户细分指标体系[13]。刘晓焜提出外在属性、内在属性和消费行为三种电力客户细分方式[14]。上述研究中,对缴费方式的描述多为传统的描述统计分析,指标较为单一,数据量少,结果的可解释性不强。

本文采用真实的用电客户缴费行为数据进行全面的指标体系构建,在传统RFM模型的基础上,提出适用于用电客户细分的WFM指标体系,更加真实全面。在方法上,由于真实的用电客户缴费行为数据

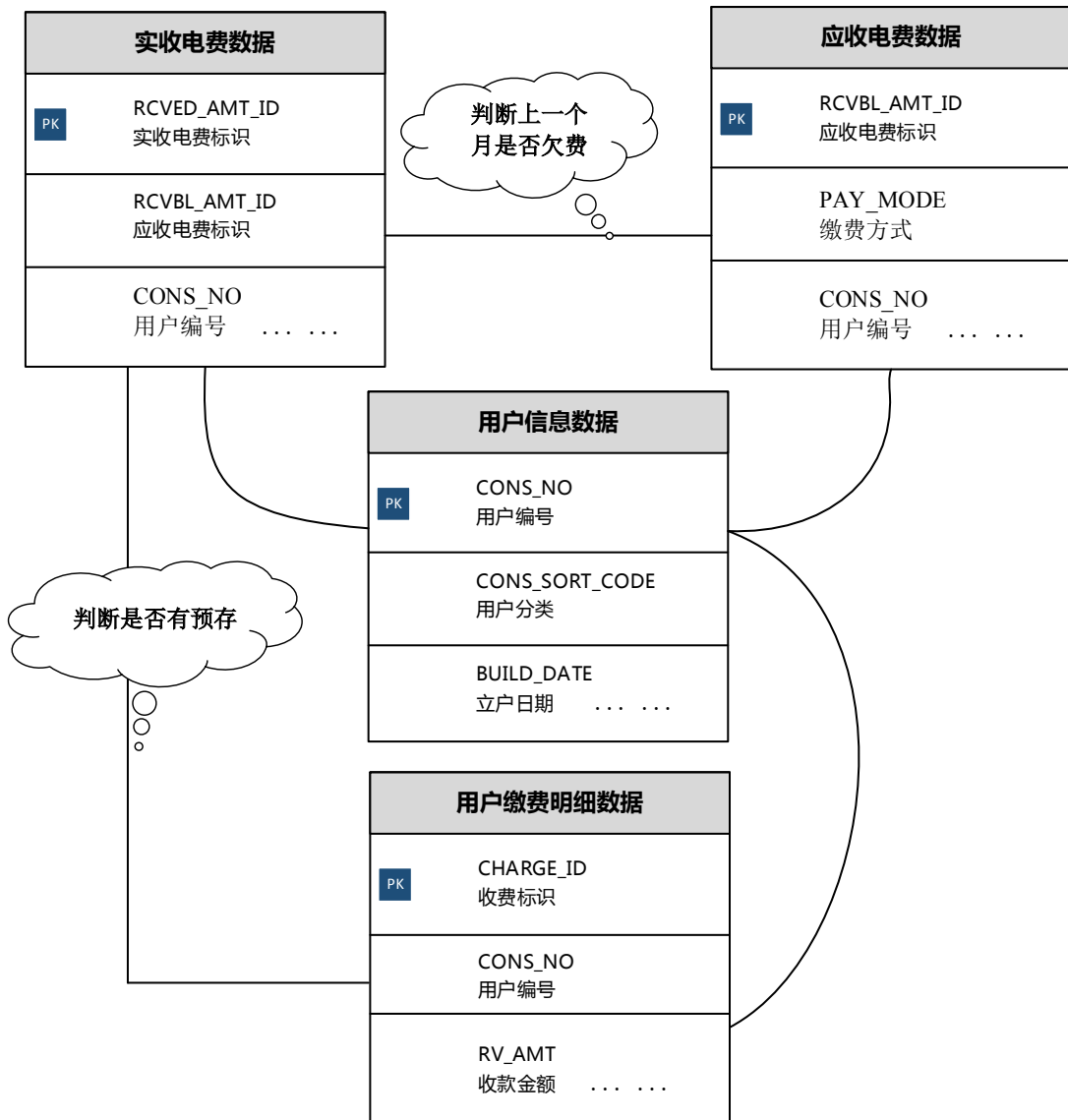


Figure 2. Relationship between data tables
图 2. 数据表间关系

Table 1. Comparison between RFM model and WFM model
表 1. RFM 模型与 WFM 模型对比

	指标 1				指标 2		指标 3
	R (Recency)				F (Frequency)		M (Monetary)
RFM 模型	客户最近一次购买距离分析点的时间				客户最后一次交费距离分析点的时间		客户一定时期内购买该企业产品的总金额
	W (Way)				F (Frequency)		M (Monetary)
	W ₁	W ₂	W ₃	W ₄	AF	PF	M
WFM 模型	电力机构坐收	电力机构自助缴费终端	金融机构代收	非金融机构代收	客户实际缴费次数	客户预存电费次数	客户一定时期内缴纳电费总额

力机构自助缴费终端使用频次,该指标在一定程度上反映了客户相对前卫的缴费习惯。 W_3 指标为金融机构代收使用频次,该指标在一定程度上反映了客户相对较前卫的缴费方式,其缴费信用较好,欠费风险小。 W_4 指标为非金融机构代收使用频次,该指标在一定程度上反映了客户与时俱进的缴费理念和方式。

2) F 指标分为实际缴费次数和预存缴费次数。涂莹等将用电客户缴费时间纳入电费回收风险模型中进行研究,并证明了其相关关系[18]。本研究中 AF 指标为实际缴费次数,虽然电费每月都会有收缴,但是客户的缴费频率却不是每月一次,该指标在一定程度上能够反映客户的缴费行为和缴费积极性。PF 指标为预存缴费次数,具体指客户实际缴纳金额大于当月应收金额,该指标在一定程度上能够反映客户的缴费行为和缴费信用度。

3) M 指标为缴费金额,本文所选客户均为低压居民,电价基本趋于稳定,所以电费的高低可以在一定程度上反映客户的用电稳定性。

3. 基于 SOM 神经网络的用电客户聚类

3.1. 聚类优度分析

考虑用电客户缴费行为数据所采用的数据量纲不同,且差异较大。因此,在对数据聚类之前首先进行具体指标值的归一化,本文所选归一化方法为最大最小值法,即:

$$X^* = \frac{x - \min}{\max - \min} \quad (1)$$

其中, x 为待归一化指标数据, \min 为指标数据的最小值, \max 为指标数据的最大值。 X^* 为归一化后所得值。K-均值聚类的最终目标是最小化每一个类内部的差异,最大化类之间的差异。设聚类结果组内平方和为 within_ss , 组间平方和为 between_ss 。聚类结果组内差距越小,组间差距越大,则表明聚类结果越好。采用聚类优度指标 P 来测量聚类效果,即:

$$P = \frac{\text{between_ss}}{\text{within_ss}} \quad (2)$$

P 越大代表聚类优度越高,即聚类结果越好。根据实际情况,选取相对较高的聚类优度 P^* 所对应的聚类类别数 K^* , 对实际数据进行分析。

3.2. SOM 神经网络聚类算法

SOM 神经网络的网络结构比较简单,由输入层和竞争层构成,输入层有 N 个神经元;竞争层有 M 个神经元。其中 N 为指标体系中所选指标数, M 为聚类类别数,即 $M = K^*$ 。网络的连接权重为 $\{w_{ij}\}, i=1,2,\dots,N; j=1,2,\dots,M$; 约束条件为: $\sum_{i=1}^N w_{ij} = 1$ 。网络的 P 个二值(0、1)输入学习模式为:

$A_k = (a_1^k, a_2^k, a_3^k, \dots, a_N^k)$; 与其对应的竞争层输出模式为: $B_k = (b_1^k, b_2^k, b_3^k, \dots, b_N^k)$, 其中 $k=1,2,\dots,P$ 。

Step 1: 初始化,按约束条件 $\sum_{i=1}^N w_{ij} = 1$, 赋予 w_{ij} $[0, 1]$ 内的随机值;

Step 2: 任选 P 个模式中的一个模式 A_k 提供给网络的输入层;

Step 3: 计算竞争层各神经元的输入值 S_j , $S_j = \sum_{i=1}^N w_{ij} a_i, i=1,2,\dots,M$;

Step 4: 按“优胜劣汰”的原则以 $S_j (j=1,2,\dots,M)$ 中最大值所对应的神经元作为胜者,将其输出状态为 1, 而其它所有神经元的输出状态置为 0, 即: $b_j = 1, S_j > S_i, (i=1,2,\dots,M, i \neq j)$; $b_i = 0, (i \neq j)$; 若出

现 $S_j = S_i$ 的现象, 则按统一约定取左边的神经元为获胜神经元;

Step 5: 修正与获胜神经元相连的各连接权重, 而其它所有连接权重保持不变 $w_{ij} = w_{ij} + \Delta w_{ij}$

$\Delta w_{ij} = \eta \left(\frac{a_i^k}{m} - w_{ij} \right)$ 其中, $i = 1, 2, \dots, N, 0 < \eta < 1$, η 为学习系数, 一般为 0.01~0.03, m 为第 k 个学习模式

向量 $A_k = (a_1^k, a_2^k, a_3^k, \dots, a_N^k)$ 中元素为 1 的个数;

Step 6: 选取另一个学习模式, 回到 Step 3, 直到 P 个学习模式全部提供给网络;

Step 7: 返回 Step 2, 直到各连接权的调整量变得很小为止。

4. 应用实例

4.1. 数据预处理与模型参数设定

应用实例数据来自国网蒙东赤峰市供电公司, 从宁城县用电客户两年半(2014年1月1日至2016年6月30日)时间内的缴费记录中, 删除噪声数据, 即删除相同 CHARGE_ID 错误数据; 此外, 为确保实验用户用电的连续性, 数据预处理时将新立户和注册销户的用电客户数据剔除。基于应用实例数据周期是 30 个月, 在原始数据异常值剔除时, 限制实际缴费频次 AF 在 10 到 60 次之间, 缴费金额 M 在 450 到 5000 元之间, 噪声数据共计 16,677 条。最终获取唯一 CHARGE_ID 用电客户连续缴费明细数据 69,305 条。所有用电客户属性均为低压用电居民, 原始数据基本描述性统计如表 2 所示。

由于本文所采用的数据量纲不同, 且差异较大, 例如缴费数据最大为 4997 元, 缴费次数只有几十次。因此, 在对数据聚类之前首先对原始数据 C_0 进行归一化得到标准化数据 C_1 。为获取聚类所需预设聚类类别数, 首先将数据使用 K-均值方法进行聚类, 并对聚类结果进行聚类优度测试, 选择相对合理的聚类簇个数。将 K 取值设为从 1 到 20 进行聚类实验, 测试聚类优度如图 3 所示。在聚类类别 K = 10 类时已经基本趋于稳定, 在此基础上增加聚类类别不能得到更好的聚类效果。当聚类类别数 K = 10 类时聚类优度为 $P = 0.976$, 再增加一类聚类类别, 聚类优度的提升值为仅为 0.1%, 考虑到本文所采用的实验数据较多, 97.6% 的聚类优度可以接受, 故选定最优聚类类别数为 10 类, 即 $K^* = 10$ 。

SOM 神经网络的输入层有 N 个神经元; 竞争层有 M 个神经元。本文中 $N = 7$, 分别为 WFM 模型中的七个细分指标, 实验数据聚类类别数 $M = K^* = 10$, 即将用电客户聚为 10 类。

4.2. 计算结果分析

SOM 神经网络聚类分析共得到 10 类不同属性的客户群, 每类客户群具体的用电客户数量如图 4 所示。其中, 第三类(cluster 3)和第五类(cluster 5)的数量较多, 均达到 11,000 个用户以上; 第九类相对较少, 只有 3268 个用户。可以直观看出, SOM 神经网络聚类的十个客户群, 在数量上存在一定的差异。为分析每个客户群具体指标的差异, 采用比较每一个聚类群体细分指标均值和总体细分指标均值的变动情况来分析, 具体数据分析如表 3 所示。

Table 2. The distribution of WFM model indicators in application data

表 2. 应用实例数据中 WFM 模型指标分布情况

	W_1 (次)	W_2 (次)	W_3 (次)	W_4 (次)	AF(次)	PF(次)	M(元)
Min	6	0	0	0	11	0	451
Mean	37.350	0.003	1.493	0.061	38	5.545	1172
Max	59	11	29	21	59	22	4997

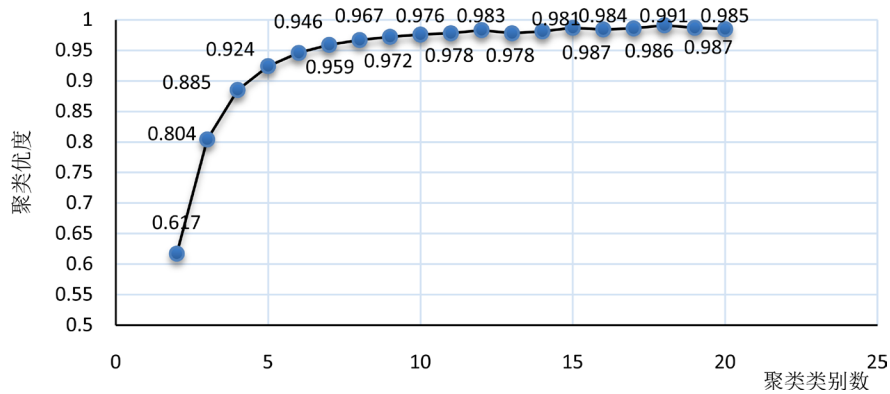


Figure 3. Selection of optimal number of clustering
图 3. 选择最优聚类类别数

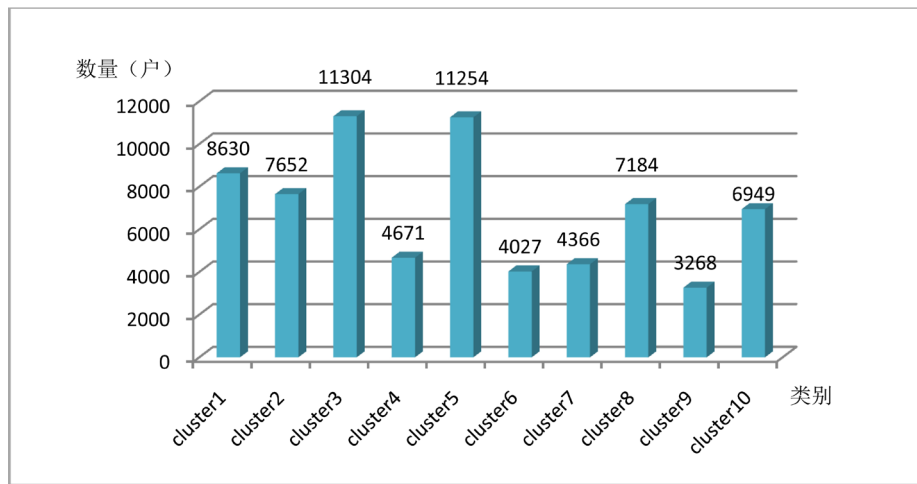


Figure 4. Clustering results of SOM neural network
图 4. SOM 神经网络聚类结果

Table 3. Index analysis of SOM neural network clustering results
表 3. SOM 神经网络聚类结果指标分析

	AF(次)	PF(次)	M(元)	W1(次)	W2(次)	W3(次)	W4(次)
总体均值	38	5.545	1172	37.35	0.003	1.493	0.061
cluster1	36.81↓	5.596↑	848.1↓	36.54↓	0↓	0.263↓	0.01↓
cluster2	42.07↑	6.978↑	1411↑	33.98↓	0.009↑	7.842↑	0.231↑
cluster3	28.26↓	2.984↓	789.1↓	27.1↓	0.014↑	1.045↓	0.089↑
cluster4	37.02↓	3.672↓	864.2↓	33.36↓	0↓	3.649↑	0.007↓
cluster5	41.94↑	5.73↑	1659↑	41.47↑	0.0002↓	0.4304↓	0.035↓
cluster6	35.25↓	1.788↓	674.5↓	35.1↓	0↓	0.1562↓	0↓
cluster7	50.04↑	12.44↑	2400↑	49.26↑	0.0014↑	0.5726↓	0.1993↑
cluster8	39.31↑	2.428↓	935.1↓	39.09↑	0↓	0.2137↓	0.0092↓
cluster9	41.14↑	6.081↑	933.9↓	40.93↑	0↓	0.2044↓	0↓
cluster10	45.42↑	9.845↑	1226↑	45.1↑	0↓	0.3094↓	0.0033↓

结合电力客户缴费行为具体分析, 可以将 SOM 神经网络聚类的 10 类结果总结为四大类群体, 即前卫高信用型客户、前卫一般客户、传统高信用型客户、传统一般客户, 即前卫类型客户比传统型客户采用电力机构自动终端缴费或非金融机构代收等新兴缴费渠道的频次更高, 信用型客户比一般客户的缴费频次和电费预存频次更高。

由表 3 中可以看出, 第二类(cluster2)和第七类(cluster7)客户缴费频次都非常高, 用电量较大, 而且预存电费的次数也高于平均水平, 特别是第七类客户预存电费次数超出平均水平的两倍, 说明此类客户信用度极高。电力机构自助终端缴费和非金融机构代收两种缴费方式均超过平均水平, 迎合了现在互联网时代发展的趋势, 将此类客户统一划分为前卫高信用型客户。

第三类(cluster3)客户从缴费频次和预存电费频次来看都显著低于平均水平, 用电量较少, 但是缴费方式较新颖, 电力机构自助缴费终端的比例在所有类别中属于最高, 传统的电力机构坐收方式不常使用, 将此类客户统一划分为前卫一般客户。

第五类(cluster5)、第九类(cluster9)和第十类(cluster10)客户他们的缴费频次和电费预存频次都非常高。在缴费方式上, 电力机构自助缴费终端等新上线缴费方式都低于平均水平, 但是传统的电力机构坐收使用率却非常高。说明此类客户在缴费方式的选择方面还是比较保守, 但是缴费积极性很高, 将该类客户统一划分为传统高信用客户。

最后, 第一类(cluster1)、第四类(cluster4)、第六类(cluster6)和第八类(cluster8)客户缴费次数和预存电费次数都普遍低于平均水平。各种新上线缴费方式使用不积极, 特别是电力机构自助缴费终端低于平均数水平。将次大类客户统一划分为传统一般型客户。

将上述四大类客户群体数据合并统计分析, 如图 5 所示。从图中可以看出, 该数据采集地区, 传统型客户占多数, 传统高信用型客户和传统一般客户型共占比 67%, 前卫高信用型客户与前卫一般型客户占比 16%, 说明该地区的在缴费方式的信息化推进方面还处于比较不发达阶段, 大部分客户采用传统的缴费方式缴费, 这与内蒙古赤峰市宁城县的经济发展现状相吻合, 再次验证了计算结果的可靠性。

4.3. 营销策略建议

结合电力公司管理实际背景以及电力客户缴费行为数据挖掘结果, 针对传统一般型客户、传统高信用型客户、前卫一般型客户和前卫高信用型客户四大类典型客户群体, 给出如下营销策略建议:

1) 对于传统一般型客户, 电力公司应该采取积极主动的方式尽力争取, 让这类群体意识到各类互联网缴费方式的便利性, 同时, 可以推出预存电费奖励机制, 鼓励客户及时缴纳电费和预存电费, 提高客

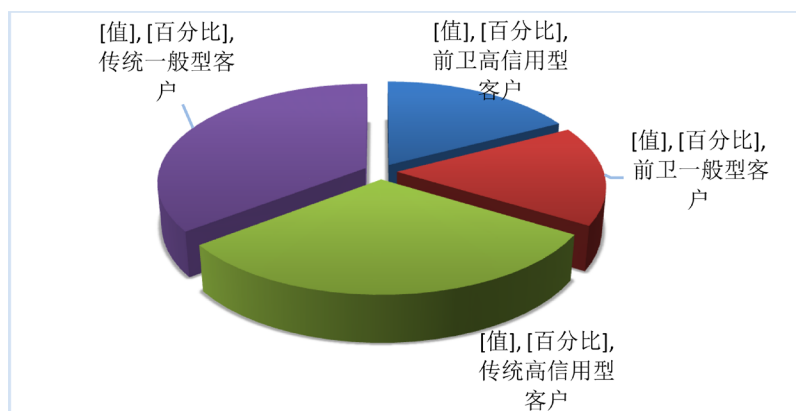


Figure 5. Final classification results

图 5. 最终分类结果

户信用度。

2) 对于传统高信用型客户, 电力公司应该抓住机会积极推广自助缴费终端、手机客户端缴费包括支付宝、电力 E 行、移动“和生活”、电信“翼支付”付等方式, 让这部分客户多接触全新的缴费模式。同时, 基于此类客户的历史缴费信用较好, 可以给这类客户一定的欠费额度。

3) 对于前卫一般型客户, 他们对于新兴事物的接受能力较强, 电力公司在推出新的缴费方式时可以优先通过各种宣传方式通知他们试用。同时, 也要告诉给他们一定的缴费和预存缴费的激励措施, 确保电费收缴的及时性。

4) 对于前卫高信用型客户, 从分类结果来看, 此类客户不仅能够快速接受新兴缴费方式而且信用度较高, 预存电费的频率高, 对传统电力机构坐收资源的占用量少。基于此, 电力公司可以采取保持其积极性的措施, 如新客户端试用免单、预存电费年终积分兑换等形式, 维持该类客户的积极性和信用度。

5. 结束语

本文依据电力客户缴费行为数据分析与挖掘的应用研究背景, 在理解传统 RFM 模型的基础上, 建立了一套适用于电力客户细分的指标体系, 该指标体系由传统 RFM 模型的三个指标扩展为 WFM 模型七个指标。利用 SOM 神经网络聚类算法对客户进行聚类分析, 通过聚类优度分析获取最优聚类簇个数。依据指标选择将聚类结果划分为十类, 接着将所得初步聚类结果属性与总体属性进行对比分析, 依据聚类群体各指标平均值相对于总体指标平均值的上升或下降情况, 对初步聚类结果进行再次的合并, 得到四类用电客户细分结果。分析四类用电客户群体的不同特征, 为电力公司提出相应的缴费渠道营销策略, 进一步提升电力公司电费收缴工作的效率。

本文所用数据均来自用电客户实际缴费记录, 数据量较大, 时间周期较长, 分析结果可信度高。由聚类结果缴费方式指标可以得出, 该地区用电客户中传统型用户约占 67%, 说明该地区的在缴费方式的信息化推进方面还处于比较不发达阶段, 大部分用电客户采用传统的缴费方式缴费, 这与内蒙古赤峰市宁城县的经济发展现状相吻合, 也验证了计算结果的合理性和可靠性。在接下来的研究中将进一步细化用电客户聚类特征选取, 例如 W 指标中 W₄ 非金融机构代收, 包含支付宝、微信等新兴缴费方式, 进行更加具体的细分研究, 准确定位客户信用和缴费渠道的使用。同时, 进一步完善用户信用评价指标, 构建用电客户信用动态评价体系。为电力公司提供更加精准的电费收缴建议, 助力新兴用电客户缴费渠道的建设与推广。

基金项目

国家自然科学基金资助项目: “动态数据挖掘中的演化聚类模型与算法研究”(71171030)。

参考文献

- [1] 田鹤. 电力多渠道缴费理论及缴费模式优化策略分析[J]. 科技创新导报, 2012(35): 41-43.
- [2] 王珣. 电力营销组织管理创新方法探索[J]. 电力需求侧管理, 2016, 18(5): 38-40.
- [3] Antreas, D. (2000) Athanassopoulos. Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behavior. *Journal of Business Research*, 47, 191-207. [https://doi.org/10.1016/S0148-2963\(98\)00060-5](https://doi.org/10.1016/S0148-2963(98)00060-5)
- [4] 赵铭, 李雪, 李秀婷, 等. 基于聚类分析的商业银行基金客户的分类研究[J]. 管理评论, 2013, 25(7): 38-44.
- [5] 褚格林. 基于聚类模型的电信客户细分研究[J]. 统计与决策, 2014(8): 176-179.
- [6] 郭崇慧, 赵作为. 基于客户行为的 4S 店客户细分及其变化挖掘[J]. 管理工程学报, 2015, 29(4): 18-26.
- [7] 蔡淑琴, 蒋士淼, Olle, G.D., 等. 基于在线客户评论的客户细分研究[J]. 管理学报, 2015, 12(7): 1059-1063.
- [8] 胡桂莉, 喻昕, 黄秋婵. 基于 SOM 神经网络的移动客户细分研究[J]. 微型机与应用, 2015, 34(23): 51-54.

- [9] 帅斌, 邓绍蔚, 黄丽霞. 基于改进 RFM 模型的铁路快捷货运客户市场细分方法[J]. 铁道科学与工程学报, 2014, 11(1): 112-117.
- [10] 黄聪, 王东. 基于 RFM 分析模式与马尔可夫链的客户行为预测模型研究[J]. 情报杂志, 2009, 28(S2): 143-146.
- [11] 宋才华, 蓝源娟, 范婷, 等. 电力企业客户细分模型研究[J]. 现代电子技术, 2014(2): 91-94.
- [12] 赵莉, 候兴哲, 胡君, 等. 基于改进 K-均值算法的海量智能用电数据分析[J]. 电网技术, 2014, 38(10): 2715-2720.
- [13] 胡晓雪, 赵嵩正, 吴楠, 等. 基于 SOM-DB-PAM 混合聚类算法的电力客户细分[J]. 计算机工程, 2015, 41(10): 295-301.
- [14] 刘晓焜. 基于电力缴费行为细分的服务提升研究[J]. 中国科技信息, 2014(23): 185-186.
- [15] Kohonen, T. (1990) The Self-Organizing Map. *Proceedings of the IEEE*, **78**, 1464-1480. <https://doi.org/10.1109/5.58325>
- [16] Vesanto, J. and Alhoniemi, E. (2000) Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, **11**, 586-600. <https://doi.org/10.1109/72.846731>
- [17] George, C. (2013) Hadjinicola. Product Positioning Using a Self-Organizing Map and the Rings of Influence. *Decision Sciences*, **44**, 431-461. <https://doi.org/10.1111/dec.12020>
- [18] 涂莹, 林士勇, 欧阳柳, 等. 基于市场细分的逻辑回归模型在电费回收风险预测中的应用研究[J]. 电力需求侧管理, 2016, 18(4): 46-49.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8763, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sg@hanspub.org