

基于改进熵值法的客户停电敏感度识别方法研究

曲 艺,李艳艳

(国家电网有限公司 客户服务中心,南京 210000)

Research on customer outage sensitivity identification method based on improved entropy method

QU Yi, LI Yanyan

(Customer Service Center, State Grid Co., Ltd., Nanjing 210000, China)

摘要:面对售电市场的放开,为提升电网企业的竞争力,客户需求得到了高度重视。停电问题是影响客户感知的首要因素。对客户停电敏感度的识别,是对客户开展精准服务的重要条件之一。提高客户停电敏感度的识别精度,是提高客户服务水平的关键。提出一种提高停电敏感度模型精度的方法,基于客户信息、行为等多维度指标,在传统熵值法基础上,利用自适应线性神经元算法修正权重,将无监督学习转化为半监督学习,提升权重赋值科学性。

关键词:熵值法;停电敏感;自适应线性神经元;梯度下降;半监督学习;综合评估

Abstract: Facing the opening of electric power market, customer demand has been highly valued in order to enhance the competitiveness of power grid enterprises. Power outage is the primary factor affecting customer perception. The identification of customer outage sensitivity is one of the important conditions for accurate customer service. Improving the identification accuracy of customer outage sensitivity is the key to improving customer service level. A method to improve the accuracy of outage sensitivity model is provided. Based on multi-dimensional indicators such as customer information and behavior, and on the basis of traditional entropy method, adaline algorithm is used to modify the weights, and unsupervised learning is transformed into semi-supervised learning, so as to enhance the scientificity of weight assignment.

Key words: entropy; power outage sensitivity; adaline; gradient descent; semi-supervised; comprehensive assessment

中图分类号:TM715;F407.61 文献标志码:B

0 引言

随着电力市场的逐步放开,电网公司面临着盘活存量市场、开拓增量市场的发展形势,以客户为中心,充分了解客户需求,提供更加专业的服务来提升客户感知,是未来供电服务的重中之重^[1]。在客户各类用电感知中,供电可靠性是影响客户感知的重要因素之一,对发生停电后的敏感度则是客户对供电可靠性感知的重要表现。不同客户类群对发生同一事件的反应有所不同,因此,如何精准分析客户感知偏好,精准定位停电敏感客户群,从而实现有针对性制定服务策略,降低停电对客户的影响,提升服务品质,具有重要意义^[2]。全网集中的客服中心,具有统一的客户服务入口,统一的客户服务标准,积累了海量客户基础信息数据及客户服务轨迹信息,利用大数据技术,可以对客户停电敏感度进行识别和预测,实现客户分群。本文在原有停电敏感度模型的

基础上,利用人工标记结果,通过利用自适应神经元算法对原有模型进行改进修正,有效提升模型识别的精准度,实现停电敏感客户的精准分群。

1 传统熵值法的局限性

1.1 敏感度综合评估方法

本文在客户停电敏感度识别方面,基于用电情况、行为轨迹等5个维度20个指标子项开展综合评估,评估的综合得分为

$$Z = \sum_{i=1}^n w_i x_i \quad (1)$$

式中: Z 为最终得分; w_i 为第 i 项指标权重; x_i 为第 i 项指标得分。

根据综合得分划分停电敏感度,其中定义综合评分在 $[0,30]$ 为低敏感客户, $[30,70]$ 为中敏感客户, $(70,100]$ 为高敏感客户。

1.2 传统熵值法

在综合评估中,指标权重有着举足轻重的地位。指标权重确定的科学、合理,预测结果的可靠性与准确性就高。客观赋权法依据指标数据信息来计算权重,熵值法是常用的客观赋权法,其根据指标相

收稿日期:2019-03-06;修回日期:2019-05-03

基金项目:国网客服中心2018年科技项目(62993117003C)

This work is supported by Science and Technology Project of Customer Service Center of State Grid of China in 2018 (No.62993117003C)

对离散程度对系统整体的影响来决定权重,离散度越大的指标具有较大的权重。

若用 P_j 表示第 j 个信息的不确定度(即出现的概率),则整个信息的不确定度为

$$S = -K \sum_{j=1}^n P_j \ln(P_j) \quad (2)$$

式中: K 为正常数,当各个信息发生概率相等时,即 $P_j = 1/n$, S 取值最大,此时熵最大。熵值法确定权重的具体步骤如下。

首先,确定各属性对各方案的贡献度。

设有多属性决策矩阵 M ,其结构如下

$$M = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1n} \\ x_{21}, x_{22}, \dots, x_{2n} \\ \vdots \\ x_{m1}, x_{m2}, \dots, x_{mn} \end{bmatrix} \quad (3)$$

则第 j 个属性下第 i 个方案 A_i 的贡献度 P_{ij} 可表示为

$$P_{i,j} = \frac{x_{i,j}}{\sum_{i=1}^m x_{i,j}} \quad (4)$$

其次,计算所有方案对属性 X_j 的贡献总量为 E_j

$$E_j = -K \sum_{i=1}^m P_{i,j} \ln(P_{i,j}) \quad (5)$$

式中: $K = 1/\ln(m)$,使得 E_j 在 $[0,1]$ 范围内。

可以看出,属性值由所有方案差异大小来决定权重系数的大小。

第三,计算权重。定义 d_j 为第 j 个属性下的各方案一致性程度,则 $d_j = 1 - E_j$,则各属性权重为

$$w_j = \frac{d_j}{\sum_{j=1}^n d_j} \quad (6)$$

1.3 传统熵值法的局限性

传统熵值法对指标个数无限制,但在应用中会出现某一指标离散度较大时,该指标对应权重也过大,导致其他重要指标被弱化的情况^[3-4]。文献[5]采加入人工干预,采用主客观赋权组合的方法来减少客观赋权的波动性。而自适应线性神经元算法能够根据已有样本的预测误差,反复修正权重,使得结果更可靠^[6-7]。

2 改进熵值法原理

改进熵值法的原理为:通过熵值法求指标初始权重 w_{0j} ($j = 1, 2, \dots, L, k$),将权重作为自适应线性神经元算法的初始权重,对已知数据集实施自适应线性神经元算法,迭代校正权重,流程如图1所示。

2.1 自适应线性神经元法

自适应线性神经元(Adaline)是1961年由美国

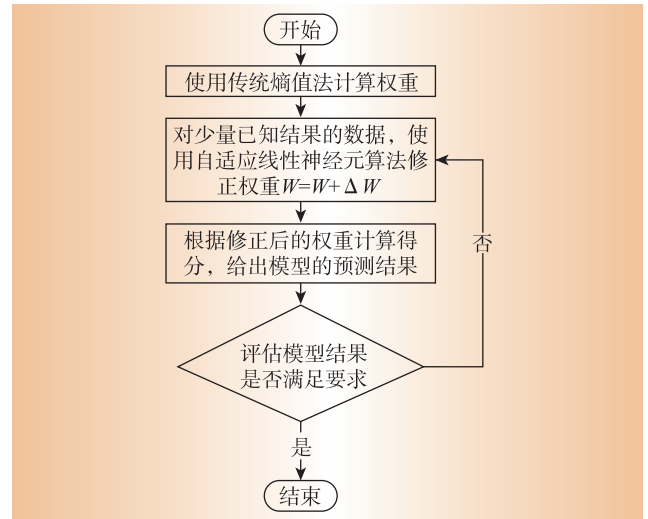


图1 改进熵值法原理

Fig. 1 Improved entropy method theory

斯坦福大学教授 Windrow 提出的,它是一个自适应可调网络,适用于信号处理中的自适应滤波、预测和模型识别,由自适应线性神经元构成的自适应滤波器可以在未知输入噪声先验知识的情况下通过神经网络系统对输入的外界干扰噪声自学习达到消除的目的^[5-6]。

2.2 权重优化

权重优化方法:定义最小二乘法最小化代价函数 $J(w)$,利用梯度下降法更新权值,熵值法求得权重作为初始权重。具体步骤如下。

通过权重计算得分

$$o_i = \sum_{j=1}^k x_j w_j \quad j = 1, 2, \dots, k \quad (7)$$

计算误差

$$J_i = \frac{1}{2} \sum_i (t^{(i)} - o^{(i)})^2 \quad (8)$$

式中: t 为真实值; o 为预测值。

利用梯度下降法更新权重

$$\begin{aligned} \frac{\partial J}{\partial w_j} &= \frac{\partial}{\partial w_j} \frac{1}{2} \sum_i (t^{(i)} - o^{(i)})^2 = \\ &= \frac{1}{2} \sum_i \frac{\partial}{\partial w_j} (t^{(i)} - o^{(i)})^2 = \\ &= \frac{1}{2} \sum_i 2(t^{(i)} - o^{(i)}) \frac{\partial}{\partial w_j} (t^{(i)} - o^{(i)}) = \\ &= \sum_i (t^{(i)} - o^{(i)}) \frac{\partial}{\partial w_j} (t^{(i)} - \sum_j w_j x_j^{(i)}) = \\ &= \sum_i (t^{(i)} - o^{(i)}) (-x_j^{(i)}) \end{aligned} \quad (9)$$

$$\begin{aligned} \Delta w_j &= -\eta \frac{\partial J}{\partial w_j} = -\eta \sum_i (t^{(i)} - o^{(i)}) (-x_j^{(i)}) = \\ &= \eta \sum_i (t^{(i)} - o^{(i)}) x_j^{(i)} \end{aligned} \quad (10)$$

式中: η 为修正参数, $w_j = w_{j-1} + \Delta w$, 其中, w_0 为通过熵值法计算的权重。

3 改进后的停电敏感度识别模型

3.1 构建评价指标体系

停电敏感度当前无统一明确的定义,客服中心倾向于按停电后拨打电话的情况认定敏感度,电力营销人员倾向于按停电造成的损失等影响来评估其敏感性^[7],企业责任人则以可能引发的舆情风险为其敏感性^[2,9]。本文考虑到指标体系的实用性、全面性和科学性,站在企业管理者的角度,将上述3类因素均考虑在内,构建了基于客户基础信息、用电情况、行为轨迹、停电情况、舆情风险5个维度的指标体系,包含20个指标子项,具体如下表1所示。

由于居民客户敏感度更倾向于用于舆情管控,非居民客户停电敏感度更倾向于用于减少损失,因此居民客户采取用电情况、行为轨迹、停电情况、舆情风险4个维度指标,非居民客户采用用户基础信息、用电情况、行为轨迹、停电情况4个维度指标。由于篇幅限制,以下只论述用非居民敏感度情况。

表1 停电敏感度评估指标体系

Table 1 Power outage sensitivity evaluation index system

一级指标	二级指标	指标定义
基础信息	行业类别	营销系统中登记的行业类别
	电源数量	单电源、双电源等
	重要客户	是否为优先服务客户
	用电容量	单位kVA
	经营范围	水产、水泥、养殖等
用电情况	近一个月用电量	营销系统中客户近一月的用电量
	近一年用电量	营销系统中客户近一年的用电量
	缴费信用	客户的缴费及时性等
行为轨迹	近3个月内停电时间内平均拨打95958次数	3个月内各次停电引发的拨打次数平均值
	近一月内拨打95598次数	向95598反映问题的倾向
	对停电问题投诉次数	停电问题引发的投诉次数
停电情况	停电类型系数	故障停电、限电、计划停电等综合评估,故障停电优先级优于计划停电等
	近一个月停电频次	近期停电次数
	近一个月平均停电时长	近期停电时长
	近一年平均停电次数	年度停电次数
	近一年平均停电时长	年度停电时长
舆情风险	城乡类别	城市维权意识远高于农村
	维权意识评分	通过环京、经济发达程度等综合评估
	风险预警次数	提出监管机构及媒体反映倾向次数
	实际发生舆情	实际向上级单位、监管机构及媒体反映次数

3.2 指标归一化处理

由于指标值不在同一单位区间,均需进行标准化处理,转换成[0,100]之间的数值。本文指标分为离散型和连续型指标。其中行业类别、电源数量、重要客户、经营范围、缴费信用、停电类型系数、城乡类别均为离散型指标,其他指标为连续型指标。

在离散型指标归一化中,为了避免值离散性过大对权重造成较大影响,本文采用专家评分法进行归一化得分。如:重要客户只要有要客和非要客,为了减少该指标对其他指标的影响,需减少2个值的差距,因此要客评为100分,非要客评为50分而不是0分;行业类别中钢铁冶炼类行业停电造成损失影响巨大,因此其得分评为100分,农业灌溉类停电影响较小,因此得分仅评为10分;其他离散指标做类似处理。

连续型指标归一化方法主要有直线型阈值法、Z-score法、折线型和曲线型算法。本文采用直线型阈值法进行处理。用电容量、用电量、拨打95598次数、投诉次数、停电次数、停电时长均为连续型正指标,其归一化公式为

$$x_i = \frac{x_i^*}{x_{\max}^*} \times 100 \quad i = 1, 2, \dots, m \quad (11)$$

式中: $x_{\max}^*$ 为该指标样本中的最大值。

3.3 通过熵值法确定初始权重

通过选取20组典型数据,利用熵值法计算非居民客户初始权重。初始权重见表2,其中停电情况、行为轨迹权重较为突出,分别为53.43%、39.52%,客户基础信息、用电情况权重较低,分别为6.02%、1.03%。显然有些指标权重过大,弱化了像经营范围等重要指标。

表2 非居民客户敏感度指标熵值法初始权重

Table 2 The initial weight of the entropy method for the sensitivity index of non-resident customers

一级指标	一级权重	二级指标	二级权重
客户基础信息	6.02	行业类别	1.74
		电源数量	0.54
		重要客户	0.53
		用电容量	0.94
		经营范围	2.28
用电情况	1.03	近一个月用电量	0.81
		近一年用电量	0.15
		缴费信用	0.07
行为轨迹	39.52	近3个月内停电时间内平均拨打95958次数	9.48
		近一月内拨打95598次数	12.79
		对停电问题投诉次数	17.25
		停电类型系数	10.62
停电情况	53.43	近一个月停电次数	13.37
		近一个月平均停电时长	12.47
		近一年平均停电次数	8.51
		近一年平均停电时长	8.47

3.4 基于自适应线性神经元算法优化权重

国网客服中心坐席使用手工标签对停电敏感客户进行打分标记,利用标记后的数据重新训练权重。使用python脚本编写自适应线性神经元算法程序,利用梯度下降法更新权重。设定更新次数为100次, $\eta=0.1$,收敛曲线如图2所示,确定最终权重如表3所示。可以看出客户基础信息部分权重较熵值法有了大幅提升,使得评估结果更具有科学性。

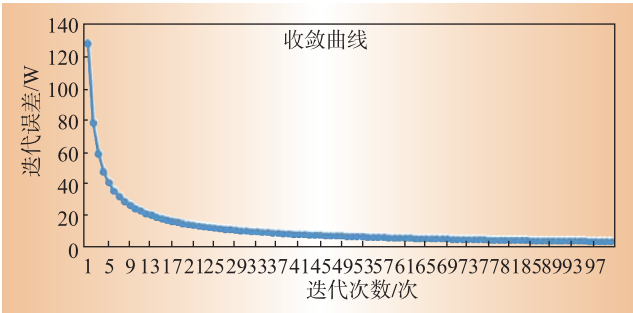


图2 迭代收敛曲线

Fig. 2 Iterative convergence curve

表3 利用自适应线性神经元算法更新后的权重

Table 3 The update weight using adaptive linear neuron algorithm

一级指标	一级权重	二级指标	二级权重
客户基础信息	30.03	行业类别	8.78
		电源数量	4.36
		重要客户	2.41
		用电容量	1.27
		经营范围	13.21
用电情况	5.73	近一个月用电量	2.32
		近一年用电量	2.34
		缴费信用	1.07
行为轨迹	31.94	近3个月内停电时间内平均拨打95958次数	9.76
		近一月内拨打95598次数	10.07
		对停电问题投诉次数	12.11
停电情况	32.30	停电类型系数	5.62
		近一个月停电次数	7.49
		近一个月平均停电时长	9.47
		近一年平均停电次数	4.98
		近一年平均停电时长	4.74

3.5 识别效果校验

本文分别完成了对国网2 166户非居民用户、居民用户的停电敏感度识别,如图3所示。

国网客服中心人员根据95598历史来电记录及回访情况,按照表4所述情景对应的分值,对上述居民、非居民各2 166位客户的停电敏感度进行手工

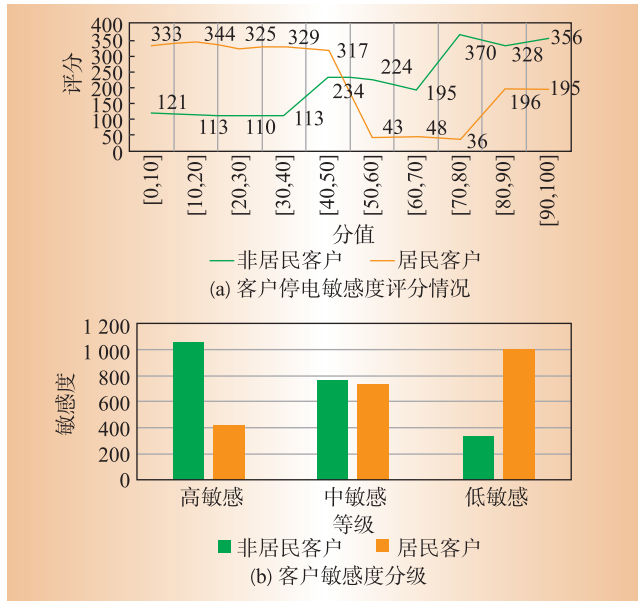


图3 客户敏感度识别情况

Fig. 3 Customer sensitivity identification situation

打分标记,分别与改进熵值法、熵值法的预测结果进行对比,计算算法识别准确率。

表4 手工标签评分表

Table 4 Manual label rating form

类型	情景	分值
来电投诉	客户对停电问题严重不满,要进行上级投诉	100
	客户投诉停电问题,但未造成损失	80
造成损失	客户投诉停电问题,造成1万元以下损失	[80,90]之间
	客户投诉停电问题,造成1万元损失	90
	客户投诉停电问题,造成10万元以上损失	100
提出意见	客户投诉停电问题,造成1万~10万元损失	[90,100]之间
	客户对停电安排及处理工作提出意见	85
回访情况	客户反映停电问题,回访客户,客户表示对停电处理不满	75
	客户反映停电问题,回访客户,客户表示对停电处理满意	50
	客户近2个月来,来电2次报修(无投诉、对处理结果满意)	50
频繁来电	回访客户,客户对停电造成的影响或将要造成的影响,表示影响不大、一般、引起一定程度的不便、影响生产生活、严重影响、不能接受,分别按右侧数值评分	[0,20,50,80,100]
	客户近2个月来,来电5次及以上报修(客户无投诉且对处理结果满意)	100
	客户近2个月来,来电2次以上、5次以下报修(无投诉、对处理结果满意)	[50,100]之间

其中,定义敏感度识别准确率为:手工标记敏感度与预测敏感度匹配数量与调查总量的比值。

其中,非居民客户、居民客户用改进熵值法识别的准确率为89.95%、85.63%,其中非居民客户由于其特征相较居民客户明显,识别准确率相对较高,相较单纯用熵值法识别准确率提高了19.75%、17.98%。

各敏感度区间的识别准确率情况见图4,可以看出改进后总体精度提升,且各区间的预测准确率较改进前波动性减小。

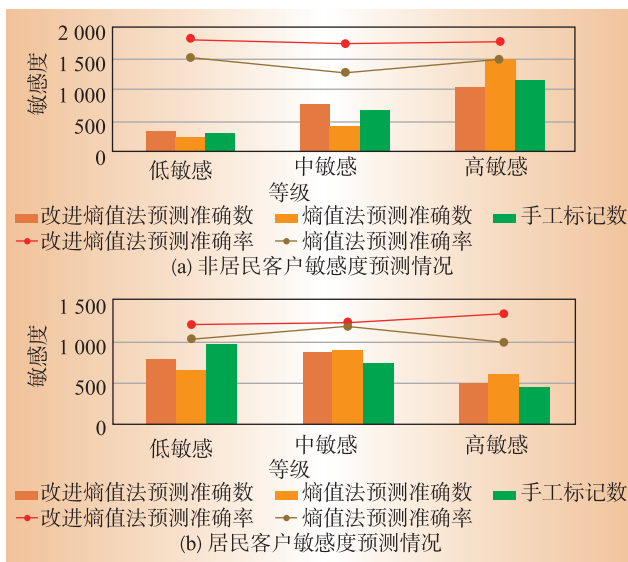


图4 预测准确度对比数据

Fig. 4 The compared data of prediction accuracy

4 结束语

本文构建了完备的非居民客户和居民客户停电敏感度评价指标体系,用改进熵值法调整权重,使得评价和识别方法更加科学。通过对比手工评分情况可以看出,改进熵值法在识别准确度上高于传统熵值法,可以更精准地识别出停电敏感客户群。

精准的停电敏感度识别结果,为国网客服中心和供电公司采取差异化服务策略提供了支撑。未来,该方法可扩展应用于同类型的客户行为预测建模,在精准识别客户、提升客户服务满意度领域将有更广阔应用。

参考文献:

[1] 魏小曼,余坤,陈星莺,等. 基于 Affinity propagation 和 K-means 算法的电力大用户细分方法分析[J]. 电力需求侧管理,2018,20(1):15-19.
WEI Xiaoman, YU Kun, CHEN Xingying, et al. Based on algorithm of affinity propagation and K-means [J]. Power Demand Side Management, 2018, 20(1):15-19.

[2] 郑芒英. 用电客户停电敏感度分析[D]. 广州:华南理工大学,2014.
ZHENG Mangying. Sensitivity analysis of power outage for electric customers [D]. Guangzhou: South China University of Technology, 2014.

[3] 黄国庆,王明绪,王国良. 效能评估中的改进熵值法赋权[J]. 计算机工程与应用,2012,48(28):245-248.
HUANG Guoqing, WANG Mingxu, WANG Guoliang. Weight assignment research of improved entropy method in effectiveness evaluation [J]. Computer Engineering and Application, 2012, 48(28):245-248.

[4] 代吉亚,宋轶,郭汝宁,等. 层次分析法和熵值法在疾控机构应急能力评估中的应用比较[J]. 中国卫生统计,2014,31(6):1 051-1 053.
DAI Jiya, SONG Yi, GUO Runing, et al. Comparison of analytic hierarchy process and entropy value method application in assessment of emergency response capacity of CDC institutions [J]. Chinese Journal of Health Statistics, 2014, 31(6):1 051-1 053.

[5] 张雷,张兵,李广强,等. 装备效能评估中一种组合赋权方法研究[J]. 空军雷达学院学报,2011,25(4):280-282,286.
ZHANG Lei, ZHANG Bing, LI Guangqiang, et al. Study of combination weighting approach for equipment efficiency evaluation [J]. Journal of Air Force Radar Academy, 2011, 25(4):280-282, 286.

[6] 张良均,王路,谭立云,等. Python 数据分析与挖掘实战 [M]. 北京:机械工业出版社,2017:95-97.
ZHANG Liangjun, WANG Lu, TAN Liyun, et al. Python practice of data analysis and mining [M]. Beijing: China Machine Press, 2017:95-97.

[7] SEBASTIAN R. Python machine learning [M]. Birmingham: Packt Publishing, 2015:33-36.

[8] 耿俊成,张小斐,袁少光,等. 基于逻辑回归模型的电力客户停电敏感度评分卡研究与实现[J]. 电力需求侧管理,2018,20(3):46-50.
GENG Juncheng, ZHANG Xiaofei, YUAN Shaoguang, et al. Research and implementation of users' outage sensitivity score card based on logistic regression model [J]. Power Demand Side Management, 2018, 20(3):46-50.

[9] 梁思博. 配电网停电影响评估的研究与应用[D]. 北京:华北电力大学,2017.
LIANG Sibao. Research and application of evaluation of power distribution network outage influence [D]. Beijing: North China Electric Power University, 2017.

(责任编辑 徐文红)