

电力客户价值细分模型研究及应用

电力客户细分对电力客户价值的分析研究是供电企业实施差异化服务、实现利益最大化的基础。设计电力价值评定指标,通过专家打分分别对客户当前价值、潜在价值标注等级,建立了基于CART决策树的客户价值决策模型,混合利用TOPSIS算法得到客户最终分值。CART决策树价值划分模型预测的结果经过对比分析证实具有较高的预测准确度和鲁棒性,是一种有效的电力客户价值评价方法。

拓广忠¹ 李晓东¹ 付薇薇² 王超¹ 姚建河¹ / 1.北京博望华科科技有限公司
2.北京科技大学计算机与通信工程学院

0 引言

随着基于大数据挖掘技术与行业经验逐步地被供电企业吸收利用,很多的国外供电企业利用此技术融合实际经营管理的特点后,构建客户价值的评价体系^[1]。如苏格兰的Scottish Power公司建立了Prime Vantage系统;法国国家电力公司EDF以客户价值为导向,积极评估分析客户的价值制定针对性的供电服务产品^[2]。国内,供电企业纷纷对现有的营销系统进行深度应用,利用系统庞大的客户行为数据为基础,通过经验打分、综合评价等方式对客户进行价值评估。

传统评估方法包括德尔菲法^[3]、层次分析法^[4]和熵值法^[5]等,遇到指标过多时权重难以确定,定量数据较少,说服力低。数据挖掘方法包括神经网络^[6]、支持向量机^[7]、层次聚类^[8]和主成分分析法^[9]等。传统方法在处理电力客户价值问题时显得过于死板,一般利用线性方法确

定客户分值。而数据挖掘的方法在给客户价值分类时准确度较高,但是不能准确地给价值评分。

本文综合了传统方法和数据挖掘算法,用专家人工打分数据作为学习样本,依据决策树算法优化数据构建决策树分类模型,混合TOPSIS方法给客户评定分值,本方法侧重于客户价值的分值细分,进一步提高了客户等级划分的差异性。

1 电力客户价值体系

1.1 价值指标体系构建

客户价值一般划分为两方面:①当前价值,即如果客户当前购电行为模式不发生改变,这类客户会给公司带来的收入价值;②潜在价值,即如果公司通过有效的差异化服务、调动客户购买积极性等,将来可能增加的客户价值^[10]。

选取客户当前价值时,考虑到客户电压等级和当前变压器容量都对用电量有线性同比增加的影响,而客户电费对当前价值有直接影响,电量一般与电费成正比,所以只分析电费指标。同样在考虑客户潜在价值时,根据价值划分理论,将为供电公司带来潜在可能收入或是可能损失的指标作为潜在价值评价的指标,包括电费趋势,最近一年年违约使用电费,最近一年年违约次数,去年年电费,三方协议签署,连续两年均没有电费。

根据文献将客户价值从当前价值和潜在价值两方面构建了本文的细分指标



拓广忠/高级工程师

关键词/Keywords

数据挖掘·
决策树·
TOPSIS·
价值细分·
指标体系·

基于大数据的客户增值服务平台关键技术研究及应用
(52018K15000G)。

体系，如图1所示。其中

$$\text{电费趋势} = \frac{\text{最近一年电费} - \text{去年年电费}}{\text{去年年电费}} \quad (1)$$

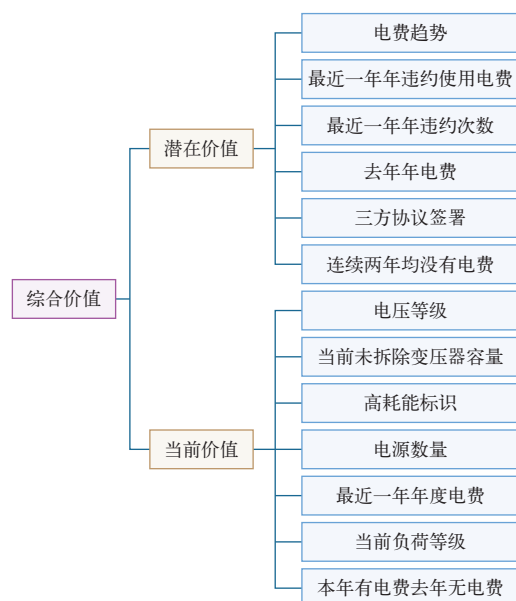


图1 电力客户价值度量的指标体系

1.2 价值指标数据描述

电力数据没有可以描述客户价值的指标，采用专家打分的方式随机选取数据生成样本容量

为2 100学习集。原始客户数据量不能直接用于分析，对各项指标的数据类型、范围及密度有一个初步了解有利于分析决策的准确性。当前价值的指标数据描述如表1所示，潜在价值的指标数据描述如表2所示。本文抽取2015年6月~2016年6月，2014年6月~2015年6月数据做实验。

通过数据描述可以看出指标数据分为数值型（即连续型）、离散型和字符型，字符型数据在数据预处理阶段需要转换成数值用于决策树模型作分析预测^[1]。

2 CART决策树价值等级评估

2.1 训练测试集预处理及生成

决策树属于概率模型，不需要归一化，但是需要进行数据分析，首先进行数据清洗。填补缺漏的数据值、平滑噪声数据、识别或更替异常值，以及解决不一致问题。

将清洗后的两套数据按照价值等级分为4类，分别是优、良、中和差。根据不同的等级分类以行向量为单位随机抽取1/4的数据作为测试集，剩下3/4数据去除价值等级属性作为训练集。当前价

表1 当前价值各项指标的数据描述

指标	电压等级 /kV	当前未拆除变压器容量 /kV·A	电源数量 /个	最近一年年度电费 /元	负荷等级	本年有电费去年无电费	高耗能标识
数据类型	连续型	连续型	连续型	连续型	字符型	离散型	字符型
最小值	6	0	0	-1.47	低：1 944	是：133	电石
1st Q_u	10	315	1	4.18	高：52	否：1 967	钢铁
中位数	10	500	1	36.81	中：104		烧碱
平均数	13	3 158	1.011	372.04			铁合金
3rd Q_u	10	1050	1	100.04			水泥
最大值	220	3 060 001	2	53 311.69			非高耗能

表2 潜在价值各项指标的数据描述

指标	最近一年违约次数/次	电费趋势(%)	最近一年年违约使用电费/元	去年年度电费/元	三方协议	连续两年均没有电费
数据类型	连续型	连续型	连续型	连续型	离散型	离散型
最小值	0	-1.151 64	0	-2.61	0：1 929	0：1 840
1st Q_u	0	-0.367 28	0	9.68	1：171	1：260
中位数	0	-0.027 07	0	42.22		
平均数	0.176 7	0.293 93	3 185	412.72		
3rd Q_u	0	0.067 46	0	108.36		
最大值	39	229.082 48	375 000	58 287.29		

值和潜在价值各有1 575条数据的训练集和525条数据的测试集，并且两种价值模型的训练测试集互不相同。

2.2 决策树建模过程

生成电力客户决策树模型的分类步骤^[12]如下所述。

1) 步骤1。对候选电力价值指标中的每一个指标用CART算法计算该指标上不同分割方法的GINI系数，选择GINI系数最小的指标作为分裂的根节点。然后比较所有候选指标上最佳划分的GINI系数，拥有最小划分GINI系数的指标成为最终测试属性。

GINI系数的计算方法。首先，样本集的GINI系数为

$$\text{gini}(T) = 1 - \sum p_j^2 \quad (2)$$

式中， p_j 为分类 j 在 T 中出现的概率。其次，若 T 被划分为 T_1 、 T_2 ，则此次划分的GINI系数为

$$\text{gini}_{\text{split}}(T) = \left| \frac{S_1}{S} \right| \text{gini}(T_1) + \left| \frac{S_2}{S} \right| \text{gini}(T_2) \quad (3)$$

式中， S 为 T 样本中的个数； S_1 、 S_2 分别为 T_1 、 T_2 中样本的个数。

2) 步骤2。若分裂指标是离散型数据，如三方协议的签署情况，样本集按照在连续指标的分割阈值，分成两部分；若分裂指标是连续型数据，则按照在该指标上的取值是否包含在该指标具有最小GINI系数的真子集中，分成两部分。

3) 步骤3。对根节点的分裂属性对应的两个样本子集 S_1 和 S_2 ，采用与步骤1相同的方法递归地

建立树的子节点。如此循环下去，直至所有子节点中的样本属于同一类别或没有可以选作分裂属性的属性为止。

4) 步骤4。剪枝生成的决策树。对于某个连续型指标 A_c ，假设在某个节点上的样本集 S 的样本数量为 total ，CART算法将对连续指标作处理。首先，对该节点上的所有样本，由小到大对连续型指标 A_c 的具体数值排序，得到指标值序列 $\{A_{1c}, A_{2c}, \dots, A_{\text{total}c}\}$ 。其次，在指标值序列中生成 $\text{total} - 1$ 个分割点。第 i ($0 < i < \text{total}$) 个分割点的取值设置为

$$V_i = (A_{ic} + A_{(i+1)c}) / 2 \quad (4)$$

可以将节点上的样本集划分为 $S_1 = \{s | s \in S, A_c(s) \leq V_i\}$ 和 $S_2 = \{s | s \in S, A_c(s) > V_i\}$ 两个子集， $A_c(s)$ 为样本 s 在属性 A_c 上的取值。最后，计算 $\text{total} - 1$ 个分割点的GINI系数，依旧选择GINI系数最小的分割点来划分样本集。

2.3 决策树模型预测

以当前价值模型为例，建立CART决策树模型。计算所有指标的GINI值，根据指标的连续性和GINI值选取合适的分割阈值。对分割后的两个字迹继续用上一步的方法取分割阈值划分指标，直到所有的数据都有一个分类结果。通过R语言编程并完成数据的计算，打印当前价值和潜在价值的预测模型如图2所示，左侧是当前价值决策树预测模型，右侧是潜在价值决策树模型。

将测试集数据输入训练好的模型中，得到测试集每个客户向量的价值等级分类结果。测试集

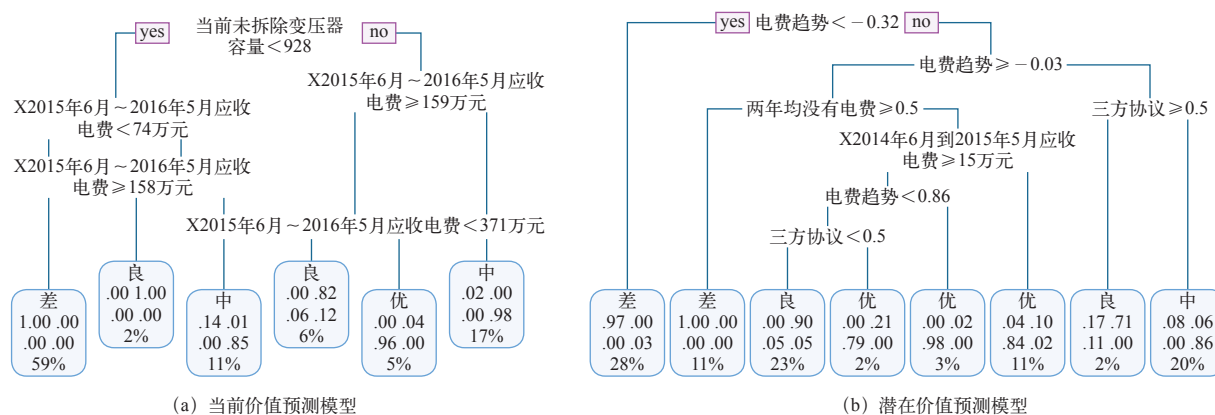


图2 决策树价值评价模型树枝构成

利用模型生成的结果与真实结果对比,正确率分布所示项将剩余数据输入预测模型,预测出的客户价值与实际专家标注价值混淆对比矩阵如表3所示,左表为当前价值模型性能评价混淆矩阵表,右表为潜在价值模型性能评价混淆矩阵表。可以看到各按照当前价值和潜在价值的不同等级抽取的共525个客户的当前价值预测正确率为96.00%,潜在价值预测准确率为94.10%,能很好地预测出客户的价值等级。

表3 模型性能评价混淆矩阵表

当前价值	优	良	中	差	潜在价值	优	良	中	差
优	48	2	0	0	优	76	5	0	1
良	3	36	1	0	良	5	119	5	1
中	0	4	106	7	中	1	4	106	6
差	0	0	4	314	差	0	1	2	193

混淆矩阵的结果显示决策树模型能够较为准确地模仿专家思路预测用户的等级。可以将剩余14 326户客户输入训练好的决策树模型中,得到这些客户等级的预测结果如表4所示。

表4 客户价值预测结果

	优	良	中	差
当前价值	8 680	3 246	1 003	1 397
潜在价值	5 711	3 061	3 598	1 956

3 TOPSIS综合评分

TOPSIS(逼近理想解排序法)的基本原理是借助离正理想解和负理想解的距离来对评价对象进行排序。正理想解是一个虚拟的最优解,各个指标均为对象中的最优值;负理想解是虚拟的最差解,各个指标值都达到对象中的最差值。如果评判对象离最优解最近同时又最远离最劣解,则为最好;否则为最差^[13]。

3.1 指标数据归一化

首先对指标数据做数据归一化工作。第一章详细描述了指标的类型和分布,数值型指标可以直接套用归一化公式预处理,离散性指标三方协议由于只有两个值0、1不需要预处理,字符型指标负荷等级需转换成对应的数值1、2和3之后进行归一化处理,利用式(5)对人工标注客户等级的

标签作归一化处理,免去量级的影响^[13],即

$$y_i = (x_i - x_{\min}) / (x_{\max} - x_{\min}) \quad (5)$$

式中, x_i 、 y_i 分别为处理前后的指标项; $x_{\min}$ 为该项指标的最小值; $x_{\max}$ 为最大值。

3.2 权重和参数设置

TOPSIS打分使用的权重由业务专家给出:电压等级为2;当前未拆除变压器容量为3;电源数量为2;年度电量为8;年度电费为8;负荷等级为2;电费趋势为4;电量趋势为4;违约使用电费为-3;去年年电费为5;三方协议为4。

TOPSIS可以根据指标的权重、相关性给客户打分,实现客户价值的进一步评估。将没有专家标记的剩余客户数据预处理后,分别输入当前价值评估模型和潜在价值评估模型,得到客户的优良中差分类,以客户当前价值为例,对不同等级的客户:设多属性决策问题的方案集为 $A = \{A_1, A_2, \dots, A_m\}$,属性集为 $F = \{f_1, f_2, \dots, f_n\}$,决策矩阵 $X = (x_{ij})_{m \times n}$,其中 x_{ij} 为第 i 个在第 j 个属性下的指标值, $i \in M, j \in N, M = \{1, 2, \dots, m\}$ 为方案的下标集, $N = \{1, 2, \dots, n\}$ 为指标的下标集。方案 A_i 可记为 $A_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}, i \in M$ 。规定 $x_{ij} \geq 0$,属性的权向量为 $\omega = \{\omega_1, \omega_2, \dots, \omega_n\}$,满足

$$\sum_{j=1}^n \omega_j = 1 \quad \omega_j \geq 0, \quad j \in N \quad (6)$$

在本方法中 $m = 2, n = 13$,分别对应当前价值、潜在价值和13项三级指标^[14]。

3.3 TOPSIS计分步骤

1) 用向量规范化法,构造规范化决策矩阵 $Y = (y_{ij})_{m \times n}$,计算方法为

$$y_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}} \quad i \in M, j \in N \quad (7)$$

2) 即构造加权规范化决策矩阵 $Z = (z_{ij})_{m \times n}$,其中 $z_{ij} = \omega_j y_{ij}, i \in M, j \in N$ 。这里权重 ω_j 由专家给出。

3) 确定正理想解 A^+ 和负理想解 A^- 为

$$A^+ = (z_1^+, z_2^+, \dots, z_n^+) \quad (8)$$

$$A^- = (z_1^-, z_2^-, \dots, z_n^-) \quad (9)$$

式中, $z_j^+ = \max z_{ij}, z_j^- = \min z_{ij}, j \in N$ 。

4) 计算各方案分别与正理想解和负理想解的Euclid距离 d^+ 和 d^-

$$d^+ = \sqrt{\sum_{j=1}^n (z_{ij} - z_j^+)^2} \quad (10)$$

$$d^- = \sqrt{\sum_{j=1}^n (z_{ij} - z_j^-)^2} \quad (11)$$

5) 计算各方案与正理想解的相对贴近度

$$C_i^+ = \frac{d_i^-}{d_i^+ + d_i^-} \quad i \in M \quad (12)$$

6) 按照 C_i^+ 的降序排列, C_i^+ 越大其对应的方案越优,排序越靠前,由此得出方案的优劣次序。

3.4 混合决策树和TOPSIS的模型实验结果

为了保留决策树模型的等级预测结果,将客户按照优良中差四个等级分开打分,这样既保证了等级的有效性,同时完成客户的进一步细分。可以得到剩余14 327名客户最终的得分如图3所示。



图3 混合决策树模型和TOPSIS算法的客户评分

由2.3节可知等级划分的正确率较高,客户在综合价值等级划分时能够抽象出专家评价思路,给客户价值划分为优良中差四大类。通过图3打分结果可知不管是当前价值还是潜在价值,客户在不同级别的高分值区域分布较少,在中低部分分布较多。综合整体分布结果同样优良的高分分布较少,中差分布较多。这种分布结果也符合二八定理,即小部分的客户为公司创造了大部分的价值,说明其评分结果的合理性。可以看出在当前价值较高的领域,客户的当前价值和潜在价

值有呈现出一种正相关的趋势。

4 结束语

本文将CART决策树算法和TOPSIS结合,为电力行业的客户细分细则提出了一种新的思路和方法。决策树更能从主观打分中提取一般方法,保留了主观赋权的优越性,通过实验数据显示准确率较高,利用算法达到由专家打分的结果。而TOPSIS法既能对各方案进行整体评判,又能反映各方案内部各因素对评价结果的影响,对预测分类后的数据进一步细化评分,克服了数据挖掘在客户价值评价方面数据集获取困难的弊端,是一种合理有效的评价方法。

参考文献

- [1] 冯晓波. 基于大数据模型搭建客户价值评价体系——对大型服务企业管理会计实践的探索[J]. 财政监督, 2015(29): 45-48.
- [2] Macfarlane D. Outage management in Scottish Power [C]. IEEE Power Engineering Society Winter Meeting, 1999: 860-865.
- [3] 张冀东, 曹连生, 燕振洲, 等. 基于层次分析法和德尔菲法的呼和浩特城区配网规划综合效益分析[J]. 内蒙古电力技术, 2009, 27(4): 10-12.
- [4] 刘新才, 曾鸣, 黄琰. 基于层次分析法的电力客户信用风险评价[J]. 电力需求侧管理, 2005, 7(6): 19-21.
- [5] 卓元志, 刘家军, 王明军, 等. 基于熵值法的组合模型用电量预测方法研究[J]. 电网与清洁能源, 2011, 27(5): 47-50.
- [6] 周林, 栗秋华, 刘华勇, 等. 用模糊神经网络模型评估电能质量[J]. 高电压技术, 2007, 33(9): 66-69.
- [7] 杨波. 基于支持向量机的电力系统短期负荷预测[D]. 锦州: 辽宁工业大学, 2015.
- [8] 胡晓雪, 赵嵩正, 吴楠. 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.

(下转第86页)

2) 通过实例分析验证了基于频域特征比较的空心电抗器匝间绝缘故障诊断判别方法能够更加准确、方便地检测电抗器匝间绝缘故障,有效避免由于空心电抗器匝间故障引起的事故,加强电网安全稳定性。

参考文献

- [1] 张国兵,张文亮.特高压交流输电系统中110 kV并联电抗器试验及匝间短路保护问题辨析[J].高电压技术,2012,38(2):303-309.
- [2] 宋喜平.脉冲振荡电压下电抗器匝间绝缘击穿特性[D].哈尔滨:哈尔滨理工大学,2016.
- [3] 廖敏夫,程显,翟云飞.干式空心电抗器脉冲振荡匝间绝缘检测系统的仿真与试验[J].高电压技术,2011,37(6):1343-1348.
- [4] 翟云飞.干式空心电抗器匝间绝缘检测系统仿真与实验[D].大连:大连理工大学,2011.
- [5] 江少成,戴瑞海,夏晓波,等.干式空心电抗器匝间绝缘检测原理及试验分析[J].高压电器,2011,47(6):67-71.
- [6] 聂洪岩,徐万里,徐雅清,等.干式空心电抗器匝间绝缘试验电压的产生方法研究[J].电力科学与技术学报,2016,31(2):45-50.
- [7] 张磊,黎大健,赵坚.干式空心电抗器匝间绝缘

检测技术探讨[J].广西电力,2015,38(1):44-45,56.

- [8] 潘臻,杨萍,江峰.用脉冲电压谐振法检测干式电抗器匝间绝缘的现场应用[J].电力电容器与无功补偿,2015,36(5):88-91.
- [9] 青晔.大型空心电抗器电磁与损耗计算[D].哈尔滨:哈尔滨理工大学,2012.
- [10] 王龙江.空心电抗器绝缘电热联合老化试验方法研究[D].哈尔滨:哈尔滨理工大学,2014.
- [11] 江少成,戴瑞海,夏晓波,等.干式空心电抗器匝间绝缘检测原理及试验分析[J].高压电器,2011(6):67-71.
- [12] 张良,吕家圣,王永红,等.35 kV干式空心电抗器匝间绝缘现场试验[J].电机与控制学报,2014,18(6):66-71.
- [13] 芦岩.66 kV干式空心电抗器匝间绝缘试验技术研究[D].哈尔滨:哈尔滨理工大学,2013.
- [14] 戴瑞海,夏晓波,林坚,等.干式空心电抗器匝间绝缘试验关键技术及装置研究[J].浙江电力,2012,31(7):1-5.
- [15] 苟烜齐,刘轩东,杨昊,等.匝间绝缘故障对干式空心电抗器电感参数影响的仿真研究[J].高压电器,2015(10):117-121,127.

(收稿日期:2017-06-27) **EA**

(上接第74页)

- [9] 邢国静.主成分分析法在电信企业客户价值评价中的应用[J].合作经济与科技,2010(7):34-35.
- [10] 刘苗,王维兵.客户价值分析综述[J].中国商界,2008(11):162.
- [11] 谢金梅,王艳妮.决策树算法综述[J].软件导刊,2008(11):83-85.

- [12] 赵培培.基于TOPSIS法的配网管理评价及优化研究[D].北京:华北电力大学,2014.
- [13] 柳小桐.BP神经网络输入层数据归一化研究[J].机械工程与自动化,2010(3):122-123.
- [14] 兰华,郑广瑞,廖志民.基于模糊TOPSIS和BP网络的电力变压器故障诊断方法[J].南方电网技术,2011,5(1):70-73.

(收稿日期:2016-12-30) **EA**