

保险业客户行为忠诚度分析

吴洪波, 张 蕾

(哈尔滨理工大学 管理学院, 黑龙江 哈尔滨 150040)

摘 要:针对保险业如何利用客户数据提高管理决策水平和销售业绩的问题,根据保险业客户忠诚度的判别条件,运用模糊数据挖掘技术,提出了通过对已有客户进行行为忠诚度聚类分析,发现公司内客户所属的类型,并据此对新客户做出忠诚度预测,从而可更好地为保险业客户管理提供帮助.

关键词:保险业; 模糊数据挖掘; 客户忠诚度

中图分类号: F272.1 **文献标志码:** A **文章编号:** 1007-2683(2010)01-0129-04

Analysis of Customer Behavior Loyalty in Insurance

WU Hong-bo, ZHANG Lei

(School of Management, Harbin University of Science and Technology, Harbin 150040, China)

Abstract: To solve at the problem how to utilize customer data to improve management decision-making level and sales in insurance, according to the discriminating conditions of insurance customer loyalty, this paper uses fuzzy data mining technology to make clustering analysis on customers based on behavior loyalty. The analysis points out customers can be classified into different groups and forecasts new customers loyalty according to the result of previous cluster analysis. The method mentioned can enhance the efficiency of customer relationship management in insurance.

Key words: insurance; fuzzy data mining; customer loyalty

由于保险业的服务性及其产品的特殊性,使得客户对于保险公司来说显得尤为重要,客户就是保险公司的生存之本. 优质客户作为一种重要资源,成了国内外各家公司争夺焦点. 如何维持已有客户,提高客户忠诚度,成为各家保险公司必须解决的问题. 数据挖掘(Data Mining,简称DM)是从大量的、不完全的、有噪声的、模糊的和随机的数据中,提取隐含在其中的、事先不知道的,但是又潜在而有用的信息和知识的过程^[1]. 数据挖掘可以快捷有效地处理大量历史和现有数据,从中提取有价值的信息^[2]. 本文运用模糊数据挖掘技术,对保险企业客户的忠诚度进行分析,发现隐藏

的模式,并对新客户的忠诚度做出预测,为企业更好地服务客户,制定有效的市场营销战略提供重要的决策依据.

1 客户行为忠诚度的含义及重要性分析

1.1 客户行为忠诚度的含义

客户忠诚(Customer Loyalty CL)是客户在较长一段时间内,对于企业产品或服务保持的选择偏好与重复性购买^[3]. 忠诚的客户通常指会拒绝竞争者提供的折扣,经常性地购买本公司的产品或服务,会向家人朋友推荐的客户. 《财富》杂志排名前1 000

收稿日期: 2009-03-17

作者简介: 吴洪波(1955—),男,教授,硕士生导师,E-mail:WHB5932@126.com;

张 蕾(1984—),女,硕士研究生.

家公司,大部分都高度重视和强化客户忠诚的培育和管理.其中,“极其清楚”最具价值客户的公司占60%,且公司年收入中排名前10位客户的贡献高达46%^[1].

客户忠诚可以依据外在行为的特征分为三个层次:行为忠诚、意向忠诚和情感忠诚.行为忠诚是客户实际表现出来的重复购买行为;意向忠诚是客户在未来可能购买的意向;情感忠诚是客户对企业及其产品或服务的态度^[1].只有当行为忠诚伴随着较高的态度取向时才产生真正的客户忠诚.态度取向代表了一个客户对一项服务积极倾向的程度,也反映了将该服务商推荐给其他客户的意愿和对首选的服务商再次惠顾的承诺.综合考虑行为和态度两个维度又可根据程度深浅将忠诚度模式分为不同的层次类别^[1].

1.2 客户行为忠诚度在保险业的重要性分析

保险业激烈的竞争,实际上是赢得顾客的竞争.保险产品不像其他有形产品那样具有明显的技术门槛和质量差别,新产品出来后,很容易被“克隆”和模仿,因此保险产品同质化的现象比较严重^[1].在这种情况下,只有通过信息化的客户关系管理,保险企业才能及时准确地把握客户需求,抢先一步开发和推广新产品,获得竞争的主动权.尽管客户行为忠诚对所有的行业来讲都是重要的,但它对保险业尤其关键,这是因为:①忠诚消费在保险客户中比消费物质产品的客户更加普遍;②保险服务提供了更多的人际间交互的机会,因此也就提供了发展客户行为忠诚的机会;③行为忠诚常常被作为降低风险的策略,同时由于购买保险的感知风险比购买物质产品时高,所以保险服务所提供的环境更易于导致客户忠诚^[1].如何准确评估和预测客户的忠诚度是更加重要的.这为企业如何更好地去识别和保留行为忠诚度高的客户,提升和挖掘意向忠诚度高的客户,以及防止客户流失都起着至关重要的作用.

2 保险业客户行为忠诚度模式发现

保险业企业CRM系统中的客户数据仓库积累了关于客户各种信息的海量数据,运用数据挖掘技术可进行客户忠诚度的分析研究.由于忠诚度不同程度的界定没有明确的界限,本文采用模糊数据挖掘技术对其进行分析,可以发现目前客户行为忠诚的不同程度的模式,另外还可以判断

一个新客户属于这些模式中的哪一类,从而采取相应的措施提升客户的态度取向,培养高价值客户.

1)在保险企业数据仓库的全体数据记录上,建立待分类的样本集 U .结合保险公司的业务特点,本文选取数据仓库中的“购买产品总数”、“目前保留情况”、“续保次数”、“向他人推荐数量”这四个指标来进行客户行为忠诚度分析.划分对象共有10个,即论域 $U = (\text{客户 } i), i = 1, 2, \dots, 10$.经过清洗转换后的数据如表1所示.

表1 客户数据表

客户名称	购买产品总数	续保次数	向他人推荐数量	目前保留情况(保留1失0)
客户1	3	1	3	1
客户2	4	2	2	1
客户3	2	0	0	1
客户4	2	0	2	0
客户5	1	0	0	1
客户6	2	2	2	1
客户7	3	2	1	1
客户8	5	3	2	1
客户9	1	1	0	0
客户10	2	2	3	1

由于这些数据不是 $[0, 1]$ 闭区间的数,所以应将这些原始数据标准化,从而使每一个指标值统一于某种共同的数值特性范围.先求其平均值.它们的平均值按式(1)计算:

$$u'_k = (u'_{1k} + u'_{2k} + \dots + u'_{nk})/n = \frac{1}{n} \sum_{i=1}^n u'_{ik}, i = 1, 2, \dots, n \quad (1)$$

由式(1)得平均向量为 $u'_k = (2.5, 1.3, 1.5, 0.8)$, $k = 1, 2, 3, 4$.

然后再按式(2)求出这些原始数据的标准差 S_k :

$$S_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (u'_{ik} - u'_k)^2} \quad (2)$$

由式(2)得标准差向量为: $S_k = (1.204, 1.005, 1.118, 0.4)$, $k = 1, 2, 3, 4$.

再按式(3)求出各数据的标准差 u''_{ik} :

$$u''_{ik} = \left| \frac{u'_{ik} - u'_k}{S_k} \right| \quad (3)$$

由式(3)计算得标准化矩阵,如矩阵 a .

$$a = \begin{bmatrix} 0.42 & 0.30 & 1.34 & 0.5 \\ 1.25 & 0.70 & 0.45 & 0.5 \\ 0.42 & 1.29 & 1.34 & 0.5 \\ 0.42 & 1.29 & 0.45 & 2 \\ 1.25 & 1.29 & 1.34 & 0.5 \\ 0.42 & 0.70 & 0.45 & 0.5 \\ 0.42 & 0.70 & 0.45 & 0.5 \\ 2.08 & 1.70 & 0.45 & 0.5 \\ 1.25 & 0.30 & 1.34 & 2 \\ 0.42 & 0.70 & 1.34 & 0.5 \end{bmatrix}$$

这时得到的标准化数据 U ,但这些还不都在 $[0,1]$ 闭区间内,还须采用下面的极值标准化公式:

$$u_{ik} = \frac{u''_{ik} - u''_{\min k}}{u''_{\max k} - u''_{\min k}} \quad (4)$$

式(4)中, $u''_{\min k}$ 和 $u''_{\max k}$ 分别表示 $u''_{1k}, u''_{2k}, \dots, u''_{nk}$ 中的最大值和最小值.

由式(4)归一化得矩阵 b .

$$b = \begin{bmatrix} 0 & 0 & 0.67 & 0 \\ 0.5 & 0.29 & 0 & 0 \\ 0 & 0.71 & 0.67 & 0 \\ 0 & 0.71 & 0 & 1 \\ 0.5 & 0.71 & 0.67 & 0 \\ 0 & 0.29 & 0 & 0 \\ 0 & 0.29 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0.5 & 0 & 0.67 & 1 \\ 0 & 0.29 & 0.67 & 0 \end{bmatrix}$$

2) 建立模糊相似关系: R 可表示为相似矩阵,一般的形式如式(5).

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \quad \begin{aligned} 0 \leq r_{ij} \leq 1 \\ i = 1, 2, \dots, n \\ j = 1, 2, \dots, n \end{aligned} \quad (5)$$

计算 r_{ij} 的方法很多,这里采用最大最小法,即式(6). 从而由式(6)建立模糊相似矩阵 c .

$$r_{ij} = \frac{\sum_{k=1}^m \min(u_{ik}, u_{jk})}{\sum_{k=1}^m \max(u_{ik}, u_{jk})} \quad (i, j \leq n) \quad (6)$$

$$c = \begin{bmatrix} 1 & & & & & & & & & \\ 0 & 1 & & & & & & & & \\ 0.49 & 0.15 & 1 & & & & & & & \\ 0 & 0.13 & 0.3 & 1 & & & & & & \\ 0.36 & 0.42 & 0.73 & 0.25 & 1 & & & & & \\ 0 & 0.37 & 0.21 & 0.17 & 0.15 & 1 & & & & \\ 0 & 0.37 & 0.21 & 0.17 & 0.15 & 1 & 1 & & & \\ 0 & 0.4 & 0.27 & 0.24 & 0.45 & 0.15 & 0.15 & 1 & & \\ 0.31 & 0.2 & 0.23 & 0.35 & 0.41 & 0 & 0 & 0.14 & 1 & \\ 0.7 & 0.2 & 0.7 & 0.12 & 0.51 & 0.3 & 0.3 & 0.11 & 0.27 & 1 \end{bmatrix}$$

3) 聚类分析. 采用最大树方法进行聚类分析,即构造一个以所有被分类的对象为顶点的图,当 $r_{ij} \neq 0$ 时,顶点 i 与顶点 j 就可以连一条边. 其方法是先画出顶点集中的某一个 i ,然后按 r_{ij} 从大到小的顺序依次连边,要求不产生回路,直到所有的顶点都被连通为止,这样就得到一棵最大树^[6]. 树的每一条边都能赋以某一权数 r_{ij} .

然后对最大树取 λ 截集,即去掉那些权数 $r_{ij} < \lambda$ 的边, $\lambda \in [0,1]$. 这样就把一棵树截成互不连通的几棵子树. 虽然最大树并不唯一,但取了截集之后,所得的子树是一样的. 这些子树即为数据仓库中归纳发现的模式. 最大树图见图 1.

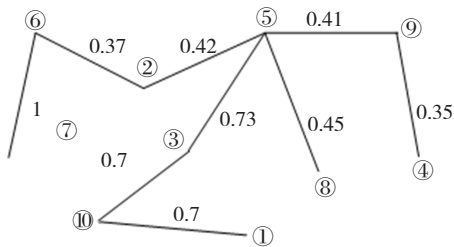


图 1 最大树图

取 $\lambda = 0.73$,将客户分成 8 类:(客户 6,客户 7),(客户 3,客户 5),其余各为一类.

取 $\lambda = 0.42$,将客户分为 4 类:(客户 4,A 类客户),(客户 9,B 类客户);(客户 6 客户 7,C 类客户);其余为一类,(D 类客户). 对所得的每一个模式,可按式(3)~(7)求其平均指标,得到表 2.

表 2 各模式汇总表

客户分类	购买产品总数	续保次数	向他人介绍数量	目前保留情况
A 类客户	2/5(0.4)	0/3(0)	2/3(0.67)	0/1(0)
B 类客户	1/5(0.2)	1/3(0.33)	0/3(0)	0/1(0)
C 类客户	2.5/5(0.5)	2/3(0.67)	1.5/3(0.5)	1/1(1)
D 类客户	2.84/5(0.57)	1.33/3(0.44)	1.63/3(0.54)	1/1(1)

$$\text{Mode}_{ij} = \sum u_{ij}/p, \quad i=1,2,\dots,s, \quad j=1,2,\dots,m \quad (7)$$

采用以上将客户分为4类的分类方式分析客户的忠诚度,这种分类方式客户的数据特征相对集中,有利于提取出有不同类别的显著差异.其中“续保次数”和“向他人推荐数量”可以从侧面反映出客户的态度取向.A类客户具有较高的购买数量和态度取向,但最后却流失了,保险公司应对这些客户的流失原因深入调查,采取必要的措施争取挽回客户;B类客户购买数量和态度取向均较低,最后流失;C类客户具有较高的购买数量和态度取向,目前处于保留状态,保险公司应该积极采取激励措施,提高这类客户的忠诚度;D类客户具有很高的购买数量和态度取向,目前处于保留状态,保险公司应深入挖掘其中的高价值客户,并通过他们争取优质客源.

3 客户的忠诚度预测

如有一新客户资料 $X = (\text{购买产品总数 } 1/5 (0.2), \text{续保次数 } 0/3 (0), \text{向他人介绍数量 } 0/3 (0), \text{目前保留情况 } 1/1 (1))$,与上述聚类分析得出的各个模式进行比对,可判断出新客户所属模式.同一类别的客户存在相同行为的概率很高,从而可以大致预测出新客户的行为.

待预测的样本 $X(1,0,0,1)$ 是在论域 U 上的1个模糊子集,与数据仓库中已分类的模式做比较,求出它们之间的贴适度,其中 Mode_i 代表上一小节中的各模式:

$$(X, \text{Mode}_i) = (1/2) [X \cdot \text{Mode}_i + (1 - X \odot \text{Mode}_i)] \quad i = A, B, C, D \quad (8)$$

式(8)中 $\cdot$ 和 $\odot$ 分别代表模糊运算中的内积和外积, $A \cdot B = \bigvee_{x \in X} (u_A(x) \wedge u_B(x))$, 内积代表 $A \cap B$ 的隶属度的上确界; $A \odot B = \bigwedge_{x \in X} (u_A(x) \vee u_B(x))$, 外积代表 $A \cup B$ 的隶属度的下确界; $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$.

根据式(8)求新客户 $X = (1,0,0,1)$ 对各类的贴适度:1)对A类:内积 = 0.2,外积 = 0,贴适度为0.6;

2)对B类:内积 = 0.2,外积 = 0,贴适度 = 0.6;

3)对C类:内积 = 1,外积 = 0.5,贴适度 = 0.75;4)对D类:内积 = 1,外积 = 0.44,贴适度 = 0.78.

根据择近原则,即 $(X, \text{Mode}_i) = \max((X,$

$\text{Mode}_1), (X, \text{Mode}_2), \dots, (X, \text{Mode}_s))$ 判定新客户 $X = (1,0,0,1)$ 与D类贴适度最大,则新客户的行为很可能跟D类客户一致.

当预测出新客户 X 归属于D类客户时,则其购买产品的总数有希望为3个,续保一次,并向其他2人推荐该保险公司.这时保险公司应根据该客户的需求积极推荐公司的其他险种,提供优质服务的同时提高转换成本,避免失误的产生,提高客户的忠诚度.

4 结 语

数据挖掘能够发现客户内部的隐藏的知识 and 规律,本文详细介绍了模糊数据挖掘算法,并将其引入保险业客户行为忠诚度分析中.通过对客户数据的分析,发现客户的潜在类型.这对保险企业的客户关系管理有着现实的指导意义.保险业通过客户的客观数据进行忠诚度分析,可以帮助企业发现有价值的客户,从而有效制定营销策略,与客户建立长期的合作关系.

参 考 文 献:

- [1] FRAWLEY W J. Knowledge Discovery in Database: An Overview [J]. AI Magazine, 1992, Fall.
- [2] 吴洪波, 赵 谦. 关联规则在大型超市中的应用 [J]. 哈尔滨理工大学学报, 2008(2): 128 - 129.
- [3] DICK A S, BASU K. Customer Loyalty: Toward and Integrated Conceptual Framework. [J]. Journal of Academy of Marketing Science, 1994, 22: 99 - 113.
- [4] 胡 丁, 吴洪波. 保险业的客户忠诚及其影响因素分析 [J]. 科技与管理, 2004, 1: 100 - 102.
- [5] KELLY Scott W, DAVIS Mark A. Antecedent to Customer Expectations for Service Recovery [J]. Journal of Academy of Marketing Science, Winter 1994, 22: 52 - 61.
- [6] 弗雷德里克·莱希赫尔德. 忠诚的价值 [M]. 常玉田, 译. 北京: 华夏出版社, 2001.
- [7] 谢友辉, 蒋新华. 数据挖掘技术及在保险领域中的应用 [J]. 信息技术, 2003, 8: 5 - 11.
- [8] 李昆红. 我国保险公司管理创新的思考 [J]. 保险研究, 2003, 2: 24 - 26.
- [9] 高洪深. 决策支持系统理论方法案例 [M]. 北京: 清华大学出版社, 2006: 106 - 108.
- [10] MARGARET H Dunham. 数据挖掘教程 [M]. 郭崇慧, 田凤占, 勒晓明, 译. 北京: 清华大学出版社, 2005: 189 - 191.

(编辑:付长缨)