

数据挖掘技术在零售企业会员价值预测中的应用研究

刘伟江,段书勇,王 勇

(吉林大学 商学院,吉林 长春 130012)

摘 要:以某零售企业 2008.06.30–2009.06.29 一年的销售数据为依据,以合理地划分会员等级为目标,根据会员一年的消费积分确定其所在的会员层级。在此基础上,采用数据挖掘技术中的决策树方法寻找衡量有价值会员的判定规则,该规则可作为对新客户和潜在客户进行价值判断的依据,为企业向客户做有针对性的营销打下基础,以期能在提高客户满意度的基础上,做到企业和客户的双赢。

关键词:数据挖掘;决策树;会员价值

中图分类号:G931.6

文献标识码:A

文章编号:1007-7634(2010)04-0584-03

Application Research on Using Data Mining Technology to Predict Retail Enterprise Membership Value

LIU Wei-jiang, DUAN Shu-yong, WANG Yong

(Business School, Jilin University, Changchun, 130012, China)

Abstract: Based on a retail enterprise's data from 2008.06.30 to 2009.06.29, this paper aims at a reasonable division of membership level, determines membership level according to their consumption scores. On the basis of this, using one of data mining technique—decision tree method to find rules which can judge valuable membership, these rules can be used to judge new customers and potential customers and lay the foundation for targeted marketing, hope this method can achieve enterprise and customers' win-win based on improving customers' satisfaction.

keywords: data mining; decision tree; membership value

1 引 言

随着信息技术的不断发展,零售企业间的竞争也不断加剧,如何向客户提供高质量的个性化服务,提高客户满意度,保持和发展高价值的客户,已成为许多零售企业所面临的一项严峻挑战^[1-2]。在迎接这一挑战的过程中,很多零售企业已经意识到自己发放的大量会员卡中蕴藏着大量有用的信息。因此,企业如何从已获取的客户数据中挖掘出有用的信息使

之能帮助企业进行决策是很多企业关注的重点^[3-4]。

数据挖掘技术^[5]作为从数据库的大量数据中揭示出隐含的、事先未知的、潜在有用的信息和知识的过程,以其挖掘建模,寻找高价值客户,并提供高质量的客户服务支持^[6]受到了人们广泛的关注,这种方式可为企业留住更多的客户并创造更高的利润。

本文以某零售企业的 2008.06.30–2009.06.29 一年的销售数据为依据,以合理地划分会员等级为目标,根据会员一年的消费积分确定其所在的会员层级。在此基础上,采用数据挖掘技术中的决策树方法

收稿日期:2009-12-12

作者简介:刘伟江(1967-),女,吉林长春人,副教授,经济学博士,主要从事电子商务、数据挖掘等研究。

寻找衡量有价值会员的判定规则^[7],该规则可作为对新客户和潜在客户进行价值判断的依据,为企业向客户做有针对性的营销打下基础。

2 企业内部不同层级会员比例的确定

一般来说,会员层级的确定应该根据该会员对企业的利润贡献来确定,而不同层级的会员在企业中会享受不同的优惠,比如说,最高层级的会员在企业购物时会享受 9 折优惠等,但普通会员却不能享受这种优惠。考虑到相对于消费金额来说,消费积分更能反映出会员对企业的利润贡献,因此根据消费积分来确定会员升级的门槛。

本文根据某企业 2008.06.30-2009.06.29 一年的会员信息数据,得出会员年消费积分分布图(图 1)。由于有少数会员的年消费积分大于 100000,将它们全部在图中表示出来会使横轴显示区域过长,此处仅以 0-100000 积分区间的会员分布图为例。

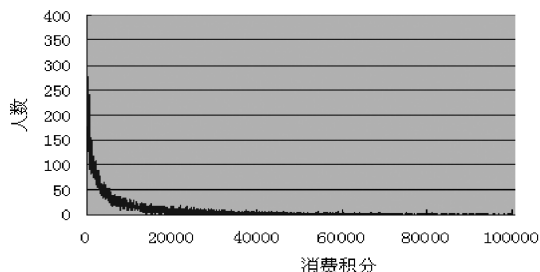


图 1 0-100000 消费积分区间的会员分布图

从图 1 中可以看出,绝大部分会员对企业的利润贡献(消费积分)相对较小,少部分会员对企业的利润贡献(消费积分)较大,而企业现有会员层级比例为:SIP 超级贵宾占总人数比例为 1.35%,VIP 贵宾占总人数比例为 54.44%,VIP 会员占总人数比例为 44.21%。即企业现有会员层级比例中 VIP 贵宾人数占很大比例,这种人数比例分布显然与图 1 的消费积分(利润贡献)的人数分布有很大差异,这是企业长年对会员进行静态管理(即会员层级只升不降)的结果,这种现象显然与企业设计会员层级的初衷不符。因此需要根据该会员在上一年度的利润贡献(消费积分)情况对会员层级进行动态调整。

动态调整的思想为:会员利润贡献(消费积分)越大,会员所处的层级就越高。根据该企业的情况,首先选定刻度,刻度的选择可以根据企业实际情况进行适当调整。本文将平均积分是全体会员平均积分的 10 倍的人群定义为 SIP 超级贵宾,即从全体会员的消费积分分布图横轴最右边开始,找一个消费

积分点 x ,使其到最右边区域人员的平均积分是全体会员平均积分的 10 倍,这段区域的人就是 SIP 超级贵宾;从 x 点开始再往左找一点 y ,使 y 与 x 区间人的平均积分是全体会员平均积分的 3 倍,这段区域的人就是 VIP 贵宾;剩余的人就是 VIP 会员(如图 2)。

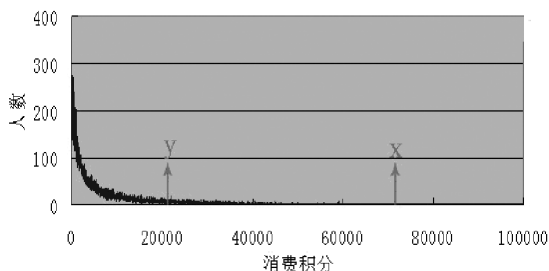


图 2 加入分界点后的不同消费积分区间的会员分布图

确定了各层级会员人数后,就可以根据各层级会员的消费积分情况动态确定下一年各层级会员的升级标准。以 SIP 超级贵宾升级标准为例:由于从 x 点开始右边区域的人数(S)是 SIP 超级贵宾,所以本文对一年中所有会员的单日最高消费积分降序排列,第 S 人的当日消费积分累计值就是成为 SIP 超级贵宾的日升级标准;取每个会员 12 个月(每个月的价值即是连续 30 天的消费积分累计值)中的消费积分最大值,根据这个最大值对所有会员的连续 30 天消费积分累计进行降序排列,第 S 人的连续 30 天消费积分累计就成为 SIP 超级贵宾的连续 30 天消费积分累计升级标准,以此类推就可得到 SIP 超级贵宾和 VIP 贵宾的单日、连续 30 天、连续 365 天的消费积分累计升级标准,而 VIP 会员的升级标准则分别是从小于 y 点向左区域中所有人单日、连续 30 天、连续 365 天消费积分累计的平均值。计算结果如表 1 所示,实际操作时,可根据此数据进行适当调整,如把小数调整为一个适当的整数等。

表 1 会员升级标准及人数比例

	VIP 会员	VIP 贵宾	SIP 超级贵宾
单日消费积分累计	1472.74	4048	11776
过去连续 30 天消费积分累计	2512.59	8392	23896.39
过去连续 365 天消费积分累计	4923.18	20386.15	70409.94
人数比例	84.05%	13.49%	2.46%

由表 1 可以看出,SIP 超级贵宾人数比例最少,VIP 会员人数比例最大,这种比例分布与图 1 所示的人数分布相符。在未来的企业动态管理中,会员消费积分的有效期为一年。任何顾客拿普通证件登记后都可成为普通会员。当普通会员的消费积分达到表 1 所示的相应标准就可以升为相应级别的会员。比方说,VIP 会员如果当日累计 4048 分,可升级为

VIP 贵宾;如果不满足当日累计积分的条件,那么如果 30 天累计 8392 分或 365 天累计 20386.15 分也可以升级为 VIP 贵宾等。

3 企业客户价值预测模型及应用

确定客户的层级后,运用数据挖掘技术建立客户价值预测模型。具体方法是运用会员客户资料的自然属性(如性别、年龄、职业、受教育程度、收入等)以及所属的会员层级,通过决策树方法建立客户价值预测模型。决策树一般是自上而下的来生成的。每个决策或事件(即自然状态)都可能引出两个或多个事件,导致不同的结果,把这种决策分支画成图形很像一棵树的枝干,故称决策树^[8]。本文根据所拥有的实际数据情况,将每个会员所属的会员层级作为目标分类类别,根据客户的性别、年龄、职业等自然属性,为每个类别挖掘出分类规则,然后用这个分类规则预测新客户或潜在客户所属的类别,从而发现潜在客户的价值,使其成为企业的忠实客户,提高企业盈利能力。考虑到 SIP 超级贵宾比例较少,如果取全部数据进行训练,必将很难得到关于 SIP 超级贵宾的规则,因此,文本从三个层级会员中,各随机抽取 100 人(共 300 人)来训练决策树,得到结果,见图 3。

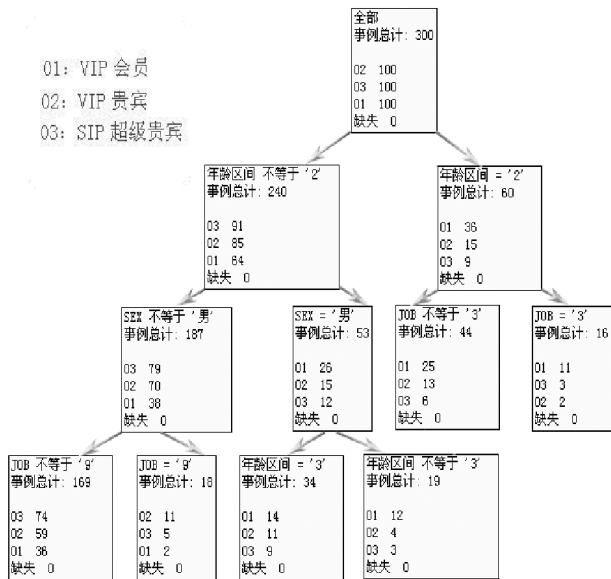


图 3 决策树

说明:

- (1)年龄区间表示的是会员的年龄范围,1 表示小于 20 岁,2 表示 20-30 之间,以此类推。
- (2)JOB 表示的是会员的职业,取值范围 1-16,3 代表学生,9 代表教师。

从图 3 中,本文得出相应有关 SIP 超级贵宾和 VIP 贵宾会员的规则。如:

规则 1:年龄不在 20-30 之间、性别为“女”并且职业不是教师,则该会员是 SIP 超级贵宾。

规则 2:年龄不在 20-30 之间,性别为“女”并且职业是教师,则该会员是 VIP 贵宾。

另外随机选取 150 个会员数据(每个层级各 50 人)进行验证,验证结果见表 2。

表 2 验证结果

预测	01(实际)	02(实际)	03(实际)
01	27	19	13
02	1	3	4
03	22	28	33

预测的准确度=(27+3+33)/150=42%

根据所得的规则,企业可以对新会员或潜在会员进行价值预测,并且根据他们可能成为的会员类型对他们采取不同的营销策略。

4 结 语

本文从某零售企业的实际数据出发,针对该企业目前会员层级比例不合理的现象,结合会员消费积分人数分布图,通过刻度设定的方法计算出每个层级会员的比例,并以此为基础根据会员的自然属性,通过决策树的方法得出客户价值判断规则,该规则为预测新客户或潜在客户的价值奠定了基础。期望这种预测活动在提供高质量的客户服务支持的同时,为企业留住更多的客户并创造更多的利润。

参考文献

1 杜朋朋. 基于数据仓库与数据挖掘的应用研究 [J]. 电子商务,2008,(4):56-60.

2 徐成杰. 个性化营销研究 [J]. 中国民营科技与经济, 2007, (1):75-76.

3 王尔平, 崔宝灵. 数据挖掘技术在电信 CRM 中的应用[J]. 管理科学, 2003, 16(5):54-58.

4 刘 涛, 吴雪亮. 数据挖掘在企业决策中的应用[J]. 中国高新技术企业,2008,(6): 41-42

5 陈京民. 数据仓库与数据挖掘技术[M]. 北京:电子工业出版社,2007:29-31.

6 S.C.Hui, G.Jha. Data mining for customer service support[J]. Information & Management, 2000,38(1):1-13.

7 王 欣,徐腾飞,唐连章. SQL Server 2005 数据挖掘实例分析[M]. 北京:中国水利水电出版社, 2008: 157-158.

8 廖开际. 数据仓库与数据挖掘[M]. 北京:北京大学出版社, 2008: 167-173.

(责任编辑:徐 波)