文章编号: 1008-8245(2008)05-0016-04

商业智能在旅游产品销售中的应用分析

余志毅 赵 青 冯运仿(黄石理工学院,湖北黄石 435003)

摘 要: 利用商业智能软件的决策树分类技术,对旅游销售企业的调查问卷数据进行分析,挖掘出最近一年内有出境旅游意愿的客户特征,从而提高营销的效率。

关键词:旅游产品:商业智能:数据挖掘

中图分类号: TP274⁺. 2 文献标识码: A

Analysis of Business Intelligence in Tourism Product Sale

YU Zhiyi ZHAO Qing FENG Yunfang

 $(\,Huangshi\,Institute\,of\,Technology\,\,Huangshi\,Hubei 435003\,)$

Abstract Using the intelligent business division and strategy technologies, this paper gave a deniled analysis of the data gathered from the survey of Travel Industries and revealed the characteristics of the tourists who had the intention of travelling abroad in the past year which enhanced the marketing efficiency.

Key words, pour im products, business Intelligence data mining

0 引言

Garme商业智能峰会重定义了商业智能,视商业智能为一个伞状的概念,它包括了分析应用、基础架构和平台以及良好的实践。此举进一步引导商业智能实现三个转折:从数据驱动转向业务驱动、从关注技术转向关注应用、从关注工具转向关注工具产生的绩效。其实质是商业智能以业务为驱动,根据商业策略及其所需要的分析以运用数据。认识的转变对技术的研发有着重大的指导作用。一个企业如果要生存和发展,就必需了解市场,了解客户,树立"以市场为导向,以客户为中心"的经营理念。旅游产品销售企业作为服务中心"的经营理念。旅游产品销售企业作为服务中心"的经营理念。旅游产品销售企业作为服务中心"的经营理念。旅游产品销售企业作为服务中心"的经营理念。旅游产品销售企业作为服务,对方等到股票,

统等不失为有效途径。而数据挖掘技术在客户分析方面有着较大优势。在营销过程中市场调查是十分重要的,市场调查将为营销策略,广告的投放提供依据,面对市场调查取得的大量纷杂、无序的数据,利用商业智能软件的数据挖掘技术进行处理,将会起到良好的结果。

1 商业智能软件

商业智能(BJ Business Intellgence),又称商务智能,就是一种将数据转变为信息、信息转变成知识的工具,并且这种工具能够在恰当的时候通过恰当的方式把恰当的信息传递给恰当的人。商业智能定义为下列软件工具的集合。

(1)简单的查询和报告工具。专门用来支持初级用户的原始数据访问,不包括适用于专业人

收稿日期: 2008-04-24

^{*}科研项目: 黄石市 2007年度科研计划项目([2007]14)。

作者简介: 余志毅(1982—) , 男, 湖北蕲春人, 讲师, 硕士生。

士的成品报告生成工具。在这一层次,商业智能仅仅是把信息进行粗加工。

- (2)在线分析处理。提供多维数据管理环境,其典型应用是对商业问题的建模与商业数据分析。
- (3) 经理信息系统。这类系统的用户希望能够在不太费力的情况下,从系统中获取大多数信息。
- (4)数据集市和数据仓库产品。包括数据清洗、数据抽取、转换、载入、数据管理和数据存取等方面的软件。
- (5)数据挖掘(Data Mining)软件。使用诸如决策树、神经网络、规则归纳等技术,用来发现数据之间的关系,做出基于数据的推断。数据挖掘是通过仔细分析大量数据来揭示有意义的新的关系、模式和趋势的过程。它使用模式认知技术、统计技术和数学技术[1]。数据挖掘的目的是为决策建模,即根据过去活动的分析预测将来的行为。比较常见的数据挖掘算法有:聚类分析、决策树、神经网络、规则归纳。一般比较好的数据挖掘工具都会支持这几种算法[2]。

2 决策树与数据挖掘工具

决策树方法起源于概念学习系统,然后发展了 D3 方法并达到高峰,最后又演化为能处理连续属性的 C4.5。决策树构造的输入是一组带有类别标记的数据,构造的结果是一棵二叉或多叉树。二叉树的内部节点(非叶子节点)一般表示为一个逻辑判断,如形式为(ai=vi)的逻辑判断,其中 ai是属性,vi是该属性的某个属性值;树的边是逻辑判断的分支结果。多叉树的内部节点是属性,边是该属性的所有取值,有几个属性值,就有几条边。树的

叶子节点都是类别标记。构造决策树的方法是采 用自上而下的递归构造。这种思路实际上就是"分 而治之"(Divide—and—conquer)的道理。二叉树的 原理与此的差别仅在于要选择一个好的逻辑判断。 在生成的决策树中可以建立一个规则基。一个规 则基包含一组规则,每一条规则对应决策树的一条 不同路径, 这条路径代表它经过节点所表示的条件 的一条连接[3]。例如市场上常见的商业智能软件 IM的 Intelligent Miner是市场上最强大和最有可伸 缩性的工具之一,正在竞争数据挖掘工具市场的领 导地位,它提供了最广泛的数据挖掘技术和算法, 在数据规模和计算性能方面具有非常高的可伸缩 性: Intelligent Miner支持分类、预测、关联规则、聚 类、顺序模式侦测和时间序列分析的算法。 Inten lie gent Miner支持 DB2 关系数据库管理系统, 并集成 了大量复杂的数据操纵函数。根据 DC的统计,ILLL telligent Miner目前是数据挖掘领域最先进的产品。 大多数算法是由 BM研究所研发的,是 BM的专有 技术,并只存在于 Intelligent Miner中。决策树使用 的是 CAU算法的二种变种, 用以产生一个分类模 型并且能够处理离散和连续数据[4]。本文决策树 分析主要依托的就是 IM的 Intelligent Miner

3 决策树技术在旅游产品销售中 的应用

下面将就某旅游销售公司的调查问卷进行分析。潜在客户的数据主要有年龄、性别、职业、收入、教育程度、是否结婚,是否曾经出境旅游等,调查客户是否会在一年内出境旅游,数据表格式如表1所示。

表 1 吴旅游销售公司的潜任							
序号	性别	年龄	收入	职业	学历	婚否	是否出境旅游
1	男	38	4 000	金融	本科及以上	是	否
2	男	40	3 000	金融	大专	是	是
3	女	31	6 000	电信	本科及以上	是	否
4	男	29	10 000	${ m I\!\Gamma}$	本科及以上	否	否
5	男	31	4 000	建筑	本科及以上	是	是
6	女	32	4 000	证券	本科及以上	是	是
7	女	31	3 000	金融	本科及以上	是	是
8	女	40	3 000	公务员	大专	是	是
9	男	64	1 500	公务员	中专	是	是

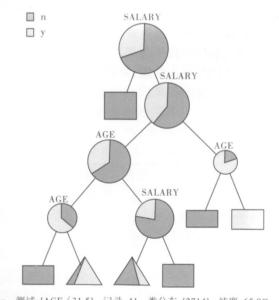
表 1 某旅游销售公司的潜在顾客调查情况表

3.1 数据预处理

数据预处理的目的是为了提高数据质量,使 数据挖掘的过程更加有效,更加容易,同时也提高 挖掘结果的质量。数据预处理的对象主要是清理 其中的噪声数据、空缺数据和不一致数据。常用 的数据预处理技术主要包括:数据清洗、相关分析 和数据变换等。数据清洗试图填补训练集中的空 缺值、识别孤立点、消除噪声、纠正数据中的不一 致。对于空缺值的处理,通常有忽略元组、人工填 写空缺值、使用全局常量填充、使用属性平均值填 充、使用与给定元组同一类的样本平均值填充、使 用最可能的值填充等方法[5]。调查问卷中,有部 分没有填写的选项,如年龄属性,对于这一部分记 录采用的是使用属性平均值填充的方法。在进行 数据挖掘工作之前,需要进行前期的数据整理工 作,比如根据直观经验去除数据中的冗余信息或 不相关信息,对于上面的数据集中的属性,像序号 等,可以在正式开始数据挖掘前去除,因为客户是 否近期内出境旅游是我们最关心的信息,我们把 属性 "是否一年内出境旅游"即 buy作为类标签属 性。

3.2 生成决策树,产生规则

整理后的数据导入到 DB 关系数据库表中,使用 EM的 Intenligent Mine提供的数据挖掘工具生成决策树,并剪枝后如图 1 所示^[6]。



标签=n,测试=[AGE (31.5],记录=41,类分布=[2714],纯度=65.9%

图 1 由训练集生成并剪枝的决策树图

在得到决策树之后,可以由其中提取分类规则,在该例中,可以提取的规则如下, If Salary =

2 750 ~6 500 and age=31. 5 ~40. 5 then buy y

也就是说,在剪去一些噪声枝节之后,在决策树的每一条支路上,都可以形成一条分类规则。可以采用这些分类规则,对潜在的客户数据进行分类,由此得出哪些客户最近有出境旅游的意愿,然后可以主动地向客户推销旅游产品,并且给予一定的优惠政策,由被动营销转变为主动营销。

3.3 决策树结果分析理解

需要说明的是这 203 份问卷是在平时来旅游公司门店咨询中获取的,来店咨询并接受问卷调查的自然多数是有出境旅游想法的,因此 31. 2%愿意出境旅游,这个比例在普通人群中是达不到的。下面我们从产生的决策树规则分析一下是否在一年内出境旅游与客户的因素之间的关系。

- (1)收入直接决定了一个人的购买力。 salary 即在月薪低于 2 550元的客户中,出境旅游是奢侈品,在近期一年内没有出境旅游的意愿;在月薪高于 6 500元的客户中,一年内也没有出境旅游的意愿,可以想象高收入人群大多已经有出境旅游的经历了。
- (2)年龄也是导致出境旅游的一个因素。接受调查的人都在 22岁~65岁之间。年轻人中愿意出境旅游的较多,年龄小于 31.5岁的人,可能是由于婚姻、购房等大额消费所困,一年内没有出境旅游的意愿;年龄在 31.5岁~40.5岁之间(占82.3%),一般来说这一部分高收入人群不受经济状况所困扰,会考虑出境旅游,因此有近期出境旅游的打算。
- (3)愿意出境旅游的多为男性。在低收入的 女性中没有考虑出境旅游问题,但男性如果收入 还可以的情况下有近期出境旅游的可能性。
- (4)从目前这些数据来看,受教育情况、婚否、工作性质与是否愿意近期出境旅游影响也不大。客户的购买行为还要从多方面进行考虑,我们仅从这几方面挖掘出外部环境与客户的购买行为之间的关系。在计划出境旅游的人群中,出境旅游的价位、旅游目的地、旅游期限等也可以进行挖掘;打算出境旅游的人群中,不愿意出境旅游的原因,也值得分析。

4 结论

采用决策树分类算法,通过对调查数据挖掘,得到一系列的分类规则,然后利用此分类规则,对潜在客户进行分析,采取主动营销,可以降低营销

成本,从而可以提高营销的成功率。随着多方面 大量数据的获得,商业智能的数据挖掘工具可以 挖掘出更有参考价值、易于理解,并具有很高的分 类准确度的规则为生产实践服务。决策树数据挖 掘技术在旅游产品销售中也有着广阔的应用前 景,值得我们进一步的研究。

参考文献

[1] 潘珩. 汽车营销决策支持系统数据仓库的设计与实现[引. 石河子大学学报: 自然科学版, 2005(5), 89

- [2] 耿晓中,张冬梅. 数据挖掘综述[3]. 长春师范学院 学报, 2006(6): 34—35
- [3] 罗海蛟, 刘显. 数据挖掘中分类算法的研究及其应用[1]. 微机发展, 2003(6), 24
- [4] 郑晓红. 商业智能提升管理信息化 [J]. 商场现代化, 2005(12): 144-145
- [5] 谭旭, 王丽珍, 卓明. 利用决策树挖掘分类规则的算法研究[J]. 云南大学学报, 2002(6); 415—419
- [6] 纪鹏, 马林. 基于信任管理模型的对比分析 [3]. 黄石理工学院学报, 2007(3): 20-22

(责任编辑 梅军进)

(上接第 12页) 根据客户提供的时间及站点数据名,利用数据库视图的逻辑无关性,建立历史数据访问视图,提供给客户端应用软件访问。通过设置中间件来实现进程间通信并提供位置透明性,而且还利用高效可靠的消息传递机制进行平台无关的数据异步传输,并基于数据通信来进行分布式系统的集成。

分布式审批系统在历史数据访问视图的建立 上使用了分布式数据库的水平分片技术,通过采 用关系代数中的连接操作来重构信息访问视图。 客户端的应用软件只要对信息访问视图进行操作 即可,从而实现了系统的逻辑无关性,不增加应用 软件的任何工作量。此架构从根本上大大减轻了 数据服务器的数据量,保证了数据服务器的长期 快速运行,同时也提供简便快捷的历史信息访问 方法。

5 结束语

本文分析了国内现有的行政审批系统仅仅依

托于一个办事大厅、彼此孤立、效率低下这一局限,从实现整个行政审批领域统一的行政审批系统的角度出发,构建了一个功能比较全面的、开放的、统一的行政审批系统原型,为下一步的实际开发应用奠定了基础。

参考文献

- [1] Andrew S Tanenbaum, Maarten van Steen 分布式系统原理与范例 [M]. 杨剑峰, 常晓波, 李敏, 译. 北京: 清华大学出版社, 2006
- [2] 邵佩英. 分布式数据库系统及其应用[M]. 北京: 科 学出版社, 2005
- [3] 杨艺. 分布式数据库中数据分配方法的研究[D]. [硕士学位论文]. 重庆: 重庆大学, 2004
- [4] 马淑娇. 2^H.并发控制的研究与实现探析[¹]. 计算机应用研究。2003 12(2): 38-40
- [5] 田嵩,梅军进,利用网络技术实现医院信息共享的设计[J].黄石理工学院学报,2007(5):35-37

(责任编辑 梅军进)