

数据挖掘在客户行为细分中的应用

摘 要

电信改革进一步深入,各大电信运营商的实力正在渐渐接近,通信市场的竞争已经十分激烈。随着经济的发展和社会的进步,人们对通信市场消费需求的多层次、多元化、个性化的特点日益突出。因此如何对客户进行准确的细分,提供有针对性的服务,已经成为了通信运营商市场经营分析的首要工作。同时,随着信息技术的发展,大规模数据处理技术的推广和应用为复杂的客户细分提供了可能。数据挖掘技术的发展,给客户细分研究带来了前所未有的突破。

本文从目前移动通信市场的竞争分析入手,提出市场细分是通信企业发展和业务创新的新出路。通过对经典市场细分理论和数据挖掘技术应用的总结,为通信企业的市场细分提供了很好的理论支持和技术支持。面对消费者需求的多元化趋势,经验表明结构化的绝对的细分研究并不足以帮助营销人员精确的掌握市场信息。针对通信企业拥有大量的客户消费数据的现状,本文指出以客户历史消费行为数据为基础的 client 行为细分是通信企业准确的掌握客户消费取向的一个非常有效的途径,并根据营销工作的需要提出了行为驱动的客户细分模型。最后,应用客户细分模型,使用 IBM Intelligent Miner 数据挖掘软件提供的聚类分析技术,完成了移动长途话音业务的客户细分,并对模型和算法做出了评价。

关键词: 移动通信 客户细分 行为细分 数据挖掘 聚类

APPLICATION OF DATA MINING IN SEGMENTATION BASED ON CUSTOMER BEHAVIOR

ABSTRACT

Since the telecommunications reform, the strength of every great telecommunications enterprises is close gradually, and the competition among Chinese mobile operators becomes fiercer than ever. On the other words, with the development of economic, consumers are becoming more and more personal and picky. Telecom business becomes more complex than before, where a key success factor is to know customers. It is customer segmentation that becomes the primary task in business analysis of telecommunications enterprises.

With the development of the information technology, the new data processing techniques allow more accurate segmentation that is based on real data from data warehouses and surveys. Data mining technologies makes it possible to combine data from different sources and understand relationships of different factors for segments.

This thesis analyses competition in mobile communication and the importance of segmentation at first. On the basis of brief introduction of traditional segmentation models and data mining technology, the thesis proposes that segmentation based on behavior is a key to know customers. On the basis of former chapters, a segmentation of mobile long call service is completed using IBM Intelligent Miner in the last chapter.

**KEY WORDS: Mobile Communication, Customer Segmentation,
Behavioral Segmentation, Data Mining, Cluster**

独创性（或创新性）声明

本人声明所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：张悦 日期：04.3.1

关于论文使用授权的说明

学位论文作者完全了解北京邮电大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属北京邮电大学。学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。（保密的学位论文在解密后遵守此规定）

保密论文注释：本学位论文属于保密在__年解密后适用本授权书。非保密论文注释：本学位论文不属于保密范围，适用本授权书。

本人签名：张悦 日期：04.3.1

导师签名：林方宁 日期：04.3.1

第一章 绪 论

1.1 竞争中的移动通信市场

1.1.1 移动通信市场发展迅速

我国的移动电话通信是在八十年代末从国外引进的现代通信工具。由于国家对信息产业的政策倾斜及扶持,中国移动通信在不到 10 年的时间内,迅速发展壮大起来。

据信息产业部统计,截至 2003 年 9 月底,我国移动交换机年增 5000 多万门,达到 3 亿多门,成为世界移动通信第一大网;移动通信用户年增 6000 万户,达到 2.5 亿户,成为世界第一。到 2003 年 10 月底,我国移动电话用户首次超过固定电话用户。到 2003 年 11 月,我国的移动电话用户和固定电话用户分别达到 26347.8 万户和 25984.2 万户,移动电话用户总数连续二个月超过固定电话用户总数,这标志着我国电信运营业的发展迎来了一个新的里程碑。

随着我国经济的发展,地区之间交流进一步加强,人的移动性大大提高,移动电话成为人们日常沟通和社会交往的重要手段。其个性化和无处不在的特性在电信市场的竞争中显示出明显优势。移动通信的有效需求持续加大,将固定电话务量大量分流。今后一个时期,我国移动电话仍有较大发展空间,移动电话用户超过固定电话用户是必然趋势。2004 年及未来五年中,移动通信业务将继续保持较快的增长,但增长率将逐渐放缓。按照信息产业部的规划,2004 年移动通信业务收入将达到 3510 亿元,占到全国电信业务收入的 48.4%,移动电话用户新增 5200 万户,移动电话普及率将达到 24.5%。

1.1.2 通信运营商间竞争加剧

我国电信运营业的发展模式由单纯的扩大投资转向创新业务、优质服务和追求效益上,经营格局逐步由垄断转向竞争与开放。继 1998 年中国电信“一分为四”的体制改革完成之后,2002 年 5 月,国务院又批准了中国电信“南北拆分”的重组方案,以中国电信、中国移动、中国网通、中国联通、中国卫星以及铁通为主要经营主体,形成了“5+1”的竞争新格局。随着电信改革的进一步深入,电信市场上已经形成了诸侯割据、群雄争霸的局面,各大电信运营商的实力正在渐渐接近。目前我国除了前面 6 家基础电信运营商外,还有超过 4400 家的增值

电信运营商，形成不同规模、不同业务、不同所有制企业间共同发展和相互竞争的格局。在这个新格局中，中国移动以业务收入最大成为市场老大，中国电信退居第二。同时，重组后的新网通实力也大大加强，进入市场份额前三甲。

截至 2003 年 6 月底，各大基础电信运营企业所占业务收入的市场份额分别是：中国移动居首位，占 37.4%；中国电信占 31.1%；中国网通 16.6%；中国联通 13.4%；中国卫星和中国铁通为 1%。从图 1-1 中我们可以看出，没有一家在竞争中处于绝对优势的地位。

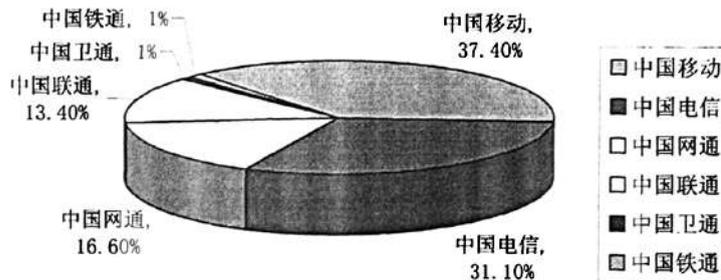


图 1-1 2003 年 1-10 月电信业务收入比例
(数据来源: CCID)

随着我国电信业向做大、做强的目标发展，以及相应电信体制改革的深化，各大电信运营企业将会进行新一轮的重组和整合，市场竞争格局将更激烈，新兴电信企业的市场份额将逐步上升，在固定电话、移动通信、数据通信领域将出现新一轮的有效竞争和市场整合。

1.2 市场细分是通信企业发展的新出路

通信市场竞争加剧的同时，人们消费水平也正在不断的提高，消费观念发生了很大转变，电信市场消费需求的多元化、个性化、多层次的特点日益突出。我国电信业已经进入一个新的时期，传统的电信业务已经愈来愈无法满足现代消费的需求，而对市场进行准确的细分，根据不同用户群体的需求提供创新性的服务，已经成为了通信业发展的一个趋势。

1.2.1 市场细分是发现市场需求的途径

随着市场竞争的加剧,广大消费者也越来越多地抱怨通信运营商提供的服务质量差强人意。企业感到经营的压力正在不断加大。究其原因除了竞争加剧之外,还有一个重要原因就是对市场细分不够,没有了解特定客户群体的特定需求,进行分级服务。

长期以来,我国的通信运营商往往只把客户群细分为大客户和普通客户两大类,只重视做好大客户的营销服务工作。但是抛开经济、人口、心理、消费行为等具体因素的差异,只把客户机械地划分为这么简单的两大类了是远远不够的。

电信运营商有必要将市场细分微型化,对通信市场的细分不仅要体现在横向的企业、政府和个体用户层面上,还要对这三个客户群进行纵向的细分。因为政府的级别,企业的规模、所属行业以及个体消费者在年龄、收入水平、文化水平等都直接影响到它们的消费能力和要求。这样才能最大程度的发现客户的不同需求,找到更大的市场空间。

1.2.2 市场细分是通信业务创新的基础

随着经济的发展,客户的消费需求将进一步个性化,只有对电信市场进行立体化深入的细分,才能进一步的发现差异化的服务需求,通信企业才能在此基础上不断地进行业务创新。

业务创新直接目的就是为了提升服务,针对不同的客户群体,实现服务的差异化、业务的个性化。而要实现差异化的个性服务首要的基础就是要对市场进行准确的细分。市场细分能够准确地发现客户的特殊需求和消费水平,有助于通信企业针对特定客户群进行业务创新,也有利于企业在业务创新时根据细分群体的消费水平更有效的进行成本控制。这样既提高了企业的盈利,又避免因盲目投资而造成的资源浪费。因此,从这个意义来讲,通信运营商细分市场的能力决定了业务创新的成功与否。

1.2.3 市场细分是提高客户满意度的前提

在市场竞争的日益激烈今天,“以顾客为中心”的服务理念已经深入人心,顾客的满意度是衡量一个通信企业的竞争实力的重要指标。如何才能达到客户的最大满意?显然仅仅依赖于微笑的服务是不够的,让客户满意的前提就是要理解客户。客户消费行为的研究和市场调研,可以帮助企业发现不同客户消费需求上的差异,对客户群进行准确的细分,从而提供针对性的优质服务。所以说,通信

运营企业细分市场的能力决定了它对客户需求的理解程度,也决定了它能否真正做到以客户为中心,提升客户满意度的能力。

1.2.4 市场细分是提高企业竞争力的手段

根据深入的市场调查进行的市场细分,会更容易帮助企业发现尚未被满足的顾客群体,并根据这一客户群体的需求特征设计出独具特色的通信服务。市场细分有助于企业挖掘竞争对手尚未发现的市场领域,并通过产品的差异化建立起竞争优势,从而提升自己的竞争力。

1.2.5 市场细分是协同合作的要求

面对市场需求的深刻变化,运营商已经发现传统的经营模式已经无法适应市场的转变,仅仅依靠自身的力量已经无法为用户提供个性化的信息服务。运营商必须在市场的发展中重新定位自己,树立一个“利益共享”的观念,寻求各方力量的合作,才能为用户提供个性化的服务。“产业链”概念的提出预示着我国的电信运营商已经深刻意识到了市场的变化,并能够更加理性和客观的看待其在市场中的定位。只有联合内容服务商、制造商、终端商、经销商,打造一个完整的产业链,才能实现一个新业务的持续发展,才能向客户提供完整的服务。产业链中每个内容服务商所对应的都是一个消费个性化的细分客户群体。从这个意义上讲,运营商在打造产业价值链时,也需要完成对客户的深入理解和市场的深入细分工作。

总之,随着社会经济的进一步发展,通信运营商再也无法创新出什么一劳永逸的新业务,只有在细分市场、细分客户、细分成本的基础上,算好边际投入、边际产出账,考虑到价值链、效益链、服务链,才能在未来的市场竞争中站稳脚跟。

1.3 论文概述

本文着眼于数据挖掘技术和市场细分理论相结合,研究移动通信客户消费行为的特殊性和市场营销工作的重点,指出以客户历史消费行为数据为基础的客户行为细分是通信企业准确的掌握客户消费取向的一个非常有效的途径。本文提出了适宜移动通信行业的行为驱动的三维客户细分模型,并且在实际工作中对模型进行了应用和评价。

本文分五部分进行阐述。第一章介绍了论文提出的背景和研究的意义。通过分析目前移动通信的竞争现状,指出市场细分是通信企业发展和业务创新的新出路。第二章本文总结了经典的市场细分理论,着重介绍了消费者市场细分的依据和基本程序,为移动客户细分提供了理论基础。在第三章,本文介绍了数据挖掘技术的发展以及在市场细分中的应用,为客户细分提供了技术支撑。第四章,根据移动通信客户消费的特点,提出了客户行为数据可以作为移动客户细分的一个有效出发点。结合移动通信企业掌握的客户信息,充分利用移动客户的多种属性,提出了以行为驱动的行为——特征——价值三维客户细分模型。第五章,将该细分模型应用到实际分析工作中,使用 IBM Intelligent Miner 数据挖掘软件提供的聚类分析技术,完成了移动长途话音业务的客户细分,并对模型和算法做出了评价。

本文应用举例中的数据来自某省 2003 年初某月的客户数据,因此模型的应用具有一定的现实意义。

第二章 经典市场细分研究体系

2.1 市场细分定义

市场上的客户和需求是千差万别的。任何企业都很难只凭自身的力量提供能够同时满足所有客户需要的产品。这不仅是由企业自身条件的限制，从整个经营效果来看也是不足取的。因此，企业必须知道哪些客户对自己来讲是最有价值的，他们的具体需求是什么。针对客户的特点提供独特的产品和行之有效的营销服务，才能获得最佳收益。很多企业的成功案例告诉我们，任何市场需求的背后都隐藏着可以被进一步明确再分的潜力和可能。

20 世纪 50 年代中期，美国市场营销学家温德尔史密斯（Wendell R. Smith）最先提出了“市场细分”的概念。

所谓市场细分（Segmentation），就是根据消费者需求的不同特征，把市场划分为若干个消费者群体的市场分类过程。具体说来，就是调查分析不同的消费者在需求、资源、地理位置、购买习惯和行为等方面的差别，然后将上述要求基本相同的消费者群体分别归并为一类，形成总体市场中的若干子市场或细分市场。不同的细分市场之间需求差别比较明显；而在每一个细分市场内部，需求差别就比较微小，基本倾向一致。

市场细分的理论基础是市场多元异质性理论。这一理论认为，消费者对大部分产品的需求是多元化的，具有不同质的要求。需求本身的异质性市场可能细分的客观基础。因此细分市场是有一定客观条件的。只有商品经济发展到一定阶段，市场上商品供过于求，消费者需求多种多样，企业无法用大批量生产产品的方式或差异化产品策略有效地满足所有消费者需要的时候，细分市场的客观条件才具备。

市场细分从其本质上看，就是将具有相同特征的客户分为一组，而将具有其他相同特征的客户分为另一组。企业对市场进行细分的目的就是服务于企业营销战略的制定与实施。细分对于制定市场营销战略有两方面的重要意义：

首先，有助于企业更准确地了解某购买群体的需要、想法和行为；

其次，资源可以更有效的配置于为满足细分市场客户需求而开展的营销组合活动中。

具体而言，市场细分的价值体现在企业在经过细分研究的市场上采取相应的营销努力可以获得销售额增长和市场份额的增长（Genschet al., 1990）。

2.2 市场细分的原则

企业可根据单一因素，亦可根据多个因素对市场进行细分。选用的细分标准越多，相应的子市场也就越多，每一子市场的容量相应就越小。相反，选用的细分标准越小，子市场就越少，每一子市场的容量则相对较大。如何寻找合适的细分标准，对市场进行有效细分，在营销实践中并非易事。一般而言，成功、有效的市场细分应遵循以下基本原则：

2.2.1 可衡量性

可衡量性是指细分的市场是可以识别和衡量的，亦即细分出来的市场不仅范围明确，而且对其容量大小也能大致做出判断。有些细分变量，如具有“依赖心理”的青年人，在实际中是很难测量的，以此为依据细分市场就不一定有意义。

2.2.2 可进入性

可进入性指细分出来的市场应是企业营销活动能够抵达的，也就是说企业通过努力能够使产品进入并对顾客施加影响的市场。一方面，有关产品的信息能够通过一定媒体顺利传递给该市场的大多数消费者；另一方面，企业在一定时期内有可能将产品通过一定的分销渠道运送到该市场。否则，该细分市场的价值就不大。

2.2.3 效益性

有效性指细分市场的容量或规模要大到足以使企业获利。进行市场细分时，企业必须考虑细分市场上顾客的数量，以及他们的购买能力和购买产品的频率。如果细分市场的规模过小，市场容量太小，细分工作烦琐，成本耗费大，获利小，就不值得去细分。

2.2.4 差异性

差异性指各不同分市场的消费者对同一市场营销组合方案会有差异性反应，或者说对营销组合方案的变动，不同细分市场会有不同的反应。如果不同细分市场顾客对产品需求差异不大，行为上的同质性远大于其异质性，此时，企业就不必费力对市场进行细分。另一方面，对于细分出来的市场，企业应当分别制定出独立的营销方案。如果无法制定出这样的方案，或其中某几个细分市场对是否采

用不同的营销方案不会有大的差异性反应，便不必进行市场细分。

2.3 消费者市场细分的依据

市场细分的指标和变量很多。一般情况下，在我们在进行消费者市场细分时的主要依据为以下四类：

2.3.1 客户基本特征

客户基本特征指标是指描述客户类别特征的指标。我们把地理因素、人口统计因素和社会经济因素称为客户的基本特征指标。

2.3.1.1 地理因素

地理因素是指消费者所处的地理位置、自然环境和社会环境，包括地域（如国家、地区、南方、北方、城市、乡村）、地形、气候、人口密度、生产力布局、交通运输和通讯条件等。地理细分之所以可行，主要是由于自然气候、传统文化、经济发展水平等因素的影响，处于不同地理环境下的消费者形成了不同的消费习惯和偏好，具有不同的需求特点，他们对产品、价格、销售渠道、广告宣传等营销组合的反应也常常存在差异。

地理因素易于辨别和分析，因此一般是人们进行市场细分的首选指标。但是地理因素都是一种相对比较静态的因素，处于同一地理位置的消费者仍然会存在很大的需求差异，因此企业要选择目标市场，还必须同时依据其他因素进一步细分市场。

2.3.1.2 人口因素

按照人口统计因素来细分市场称为人口细分。这方面的具体变量很多，包括年龄、性别、职业、收入、教育、家庭人口、家庭生命周期、国籍、民族、宗教、社会阶层等。很明显，这些人口变量与需求差异之间存在着密切的因果关系。不同年龄组、不同文化水平的消费者，有不同的生活情趣、消费方式、审美观和产品价值观，因而对同一产品必定产生不同的消费需求；而经济收入的高低不同，则会影响人们对某一产品在质量、档次等方面的要求差异，如此等等。因此，依据人口变量来细分市场历来为人们所普遍重视。

2.3.2 客户心理特征

心理特征指标用来描述客户的个性特征，这些特征与客户对产品、包装、广告、公关行为的反应密切相关。它覆盖了从自我认识、生活方式到对产品的态度、兴趣和观点以及产品属性等一系列广阔的范畴。研究消费者心理的核心之处在于把个性数据与动机和行为数据相结合，有助于企业了解客户的行为特点。

与客户基本特征不同，客户心理特征是内在的、非显性的生理和情感反应。由于人们无法进入客户的内心世界去观察，所以我们在运用心理因素对客户进行细分时，必须借助心理学中的一些分析指标。如表 2-1 所示，心理指标主要包括生活方式、个性、社会阶层、社会风格等几方面。

表 2-1 心理行为细分指标

生活方式	朴素型、追求时髦型、大众型
社会阶层	下层、中层、上层、上上层
个性	保守或激进、内向或外向、独立或依赖
社会风格	剖析型、表现型、驱动型、和藹型

➤ 生活方式

生活方式是一个人在世界上所表现出的活动、兴趣和看法的集中体现，它对消费需求有着深刻的影响。它不仅决定消费者参加什么样的活动，而且还决定他们购买什么品牌的产品或服务。

➤ 社会阶层

社会阶层是指具有相似社会经济地位、价值观念和生活方式的人们组成的群体。社会阶层是在一个社会中具有相对的同质性和持久性的群体，他们是按等级排列的，每一阶层成员具有类似的价值观、兴趣爱好和行为方式。

➤ 个性

个性是一个人所具有的心理特征，它导致了一个人对它所处的环境的相对一致和持续不断的反应。我们可以用很多性格特征来描述一个人的个性，如内向、外向、开拓、保守、文静、急躁等等。

➤ 社会风格

主要影响人们如何对工作或工作之外的刺激做出反应。这个指标主要看两个方面：自信和反应性。自信是一个人把握地陈述观点、信心、力度以及试图引导他人行动的程度；反应性一个人接受外界诱惑、影响或刺激的程度以及在

他人面前如何表现自己的感受、情绪和印象。

客户心理特征是市场细分中比较复杂的一类指标，必须对消费者不同的心理进行充分的调查研究，才能获得可靠的数据。在进行客户心理细分时也可以聘请一些外部专家来担当客户心理分析的角色，他们的经验能够提高心理细分的准确性，弥补营销人员的不足。

2.3.3 客户行为特征

客户的购买行为是与购买商品有关的各种活动，如收集信息、比较评估、购买和购买后的有关活动等。由于所追求的利益和经济收入等因素的不同，客户购买及使用某一产品的程度通常会表现出很大的不同。由于客户在使用程度上通常会存在差异，我们就可以按照这一特征对客户进行区分。

客户行为特征是指描述客户的购买和使用商品的情况的一系列指标，如购买动机、购买地点、使用状况、使用频率等。

➤ 购买时机

根据消费者产生需要、购买或者使用产品、服务的时机，可以将整个市场划分为若干具有不同需求特点的子市场。例如，每天购买使用的时间（上午、下午、晚上）、每年中使用的时间不同（春、夏、秋、冬）。又如，春节、中秋节等假日前后。

➤ 购买地点

这类变量是假定客户在不同购买或使用场合会产生不同的行动。例如，购买者可能在办公室里是个保守、实用主义的购买者，但在家里却肆无忌惮。

➤ 使用状况

许多产品可以按照使用状况将消费者分为从未使用者、曾经使用者、潜在使用者、首次使用者、经常使用者五中类型，即五个细分市场。

对于使用状况不同的客户，不同的企业将有不同的营销策略。一般来说，资金雄厚、市场占有率较高的企业应该特别注重吸引潜在使用者的购买，此时企业可以通过各种营销策略组合，把潜在的使用者转变为首次使用者，进而再努力使他们成为本公司产品或服务的经常使用者。而对于一些中小公司，特别是那些无力开展大规模市场营销活动的小公司来说，应该设法把市场领袖不重视的、散落于市场各处、无特别品牌偏好的经常使用者争夺过来。

➤ 使用频率

根据消费者对特定商品或服务的使用次数和数量,可以将消费者划分为大量使用者、中量使用者和少量使用者几个客户群体。大量使用者的数目不多,但其消费量在消费掉的产品中占较大的比重。因此,许多企业自然想以大量使用者作为目标市场。当然,反其道而行之以取得经营上的成功,也是极有可能的。关键在于对大量、中量、少量用户的消费特点和购买行为要有透彻的了解,不仅要推出适宜的产品,在价格、包装、销售渠道、广告宣传等方面也要区别对待,精心安排。

2.3.4 客户价值特征

在现在的营销环境下,企业实施市场细分的动因已不仅仅是要拓展企业经营的触角和确定市场突破的方向,还应当强化企业与高价值客户间的互动关系,其最终的目的是要通过增量销售(up-selling)和交叉销售(cross-selling)的方式,来扩大企业的市场份额和客户份额。这个时候就需要企业辨别出那些高价值的客户。客户价值特征主要包括客户终身价值、忠诚度等。

► 客户终身价值

广义的客户终身价值,是指客户在与企业保持买卖关系的全过程中企业从客户处所获得的全部利润的现值。对现有客户来说,其终身价值可分成两部分,一是历史价值,也就是到目前为止客户给公司带来的净利润;二是客户的未来价值,即客户在将来可能为企业创造的利润。客户的历史价值计算起来比较简单,因此,我们更关注的是客户的未来价值。

客户的未来价值也可以看成由两个部分构成:第一部分为假定客户现行购买行为模式保持不变时,客户未来可以给公司创造的利润。这部分是根据客户关系的当前状态做出的对客户未来价值的一种保守估计。第二部分为假定公司采用更积极的客户保持策略,使客户购买行为模式向着有利于企业利润的方向发展时,客户未来可以给企业带来的利润总和。这部分是对客户增值潜力的一种估计。

► 客户忠诚度

客户忠诚实际上是一种客户行为的持续性,客户忠诚度是指客户忠诚于企业的程度。单就消费者对品牌的忠诚度来看,许多消费者都不同程度的存在着“品牌偏好”这一种购买行为。按照消费这对某品牌产品的忠诚度来细分市场,我们可以把消费者市场划分为3个消费群体:坚定的忠诚者,这类消费者始终不渝的购买某一品牌,在任何时机、任何场合都绝对忠诚于某一品牌;中度的忠诚者,这类消费者一般不只是忠于一种品牌,而是忠于两个或多个品牌,但是除了他们所忠诚的品牌以外,他们不会或很少购买其他品牌的产品或服务;多变者,这类

消费者对任何品牌都无所谓，购买具有很大的随意性。

研究客户的忠诚度对企业的营销很有意义。比如，在“坚定的忠诚者”所占比重较高的市场上，其他公司很难进入，相反，对于那些多变者占比重较大的市场，企业应重新审视原来的品牌定位和目标市场确立等是否准确，以及随着市场环境、竞争环境变化，需要对营销组合的哪些因素重新进行调整。

在市场细分时并不是要求这些指标面面俱到，应该根据细分的目的不同，有针对性地选取，以“常用市场细分指标”为基础结合以往市场研究经验，重新构造细分变量指标。值得一提的是，对中国市场进行细分时，一些不同于欧美的变量尤其值得关注，如地理因素中的“经济发达程度”，人口因素中的“单位性质”等等。这些变量对中国消费者的行为有很大的影响。

2.4 市场细分的基本程序

在进行市场细分时，不同的企业根据产品和市场的不同特点，所采取的程序可能不尽相同，但这些方法中也总有一些关键性的相同点，这些相同点构成了市场细分的基本程序。

市场细分是一个动态的过程，主要可以分为五个阶段：确立细分研究目标、确定细分变量、数据收集、数据分析和细分市场、营销方案制定。如图 2-1 所示。

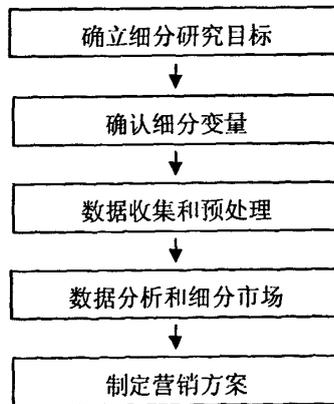


图 2-1 市场细分的基本程序

2.4.1 确立细分研究目标

在进行市场研究开始时，市场研究人员需要确定市场细分的目标：是增加现

有顾客对产品的忠诚度，是吸引新的顾客，还是把客户从竞争对手那里吸引过来；研究是为短期规划服务，还是为长期战略服务；公司管理者和销售者对现有市场结构的看法等等。

这个时候研究人员必须要确定以下几个重要问题：

- 客户对产品或服务介入的程度有多深？
- 客户对这种产品、服务或该行业了解有多深？
他们愿意而且能够认识到何种程度？
- 这是一种新产品还是现有产品？
- 市场细分的目的是什么？
- 市场细分是为短期规划服务还是为长期战略服务？
- 公司管理者对现有市场结构的看法如何？

2.4.2 确认细分变量

在确定了市场细分的研究目的以后，我们面临的问题就是如何选择市场细分变量。这是市场细分研究关键的一步，这一阶段的工作直接决定了市场细分的成功与否。

事实上，细分指标就是为研究人员观察客户提供了各种窗口。通过对这些细分指标的分析，企业可以很容易的对自己所面对的客户形成一个综合的认识：了解企业的客户在哪个区域、了解客户的需求结构和购买力结构、了解客户的人口特征和心理特征等等。只有具备这些认识，企业才能真正的清楚客户需要什么，应该向客户提供什么。

一般而言，细分变量选择必须满足三方面的要求：

- 细分变量应当能够根据客户对营销变量的反应区分出不同的细分市场；
- 细分变量的选择应当与企业实施的营销战略有直接的联系；
- 细分变量可以概括出细分市场的关键差别性特征。

2.4.3 数据收集和预处理

选定了市场细分指标以后，就需要有针对性的采集客户数据。细分不是营销人员根据细分指标进行的主观推断，而是在收集大量客户信息的基础上做出的客

观分析。数据收集的充分性和有效性将对细分结果产生直接的影响，因此数据收集和预处理工作是细分分析的基础。

在进行采掘工作时要依照以下 3 个原则：

➤ 针对性

数据的采集不是漫无目的的，应该根据细分指标和拟定的细分方法来设计一些有针对性的问题；

➤ 系统性

数据收集工作不仅仅为一次细分服务，而且对长期工作有深远影响。因此我们要建立一个完整的信息采集系统，为细分提供有力的信息支撑；

➤ 灵活性

很多企业在细分完成之后，就把采集的信息随意处置，这无疑是对资源的极大浪费。所以，我们主张要把客户信息以正确的方式放在正确的位置，以便营销者可以随时调用。

2.4.4 数据分析和细分市场

在这个阶段营销人员根据采集的数据和市场细分的目的不同选取不同的数据分析方法。

随着营销统计方法的日益精确化以及服务的日益个性化，市场细分在不同情况下常常精确到需要应用多种统计方法。目前的市场细分工作中，简单的单变量细分已经不能满足需要，研究人员越来越倾向于使用一系列描述消费者特征的指标进行市场分割，这样可以全面准确、细致的描述消费者特征，其研究成果相对比较准确、精细。

随着信息技术的发展，大规模数据处理技术的推广和应用为复杂的市场细分提供了可能。数据挖掘技术的发展，给市场细分研究带来了前所未有的突破。在下一章中我们将着重介绍数据挖掘算法在市场细分中的应用。

2.4.5 制定营销方案

完成以上的工作后，我们就已经明确了企业目前所处的细分市场和即将要进入的细分市场，接下来的工作是制定什么样的营销战略来占领市场。

通过分析，可发现有利可图的细分市场往往不止一个，但企业的资源和短期生产能力又是有限的。因此，细分人员应将细分出来的市场按盈利程度排序，从

盈利程度最高的细分市场顺序选择目标市场，直至企业的能力不能在满足为止。然后，就应当有针对性的分别指定市场营销方案，包括产品开发、价格策略、渠道选择、促销政策等，以便有效地进入每一个选定的目标市场。

营销战略的制定除了考虑到运用各种各样的战略策略之外，还应考虑到企业对每个方案的执行能力和执行程度。实际上有很多方案设计都束之高阁，就是因为没有站在企业现实情况的角度上去制定可操作性的有用的营销方案。

第三章 数据挖掘在市场细分中的应用

近十几年,随着计算机和网络的广泛应用,很多企业主动积累的客户数据越来越多,而且通过各种渠道反馈而来的各种市场信息也开始呈爆炸式增长。尽管客户数据是一个很有价值的资源,但要将这些资源用于客户细分并不是件容易的事。如何从大量的数据中挖掘出有用的信息,来分析和区别客户,是营销工作中亟待解决的问题。目前大规模数据处理技术的推广和应用为复杂的市场细分提供了可能。数据挖掘技术的发展,给市场细分研究带来了前所未有的突破。在这一章中我们将介绍数据挖掘算法在市场细分中的应用。

3.1 数据挖掘技术

数据挖掘(Data mining)技术是人们长期对数据库技术进行研究和开发的结果。起初各种商业数据是存储在计算机的数据库中的,然后发展到可对数据库进行查询和访问,进而发展到对数据库的即时遍历。数据挖掘使数据库技术进入了一个更高级的阶段,它不仅能对历史数据进行查询和遍历,并且能够找出历史数据之间的潜在联系,从而促进信息的传递。

3.1.1 数据挖掘的概念

3.1.1.1 商业层面上的定义

数据挖掘是一种新的商业信息处理技术,其主要特点是对商业数据库中的大量业务数据进行抽取、转换、分析和其他模型化处理,从中提取辅助商业决策的关键性数据。

简而言之,数据挖掘其实是一类深层次的数据分析方法。数据分析本身已经有很多年的历史,只不过在过去数据收集和分析的目的在于科学研究,另外,由于当时计算能力的限制,对大数据量进行分析的复杂数据分析方法受到很大限制。现在,由于各行业业务自动化的实现,商业领域产生了大量的业务数据。分析这些数据也不再是单纯为了研究的需要,更主要是为商业决策提供真正有价值的信息,进而获得利润。但所有企业面临的一个共同问题是:企业数据量非常大,而其中真正有价值的信息却很少,因此从大量的数据中经过深层分析,获得有利于商业运作、提高竞争力的信息,就像从矿石中淘金一样,数据挖掘也因此而得

名。

综上,数据挖掘可以描述为:按企业既定业务目标,对大量的企业数据进行探索和分析,揭示隐藏的、未知的或验证已知的规律性,并进一步将其模型化的先进有效的方法。

3.1.1.2 技术层面上的定义

概括地说,数据挖掘就是从大量的、不完全的、有噪声的、模糊的、随机的数据中,提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程。作为一种数据处理技术,数据挖掘允许数据处理者从不同的角度去分析数据。找出数据集中个数据子集所蕴涵的指示,并能发现这些数据子集间的关系,从而找出隐藏在被处理数据中的数据关系规则。

数据挖掘是一门交叉学科,它把人们对数据的应用从低层次的简单查询,提升到从数据中挖掘知识,提供决策支持。

需要说明的是,数据挖掘并不是一个新兴的技术,许多年前,统计学家就开始手工挖掘数据库,寻找符合统计学规律的有意义的模式。今天的技术是可以使挖掘过程自动化,把数据挖掘与商业数据仓库相结合,并以适当的形式把结果表示给从事商业活动的用户看。更好的数据挖掘产品不仅使用强大的算法建模,而且还涉及到更广泛的商业和技术方面问题——如把它们集成到当今复杂的信息技术应用环境中去。

3.1.2 数据挖掘的步骤

数据挖掘的过程就是一个不断探索数据特征、建立和检验模型,发现模型和数据间的关系的过程。这个过程大体上可以分为:问题定义、数据准备、模型建立、模型的验证评价。

3.1.2.1 问题定义

在该过程中,数据挖掘人员必须与领域专家及最终用户紧密协作,一方面明确实际工作对数据挖掘的要求,另一方面通过对各种学习算法的对比进而确定可用的学习算法。后续的学习算法选择和数据集的准备都是在此基础上进行的。

3.1.2.2 数据准备

在问题定义之后,我们需要为数据挖掘准备一个数据集。数据准备过程分为三个阶段:

1. 数据的选择

搜索所有与业务对象有关的内部和外部数据信息,并从中选择出适用于数据挖掘应用的数据。根据数据挖掘的主题,选择相关的数据进行数据取样,选择与数据挖掘任务相关的数据集,减少数据挖掘要处理的数据量,而又不降低数据挖掘的精确度。这个过程是数据集的定位过程。

2. 数据的预处理

研究数据的质量,为进一步的分析做准备,并确定将要进行的挖掘操作的类型。包括取出数据集中的噪音数据,为构造只是模型收集必要的信息,决定处理残缺数据的策略,报告时间序列信息等。也就是对数据进行过滤,并且对数据的空间属性赋值。

3. 数据的转换

将数据转换成一个分析模型。这个分析模型是针对挖掘算法建立的。建立一个真正适合挖掘算法的分析模型是数据挖掘成功的关键。这个过程的主要任务是使用减低维数或投影的方法来减少要考察的有效变量数目,选取数据集的维数。

3.1.2.3 模型建立

建立数学模型是数据挖掘工作的核心环节。该阶段首先根据对问题的定义明确挖掘的任务或目的,如分类、聚类、关联规则发现或序列模式发现等。之后要决定使用什么样的算法。数据挖掘中具体使用哪一种方法,取决于数据集市的特征和需要实现的目标。在实际应用中,往往是对多种建模方法的比较和综合。目前比较常用的建模方法有神经网络模型、决策树模型和回归模型。

3.1.2.4 模型检验与评估

数据挖掘阶段发现的模式,经过评估可能会发现存在冗余或无关的模式,这时需要将其剔除;模式也有可能不满足用户要求,这时则需要整个发现过程回退到前面阶段,如重新选取数据、采用新的数据变换方法、设定新的参数值,甚至换一种算法等。模型检验阶段的主要工作是把检验数据代入已经建立的模型中,观察模型的响应,通过比较模型的响应和真实的数据,从而评估模型的准确程度。

如果模型的准确性比较差，就需要重新进行数据探索、建立新的模型、直至新模型检验。因此，在实际应用中，数据探索、建立模型、模型检验是反复迭代的过程。

3.1.2.5 模型的实施

模型建立并经验证之后，可以有两种主要的使用方法。第一种是提供给业务人员或分析人员做参考，通过察看和分析这个模型之后提出行动方案建议。例如可以把模型检测到的聚集、模型中蕴含的规则、或表明模型效果的 ROI 图表拿给分析人员看。另一种是把此模型应用到不同的数据集上。模型可以用来标示一个事例的类别，给一项申请打分等。还可以用模型在数据库中选择符合特定要求的记录，以用 OLAP 工具做进一步的分析。

3.1.3 数据挖掘常用算法

在模型建立阶段的主要任务是选择数据挖掘的算法。选择实现算法要考虑两个因素：一是不同的数据有不同的特点，因此，需要用与之相关的算法来挖掘；二是要根据用户或实际运行系统的要求，有的用户可能希望获取描述型的、容易理解的知识，而有的用户只是希望获取预测准确度尽可能高的预测型知识，并不在意获取的知识是否易于理解。下面简要介绍几种常用的数据挖掘算法。

3.1.3.1 神经网络算法（Neural Networks）

神经网络近来越来越受到人们的关注，因为它为解决高复杂度问题提供了一种相对来说比较有效的简单方法。神经网络可以很容易的解决具有上百个参数的问题。

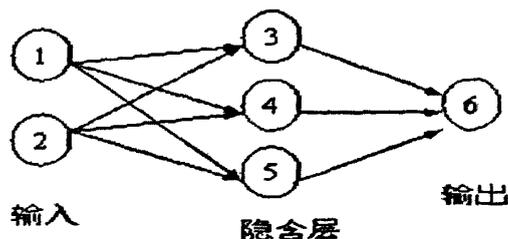


图 4-2 一个神经元网络

在结构上,可以把一个神经网络划分为输入层、输出层和隐含层(见图4-2)。输入层的每个节点对应一个个的预测变量。输出层的节点对应目标变量,可有多。在输入层和输出层之间是隐含层(对神经网络使用者来说不可见),隐含层的层数和每层节点的个数决定了神经网络的复杂度。

调整节点间连接的权重就是在建立(也称训练)神经网络时要做的工作。最早的也是最基本的权重调整方法是错误回馈法,现在较新的有变化坡度法、类牛顿法等。无论采用那种训练方法,都需要有一些参数来控制训练的过程,如防止训练过度和控制训练的速度。

3.1.3.2 决策树算法 (Decision Trees)

决策树提供了一种展示类似在什么条件下会得到什么值这类规则的方法。数据挖掘中决策树是一种经常要用到的技术,可以用于分析数据,同样也可以用来做预测。它是利用信息论中的互信息(信息增益)寻找数据库中具有最大信息量的字段,建立决策树的一个结点,再根据字段的不同取值建立树的分支;在每个分支子集中,重复建立树的下层结点和分支的过程,即可建立决策树。常用的算法有CHAID、CART、Quest和C5.0。

3.1.3.3 关联规则算法 (Association Rules)

关联规则是发现交易数据库中不同商品(项)之间的联系,这些规则找出顾客购买行为模式,如购买了某一商品对购买其他商品的影响。数据关联是数据库中存在的一类重要的可被发现的知识。若两个或多个变量的取值之间存在某种规律性,就称为关联。关联可分为简单关联、时序关联、因果关联。关联分析的目的在于找出数据库中隐藏的关联网。有时并不知道数据库中数据的关联函数,即使知道也是不确定的,因此关联分析生成的规则带有可信度。

关联规则发现可分为两步。第一步是迭代识别所有的频繁项目集,要求频繁项目集的支持率不低于用户设定的最低值;第二步是从频繁项目集中构造可信度不低于用户设定的最低值的规则。识别或发现所有频繁项目集是关联规则发现算法的核心,也是计算量最大的部分。

常用的关联规则挖掘算法有Apriori算法、频繁模式增长、多层次关联规则挖掘、多维关联规则挖掘、基于约束的关联规则挖掘等。

3.1.3.4 聚类算法 (Clustering)

将物理或抽象对象的集合分组成为有类似的对象组成的多个类的过程,称为聚类。由聚类所生成的簇 (Cluster) 是一组数据对象的集合,同一簇中的对象彼此相似,与其他簇中的对象相异。在许多应用实践中,可以将一个簇中的数据对象作为一个整体来对待。聚类与分类的主要不同点在于聚类不依赖于定义的类和带有标号的训练数据集,它要划分的类是未知的,故将聚类称为观察式学习,而不是示例式学习。

主要的聚类方法有:划分方法;基于密度的方法;基于网格的方法;基于模型的方法,如统计学方法、神经网络方法等。

3.1.3.5 遗传算法 (Genetic Algorithms)

遗传算法是由美国密执根大学的 Holland 教授创建的。它是一种借鉴生物界自然选择和自然遗传机理的高度并行、随机、自适应搜索算法,具有规则描述和程序实现相对简单、表现行为却相当复杂的特点,而且收敛速度快,计算时间较短。采用遗传算法,在普通计算机上用最大众化的编程语言就能实现计算机虚拟,因此在企业管理和决策过程中有着广阔的应用前景。

3.1.4 数据挖掘常用软件

目前市场上发布的数据挖掘软件大致分为三类:通用工具类、综合工具类、面向特定应用工具类。其中:

1. 通用工具类

这类工具目前占有较大和成熟的一块市场。从定义上说,它们是非面向特定应用的;从本质上说,它们的应用范围是水平的。

主要包括: SAS Enterprise Miner, SPSS Clementine, IBM Intelligent Miner, Oracle Darwin, Angoss KnowledgeSeeker.

2. 综合工具

这类工具是指商家要求该类工具能提供管理报告、在线分析处理和和普通结构中的数据挖掘能力。

著名的有 Cognos Scenario 和 Business Objects。

3. 面向特定应用工具类

这类工具发展迅速,在这一领域的厂商设法通过提供商业方案而不是寻求方案的一种技术来区分自己和别的领域的厂商。

重要的工具有: KDI(重点在零售业); Options & Choices(重点在保险业); HNC(重点在欺诈行为探查); Unica Model 1(重点在市场)。

目前的数据挖掘软件分类如下表:

表 4-1 数据挖掘软件归类表

实现技术	数据挖掘软件
统计方法	SAS/EM, Clementine, DataEngine, Partek, Matlab
可视化技术	SAS/EM, Clementine, Visualication Data Explorer, IRIS, Partek, PV-WAVE, WinViz, MineSet, AVS/Express, NetMap, CrossGraphs,
决策树	SAS/EM, Dawin, CART, KnowledgeSEEKER, KnowledgeSTUDIO, Business Miner, Scenario, Intelligent Miner, Decision Series, Minset, ALICE d' I Soft, SE-Learn, MinSet, NCR
神经网络	SAS/EM, Clementine, 4Thought, Intelligent Miner, Decision Series, NeuralSIM, Darwin, DataEngine, DataScope, dbProphet, Partek, KnowledgeSTUDIO, Scenario, HNC, NRC, Unica Model 1
遗传算法	Partek, Aegis Development System, OMEGA, Unica Model 1
关联规则	SAS/EM, MineSet, Clementine, Scenario, Decision Series. Intelligent Miner, NCR, KDI, Options & Choices
最邻近算法	SAS/EM, Darwin, KnowledgeSTUDIO, Intelligent Miner

从上面的表格我们可以看出目前的数据挖掘软件主要实现了统计、可视化技术、神经网络、决策树等方法,其他方法还有待开发。

3.2 市场细分常用的数据挖掘算法

数据挖掘技术在企业市场营销中得到了比较普遍的应用,它的发展给市场细分研究带来了前所未有的突破。有多种方式可以在细分上运用数据挖掘,很多数据挖掘技术都可以履行分类任务从而被用来将客户数据库分成细分群,每个细分群由这些被归入同一类的客户组成。此外,类似聚类的方法也可以用来进行市场细分,在这种情况下,将客户归为某一细分群是因为客户和同一群体中的客户更像,和其他群体的客户差别较大。

下面我们来详细介绍几种在市场细分中常用的分类和聚类的方法。

3.2.1 聚类算法

聚类技术最广泛的用于客户细分,它是将一组个体按照相似性的大小分成若干个类别,即“物以类聚”。聚类是一种发现型数据挖掘技术。通常人们讨论的

发现型数据挖掘技术是指在做决策是前没有任何对客户细分的预备知识的要求。这种技术使那些具有相同特征的客户组成客户群，同时使不同客户群中的客户差异性最大化。

3.2.1.1 聚类概念

聚类（Clustering）是将物理或抽象对象进行分组并将相似对象归为一类的过程。数据聚类将物理的或抽象的对象分成几个群体，在每个群体内部，对象之间具有较高的相似性，而在群体之间相似性则比较低。

聚类问题可一般性的描述为：待聚类样本空间 $X = \{X_1, X_2, \dots, X_n\}$ ，每个样本 X_i 由一组特征数据组成的 m 维向量， $(x_{i1}, x_{i2}, \dots, x_{im})$ 表示， X 的样本聚类即是 X 的一个划分 A_1, A_2, \dots, A_t 满足

$$\bigcup_{i=1}^t A_i = X, \quad A_i \cap A_j = \Phi$$

（ i 与每个划分 A 的元素的相互之间的距离都很小）

当 t 为定值时的聚类是静态聚类，即决策者已事先定出聚类的类别数； t 为变量时的聚类是动态聚类，即决策者事先不指定聚类数， t 的大小由样本空间的客观情况而定。

作为统计学的一个分支，在统计方法中，聚类称聚类分析，它是多元数据分析的三大方法之一（其它两种是回归分析和判别分析）。它主要研究基于几何距离的聚类，如欧式距离、明考斯基距离等。传统的统计聚类分析方法包括系统聚类法、分解法、加入法、动态聚类法、有序样品聚类、有重叠聚类和模糊聚类等。这种聚类方法是一种基于全局比较的聚类，它需要考察所有的个体才能决定类的划分；因此它要求所有的数据必须预先给定，而不能动态增加新的数据对象。聚类分析方法不具有线性的计算复杂度，难以适用于数据库非常大的情况。

在机器学习领域，聚类是无监督的学习（unsupervised learning）的一个例子。与分类学习相比，分类学习的例子或数据对象有类别标记，而要聚类的例子则没有标记，需要由聚类学习算法来自动确定。而且与有监督的分类方法相比，聚类的时间复杂度要比分类大得多。

主要的聚类技术包括了划分方法，层次的方法，基于密度的方法和基于模型的方法等。

3.2.1.2 基于模型的聚类方法

在完成细分时，我们可以选择已经成形的数据挖掘工具来完成客户聚类。例如 IBM 的商业智能软件 Intelligent miner 中使用的聚类方法是基于模型聚类方法。

基于模型的聚类方法试图优化给定的数据和某些数学模型之间的适应性。这样的方法经常是基于“数据是根据潜在的概率分布生成的”假设。基于模型的方法主要有两类：统计学方法和神经网络方法。

1. 统计学方法

概念聚类 (conceptual) 是机器学习中的一种聚类方法，绝大多数方法采用了统计学的途径，在决定概念或聚类时使用概率度量。与传统的聚类不同，概念聚类除了确定相似对象的分组外，还要为每组对象发现特征描述。也就是说，只有当一组对象可以被概念描述时才形成一个簇。这不同于基于几何距离度量相似度的传统聚类。因此概念聚类由两部分组成：1) 发现合适的簇；2) 形成对每个簇的描述。在这里，追求较高类内相似度和较低类间相似度的指导原则仍然适用。比较流行的概念聚类的方法有 COBWEB、CLASSIT、AutoClass 等。

2. 神经网络方法

神经网络方法将每个簇描述为一个标本 (exemplar)。标本作为聚类的“原型”，不一定对应一个特定的数据实例或对象。根据某些距离度量，新的对象可以被分配给标本与其最相似的簇。被分配给一个簇的对象的属性可以根据该簇的标本的属性来预测。两个比较著名的神经网络聚类是竞争学习 (competitive learning) 和自组织特征映射 (self-organizing feature map)。

3.2.2 分类算法

虽然聚类算法是当前市场研究中必不可少的分析方法之一，在市场研究中的应用已经非常普遍且研究与应用文献很多，但在市场研究过程中远远无法满足需要，尤其是聚类的目的性需要主观判断，并且客观上一般认为所有指标所描述的特征的重要性是相同的，不具备对指标的筛选和评判功能。因此市场细分应用上受到一定限制。

分类算法也是我们在行为细分中常用的一种方法。这里我们介绍决策树中的 AID 和 CHAID 方法。AID (Automatic Interaction Detector) 和 CHAID (CHI-Squared Automatic Interaction Detector) 方法的目的与聚类分析相似，但它们在市场细分中具有独特作用。AID 和 CHAID 的核心思想都是根据一个给定的目标变量和其他经过筛选的样本特征指标对样本进行最优分割的方法。AID 用于对连续或序次

等级变量进行优化分割的方法，它具有互动和自适应或自学习功能。与 AID 相比 CHAID 是比较新的市场细分方法。由于市场调查和市场研究多为分类或定性数据，无法用连续变量的方法处理，因此需要相应的分析方法加以研究。CHAID 就是专门用于对分类和序次等级数据进行分析的有效方法，是针对分类或序次等级变量，根据给定目标变量和样本特征变量进行优化分类的方法 AID 和 CHAID 的自学习和自适应功能在模式的自动识别和优化方面起到重要的作用。

3.2.2.1 AID 算法

AID 方法的分类过程是：首先根据市场细分目的选定被分割样本的一个特征变量作为分类的目标变量，将全部待分样本看作一类。然后选定除了目标变量以外的其他指标作为分类变量并依次分别进行分割，以被分割目标变量确立的最优目标为判据，确定分割的最优变量。使用最优分割变量进行分割，产生新类。对新类分别根据目标函数，确定下一次分割群体，然后对待分样本重复上述步骤，直到分类条件满足为止。

从 AID 的分割过程可以看到，所选定的变量可能仅仅在某一个分割阶段起作用，或许在全过程都没有参与分割，或许不止一次参与分割，从而使各个阶段的分割在优先信息标准下达到最优。尤其是分类的目的性在市场研究过程中与系统聚类和动态聚类相比具有突出的优势。

3.2.2.2 CHAID 算法

CHAID 方法的分类过程与 AID 相似。CHAID 以原始数据处理为出发点。首先选择分类的目标变量，然后分类指标与分类目标变量进行交叉分类，产生一系列二维频数表；分别计算所生成二维分类表的 χ^2 或似然估计统计量，比较统计量的大小，以最大统计量的二维表作为最佳初始分类表，在最佳二位分类的基础上继续使用分类指标对目标变量进行分类，重复上述过程直到分类条件满足为止。

从 CHAID 的算法可以看出：CHAID 的分类过程与 AID 相似，也是采用逐步探查的方法。由于 CHAID 采用定性数据，目标变量可以根据分类特征指标进行多种分类，CHAID 方法就是对众多分类加以比较并找到最佳分类变量和最佳分类结果。按照最优分类线索找到的最佳结果成为继续进行最优分类的依据。在分类过程中，虽分类维数的增加，对样本特征的描述越来越准确，同时，样本对目标的反应也越来越突出。

第四章 行为驱动的移动通信客户细分模型

4.1 客户行为细分

4.1.1 客户行为细分的提出

在第一章中我们介绍了 5 种细分指标,到底哪些细分指标对细分市场的作用最大呢?为具体的产品或服务选择一个或一系列细分指标作为细分的基础通常是一件非常困难的事情。

每种指标在进行细分时都有它不可替代的作用。但是这并不等同于这些细分指标在进行细分时都具有相同的权重。在使用它们时要有一定的优先顺序,否则就会造成主次不清,不能有效地进行细分。

在今天的营销环境下,企业的目的不仅在于要寻找到有潜力、值得为之服务的客户群体,更重要的是,要千方百计取悦他们,让他们满意,并最终形成持续的购买行为,这就要求企业的经营活动具有充分针对性。经验表明当结构化的绝对的细分研究在提供一些分析甚至一些偶然的推测型预测能力时,它们得到的结果与精确的结果相去甚远。

当我们想要进行精确和精细的预测时,有一个非常有效的途径就是根据客户过去行为的历史来分析客户。就像很多人从历史中认识到的那样,得出人们真实想法和感受的最好途径就是观察他们在做什么。从市场的角度看,这尤其是真的。

4.1.1.1 细分工作效益的要求

市场细分是一项成本昂贵的营销活动。通常我们在进行细分时,需要开展很多工作,如市场调研、市场分析、市场预测等等,这些工作都需要花费大量的资金,而且它还需要营销人员浮出足够多的时间和营销努力,因而会给一个企业带来很高的机会成本。所以,保证细分工作的有效进行就成为营销人员的重要使命。而决定细分方向性的关键一步就是细分指标的确定。如果指标选择不当,后面细分工作无疑就是在错误的基础上开始另一个错误。

4.1.1.2 激烈市场竞争的要求

在产品推出的初始阶段,企业掌握的信息不是很多,也不系统。这时的市场

细分主要是凭借营销人员对市场的感性认识和他们多年的营销经验进行的。常常采用的是传统市场细分中的地理因素和人口统计因素来细分市场。这个时候的市场细分是对市场的广度细分，可以称之为一种艺术。

但是，当产品已经渗透到每个细分市场中时，这个市场也就进入了表层的饱和和竞争状态。这时，很多企业就陷入激烈而又残酷的价格战中无法自拔，这就是企业对市场细分的认识不足所产生的无奈表现。事实上，这时候，企业应该改变细分指标，选择更为复杂的客户行为细分变量，进一步了解客户的消费取向，完成对市场的深度细分。这非常符合市场细分的科学性原则。

4.1.1.3 大量客户信息的要求

电信企业最大的优势就是掌握了大量的客户信息，也就是说拥有了比较全面地客户消费行为相关的内部数据资料和外部数据资料。这些数据为电信企业进行客户细分提供了很好的依据。将这些数据进行科学的提炼、加工和集成，组织到统一的中央数据仓库系统，并运用各种先进的统计分析工具对这些数据进行灵活的、动态的、多角度、多层面的统计和分析，挖掘出对营销人员有用的信息，为市场细分提供有力的支持，以针对不同的客户群，开展有针对性、精确的市场营销活动，更大程度的满足客户的个性化需要，提高业务使用量，使客户为企业带来更大的价值。

综上，虽然市场人员可以研究客户的不同特征——他们的年龄、性别、参数和态度——最终，大家所关心的唯一事实就是客户真正的购买行为。我们只在其他特征能帮我们更好的理解真实行为时，才研究他们。

4.1.2 客户行为细分的定义

客户行为细分是指基于对客户历史行为数据的分析，识别具有相同行为的客户群，并且描述每个客户群的相同行为和其他辅助变量，从而完成对产品市场的细分。

客户行为细分的主要细分指标就是我们在前面提到的客户行为特征。客户行为特征是反映客户消费行为与需求差异的指标，主要用来描述客户对产品的了解程度、态度、使用情况及反应等。

4.2 客户行为细分的研究方法

4.2.1 RFM 分析方法

基于许多数据库营销的经验，直接营销专家 Bobstone 提出了 RFM 客户细分模型。RFM 分析是广泛应用于数据库营销的一种客户细分方法。

RFM 是指三个变量，市场研究人员依靠这三个变量作为主要客户行为的预测因素。这三个变量每个词代表客户在购买时的一种不同的行为特征，合并在一起，就描绘了一个巨大的客户群体购买行为的画卷。

R (Recency): 最后一次购买时间变量；指从该客户上次购买到现在的时间长短。时间越短，R 值越大。这个特征告诉我们客户多久前在我们这里购买过商品。

F (Frequency): 购买频率变量；指在某一时间段内客户消费的次数，交易次数越多的客户越有可能继续消费企业提供的产品。

M (Monetary): 购买金额变量；指在某一时间段内客户购买商品的消费金额。M 越大，说明客户为企业带来的价值越大，也越有可能再次响应企业的产品与服务。

RFM 分析的所有成分都是行为方面的，这些因素在客户的消费记录中容易得到。以最近的客户行为为依据可以更加准确和有效的预测的客户的后续消费行为。RFM 分析的流程如图 4-1 所示。

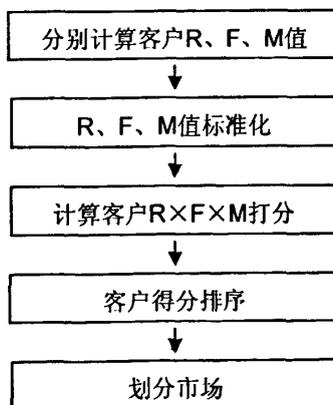


图 4-1 RFM 模型分析步骤

根据客户购买的历史交易记录，分别计算每个客户的 R、F、M 值。描述 RFM 分析的商业标准方法是利用古老又实用的交叉表报告。这种报告有代表性的从最

高消费开始列出了消费额的范围,在旁边还标注了客户购买的最近日期和购买频率值。

另一种方法是可以按照最近的购买日期将客户记录排序,最近的在顶端,最远的在底端。这样就可以把全部客户分成若干相同的部分。位于顶端的客户赋予高分,以此类推。这样每个客户都被定位在一个三维空间里。例如我们将客户分成5个组,这样从(1, 1, 1)到(5, 5, 5),合计有125个客户群。

在计算了所有客户的 $R \times F \times M$ 后,把计算结果从大到小排序,前面的20%是最好的客户,企业应该尽力保持他们,后面的20%客户企业不用花大量的经理在他们身上,企业还应该大力投资于中间的60%的客户,使他们向前面的20%迁移。向上迁移(Migrateup)的客户提高了他们的消费和忠诚度。另外,企业应关注那些拥有与前面20%的客户相同特性的潜在客户。

RFM分析对于直销市场人员接触客户行为是一种有效的客户细分方法。在企业开展促销活动后,重新计算出每个客户的RFM,对比促销前后的RFM,可以清楚的看出每个客户对于该活动的响应情况,为企业开展更加有效的营销提供可靠的依据。其缺点是分析过程复杂,需要消耗很多时间,而且细分后得到的客户群过多,如每一种变量是用3个值就会得到27个客户群,以至于难以形成对每个客户群的准确理解,也就难以针对每个细分客户群制定有效的营销策略。RFM分析的另一个缺点是购买次数(F)与同期总购买额(M)这两个变量间存在多重共线性,即如果某客户每多一次购买,他的总购买额也会相应的增加。

4.2.2 客户价值矩阵分析

为了消除购买次数与总购买金额之间的多重共线性,Mareas提出对传统的RFM分析模型进行修正,用平均购买金额代替总购买金额。另外,为了解决传统RFM分析过多细分客户群的缺陷,Mareas提出用购买次数(F)与平均购买金额(A)构造的客户价值矩阵简化细分的结果,如图4-2所示

产生客户价值矩阵需要的信息有:客户代码、购买日期、日购买金额,购买次数由不同购买日期的数目确定,日购买金额用来计算平均购买额。在Mareas提出的客户价值矩阵中,确定购买次数与平均购买金额的基准是各自的平均值,一旦确定每一个坐标轴的平均值,每个客户就被定位于客户价值矩阵的某个象限中的客户群的关键差异。

客户价值矩阵的优点在于这种细分方法可以对每一个客户群的提出了有针对性的营销战略/战术。最好的客户代表了企业的基础,保持是必要的;对于乐于消费性,最适当的战略是增加他们的购买频率;对于经常性客户,他们通过不

断的购买证明了自己的忠诚，最好的战略是通过交叉销售、向上销售增加他们的平均购买额；对于不确定性客户，最好的战略可以描述为慎重的精选，把营销努力集中与不确定性的新客户和那些对某种具体的产品感兴趣的客户。结合其它的客户信息，客户价值矩阵能够制定跨越客户群的营销战略/战术，强化客户保持。

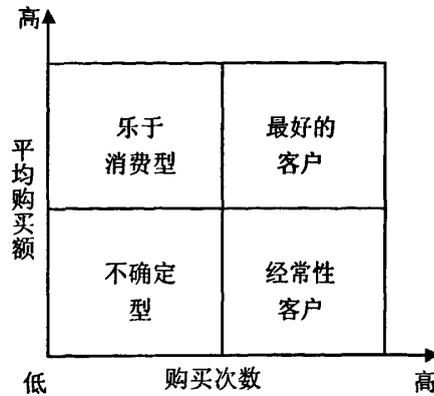


图 4-2 客户价值矩阵

4.2.3 多维度客户行为细分

RFM 分析与客户价值矩阵利用数据库中的客户购买信息进行低维度的客户细分。但是目前的客户数据库正变得更加庞大和复杂，市场细分越来越需要从更多的角度对客户进行分析。在市场细分过程中营销人员需要使用一系列的描述客户行为的指标进行市场分割。多维度的细分可以全面准确、细致的描述消费者特征，其研究结果比较精细。因此多维度的市场细分越来越受到市场研究和营销人员的重视，逐渐成为客户行为细分的重要研究方法。

我们在进行客户行为细分时不仅要考虑客户使用的频次和使用量，还要将客户使用的时机、使用的时间长短等因素考虑进来。这种需求在通信行业的客户分析中更为明显，我们需要在考虑客户呼叫次数、呼叫时长的同时，考虑客户呼叫的时间段、每次呼叫的间隔。

同时市场分析过程中，营销人员还要考虑产品购买和使用的替代性和相关性问题。客户在购买一种产品的同时可能还会去买另一种产品配套使用或者不会去买另一种产品。这样就进一步增加了客户行为细分复杂性。

我们应用数据挖掘技术中的聚类和分类等就可以处理这些复杂的、大量的客户行为数据，完成客户的行为细分工作。主要应用的数据挖掘技术我们在上一章中已经详细的介绍，这里就不再赘述了。

4.3 行为驱动的移动客户细分模型

客户行为细分是市场细分的一个很好的出发点,但是客户行为细分只能满足企业了解客户消费的方向。为了更好的了解客户的行为,制定合理的市场细分方案,我们还必须依据传统的主流细分指标。

企业通过对于客户历史行为的研究完成了客户行为细分,针对每个细分群体可以提供符合他们消费行为的产品或服务。但是企业要想对这些客户开展营销活动,首先要确定这些客户的活动范围和人口特征,因此利用客户基本特征指标对具体的客户群进行描述是必须的。同时,企业在制定营销方案时还要考虑到具体的细分市场能给企业带来多大的利润,需要企业花费多大的精力去进行营销活动等,这就需要利用客户的价值特征指标来进行衡量。由此,如果我们只分析客户的行为而放弃其它几种细分指标,或者将行为指标与其他指标孤立分析,那么细分活动就会大打折扣。

如何在客户行为细分的基础上整合其它的客户细分指标,是我们在完成行为细分研究时需要解决的另一个关键问题。这里我们根据客户的行为、基本特征和客户价值三种指标之间的关系,构建了一个基于客户行为细分的行为—特征—价值三维市场细分模型。如图 4-3 所示。也就是说,行为驱动的市场细分模型是综合利用市场细分中的三类细分指标完成市场细分和营销方案制定的。

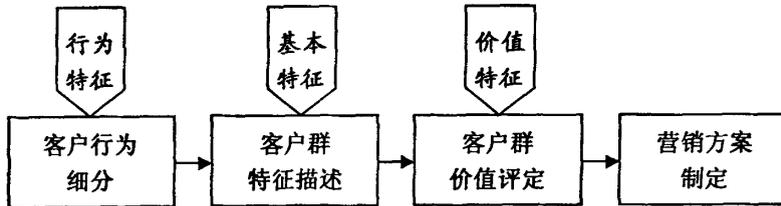


图 4-3 行为驱动的三维客户细分模型

4.3.1 模型主要特征变量

4.3.1.1 客户行为特征

1. 签约行为 (Subscribe Behavior)

签约行为这个词用来描述当客户决定使用或改变通信业务而与移动通信公司签订的某一项业务的使用合约事件。移动通信行业的客户签约行为与零售业客户的“进入商店”行为很相似。不论客户是否在零售商店中买东西,但如果当时

他们决定不进入商店，他们将不可能购买任何东西。通信运营商提供的很多业务（商品线）如长途电话、移动电话、数据通信等，在这些业务带来利润前，运营商一定要与客户签订协议。

描述客户签约行为的主要指标有：入网时间、在网时长、离网时间、离网原因、手机类型、服务类型、服务标识、服务号码、附加功能、转网标识、市场促销活动标识、数据业务服务等。

2. 使用行为 (Utilization Behavior)

使用行为这个词是用来从不同角度来描述客户如何使用服务的。使用行为可以用服务的使用分钟数或秒数、特定呼叫的时长、呼叫的次数或其他使用模式的种类来表示。这些信息是由网络设备、详细话单等捕捉到并记录的。在处理行为属性时要注意的，因为详细行为数据涉及的数据源很广泛，数据量很庞大，因此在选择具体的分析属性时一定要选择有业务意义的行为属性，增加分析的针对性和准确性。

描述客户使用行为的主要指标有：主叫号码、呼叫类型、被叫号码、呼叫发生地、呼叫时长、计费时长、平均呼叫时长、呼叫次数、呼叫金额、漫游呼叫次数、非漫游呼叫次数、呼叫时间、长途方式、国内长途呼叫次数、国际长途呼叫次数、高额呼叫次数、联系号码个数、活动地区数、呼叫其它运营商号码次数等。

3. 支付/信用行为 (Payment / Credit Behavior)

对通信运营商来说另一个重要的客户行为是客户支付账单的形势、频率和可信性。客户支付的速度和规律性对电信公司来说关系到客户的价值和信用度。支付和信用行为就是对客户支付的形式或者他们的信用度等级的描述。

描述客户支付/信用行为的主要指标有：账户 ID、担保信息、帐户类型、开户时间、交费渠道信息、付费计划、持有服务或产品、付费方式、欠费等级、催付状态、信用额度、交费金额、月均交费额、交费总额、销售渠道、套餐计划、预付费、押金、优惠方式等。

4. 维护行为 (Maintenance Behavior)

在描述通信行业消费者行为时，最近热点的话题是维持行为。维持行为告诉我们在维持客户的最低需求服务水平的费用。

维护行为从商业角度讲至关重要。很多商业客户经常要求大量的特殊的客户维持需求。我们在考察客户的贡献时，要将低利润与高维持成本的结合。有时客户可能看起来贡献很大（因为高使用率），但是事实上，从总成本的角度来讲，给企业带来的利润是非常小的。

从客户的维护行为中还可以看出客户的关心焦点业务是什么,如果投诉或咨询的次数增多,企业就要对该项业务或服务进行仔细的研究了。

描述客户维护行为的主要指标有基本记录:客户与服务中心联系次数、投诉次数、投诉原因等。

这四种行为没有一个客户行为变量独自就可以提供我们所需的信息。依据我们想要知道的和我们想要处理的信息,我们需要对这些因素进行综合。

行为因素的实际综合和它们之间的关系给我们提供了一个客户行为的直观图形描述,这就是所说的客户行为函数(B_f)。这个函数由不同行为变量关系的定义给出。

$$B_f(\text{Sub}, \text{Util}, \text{Cr}, \text{M})$$

其中, Sub: 签约行为; Util: 使用行为;

Cr: 信用/支付行为; M: 维持行为。

在客户行为细分中,根据不同的细分目标,行为函数中属性的选择也不相同。例如:

在对离网客户进行细分时,这四种行为都是我们比较关心的,这时的客户行为函数为: $B_f(\text{churn}) = B_f(\text{Sub}, \text{Util}, \text{Cr}, \text{M})$;

在研究进行的客户长途话音业务使用情况的细分,需要选择的主要是客户的使用行为。因此长途话音业务使用行为函数为: $B_f(\text{longcall}) = B_f(\text{Util})$ 。

4.3.1.2 客户基本特征

移动客户基本特征指标包括:客户背景属性、客户联系属性

1. 客户背景属性

客户背景属性是指客户的姓名、年龄、性别等人口统计信息,这些信息是由客户在首次申请服务时填写的,可以用于识别和描述客户群的特征。由于其中的一些信息会随时间的迁移而发生变化,造成信息的过时或不准确,因此要保证这些信息的准确,就必须持续不断的收集和维护这些信息,对信息变更保留历史记录。

主要的客户背景属性有:姓名、性别、身份证件号码、年龄、居住地址、职业类别、工作单位、婚姻状况、教育程度、收入范围等。

2. 客户联系属性

客户联系属性是指与客户联系方式有关的信息, 这些信息是由客户在首次申请服务时填写的, 为服务和市场营销提供与客户联系的渠道信息。这些信息会随时间的迁移而变化, 因此, 要不断更新信息, 并对变更保留历史记录。

主要的客户联系属性有: 联系电话、邮政编码、通信地址、传真号、Email 地址、联系人、联系人电话。

4.3.1.3 客户价值特征

客户价值属性是指根据客户的背景资料、消费及交费历史、呼叫记录等资料, 经过计算、分析和数据挖掘得到的用于对客户进行评价的指标。客户价值属性是对客户进行评价的一个重要组成部分, 是和业务决策结合最紧密的一种属性, 也是市场和客户研究的核心部分。

主要的客户价值属性有: 消费层次、信用度、活跃程度、客户价值、成为大客户的概率、离网概率、挽留价值等。

4.3.2 客户细分模型描述

行为驱动的客户细分模型实质是将客户定位在一个三维空间里。在确定客户的行为特征后, 将客户群体映射到“基本特征——客户价值”平面上, 根据客户在平面上的投影, 对客户进行描述定位, 指定相关的营销策略。

4.3.2.1 客户行为细分阶段

在细分的开始阶段, 我们选用客户行为特征指标来细分客户市场。根据客户的不同消费需求和行为将客户划分为不同的客户群组。每个客户群组中具有相同的消费行为, 不同客户群间消费行为差异显著。就像前面所说的, 将客户行为特征指标作为细分的初始变量, 可以使企业从客户的历史行为中了解客户的消费方向。这对于企业在分析已经推出的产品进行深度市场细分是非常必要的。

4.3.2.2 客户群特征描述阶段

第二步, 在完成客户行为细分之后, 我们选用客户基本特征指标对每一个细

细分市场进行客户群特征描述,在必要的时候可以根据基本特征属性对客户群进行再细分。使用客户行为特征指标进行市场细分时,存在着细分市场形象描述困难的问题。我们应用统计学的方法,对每个客户群体的基本特征指标进行统计分析,根据客户群的不同地理属性、人口统计特征以及心理特征对客户群进行描述,这样便于营销人员找到各个细分市场,同时根据不同市场的特点分析制定营销策略。此外,利用客户基本特征指标进行市场描述的同时,也是对前面行为细分市场的一种检验。如果在细分市场描述时,没有发现客户群体间的较大差别,就有必要检查前一步骤的正确性和合理性了。

4.3.2.3 客户群价值评定阶段

完成客户群特征描述后,营销人员已经基本找到了具有不同消费行为特征的客户群,并且知道他们的特征和营销努力的方向。在制定营销方案之前,营销人员需要完成每个营销方案的成本预算。具体针对某一客户群应该花多大的代价去完成市场推广工作呢?这个时候我们就需要考虑这个客户群到底能给企业带来多大的价值。

利用客户的价值特征指标,根据客户群中每客户的价值来推算客户群体的价值,这就是我们在第三步中需要完成的工作。这是制定营销方案中关键的一步,也是企业营销活动评定中的重要方面。

第五章 移动通信客户行为细分举例

到目前为止，我们讨论了行为驱动移动客户细分模型。下面，我们根据中国移动经营分析系统的专题分析需求，以移动长途语音业务为例，利用行为驱动的客户细分方法，完成客户细分。

5.1 商业目标理解

根据经典市场细分理论，我们在进行市场细分的第一步就是要明确的定义营销研究的商业目标。

5.1.1 基本商业问题

运营商在经营过程中基本的商业问题是：如何获得利润？利润主体是收益，收益由收入和成本决定，收入来源于客户使用业务而支付的费用，成本在于发展客户、系统建立和运营维护等方面。根据客户使用业务而支付的费用，再将商业问题分解为：“如何提高客户支付的费用？”，并进一步分解为：“提高客户的数量？”和“提高单个客户的平均价值？”。对于如何提高客户数量问题可以分为：“如何吸引新增客户？”以及“如何减少客户的流失？”。

5.1.2 具体商业目标

目前国内长途话音业务市场竞争比较激烈，除了几大运营商提供长途话音业务以外，用户也可以通过互联网等方式进行长途通话。如何在激烈的市场竞争中提高长途话音业务的市场份额，成为中国移动的市场经营中的一个目标。

中国移动用户在使用长途话音业务时可以选择的有：国内/国际长途电话直拨业务、移动 IP 一次拨号业务（17951）、中国移动 IP 二次拨号业务（IP 记帐卡 17950）、其他运营商 IP 记帐卡业务。具体中国移动用户在使用长途话音业务时喜欢使用那种方式，如何提高客户在移动 IP 电话业务商的使用量，减少使用其他运营商的 IP 业务，使中国移动在长途话音业务中提高市场份额。这是营销分析人员在长途话音业务市场细分中需要解决的问题。

因此，在进行移动长途话音业务市场细分时，我们的目标是通过对客户合理的分类，并对当前客户以及预期的客户群作区段分析，判断不同区段的优势与弱

势，对客户总体构成有准确的认识，对客户的服务和营销更具针对性。具体目标是：

1. 辨别客户在长途话音业务上的使用习惯，
利用客户群体特征制定长途业务推广方案；
2. 吸引新客户使用移动长途业务，
提高客户长途话音业务的使用量；
3. 提高中国移动在长途话音业务上的市场份额。

5.2 细分变量选择

1. 客户行为特征指标

根据商业目标，我们在选择客户行为指标时主要选取了客户的使用行为指标。初步选定的指标如下表：

表 5-1 客户行为特征指标

序号	指标名称	指标描述
1	月长途呼叫次数	统计每用户个体当前月的直拨的长途呼叫次数
2	月长途呼叫时长	统计每用户个体当前月的直拨的长途呼叫时长
3	月国内长途呼叫次数	统计每客户个体当前月的直拨的国内长途呼叫次数
4	月国内长途呼叫时长	统计每客户个体当前月的直拨的国内长途呼叫时长
5	月国际长途呼叫次数	统计每客户个体当前月的直拨的国际长途呼叫次数
6	月国际长途呼叫时长	统计每客户个体当前月的直拨的国际长途呼叫时长
7	月 17951 IP 国内区域呼叫次数	客户使用 17951 直拨拨打国内区域的次数
8	月 17951 IP 国外区域呼叫次数	客户使用 17951 直拨拨打国外区域的次数
9	月 17951 IP 国内区域呼叫时长	客户使用 17951 直拨拨打国内区域的时长
10	月 17951 IP 国外区域呼叫时长	客户使用 17951 直拨拨打国外区域的时长
11	月 17950 IP 业务呼叫次数	手机客户使用 17950 卡业务通话次数
12	月 17950 IP 业务呼叫时长	手机客户使用 17950 卡业务通话时长
13	月其他运营商 IP 业务使	手机客户使用其他运营商 IP 业务通话次数

	用次数	
14	月其他运营商 IP 业务使用时长	手机客户使用其他运营商 IP 业务通话时长
15	月 IP 业务总呼叫次数	Σ IP 通话次数
16	月 IP 业务总呼叫时长	Σ IP 通话时长
17	漫游使用 IP 次数	Σ IP 通话次数（漫游）

这些指标为初步选定的细分指标，在数据准备和实际建模过程中，我们需要对这些指标做进一步的分析，选定其中合适的指标或者组合其中的多个指标作为建模需要的变量。

2. 客户基本特征指标

客户背景属性和客户联系属性是此次客户细分模型所需要的客户基本特征指标。由于企业拥有的客户背景属性的不全面，我们在此次建模中仅使用客户的年龄和性别特征，其他客户地域属性、职业教育等信息缺失。

3. 客户价值特征指标

客户价值特征中提到的客户流失概率、信用度、客户潜在价值等指标都是目前移动经营分析其他专题同时研究的课题。在此次建模中，没有这些信息，所以客户价值特征指标主要选取客户月消费额和消费层次两个指标。

5.3 数据采集和预处理

5.3.1 数据来源

在进行细分时所用的数据主要来源于客户的呼叫记录数据、客户基本资料数据和客户缴费记录数据等。这些数据的获得有两种途径。一是从已经建成的经营分析系统数据仓库中抽样获得。在没有建成经营分析系统时，可以抽取移动 BOSS 数据，经过 ETL 生成建模数据集。

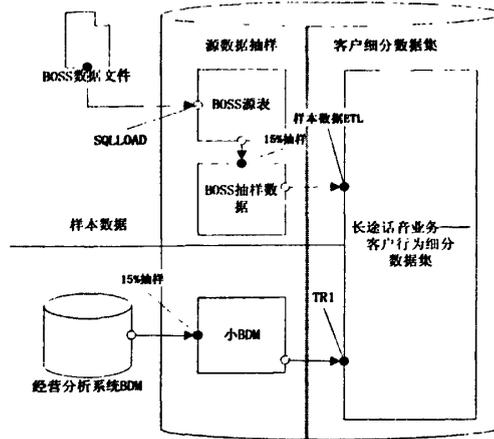


图 5-1 客户细分数据来源

5.3.2 数据评估

在建立客户细分数据集之后，我们需要执行一次最初数据自身的评估。根据对数据最初的理解从下面几项来检查数据集：

- 数据分配
- 缺失和不能用的数据
- 属性间关联

这里我们使用的数据分析工具是 IBM 的数据挖掘工具 Intelligent Miner（以下简称 IM）。用单变量统计或数据特征可视化 IM，可以获得对数据集的统计总揽。统计的结果用一系列柱状图和饼图表示，用可视的结果来分配数据和描述统计量，例如平均值、最小值、最大值。在这一步中，我们可以看到错值、缺失值和不可用的数据，以及一些可以指出数据错误的非正常分配。

下一步是看是否变量之间存在关联性。如果存在高度关联性的变量，可能会造成神经网络或其它数据挖掘方法得到结果的偏差。如果在细分数据库中存在高度关联的变量，那么结果可能就不再适宜了。因为两个高度关联的输入变量相当于同一个输入变量赋予了两次权值，如果输入是相同信息的多元相关变量，这样就会造成这个信息比其他输入更重要了。IM 提供了两种处理属性间关联的统计学方法：主成分分析和因子分析。

5.4 细分模型建立

5.4.1 建模环境

硬件配置： PC 台式机一台

软件配置： Windows 2000, DB2 数据库

建模工具： IBM 的数据挖掘工具 Intelligent Miner (IM)

用到的辅助工具和程序：

Quest Central for DB2 命令窗口、UltraEdit 文字编辑器

5.4.2 建模算法的选择

Intelligent miner 中提供的聚类方法是基于模型的两种聚类方法：统计学方法和神经网络方法。

在大多数变量都是数值时一般选用神经聚类, 如果大多数变量是陈述式的则一般选用统计聚类方法, 因为统计聚类具有不将陈述性变量转化成数字就可以进行聚类的独特功能。但是, 根据算法特征对两种算法进行考察, 两种算法也可以结合使用。

5.4.2.1 统计聚类

统计聚类最初由分析存在陈述型变量的典型统计数据发展而来。由于这个原因, 这种技术能最好的处理陈述型变量。统计聚类技术是自动发现聚类的个数, 而神经聚类需要指出你要分的聚类数, 但是要求指定同一类中客户的相似程度——condorcet。这个门限值的范围在 0 到 1 之间, 取值为 1 表示在这一类中所有的客户必须是同样的, 取值为 0 表示客户完全不同。如果两个客户比给定的 condorcet 更相似, 那么他们将被分配到统一的类中。

两个客户的相似程度可以通过比较每个客户特征值给出变量符合程度的得分来计算。然后将这个得分进行求和, 通过比较变量的数值来划分。如果所有的变量都是陈述型变量那就很容易理解了。以性别为例, 如果你要比较两个都是男性的客户, 那么它们的性别特征变量是相同的, 他们的得分为 1.0。如果他们的性别是不同的(男和女), 他们是不同的, 得分为 0.0。所以如果用 10 个特征变量来描述一个客户, 我们在比较两个客户时, 如果有 5 个变量是相同的 5 个是不同的, 那么这两个客户有 50%是相似的, 可以被分在同一组。

对于连续变量，相同的概念略有不同。如果取值是相同的，那么它们的得分是 1.0，但是如果他们不同，将根据不同的程度得到一个得分。典型的，如果依据所有客户的标准差来描述不同程度，如果两个变量值的差是 0.5 个标准差，则得分为 0.5。虽然，通常使用的是标准差，也可以定义一些其他的方法。

明显的，如果我们给定相似指数为 1.0，那么我们相信被分到同一组中的客户应是完全相同的。如果我们的客户特征变量很多或变量取值在一个很大的范围，那么结束时我们得到每个客户作为一类。在另一个极端，如果将相似指数设为 0，聚类结果是我们将得到一个包括所有客户的巨大的类，因为所有的客户都是候选。

因此，如果我们给统计聚类算法一个 condorcet，并不限制所产生类的个数，可以得到满足相似指数的最小类个数。

5.4.2.2 神经聚类

神经聚类在处理连续性变量或将陈述型变量转化成数值后是最好的技术。神经网络用于聚集的方法是将每一个类看作一个标本，它是这个类的“典型”，但不需和某个具体的记录或例子相对应。根据新记录和这个标本之间的距离，就可以将这个记录进行分类了。这里使用的是一个 kohonen 特征映射的神经网络。kohonen 特征映射使用一个称作自组织的进程来将相似的输入记录组合在一起。用户指定群集个数和便利数据的遍数。这些参数控制进程时间和将数据记录分配到群集时使用的粒度程度。与统计聚类不同的是，神经聚类需要建模人员给定想要分成的类的个数。通常，神经聚类要求输入数据的标准化，将数据标准化在 0.0 到 1.0 之间。

两种聚类算法所产生的分数在 0 到 1 之间，用来描述这个纪录与从属的群体之间的适合程度。换句话说，是与同类中的其他成员的相同程度。在统计聚类中，如果得分接近 1 则说明这个纪录与这个类很适合。而在神经聚类中如果得分为 0 才说明纪录很适合该类。

5.4.3 行为细分过程

我们采用的行为细分指标都是定量变量，因此在分群方法上我们选择 IM 中的神经聚类函数。

表 5-2 行为细分建模记录

数据集描述	数据集	MOD_LONGCALL_BEHAVIOR
	时间段	一个月

记录描述	最大遍数	7	最大群集数	9
	最大行数	3	最大列数	3
结果记录	最终群集方差 = 0.01356			

参数说明：

1. 最大遍数

指定的遍数增大可以改进所生成群集的质量。限制遍数会减少执行群集所需的进程时间，但它也降低了群集的准确度。一般来说，最多 5 到 10 遍就已经足够了。

2. 最大群集数

通过改变在最大群集数字段的值，可以更改神经网络函数生成的群集数目。在使用高级页面和控件时，此字段由最大行数和最大列数字段代替。

3. 最大行数和最大列数

在群集方式中，挖掘函数用矩形网络的形式来创建指定群集数。最大行数是网络的一维，最大列数是网络的另一维。网格等于提供的最大行数，小于等于最大列数。

4. 最终群集方差

在应用聚类算法后，得到的在 0—1 之间的分数，用来描述这个纪录与从属的群体之间的适合程度。换句话说，是与同类中的其他成员的相同程度。在神经聚类中如果得分为 0 才说明纪录很适合该类。

模型建立后的输出字段为：

1. 记录记分字段

该字段中的值将是最拟合群集的相应输入记录的拟合质量。

2. 群集标识字段

该字段中的值将是相应输入记录的次拟合群集的标识。

3. 记录记分字段

该字段中的值将是次拟合群集的相应输入记录的拟合质量。

4. 可信度字段

该字段的值将是记录分配的可信度。

5.4.4 客户群特征描述

根据行为细分中得到的群集标识字段,可以分别统计各行为细分群体的客户基本特征指标的分布情况,完成客户群特征描述。这里采用 IM 工具中的双变量统计分析方法。

我们可以使用双变量统计函数来执行以下任务:

- 计算数字字段的基本统计,例如,最大值、最小值、平均值、方差和频数;
- 为类别字段和离散数值字段计算频率;
- 对已选定字段执行 χ^2 平方测试。如果选择了一个字段,则将为其它字段与该字段的所有现有组合计算 χ^2 平方值;
- 执行 F 测试。F 测试适用于数字字段的所有偶对;
- 计算已选定数字字段的分位点;
- 获得输入数据之外的样本;
- 将输入数据作为一个示例复制到一个输出表中。

5.4.5 客户群价值评定

同样使用双变量统计分析,对各个客户群的消费额、客户积分、客户消费层次、客户潜在价值等进行统计。给出每个客户群在总利润中的贡献度和测算每客户群的营销成本门限。

5.5 模型结果分析

完成客户细分模型的建立之后,下一步就是解释我们得到的结果来决定如何将他们应用到我们的商业计划中。

IM 工具提供和很好的数据挖掘结果的可视化, 我们可以用图形来表示模型结果。

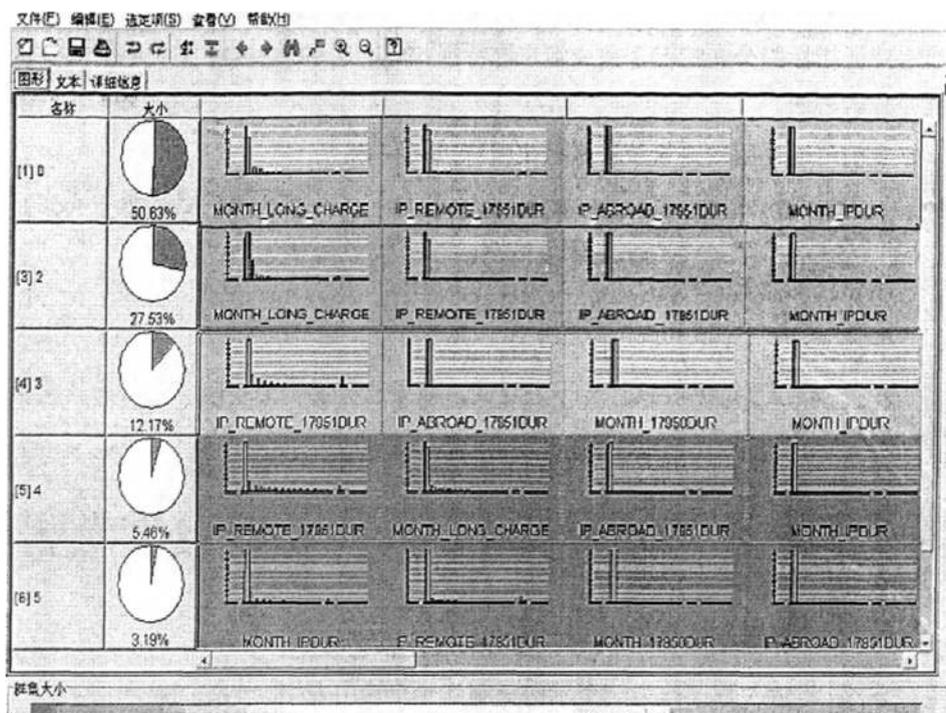


图 5-2 行为细分聚类结果

5.5.1 模型结果技术评定

评价细分结果的最重要的标准是要从商业的角度来理解结果。检验这方面的最好的途径是检验不同的细分并且尽量给每个细分一个从商业角度来看有意义的名字。

还有一些指标可以用来评价每个细分结果的质量。作为细分的结果, 我们可以得到客户个体被分到特定的群体和替换群体 (第二选择) 中的可信度。

模型结果的技术评定应该检查可信度是否过低 (接近 0.5)。如果是这样的话, 进行额外的反复细分是有必要的。检验指示变量第二选择聚类的变量是用来表明与所分到的类最接近的类是哪一个。如果一个簇的信用度很低, 并且与其他类具有很高的相似性, 那么就需要进行更多的反复挖掘。

5.5.2 模型结果商业分析

整个挖掘的过程中最有意义的一个步骤就是从商业的角度来对结果做出解释，为使用者提取出有用的信息或知识。

如图 5-2，我们得到的细分结果，将客户分成九个群体。可视化的细分结果从总体上根据客户的行为描述了客户在使用长途话音业务时的不同消费行为。例如：

- 长途业务需求度较低，使用较少的客户群体
- 国内长途直拨次数高，IP 使用次数较少的客户群
-

下面我们介绍一个感兴趣的客户细分群体的详细结果资料，并解释这个细分群的客户特性以及这些特性中隐藏的商业意义。

IP 业务缺乏了解的一组

这个客户群占客户总数的 27.53%。

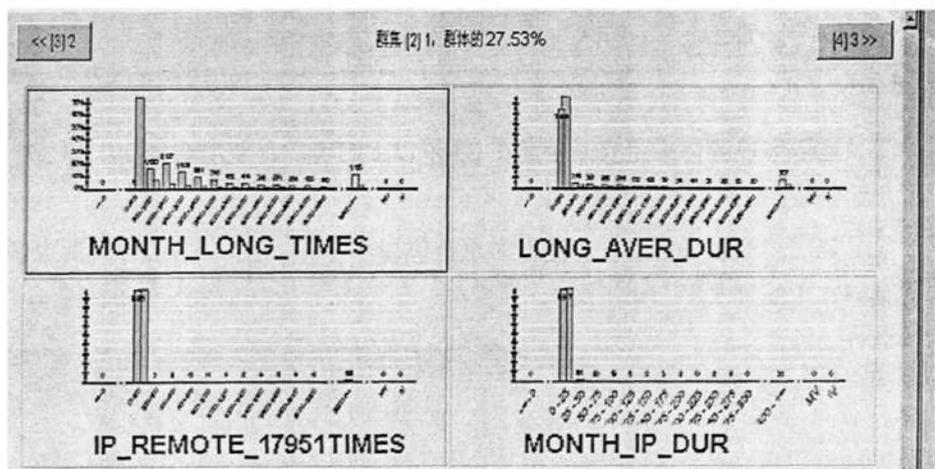


图 5-3 客户群行为特征

这个客户群在长途话音业务上的消费行为特点是：

- 月长途呼叫次数较高 (MONTH_LONG_TIMES)
- 月平均长途呼叫时长相对较短 (LONG_AVER_DUR)
- IP 业务使用次数和时长都很短 (IP_REMOTE_17951TIMES、MONTH_IP_DUR)

这些特征说明这个客户群体的客户长途业务的需求并不低,但是在进行长途呼叫时,主要采取的是直拨呼叫的方式,呼叫时长都比较短。这说明该客户群体可能还不了解 IP 电话业务。

对这个客户群进行基本特征分析得到

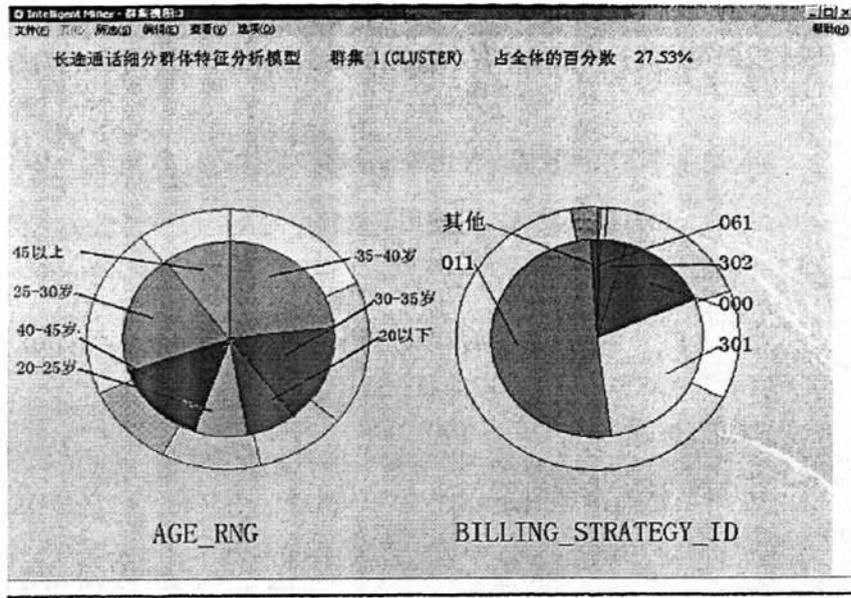


图 5-4 客户群体基本特征分析

这个客户群体中主要是年龄在 35 岁以上的客户,使用套餐 301 和 302 的客户比例较高。

第三步对这各客户群体进行客户价值指标分析,这个客户群体月消费额主要集中在中低等水平。如下图

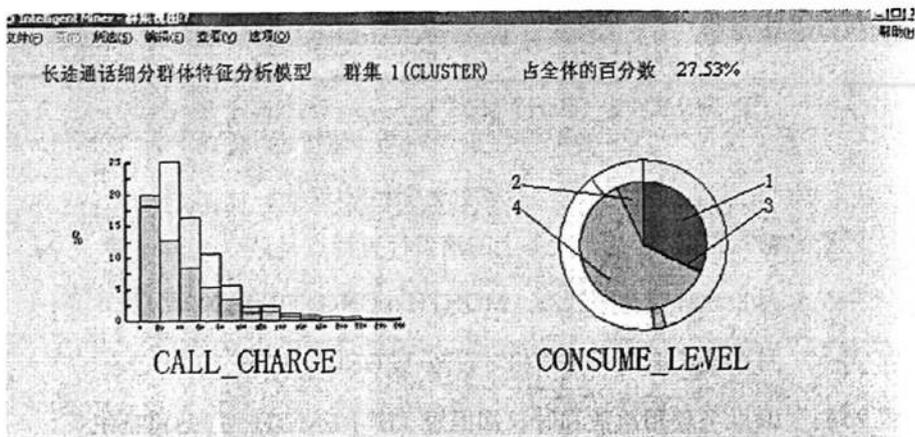


图 5-5 客户群价值特征分析

经过上面的分析，我们基本上可以描述这个群体为 IP 业务缺乏了解的客户，这些客户年龄较大，主要选用 301 和 302 套餐，消费层次处于中等水平。

5.6 细分模型应用

通过第四节的数据挖掘和第五节的结果解释后，我们要进行客户细分工作的最后一步，这也是所有步骤中最重要的一步。如何将挖掘的结果应用的商业问题中直接关系到数据挖掘提供的商业价值。因为我们经常将数据挖掘只看作是一种分析工具，很难用将它与现存的 CRM 系统结合到一起来获得更多的商业利润。在这一节中我们来讨论如何将数据挖掘在商业问题上进行展开，将细分的结果输入和应用到已有的 CRM 系统中。

5.6.1 模型的展开

一旦建立了有客户细分模型，我们就可以利用 IM 的应用模式为全部客户设计细分标识，能使将客户细分模型应用到每个客户，在需要时设计个性化的细分标识。

我们可以在不同应用中能使用细分标识，例如：

- 呼叫中心：当需要处理客户呼叫时，呼叫中心工作人员就能将客户的咨询投诉信息与细分相关联。
- 市场：我们可以使用其他的工具，如 OLAP 或电子数据表格，利用细分标识进一步做其他尺度的分析。

5.6.2 模型在市场战略中的应用

细分模型能应用的目标战略过程中：

- 定义目标客户群
- 定义战略方法
- 执行和评价战略

5.6.2.1 定义目标客户群

定义目标客户群是与企业战略目标和策略主要相关的因素。我们可以是从细

分模型中得到这个目标客户群的主要特征。

例如上面的例子中，我们就可以将目标客户定位在套餐 301、302 客户中。

5.6.2.2 定义战略方法

这里我们可以根据细分设计调查问卷来获得信息，开展市场调查来发现特殊细分群体客户的需求。这些需求和我们从挖掘中得到的客户特征是进一步发展为客户提供的业务和选择媒体的基础。

例如，我们可以针对套餐 301 和 302 的客户进行调查，得出他们主要接触的媒体宣传渠道。然后通过这个渠道向客户宣传 IP 一次直拨业务（17951）。

5.6.2.3 执行和评价战略

我们可以通过优先权执行客户战略活动，这个优先权是根据细分分数和客户价值相结合而制定的。实行客户优先权的原因是渠道容量限制了战略目标中客户的数量。

在活动实施后，我们可以评价活动的表现，同时活动的结果也可以用来进行其他的挖掘分析。

5.6.3 模型的维护

由于市场环境的动态变化，模型的执行会随时间逐渐降级。客户细分模型应给进行更新来适应变化和维护执行。当要更新模型时，我们需要考虑以下几点。

- 回顾用过的变量：变量的新的应用，例如新的业务的增加和取消；
- 回顾用过的技术：利用以往建模的经验。

总 结

模型的评价

对市场细分方法和移动通信行业客户消费特点的研究表明,行为驱动的客户细分模型是个较为理想的客户细分方法。因为它不仅帮助企业了解客户的消费方向,也为企业找出客户的特征并且为进一步的营销方案指定提供了费用预算方面的支持。也就是说此模型可以帮助企业从客户行为、基本特征、客户价值多角度的了解客户。企业对消费者的认识越全面,对细分市场间差异认识也就越深刻,细分的效果就会越好。此模型能够较好的满足了移动通信企业的市场细分需求,基本符合移动经营分析系统对客户行为分析的要求。

行为驱动的客户细分模型对用户信息的要求较高,客户信息的真实性和细分结果的准确性成因果关系。在此次建模中使用的客户基本特征只有客户的年龄和性别指标,缺失客户的地理特征、教育程度和工作等指标,影响了客户群体描述的精确性。移动运营商在今后的工作中应该注重客户信息的收集和更新工作,为客户细分提供准确全面的信息。

算法的评价

本文使用聚类分析算法研究了移动通信客户细分问题,应用 IBM 公司的 Intelligence miner 工具完成了细分模型建立的过程。聚类算法通过辨别客户在行为特征上的相似性,将客户划分为若干群。通过对移动长途话音业务客户细分的应用发现,聚类算法在分析客户细分问题便于使用和发现客户行为特征。

但是聚类算法也存在着一些不足之处。聚类的目的性需要主观判断,并且客观上一般认为所有指标所描述的特征的重要性是相同的,不具备对指标的筛选和评判功能。因此在进行聚类建模之前,所用指标的选择需要营销人员的经验,同时在数据预处理阶段需要对指标进行相关性分析,选择主要指标或用因子作为聚类分析的变量。

进一步的研究工作

1. 市场调查工作

企业在进行客户细分中,有时会需要了解客户的消费心理特征和生活方式、

利益追求和产品态度等信息，这些信息不可能也不必要通过企业内部数据库收集。大量的心理特征以及产品态度信息，需要企业通过外部的渠道信息获得，通常的手段是市场调查。为保证市场调查的有效性和准确性，在市场调查时要求拥有严谨的研究态度和科学的研究方法。

因此对市场调查方法的研究，以及如何将客户心理特征与其他特征相结合是移动通信客户细分研究的必要工作。

2. 市场细分绩效评价

市场细分的前提是建立在差异化的基础上，但是市场细分并不是越细越好。在企业进行市场细分时会涉及到市场、财务、人力等多部门多方面的决策。如何评估市场细分策略的可行性和细分工作的绩效，是一个需要研究的问题。

参考文献

- [1] Chengyi Sun,Hongxing Chao,Yan Sun "Genetic-Based Clustering Neural Networks and Applications" 1997 IEEE international Conference on Intelligent Systems
- [2] Martin Natter "Conditional Market Segmentation by Neural Networks" Department of Industrial Information Processing 1997
- [3] C.B.Kappert,S.W.F.Omta "Neural Networks and Business Modelling-An Application of Neural Modelling Techniques to Prospect Profiling in the Telecommunications Industry" 1997 IEEE
- [4] Sheng-Chai Chi,Ren-jien Kuo,Po-Wen Teng "A Fuzzy Self-Organizing Map Neural Network for Market Segmentation of Credit Card" National Taipei University of Technology 2000
- [5] Mattison, Robert M. "Winning telco customers using marketing databases" British Library Cataloguing in Publication Data 2001
- [6] "Mining your own business in telecoms using DB2 Intelligent Miner for Data" International Technical Support Organization July 2001
- [7] "IBM DB2 Intelligent Miner for Data" International Business Machines Corporation 1996, 2002.
- [8] Blattberg, Robert, Getz et al Managing Customer Retention[J] Incentive Apr.2002 Vol.176 Issue 4: 114-116
- [9] Jim Sellers, Arthur Hughes RFM Analysis: A New Approach to a Proven Technique [DB/OL] www.relationshipmktg.com/FreeArticles/rmr017.pdf
- [10] Arthur Middleton Hughes Making Your Database Pay Off Using Recency Frequency and Monetary Analysis [DB/OL] <http://www.dbmarketing.com/articles/Art104a.htm>
- [11]高云龙 于在青 郇启扬 《市场营销操作手册》社会科学文献出版社 2000 P206-214
- [12]Naresh·K·Malhotra 著 徐平等译 《市场营销研究应用导向（第3版）》电子工业出版社 2002
- [13]宝利嘉顾问 《细分——从客户区隔中谋取利润》中国社会科学出版社 2003
- [14]Alex Berson 等著 贺奇等译 《构建面向 CRM 的数据挖掘应用》人民邮

- 电出版社 2001
- [15]康晓东 《基于数据仓库的数据挖掘技术》 机械工业出版社 2004
- [16]韩 《数据挖掘：概念与技术》 机械工业出版社 2001
- [17]中国移动 《中国移动经营分析系统业务规范》 中国移动集团 2002.6
- [18]中国移动 《中国移动经营分析系统技术规范》 中国移动集团 2002.6
- [19]周志元 《移动经营分析系统之专题分析需求规格说明书》 北京新宇软件公司 2003.8
- [20]张悦 《IP 业务使用情况专题分析建模方案》 北京新宇软件公司 2003.8
- [21]张悦 《IP 使用情况分析专题结果分析报告》 北京新宇软件公司 2003.8
- [22]王广州 AID 和 CHAID 在多变量市场细分中的应用研究 市场与人口分析 第 5 卷第 6 期 1999.11 16-18
- [23]陈静宇 价值细分——价值驱动的细分模型 中国流通经济 2003,17(6) 53-56
- [24]傅志远 多元统计工具在电信市场研究中的应用 信息网络 2003(3) 10-15
- [25]刘义 万迪 张鹏 基于购买行为的客户细分方法比较研究 管理科学(哈尔滨) 2003(1) 69-72
- [26]张国方 金国栋 客户细分理论及应用策略研究 华中科技大学学报: 社会科学版 2003 17(3) 101-104
- [27]傅志远 市场细分标准的探讨及其对电信营销的启示
<http://www.chinatelecom.com.cn> 2003.6
- [28]张茂州 电信业务创新应基于市场细分
<http://tel.21cn.com/news/2002-08-15/744378.html> 2002.8
- [29]郑涌波 潘志 宽带用户市场细分初步分析 <http://www.cnii.com.cn/> 2002.7
- [30]胡世良 市场细分是固网增值新出路
http://www.ctiforum.com/forum/2003/12/forum03_1222.htm 2003.12
- [31]孙凤 中国城镇居民消费行为的数量研究 [博士学位] 国家图书馆 2000\F126\3
- [32]薛薇 数据挖掘的客户关系管理应用研究 [博士学位] 国家图书馆 2003\F274\14
- [33]易雪峰 基于顾客行为的战略规则研究 [博士学位] 国家图书馆 2003\F270\120
- [34]丁权 基于 CRM 数据挖掘的移动客户细分模型建立与分析 [硕士学位论文] 北京邮电大学图书馆 2003

