

文章编号:1001-9081(2010)02-0495-04

## 基于动态聚类的证券业客户细分实证研究

钱维佳,王延清

(华东理工大学 管理科学与工程系,上海 200237)

(kidsherry@163.com)

**摘要:**在客户关系管理理论基础上,建立了一个包含13个行业特色指标的证券业客户多维细分模型,并利用聚类分析对国内某知名券商的具体客户信息和交易数据进行了实证研究,有效识别出了具有不同特征以及偏好的客户群,并在此基础上提出了相应的营销策略。

**关键词:**动态聚类;客户关系管理;客户细分;证券业;个性化营销

**中图分类号:** TP311.13 **文献标志码:** A

## Empirical research on customer segmentation of securities based on clustering

QIAN Wei-jia, WANG Yan-qing

(Department of Management Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)

**Abstract:** This paper proposed a multidimensional customer segmentation model applied in securities industry, and carried out an empirical research in real customer data from a domestic famous security company by the clustering analysis of data mining. Finally, some marketing strategies were put forward in correspondence with different customers of different traits and preferences.

**Key words:** dynamic clustering; Customer Relationship Management (CRM); customer segmentation; securities industry; personalization marketing

### 0 引言

近年来,在佣金浮动制的催化作用下,中国证券行业的竞争日趋激烈,其经营模式已由过去的以“产品”为中心转变成为以“客户”为中心,客户关系管理(Customer Relationship Management, CRM)也由此引起了券商的广泛关注<sup>[1]</sup>。为了进一步提高服务质量,根据20/80法则,券商有必要进行客户细分,但是,按照资产、地域等明显属性将客户分为大、中、小、散户的传统客户细分方式已经无法满足要求<sup>[2]</sup>,面对券商长期积累的大量电子化数据,依靠传统的统计分析方法已经无法从中获取更多有价值的信息,而数据挖掘的提出则为深入进行客户分析提供了有效的途径<sup>[3-6]</sup>。目前,已有很多研究将数据挖掘技术应用到客户细分中,综合分析这些研究,可以发现它们具有以下问题:1)客户细分的维度不足。使用单一或者少量的维度对客户进行细分,完全不足以全面地描述客户特征,从而无法有效指导个性化营销。2)缺乏针对性。技术类研究者偏重于研究通过改进数据挖掘算法或者组合使用挖掘方法来提高细分的质量<sup>[7-11]</sup>,而实证分析偏好者的研究多针对零售业、电信业等领域,对证券业的客户细分研究可算是凤毛麟角<sup>[12-13]</sup>,导致指标体系的建立缺乏行业特色。基于此,本文拟在客户价值理论的基础上,深入研究影响客户投资的各种可能因素,构建证券客户多维细分模型,并利用动态聚类技术对客户进行细分。

### 1 证券客户多维细分模型

#### 1.1 客户多维细分指标体系

综合分析相关研究成果<sup>[3-4,7]</sup>,提出“证券客户多维细分

指标体系”,如图1所示,该体系以业务利润最大化为最终目标,具体包括两级指标:一级指标分别从四个方面来分析客户行为,即客户价值、客户关系强度、客户风险和客户偏好。一级指标下面又包含13个可量化的二级指标,分别代表客户的13个不同属性。

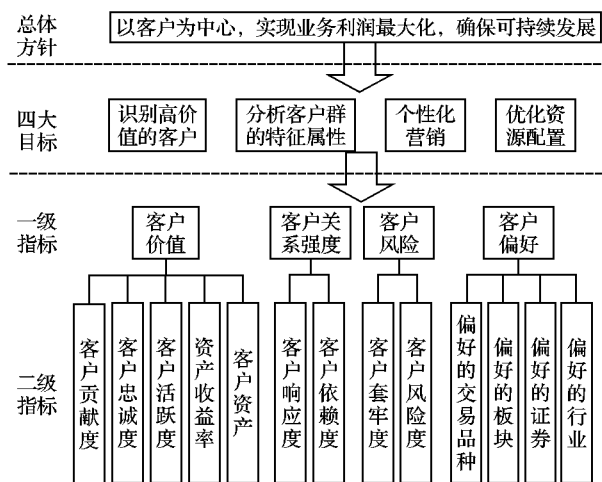


图1 证券客户多维细分指标体系

**客户价值** 为了识别出高价值客户,必须对客户价值进行分析。客户价值包括当前价值和潜在价值<sup>[14]</sup>,前者主要表现为客户的贡献度,而后者则可以通过客户活跃度、资产收益率和客户资产三个指标来进行综合评定。对客户价值的分析直接关系到客户关怀的资本投入,有效指导券商减少对于低价值客户的成本支出,将更多的资源用于那些真正有商业价值的客户上,实现资源的优化配置。

收稿日期:2009-08-19。

**作者简介:**钱维佳(1984-),女,上海人,硕士研究生,主要研究方向:金融电子商务、商务智能;王延清(1968-),男,山东济宁人,副教授,博士,主要研究方向:信息系统混沌理论、信息系统规划、金融电子商务。

**客户关系强度** 客户关系的强弱主要表现在客户对于经纪人所提供的投资建议的接受程度和响应速率上,关系强度高直接会影响到客户对于营销服务的接受程度、忠诚度以及对服务的满意程度,对其分析能够有针对性地提高营销服务的质量。

**客户风险** 风险分析的目的是掌握不同客户的投资风险偏好以及当前的投资风险状况。一方面,可以根据客户的风险偏好推荐不同的产品;另一方面,对于投资状况不理想的客户给予针对性的投资建议,不但能帮助其改善投资现状,还能提升客户满意度和忠诚度。

**客户偏好** 偏好分析则是实现个性化营销的基础之一,可以根据客户的特殊偏好来推荐产品<sup>[2]</sup>,提供针对性强的行情分析报告等手段来提高经纪人的服务质量,从而达到增加交易量和交易次数,增加客户满意度的目的。

## 1.2 指标计算

本文提出的模型共包含13个二级指标,各个维度的含义、计算方式和维度取值范围如下:

### 1) 客户贡献度。

客户贡献度是指过去每个时间单元的利润总和,假设不考虑服务成本支出:

$$\text{贡献度} = \text{佣金} + \text{利差}$$

根据样本数据和二八法则确定阈值1600,划分维度区间为{高,低}。

### 2) 客户忠诚度。

客户忠诚度可以通过计算各经纪人的客户保留率来衡量,其计算方式采用HarrisBlack顾客保留国际资讯公司的研究:

$$\text{客户保留率} = (\text{期末客户数} - \text{期间新增客户数}) \div \text{期初客户数}$$

根据样本数据,划分维度区间为{高,低}。

### 3) 客户活跃度。

活跃度按照周转率来计算:

$$\begin{aligned} \text{活跃度} &= \text{周转率} = \text{日均成交量} \div \text{日均资产} \div \text{日均成交量} = \\ &= \sum \text{每日成交量} \div \text{期间天数} \div \text{日均资产} = \\ &= \sum \text{每日资产} \div \text{期间天数} \end{aligned}$$

参照行业标准确定阈值{0.3,0.1,0.01},划分维度区间为{很活跃,比较活跃,不太活跃,很不活跃}。

### 4) 资产收益率。

资产收益率是客户交易能力的某种体现,只有保持良好的资产收益率才能保证客户资产不会萎缩,不至于打击投资积极性,是衡量交易持久性的重要指标:

$$\begin{aligned} T \text{ 时期客户 } J \text{ 的投资收益率 (ROA)} &= \\ &= (\text{期末资产} - \text{期初资产} - \text{资金净流入}) \div \text{期初资产} \end{aligned}$$

维度取值的范围为{平均以上,平均以下}。

### 5) 客户资产。

客户资产主要是指投资的市值以及保证金余额之和,资产大小决定了其交易能力的大小:

$$\text{资产} = \text{市值} + \text{保证金余额}$$

维度区间的划分为{大于100万,20~100万,1~20万,小于1万}。

### 6) 客户响应度。

客户响应度是指客户对券商经纪人投资建议的响应速度:

$$\text{客户响应度} = (\text{买入响应度} + \text{卖出响应度}) \div 2$$

$$\text{买入响应度} = \left\{ \sum \left\{ (\text{推荐品种买入成交金额} \div \text{前日资产总量}) \times [(\text{计算频度对应天数} - T + 1) \div \text{计算平度对应天数}] \right\} \right\} \div \text{推荐次数}$$

$$\text{卖出响应度} = \left\{ \sum \left\{ (\text{推荐品种卖出成交金额} \div \text{前日资产总量}) \times [(\text{计算频度对应天数} - T + 1) \div \text{计算平度对应天数}] \right\} \right\} \div \text{推荐次数}$$

$$T = \text{有效交易日} - \text{推荐日} (1 \leq T \leq \text{计算频度对应天数})$$

参照行业标准确定阈值0.3,划分维度区间为{高,低}。

### 7) 客户依懒度。

客户依懒度是指客户对券商经纪人投资建议的信任程度:

$$\text{客户依懒度} = (\text{推荐买入依懒度} + \text{推荐卖出依懒度}) \div 2$$

$$\text{推荐买入依懒度 (按日求和)} = \sum (\text{推荐品种成交金额} \div \text{推荐品种成交前日资产总量}) \div \text{推荐次数}$$

$$\text{推荐卖出依懒度 (按日求和)} = \sum (\text{推荐品种成交金额} \div \text{推荐品种前日持股数量}) \div \text{推荐次数}$$

参照行业标准确定阈值0.3,划分维度区间为{高,低}。

### 8) 客户套牢度。

客户套牢度反映了客户买卖证券的亏损情况,该维度很大程度上反映了客户的投资心理,套牢度越高,说明客户的亏损程度越高。

$$\text{套牢度} = (1 - \text{市值} \div \text{成本}) \times 100\%$$

参照行业标准确定阈值{0.8,0.6,0.4,0},划分维度区间为{深度套牢,高套牢,中套牢,低套牢,不套牢}。

### 9) 客户风险度。

客户风险度是指持仓风险,其中证券风险系数由券商设定,目前由于单个证券的风险系数尚无法确定,风险系数仅设置到品种:

$$\text{客户风险度} = \sum (\text{证券买入金额} \times \text{证券风险系数}) \div \text{期间总买入金额}$$

参照行业标准确定阈值{80,60,40},划分维度区间为{高风险,中高风险,中风险,低风险}。

### 10) 偏好的交易品种、板块、行业、证券。

客户的交易偏好会通过交易量的大小直接反映出来:

$$\text{某交易品种偏好度} = \text{某交易品种购买金额} \div \text{总购买金额}$$

$$\text{某板块偏好度} = \text{某板块购买金额} \div \text{总购买金额}$$

$$\text{某行业偏好度} = \text{某个行业证券购买金额} \div \text{总买入金额}$$

$$\text{某证券偏好度} = \text{某证券总买入量} \div \text{总买入量}$$

根据样本数据,划分维度区间为{低,中,高}。

## 2 证券客户细分的实证分析

### 2.1 数据准备

#### 2.1.1 样本数据

聚类采用的维度数量越多,对于数据的精度要求也就越

高,因此选取了国内电子化程度较高的某知名证券公司的数据资料作为数据来源,共收集了5000名样本客户的基本信息和581885条交易记录,时间跨度为2003年9月3日—2005年1月27日。在此基础上进一步对数据进行采样,选取其中代表性较强的数据,去除数据集中的“噪声”,从原数据中剔除一部分交易异常的客户,如在2004年全年都未曾进行过交易的死户共计2359个,以及当年销户客户和小部分异常事例等,最后保留2011个样本客户的基本信息和452011条交易记录,时间跨度修正至2004年1月1日—2004年12月31日。

### 2.1.2 数据的重构和整合

根据前文中列出的13个二级指标计算样本客户的指标值。由于样本数据集的数据完整性限制,我们将有选择地删减指标数量,去除掉“客户忠诚度”、“响应度”、“依赖度”和“证券的偏好”这四个指标。

### 2.1.3 聚类过程的数据处理

比例缩放和加权的概念在聚类过程中都起着重要作用。考虑到不同变量是以不同单位或者在不同范围中测量的事实,比例缩放可用于调整变量的值。加权方式则提供了对变量的一种相对调整,使得其中的一些变量变得比另一些变量更重要<sup>[15]</sup>。

各属性中属连续型的有:贡献度、活跃度、收益率、风险度、套牢度、客户资产,且这些数值无论在度量单位和数量级上,都存在较大差异,需进行标准化处理,其结果是使得各属性的均值为0,标准差为1。同时,本次聚类的目的在于识别出有价值的客户的基础上提供营销方案,因此,客户的价值分析显得相对重要,因此将客户贡献度的权重设定为2,其余为1。

## 2.2 动态聚类过程

通过SQL Server2005中的数据挖掘组件对客户数据进行挖掘<sup>[16]</sup>,过程主要有如下几个步骤。

1)使用不同的 $k$ 值(即客户分类数)进行自动聚类探测,比较不同 $k$ 值下的聚类结果的质量,初步确定最优 $k$ 值:聚类区别于分类的一大特征就是分类数的不确定性,在聚类开始之前,很难预先制定一个理由去选择一个特定的 $k$ 值,为了避免最后聚类得到的类别过于繁多,导致不可解释性和不符合业务需求,需要预先使用不同的 $k$ 值进行自动聚类探测,对于每一次的聚类结果,通过评价簇(cluster)的好坏、可解释程度、业务需求的契合度等因素,初步确定 $k=8$ (如图2)。

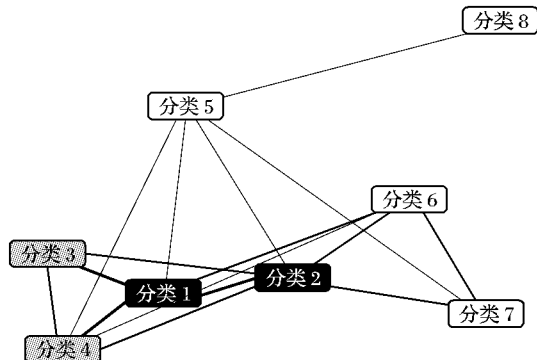


图2 显示聚类之间连接的关系图

2)分析分类关系图,评价簇的好坏,提出改进方案:簇好坏的评价主要有两个指标:“簇内形似”和“簇间差异”<sup>[15]</sup>。从几何学的角度看,就是簇内各个记录之间彼此相近,而各个

簇之间分开得很远。图3是 $k=8$ 时所获得的分类关系视图,表现出了各簇的强度(簇内形似)以及簇与簇之间的连接强度(簇间差异)。图中连线的深浅程度标识簇间差异,各节点的深浅代表差异的大小。

在每次聚类过程中,时常会产生一个或者若干个强簇——即样本数量相当大的簇,其中的记录非常相似。图3中,除了分类1、2属于强簇且表现良好之外(见表1),分类3(样本数294)、4(284)属于次强簇,且组内差异较小。余下的4个较弱的簇中分类7虽然包含的样本数比较小,但是其簇内记录相似性较好,属表现良好;分类5、分类6和分类8与其余5个簇之间的关联性比较弱,且簇内差异也比较大,表现较差,为了提高聚类质量,将这三个簇从总体样本中分离出去,进行单独聚类。

表1 分类1和分类2的维度变量数值统计表

维度变量	分类2(样本量:531)			分类1(样本量:523)		
	均值	标准差	说明	均值	标准差	说明
贡献度	-0.38	0.02	低	-0.34	0.04	低
活跃度	-0.28	0.02	很不活跃	-0.21	0.07	很不活跃
资产值	-0.11	0.02	1万~20万元	-0.10	0.02	1万~20万元
风险度	-1.34	0.03	低风险	0.78	0.00	低风险
套牢度	0.07	0.05	中套牢	0.04	0.03	低套牢
收益率	0.16	0.58	平均以上	-0.09	0.99	平均以下

3)根据改进方案,适当调整 $k$ 值,得到最终客户群:除去强簇和表现良好的簇的强力牵引,在剥离的样本中,存在较原先表现更为良好的簇;但同时,这部分样本数量较少、个体差异较大,导致某些离群值对聚类的影响也会放大。因此,在新一轮聚类中,我们将原先剥离的3个簇按 $k=4$ 进行单独聚类(见图3)。最终,总共得到了9个簇,即9个客户类。

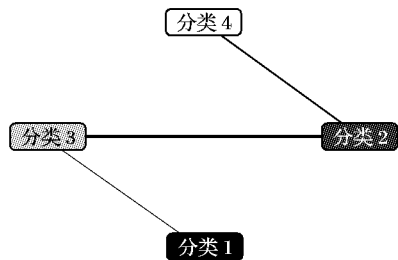


图3 聚类之间连接的关系图

## 2.3 聚类结果与分析

根据聚类结果,在对每个指标值进行统计的基础上,按照1.2节中的取值范围说明,可以得出9个客户类的特征描述(见表2)。

针对每类客户的特征描述,对类内样本单独做深入数据分析之后,提出以下具有针对性的营销策略(见表3),其中客户类按照重要程度排序。

## 3 结语

本文提出了一个具有证券行业特色的多目标多维度的证券客户细分模型,并利用动态聚类方法进行了实证分析,最终挖掘出具有现实意义的客户群,由此帮助券商制定出针对不同客户群的营销策略,把握有价值顾客,减少客户流失,实现业务利润的最大化。当然,在指标相关性研究及通过建立客户类型预测模型来识别数据不足的新客户类型等方面,还有待今后进一步深入研究。

表 2 聚类结果

分类 (样本数)	结果	价值分析				风险分析		偏好分析		
		贡献度	活跃度	收益率	客户资产	套牢度	风险度	行业	板块	种类
分类 1 (523)	指标值	-0.34	-0.21	-0.08	-0.10	0.04	0.78	0.11	0.62	0.74
	特征描述	低	很不活跃	平均以下	1 万~20 万元	低套牢	高风险	低	高	高
分类 2 (531)	指标值	-0.38	-0.28	0.17	-0.11	0.07	-1.34	0.57	0.15	0.84
	特征描述	低	很不活跃	平均以上	1 万~20 万元	中套牢	低风险	高	低	高
分类 3 (294)	指标值	0.158	-0.112	-0.016	0.071	0.029	0.775	0	0	0.19
	特征描述	高	不太活跃	平均以下	20 万~100 万元	不套牢	高风险	低	低	低
分类 4 (284)	指标值	-0.248	0.073	0.051	-0.098	-0.020	0.778	0	0	0.74
	特征描述	低	不太活跃	平均以上	1 万~20 万元	不套牢	高风险	低	低	高
分类 5 (99)	指标值	0.130	0.110	0.315	0.123	0.032	-0.906	0.24	0.24	0.21
	特征描述	高	不太活跃	平均以上	20 万~100 万元	低套牢	低风险	中	中	中
分类 6 (121)	指标值	0.68	1.12	-0.84	-0.05	-0.01	0.08	0	0	0.32
	特征描述	高	比较活跃	平均以下	1 万~20 万元	不套牢	中高风险	低	低	中
分类 7 (84)	指标值	-0.35	-0.25	0.44	-0.08	-0.55	-0.47	0.28	0.24	0.59
	特征描述	低	很不活跃	平均以上	1 万~20 万元	不套牢	中风险	中	中	高
分类 8 (49)	指标值	6.93	3.05	-0.31	0.66	0.02	-0.21	0.21	0.25	0.07
	特征描述	高	很活跃	平均以下	>100 万元	不套牢	中风险	中	中	低
分类 9 (26)	指标值	-0.09	-0.16	-0.35	3.16	-0.72	-0.54	0.43	0.45	0.56
	特征描述	高	不太活跃	平均以下	>100 万元	不套牢	低风险	中	中	高

表 3 客户分析与营销策略

客户类	客户分析	营销策略
分类 2	属睡眠客户,是最不关注客户	应尽量减少该类客户的比例
分类 1	虽贡献度不高,但为券商主要客户群	在缩减服务成本之上以引导为主,可适当安排面向大众的投资讲座
分类 4	属风险偏好型,潜在价值较高	控制成本和适度开发,可提供佣金优惠等措施增加其活跃度,使其发展成优质客户
分类 7	交易次数少但盈利能力强,属理性投资者,有一定专业知识,偏好中长线投资,潜在价值较高	投资建议应当是高质量的,甚至可以是技术类型的,以增加其交易量和活跃度
分类 5	属于稳健型投资者,盈利较高但不太活跃,属潜在优质客户	尽量提供稳健型的投资方案,在确保收益率的前提下,增加其投资积极性
分类 3	现有价值和潜在价值较高,又具有特殊的风险偏好,投资渠道少,多偏好中长线交易或短线投机	经纪人应当注重沟通,提供可靠的分析报告,使之操作更加果敢,提升其对经纪人的信任程度,使其发展成为核心客户
分类 6	公司利润的主要贡献者之一,偏好短线交易,投资不够理性,潜在流失风险大	应尽量吸引更多的这类客户加入,在流动性大的前提下维持这部分客户的期间总量,适当进行客户关怀,增加其忠诚度
分类 8	属核心优质客户	提升其潜在价值,尽可能延长客户生命周期,券商应做好客户沟通,提供方便的交易方式和场所以及适当的佣金折扣,甚至可以提供专业咨询公司的分析报告,配备专员进行一对一的营销,关注其偏好,提供优质投资建议,在硬件服务上,除大户室等常规服务之外,还可组织休闲娱乐健身活动
分类 9	拥有资金量很大,但是差异性也很大	可视作特殊离群客户做个别分析

参考文献:

[1] 宗传宏,聂祖荣,李波. 证券公司客户关系管理研究[J]. 商业研究, 2004(6): 125-127.

[2] 江晓东. 证券公司客户分类体系问题探析[J]. 内蒙古财经学院学报, 2004(1): 41-44.

[3] 江晓东. 基于证券营业部客户关系管理的统计指标体系[J]. 统计与决策, 2003(9): 08-09.

[4] 张国政. 客户关系管理中基于数据挖掘的客户细分研究[J]. 商业研究, 2006(13): 153-155.

[5] 段文文. 数据挖掘在证券业 CRM 中的实施研究[J]. 集团经济研究, 2006(4): 152-153.

[6] RYGIELSKI C, WANG J-C, YEN D C. Data mining techniques for customer relationship management [J]. Technology in Society, 2002, 24(4): 483-502

[7] 李斌,郭剑毅. 聚类分析在客户关系管理中的研究与应用[J]. 计算机工程与设计, 2005, 26(2): 540-542.

[8] 赵宏霞,杨皎平,陈宗娇. 面向客户需求细分的数据挖掘研究[J]. 科技管理研究, 2005(11): 207-210.

[9] 李卫东,宋威,杨炳儒. 利用数据挖掘方法分析客户生涯价值[J]. 计算机工程与应用, 2005, 41(6): 17-18, 36.

[10] 郭蕴华,陈定方. 基于模糊聚类分析的客户分类算法研究[J]. 计算机应用研究, 2005, 22(4): 52-53.

[11] 宋艳,梁静国. 基于模糊聚类的客户分类应用研究[J]. 物流科技, 2005, 28(1): 26-28.

[12] 罗本德,彭小兵. 股票投资群体动态聚类研究[J]. 经济体制改革, 2005(3): 108-111.

[13] 谢寰红. 数据挖掘在证券公司 CRM 客户细分的应用[J]. 计算机工程, 2004, 30(B12): 553-554, 585

[14] SHIN H W, SOHN S Y. Segmentation of stock trading customers according to potential value[J]. Expert Systems with Applications, 2004, 27(1): 27-33.

[15] BERRY M J A, LINOFF G S. 数据挖掘技术[M]. 别荣芳,尹静,邓六爱,译. 北京:机械工业出版社, 2006: 134-139.

[16] TANG Z H, MACLENNAM J. 数据挖掘原理与应用——SQL Server2005 数据库[M]. 邝祝芳,焦贤龙,高升,译. 北京:清华大学出版社, 2007: 98.