

# 基于聚类分析的银行客户关系管理策略研究

肖 凤,郑海健,卢 闯

(重庆大学 经济与工商管理学院,重庆 400030)

**摘 要:** 本文将数据挖掘聚类技术应用于银行客户分类,分析了银行交易特征数据;构建了数据模型,对 60 万条样本数据进行了 K-means 聚类;采用轮廓系数对聚类结果进行了评估,并基于聚类结果进行了客户二维分类,据此制定出更有效的客户关系管理策略。

**关键词:** 数据挖掘;聚类分析;银行客户分类;客户关系管理

**中图分类号:** F270.7 **文献标识码:** A **文章编号:** 1002-980X(2010)01-0087-07

银行业是典型的以客户为导向的服务行业<sup>[1]</sup>。信息化程度的提高使中国金融业获得了大量客户电子数据资源,而客户数据信息的理解和利用对提高银行服务水平具有重要作用<sup>[2]</sup>。传统的技术支持已无法满足用户的要求,数据挖掘技术是对海量数据进行处理,从不完整的、有噪声、模糊的数据中提取隐含在其中的信息<sup>[3]</sup>。应用数据挖掘技术可将银行市场和客户分成有意义的群组,协助银行更好策划活动和设计新的市场运动<sup>[4]</sup>,通过获得客户类别来分析并预测客户的消费模式<sup>[5]</sup>。准确的客户分类使企业能有效实施客户关系管理策略。因此,研究数据挖掘技术,对客户进行分类,根据分类结果制定 CRM 策略具有重要的理论价值与现实意义。

本文以某银行客户一年交易数据为依据,无放回抽样基础上使用数据挖掘 K-mean 聚类方法建立模型,进行轮廓系数的评估。基于聚类结果进行银行客户二维分类研究,解决传统统计分析中海量数据噪声问题和判断欠准确性问题,有很强适应性。与银行原有的以经验为主的分类方法相比较,提高了客户正确识别率。基于研究结果针对性地提出客户关系管理策略建议。

## 1 文献回顾

金融行业的竞争主要是对客户的争夺,而对客户数据信息进行挖掘,并在此基础上分类出不同价值客户,对服务质量的提高和企业的成败有决定作用<sup>[6]</sup>。数据挖掘技术工具能有效分析客户信息,降低分类风险<sup>[1]</sup>。如果面对当前的市场状况,银行能建立一个有效的客户划分模型,这样便可以更个性

化地服务于客户,降低有关成本<sup>[6]</sup>。所以,如何有效对客户进行分类就成了制定客户关系管理策略的重要课题。在应用数据挖掘技术进行客户关系管理问题上,检索发现,国内外相关研究成果主要有:

Mellon 银行用 IBM 公司的 Intelligent Miner 数据挖掘技术对信贷客户进行信用评估<sup>[7]</sup>,将其分为低、中、高危险者,对每类客户量身定做金融产品,实现了“留住优质客户,争取有潜力的新客户”,做到了以客户为中心。Firststar 银行使用 Masksman 数据挖掘工具,根据客户消费模式预测应在何时为客户提供何种产品<sup>[8]</sup>。基于聚类算法的相关研究有,赖玉霞使用遗传算法提高聚类质量,使每类客户具有更相近的交易模式,从而更好地制定企业服务策划<sup>[9]</sup>。刘满芝使用因子分析统计聚类,于曼曼使用 ANOVE 方法研究电子商务,针对客户对服务的感觉对在线商品进行分类,对客户进行个性化服务<sup>[10]</sup>。湛宁改进传统的决策树分类挖掘算法即 ID3 算法<sup>[11]</sup>,对多个分类分别构建决策树,标明分类概率对金融客户进行一维分类。胡春对旅游公司的客户进行分类,预测哪类客户将参加哪类旅行团<sup>[12]</sup>。基于神经网络算法的有:陈伯成用自组织映射神经网络(self organization map, SOM),综合客户的利润贡献和行为倾向划分客户等级和类型,建立一个模型来预测未来可能流失的这类客户<sup>[13]</sup>。

基于聚类、决策树、神经网络等算法的研究中大多是对于研究方法和模型的改进,而且在客户分析应用上进行的几乎都是客户一维分类。为此,本文在前人研究基础上,以 XX 商业银行客户分析作为研究对象,运用基于原型的聚类方法对客户进行二

收稿日期:2009-09-17

基金项目:重庆市自然科学基金资助项目(CSTC,2008BB2041)

作者简介:肖凤(1982—),女,内江人,重庆大学经济与工商管理学院硕士研究生,研究方向:情报学、数字资源处理与利用、信息管理与系统;郑海健(1985—),男,安徽黄山人,重庆大学经济与工商管理学院硕士研究生,研究方向:企业管理;卢闯(1969—),男,辽宁抚顺人,重庆大学经济与工商管理学院在职工商管理硕士。

维分类,在银行客户关系管理策略制定上进行实证研究。

## 2 研究方法

本文客户分类应用数据挖掘技术聚类原理 (data mining, DM), 从大量的、不完全的、有噪声的、随机的数据中提取隐含在其中事先不知道的有用的信息,将收集的数据进行聚类 (clustering) 分析,把一组个体按照相似性归成若干类别,使得属于同类别的个体之间的距离尽可能小,不同类别个体之间的距离尽可能大<sup>[14]</sup>。本文针对 XX 银行客户现状进行分析,抽样提取数据,进行 SAS Enterprise Miner (以下简称 SAS/EM) 聚类 and 轮廓系数评估,对客户聚类结果进行细分,建立差异化的服务模式提供客户关系管理策略。采用 SAS/EM 为数据挖掘工具平台。

### 2.1 研究框架

如图 1 所示,采用 SAS/EM 为数据挖掘工具平台,建立本文的研究框架。

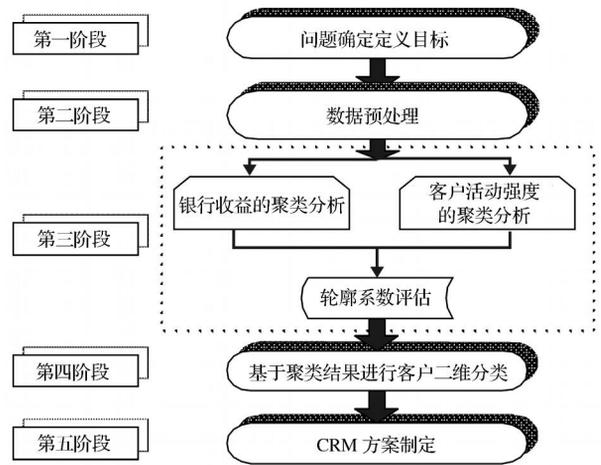


图 1 基于聚类的银行客户分类模型研究框图

### 2.2 数据模型

数据模型是建立客户聚类数据挖掘模型的前提和条件。在银行收益数据模型中的变量有两类: 客户代码,代表客户名称; 银行收益数据,主要包括每月的账户余额、每月 POS 消费情况、每月存取款情况、每月交费业务情况等乘以银行的收益率汇总后得到各季度的收益数据。数据模型结构如表 1 所示。

表 1 银行收益数据模型

序号	属性名	属性含义	属性类型
1	客户代码	能唯一识别客户的编码。	ordinal
2	第一季度收益	某个客户在第一季度为银行带来的收益额。	interval
3	第二季度收益	某个客户在第二季度为银行带来的收益额。	interval
4	第三季度收益	某个客户在第三季度为银行带来的收益额。	interval
5	第四季度收益	某个客户在第四季度为银行带来的收益额。	interval

本数据模型表示某客户在每季度的不同交易行为为银行带来的收益,其中客户代码代表银行在这段时间里随机抽样出来的客户;每一季度收益表示银行抽样出的某客户在第每季度为银行带来的收益金额,包括存款收益、缴费收益、扣费收益、网上消费

收益、POS 消费收益等。

在银行客户活动强度数据模型中输入变量有两类,一类是客户代码,代表客户名称,另一类是交易强度数据。数据模型结构如表 2 所示。

表 2 客户活动强度数据模型

序号	属性名	属性含义	属性类型
1	客户代码	能唯一识别客户的编码。	ordinal
2	第一季度交易强度	某客户第一季度在银行交易的次数合计,主要包括存款次数、网上消费次数、POS 消费次数、取款次数、缴费次数等。	interval
3	第二季度交易强度	某客户第二季度在银行交易的次数合计。	interval
4	第三季度交易强度	某客户第三季度在银行交易的次数合计。	interval
5	第四季度交易强度	某客户第四季度在银行交易的次数合计。	interval

客户活动强度的数据模型用于表示某客户每季度在银行的交易次数行为特征,反映该客户在银行的活动强度。客户代码代表银行在制定时间里随机抽样出来的客户;每季度交易强度表示抽样出的某客户第一季度在银行所有交易次数的合计,其中包括存款次数、网上消费次数、POS 消费次数、取款次

数、缴费次数等所有交易活动的次数。

### 2.3 聚类方法

1) K-means 算法。

K-means 能有效处理类内密集类间区别明显的的数据,其时间复杂度为  $O(nkt)$  (其中  $t$  是迭代次数),因此有相对较高的可伸缩性和高效率。K-

means 的实现如下:

假定  $N_i$  为第  $i$  类聚类  $C_i$  中的样本数目,  $m_i$  是样本的均值。若分成  $k$  类, 则各聚类的均值表示为:

$$m_i = \frac{1}{N_i} \sum_{p \in C_i} p \quad (1)$$

误差平方和为:

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (2)$$

式(2)中:  $E$  是观测数据中所有对象平方误差综合;  $P$  是空间中的点;  $m_i$  是聚类  $C_i$  的平均值 ( $P$  和  $m_i$  是多维的)。对于不同的分类,  $K$  值不同, 则  $E$  值就不相同, 使  $E$  达到最小的分类数  $K$  认为是当前的目标数<sup>[15-16]</sup>。

### 2) 轮廓系数评估。

本文采用轮廓系数来评估聚类效果。轮廓系数方法结合了凝聚度和分离度来评价聚类优良的总度量。簇的凝聚度 (cluster cohesion) 用于度量簇中对象密切关系。

$$cohesion(C_i) = \frac{x}{y} \frac{c_i}{c_j} proximity(x, y) \quad (1)$$

分离度 (cluster separation) 用于度量确定一个簇不同于其他簇。<sup>[17-19]</sup>

$$separation(C_i, C_j) = \frac{x}{y} \frac{c_i}{c_j} proximity(x, y) \quad (2)$$

计算个体点的轮廓系数:

对第  $i$  个对象, 计算它到簇中所有其他对象的平均距离。记  $a_i$ 。

对第  $i$  个对象和不包含该对象的任意簇, 计算该对象到给定簇中所有对象的平均距离。关于所有簇, 找出最小值, 记  $b_i$ 。

对第  $i$  个对象, 轮廓系数为  $s_i = (b_i - a_i) / \max(a_i, b_i)$ 。

## 3 案例分析

对 XX 银行重庆分行随机抽取客户样本, 并提取样本客户 2008 年全部的交易数据。用 EXCEL 软件对数据进行预处理, 把干净的数据导入 SAS/EM, 对样本数据进行多次聚类, 选择评估结果较好的一次作为聚类结果, 基于聚类结果进行客户二维分类。

### 3.1 客户数据

以 XX 银行提供的个人储值卡客户历史数据来构建客户分类模型。在该行中心数据库中从 2008 年 1 月至 12 月随机抽取 6079 个客户的 60 万条交易数据。对初始数据按表 1 和表 2 所示数据结构的要求进行了各种统计和汇总处理。

### 3.2 数据预处理

1) 基于银行收益的数据预处理。每个客户每次交易记录都记载在原始数据中, 有多少次交易客户名就会重复多少次。必须进行属性约简。同一客户的交易记录在一个季度中只出现一次。

通过季度交易总额表计算出收益表。根据银行每项的收益率计算出该项的银行收益金额。为了减小样本数据间的差异性对数据进行归一化处理, 处理后的收益表数据模型节选见下表 3 所示。

表 3 归一化银行收益表

客户	一季度收益金额	二季度收益金额	三季度收益金额	四季度收益金额
p1	0.0836	0.0054	0.0134	0.0571
P2	0.0543	0.0384	0.0195	0.0421
P3	0.0161	0.00914	0.0121	0.0084
P5	0.0075	0.0054	0.0071	0.0047
.....	.....	.....	.....	.....
P6079	0.0017	0.0141	0.0036	0.0014

2) 基于客户活动强度的数据的预处理。原始数据中, 记录了客户在 2008 年 12 个月中的每次交易记录, 以季度为时间段, 得出每季度交易次数的合计。对基于客户活动强度的数据进行归一化后, 模型节选如表 4 所示。

表 4 归一化客户活动强度表

客户	一季度客户活动强度	二季度客户活动强度	三季度客户活动强度	四季度客户活动强度
P1	0.1112	0.0258	0.0432	0.0243
P2	0.0764	0.0845	0.0384	0.0197
P3	0.0342	0.0254	0.0403	0.0101
P4	0.0094	0.0024	0.0045	0.0045
.....	.....	.....	.....	.....
P6079	0.0746	0.1610	0.0235	0.0874

### 3.3 模型建立

建立 SAS/EM 数据分析模型, 对预处理好的数据进行聚类分析和评估。

在 SAS/EM 中, 建立“SEMMA”分析流程图, 将处理好的数据导入 SAS/EM, 进行聚类分析, 见图 2。

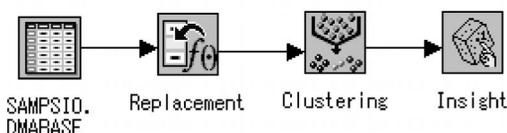


图 2 “SEMMA”流程图

1) 银行收益数据聚类分析。通过对聚类参数不断调整, K-means 聚类对数据测量标准非常敏感<sup>[20]</sup>, 数据先去掉缺省值再进行标准化 (stdDev.) 处理。首先, 确定聚类数。根据银行以往对客户的分类等级的分析和总结, 实验阶段选取聚类数 10、

8、6、3,通过不同的聚类结果进行比较分析和轮廓系数对聚类结果的评价,当聚类数为 6 的时候轮廓系数为 0.53 最大。因数据太分散,聚类结果中第 4 类含客户量占总客户的 87%,不符合银行现实分类比例情况,因此把第 4 类拿出进行第二次聚类。表 6 二次聚类结果中当类  $K$  为 6 的时候,轮廓系数取值最大为 0.713,结果较为满意也符合银行实际要求。

表 5 银行收益一次聚类结果

Cluster	Frequency of Cluster	Root-mean-power Deviation	Maximum Distance from Cluster Seed	Nearest Cluster
1	6	12.047361446	37.618711225	2
2	61	3.1414056448	15.373779127	5
3	2	3.1274425082	4.4221056348	2
4	5568	0.1648231507	2.5199482449	5
5	438	1.0258572564	6.358469177	4
6	2	6.7310451898	9.5174868231	2

表 6 二次聚类结果 output 表

Cluster	Frequency of Cluster	Root-mean-power Deviation	Maximum Distance from Cluster Seed
1	54	1.6380277033	8.879758704
2	3946	0.2494511527	2.1827129616
3	91	1.5710091302	11.39585013
4	83	1.3454796072	8.4317031605
5	1051	0.7274778259	3.6787334364
6	343	1.1720186904	6.491522839

2)对客户活动强度的数据进行聚类分析。根据银行对客户活动强度的分类等级的分析和总结,选取聚类数  $K$  为 15、8、6、3,当  $K$  为 15 时轮廓系数最大为 0.762,聚类结果满意。

表 7 聚类结果 output 表

Cluster	Frequency of Cluster	Root-mean-power Deviation	Maximum Distance from Cluster Seed	Nearest Cluster
1	23	2.4664811683	9.2563721332	11
2	3	11.013757472	25.282936474	1
3	308	0.4242902612	2.1948133714	15
4	18	2.6081086966	17.129582297	11
5	2237	0.103229775	0.7073846757	7
6	196	0.6029245412	4.1995173917	15
7	1194	0.1544685533	1.3000597999	5
8	1	-	24.85507533	4
9	739	0.2293677258	1.597408397	7
10	174	0.7230081532	4.2310274267	6
11	108	1.0312470758	4.6252125233	10
12	487	0.2580808101	1.3559059408	7
13	103	0.5703573814	2.6936531836	15
14	26	1.1767390749	3.6908110101	13
15	461	0.3222274493	1.6712122994	9

3.4 聚类结果评估

1)基于银行收益聚类的评估。

由于采用的是非监督方式的聚类,聚类结构的优良性度量不考虑外部信息,无法确定类的数目<sup>[21]</sup>,因此采用轮廓系数来评估分类效果。轮廓系

数的计算先考虑各个分类数  $K$ ,每个样本点被分到适当类中的效率为  $s_i$ ,然后对  $s_i$  求平均值  $s_k, s_k$  的值越接近 1 越说明把样本分成  $k$  类的效果越好<sup>[13-15]</sup>。表 8 显示了轮廓系数的实现过程,表 9 显示了轮廓系数与聚类效果的关系。

表 8 轮廓系数的实现过程

$\forall k, k \in [N_L, N_U], N_L$ and $N_U$ are the lower and upper bounds of the numbers of clusters
$\forall X, X$ is a cluster in $k$ clusters
$\forall i, j \in X, i$ and $j$ are data in cluster $X$
$a_i = d(i, j)$ is the average dissimilarity of $i$ to all data $j$ in cluster $X$
$b_i = \min\{d(i, j)   j \in X, j \neq i\}$
$s_i = \frac{b_i - a_i}{\max\{b_i, a_i\}}$ is the silhouette of data $i$
$s_x = \text{average}(s_i)$ is the average silhouette of cluster $X$
$s_k = \text{average}(s_x)$ is the average silhouette width for cluster $k$
$SC = \max_k s_k$ is the silhouette coefficient for the clustering result

表 9 轮廓系数与聚类效果关系

轮廓系数 $s_k$	聚类效果	轮廓系数 $s_k$	聚类效果
0.71 ~ 1.00	聚类是高效率的,令人信服	0.26 ~ 0.50	聚类缺乏说服力
0.51 ~ 0.71	聚类合理	0.25	没有任何实际的分类

对银行收益数据进行聚类时,一次聚类  $K$  为 6 时,轮廓系数最大值为 0.53,二次聚类  $K$  为 6 时,轮廓系数最大为 0.713。对银行客户活动强度聚类时,当  $K$  为 15 时,轮廓系数最大为 0.762,聚类结果满意。

## 4 客户分类与客户关系管理策略

### 4.1 客户分类

根据银行实际需求把银行收益聚类结果和客户活动强度聚类结果按照类的标准均值对比图特征进行类类组合分别划分为高价值客户类、中价值客户类、低价值客户类和强活跃客户类、中度活跃客户类、弱活跃客户类。

以银行收益第二次聚类中的第二类 Cluster2-2 和客户活动强度聚类结果中的 Cluster10 的标准化均值对比图为例进行特征分析类类组合,如图 3、图 4 所示。

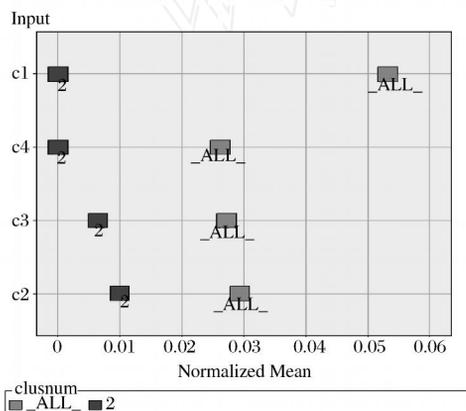


图 3 银行收益 cluster2-2 标准化均值对比图

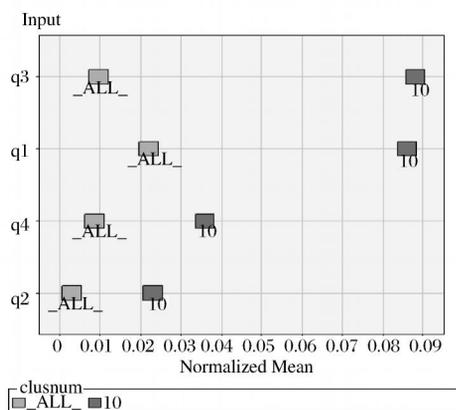


图 4 客户活动强度 cluster10 标准化均值对比图

图 3 中为“2”的方块表示 Cluster2-2 这类客户给银行带来收益的平均金额,位 ALL 的方块表示所有客户给银行带来收益的平均金额。从图 3 中可以看出该类客户给银行带来的平均收益远小于总客户的平均收益。因此该类客户归为低价值客户,即为银行带来收益很小或几乎收益为零的一类客户。因此 Cluster2-2 归为低价值客户。同理 Cluster1、2、3、6、2-1、2-3、2-4、2-6、5 按照标准化均值对比图归为高价值客户类。Cluster2-5 归为中价值客户。图 4 中该类客户在给银行交易的平均强度远大于每个客户在银行的平均交易强度。Cluster1、2、4、6、8、10、11、13、14、15、与 Cluster10 有相似特征,把这些类归为强活跃客户类。同理:Cluster3、9 归为中度活跃客户类,Cluster12、7、5 归为弱活跃客户类。

### 4.2 客户分类与银行现有分类情况对比分析

基于聚类结果的客户分类与原有经验为主的客户分类结果有明显差异,具体情况如下:

客户价值分类。表 10 显示了银行现行客户分类和按客户交易信息进行聚类划分后的对比结果。

表 10 基于聚类的客户分类与银行现行分类对比分析表

银行现行对 6079 个样本客户分类	聚类结果与银行现行分类对比	聚类结果与银行现行分类的匹配程度(%)
高价值客户 1081 位	银行现行划分的 1081 位高价值客户按聚类后的划分	48
普通客户 4999 位	银行现行划分的 4999 位普通价值客户按聚类后的划分	88.7

银行现行分类是按客户办卡填表信息定位不同价值客户(如客户在什么单位,客户的职称等),但这类客户实际能否为银行带来收益,带来收益的多少

都是不确定的。根据客户的交易数据进行客户聚类分析来确定什么客户能够给银行带来多少收益因此来定位客户更加科学。

交易强度分类。交易强度是是客户在银行的交易行为特征的重要因素,客户活动频率的高低一定程度上也反应客户对银行的满意度等,但原有分类没有给与考虑。

### 4.3 客户关系管理策略

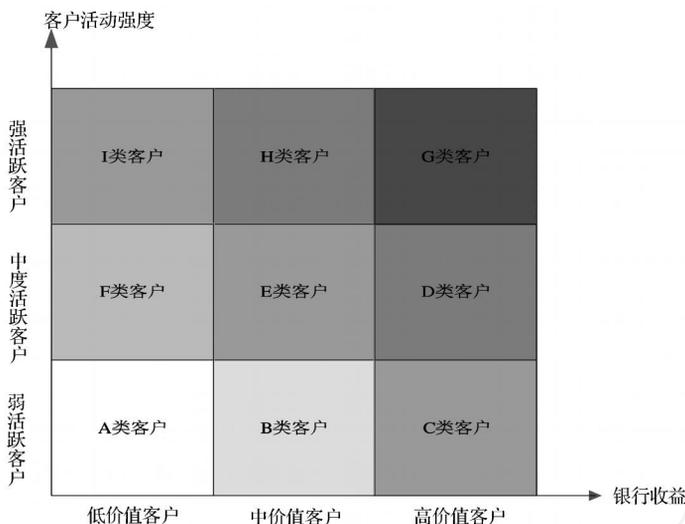


图 5 客户二维分类模型图

客户二维分类后形成 9 类客户群,如 G 类客户是强活跃高价值客户是银行重点服务对象。A 类弱活跃低价值客户为银行尽量淘汰的客户类。银行可为每类客户制定不同的 CRM 策略。以下以 G 类 A 类客户为例进行说明。

G 类客户(高活跃高价值客户)。这类客户交易直接影响银行的收益和发展。银行对他们实施个性化客户关系管理策略是:免年费、保险服务、持卡消费积分奖励、各种形式的抽奖、持卡会员服务等等一系列优惠措施和附加服务。具体措施如下:

1) 多种市面上商业活动享受贵宾折扣优惠,与商场、保险、饭店等消费机构合作,只要客户使用 XX 银行 VIP 贵宾卡支付即可享受优于中价值客户和低价值客户的折扣,从而刺激银行卡消费。

2) 从数据库中采集到客户生日信息,定期从这类客户中抽出一定量的客户,在他们生日当天送去蛋糕和鲜花。银行还定期搞回馈活动,在过节时送红酒,月饼或者让客户的小孩免费参加高尔夫球培训。

3) 附加服务手段,如对于航空旅行中的贵宾礼遇服务、停车优惠服务等。这类客户坐飞机时持有该银行的 VIP 卡可以在登机检票时从快速通道进行检票,节约了客户登机的时间。

4) 研究调查表明高价值客户不愿意排队等候时间超过 5 分钟,因此这类客户在银行的存取款数量相对较大,为了节省他们在大厅排队等候的时间银

通过划分出不同客户类型,确定要服务的对象进行差异化服务。以收益为横坐标,活动强度为纵坐标构建客户二维分类模型。以下为客户二维分类模型图 5。

行为他们提供了专门的大客户室,可以安全快捷的进行任何银行业务。

5) 刷卡消费计积分,客户可以在年底按积分兑换礼品,如家电用品、玩具、饭店住宿餐饮的优惠礼卷、航空机票较高价值的奖品等。

6) 银行与该客户建立伙伴型关系。客户实施“一对一”工作机制,由高级客户经理进行专门维护。客户有任何问题或不满意,银行要在最短时间内进行解决。把主要精力用在这类客户身上。

A 类客户(低价值弱活跃客户):行为特征是客户活跃度很低,甚至几乎长时间不产生任何交易,给银行带来的价值也很低,或者价值为零。这类客户虽然持有银行的卡,但很可能没有进行使用。对于这类客户银行可考虑通知该类客户销卡或将该类客户归入睡账户中,以减少主机系统资源的占用。

## 5 结论

本文通过对某银行客户数据进行数据挖掘聚类分析,取得了以下一些成果: 基于 XX 银行重庆分行客户交易数据,建立模型,用数据挖掘聚类技术进行客户分类,优化了银行客户分类方法。应用 K 均值进行聚类分析,轮廓系数进行评估,聚类结果可接受性强。聚类结果和银行现行客户分类存在差异,评估证明聚类后的客户分类更符合实际情况,使客户分类更合理、客户行为特点分析更加有效。

对不同类型的客户分别制定相应的客户关系管理

策略,使得客户关系策略实施性更强,以此来增加该银行的竞争力,更好地保留客户。

### 参考文献

- [1] ASPINALL E,NANCARROW C,STONE M. The meaning and measurement of customer retention[J]. Journal of Targeting Measurement and Analysis for Marketing, 2001,10(1):79-86.
- [2] CARDELL N S,GOLOVNA YA M,STEINBERG D. Churn modeling for mobile telecommunications: Winning the Duke/NCR Teradata center for CRM competition [EB/OL]. 2003,http://www.salford-systems.com.
- [3] BREIMAN L ,FRIEDMAN J ,OLSHEN R ,et al. Classification and regression trees [ R ]. Pacific Grove: Wadsworth,1984.
- [4] DAYKJN C D,BERNSTEIN G D. A simulation model to examine questions of solvency in the light of asset and run-off risks[ R ]. Biarritz:ASTIN Colloquium,1985.
- [5] SHARPE W F. Capital asset prices:a theory of market equilibrium under conditions of risk[J]. Journal of Finance, 1964(19):341-360
- [6] WEI C P,CHIU I T. Turning telecommunications call details to churn prediction:a data mining approach[J]. Expert Systems with Applications,2002,23:103-112.
- [7] COLGATE M,STEWART K, KINSELLA R. Customer defection:a study of the student market in Ireland[J]. International Journal of Bank Marketing, 1996,14(3):23-29.
- [8] AHMAD R,BUTTLE F. Retaining telephone-banking customers at frontier bank[J]. The International Journal of Bank Marketing, 2002,20(1):5-16.
- [9] WANG,HSIAO F. Managing customer profitability in a competitive market by continuous data mining[J]. Business Source Elite,2006,35(6):715-723.
- [10] FIONA C. Mining customer data into intelligence[J]. Telecom Asia,2007(3):24-27.
- [11] 蒙肖莲,蔡淑琴,杜宽旗,寇建亭. 商业银行客户流失预测模型研究[J]. 系统工程,2004(12):67-71.
- [12] 钟晓. 数据挖掘综述[J]. 模式识别与人工智能,2001,14(1):48-54.
- [13] PANNING T,MICHAEL S,VIPIN K. 数据挖掘导论[M]. 北京:人民邮电出版社,2006.
- [14] 尉景辉,何丕廉. 基于 K-Means 的文本层次聚类算法研究[J]. 计算机应用,2005(10):2323-2324
- [15] 宋青粉,李文璋. 基于 SOM 网络的服装号型设置[J]. 国际纺织导报,2008(5):70-80.
- [16] 郭景峰,赵玉艳. 基于改进的凝聚性和分离性的层次聚类算法[J]. 计算机研究与发展,2008,45(S1):202-206.
- [17] 杨兰,卢润德. 基于客户价值的客户分类方法研究[J]. 现代管理科学,2007(11):95-96.
- [18] 谭元戎,孙剑平. 聚类模型在客户关系管理中的应用以及对特征提取的探讨[J]. 技术经济,2007,26(5):51-56.
- [19] HAN J W,KAMBER M. 数据挖掘[M]. 机械工业出版社,2007:108.
- [20] 张云涛,龚玲. 数据挖掘原理与技术[M]. 北京:电子工业出版社,2004:171-172.
- [21] 唐志航,杨保安. 系统聚类在客户关系管理中的研究与应用[J]. 计算机工程与应用,2007,43(13):220-223.

## Study on CRM of Bank Based on Clustering Analysis

Xiao Feng,Zheng Haijian,Lu Chuang

(School of Economics and Business Administration,Chongqing University,Chongqing 400030,China)

**Abstract :** This paper applies the technology of clustering data mining to the classification of bank's customers ,and analyzes the feature data of bank transaction. Then it builds a data model ,and makes the data K-means clustering for 600 thousand pairs of samples. And it uses the contour coefficient to assess the clustering results ,based on which ,it makes the two-dimensional classification for bank's customers basing on the clustering results. Finally ,it puts forward more effective strategies on customer relationship management of bank.

**Key words :** data mining ;clustering analysis ;bank's customer classification ;CRM