

# 客户关系管理中的动态客户细分方法研究<sup>①</sup>

叶 强, 卢 涛, 闫相斌, 李一军  
(哈尔滨工业大学管理学院, 哈尔滨 150001)

**摘要:** 在分析客户行为的随机性和非确定性的基础上, 指出现有的确定性客户细分方法不能很好地适应客户细分问题的这些特点. 为此, 提出基于云模型的动态客户细分模型, 该模型将客户细分过程表示为一个  $C$  过程与一个  $P$  过程, 并将描述非确定关系的云模型理论引入到客户细分的  $P$  过程中, 从而实现了客户细分的动态性, 提高了模型对客户行为描述的客观性. 文章采用来自 UCI 的合成数据及来自银行的实际客户数据进行了数据实验, 实验结果表明了该方法的有效性.

**关键词:** 客户关系管理; 动态客户细分; 云模型

**中图分类号:** TP18

**文献标识码:** A

**文章编号:** 1007-9807(2006)02-0044-09

## 0 引 言

从 20 世纪 90 年代, 伴随着以客户为中心的管理思想的发展, 由 Gartner Group 提出的客户关系管理 (customer relationship management, CRM) 在研究与应用上受到了广泛的重视<sup>[1,2]</sup>. CRM 包含了三个层次的含义, 首先, CRM 代表了一种新的管理理念; 其次, CRM 代表了一种新的管理机制; 最后, CRM 又代表了一种新兴的企业信息系统. 作为管理信息系统的 CRM 目前主要有操作型、分析型和综合型等 3 种类型. 随着管理理论与管理手段的发展, 深入的客户数据分析将在管理中发挥越来越重要的作用, 因此分析型或综合型 CRM 已经成为一种发展趋势. 作为客户关系管理的核心概念之一, 客户细分已成为 CRM 的重要组成部分和基础性的分析功能<sup>[3]</sup>, 并将对包括操作层、战术层以及战略层在内的企业管理提供全面的信息支持. 正如 2002 年哈佛商业评论的一篇文章所说, 任何高效的客户关系管理都将以扎实的客户细分为基础<sup>[4]</sup>. 客户细分 (customer segmentation), 是企业在明确的战略、业务模式和特定的市场中,

根据客户的属性、行为、需求、偏好以及价值等因素对于客户进行分类, 并提供针对的产品、服务和营销模式的过程<sup>[2]</sup>. 客户细分的概念来源于 20 世纪 50 年代出现的市场细分概念, 目前已经成为广泛使用的管理工具之一. 2002 年, 美国的管理咨询公司 Bain & Company 对来自世界各地的 451 位高级管理人员的一项调查表明, 客户细分工具已经成为全球使用率最高的 10 大管理工具之一<sup>[5]</sup>. 目前, 常用的客户细分方法主要有: 经验描述法, 传统统计法, 非传统统计法等<sup>[6-8]</sup>. 最初的客户细分是由经验描述法开始的, 这种方法一般由决策者根据以往的经验对客户进行类别划分, 因此具有较强的主观性; 基于传统统计方法的客户细分一般是根据对客户属性特征的简单统计来划分客户类别; 非传统统计方法主要是近年来兴起的数据挖掘等基于人工智能技术的非数值计算方法. 国内外对客户细分的研究主要集中于用现有的分类方法在某些具体行业与领域的应用<sup>[9]</sup>. 这些领域既包括传统的旅游业、餐饮业、零售业、传媒业, 也包括新兴的网络营销、网络服务及其它电子商务领域. Jaesoo Kim 等研究了神经网络在旅游业客

① 收稿日期: 2003-12-26; 修订日期: 2005-10-14.

基金项目: 国家自然科学基金资助项目(70501009); 黑龙江省自然科学基金资助项目(C0304).

作者简介: 叶 强(1972—), 男, 黑龙江尚志人, 博士, 副教授.

户细分中的应用<sup>[10]</sup>, Vellido 等研究了神经网络在网上客户细分的应用<sup>[11]</sup>, Heilman 等研究了利用客户购买数据进行客户细分的问题<sup>[12]</sup>. 这些研究在解决客户分类问题中所使用的数据挖掘分类模型多为确定性分类模型, 即条件属性相同的对象被确定地分入某一特定类; 同一待分类模式用同一个分类模型进行多次分类时, 结果也将是一成不变的. 鉴于客户行为的复杂性、不确定性及现有方法在解决客户细分问题上的不足, 本研究将提出一种建立在云模型思想基础上的动态客户细分模型, 用于解决客户细分中的非确定性问题.

## 1 问题的提出

用现有的分类方法进行客户细分时, 一般得到的是确定的客户细分结果, 即(1)对于一个特定客户, 其所属类别是确定的、非此即彼的; (2)同一个客户的数据, 用同一个分类器进行多次分类, 每次得到的结果将是相同的; (3)两个属性数据相同

的客户, 分类结果必然是相同的.

客户细分中得到确定的结果, 主要原因是现有分类模型都是确定性分类模型. 根据确定的分类模型, 如果一个对象满足某一类的分类条件, 就将被绝对地划分到该类中.

例 1 根据月收入来区别高收入(I类客户)、中等收入(II类客户)及低收入(III类客户)3类客户, 一个分类模型可以简单表示如下

$$\begin{aligned} \text{income\_class}(x) \\ = \begin{cases} A \text{ 类客户, } \text{salary}(x) \geq 5000 \\ B \text{ 类客户, } 5000 > \text{salary}(x) \geq 1000 \\ C \text{ 类客户, } \text{salary}(x) < 1000 \end{cases} \end{aligned}$$

按照这个标准, 如果一个客户的月收入为 4 999 元, 那么他将被硬性地划归为 B 类客户.

即使有些分类方法采用基于隶属函数和隶属度的模糊方法, 然而一旦隶属函数和隶属度确定, 那么分类结果仍将被“硬化”.

例 2 现有一个具有  $m$  个条件属性  $\{x_1, x_2, \dots, x_m\}$  的客户 X 的数据经某一模糊分类器(隶属函数)计算后得出图 1 所示的隶属度结果.



图 1 基于模糊理论的客户细分

Fig. 1 Customer segmentation based on fuzzy theory

X 属于 I 类客户、II 类客户和 III 类客户的隶属度分别为 0.49, 0.82, 0.83. 由于  $\max\{Y_i(x)\} = Y_3(x) = 0.83$ , 因此该客户将被确定地划分为 III 类客户.

然而, 由于受心理和社会等复杂因素的影响, 客户行为往往存在很大的非确定性和随机性, 而确定性的分类模型无法客观地描述客户的这种行为特征. 因此, 在客户关系管理的客户细分中, 采用确定性分类模型, 会存在一定的不足. 这些不足表现如下.

(1)一旦一个客户被归入某一类, 就可能永远失去了对该客户适用其它客户策略的机会. 如例 1 中月收入为 4 999 元的客户, 其月收入很接近 I 类客户, 因此该客户的行为模式很可能也接近于 A 类客户的行为模式. 如果应用传统的分类方法, 在其属性数据与分类模型不变的情况下, 企业可

能会一直对其使用 II 类客户策略, 而失去尝试使用 I 类客户策略的机会; (2)对于同一类别中的客户没有区分不同客户属于该类的强度差异. 如一个客户经过例 2 模型处理后, 得到  $\{0.4, 0.5, 0.3\}$ , 另一个客户的数据经处理后得到隶属度  $\{0.01, 0.9, 0.1\}$  两个客户都将被归入 II 类, 但是从隶属度的差异可以判断, 这两个客户属于 II 类的程度将存在一定的差异.

鉴于现有的分类方法没有解决客户细分中存在的上述问题, 本研究借鉴云模型的思想, 提出动态分类模型.

## 2 云模型的基本思想

云模型作为一种描述非确定性关系的数学方

法,最早由李德毅<sup>[13-15]</sup>等人提出,它把模糊性和随机性相结合,构成定性描述和定量描述间的相互映射。

设  $U = \{u\}$  是一个普通集合,称为论域。 $T$  是论域上的概念,论域  $U$  中的元素  $u$  对  $T$  的隶属程度  $C_T(u) \in [0,1], C_T(u) \rightarrow [0,1], \forall u \in U, u \rightarrow C_T(u)$  是一有稳定倾向的随机数。概念  $T$  的云模型  $C$  是从论域  $U$  到区间  $[0,1]$  的映射。

文献<sup>[13]</sup>提出了基于正态分布的正态云模型。该模型可以用一个三元组进行描述,

$C \langle E_x, E_n, H_e \rangle$  其中,  $E_x, E_n, H_e$  分别称为期望值、熵及超熵(该参数反映熵的分布状态)。

$$MEC(u) = e^{-\frac{(u-E_x)^2}{2E_n^2}}$$

$$E_n = f(H_e)$$

期望值  $E_x$  反映概念的集中性;熵  $E_n$  反映概念的离散性;超熵  $H_e$  反映概念的离散性。

如果用云模型表示少年、青年、中年等年龄概念与年龄之间的关系,可以使用正态云模型表示为如图2所示的形式。

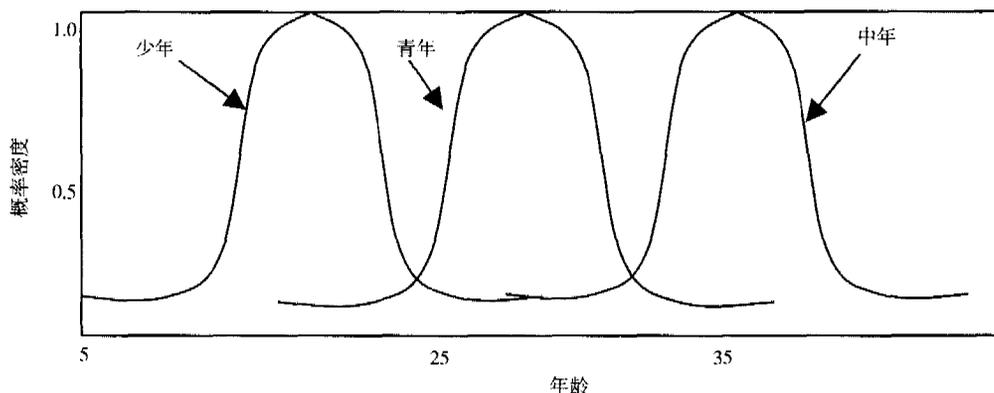


图2 云模型示意图

Fig.2 Example of loud model

云模型具有以下特点。

(1) 云模型体现的是概念  $T$  到区间  $[0,1]$  上一对多映射,而不是一条明晰的隶属曲线,从而产生了云的概念。

(2) 云具有可以伸缩、无边缘、有弹性的特点。

(3) 云模型强调的是云的整体特性而不是特定云滴的状态。

(4) 云模型的应用需要在计算机的支持下完成。

本研究将尝试在客户细分中采用云模型来描述具有动态特点的非确定性客户分类问题。

### 3 动态客户细分模型

#### 3.1 基于云模型的动态细分

根据客户关系管理问题的特点,客户细分应该能够体现客户行为的不确定性和随机性特征。为此,本文提出基于云模型的动态分类模型。

在数据集上与分类模型相关的属性集合为条件属性集和决策属性集,设条件属性集  $C = \{c_1,$

$c_2, \dots, c_m\}$ ,  $c_i$  的值域为  $X_i, i = 1, 2, \dots, m$ ;  $D = \{d_1, d_2, \dots, d_k\}$  是决策属性集,  $d_i$  的值域为  $Y_i, i = 1, 2, \dots, k$ 。数据集中的记录根据决策属性的不同取值构成了不同的类。

如某个数据集上的决策属性  $D = \{\text{年龄阶层, 购物倾向}\}$ , 年龄阶层的值域为  $\{\text{青年, 中年, 老年}\}$ , 购物倾向的值域为  $\{\text{时尚型, 保守型}\}$ , 则在  $D$  上取值为 (青年, 时尚型) 的数据记录构成了一个客户类。设  $L$  是论域上所有类的集合, 显然  $L \subseteq Y_1 \otimes Y_2 \otimes \dots \otimes Y_k$ 。

所谓分类方法或分类模型,就是属性  $C$  取值与属性  $D$  取值的对应关系。

设  $s$  是一个待分类的数据集,  $X = \delta_c(s)$ ,  $\delta_c(s)$  表示  $s$  在属性集  $C$  上的投影运算结果, 显然,  $\forall x \in X, x$  是一个待分类模式。

对于一个待分类模式进行分类,有的分类模型直接给出确定的类,如决策树模型等,称之为类标号型分类;有的分类模型是首先得到该待分类模式属于各类的隶属度,然后再根据隶属度确定最后的类,如神经网络模型及贝叶斯分类模型等,可以称之为度量值型分类。

**定义 1** 设  $l \in L$ ,  $l$  是一个类,  $f_l: X \rightarrow [0, 1]$  是一个函数,  $\forall x \in X, f_l(x)$  为  $x$  对于类  $l$  的隶属度, 称  $f_l$  为对应类  $l$  的隶属函数。

对于类标号型分类, 每个类的隶属函数取值只能为 0 或 1. 在度量值型分类中, 隶属函数的  $f$  取值为  $[0, 1]$  之间的实数。

设  $F = \{f \mid \text{存在一个类 } l \in L, f \text{ 是 } l \text{ 的隶属函数}\}$ , 若  $|F| = n$ , 则对一个待分类模式  $x$  用  $F$  中函数处理后, 可以得到一个  $n$  元向量  $[f_{l_1}(x), f_{l_2}(x), \dots, f_{l_n}(x)]$ , 其中每个分量上的隶属函数  $f_{l_i}$  对应一个目标类  $l_i, i = 1, 2, \dots, n$ .

**定义 2** 本研究将形如  $\{[l_1, f_{l_1}(x)], [l_2, f_{l_2}(x)], \dots, [l_n, f_{l_n}(x)]\}$  的有序对集合称为  $x$  的分类状态集, 记为  $L_x$ , 其中  $x \in X$  为一个待分类模式。

在例 1 中一个收入为 4 999 元的客户, 其分类状态集可以表示为

$\{(高收入类, 0), (中等收入类, 1), (低收入类, 0)\}$

在例 2 中的客户, 其分类状态集将是如下形式

$\{(I \text{ 类}, 0.49), (II \text{ 类}, 0.82), (III \text{ 类}, 0.83)\}$

在具体的分类中,  $F$  通过分类算法来表达, 本文将这个过程称为分类算法过程或  $C$  过程 (classification algorithm process). 一个待分类模式经过  $C$  过程将得到相应的分类状态集, 最终的结果将根据待分类模式在各目标类上的隶属度

取值, 按照某种原则经过进一步处理得出, 这个过程本研究称之为后处理过程或  $P$  过程 (post classification process). 定义 3 给出了  $P$  过程的形式化定义。

**定义 3** 设  $X$  是待分类模式的集合,  $Z$  是一个集族,  $Z = \{L_x \mid x \in X\}$ , 设  $L$  是类的集合, 函数  $P: Z \rightarrow L$ , 将一个分类状态集映射为一个分类结果, 应用函数  $P$  进行处理的过程, 称为  $P$  过程或分类的后处理过程。

**定义 4** 分类器为一个四元组  $(R, L, C, P)$ , 其中  $R$  为一个关系模式,  $L$  为论域上所有类的集合,  $C$  表示应用分类算法对待分类模式进行处理的过程,  $P$  为后处理过程。

于是分类过程便可以被描述成一个包含  $C$  过程和  $P$  过程的  $C$ - $P$  模型 (图 3)。

目前的确定型分类之所以分类结果是确定的, 其原因便是  $P$  过程为确定性模型, 即是一个确定的函数或方法, 例如极大值法等, 所以无论分类器采用什么算法, 分类结果都是确定的。

本研究将云模型的动态思想引入到客户分类中, 建立动态的分类过程模型, 在动态分类过程中, 云模型的特点将主要体现在客户细分的  $P$  过程中。

**定义 5** 一个分类器  $Classifier = (R, L, C, P)$ , 若  $P$  是一个随机过程, 则  $Classifier$  称为动态分类器。

在本研究提出的动态分类器中,  $P$  过程为基于云模型动态过程, 如图 4 所示。

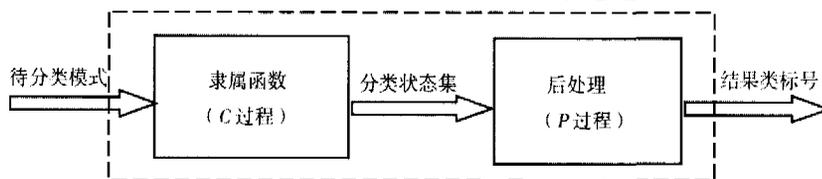


图 3 分类过程示意图

Fig. 3 Classification process

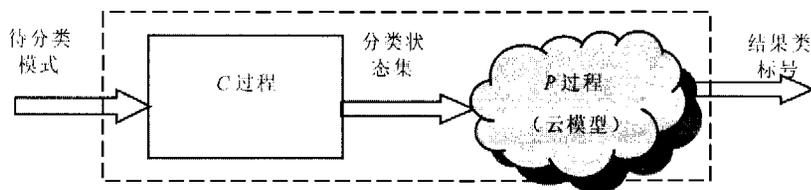


图 4 动态客户细分过程示意图

Fig. 4 Dynamic customer segmentation process

### 3.2 动态客户细分模型的 $P$ 过程

动态分类器与传统的分类器的不同之处主要

在于对分类状态集的后处理过程不同, 动态分类器中的后处理是一个随机过程, 因此针对分类后

产生的分类状态集,应用该过程所得到的客户类别可能会随运行时间的不同而不同.动态分类器的后处理过程设计的基本原则是隶属函数度高的类别被确定为客户类别的概率大.

针对客户细分问题的特点,设计动态分类器的后处理过程还需要解决这样两个问题.

1) 控制最小隶属度

在该分类模型中,可以根据需要对分类结果类别的最小隶属度加以限制.

本研究采用的方法是通过参数控制,设定一个隶属度阈值  $\epsilon$ ,对分类状态集中的每个元素经过计算,决定候选类别.

$\epsilon$  的设定可以有以下两种方法.

a.  $\epsilon$  取候选类别应满足的最小隶属度

这时隶属度高于  $\epsilon$  的类别可以作为候选类别,隶属度低于  $\epsilon$  的元素将不被加入候选类别,如果取  $\epsilon = 0.7$ ,则分类状态集  $\{(c_1, 0.1), (c_2, 0.9), (c_3, 0.8)\}$  中  $c_2$  和  $c_3$  作为候选类,  $c_1$  则不被加入候选类别集.

采用这种方法可能会出现某些分类模式最终哪一类都不属于的情况,如一个分类状态集为  $\{(c_1, 0.4), (c_2, 0.75), (c_3, 0.6)\}$ ,若取  $\epsilon = 0.8$ ,则候选类别的集合为空.在客户细分中,这种情况经常出现,当人们根据客户的行为为其分类的时候,因为客户并不是每一次行为都是理性的,所以可能导致属于各种类别的特征都不明显的情况,这时需要在分类的最后输出结果中增加一个值:“类别无法确定”.

b.  $\epsilon$  取与最大隶属度的绝对差或相对差

对于分类状态集  $\{(c_1, p_1), (c_2, p_2), \dots, (c_n, p_n)\}$  中的每个元素  $(c_i, p_i)$ ,若  $\epsilon$  取绝对差,当  $|p_i - \max(p_1, p_2, \dots, p_n)| \leq \epsilon$  时,将  $c_i$  作为候选类别;若  $\epsilon$  取相对差,当  $\frac{|p_i - \max(p_1, p_2, \dots, p_n)|}{\max(p_1, p_2, \dots, p_n)} \leq \epsilon$  时,将  $c_i$  作为候选类别.

2) 决策者偏好的反应

$P$  过程还可以体现决策者在不同决策环境下的偏好.如决策者希望对 I、II、III 三类客户采取不同的策略,假设一个客户经处理后得到的分类状态集是  $\{(I, 0.1), (II, 0.8), (III, 0.9)\}$ ,客户属于 III 类的可能性比较大,如果这次采取的策略是“给 II 类客户邮寄贺年卡,对 III 类客户没有行

动”,那么尽管客户属于 III 类的可能性较大,但为了争取潜在的客户,决策者可能还是希望某些这种情况的客户能归入 II 类客户的机会大一些;但如果这次采取的策略是:“给 B 类客户打 4 折,对 III 客户没有优惠”,决策者严格控制享受此优惠的客户,这时就希望该客户归入 III 类的可能性更大一些.

控制候选类别的参数可以反映决策者的偏好,此外还可以对隶属度进行适当的函数变换以反应决策者的偏好.例如可以对各个类别的隶属度赋予不同的权值,决策者希望放宽限制的类别权值稍高.

在一维均匀云模型下,客户  $x$  对应类别  $C_i$  的概率与相应各类的隶属度  $y_i$  之间的关系可以用公式(1)表示

$$P_i(x) = P(x \in C_i) = \frac{l_i}{\sum_{j=1}^N l_j}, \quad i = 1, \dots, N \quad (1)$$

这时一个客户最终被判断为某一类的概率等于该类相应隶属度占隶属之总和的比率.考虑到隶属度为  $[0, 1]$  间的实数,引入参数  $\pi$ ,并采用下式来表示客户的分类概率分布

$$P_i(x) = P(x \in C_i) = \frac{y_i^\pi}{\sum_{j=1}^N y_j^\pi} \quad (2)$$

其中,  $\pi$  为自然数.当  $\pi$  取 1 时公式(2)等同于公式(1),随着  $\pi$  取值的不断增大,客户分类的概率分布不断倾向于具有最大隶属度的类别,当  $\pi$  取较大值时,由公式(2)所反映的动态分类则等同于传统的基于最大隶属度原则的确定分类.

例如某一客户具有分类状态集  $\{(第 1 类, 0.75), (第 2 类, 0.5), (第 3 类, 0.25)\}$ ,利用公式(2),则不同  $\pi$  取值时的分类概率如表 1 与图 5 所示.

表 1 客户隶属不同类的概率与参数  $\pi (PI)$  取值的关系  
Table 1 Relationship between  $\pi (PI)$  and the possibility of a customer belonging to different classes

PI 取值	第 1 类的概率	第 2 类的概率	第 3 类的概率
PI = 1	50%	33%	17%
PI = 3	75%	22%	3%
PI = 15	100%	0%	0%

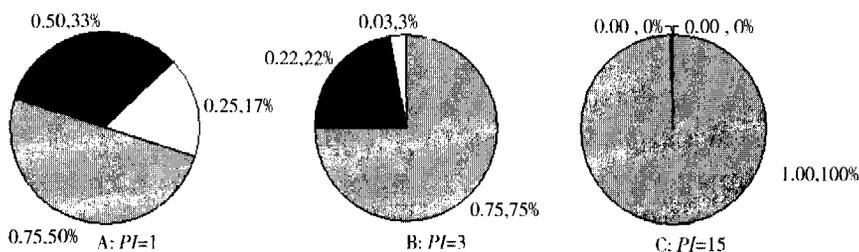


图 5 客户隶属不同类的概率与参数  $\pi(PI)$  取值关系示意图

Fig. 5 Pie chart of the relationship between  $\pi(PI)$  and the possibility for each class

综合上述想法, 本文基于一元均匀分布云模型提出动态分类器的  $P$  过程如下:

输入 分类状态集  $\{(c_1, l_1), \dots, (c_n, l_n)\}$ , 参数  $\epsilon$ , 参数  $\pi$ , 对应各个类别的一组权值  $\{w_1, \dots, w_n\}$

输出 带分类模式(客户)的类别名称

步骤 1 根据  $\epsilon$  确定候选类别集合  $C = \{(c_1, l_1), (c_2, l_2), \dots, (c_j, l_j)\}$ .

步骤 2 将各候选类别的隶属度与该类别的权重相乘, 得到

$$C' = \{(c_1, y_1), (c_2, y_2), \dots, (c_j, y_j)\}, \text{ 其中, } y_n = l_n \times w_n$$

步骤 3 计算分类模式属于各类的概率.

设  $X$  的分类概率向量为  $P(x) = [p_{i1}, p_{i2}, \dots, p_{ij}]^T$ , 其中

$$P_i(x) = P(x \in C_i) = \frac{y_i^x}{\sum_{j=1}^N y_j^x}$$

步骤 4 调用动态类生成算法, 确定该待分类模式所对应的类别.

在设计动态类生成算法时, 本研究参考云模型思想, 在客户细分模型的  $P$  过程中采用一维均匀云发生器, 该发生器由一个服从  $[0-1]$  间均匀分布的随机函数构成. 基于该均匀云发生器, 可以产生动态细分结果. 下面给出生成动态类的云模型算法.

算法 基于云模型的动态类生成算法

输入:

待分类模式  $X$ ;

目标类集  $C = \{c_1, c_2, \dots, c_N\}$ ;

分类状态集  $\{(c_1, m_1), (c_2, m_2), \dots, (c_N, m_N)\}$

输出:  $X$  所属的类

- (1)  $P = 0$ ;
- (2)  $C = Null$ ;
- (3)  $Cloud = RAN()$ ;
- (4) For ( $i = 1$ ;  $P < Cloud$ ;  $i++$ ) {
- (5)  $P = P + m_i$ ;
- (6)  $C = C_i$ ;
- (7) }
- (8) Return  $C$ ;

在上述算法中,  $RAN()$  为服从一维均匀分布的随机函数. 因为分类状态集中待分类模式属于各类的概率各不相同, 可以认为不同的类将  $[0, 1]$  区间划分为大小不等的段, 算法中(4)即是确定随机函数落在哪一个段内, 从而确定待分类模式属于哪一类. 概率大的类所属的区间段也比较大, 因而被选中的概率也大.

## 4 数据实验

本研究分别采用合成数据及来自某商业银行的实际数据进行了数据试验, 以验证本研究提出的动态客户细分方法.

### 4.1 合成数据

1) 数据来源

数据试验的目的是验证本文提出的动态分类方法的性能, 为保证数据质量, 本文采用了人工合成数据进行实验.

该人工合成数据取自美国加州大学 Irvine 分校机器学习知识库 (machine learning repository, UCI). 这是一个供全球数据挖掘与机器学习研究者交流研究成果的开放共享数据库, (网址为: <http://www.ics.uci.edu/~mlearn/MLRepository.html>).

本研究选用该知识库中一个包含 5 000 条记

录的分类学习与测试数据集,每条记录包含 21 个连续值条件属性及一个类标号属性,这些数据分别属于 3 类,且类标号均已知.本研究用该数据的 3 种类标号分别表示客户的 3 个信用度级别,即 {信用度低,信用度中,信用度高};条件属性的原始取值范围为 [0,6],并在此基础上对所有数据进行了期望值为 0,标准差为 1 的噪声处理(UCI 数据库中的数据已经包含了这一噪声处理结果),本

研究用该数据的 21 个属性字段表示客户的 21 个条件属性.

2) 客户细分模型结构

本数据实验所用的客户细分模型结构采用如图 6 所示的动态多分类器融合模型,该模型包含 4 个 BP 神经网络分类器,和一个分类融合器,并采用遗传算法作为分类融合模块的系数矩阵优化算法<sup>[16]</sup>.

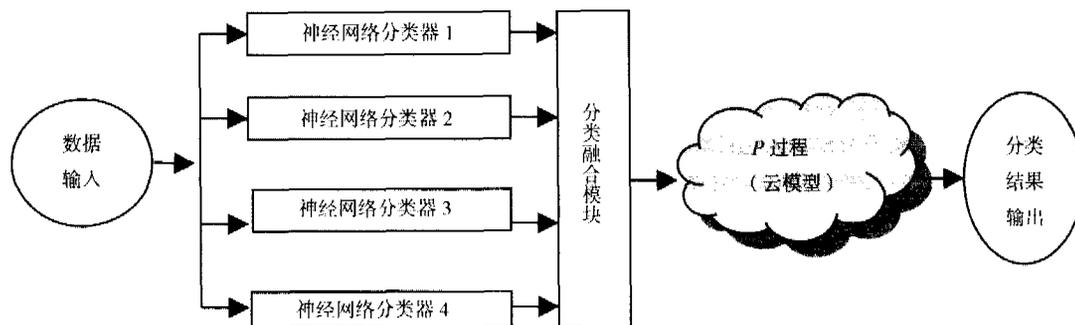


图 6 动态客户细分模型结构图

Fig.6 The structure of dynamic customer segmentation model

多分类器融合分类模型中的每个单分类器的结构如表 2 中所列.

表 2 神经网络单分类器的结构

Table 2 Structure of single ANN classifiers

分类器名称	隐层个数	隐层节点数
神经网络分类器 1	1	18
神经网络分类器 2	1	25
神经网络分类器 3	1	32
神经网络分类器 4	1	40

3) 实验结果

本文在 5 000 条记录中随机抽取 2 000 进行实验,其中 1 000 条记录组成训练样本集,另外 1 000 条记录作为测试样本集.经 1 000 个训练数据训练得到动态分类模型,其中分类融合器矩阵由遗传算法训练获得.

利用经训练样本集训练获得的动态分类器,对另外 1 000 个测试数据进行两次平行分类测试,如果在分类测试时,分类模型给出的客户类与已知的该客户类标号一致,则表明分类模型对这位客户的分类是正确的,实验结果如表 3 所示.

表 3 合成数据实验结果

Table 3 Experimental result on the synthetic data

分类实验编号	正确分类客户数	分类正确率
A	864	86.4%
B	866	86.6%
至少有一次被正确分类	923	92.3%

实验数据表明,在两次实验中至少有一次被正确分类的比率高达 92.28%,这表明由于动态分类模型的分类结果不是确定不变的,因此大大提高了对客户实现正确分类的机会.

4.2 银行客户风险评估数据

1) 数据来源

本文的数据来源于文献<sup>[17]</sup>,由某商业银行获得的 100 家短期贷款企业客户的风险评估数据.该研究参考其他研究者的经验,根据 LAPP 原则(流动性(liquidity)、活动性(activity)、盈利性(profitability)、潜力(potentialities))选择了文献<sup>[17]</sup>提出的风险评估因素体系.该体系根据销售收入/总资产、总资产周转率、流动资产周转率、固定资产周转率、存货周转率、应收款周转率、流动比率、营运资金/总资产、速动比率、超速动比率、资产负债率、成本费用利润率、销售净利率、资产收益率、净资产收益率、贷款方式等 16 个指标经主因素分析得出 4 项客户指标<sup>[18]</sup>.这 4 项指标分别是营运能力因子、偿债能力因子、盈利能力因子及贷款方式因子.

本文获得的银行客户数据中包含 100 家企业客户的短期贷款数据,其中风险程度按高低两极评价,数据示例见表 4.

表4 某商业银行客户风险数据示例

Table 4 Example of customer risk data in a commercial bank

样本	营运能力因子	偿债能力因子	盈利能力因子	贷款方式因子	信用风险类别
1	-0.226 4	-0.566 3	0.389 3	-0.566 0	低
2	-0.630 2	-0.118 0	0.457 3	0.355 1	低
3	-0.475 6	-0.271 5	0.572 4	0.848 1	低
...	...	...	...	...	...
98	-0.702 9	0.576 3	0.477 5	0.361 3	高
99	-0.364 6	-0.644 9	0.426 0	0.208 3	高
100	-0.606 2	1.465 4	0.917 7	0.078 7	高

本文利用从该银行的100个企业客户中随机选出的70家客户组成训练样本,剩下的30家客户作为测试样本,进行了银行客户信用风险分类模型的训练和测试。

### 2) 分类模型结构

本实验采用具有4个输入层节点,2个输出层节点,以及分别具有5个隐层节点(Ann-5)、8个隐层节点(Ann-8)、11个隐层节点(Ann-11)的3个BP神经网络作为基本分类器,按照4.1节的模式构建动态分类器。

### 3) 试验结果

实验结果如表5所示。

表5 银行客户风险数据实验结果

Table 5 Experimental result on the customer risk data

分类实验编号	正确分类客户数	分类正确率
A	26	86.67%
B	25	83.33%
至少有一次被正确分类	29	96.67%

实际数据的实验结果再一次表明,动态客户

细分方法可以大大提高多次客户分类中,对客户实现至少一次被正确分类的机会。

## 5 结 论

客户细分作为客户关系管理中的核心内容之一,已经日益受到人们的关注.由于受社会、心理及环境等复杂因素影响,客户行为往往具非确定性和非一致性.然而,现有的客户细分方法大多为确定性细分方法,无法客观描述客户行为的随机性和不确定性.为此,本文提出了基于云模型的动态客户细分方法,该方法将描述非确定关系的云模型理论引入到客户细分方法中,从而实现了客户细分的动态性,提高了模型对客户行为描述的客观性.本文分别采用合成数据与银行客户风险数据,对文中提出的动态客户细分模型进行了数据实验,实验结果表明该方法可以显著提高在非确定因素影响下,企业对客户进行正确分类的机会。

## 参 考 文 献:

- [1] Brown S A. Customer Relationship Management: A Strategic Imperative in the World of E-Business[M]. New York: John Wiley & Sons, 2000. 6—26.
- [2] 齐佳音, 韩新民, 李怀祖. 我国客户关系管理研究的紧迫性和方向分析[J]. 管理科学学报, 2002, 5(4): 88—94.  
Qi Jiayin, Han Xinmin, Li Huaizu. Emergency and directions for CRM in China[J]. Journal of Management Sciences in China, 2002, 5(4): 88—94. (in Chinese)
- [3] 张国方, 金国栋. 客户细分理论及应用策略研究[J]. 华中科技大学学报(社会科学版), 2003, (3): 101—104.  
Zhang Guofang, Jin Guodong. Research on customer segmentation theories and their application strategies[J]. Journal of Huazhong University of Science and Technology (Social Science Edition), 2003, (3): 101—104. (in Chinese)
- [4] Darrell Rigby, Frederick F, Reichheld. Avoid the four perils of CRM[J]. Harvard Business Review, 2002, (1): 101—109.
- [5] Rigby D K. Management Tools 2001 Global Results: Annual Survey of Senior Executives[R]. BAIN & COMPANY Inc Research Report, 2002. 1—4.
- [6] Per Vagn Freytag, et al. Business to business market segmentation[J]. Industrial Marketing Management, 2001, 30(6): 473—486.

- [7]Chen Lei-da, Soliman K S, Mao En, Frolick M N. Measuring user satisfaction with data warehouse: An exploratory study[J]. *Information & Management*, 2000, 37(3): 103—110.
- [8]Suh E H, Noh K C, Suh C K. Customer list segmentation using the combined response model[J]. *Expert Systems with Applications*, 1999, 17(2): 89—97.
- [9]刘 义, 万迪昉, 张 鹏. 基于购买行为的客户细分方法比较研究[J]. *管理科学*, 2003, 16(1): 69—72.  
Liu Yi, Wan Difang, Zhang Peng. Comparative research of solutions for the customer segmentation based on the purchase behavior [J]. *Management Sciences in China*, 2003, 16(1): 69—72. (in Chinese)
- [10]Kim Jaesoo, *et al.* Segmenting the market of West Australian senior tourists using an artificial neural network[J]. *Tourism Management*, 2003, 24(1): 25—34.
- [11]Vellido A. Segmentation of the on-line shopping market using neural networks[J]. *Expert Systems with Applications*, 1999, 17(4): 303—314.
- [12]Heilman C M, Bowman D. Segmenting consumers using multiple-category purchase data[J]. *Intern. J. of Research in Marketing*, 2002, 19(3): 225—252.
- [13]Li D, Han J, Shi X M, Cheng M. Knowledge representation and discovery based on linguistic atoms[J]. *Knowledge-Based System*, 1998, 10(7): 431—440.
- [14]李德毅. 知识表示中的不确定性[J]. *中国工程科学*, 2000, (10): 73—79.  
Li Deyi. Uncertainty in knowledge representation[J]. *Engineering Science*, 2000, (10): 73—79. (in Chinese)
- [15]蒋 嵘, 李德毅. 基于形态表示的时间序列相似性搜索[J]. *计算机研究与发展*, 2000, (5): 601—608.  
Jiang Rong, Li Deyi. Similarity search based on shape representation in time-series data sets[J]. *Journal of Computer Research and Development*, 2000, (5): 601—608. (in Chinese)
- [16]Eunju Kim a, Wooju Kim b, Yillbyung Lee. Combination of multiple classifiers for the customer's purchase behavior prediction [J]. *Decision Support Systems*, 2002, (34): 167—175.
- [17]于立勇. 我国国有商业银行信用风险评估研究[D]. 哈尔滨: 哈尔滨工业大学, 2002. 89—110.  
Yu Liyong. Study on Credit Risk Assessing in State Own Commercial Banks of China [D]. Harbin: Harbin Institute of Technology, 2002. 89—110. (in Chinese)
- [18]于立勇. 商业银行信用风险评估预测模型研究[J]. *管理科学学报*, 2003, 6(5): 46—52.  
Yu Liyong. Study on credit risk assessing and forecasting model in commercial bank[J]. *Journal of Management Sciences in China*, 2003, 6(5): 46—52. (in Chinese)

## Research on dynamic customer segmentation in customer relationship management

YE Qiang, LU Tao, YAN Xiang-bin, LI Yi-jun

School of Management, Harbin Institute of Technology, Harbin 150001, China

**Abstract:** Customer segmentation is one of the basic functions in customer relationship management. Most of the customer segmentation methods definitely classified the customer into different classes according to their attributes. But these definite classification methods can't fit the customer segmentation tasks well, for the uncertain and stochastic characteristics of customer behavior. The shortcoming of current definite customer segmentation methods was pointed out in this paper. To solve this problem, cloud model theory was introduced into customer segmentation and a dynamic customer segmentation model was built up in this paper. This dynamic customer segmentation model consisted of a *C*-process and a *P*-process according to the uncertainty and the heterogeneity in data space of customers' behaviors. The proposed new model could meet the requirement in presenting the uncertainty of customers' behaviors by introduced the cloud model theory in to the *P*-process. A synthetic data experiment and a real data experiment indicated the effectiveness of this method.

**Key words:** customer relationship management; dynamic customer segmentation; cloud mode