

基于案例推理的银行 零售客户价值细分模型构建

曹国 沈利香

(常州工学院经济管理学院 江苏常州 213022 常州工学院计算机信息工程学院 江苏常州 213002)

【摘要】 本文首先结合 RFM 理论建立了多层多维的商业银行零售客户价值的评价指标体系。通过拓宽 RFM 模型的指标体系,改变了 RFM 模型简单的特点。接着利用数据挖掘技术构建商业银行零售客户价值智能分类模型,针对数据仓库中的零售客户原始数据,利用 K-means 算法获得案例库中的历史案例客户价值分类类型;在构建案例相似度模型中,本文首次利用 ANP 模型获得评价指标的权重,强调了评价指标之间的相互影响。最后,利用灰色关联分析原理建立了多层灰色关联方法以适应所构建的商业银行零售客户评价指标体系。

【关键词】 客户细分 RFM 模型 案例推理 灰色关联分析 网络层次分析法

零售客户价值细分研究主要是探讨零售客户在某一时间点的价值表现,价值细分的主要目的就是对客户在该时间点的行为进行管理。通过对零售客户价值进行细分研究,可以帮助商业银行正确认识客户的价值贡献,并适当地进行分类,从而辅助商业银行长期识别、保留和发展零售客户。对此,众多学者进行了研究。

冯永等(2004)提出了一种基于动态 SOM 神经网络和 RFM 指标的客户分类方法,首先利用动态 SOM 神经网络聚类分析模型产生客户簇,使企业能够有针对性地对不同客户实施差别化服务策略,为企业的客户战略提供了有力的支持。孟钊兰等(2008)运用多元统计的因子聚类分析方法,从风险要素和价值要素两个维度来重新设置银行信贷客户分类。殷

琪(2008)利用模糊数学的有关理论建立了一套多因子客户分类评价模型,并应用此模型对企业的客户分类进行评价。张庭溢等(2008)基于数据挖掘的客户细分方法,提出了一套电信行业客户分类的数据挖掘技术应用解决方案,并对电信公司小灵通客户进行了应用分析。Achim Walter, Thomas Ritter 和 Hans Georg Gemunden(2001)以货币因素与非货币因素将客户价值分为直接价值、间接价值和社会价值。

与此不同,本文在前人研究的基础上拓宽了传统的 RFM 评价指标体系,并结合案例推理知识建立零售客户价值的智能细分模型,从而为企业的客户价值和群体信息的获取提供一个新的方法,以便能及时正确地识别零售客户群体信息,制定相关的营销策略,抢占市场。

挥了薪酬的激励作用。而当公司业绩较差时,薪绩敏感度相对较低,上市公司选择与业绩相关度不大的薪酬机制,薪酬的惩戒作用发挥得不够。同时,当公司表现为正向业绩时,国有企业薪绩敏感度较低,表明在国有企业的高管绩效评价体系中,业绩所占比重较民营企业小,国有企业担负着更大的社会责任。而在公司表现为负向业绩时,民营企业薪酬的惩戒功能有待提高。

本研究可能存在以下局限性:①只考虑了高管的货币薪酬,而未考虑高管以股权及实物形式存在的薪酬,结果可能存在偏差。②本文未对业绩及薪酬本身的高低及变化的大小所存在的不同维度进行更深入的研究。对于上述局限,也是未来研究需要考虑的问题。

【注】 本研究得到国家自然科学基金重点立项(编号:71032006)的指导。

主要参考文献

1. Gibbons, Murphy, K.. Executivesial incentives and the horizon problem. Journal of Accounting and Economics, 1990;

14

2. McGuire, J., J. Chiu. A. Elbing. Executive Incomes, Sales and Profits. American Economic Review, 1962; 5

3. Yermack D.. Higher market valuation for firms with a small board of directors. Journal of financial Economics, 1996; 4

4. 陈冬华, 陈信元, 万华林. 国有企业中的薪酬管制与在职消费. 经济研究, 2005; 2

5. 杜胜利, 翟艳玲. 总经理年度报酬决定因素的实证分析——以我国上市公司为例. 管理世界, 2005; 8

6. 江伟. 行业薪酬基准与管理者薪酬增长——基于中国上市公司的实证分析. 金融研究, 2010; 4

7. 李琦. 上市公司高级经理人薪酬影响因素分析. 经济科学, 2003; 6

8. 李增泉. 激励机制与公司绩效——一项基于上市公司的实证研究. 会计研究, 2000; 1

9. 刘斌, 刘星, 李世新, 何顺文. CEO 薪酬与企业业绩互动效应的实证检验. 会计研究, 2003; 3

一、零售客户价值细分指标体系

传统的 RFM(Recency Frequency Monetary)指标体系主要是依据客户的最近交易日、交易频率、交易金额来判断该客户是企业的黄金客户还是潜在客户或者是即将流失的客户。RFM 模型由于其思想比较简单,又能描述客户的交易行为,因此很早就在许多公司中得以应用。但是,理论界对此的关注却大大晚于实务界,直到信息技术的发展使得数据库营销技术得到大力发展的时候,该模型才开始得到广泛的研究和应用。考虑到 RFM 模型较为简单,本文采用拓展的 RFM 模型来构建商业银行零售客户价值评价的指标体系(见表 1)。

表 1 基于 RFM 的商业银行零售客户分类指标体系

控制指标	网络指标	指标描述
最近交易日	TR	客户最近一次的交易日期到分析日之间的时间差
	MR	客户各次交易日距离分析日时间长度的平均值
	StdR	客户各次交易日距离分析日时间长度的标准方差
	CVR	StdR与MR的比率
交易频率	TF	观察期内交易的次数
	MF	客户最近一个月的交易次数
	YF	客户最近半年的交易次数
交易金额	TM	观察期内的客户交易总金额
	AVM	观察期内的客户交易金额平均值

二、基于 CBR 的零售客户价值细分模型

不同类型的客户所需要的服务是不同的,如果对所有客户都采取相同的服务策略,则只能满足部分客户的需求,而其他客户却很难得到自己所需要的服务,这样就造成了资源的巨大浪费。然而,商业银行的资源是有限的,要使有限的客户服务资源发挥最大的作用,必须对客户进行分类。因此,本文利用案例推理技术构建商业银行零售客户价值的智能分类模型。在构建客户价值智能分类模型时,针对数据库中的零售客户原始数据,利用 K-means 算法和零售客户价值模型,获得案例库中的历史案例客户价值分类类型。在构建案例相似度模型时,本文首次利用 ANP 模型获得评价指标的权重,强调了评价指标之间的相互影响。与此同时,利用灰色关联分析原理建立多层灰色关联方法以适应所构建的商业银行零售客户评价指标体系。

1. 案例的描述。案例是能导致特定结果的一系列特征属性的集合,也是形成问题求解结构的子案例的关联集合。因为案例推理在很大程度上依赖于它所收集案例的表示结构和内容,故案例表示是 CBR(基于案例的推理技术)的基础,它需要决定每个案例应当存贮什么内容。一个典型的案例可以表示为一个多元组: (A, V, W_A, O) 。其中: $A=(a_1, a_2, \dots, a_m)$ 是多维客户价值细分指标体系; V 是零售客户在评价指标体系下的指标值; W_A 是零售客户主体对评价指标体系的偏好结构,在本文中此偏好结构表现为各评价指标的权重; O 为输出的某种价值类别,即目标零售客户属于哪一个类别。

2. 建立案例库。本文根据案例的描述方式建立案例库,它是用各个案例的指标值进行描述的,并采用关系数据库技

术来建立案例库。案例库中采用各个指标作为域名,案例号作为主关键字。利用关系数据库存储案例后,对案例库的管理和维护(包括案例的增加、删除、修改等)大部分由数据库管理系统完成。当然,在实际应用中,需要将一个完整的、有代表性的事例作为一个案例保存到案例库中,针对这类应用需求,本文利用 K-means 算法从客户数据库中选择典型案例客户来构造案例库。

3. 案例的相似度。相似性度量在 CBR 系统中十分重要,合适的度量方法可以迅速、准确地找到所需要的案例。相反,如果度量方法定义不好,检索的结果就不理想,也就谈不上成功应用。因此,相似性度量方法的选取十分重要。本文利用灰色理论并结合网络层次法构建案例的相似度。通过引入网络层次分析法计算分类指标权重,弥补了传统案例推理中假设指标间相互独立的缺点,强调了分类指标之间的相互联系和相互影响。

(1)属性的权重。对指标权重的确定,通常可以采用两种方法:一类是客观赋权法,其源信息来自于统计数据本身,属于这一类的有主成分分析法、因子分析法等,其突出优点是权值的客观性强,主要缺点是确定的权值有时与实际相悖;另一类是主观赋权法,其源信息来自专家咨询,属于这一类的有排队计分法、层次分析法等,其主要优点是分类人员可以根据实际问题合理确定各个指标权重之间的排序,主要缺点是主观随意性较大。本文采用网络层次分析法来计算分类指标的权重,与 AHP 分析方法相比较,最大的不同点是 ANP 方法中不同层次之间的信息反馈和同一层次元素之间具有相互依赖的关系。ANP 算法如下:

步骤一:建立评价指标的 ANP 模型。

设 ANP 模型中控制层有元素 (P_1, P_2, \dots, P_m) ,网络层中有元素组 (C_1, C_2, \dots, C_N) ,其中 C_i 中有元素 $(e_{i1}, e_{i2}, \dots, e_{in})$, $i=1, 2, \dots, N$ 。以控制层元素 $p_s(s=1, 2, \dots, m)$ 为准则,以 C_j 中的元素 $e_{j1}(1=1, 2, \dots, n_j)$ 为次准则,逐层分级设计。

步骤二:设计控制元素的判断矩阵并求出排序向量。

对网络层元素组 C 中的其他元素按照它对于元素 e 的影响力的大小进行比较分析,构建控制层元素 $p_s(s=1, 2, \dots, m)$ 下的判断矩阵,并根据特征根法可求得排序向量 $(w_{i1}^{(j1)}, w_{i2}^{(j1)}, \dots, w_{in_j}^{(j1)})^T$,记为:

$$W_{ij} = \begin{bmatrix} w_{i1}^{(j1)} & w_{i1}^{(j2)} & \dots & w_{i1}^{(jn_j)} \\ w_{i2}^{(j1)} & w_{i2}^{(j2)} & \dots & w_{i2}^{(jn_j)} \\ \dots & \dots & \dots & \dots \\ w_{in_j}^{(j1)} & w_{in_j}^{(j2)} & \dots & w_{in_j}^{(jn_j)} \end{bmatrix} \quad (1)$$

矩阵 W_{ij} 中的列向量是 C_i 中元素对元素 C_j 中元素的重要度的排序向量。

步骤三:构建并计算超矩阵和加权超矩阵。

对于 ANP 模型中准则层与网络层中相互无关准则或元素的权重采用 AHP 通过两两判断矩阵获得,而把网络层元素相互影响的排序向量集合起来得到在准则 p_s 下的超矩阵 $W=(w_{ij})_{N \times N}$ 。然后以 p_s 为准则,将 p_s 下的各组元素对元素

组 $C_j(j=1,2,\dots,N)$ 的重要性进行比较,得到归一化的特征向量 $(a_{1j},a_{2j},\dots,a_{nj})^T$ 。在比较中,与 C_j 无关的元素对应的排序向量分量为零,得到矩阵 $A=(a_{ij})_{N \times N}$ 。最后对超矩阵 \bar{W} 进行 $(2k+1)$ 次幂运算,即 $W(k)=W^{2k+1}(k)(k=1,2,\dots,N)$,直到满足收敛条件: $W(k)=W(k+1)$,则 W 中的列元素就是网络层各元素的权重。

(2)相似度。相似性度量是衡量目标案例与源案例之间的相似程度,相似度的定义和判定是案例推理中的一个关键问题。目前,计算案例之间相似度的方法很多,主要有主观评分法、数量积分法等。本文提出了一种新方法——灰色层次关联分析法。算法如下:

步骤一:构建网络层的决策矩阵。

假设案例库中历史案例零售客户的价值评价指标向量为: $(V_{i1}^{(f)},V_{i2}^{(f)},\dots,V_{ij}^{(f)})$ 。其中: $i=1,2,\dots,m;j=1,2,\dots,n;f=1,2,\dots,t$ 。 $V_{ij}^{(f)}$ 表示历史案例 f 的控制指标的第 i 个下第 j 个分类指标的价值。 f 表示历史案例客户数目, i 表示控制指标数目, j 表示控制指标 i 的网络指标的数目。

假设查询案例零售客户的价值评价指标向量为: $(V_{i1}^{(0)},V_{i2}^{(0)},\dots,V_{ij}^{(0)})$ 。式中: $i=1,2,\dots,m;j=1,2,\dots,n;V_{ij}^{(0)}$ 表示案例库中查询案例客户 0 的第 i 个控制指标下第 j 个评价指标的价值。 i 表示控制指标的数目, j 表示控制指标 i 的网络指标的数目。则控制指标 i 的网络指标决策矩阵可表示为:

$$G_i = \begin{bmatrix} v_{i1}^{(0)} & v_{i2}^{(0)} & \dots & v_{in}^{(0)} \\ v_{i1}^{(1)} & v_{i2}^{(1)} & \dots & v_{in}^{(1)} \\ \dots & \dots & \dots & \dots \\ v_{i1}^{(t)} & v_{i2}^{(t)} & \dots & v_{in}^{(t)} \end{bmatrix} \quad (2)$$

因此,控制指标 i 的距离矩阵为:

$$G_i' = \begin{bmatrix} \Delta_{i1}^{(01)} & \Delta_{i2}^{(01)} & L & \Delta_{in}^{(01)} \\ \Delta_{i1}^{(02)} & \Delta_{i2}^{(02)} & L & \Delta_{in}^{(02)} \\ M & M & M & M \\ \Delta_{i1}^{(0t)} & \Delta_{i2}^{(0t)} & L & \Delta_{in}^{(0t)} \end{bmatrix} \quad (3)$$

其中: $j=1,2,\dots,n;f=1,2,\dots,t$ 。

$$\Delta_{ij}^{(0f)} = |v_{ij}^{(0)} - v_{ij}^{(f)}| \quad (4)$$

步骤二:构建控制指标 i 的灰色关联系数矩阵。

根据灰色关联理论,控制指标 i 的灰色关联系数可表示为:

$$\zeta_{ij}^{(0f)} = \frac{\rho \max_{f \in t} \max_{j \in n} \Delta_{ij}^{(0f)}}{\Delta_{ij}^{(0f)} + \rho \max_{f \in t} \max_{j \in n} \Delta_{ij}^{(0f)}} \quad (5)$$

式中, $j=1,2,\dots,n;f=1,2,\dots,t$ 。 ρ 为分辨率,其取值一般在区间 $[0,1]$ 中,本文取 $\rho=0.5$ 。

$$\begin{bmatrix} \zeta_{i1}^{(01)} & \zeta_{i2}^{(01)} & L & \zeta_{in}^{(01)} \\ \zeta_{i1}^{(02)} & \zeta_{i2}^{(02)} & L & \zeta_{in}^{(02)} \\ M & M & M & M \\ \zeta_{i1}^{(0t)} & \zeta_{i2}^{(0t)} & L & \zeta_{in}^{(0t)} \end{bmatrix} \quad (6)$$

其中: $j=1,2,\dots,n;f=1,2,\dots,t$ 。 $\zeta_{ij}^{(0f)}$ 表示查询案例客户 f 和案例库中历史案例客户 0 在控制指标 i 下的网络指标 j 的灰色关联系数。

步骤三:构建控制层决策矩阵。

由灰色关联理论可得到控制指标 i 的灰色关联度的计算公式,并可表示为:

$$\zeta_i^{(f)} = \frac{1}{n} \sum_{j=1}^n w_{ij} \zeta_{ij}^{(0f)} \quad (7)$$

式中: $\zeta_i^{(f)}$ 表示第 f 个查询案例客户的灰色关联度, w_{ij} 表示控制指标 i 下的网络属性 j 的权重, $\zeta_{ij}^{(0f)}$ 为相应指标的灰色关联系数。

将网络属性的灰色关联度作为相对应的控制属性的价值,并构建控制层属性的价值矩阵如下:

$$G = \begin{bmatrix} \zeta_{i1}^{(1)} & \zeta_{i2}^{(1)} & L & \zeta_{im}^{(1)} \\ \zeta_{i1}^{(2)} & \zeta_{i2}^{(2)} & L & \zeta_{im}^{(2)} \\ M & M & M & M \\ \zeta_{i1}^{(t)} & \zeta_{i2}^{(t)} & L & \zeta_{im}^{(t)} \end{bmatrix} \quad (8)$$

式中, $i=1,2,\dots,m;f=1,2,\dots,t$ 。 $\zeta_i^{(f)}$ 表示第 f 个查询案例客户的第 i 个控制指标价值。

步骤四:案例的相似度。

查询案例客户与源案例客户之间的综合灰色关联度可表示为:

$$\begin{aligned} sim_f &= \frac{1}{m} \sum_{i=1}^m w_i \zeta_i^{(f)} \\ &= \frac{1}{m} \sum_{i=1}^m w_i \left[\frac{1}{n} \sum_{j=1}^n w_{ij} \zeta_{ij}^{(0f)} \right] \end{aligned} \quad (9)$$

式中, sim_{0f} 表示第 0 个查询案例客户与第 f 个查询案例客户之间的相似度, w_i 表示第 i 个控制指标的权重。

4. 案例的检索。CBR 系统是否为可行的衡量标准,取决于该系统能否检索出最相似的案例,以及其为新案例提供解决案例的能力。著名的案例检索方法有:最邻近算法、归纳推理、知识引导推理和模板检索法等。而其中最常用的算法为最邻近算法和层次相似算法。本文采用最邻近算法模型。

最邻近算法的 NN 算法成立的假设是案例的属性间存在合适的匹配规则和程序。

三、实证分析

1. 数据的标准化处理。本研究的数据来源于南京市 X 行零售客户数据库,研究的观察期定义为 1998 年 1 月至 2003 年 12 月,共 679 名零售客户。在本文的实证中,利用 K-means 算法将 679 名零售客户随机分为两类,第一类客户共 324 名,该类客户将通过 K-means 算法形成案例库中的历史案例;第二类客户共 355 名,这些客户将作为查询案例来检验算法的可行性。考虑到银行数据的保密性,本文仅从第二类客户群中随机抽取 18 名案例客户作为算例来验证算法的可行性。考虑到原始样本数据中价值指标的量纲不同,因此必须对数据进行标准化处理。

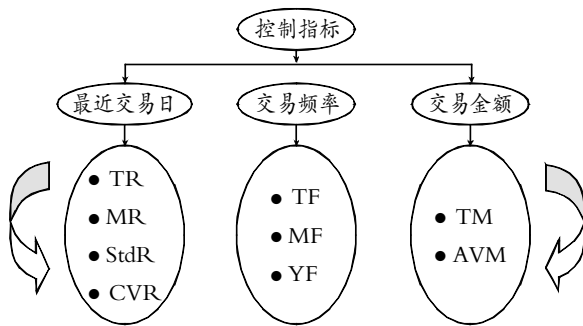
经过归一化处理后的指标值如表 2 所示:

表2 经预处理后的属性价值

	最近交易日(R)				交易频率(F)			交易金额(M)	
	TR	MR	StdR	CVR	TF	MF	YF	TM	AVM
C ₁	0.607 1	0.449 1	0.390 3	0.043 2	0.000 0	0.000 0	0.000 0	0.044 4	0.156 4
C ₂	0.250 0	0.500 0	0.548 9	0.136 7	0.117 6	0.666 7	1.000 0	0.051 9	0.102 9
C ₃	0.285 7	0.861 1	0.819 9	0.374 1	0.142 2	0.000 0	0.416 7	0.476 4	0.958 2
C ₄	0.285 7	0.731 5	0.655 0	0.664 3	0.049 0	0.000 0	0.000 0	0.165 5	0.386 7
C ₅	0.035 7	0.231 5	0.276 6	0.467 6	0.004 9	0.333 3	0.333 3	0.000 0	0.070 3
C ₆	0.821 4	1.000 0	0.976 1	0.071 9	0.073 5	0.000 0	0.000 0	0.368 0	0.601 4
C ₇	0.535 7	0.768 5	0.774 6	0.829 7	0.053 9	0.000 0	0.083 3	0.087 3	0.136 4
C ₈	1.000 0	0.782 4	1.000 0	0.573 1	0.098 0	0.333 3	0.500 0	0.121 7	0.067 6
C ₉	0.000 0	0.791 7	0.922 3	1.000 0	0.034 3	0.000 0	0.000 0	0.059 2	0.000 0
C ₁₀	0.571 4	0.662 0	0.624 5	0.712 2	0.029 4	0.000 0	0.000 0	0.101 6	0.209 5
C ₁₁	0.214 3	0.115 7	0.171 2	0.472 4	0.205 9	1.000 0	0.500 0	0.140 2	0.201 1
C ₁₂	0.142 9	0.314 8	0.315 1	0.479 6	0.225 5	0.333 3	0.666 7	0.168 4	0.151 7
C ₁₃	0.142 9	0.726 9	0.666 8	0.700 2	1.000 0	0.666 7	0.250 0	1.000 0	1.000 0
C ₁₄	0.035 7	0.435 2	0.545 5	0.307 0	0.715 7	0.333 3	0.416 7	0.395 0	0.361 5
C ₁₅	0.035 7	0.000 0	0.000 0	0.738 6	0.205 9	0.000 0	0.833 3	0.040 7	0.069 3
C ₁₆	0.142 9	0.694 4	0.764 3	0.237 4	0.754 9	0.666 7	0.750 0	0.791 9	0.626 6
C ₁₇	0.000 0	0.421 3	0.446 2	0.000 0	0.725 5	0.666 7	0.500 0	0.211 8	0.107 7
C ₁₈	0.091 1	0.928 6	0.351 9	0.463 1	0.294 1	0.333 3	0.000 0	0.303 5	0.163 5

2. 基于 CBR 的零售客户价值细分模型。

(1) 指标权重。根据 ANP 原理,为计算客户价值指标的权重,首先需要构建各层价值指标之间的网络层次结构。在本文的客户价值评价网络层次结构中,控制层中的一级指标是商业银行零售客户的最近交易日、交易频率以及交易金额三个指标,在三个控制指标下各有相应的二级子指标(即网络指标),其商业银行零售客户价值细分指标的网络层次结构示意图如下图所示。



客户价值评价属性网络层次结构图

在网络结构示意图中,本文假设三个控制指标之间相互独立,并且网络层中三类指标体系之间也相互独立。同时,在三个一级指标中,假设最近交易日和交易频率的二级子指标是内部之间相互影响。图中半环表示这两类指标的二级指标内部之间相互关联。由于本研究中假设控制指标之间相互独立,故可根据 AHP 来计算各自相应指标的权重。根据 AHP 的原理可获得的控制层分类指标的权重 $w_i=(w_1, w_2, w_3)=(0.131 2, 0.792 8, 0.076 0)$ 。由于交易金额中的二级指标只有

两个,根据网络层次分析法的原理,它们的相对权重需用 AHP 算得: $w_{3j}=(w_{31}, w_{32})=(0.5, 0.5)$,并且它们直接用于后面的评价。

由于控制指标交易频率及客户最近交易时间长度的二级指标内部之间是相互影响的,故需利用网络层次分析法。本文仅以交易频率下的二级指标 ATF 为例说明网络层次中各指标权重的计算方法(见表 3),其他指标的权重计算类似。将算得的网络指标之间的权重输入到超矩阵,如通过表 3 算得的权重即为超矩阵的第六列与第七、八两行的交叉元素。同时,在超矩阵的第六行第六列表示元素 ATF 自己与自己相互比较,故其元素值为零。又由于其他控制指标的子指标与指标 ATF 之间相互独立,故第六列中其他元素值都为零。

表3 MF、YF 相对于 TF 的相对权重

TF	MF	YF	相对权重
MF	1	7	0.725 7
YF	1/7	1	0.274 3

利用同样的方法可构建评价指标的超矩阵(见表 4),在该矩阵中,主对角线上的元素值都为零。为获得持续、稳定的权重,需对超矩阵求极限超矩阵(见表 5)。由极限超矩阵可得到网络指标长期稳定的权重,即: $w_{1j}=(w_{11}, w_{12}, w_{13}, w_{14})=(0.328 1, 0.187 2, 0.224 4, 0.260 3)$, $w_{2j}=(w_{21}, w_{22}, w_{23})=(0.341 4, 0.423 8, 0.234 7)$ 。

表4 超矩阵

	TR	MR	StdR	CVR	TF	MF	AYF
TR	0	0.285 3	0.705 0	0.466 7			
MR	0.142 9	0	0.084 1	0.466 7			
StdR	0.571 4	0.104 7	0	0.066 6			
CVR	0.285 7	0.637 0	0.210 9	0			
TF					0	0.667 1	0.250 1
MF					0.725 7	0	0.749 9
AYF					0.274 3	0.332 9	0

表5 极限超矩阵

	TR	MR	StdR	CVR	TF	MF	AYF
TR	0.328 1	0.328 1	0.328 1	0.328 1			
MR	0.187 2	0.187 2	0.187 2	0.187 2			
StdR	0.224 4	0.224 4	0.224 4	0.224 4			
CVR	0.260 3	0.260 3	0.260 3	0.260 3			
TF					0.341 4	0.341 4	0.341 4
MF					0.423 8	0.423 8	0.423 8
AYF					0.234 7	0.234 7	0.234 7

(2) 相似度。灰色关联是灰色系统的基本概念。灰色关联是指事物之间的不确定性关联,或系统因子之间,因子对主行为之间的不确定性关联。灰色关联分析实质上是通过比较数据序列的曲线几何形状的接近程度来判断其联系紧密程度。一般来说,几何形状越接近,变化趋势也就越接近,关联度就愈大,反之就愈小。本文以客户 1 为例来检验算法,所建立的

网络指标价值矩阵见表6。

表6 网络指标的价值矩阵

ID	TR	MR	StdR	CVR	TF	MF	YF	TM	AVM
C ₁	0.607 1	0.449 1	0.390 3	0.043 2	0.000 0	0.000 0	0.000 0	0.044 4	0.156 4
类1	0.826 8	0.433 0	0.717 8	0.760 0	0.479 7	0.162 6	0.281 8	0.487 2	0.487 2
类2	0.837 5	0.149 9	0.717 9	0.800 8	0.522 8	0.269 3	0.322 3	0.253 7	0.253 7
类3	0.336 7	0.117 5	0.267 7	0.319 5	0.360 3	0.576 5	0.616 8	0.395 2	0.395 2
类4	0.373 2	0.221 5	0.337 0	0.370 0	0.236 3	0.225 2	0.282 3	0.122 8	0.122 8
类5	0.207 7	0.154 0	0.194 3	0.215 4	0.147 7	0.260 7	0.280 7	0.229 4	0.229 4
类6	0.120 7	0.302 4	0.124 8	0.123 4	0.076 5	0.195 2	0.187 2	0.072 8	0.072 8
类7	0.791 6	0.146 4	0.683 7	0.683 8	0.683 9	0.684 0	0.684 1	0.684 2	0.684 2
类8	0.563 0	0.246 9	0.492 8	0.537 9	0.322 9	0.215 6	0.260 2	0.275 8	0.275 8

表6中:C1表示查询案例客户的ID为1号,类1至类8表示由K-means方法从历史案例客户所构建的八类客户中随机抽取的典型客户(每类抽取1个典型客户),并作为历史案例客户使用。同时,由公式3可得距离矩阵,见表7。

表7 距离矩阵

	TR	MR	StdR	CVR	TF	MF	YF	TM	AVM
D(C _{1,1})	0.219 7	0.016 1	0.327 5	0.716 8	0.479 7	0.162 6	0.281 8	0.442 8	0.330 8
D(C _{1,2})	0.230 4	0.299 2	0.327 6	0.757 6	0.522 8	0.269 3	0.322 3	0.209 3	0.097 3
D(C _{1,3})	0.270 4	0.331 6	0.122 6	0.276 3	0.360 3	0.576 5	0.616 8	0.350 8	0.238 8
D(C _{1,4})	0.233 9	0.227 6	0.053 3	0.326 8	0.236 3	0.225 2	0.282 3	0.078 4	0.033 6
D(C _{1,5})	0.399 4	0.295 1	0.196 0	0.172 2	0.147 7	0.260 7	0.280 7	0.185 0	0.073 0
D(C _{1,6})	0.486 4	0.146 7	0.265 5	0.080 2	0.076 5	0.195 2	0.187 2	0.028 4	0.083 6
D(C _{1,7})	0.184 5	0.302 7	0.293 4	0.640 6	0.683 9	0.684 0	0.684 1	0.639 8	0.527 8
D(C _{1,8})	0.044 1	0.202 2	0.102 5	0.494 7	0.322 9	0.215 6	0.260 2	0.231 4	0.119 4

由公式(5)可算得网络指标的灰色关联系数矩阵,其中:分辨率 $\rho=0.5$ 。那么,网络指标的灰色关联系数矩阵见表8。

表8 网络指标灰色关联系数矩阵

GRA	TR	MR	StdR	CVR	TF	MF	YF	TM	AVM
C _c (C _{1,1})	0.632 9	0.959 2	0.536 3	0.345 7	0.441 2	0.699 7	0.573 4	0.461 1	0.533 8
C(C _{1,2})	0.621 8	0.558 7	0.536 2	0.333 3	0.420 1	0.584 5	0.540 3	0.644 1	0.795 6
C(C _{1,3})	0.583 5	0.533 2	0.755 5	0.578 2	0.512 5	0.396 5	0.380 5	0.519 2	0.613 3
C(C _{1,4})	0.618 2	0.624 7	0.876 6	0.536 8	0.615 8	0.627 2	0.573 0	0.828 5	0.918 5
C(C _{1,5})	0.486 8	0.562 1	0.659 0	0.687 5	0.719 5	0.592 3	0.574 4	0.671 9	0.838 4
C(C _{1,6})	0.437 8	0.720 8	0.587 9	0.825 3	0.832 0	0.660 0	0.669 3	0.930 3	0.819 2
C(C _{1,7})	0.672 5	0.555 8	0.563 5	0.371 6	0.356 5	0.356 4	0.356 4	0.371 9	0.417 8
C(C _{1,8})	0.895 7	0.652 0	0.787 0	0.433 7	0.539 8	0.637 3	0.592 8	0.620 8	0.760 3

将网络指标的灰色关联度作为控制指标的价值,并输入到控制指标的价值矩阵。由公式(9)经计算可得客户1与源案例客户之间的灰色关联度(见表9)。根据公式(9)的物理意义,相似度越大,说明查询案例客户与历史案例客户的相似程度越大。因此,本文采用最大灰色相似度所对应的案例作为价值评价的结果。从表9可以看出,C₁₃、C₁₆和C₁₇属于高价客户,而其他15个查询案例客户则属于较低价值客户。

为检验查询结果的可靠性,根据客户可计算出18个查询案例的各自评价价值,结果表明查询客户C₁₃、C₁₆和C₁₇的评价价值明显高于其他各查询客户。这与本文设计的客户价值评价结果是相吻合的。

表9 案例客户的相似度

Sim	类1	类2	类3	类4	类5	类6	类7	类8	Max
C ₁	0.108 6	0.060 4	0.051 9	0.072 0	0.071 7	0.081 4	0.040 7	0.068 7	0.108 6
C ₂	0.052 4	0.055 9	0.074 6	0.067 5	0.071 4	0.071 4	0.074 8	0.060 4	0.074 8
C ₃	0.073 7	0.066 8	0.064 3	0.077 1	0.080 9	0.075 6	0.053 0	0.073 1	0.080 9
C ₄	0.069 2	0.064 5	0.057 7	0.075 6	0.073 3	0.082 1	0.047 2	0.071 9	0.082 1
C ₅	0.069 9	0.079 7	0.067 6	0.086 1	0.089 3	0.087 5	0.053 5	0.078 9	0.089 3
C ₆	0.069 4	0.062 1	0.056 3	0.069 5	0.073 7	0.081 4	0.047 6	0.069 2	0.081 4
C ₇	0.072 0	0.068 3	0.054 1	0.076 1	0.075 6	0.086 4	0.048 0	0.072 6	0.076 1
C ₈	0.068 7	0.077 3	0.070 0	0.082 7	0.086 0	0.084 3	0.058 3	0.077 1	0.086 0
C ₉	0.066 8	0.062 4	0.052 2	0.071 8	0.071 7	0.082 3	0.045 7	0.068 4	0.082 3
C ₁₀	0.068 8	0.064 0	0.054 0	0.073 6	0.074 3	0.082 0	0.047 5	0.072 2	0.082 0
C ₁₁	0.056 1	0.060 0	0.072 5	0.074 1	0.072 7	0.065 8	0.060 7	0.067 7	0.074 1
C ₁₂	0.068 1	0.076 7	0.082 0	0.090 4	0.086 7	0.076 8	0.066 4	0.081 6	0.090 4
C ₁₃	0.063 2	0.062 2	0.064 7	0.058 1	0.058 0	0.053 4	0.079 8	0.061 7	0.079 8
C ₁₄	0.076 8	0.085 2	0.074 0	0.075 1	0.076 6	0.066 5	0.074 5	0.077 8	0.085 2
C ₁₅	0.065 4	0.063 3	0.064 2	0.077 8	0.073 3	0.074 9	0.054 6	0.070 7	0.077 8
C ₁₆	0.060 3	0.062 5	0.076 0	0.053 1	0.053 8	0.048 6	0.098 1	0.054 9	0.098 1
C ₁₇	0.061 1	0.069 3	0.078 0	0.063 9	0.062 7	0.058 0	0.088 9	0.062 9	0.088 9
C ₁₈	0.085 8	0.088 4	0.092 6	0.096 2	0.092 7	0.086 4	0.074 5	0.098 7	0.098 7

四、小结

本文主要研究商业银行零售客户的价值评价和客户终身价值评估,期望可以帮助银行识别、保留和发展零售客户,为银行正确认识客户的价值贡献提供决策支持。本文首次利用数据挖掘技术构建商业银行零售客户价值智能分类模型。数据挖掘技术不仅仅是研究工具和分类手段,它还与客户价值分类的理论研究一起构成的系统方法。在构建客户价值智能分类模型时,针对数据仓库中的零售客户原始数据,利用K-means方法获得案例库中的历史案例客户价值分类类型;在构建案例相似度模型中,本文首次利用ANP模型获得评价指标权重,强调了分类指标间的相互影响,同时利用GRA原理建立多层灰色关联方法以适应所构建的分类指标体系。

【注】本文系江苏省教育厅哲学社会科学基金项目(项目编号:09SJD630006)、教育部人文社会科学研究青年项目“面向商务智能的商业银行移动商务消费者信息挖掘与应用研究”(项目编号:11YJJCZH005)的阶段性研究成果。

主要参考文献

1. 权明福, 齐佳音, 舒华英. 客户价值评价指标体系设计. 南开管理评论, 2004; 7
2. 冯永, 钟将, 李志国, 叶春晓. 面向客户知识管理的动态SOM客户分类方法. 重庆大学学报(自然科学版), 2007; 11
3. 孟钊兰, 邵洪选. 商业银行客户分类新体系探索. 金融理论与实践, 2008; 3