

Customers Segmentation Based on Factor Analysis and Cluster

Qihui Zhang

College of Economics and Management, Shandong Jiaotong University, Jinan Shandong
Email: zqhunisun@163.com

Received: Apr. 16th, 2019; accepted: Apr. 30th, 2019; published: May 7th, 2019

Abstract

In order to individual customer relationship maintenance, this paper selects auction customers as the research object, using the method of factor analysis to summarize the eleven evaluation index as five evaluation indexes that are more easy to explain and comprehensive evaluation of customer and clustering are based on factor analysis. Aiming at the blindness set the number of clusters, the quality clustering criterion function to measure the number of clusters is put forward. Through calculation and analysis of the actual data, the experimental results show that the number of customer clusters is more ideal based on factor analysis and clustering criterion function. The clustering results has better explain ability and scientific customer segmentation ability on customer bidding behavior.

Keywords

Factor Analysis, Cluster Analysis, Customer Segmentation

基于因子分析和聚类的客户细分研究

张启徽

山东交通学院经济与管理学院, 山东 济南
Email: zqhunisun@163.com

收稿日期: 2019年4月16日; 录用日期: 2019年4月30日; 发布日期: 2019年5月7日

摘要

为了有针对性地进行客户关系维护, 文章选取在线竞拍客户为研究对象, 运用因子分析法把评价客户的

十几项指标概括为5项更容易解释的评价指标，并在因子分析的基础上对客户进行综合评价和聚类；针对聚类数目设置的盲目性，提出运用聚类准则函数来衡量聚类数目的质量；通过对实际竞拍数据的计算分析，实验结果表明基于因子分析和聚类准则函数的客户聚类数目更为理想，聚类结果表明该方法有更好的客户竞拍行为解释能力和科学的客户细分能力。

关键词

因子分析，聚类分析，客户细分

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

客户细分是现代市场营销里的重要理念和客户关系管理的重要策略，它既是客户关系管理中的核心方法之一，也是分门别类研究客户、进行有效客户评估及实施客户策略的基本原则之一[1]。不同类型的客户，需求、消费能力不同，不同消费能力的客户对企业来说客户价值不同。企业要想了解哪些是有价值客户，哪些是潜在客户，哪些是成长性客户，哪些是无价值客户就必须根据客户的属性，需求，行为，偏好等因素对客户进行分类，客户细分能够在复杂的消费者群体中准确识别出最有可能成为主导型客户群体[2]。同一组的客户具有相识的需求和爱好，根据分类特点提供有针对性的服务可以使企业产品和服务更适合那些可能购买他们产品的客户，从而实现客户价值的最大化。聚类是客户细分的主要方法，聚类能够将类别之间的数据特征差异度最大化，这也是衡量聚类算法有效性的一个重要指标[3]。国内外学者对因子分析和聚类在不同行业的应用进行了认真研究，但鲜有文献对在线竞拍客户进行研究，且多是使用 SPSS，Eview 等统计分析软件进行聚类，在聚类时需要事先设置分类数，SPSS，Eview 等软件根据用户设置的分类数统计出分类结果。用户分类数设置多是根据经验，但分类数设置是否合理鲜有文献给出验证，文章通过实现聚类准则度量函数程序验证了竞拍客户聚类数目的合理性。

2. 客户评价指标

以网上竞拍客户为例，网上拍卖客户竞拍目的的不同造就了竞拍行为的多样性，为便于客观评价，就要从多方面选取客户评价指标。从客户注册网站开始到不断参与竞拍的过程中产生了多项反映客户状况的指标，如注册日期、最后登录日期、参与竞拍数、得拍数、竞得后在规定期限内不付款迫使订单流拍的流拍数、竞得后拒绝付款取消订单的拒拍数、收货后不满意的退拍数、拍卖成交订单数、拍卖成交额、零售订单数、零售额(多数在线竞拍平台为了增加佣金收入允许零售)等多达十几项反映客户状况的指标。由于注册日期、最后登录日期不是尺度变量，用最后登录日期减去注册日期计算出客户持续使用在线拍卖平台的活动天数，用当前日期减去客户注册日期得出客户注册天数，用注册天数、活动天数分别代替注册日期和最后登录日期作为反映用户活跃程度的两个指标；用流拍数、拒拍数、退拍数与得拍数的比值：流拍率、拒拍率、退拍率来反映竞拍者的拍卖信誉。

3. 因子分析评价法

因子分析法以较少的几个公共因子变量反映原始指标变量的大部分信息，通过各指标的贡献度大小

确定权重，降低了分析问题的难度[4]，因子分析的方程为：

$$\begin{aligned}x_1 &= a_{11}F_1 + a_{12}F_2 + a_{13}F_3 + \cdots + a_{1m}F_m + \varepsilon_1 \\x_2 &= a_{21}F_1 + a_{22}F_2 + a_{23}F_3 + \cdots + a_{2m}F_m + \varepsilon_2 \\x_3 &= a_{31}F_1 + a_{32}F_2 + a_{33}F_3 + \cdots + a_{3m}F_m + \varepsilon_3 \\&\vdots \\x_p &= a_{p1}F_1 + a_{p2}F_2 + a_{p3}F_3 + \cdots + a_{pm}F_m + \varepsilon_p\end{aligned}$$

其中 $X = (x_1, x_2, \dots, x_p)$ 是可观测随机原始变量， $F = (F_1, F_2, \dots, F_m)$ 为公共因子； a_{pm} 为第 P 个变量在第 m 个因子上的载荷，表示第 P 个变量 x_p 对第 m 个因子 F_m 的依赖程度和相对重要性， ε_p 为误差项，分析因子分析的一般步骤为[5]：

- 1) 确认待分析的数据变量是否适合因子分析；
- 2) 进行初步因子分析，选取因子变量；
- 3) 因子旋转和对因子变量进行解释；
- 4) 计算因子得分和综合得分进行进一步分析。

4. 在线竞拍客户评价

本文选取竞拍者持续活跃天数、注册天数、零售订单成交额、零售成交额、拍卖成交订单数、拍卖成交额、参与竞拍数、得拍数、流拍率、拒拍率、退拍率共 11 个指标作为分析指标，采用因子分析法对在线竞拍客户进行评价。从中国收藏热线网站公开的客户信息中随机选取 180 位客户的上述 11 个指标数据作为样本数据进行因子分析，样本数据满足因子分析样本量为变量数的 10-25 倍的理想要求。

4.1. 原始数据的标准化

原始数据指标单位有的是天数、有的是个数、有的是金额、有的是比值，指标单位不同，如果不通过标准化处理来消除不同指标之间量纲的不一致和数量级别的差异容易造成分析结果的失真[6]，标准差标准化是最常用的数据标准化方法之一，经过这种方法处理后的数据符合均值为 0、标准差为 1 的正态分布，标准化公式为： $x^* = (x - \mu) / \sigma$ ， μ 为所有样本的均值、 σ 为所有样本的标准差。

4.2. KMO 和 Bartlett 检验

因子分析要求原始变量之间具有较强的相关性，因此首先要对原始指标变量进行相关性检验。因子分析常用的相关性检验是 KMO 检验和 Bartlett 检验，KMO 检验是为了看数据是否适合进行因子分析，Bartlett 球形检验是为了看数据是否来源于服从多元正态分布的总体。因子分析要 Bartlett 球形检验显著性水平 sig 小于 0.05，KMO 取值在 0 和 1 之间，0.5 以下表示不能做因子分析，0.5 以上表示适用[7]。利用 SPSS 统计软件对样本数据进行相关性检验得出样本数据的 KMO 为 0.753，球形检验 Bartlett 的 sig 为 0，如表 1 所示，表示样本数据比较适合因子分析。

Table 1. KMO and Bartlett Test
表 1. KMO 和 Bartlett 检验

提取足够量的 Kaiser-Meyer-Olkin 度量		0.753
	近似卡方	1575.28
Bartlett 球形检验	df	21
	Sig	0

4.3. 提取公共因子变量

首先对原始数据进行无量纲化处理, 然后利用 SPSS 软件对 180 位竞拍客户的竞拍数据进行因子分析, 分析结果见表 2 所示。

Table 2. Explained total variance
表 2. 解释的总方差

原始因子	初始特征值		
	特征值	方差的%	累计%
活动天数	3.937	35.787	35.787
注册天数	2.045	18.592	54.38
零售单数	1.351	12.282	66.661
零售额	1.018	9.253	75.914
拍卖单数	1.007	9.152	85.066
拍卖额	0.817	7.424	92.490
参拍数	0.422	3.840	96.330
得拍数	0.197	1.788	98.118
流拍率	0.162	1.472	99.59
拒拍率	0.042	0.383	99.973
退拍率	0.003	0.027	100.00

因子分析要求提取的因子对整体数据解释的累积方差贡献率不小于 85%, 前 5 个因子变量方差累计贡献率超过了 85%, 说明这 5 个因子变量涵盖的信息足够表达 11 个原始指标数据所能表达的信息, 因此选取这 5 个因子作为公共因子。

公共因子变量对原始变量进行了综合, 其所代表的实际意义不一定容易解释, 解决公共因子不能反映实际意义的一种方法是因子旋转, 因子旋转是将因子载荷矩阵(原始因子在提取的公共因子上的得分构成的矩阵)旋转后得到旋转成份矩阵, 11 个指标在 5 个因子上的载荷所构成的因子载荷矩阵如表 3 所示, 因子载荷矩阵旋转后得到的旋转成份矩阵如表 4 所示。

Table 3. Matrix of load of factors
表 3. 因子载荷矩阵

原始因子	提取的公共因子				
	1	2	3	4	5
活动天数	0.501	0.735	0.418	0.097	-0.037
注册天数	0.492	0.735	0.426	0.101	-0.046
零售单数	0.458	0.460	-0.614	0.158	-0.055
零售额	0.590	0.179	-0.674	0.250	0.031
拍卖单数	0.850	-0.308	0.181	-0.120	-0.016
拍卖额	0.833	-0.181	-0.132	0.056	0.104
参拍数	0.824	-0.391	0.151	0.026	0.011
得拍数	0.843	-0.446	0.185	-0.030	-0.048
流拍率	-0.107	-0.488	0.063	0.511	-0.232
拒拍率	-0.250	-0.042	0.206	0.752	-0.164
退拍率	-0.061	-0.044	0.099	0.254	0.951

Table 4. Rotating component matrix
表 4. 旋转成份矩阵

原始因子	提取的公共因子				
	1	2	3	4	5
活动天数	0.130	0.969	0.122	-0.072	-0.009
注册天数	0.124	0.971	0.113	-0.064	-0.016
零售单数	0.028	0.215	0.870	-0.125	-0.081
零售额	0.265	0.022	0.909	-0.025	0.023
拍卖单数	0.911	0.136	0.053	-0.108	-0.050
拍卖额	0.765	0.082	0.393	-0.067	0.085
参拍数	0.918	0.066	0.085	0.035	0.014
得拍数	0.970	0.043	0.031	0.018	-0.054
流拍率	0.139	-0.296	-0.093	0.670	-0.068
拒拍率	-0.187	0.101	-0.050	0.805	0.066
退拍率	-0.012	-0.018	-0.042	0.006	0.992

4.4. 公共因子变量的命名及解释

从表 4 旋转成份矩阵可以看出，旋转成份矩阵提高了提取的公因子的可解释性，具有较高因子载荷的公因子均匀分布在原若干评价指标上，说明旋转对因子载荷起到了明显的分离作用，使提取的各公因子都有了某种意义，据此可以对这 5 个公共因子变量进行命名与解释。

从表 4 旋转成份矩阵可以看出：第一因子(第一列)中的拍卖成交订单数、拍卖金额、参与竞拍次数、得拍数相关系数较大，说明第一因子主要反映了客户通过竞拍方式购买情况，命名为“拍卖因子”；第二因子中的持续活动天数、注册天数相关系数较大，说明第二因子主要反映了客户使用在线竞拍平台的活跃程度，命名为“活跃因子”；第三因子中零售订单、零售额相关系数相对较大，说明第三因子主要反映了客户的零售购买情况，命名为“零售因子”；第四因子中的流拍率、拒拍率相关系数相对较大，说明客户得拍后没有支付款项，订单没有成功交易，表明第四因子主要反映了客户的拍卖信誉，命名为“信誉因子”；第五因子中的退拍率相关系数较大，退拍反映了客户对商品的挑剔和苛刻要求程度，这是卖家最不愿意碰到的客户，命名为“挑剔因子”，旋转后，每个公因子变量都有了较明确的含义。

4.5. 计算因子得分

因子得分是每一样本数据在各因子上的具体数据值，这个数值就是因子得分。既然各变量可以表示为各因子的线性组合，反过来，各因子也可以表示为各观测变量的线性组合：

$$F_{ij} = w_{j1}x_{i1} + w_{j2}x_{i2} + w_{j3}x_{i3} + \dots + w_{jp}x_{ip} \quad (j = 1, 2, 3, \dots, m)$$

F_{ij} 为第 i 行样本在第 j 个因子上的得分， w_{jp} 为第 j 个因子在第 p 个变量处的因子得分系数， x_{ip} 为第 i 行样本的第 p 个变量标准化后的数据，因此根据标准化后的指标数据(表 5)和成份得分系数矩阵(表 6)就可以计算出各样本数据在各因子上的得分(表 7)。

由于流拍率、拒拍率、退拍率是反映竞拍者拍卖信誉不好的因子，因此在计算因子综合得分时，流拍率、拒拍率、退拍率的因子得分系数应加减号，这样高流拍率、高拒拍率、高退拍率的客户对应项的因子得分才低，综合得分也低，计算出的综合得分才能真实反映客户竞拍情况。如第一个样本数据在第一个因子拍卖因子上的得分 F_{11} 等于第一个样本对应的 11 个指标数据与第一个因子拍卖因子对应的 11

个指标得分系数乘积累计和，计算如下：

$$\begin{aligned}
 F_{11} &= -0.5043 * -0.018 + (-0.5337) * -0.019 + (-0.2165) * -0.111 + (-0.0017) * \\
 &\quad -0.026 + 0.3794 * 0.289 + 0.5298 * 0.202 + (-0.1731) * 0.293 \\
 &\quad + 0.1185 * (0.317 - 0.7431) * (0.079 - 0.3334) * (-0.051 - 0.4242) * 0.01 \\
 &= 0.2008
 \end{aligned}$$

同理可以计算出第一个样本在其它 4 个公因子上的得分 F_{12} 、 F_{13} 、 F_{14} 、 F_{15} 和其余样本数据在 5 个因子上的得分 F_{i1} 、 F_{i2} 、 F_{i3} 、 F_{i4} 、 F_{i5} 。

Table 5. Standardized customer data

表 5. 标准化后的客户数据

客户指标	客户编号				
	1	2	3	4	...
活动天数	-0.5043	-0.3079	-0.3466	-0.3452	...
注册天数	-0.5337	-0.3352	-0.3743	-0.3729	...
零售单数	-0.2165	-0.2311	-0.1824	0.2460	...
零售额	-0.0017	-0.5294	-0.3062	-0.0533	...
拍卖单数	0.3794	-0.3309	1.4810	-0.4723	...
拍卖额	0.5298	-0.5716	-0.1626	-0.6246	...
参拍数	-0.1731	-0.5072	0.0618	-0.3610	...
得拍数	0.1185	-0.4101	1.2284	-0.4935	...
流拍率	0.7431	0.3548	0.4927	-0.2457	...
拒拍率	0.3334	0.4628	0.4110	-1.4083	...
退拍率	0.4242	-1.096	0.2505	-2.0876	...

Table 6. Component score coefficient matrix

表 6. 成份得分系数矩阵

原始因子	提取的公共因子				
	1	2	3	4	5
活动天数	-0.018	0.496	-0.057	0.047	0.002
注册天数	-0.019	0.499	-0.061	0.055	-0.005
零售单数	-0.111	0.000	0.531	0.003	-0.054
零售额	-0.026	-0.109	0.564	0.083	0.050
拍卖单数	0.289	0.013	-0.107	-0.070	-0.039
拍卖额	0.202	-0.044	0.151	-0.003	0.102
参拍数	0.293	-0.013	-0.061	0.060	0.024
得拍数	0.317	-0.024	-0.105	0.037	-0.045
流拍率	0.079	-0.087	0.028	0.587	-0.078
拒拍率	-0.051	0.152	0.070	0.753	0.054
退拍率	0.010	0.002	0.003	-0.012	0.980

Table 7. Customer factor score and comprehensive score table
表 7. 客户因子得分及综合得分表

公共因子	客户编号				
	1	2	3	4	...
拍卖因子	0.2008	-0.4318	0.8238	-0.5690	...
活跃因子	-0.5221	-0.2623	-0.3493	-0.1167	...
零售因子	-0.0625	-0.3992	-0.5861	0.2848	...
信誉因子	-0.7702	-0.6682	-0.7126	1.1342	...
挑剔因子	-0.3327	1.0247	-0.3617	2.0562	...
综合得分	-0.1582	-0.2591	0.0692	0.1208	...

4.6. 计算综合得分

综合得分反映了每个样本数据在 5 个因子上的综合评价情况，样本数据综合得分 = Σ (各因子所对应的方差贡献率 * 各因子得分) \div Σ 各因子方差贡献率合计，例如第一行样本数据的综合得分：

$$ZF_1 = \left(\frac{0.35787 * 0.2008 + 0.18592 * (-0.5221) + 0.12282 * (-0.0625) + 0.09253 * (-0.7702) + 0.09152 * (-0.3327)}{0.85066} \right) = -0.1582$$

同理可计算出其他样本数据的综合得分(表 7)。

总的来说，综合得分高的竞拍者拍卖成交额、零售额、得拍率、活动天数相对较高、流拍率、拒拍率和退拍率相对较低。根据综合得分的高低可对竞拍者排名，但综合得分的高低还要辩证看待，具体分析是哪个因子影响了综合得分，例如有些注册时间较短的客户拍卖因子得分高，活动因子得分低，这样综合得分也低，但这些客户可能发展为长期重要客户；还有些拍卖因子得分低，但零售因子得分高的客户，这些客户可能没有充足的时间参与竞拍。显然这些综合得分低的客户不同于因拒拍率、流拍率和退拍率高而使综合得分低的客户，因此有必要在因子分析的基础上对客户进行聚类分析。

5. 客户聚类实证

常用聚类算法把目标客户划分为几个具有明显特征区别的细分群体，同一群体的客户具有较大的相识性，不同群体的客户具有较大的差异性[8]，针对不同客户群体提供个性化的服务以提升运营效率和商业效果。虽然使用 SPSS、Eview 软件很容易得出聚类结果，但聚类数目的合理性无从检验，下文通过实证检验了聚类数目的合理性。

5.1. 相识性度量方法-距离

对于数据集 $X = \{x_1, x_2, x_3, \dots, x_n\}$ ，其中每个对象， $x_i = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip}\}$ ，即每个对象有 p 个属性值构成(如竞拍客户有在线天数、竞拍次数、竞拍额、退拍次数等指标属性)。任意两个对象 x_i 和 x_j 之间的相识程度常用两个对象之间的距离 $d(x_i, x_j)$ 来衡量，计算距离的方法很多，在聚类算法中以欧式距离最为常用[9]，欧式距离的计算公式为：

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2}$$

5.2. 聚类准则函数

常用误差平方和准则函数来衡量聚类结果的质量，数据集 $X = \{x_1, x_2, x_3, \dots, x_n\}$ 分为 C 个聚类， $W_j \{j = 1, 2, 3, \dots, C\}$ 为 C 个聚类中的第 j 个聚类， x_j 表示 W_j 中的任意数据对象， Z_j 为聚类 W_j 的中心，

$$z_j = \frac{1}{n} \sum_{j=0}^n x_j$$

误差平方和准则函数公式为:

$$J_c = \sum_{i=1}^c \sum_{x \in w_j} (x_j - z_j)$$

从公式可以看出误差平方和表示的是把 n 个数据对象分为 C 个类时, 分类样本与其所在类中心的差值平方和。 J_c 越小, 说明各类内对象与其中心越紧凑, 类内数据对象间相似程度越大, 聚类质量越好。

5.3. C-均值聚类算法

C-均值聚类算法是目前应用最广泛最成熟且快速简单的聚类分析方法, C-均值聚类算法的步骤为:

- (1) 按最大距离原则选取 C 个初始聚类中心;
- (2) 按最小距离原则将待分类对象逐个分到离它最近的 C 个类的某一类, 并计算误差平方和 J_c ;
- (3) 重新计算分配后的各类心和误差平方和 J_c ;

若 J_c 收敛, 则算法结束, 输出结果; 否则转至(3)。

5.4. 客户聚类实证

用 C#语言实现的 C-均值聚类算法, 主要代码如下:

```

Console.WriteLine("请输入预分类数:");
String tempnum = Console.ReadLine();
int num =int.Parse(tempnum);
double[,] arrxin = new double[num,5];
arrxin = datadeal.getchushi(arr1,num);
double jc1 = 0,jc2 = 0;//保存误差平方和
jc1= datadeal.getfenlei(arr1,arrxin );
jc1=Math.Round(jc1,3);
Console.WriteLine("分类数为{0};收敛时-", num);
for (int i = 1; i < 20;i++)//重置类心
{
    arrxin = datadeal.newcenter(arr1, num);
    jc2 =datadeal.getfenlei(arr1,arrxin);
    jc2 =Math.Round(jc2,3);
    if (jc2== jc1)
    {
        Console.WriteLine("误差平方和为:{0}\n",jc2);
        break;}
    jc1 = jc2;
}
string outstr ="分类数为{0}时的分类结果:\n";
Console.WriteLine(outstr, num);
Console.WriteLine("{0}\t{1} {2} {3} {4} {5} {6}\n","客户号","拍卖因子","活动因子","零售因子","信誉因子","挑剔因子","分类");
for (int i = 0; i < arr1.GetLength(0); i++)
{
    for (int j = 0; j < arr1.GetLength(1); j++)
        Console.WriteLine("{0}\t", arr1[i, j]);
    Console.WriteLine("\n");
}

```


不同分类数的误差平方和如表 8 所示, 由表 8 可知在类为 4 时, 误差平方和最小, 因此分类前设置分类数为 4 是最合理的, 分类数为 4 时的各客户具体归类结果如表 9 所示。

Table 8. Sum of squares of errors for different classifications

表 8. 不同分类数的误差平方和

分类数	迭代收敛时误差平方和
3	10.687
4	3.114
5	12.349
6	4.331

Table 9. Clustering results when the number of classifications is 4

表 9. 分类数为 4 时的聚类结果

客户编号	拍卖因子	活动因子	零售因子	信誉因子	挑剔因子	客户聚类类别
1	0.2006	-0.5216	-0.6287	0.7696	0.3319	III
2	-0.4315	-0.2626	-0.3991	0.668	-1.0258	I
3	0.8232	-0.3486	-0.586	0.713	0.3599	III
4	-0.5683	-0.1164	0.285	-1.134	-2.0566	I
5	-0.0104	2.0672	-0.8479	-0.6934	-3.5075	I
6	2.5442	0.0873	1.7612	0.1532	0.0558	IV
7	-0.3893	-0.7604	-0.4841	0.508	0.4346	III
8	0.3549	-0.2902	0.6945	-0.1396	-0.2068	IV
9	-0.8817	0.4878	0.2461	-7.6969	0.0565	II
10	0.057	0.3241	-0.481	0.4529	-1.382	I
11	-0.3849	-0.7726	-0.1838	0.609	-1.4849	I
...

5.5. 客户聚类解释

客户聚类分类结果如表 9 所示, 因篇幅限制仅列出 11 个客户, 根据这 11 个客户的因子得分及因子得分对应的原始数据可以将这 4 类客户概括如下:

I 类: 低信誉客户。这类客户的拒拍率、流拍率和退拍率都较高。拒拍率、流拍率较高的客户多是自己有商品在拍卖, 通过托拍来提升自己拍品的拍卖价格, 由于无人追拍而使自己得拍, 得拍后只能以流拍或拒拍的方式来结束拍卖, 影响了正常的竞拍秩序。退拍率较高的客户对拍品品相要求严格, 这类客户得拍后, 可提醒卖家在发货前就拍品品相与买家多沟通。低信誉竞拍客户无论注册时间长短, 活跃程度如何, 由于真实成交少, 对贡献利润低, 对于低信誉客户如果得拍后拒绝交易可以扣除其一定的押金或进一步降低其信誉值或授权卖家将其拉入黑名单。

II 类: 一般价值客户。一些客户虽然注册时间较长, 也有一些拍卖或零售成交记录, 但现在很少登录网站或虽经常登录网站但较少参与竞拍或零售购买商品。还有一些客户注册时间不长, 也不经常登录网站或虽经常登录网站但较少参与竞拍或零售购买商品。这些客户的特征是不管注册时间长短, 登录频繁与否, 现在较少参与竞拍或零售购买商品, 这类客户可能受制于自身经济条件只关注喜欢的商品, 对这类客户可以分析其以前的竞拍喜好, 通过手机短信的方式欢迎用户常来网站看看和有针对性地推荐其可能喜欢的拍品。

III类：新价值客户。这类客户虽然注册时间不长，但经常登录网站参加竞拍或者以零售方式购买商品，累计拍卖成交或零售成交逐渐增大。这类客户是在线竞拍网站客户的新生力量，这类客户有的是真正喜欢收藏，有的则可能是一时兴趣，他们可能分化为稳定价值客户或一般价值客户。对这部分客户要实时跟踪，关注哪些客户可能转化为稳定价值客户，对其佣金可以根据其购买情况浮动管理，发展壮大这部分客户并使之转化为稳定价值客户是提升在线竞拍网站佣金收入的关键。

IV类：稳定价值客户。这类客户注册时间较长，经常登录网站参加竞拍或以零售方式购买商品，并且拒拍率、流拍率和退拍率较低，拍卖成交额或零售成交额都较大，这类客户是网站拍卖藏品的真正爱好者，有较强的购买欲望和购买能力，是在线竞拍网站佣金收入的主要来源，稳定这部分客户对维持在线竞拍网站的运营至关重要，可以降低其佣金比例或根据其成交额返回一定的佣金或赠送自制的有意义的纪念品。

6. 结语

文章利用因子分析法将竞拍客户十多个评价指标综合为5个指标，使客户竞拍行为更容易解释，并在因子分析的基础上对客户进行聚类分析，并用程序检验了聚类数目的合理性。文章为在线竞拍网站提供了一种有理论基础的客户细分解决方案，提高了客户细分的科学性，便于在线竞拍网站针对不同的客户分类提供有效的客户关系管理措施。当今社会进入大数据时代，利用数据挖掘技术分析客户的消费行为将会极大化客户价值，有利于企业目标的实现。

参考文献

- [1] 何媛. 基于客户关系管理中客户细分的研究与应用[J]. 现代营销, 2017(8): 53.
- [2] 梁妮. 网络营销环境下的客户细分与营销策略[J]. 科技经济导刊, 2016(32): 247.
- [3] 黄亚萍, 李垣江. 基于 K-means 算法的电子商务客户细分研究[J]. 电子设计工程, 2017, 25(2): 63-66.
- [4] 马健悦. 我国城镇居民消费水平的聚类分析和因子分析[J]. 中国商论, 2018(2): 74-75.
- [5] 麻雅静, 武刚. 基于因子分析与风险矩阵的 IT 外包风险评价[J]. 计算机工程, 2013, 39(9): 281-284.
- [6] 卢辉. 数据挖掘与数据化运营实战思路、方法、技巧与应用[M]. 北京: 机械工业出版社, 2013: 130-131.
- [7] 贺盛瑜, 马会杰, 滕喜华. 基于因子分析和聚类分析的我国电子商务发展水平研究[J]. 经济体制改革, 2017(2): 196-200.
- [8] 曾小青, 徐秦. 基于消费数据挖掘的多指标客户细分新方法[J]. 计算机应用研究, 2013, 30(10): 44-47.
- [9] 吴明礼, 黄亚非. 基于聚类的多指标客户细分方法[J]. 电脑知识与技术, 2018, 14(5): 18-21.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2168-5843, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: ecl@hanspub.org