

基于改进 RFM 模型与聚类算法的民航客户细分研究

白 钊¹, 王神武², 王梦杨²

(1. 中国民用航空飞行学院 经济与管理学院, 四川 广汉 618307;

2. 中国民用航空飞行学院 机场学院, 四川 广汉 618307)

摘要:对于传统的 RFM 模型客户细分而言,由于缺乏科学的指标而无法针对航空市场进行精确的细分,因此提出 TLFM 模型作为细分的指标,利用自组织映射(SOM)算法和“手肘法”确定 k -means 算法的最佳 k 值,从而得到航空市场客户细分的结果。最后根据客户细分的情况和各类客户的行为特征,提出差异性的服务和不同的营销策略来提高航空公司的竞争力。

关键词:TLFM 模型;自组织映射(SOM)算法; k -means 算法;客户细分

中图分类号:F562.6 **文献标志码:**A **文章编号:**1671-1807(2023)19-0200-04

在传统的营销模式中,大部分企业会对市场进行详细的调查,针对客户已有产品的偏好进行相应的供需调整。但这种营销模式存在一定的误差性,会导致营销资源的浪费和营销效率低下。针对传统的营销模型,美国学者温德尔·史密斯在1956年就提出客户细分的理论,根据客户的行为价值和自身的优劣势来制定对应的营销策略来提高企业的收益,并在各个行业得到运用和发展^[1]。在航空市场中,对于市场客户的细分已经有了一定的研究。赵崇希^[2]尝试用 RFMP 模型和聚类算法对客户进行了细分,并为某企业制定了一套精准的营销策略;杨琳等^[3]用改进的 RFM 模型和 k -means 算法对航空市场客户进行了分类,并针对各类航空客户提出了相应的营销策略;邢容^[4]在 2022 年利用客户细分和客户关系管理理论,对客户关系发展、维护、流失管理策略进行了重构和优化。但还存在以下两个问题:

1)对于指标的选取。划分不同的客户群体时利用特定指标来进行聚类分析是整个客户群体划分过程中的重要部分,传统的 RFM 模型被广泛地运用在各个行业的客户价值评价,但在航空市场客户细分中,需要改进某些指标使航空市场细分更加具有针对性,便于在后续的营销和服务中进行差异化管理。

2)传统的 k -means 算法对于 k 值的选取一般是

人为设定来进行聚类分析。由于是人为设定的 k 值,所以存在一定的主观性,对于聚类的分析过程存在一些人为主观性,会导致在后续制定营销计划时得不到预期的收益。

因此主要从以下两点来进行解决和改善:

1)对于 RFM 模型在航空市场中客户细分指标的针对性,选取的指标有:入会时间(T-time),指航空市场客户对于某个航司的第一次接触时间;会员级别(Level),是指某个客户在本航司的会员级别;订票总金额(Monetary),是指该会员在该航司消费的总金额;飞行次数(Frequency),是指某客户在该航司乘坐的次数,从而构建客户价值评价指标 TLFM 模型。

2)对于传统的 k -means 算法而言,初始 k 值的确定是通过人为主观进行确定。在本文中采取自组织映射(self-organizing map, SOM)算法对所有的样本进行初次聚类,再对比“手肘法”来进行综合分析得到初始聚类 k 值。

1 模型与算法介绍

1.1 改进的 RFM 模型

RFM 模型是衡量客户价值和客户创造利益能力的重要工具和手段,在众多的客户关系管理(customer relationship management, CRM)的分析模式中,RFM 模型应用最广泛^[5]。RFM 模型最早被提出是在 1994 年, Hughes^[6]站在企业的角度来综合

收稿日期:2023-06-28

基金项目:中国民用航空飞行学院研究生科研创新项目(XSY2022-24)。

作者简介:白钊(1971—),男,四川成都人,中国民用航空飞行学院经济与管理学院,副教授,研究方向为民航运输管理、人力资源管理;王神武(1998—),男,湖北荆州人,中国民用航空飞行学院机场学院,硕士研究生,研究方向为民航运输管理;王梦杨(1997—),女,山东淄博人,中国民用航空飞行学院机场学院,硕士研究生,研究方向为民航运输管理。

考量客户的购买行为的分析模型。主要是通过 3 个指标 R (recency, 客户最近一次的购买时间)、 F (frequency, 客户在某时间内的购买频率)、 M (monetary, 客户在某个时间段的消费总金额) 来决定客户价值的高低。在民航市场客户价值的分析中, 应该加以改进这 3 个指标, 本文中所选取的是入会时间(T)、会员级别(L)、飞行次数(F)和订票总金额(M)这 4 个指标构建 TLFM 模型, 利用 k -means 算法对 U 航空公司的多个变量进行聚类分析得到旅客的分类。

1.2 改进的 k -means 算法与 SOM 神经网络

k -means 算法是指把数据的样本聚集成预定的聚类数, 先在样本数据中随机抽取几个样本作为原始的聚类中心, 然后通过距离的计算将最靠近聚类中心的数据点作为一类, 在此基础上通过不断地迭代, 根据迭代情况来更换聚类中心的值, 直到收敛到最优的聚类中心, 达到最合适的聚类结果^[7]。对于聚类中心 k 值的确定在传统的 k -means 算法中, 聚类数一般是需要预先给定, 存在一定的人为主观性。因此在确定原始 k 值聚类中心的个数时需要根据数据进行判断。本文中综合运用自组织映射(SOM)神经网络和“手肘法”来确定初始 k 值。SOM 神经网络是一种无导师学习方法, 具有很好的自组织、可视化和简单聚类的特性, 目前已经得到了学术界的广泛运用和研究^[8]。

1.3 模型的构建

根据改进的 RFM 模型, 通过 SOM 算法进行初步聚类并通过“手肘法”进行最佳 k 值的确定, 将最佳 k 值代入算法中进行计算得到聚类结果。

SOM 神经网络的基本步骤如下。

设 $x = \{x \in R^d\}$ 为 d 维样本向量的集合, A 为所有神经元的集合, m_i 为第 i 个神经元的权重。

1) 输出层的权值初始化。用小的随机数初始化权值。

2) 在 t 时刻, 随机加入一个样本记为 $X(x_t)$ 。

3) 利用欧式距离计算得到当前的获胜节点 C 。此时获胜节点 C 为

$$C = \min\{\|x(t) - m_i(t)\|, i \in A\} \quad (1)$$

式中: $m_i(t)$ 为在 t 时刻第 i 个神经元的权重。

4) 对于权值的竞争训练。对于所有的神经元点, 用下面的准则来进行更新:

$$m_i(t+1) = m_i(t) + a(t)h_{ci}(t)d[x(t), m_i(t)] \quad (2)$$

式中: $a(t)$ 为学习的步长; $d[x(t), m_i(t)]$ 为两个向量之间的欧式距离; h_{ci} 为节点 i 与节点 c 之间的近邻函

数值; $m_i(t+1)$ 为下一时刻第 i 个神经元的权重。

5) 更新 $a(t)$ 和领域 $NC(t)$ 。若达到终止的要求, 则算法停止; 否则重置 $t = t + 1$ 重复上面的步骤。

用手肘法确定 k 值。手肘法是一种常规的 k 值确定的方法, 其主要的步骤如下。

1) 计算所有 k 值对应的误差平方和 (sum of squared errors, SSE)。

$$SSE = \sum_{i=1}^k \sum_{x \in m_i} |d(X, C_i)| \quad (3)$$

式中: SSE 为所有样本的聚类误差; X 为某个样本; C_i 为第 i 个质心; $d(X, C_i)$ 为样本到质心的距离。

2) 利用工具制作 SSE 与 k 值对应关系折线图。

3) 通过关系折线图找到最佳的 k 值。

根据 SOM 算法和手肘法确定最优值, 将最优值作为最优 k 值代入 k -means 算法, 其重要的步骤为: ①确定最优 k 值(由 SOM 算法和手肘法确定); ②计算样本到每个聚类中心的欧式距离; ③根据计算得到样本到每一个聚类中心的欧式距离, 距离最小的就归于某一类。

2 U 航客户细分实例分析

2.1 数据获取

对 U 航空公司的客户进行细分, 从 U 航空公司中随机抽取 2019 年 1 月 1 日至 2019 年 6 月 30 日 6 000 名客户的订票记录, 其中包含会员编号、性别、入会时间、会员等级、飞行总里程、订票总金额和飞行次数等。从中选取 TLFM 模型所需要的会员编号、入会时间、订票总金额和飞行次数指标数据。

2.2 数据处理

由于原始数据具有噪声大、信息缺失和不一致的情况, 需对原始数据进行清洗。将 6 000 条数据处理筛查之后得到 4 520 条数据。因数据量较大采用 IBM SPSS Modeler 来对数据进行进一步的处理, 为了满足计算工具的需要将数据进行量化处理。具体的指标划分及量化情况见表 1。

对于 SOM 算法和 k -means 算法计算, 由于各指标的单位不同可能会影响聚类结果, 因此将量化之后的数据集利用 spss 进行标准化处理, 表 2 为部分标准化之后的数据。

2.3 算法实现

将数据使用 MATLAB 进行 SOM 算法计算得到第一次的简单聚类, 共有 13 个, 再计算所有 k 值对应的误差和 SSE 值得到对应关系折线图, 如图 1 所示。

表 1 指标划分及量化

指标	指标划分	量化值
客户入会时间	2010 年前	4
	2011—2013 年	3
	2014—2016 年	2
	2017—2019 年	1
会员级别	普通卡	1
	银卡	2
	金卡	3
	白金卡	4
	终身白金卡	5

表 2 部分标准化处理之后数据

入会时间	会员级别	订票总金额	飞行次数
1.958 67	0.554 69	0.284 81	-0.032 22
0.845 41	0.554 69	-0.585 27	-0.681 12
-0.267 85	1.981 62	-0.152 72	-0.032 22
0.845 41	0.554 69	-0.430 15	-0.551 34
0.845 41	0.713 46	0.139 63	0.227 35
0.845 41	3.249 78	4.428 36	3.082 52

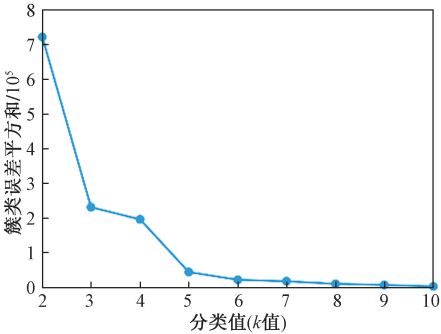


图 1 “手肘法”折线图

通过观察折线图能观察到图中出现了二次较大的“肘处”，其分别是 $k=3$ 和 $k=5$ 处，对于 $k=5$ 处其误差更小，因此根据手肘法所选取的最佳 k 值应该为 5 类。对于 SOM 算法简单聚类得到的聚类结果为 13 类，通过观察“手肘法”的折线图能看到当 $k=13$ 时误差是相对于 $k=5$ 更小的，但是在实际营销过程中，针对 13 类客户制定不同的营销策略会导致成本过大，因此最佳的 k 值可以选取为 5，然后对于处理之后的数据利用 k -means 算法进行细分计算，将细分之后的群体定义为核心客户(A 类)、高价值客户(B 类)、次价值客户(C 类)、重要客户(D 类)和一般客户(E 类)。客户特征雷达图如图 2 所示。

2.4 结果分析

根据图 2 可知，客户群体 1 的所有指标属性都较低，因此将其定义为一般客户(D 类)；客户群体 2 的 T 、 L 和 F 指标属性都较高，但出行次数不是特别高，因将其定义为高价值客户(B 类)；客户群体 3 的 L 、 F 和 M 的指标属性和第一类较相似，但属于航空

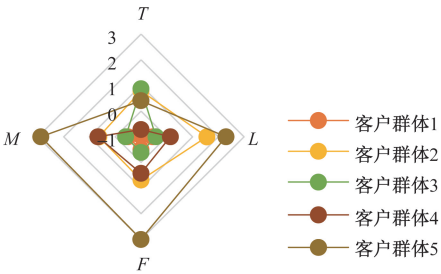


图 2 客户特征雷达图

公司的老客户，可能在这个时间段乘坐较少，因此将其定义为重要客户(D 类)；客户群体 4 相较于群体 2 只有 T 偏低一点，可以定义为次价值客户(C 类)；客户群体 5 所有指标属性都是最高的，因此将其定义为核心客户(A 类)。客户的最终分类和聚类结果见表 3。

根据表 3 的聚类结果可以得到各类客户群中 A 类客户群体占据 5.8%，其他各类占比分别为 16.3%、6.4%、37.6% 和 33.4%，具体各类客户群体占比如图 3 所示。

通过聚类结果和各类客户群体的指标属性，航空公司可针对客户细分的结果来制定相应的营销策略和提供差异性的服务^[4]。

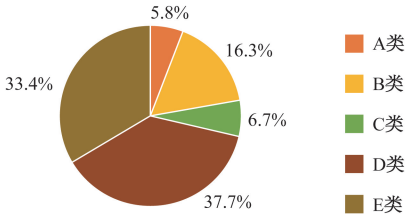


图 3 客户种类占比

表 3 聚类结果

客户分类	T	L	F	M	聚类数
核心客户(A 类)	0.405 2	2.299 9	2.998 9	2.882 7	263
高价值客户(B 类)	0.789 5	1.558 3	0.682 5	0.634 8	737
次价值客户(C 类)	-0.717 5	0.139 3	0.425 3	0.671 6	302
重要客户(D 类)	0.869 6	-0.420 7	-0.392 7	-0.403 5	1 703
一般客户(E 类)	-0.963 1	-0.539 6	-0.499 0	-0.493 4	1 510

3 客户细分结果分析及建议

核心客户(A 类)只占 5.8%，但能带来 27.7% 的收益。此类客户消费频率较高，在半年时间里平均消费次数达到了 30 次，人均消费了 3.82 万元，且

这类客户入会的时间较长,客户的黏性较大;同时这类客户会员级别也很高,主要以白金卡和金卡会员为主。因此航空公司对于 A 类客户应该视为优先考虑的对象,针对这类客户需要差异化管理,多与客户沟通获取客户的需求,如为这类客户提供地面和空中 VIP 服务、开辟绿色通道减少候机时间和在空中增加智能点餐和免费电影点播等差异性服务,由此使这类客户对航空公司的满意度和忠诚度得到保持和提高。

高价值客户(B类)占据总客户的 16.3%,能为航空公司带来 30%的收益,这类客户半年内的人均消费达到了 1.5 万元,平均出行次数为 12 次,并且大部分客户入会时间都相对较长,主要以金卡会员为主。这类客户是航空公司的中坚力量,同样需要有一定的“特权”,如提供准点高效的航班和行李托运差异化等方式来加强这类客户对公司的信任度,延长 B 类客户的生命周期。

次价值客户(C类)虽只占据总客户量的 6.7%,却带来了 10.4%的收益,并且半年内人均出行数达到了 12.4 次,但这类客户的入会时间较短,是具有优质的潜力客户,航空公司需要投入一部分的资源,如提供更多的购买渠道和较大的商品折扣来吸引这类客户成为自己的忠实客户。

重要客户(D类)占据总客户的 37.7%,却只能为公司带来 19.2%的收益,人均出行量 4.1 次,但从客户入会时间来看,时间较长。因此对于这类客户应该在保持原有的基础上增加和老客户的沟通,推出一些反馈老客户的活动,如赠送一些纪念品、周年折扣票等来避免客户的流失。

一般客户(E类)占据 33.4%,给航空公司带来的收益只占 12.7%,这类客户人群占比较多,一般为冲动型消费人群,在较大的折扣机票时会选择出行。因此对于这类人均消费较低、出行频次低的客

户可提供一些团队购票的活动去吸引出行或者提高客座率。

4 总结

首先分析传统的营销模式和客户细分的特点,并对比民航企业与传统行业客户细分的区别提出了 TLFM 模型。随后采取 k -means 算法对客户进行细分,在 k -means 算法中最佳 k 值采取了“手肘法”和 SOM 算法来进行确定,SOM 算法主要是对原始数据进行初次聚类,“手肘法”主要是计算所有的 k 值来观察对应的误差,最终通过 SOM 算法和“手肘法”共同确定最佳的 k 值,从而将航空市场客户细分为核心客户(A类)、高价值客户(B类)、次价值客户(C类)、重要客户(D类)和一般客户(E类)这 5 类,并对这 5 类客户进行分析和解释,给航空公司提供了对应的营销策略和改进的方法。这些策略不仅能提升航空公司的收益,还能提高航空公司在竞争激烈的民航企业中保持自己的竞争力。

参考文献

- [1] ULAGA W. Customer value in business markets:an agenda for inquiry[J]. Industrial Marketing Management, 2001,30(4):315-319.
- [2] 赵崇希. 基于客户价值细分的 A 公司精准营销策略[D]. 绵阳:西南科技大学,2019.
- [3] 杨琳,寇勇刚,白钊. 基于改进 RFM 模型对民航客户的细分研究[J]. 数学的实践与认识,2021,51(1):33-39.
- [4] 邢容. Y 航空地面保障服务公司客户关系管理研究[D]. 呼和浩特:内蒙古财经大学,2022.
- [5] 杨超凡. 考虑用户行为细分的定向短信广告效果影响因素研究[D]. 南京:南京理工大学,2020.
- [6] HUGHES A M. Strategic database marketing[M]. New York:Mc Graw-Hill Pub. Co.,2005:15-20.
- [7] 杨倩倩,生佳根,赵海田. K-means 聚类算法在民航客户细分中的应用[J]. 电子设计工程,2015,23(12):25-27.
- [8] 付丽梅. 基于 K-means 优化的 SOM 神经网络算法的视频推荐系统[J]. 软件工程,2022,25(10):17-19,7.

Research on Civil Aviation Customer Segmentation Based on Improved RFM Model and Clustering Algorithm

BAI Zhao¹, WANG Shenwu², WANG Mengyang²

(1. School of Economics and Management,Civil Aviation Flight University of China,Guanghan 618307,Sichuan,China;

2. Airport College,Civil Aviation Flight University of China,Guanghan 618307,Sichuan,China)

Abstract: As for the customer segmentation of the traditional RFM model, due to the lack of scientific indicators, it is impossible to conduct accurate segmentation for the aviation market. Therefore, the TLFM model is proposed as the index of segmentation. SOM algorithm and “elbow method” are used to determine the best k value of k -means algorithm, so as to obtain the results of the customer segmentation of the aviation market. Finally, according to the situation of customer segmentation and behavior characteristics of various customers, different services and different marketing strategies are proposed to improve the competitiveness of airlines.

Keywords: TLFM model;self-organizing map(SOM)algorithm; k -means algorithm;customer segmentation