

基于RFMD模型和聚类分析的 健身休闲行业客户分类

陈燕萍, 王文杰

东华大学旭日工商管理学院, 上海

收稿日期: 2021年4月18日; 录用日期: 2021年5月12日; 发布日期: 2021年5月19日

摘 要

本研究改进了原始RFM模型, 构建了新的客户价值识别模型RFMD, 用于对健身休闲行业客户进行分类, 并根据所提出的模型识别出该行业的不同价值客户群体。研究使用了上海一家民营体育场馆的真实数据, 在RFM模型中新增客户平均消费持续时长指标D, 并采用两步聚类与K-means聚类相结合的两阶段聚类算法对客户进行聚类。研究结果表明, 基于RFMD模型可将健身休闲行业客户分为“重要价值客户”、“一般价值客户”和“低价值客户”。

关键词

客户价值, 聚类分析, 健身休闲行业, 客户分类, RFM模型

Customer Segmentation of Fitness and Leisure Industry Based on RFMD Model and Cluster Analysis

Yanping Chen, Wenjie Wang

Glorious Sun School of Business Management, Donghua University, Shanghai

Received: Apr. 18th, 2021; accepted: May 12th, 2021; published: May 19th, 2021

Abstract

This paper improves the original RFM model and builds a new customer value recognition model RFMD, which is used to classify customers in the fitness and leisure industry and identify different value customer groups in the industry based on the proposed model. The study used real data

from a private sports stadium in Shanghai, added a customer average consumption duration indicator D in the RFM model, and used a two-stage clustering algorithm combining two-step clustering and K-means clustering to gather customers kind. The research results show that based on the RFMD model, customers in the fitness and leisure industry can be divided into “important value customers”, “general value customers” and “low value customers”.

Keywords

Customer Value, Clustering Analysis, Fitness and Leisure Industry, Customer Segmentation, RFM Model

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

在人工智能热潮席卷全球之时,越来越多的研究利用机器学习算法从海量数据中探寻隐藏的价值信息,实现客户类型识别,预测客户的行为,以及发现未来的规律信息,为企业针对不同类型客户采取针对性保留措施提供指导,推动企业客户关系管理战略的实施,减少企业在吸引新客户和维系老客户方面的成本。

在国家经济的发展下,人们生活水平不断提高,对美好生活的需要日益增长。人们的消费观念也逐渐发生了转变,消费需求也愈加旺盛。随着健康消费理念的深入人心,越来越多的消费群体在闲暇时间走进体育健身、休闲娱乐场所,健身休闲成为个人消费的新热点,成为潮流趋势。2016年10月,国务院办公厅印发《关于加快发展健身休闲产业的指导意见》,提出大力支持健身休闲产业发展,积极推动“互联网+健身休闲”、“大数据+健身休闲”的发展,产业总规模在2025年将达到3万亿元。在国家的大力助推下,功能逐渐完善、门类逐渐齐全的健身休闲场所也越来越多,健身休闲企业间的竞争也逐渐加剧,如何实现现有客户的分类管理和保留,为客户提供更具针对性的服务,提高他们对企业的满意度和忠诚度也是当前健身休闲企业应该面临的问题。

本文基于上海L体育场馆的客户消费数据,探究健身休闲行业客户分类问题。根据行业客户消费特点改进原始客户价值识别模型RFM模型,形成具有行业客户消费特点的RFMD模型,并基于改进模型采用聚类分析方法对所有客户群体进行价值划分。通过将所有客户划分成了多个更小的、有意义的、相对同质化的客户群体,识别出该行业的重要价值客户,有利于企业了解不同价值客户的消费偏好,为企业实现更精准的客户分类提供指导,使企业能更好地响应客户需求,提高竞争力。

2. 文献综述

市场细分的概念最早由Smith于1956年提出,他认为整个市场可以看作是一组内部具有同质性的异质细分市场[1]。自被提出以来,市场细分得到了许多学术研究的进一步支持,被认为是学术界和实践中日益重要且被广泛接受的概念之一。

细分对于了解相应市场的特征、预测客户行为、发现和探索新的市场机会以及识别值得追求的群体具有显著贡献[2][3]。客户行为细分基于预测理论,根据客户过去和当下的消费行为对客户未来行为进行预测,是一种基于行为模式数据和信息技术的细分方法[4]。

RFM 模型最初由 Hughes (1996)提出, 它根据数据库中过去的客户行为活动来分析和预测客户的行为 [5]。它由三个度量组成, 即近度 R (Recency)、频度 F (Frequency)和值度 M (Monetary), R 指自上次购买以来的时间间隔(如天或月), F 衡量在特定时期内的购买次数, M 代表在特定时期内的购买总金额。随着以客户为导向的细分方法的出现, RFM 模型在分析和预测客户行为, 进而进行客户细分方面的应用越来越广泛。

RFM 模型在过去的二十几年中得到了发展, 一些研究尝试通过考虑额外的变量或根据产品或服务的性质排除一些变量来开发新的 RFM 模型, 验证它们是否比传统的 RFM 模型表现更好。很多学者在考虑到原始模型指标的不足以及不同行业客户行为特点时, 通过移除或替换部分指标对 RFM 模型进行了扩展, 使之在特定行业的客户行为分析上具有更强的适用性。如郭崇慧等(2015) [6]认为 R 指标在汽车维修客户细分中的作用不大, 因此移除 R, 并新增客户保留时长 Y 和客户汽车行驶里程数 K 两个指标, 构建了新的 YKFM 模型。张斌等(2017) [7]根据铁路货运客户行为的特点, 以近期发货行为能力 K 替代 R, 以货运收入贡献率 A 替代 M, 并加入了发货倾向预测值 V, 使改进模型更适用于铁路货运客户的细分。任春华等(2019) [8]基于 F 和 M 存在共线性的问题, 以平均购买金额 A 替代 M, 并新增平均购买时间 T, 发展了 LRFAT 模型。

近年来, 不同类型的 RFM 模型在美发、汽车、旅游、铁路货运等各种行业的客户细分中得到应用并表现良好, 这些关于不同领域的大量研究已经表明了学术界对这个主题的持久兴趣, 以及 RFM 模型在理解客户消费行为和细分客户方面的效率。

聚类是实现根据客户的需求和偏好的相似性对他们进行分组的常用工具, 客户聚类, 又称客户细分, 是现代营销和客户关系管理成功的关键因素之一 [9]。聚类分析是市场细分分析中最常用的技术, 它根据客户之间的相似性或差异性对客户进行分组。基于 RFM 模型细分基础和聚类技术的客户细分方法已在不同行业的客户划分中得到了应用(见表 1)。如基于 RFM 模型, Wei 等(2013) [10]结合 SOM 与 K-means 聚类研究了美发客户的类别, 而 Zhou 等(2020) [11]对 K-means 算法进行了改进, 并将其应用于中国移动电信网络客户的分类问题研究。基于 LRFM 模型, Li 等(2011) [12]运用层次聚类与 K-means 聚类对纺织品购买客户进行分类研究; Wei 等(2012) [13]运用 SOM 聚类实现了牙科诊所客户的分类管理。K-means 聚类算法在商业银行基金 [14]、铁路货运 [7]等行业的客户细分中也得到了应用。

Table 1. Application of customer segmentation based on RFM model and clustering technology

表 1. 基于 RFM 模型和聚类技术的客户细分应用

作者(年份)	期刊	细分模型	新增指标含义	技术	研究对象
Li D C 等(2011)	Expert Systems with Applications	LRFM	L (客户关系长度)	层次聚类、K-means	纺织品购买者
Jo-Ting Wei 等(2012)	Expert Systems with Applications	LRFM	L (客户关系长度)	SOM	牙科诊所客户
Wei J T 等(2013)		RFM	/	SOM、K-means	美发客户
Jian Zhou 等(2020)	Expert Systems with Applications	RFM	/	Sparse K-means	中国电信网络客户
张斌等(2017)	交通运输系统工程与信息	KFAV	K (近期发货行为能力)、 A (货运收入贡献率)、 V (发货倾向预测值)	K-means	铁路货运客户
赵铭等(2013)	管理评论	RF + 三个细分变量群	购买方式、基金类型、 分红方式变量群	K-means	商业银行基金客户
郭崇慧等(2015)	管理工程学报	YKFM	客户保留时长(Y)、 汽车行驶里程数(K)	SOM	4S 店客户

RFM 模型特征被有效地用于理解和分析客户行为特征,但它忽略了提供客户消费意愿的客户的消费持续时间,而在一些行业中,客户的消费持续时长对企业评估客户价值、开展营销活动所要参考的关键信息。此外,在经典的 RFM 中,定义的近度变量 R 只考虑客户的最近一次交易,但在很多情况下,由于客户的行为呈现出时间变化,近度可能不能正确地反映客户的重复消费或访问倾向。除此之外,尽管在不同的应用领域中存在着一些关于如何细分客户以改善客户关系和营销策略的研究,但在健身休闲行业中,如民营体育场馆、KTV 等,关于客户细分的研究却很少。

3. 基于客户价值的健身休闲行业客户分类模型设计

3.1. 改进 RFM 模型评价指标

经典的 RFM 模型中的 R (近度)特征仅是基于客户的最后一次交易计算最后一次交易时间距离观察时间的长度,在健身休闲行业中由于具有会员制度,客户可能会为了消耗账户内的少量余额而进行最后一次消费,若其计划消耗完余额后没有再次消费计划,近因 R 无法反映其重访倾向。因此,为了能够更准确地揭示客户重访行为,本文对 RFM 模型中的近因变量 R 进行修改,考虑客户的最后 N 次消费,将 R 修改为最近 N 次消费与观察期截止日期的平均时间间隔。本文将新提出的 RFMD 模型:客户消费新鲜度(R)、客户消费频度(F)、客户消费金额(M)和客户消费持续时长(D)的定义如下:

Recency (R): 客户消费近度,本文将 R 指标修改为客户的 n 次最近消费与观察期的最后日期之间的平均间隔天数。因此,在 RFMD 模型中,近度 R 计算公式如下:

$$Recency(n) = \frac{1}{n} \sum_{i=1}^n date_diff(t_{enddate}, t_{m-i+1}) \quad (1)$$

其中, $date_diff(t_{enddate}, t_{m-i+1})$ 表示客户距离观察期结束日期 $t_{enddate}$ 的最近消费日期与 $t_{enddate}$ 之间的间隔天数, t_m 为最后一次消费时间, m 为总消费次数, n 为考虑的距离观察期结束日期的最近消费次数。值得注意的是,在式子(1)中,当考虑最近消费次数为 1 次,即 $n=1$ 时, R 即为原 RFM 模型中的最近一次消费与观察期结束日期的时间间隔,因此,本文改进的 R 指标也包含了原模型的 R 指标。平均时间间隔越短, R 值越小,表明客户的消费就越近,因此该客再次重访或消费的可能性就越高。

Frequency (F): 客户消费频度,指客户在观察期内到健身休闲场所消费的总次数,如一季度 10 次,一个月 5 次或一周 3 次等。 F 指标值部分反映了客户的健身习惯与需求,客户的重复消费行为是客户忠诚的基本行为表现[13]。频度越高, F 值越大,客户忠诚度越高。

Monetary (M): 客户消费总金额,指在观察期内客户的总消费金额,反映了客户对企业收入的贡献。一般地,客户的消费金额越高, M 值越大,客户对企业的价值贡献越大,客户越重要。

Duration (D): 指客户在观察期内每次消费平均持续时间(以分钟为单位),该指标与观察期内频度 F 和每次消费持续时间有关。对于按时间向客户收取费用的服务,客户消费平均持续时间越长, D 值越大,客户的贡献度越大,客户对该服务的需求和黏性越大。 D 计算公式如下:

$$Duration(D) = 1/F \sum_{i=1}^F minu_diff(t_{i_end}, t_{i_start}) \quad (2)$$

其中, $minu_diff(t_{i_end}, t_{i_start})$ 表示客户第 i 次消费的持续时间(以分钟为单位), t_{i_end} 和 t_{i_start} 分别表示客户在观察期内第 i 次消费开始时间和结束时间, F 为频度。

基于 RFM 模型改进的 RFMD 模型,加入了指标 D ,形成了新的客户价值分类指标体系,本研究使用实际客户数据作为 RFMD 模型指标值,并假设所有模型指标具有同等权重。各指标及含义如表 2 所示:

Table 2. RFMD model indicators explanation**表 2.** RFMD 模型指标及含义

指标	含义
R (Recency)	客户消费近度, 客户的 n 次最近消费与观察期的最后日期之间的平均间隔天数
F (Frequency)	客户消费频度, 客户在观察期内消费的总次数
M (Monetary)	客户消费总金额, 客户在观察期内的总消费金额
D (Duration)	客户消费平均持续时间, 客户在观察期内每次消费平均持续时间(以分钟为单位)

3.2. 基于改进 RFM 模型的客户分类两阶段聚类算法设计

在使用 RFM 模型时, 聚类分析方法已被广泛用于客户细分。通过聚类技术, 根据客户的历史消费记录将客户划分成不同群体, 了解客户特征与客户消费之间的联系, 从而有效地将资源分配给不同类别客户, 做出营销决策。

两步聚类算法是 SPSS 中使用的一种聚类方法, 是 BIRCH 层次聚类的改进。两步聚类第一阶段为预聚类, 依据贝叶斯信息准则(BIC, Bayesian Information Criterion), 并利用对数似然距离作为相似度标准对原始数据进行聚类, 得到最佳聚类数量; 第二阶段为聚类, 利用 Schwarz 准则下的凝聚层次聚类对预聚类进行合并[15], 直至达到最佳的聚类数目。这种方法的主要优点是具有处理分类和连续变量, 自动选择集群的数量, 以及分析大型数据集的能力。然而第二阶段中的凝聚法在处理数据量较大时速度较慢, 时间复杂度较高[16]。

K-means 聚类由于易于实现而非常流行, 其基本思想是给定样本集 $D = \{x_1, x_2, \dots, x_m\}$ 和聚类簇数 k , 随机选取 k 个样本作为初始均值向量, 然后计算其余样本到 k 个均值向量的距离, 将其分配给最近的均值向量对应的簇, 并计算更新后簇的均值向量, 通过不断迭代直到当前均值向量不再更改为止。即 K-means 算法针对聚类所得簇划分 $C = \{C_1, C_2, \dots, C_k\}$ 最小化平方误差

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - u_i\|^2 \quad (3)$$

其中 $u_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$ 是簇 C_i 的均值向量。E 值越小则簇内样本相似度越高。K-means 算法采用了贪心策略, 通过迭代优化来近似求解式(5)。

K-means 聚类算法原理简单易懂, 是解决聚类问题的一种经典, 简单, 快速的算法[9], 但也存在需要预设聚类数目、容易陷入局部最优等缺点。

针对单一聚类算法的不足, 以往文献提出采用两阶段聚类方法进行客户细分研究[16]。因此, 本文采用 K-means 技术, 从两步聚类的预聚类结果中获得聚类簇数。第一阶段采用两步聚类的预聚类, 利用 BIC 准则对数据集进行聚类, 确定聚类簇数(k); 第二阶段采用 K-means 方法, 以第一阶段获得的最佳聚类数作为 K-means 的初始聚类中心 k 执行 K-means 聚类算法以实现体育场馆客户的划分。基于两步聚类和 K-means 的两阶段聚类算法主要步骤如下:

第一阶段: 两步聚类第一步预聚类, 由 BIC 准则计算获取最佳聚类簇数 k ;

第二阶段: K-means 聚类, 1) 从数据集中随机选择 k 个数据作为初始质心; 2) 基于欧式距离计算数据集中各个样本到 k 个初始质心之间的距离, 并将其分配至距离最近的簇中; 3) 重新计算新的簇的质心; 4) 判断新的质心是否发生变化, 若质心不变, 则结束聚类, 获得 k 个类, 否则重复 2) 和 3)。

该两阶段聚类方法在处理大数据集时既加快了计算速度, 同时弥补了 K-means 算法无法自动确定聚类数的不足。

4. 健身休闲行业客户分类模型验证与结果分析

4.1. 数据准备

这项研究调查的 L 体育场馆是一家全国著名的连锁体育场馆品牌, 本文所使用的原始数据集是从该公司的系统中提取的上海场馆的相关数据, 它包含了 2020 年 9 月 1 日至 2020 年 11 月 30 日期间 52,122 名客户的近 38 万条消费记录。在此期间, 有几个客户只去过一两次场馆运动, 因为改进 RFM 模型 RFMD 的改进指标 R 对于去场馆进行运动消费不到 3 次的客户来说被认为是无用的, 所以本文排除了去过场馆消费不到 3 次的客户, 最后得到 50,672 名客户有效数据。

由于数据是从企业交易系统中导出的原始数据, 可能某些交易记录数据存在缺失、异常等情况, 脏乱数据的直接使用会导致模型效果出现偏差, 因此需要对原始数据集进行预处理, 包括缺失值和异常值的检验和处理、数据类型的转换、数据归一化等操作。

对预处理后的数据进行初步探索性分析, 以了解和掌握数据的概况, 分析变量之间的关系提供参考, 也有利于后续进行客户分类结果分析。基于客户消费的开始时间对场馆一天内某小时的消费情况进行考察, 分时段统计落在各时段内的消费记录数, 按星期日期统计一周内每天的消费情况, 结果如图 1 所示。图 1 显示, 客户去场馆运动的时间集中于晚上 20~23 点和夜间 0~5 点。客户的高消费日分布于周六和周日, 周一到周五每天的消费记录数差异不大。企业可以根据客户的消费时间分布规律开展相关活动, 如娱乐篮球赛、小组 pk 赛等, 以吸引到更多客户参与, 使活动效果达到最大化, 同时提升客户的体验和兴趣。

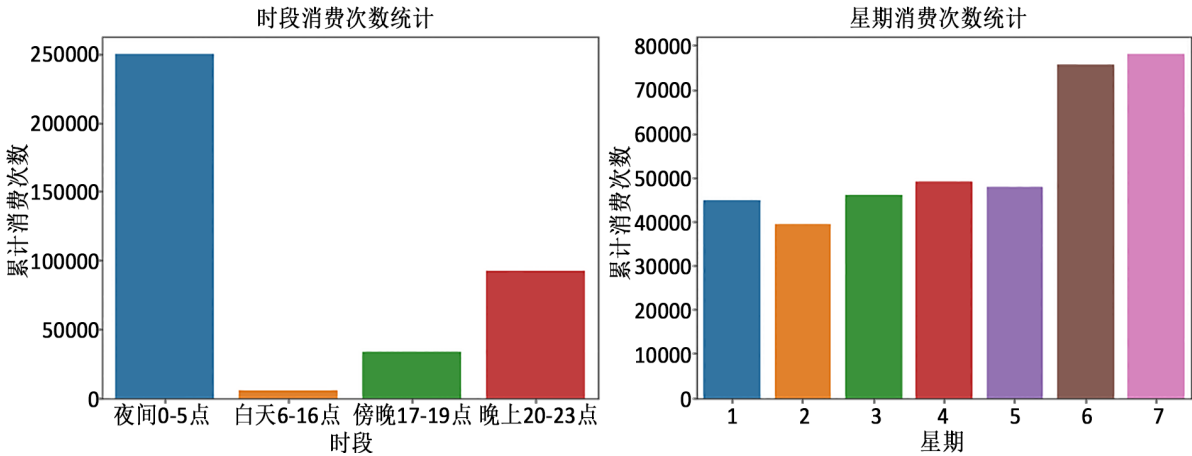


Figure 1. Consumption frequency binning statistics distribution chart by time period (left) and week date (right)

图 1. 消费频次按时段(左)和星期日期(右)分箱统计分布图

该企业客户所持有卡的类型有散客卡、会员卡、白领卡、散客包场卡、A 类卡和 B 类卡 6 种, 分别统计持有不同类型卡客户的消费频次和累计交易额、累计持续时长数据, 结果如图 2 所示。图 2(左)显示, 在统计周期内, 会员卡客户的消费次数远大于其他类型卡客户, 散客卡的消费次数次之, A 类卡和 B 类卡客户很少到场馆进行消费。图 2(右)中, 与消费次数相一致, 会员卡客户的累计交易额和持续时长均居第一, 白领卡客户具有消费次数少、持续时长短, 但交易额高的特点; 散客卡的客户虽然运动累计持续时长比较长, 但交易金额却比较少; A 类卡和 B 类卡客户的累计交易额和持续时长均很少。由此可见, 不同卡类型客户的消费频次与交易额和持续时长无明显必然联系, 某类型卡的客户消费次数很多, 但不代表其能产生较大的交易额。

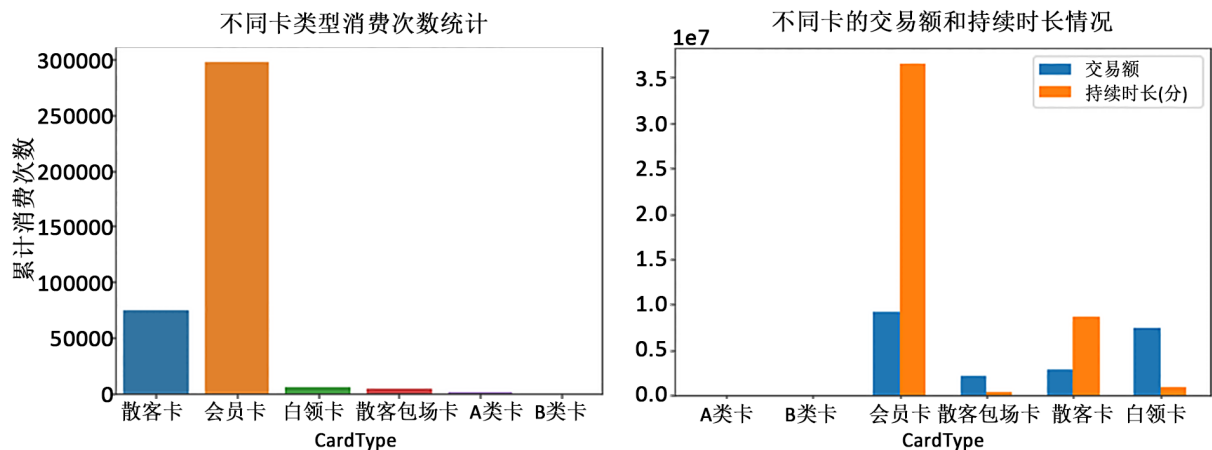


Figure 2. Distribution: consumption frequency (left) and transaction amount and duration (right) of different card types

图 2. 不同卡类型消费频次(左)及交易额和持续时长(右)分布统计图

4.2. 健身休闲场所客户分类模型验证与结果分析

4.2.1. 改进 RFM 模型指标数据获取

作为客户分类的第一步, 为每个客户计算 RFMD 模型的特征。根据 3.1 节中改进后的 RFMD 模型各指标的含义和计算方式, 从原始数据集中提取各指标数值。本文以样本数据获取的截止日期 2020 年 11 月 30 日作为考察时间距离的结束日期, 各指标具体计算方式为:

R 为最近 3 次消费时间距 2020 年 11 月 30 日的平均间隔天数, 单位: 天;

F 为客户在 2020 年 9 月~11 月时间段内的累计消费次数, 单位: 次;

M 为客户在 2020 年 9 月~11 月时间段内的累计消费金额(交易额), 单位: 元;

D 为客户平均每次消费持续的时长, 与客户的累计消费持续时长和消费次数 F 有关, $D = \text{客户的累计消费持续时长} / \text{消费次数 } F$, 体现了客户消费的规律性, 单位: 分钟。

提取五个指标后, 需要分析每个指标的数据分布, 数据范围如表 3 所示。

Table 3. The data range of the RFMD indicator

表 3. RFMD 指标数据范围

属性	R	F	M	D
最大值	89	111	273,400	786
最小值	0	3	0	14.67
平均值	30.71	7.46	423.65	123.13

为避免属性之间的尺度差异对聚类分析结果的影响, 在聚类之前, 本文采用“最大最小标准化”方法对 RFMD 模型的五个指标数据进行标准化, 使得变量数据取值映射到[0, 1]之间。由于 R 越大代表客户最近的消费时间距离越远, R 越小越好, 所以将 R 转化成 $1/R$ 进行计算, 部分数据处理的结果如表 4 所示。

Table 4. RFMD model data example

表 4. RFMD 模型指标数据示例

Card_ID	1/R	F	M	D
01	0.0090	0.0556	0.0008	0.0975

Continued				
02	0.0070	0.0000	0.0003	0.1016
03	0.0110	0.0370	0.0006	0.1080
04	0.0109	0.0278	0.0008	0.2275
05	0.0082	0.0093	0.0003	0.0815
06	0.0169	0.0556	0.0016	0.1664
07	0.0400	0.0463	0.0004	0.0709
08	0.0175	0.0093	0.0003	0.1369
09	0.0149	0.0093	0.0007	0.1631
10	0.0068	0.0000	0.0002	0.1504

本研究的数据来源于真实企业的日常交易记录,数据真实可靠,因此本文在计算 RFMD 模型指标数据时直接使用实际客户数据,不对各指标做权重的区分,并假设所有模型变量的权重相等(即同等重要)。

4.2.2. 基于聚类分析的客户分类结果

1) 聚类结果分析

本文利用构建的两步聚类算法对标准化后的 RFMD 模型指标数据进行聚类分析,首先在 SPSS Modeler 中执行“二阶聚类”(两步聚类)的第一步,由于 RFMD 各指标均为连续变量,距离测量选择“欧式(N)”,获得最佳聚类数目 $k = 2$,此时凝聚和分离的轮廓测量(轮廓指数)值接近 1,聚类质量为较好,代表聚类模型效果较好。当 $k = 3$ 时,凝聚和分离的轮廓测量值也大于 0.8。同时利用 The Elbow Method 手肘法得出结果如图 3 所示,拐点在 2 和 3 的位置最明显。因此,结合企业的实际客户价值情况和现实意义,本文最终确定 $k = 3$,即将客户划分为 3 个类别,分别用标签 1、2、3 表示。

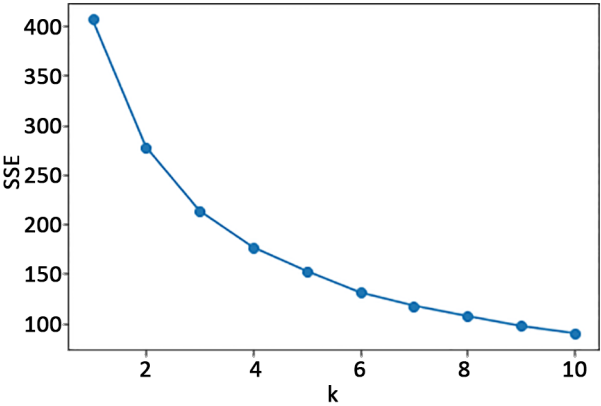


Figure 3. The change trend graph of the number of clusters (k) and the error variance within clusters (SSE)

图 3. 簇数(k)与簇内误差方差(SSE)的变化趋势图

Table 5. ANOVA table
表 5. ANOVA 表

	聚类		误差		F	显著性
	均方	自由度	均方	自由度		
R	12.767	2	0.001	50669	11090.163	0.000

Continued

F	71.542	2	0.001	50669	80497.479	0.000
M	0.119	2	0.000	50669	911.862	0.000
D	0.657	2	0.002	50669	264.637	0.000

表 5 中方差分析显示, 显著性都为 0.000, 都小于显著性水平, 即 $p < \alpha = 0.05$, 说明每一个变量在最后得到的 3 个类别间都存在显著差异。最终聚类类别和客户数量如表 6 所示, 有效客户总数为 50,672, A 类卡、B 类卡、白领卡、会员卡、散客包场卡、散客卡的客户总数分别为 151、57、476、38,213、494、11,281, 3 个类别的客户总数量分别为 38,962、10,307、1403, 分别对应标签 1、2、3。其中在 3 个类别客户中人数最多的都是会员卡客户, 各约占该类别客户总数的 73%、85%和 68%; 其次都是散客卡客户数, 各约占该类别客户总数的 25%、12%和 27%; 其余卡类型客户数在各类别客户总数的合计占比分别为 2%、3%和 5%。

Table 6. Number of customers in each cluster

表 6. 每个聚类中的客户数目

聚类类别	A 类卡客户数	B 类卡客户数	白领卡客户数	会员卡客户数	散客包场卡客户数	散客卡客户数	客户总数	消费总金额占比
1	147	51	224	28,488	417	9635	38,962	39.8%
2	4	6	201	8769	63	1264	10,307	36.0%
3	0	0	51	956	14	382	1403	24.2%
总计	151	57	476	38,213	494	11,281	50,672	100%

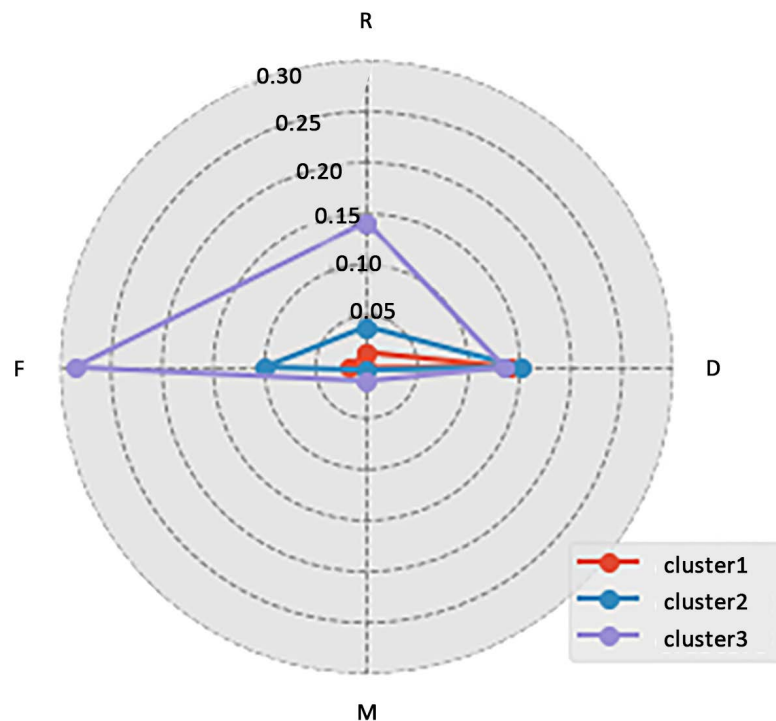


Figure 4. Customer characteristics analysis

图 4. 客户特征分析

2) 客户价值分析

根据聚类中心进行特征分析如图 4 (雷达图)所示, 由于在聚类时将 R 转化为了 1/R, 因此 1/R 越大(即 R 越小)越好, 则客户群 1 的 1/R 属性最小(R 最大), 客户群 3 的 1/R、F、M 属性最大, D 属性较小; 客户群 2 具有最大的 D 属性。根据体育运动企业的具体业务分析, 通过比较集群间各指标的大小, 评价一个集群的特征。例如, 客户群 3 具有最大的 1/R (最小的 R)、F、M 属性和最小的 D 属性, 可以说 R、F、M 是客户群 3 中的优势特征。在客户群 2 中, D 和 F 属性重要程度更高, 在客户群 1 中主要是 D 属性, 其他三个属性都很小。

表 7 展示了每个类别客户的 RFMD 属性的平均值和 RFMD 分数, 以总结各集群客户的 RFMD 特征。在形成最后一列 RFMD 分数时, 本文参考了 Ha 和 Park (1998)的方法。在该方法中, 如果聚类的平均 R、F、M 或 D 值大于所有客户的该指标数据的总平均值, 则该指标以向上箭头符号(↑)表示, 否则以向下箭头符号(↓)表示。

Table 7. Two-stage clustering results based on RFMD
表 7. 基于 RFMD 的两阶段聚类结果

集群	类别大小	平均 R	平均 F	平均 M	平均 D	RFMD 分数
集群 1	38,962	35.75	4.81	219.25	121.24	R↑ F↓ M↓ D↓
集群 2	10,307	14.80	13.82	750.67	130.86	R↓ F↑ M↑ D↑
集群 3	1403	7.62	33.86	3697.30	118.76	R↓ F↑ M↑ D↓
平均值		30.71	7.45	423.65	123.13	

根据上述特征分析的图表显示, 每个客户群都有显著不同的表现特征。为了解释基于 RFM 模型的客户细分结果, Marcus (1998)提出了基于频率(F)和货币(M)变量的客户价值矩阵, 形成了最佳客户(F↑M↑)、挥霍客户(F↓M↑)、不确定客户(F↓M↓)和频繁客户(F↑M↓)四大客户类型。通过将消费持续时长(D)纳入 RFM 模型, 并以前述研究为基础, 本文将确定的客户集群定义为三个级别的客户类别: 低价值客户、一般价值客户、重要价值客户, 如表 8 所示。

Table 8. Name and overview of each customer cluster
表 8. 各客户集群名称和概述

集群	名称	RFMD 分数	客户数占比	消费金额占比
集群 1	低价值客户	R↑ F↓ M↓ D↓	76.9%	39.8%
集群 2	一般价值客户	R↓ F↑ M↑ D↑	20.3%	36.0%
集群 3	重要价值客户	R↓ F↑ M↑ D↓	2.8%	24.2%

根据 RFMD 分数, 集群 2 和 3 的 R 值均低于平均值, 表明两个集群中的客户都与企业有着密切的关系, 这些集群中的客户也可以被视为亲近客户, 并且, 两个集群客户的 F 和 M 值都高于平均值, 运动较为频繁, 对企业收入的贡献度较大。然而, 这两个群体的平均消费持续时长(D)不同, 集群 2 客户的 D 值高于平均值, 集群 3 客户的 D 值小于平均值。此外, 尽管集群 3 是所有集群中最小的, 客户数仅为 2.8%, 但该组中的客户在所选时间跨度内的总消费额(M)占比最大, 即不到 3%的客户数却贡献了企业约 25%的收入。因此, 集群 3 可以被定义为“重要价值客户”, 集群 2 以 20.3%的客户数贡献了企业 36%的收入, 定义为“一般价值客户”。在集群 1 中, 消费近度 R 大于平均值, 并且 F、M、D 值均小于平均值, 说明这些客户最近没有去且平时也不怎么经常去该企业的场所运动消费, 而且该集群客户的每次运动时间

也比较短, 单个个体对企业的贡献度不大, 约 77% 的客户数仅给企业带来 40% 的收入, 因此, 将这个集群客户划分为“低价值客户”。

对于具有不同价值的客户类型, 在提供同等资源的条件下他们对企业的收入贡献是不同的。重要价值客户数量虽少, 但他们对企业的收益起着不可或缺的重要作用, 值得企业在投入资源时给予更多关注和资源倾斜, 并注重积极维护好与这些客户的关系。一般价值客户数量不大, 对企业的收入贡献有限, 但通过有效的经营方法和营销策略, 可以进一步挖掘他们的价值将其转为重要价值且更忠诚的客户。低价值客户数量占比很大但对企业的收入贡献却不足, 这些客户跟企业关系没那么紧密, 运动消费也比较零散, 没有特定的规律可循, 若企业对其投入的资源过多, 客户却给不到应有的利益回报, 那只会增加企业自身的运营成本。低价值客户较高的近度和如此低的频率和运动时间也许是这些客户可能离开公司的信号, 这表明该集群中的客户比较容易成为流失客户。但该集群客户占了该企业客户数的绝大部分, 到底是哪些具体的客户会流失, 需要进行更准确的识别。

由此可知, 客户的合理化分类并在客户管理中实施差异化策略, 是企业维系好客户关系, 在同类市场中增强自身竞争力的关键。并且, 对具有不同价值的客户, 尤其是可能具有流失信号的客户进行流失情况的探索, 并将流失可能性大的客户转化为忠诚的高贡献度客户, 对于一家企业提高盈利能力也是至关重要的。

5. 总结与展望

本文围绕健身休闲行业客户分类问题进行研究, 通过文献分析方法来梳理国内外研究现状并从中发现考虑客户的消费时间问题的客户价值分析相关研究较少, 尤其是在健身休闲行业客户分类的研究相对空白。针对现研究的不足, 本文提出考虑客户消费持续时间价值的客户价值识别模型 RFMD, 并将两步聚类和 K-means 聚类算法相结合, 对健身休闲行业客户进行分类, 帮助企业识别价值客户, 为企业进行差异化精准营销创造前提条件。研究结果发现重要价值客户数量虽很少但贡献了企业较多的利润, 而低价值客户数量占比非常高但消费额占比却未能与客户数量级别相匹配, 能为企业创造的利润少之又少。为了实现资源的合理配置, 企业应该优先考虑重要价值客户的管理与运营。

结合本文研究, 对健身休闲行业相关企业的客户管理工作提出了以下建议:

- 1) 激发和增强客户的自主消费意愿, 使客户延长消费时间。客户在馆内消费时间越长, 其产生的消费额越大, 对于那些几乎每次消费时间不超过半小时或者在半小时左右的客户, 企业可采取调研、访谈等形式获取和分析其消费时间较短的原因, 并根据原因采取针对性改进措施。
- 2) 通过分析客户特征, 探索群体特点, 对客户进行更准确的细分。对客户细分群体采取个性化、创新式的运营及管理, 能够让客户产生被重视的感觉, 从而提高客户的满意度, 激发客户的购物潜能, 进而提高企业的收益。

由于数据收集和研究者能力的局限性, 研究中仍有一些不足之处, 如由于数据收集上的不足, 未能获取客户的基础信息, 如客户性别、年龄、职业等, 在客户分类后未能结合客户的统计学信息进行进一步分析客户群体的特征, 未来可以考虑在可以获取完整客户数据的情况下在客户分类后从多维度对不同价值客户进行特征分析。

参考文献

- [1] Smith, W.R. (1956) Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of Marketing*, 21, 3-8. <https://doi.org/10.1177/002224295602100102>
- [2] Huerta-Muoz, D.L., Rios-Mercado, R.Z. and Ruiz, R. (2017) An Iterated Greedy Heuristic for a Market Segmentation Problem with Multiple Attributes. *European Journal of Operational Research*, 261, 75-87.

- <https://doi.org/10.1016/j.ejor.2017.02.013>
- [3] Liu, J., Liao, X., Huang, W. and Liao, X. (2019) Market Segmentation: A Multiple Criteria Approach Combining Preference Analysis and Segmentation Decision. *Omega*, **83**, 1-13. <https://doi.org/10.1016/j.omega.2018.01.008>
 - [4] 刘英姿, 吴昊. 客户细分方法研究综述[J]. 管理工程学报, 2006, 20(1): 53-57.
 - [5] Hughes, A.M. (1996) Boosting Response with RFM. *Marketing Tools*, **5**, 4-10.
 - [6] 郭崇慧, 赵作为. 基于客户行为的 4S 店客户细分及其变化挖掘[J]. 管理工程学报, 2015, 29(4): 18-26.
 - [7] 张斌, 彭其渊. 基于 KFAV 的中国铁路货运客户细分方法研究[J]. 交通运输系统工程与信息, 2017, 17(3): 235-242.
 - [8] 任春华, 孙林夫, 吴奇石. 基于 LRFAT 模型和改进 K-means 的汽车忠诚客户细分方法[J]. 计算机集成制造系统, 2019, 25(12): 3267-3278.
 - [9] Chen, X., Fang, Y., Yang, M., et al. (2018) PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data. *IEEE Transactions on Knowledge & Data Engineering*, **30**, 559-572. <https://doi.org/10.1109/TKDE.2017.2763620>
 - [10] Wei, J.T., Lee, M.C., Chen, H.K., et al. (2013) Customer Relationship Management in the Hairdressing Industry: An Application of Data Mining Techniques. *Expert Systems with Applications*, **40**, 7513-7518. <https://doi.org/10.1016/j.eswa.2013.07.053>
 - [11] Zhou, J., Zhai, L. and Pantelous, A.A. (2020) Market Segmentation Using High-Dimensional Sparse Consumers Data. *Expert Systems with Applications*, **145**, Article ID: 113136. <https://doi.org/10.1016/j.eswa.2019.113136>
 - [12] Li, D.-C., Dai, W.-L. and Tseng, W.-T. (2011) A Two-Stage Clustering Method to Analyze Customer Characteristics to Build Discriminative Customer Management: A Case of Textile Manufacturing Business. *Expert Systems with Applications*, **38**, 7186-7191. <https://doi.org/10.1016/j.eswa.2010.12.041>
 - [13] Wei, J.-T., Lin, S.-Y., Weng, C.-C. and Wu, H.-H. (2012) A Case Study of Applying LRFM Model in Market Segmentation of a Children's Dental Clinic. *Expert Systems with Applications*, **39**, 5529-5533. <https://doi.org/10.1016/j.eswa.2011.11.066>
 - [14] 赵铭, 李雪, 李秀婷, 吴迪. 基于聚类分析的商业银行基金客户的分类研究[J]. 管理评论, 2013, 25(7): 38-44.
 - [15] Ballestar, M.T., et al. (2018) Customer Segmentation in e-Commerce: Applications to the Cashback Business Model. *Journal of Business Research*, **88**, 407-414. <https://doi.org/10.1016/j.jbusres.2017.11.047>
 - [16] 孙冰, 沈瑞. 基于在线评论的产品需求偏好判别与客户细分——以智能手机为例[J/OL]. 中国管理科学, 1-10. <https://doi.org/10.16381/j.cnki.issn1003-207x.2020.0164>, 2020-11-28.