

doi:10.16104/j.issn.1673-1891.2022.01.010

基于改进的 RFM 模型客户价值分层研究

龙冰婷

(广州华商学院数据科学学院, 广东 广州 511300)

摘要:对传统零售行业客户进行价值分层,进一步对不同价值客户采取相应的营销策略,既能降低营销成本,又能提升营销效果,最终将显著提升企业利润。基于经典 RFM 模型进行改进,修改了指标 F 的计算方式,然后对模型指标以熵权法确定其权重,最后进行 K-means 聚类分析,将客户划分为不同价值层级;以某大型超市相关数据进行实证研究,对超市观察期内的客户进行价值分层后,进一步分析不同类型客户的创利情况,验证了聚类结果的合理性,客户价值层级越高,实际创利情况越佳;最后对不同类型的客户给出相应的营销策略。

关键词:RFM 模型;K-means 聚类;客户价值;熵权法;创利分析

中图分类号:F274 **文献标志码:**A **文章编号:**1673-1891(2022)01-0052-06

Research on Customer Value Stratification Based on Improved RFM Model

LONG Bingting

(School of Data Science, Guangzhou Huashang College, Guangzhou, Guangdong 511300, China)

Abstract: This paper carries out customer value stratification in the traditional retail industry and adopts corresponding marketing strategies for customers of different values, which not only reduces the marketing cost, but also enhances the marketing effects, and ultimately increases the enterprise profit significantly. Based on the classical RFM model, the calculation method of index F is modified, then the weights of model indexes are determined by entropy weight method and finally the customers are divided into different value levels by K-means clustering analysis. According to the relevant data of a large supermarket, with the value stratification of customers in the observation period, this paper analyses the profit-making situation of different types of customers, and verifies the rationality of the clustering results, which proves customers with higher value level create more profits. Finally, corresponding marketing strategies are given to different types of customers.

Keywords: RFM model; K-means clustering; customer values; entropy weight method; profit creation analysis

0 引言

电子商务的发展给传统零售行业带来了巨大的冲击,一方面,网购的便捷性改变了人们的消费习惯,更多的人倾向于网上下单送货上门的便利;另一方面,电商平台“大数据营销”策略实施效果显著,通过分析客户用户画像进行精准营销,智能推荐商品,进一步提高销售利润。传统零售行业对于客户的研究更多停留在会员制层面,采取“一刀切”的营销策略,会员拉新和会员流失的问题在电子商务的冲击下尤为突出,这个问题亟待解决。如果零售商能借鉴电商行业的“大数据营销”方式,充分利用客户的数据信息来了解客户的消费特性,针对性

地策划营销活动,就能在一定程度上减少老会员流失并且发展更多的新会员,进而为企业获取更高的利润。因此,笔者以某大型超市为例进行相关分析,利用超市客户的消费数据研究其价值,进而对客户进行价值层级划分,最后针对不同类型的客户提出差异化的营销策略,使得超市创利更佳。

对客户价值的研究更多的是基于机器学习的聚类分析,以 RFM 模型为基础,对客户消费行为数据进行分析,最终对客户进行类别划分。林盛等^[1]通过 AHP 法得到电信行业 RFM 指标的权重,结合 K-means 聚类对客户进行分类,但是 AHP 赋权存在很大的主观性,且对聚类结果没有进行合理性分析;徐翔斌等^[2]在传统 RFM 模型基础上引入了总利

收稿日期:2021-09-18

基金项目:广东省普通高校人文社科项目(2017KQNCX266)。

作者简介:龙冰婷(1993—),女,江西吉安人,助教,硕士,研究方向:高维数据统计,数据挖掘。

润属性作为模型指标对客户进行细分,但是对于模型指标的权重是主观给出的;罗彪等^[3]采用 ANP 确定指标权重,使用 K-means 聚类进行客户分类,并对每个类别的客户给出相应的营销策略,但是其研究中假设客户的贡献度、重要度、忠诚度和信用度之间相互独立,在实际应用中往往得不到满足;熊兰等^[4]对企业的所有产品分类,创建基于 RFM 的多层级客户价值模型对客户进行分类,但是各类产品的相对权重和 RFM 模型指标的权重都存在很大的主观性;靖立峥等^[5]在 RFM 模型基础上引进了消费行为特征,通过 CH 评估指标确定最佳聚类数目,但是将所有指标的权重视为一样;陈方芳等^[6]将 RFM 模型与 K-means 聚类算法结合,对百货商场客户进行价值研究,对比 R 、 F 、 M 指标在每个类别中的质心值与整体均值,判定不同类别客户的价值,但是没有对客户给百货商场带来的实际利润进行分析,没有验证“价值客户”的实际价值。

针对国内外学者对客户价值研究上存在的 4 大类问题,本文提出改进的 RFM 模型。(1) RFM 模型指标权重采用熵权法^[7]进行客观科学赋权,相比于 AHP 减少主观因素的影响;(2) RFM 模型指标,在原有的 F 基础上考虑时间累积效应,修改为观察期内年均消费次数;(3) 最佳聚类数目,采用手肘法,以误差平方和骤降的点作为最优 K 值;(4) 引入客户利润分析,对于 K-means 聚类得到的不同类型客户进行创利分析,对比每一类客户给企业带来的价值,观察划分的“价值层级”与真实的“创利等级”是否保持一致,从而验证模型对不同客户进行价值分层结果的合理性。综合考虑 RFM 模型指标权重和不同类别客户的利润,来说明本文提出的模型的实际应用价值,最终对不同价值客户给出合理的营销策略。

1 模型理论

1.1 RFM 模型

对于客户价值的研究^[8-10],RFM 模型被广泛应用,作为衡量客户价值和客户创利能力的重要工具和手段,其主要原因在于 RFM 模型简单易懂,与实际应用场景吻合度高^[11-14]。

RFM 模型最初由 Hughes^[15]于 1994 年提出,是衡量客户价值和客户创利能力的重要工具和手段。RFM 模型包含 3 个指标: R (Recency)、 F (Frequency)和 M (Monetary), R 表示客户最后一次消费距观察点的天数, F 表示客户在观察期内消费的次数, M 表示客户在观察期内消费的总金额。最

近消费时间(R)越近、消费的次数(F)越多、消费的金額(M)越高,则相应的客户价值越高。最终通过 3 个指标的加权和得到相应客户的价值得分,RFM 作为定量分析模型,在一定程度上削减了主观因素对客户价值的影响。客户价值得分(score)如式(1)所示:

$$\text{score} = w_R \times R + w_F \times F + w_M \times M \quad (1)$$

式中: w_R 、 w_F 、 w_M 分别表示 R 、 F 、 M 这 3 个指标的权重。

1.2 K-means 聚类

K-means 作为聚类中的经典算法,思想简单、计算速度快、计算成本低,从而被广泛应用于实际中。K-Means 算法基于给定样本之间的距离大小,来衡量样本间的相似度,样本间距离越小,相似度越高,划分为同一个簇的可能性越大;反之,则越可能被划分到不同的簇中,使得簇内的点尽量紧密地连在一起,而让簇间的距离尽量地大,最终将样本集划分到 K 个簇中。

1.3 熵权法

在信息论中,“信息熵”是对样本不确定性的一种度量。不确定性越大,熵值越大,则包含的信息量越多;不确定性越小,熵值越小,则包含的信息量就减少。基于熵的特性,可以通过计算熵值来判断一个事件的随机性及无序程度,也可以用熵值来判断某个指标的离散程度,指标的离散程度越大,该指标对综合评价的影响(权重)越大。比如样本在某指标下取值都相等,则该指标对总体评价的影响为 0,权值为 0。

熵权法基于熵的特性,通过计算不同指标的“信息熵”,来判断该指标的重要程度。它是一种客观赋权法,结果仅由数据本身的离散性所决定,通过数据体现出相应指标的重要程度进而赋予不同的权重,更具有科学性。熵权法计算步骤如下。

1) 归一化处理。

将 RFM 模型的 3 个指标区分为正向指标和负向指标, F 、 M 两个指标的值越大越好,为正向指标;而指标 R 值越小越好,为负向指标。由于正向和负向指标的差异性,需要对其分别采用式(2)和(3)进行归一化处理,最终解决各项异质指标值的同质化问题。

正向指标:

$$z_{ij} = \frac{x_{ij} - \min(x_{1j}, x_{2j}, \dots, x_{nj})}{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - \min(x_{1j}, x_{2j}, \dots, x_{nj})} \quad (2)$$

负向指标:

$$z_{ij} = \frac{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - x_{ij}}{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - \min(x_{1j}, x_{2j}, \dots, x_{nj})} \quad (3)$$

式中: x_{ij} 表示第 i 个样本的第 j 个指标; $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$ 。

2) 计算指标的信息熵。

第 i 个样本在第 j 项指标下的占比:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}} \quad (4)$$

第 j 项指标的信息熵:

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}) \quad (5)$$

式中: $k = \frac{1}{\ln(n)}$ 。

3) 计算指标的权重。

第 j 项指标的权重:

$$w_j = \frac{1 - e_j}{m - \sum_{j=1}^m e_j} \quad (6)$$

2 模型建立与分析

2.1 数据处理

2.1.1 改进的 RFM 模型指标选取

传统 RFM 模型中, 采用的 F (Frequency) 是所有时间内客户下单的次数, 存在时间累积效应的误差。对于长时间均等消费和偶尔大量消费的客户没有区分度, 然而在只考虑该指标的情况下, 前者客户黏性更高, 价值更高, 更值得关注。所以为了在一定程度上消除时间累积效应的误差, 对传统 RFM 模型在指标选取上进行改进, 采用下单总次数/客户产生交易的年限, 即年均下单次数, 作为新的 F 。改进后的 RFM 指标与传统的 RFM 指标含义比较如表 1 所示。

表 1 传统 RFM 模型与改进的 RFM 模型指标含义对比

模型	R	F	M
传统 RFM 模型	客户最后一次消费距观察点的天数	客户在观察期内消费的次数	客户在观察期内消费的总金额
改进的 RFM 模型	客户最后一次消费距观察点的天数	客户下单总次数/客户产生交易的年限	客户在观察期内消费的总金额

2.1.2 RFM 指标计算

基于某大型超市 2015—2018 年零售交易数据, 进行数据清洗后得到有效客户共 793 名。首先统计观察期内每个客户的 R 、 F 、 M 值, 得到 3 个指标值对应的描述性统计量, 结果如表 2 所示。

表 2 RFM 指标原始值的描述性统计量

统计量	R	F	M
样本量	793	793	793
均值	148.77	1.82	2 896.85
标准差	186.11	0.67	2 628.67
最小值	2.00	0.50	4.83
下四分位数	32.00	1.25	1 146.05
中位数	77.00	1.75	2 256.39
上四分位数	185.00	2.25	3 785.28
最大值	1 167.00	5.00	25 043.05

2.1.3 归一化处理

为了消除 RFM 模型各指标的计量单位对聚类分析结果产生不合理的影响, 不直接采用 R 、 F 、 M 指标的原始数据进行聚类, 需要对数据进行归一化处理。对于 R 、 F 、 M 这 3 个指标, F 、 M 属于正向指标, R 属于负向指标, 分别采用式 (2) 和式 (3) 进行归一化处理, 得到的部分数据如表 3 所示。

表 3 归一化处理后 RFM 模型的部分数据

客户 ID	R	F	M
AA-10315	0.842 060	0.166 667	0.222 010
AA-10375	0.983 691	0.388 889	0.041 998
AA-10480	0.777 682	0.111 111	0.071 318
AA-10645	0.952 790	0.222 222	0.202 974
AB-10015	0.643 777	0.111 111	0.035 199

2.2 熵权法确定模型各指标权重

采用熵权法对归一化后的数据计算 R 、 F 、 M 这 3 个指标的权重 w_R 、 w_F 、 w_M , 最终得到的结果如表 4 所示。

表 4 RFM 模型指标权重

	w_R	w_F	w_M
取值	0.042 892	0.270 020	0.687 088

结果显示, M 的权重最高, R 的权重最低, 客户最近一次消费时间距离观察点的天数差异并不显著, 而观察期内的消费金额相差较大。

2.3 K-means 聚类

2.3.1 RFM 模型指标加权处理

传统的 K-means 聚类分析, 默认 R 、 F 、 M 3 个指标的权重相同, 然而在实际应用中 3 个指标的权重并非如此。此次研究根据熵权法得到的 3 个指标的权重结果, 进一步赋予 3 个指标不同的权重进行加权处理, 得到 K-means 聚类分析所需数据分别为 R_w 、 F_w 、 M_w 。

$$R_w = \frac{R \times w_R}{R \times w_R + F \times w_F + M \times w_M} \quad (7)$$

$$F_w = \frac{F \times w_F}{R \times w_R + F \times w_F + M \times w_M} \quad (8)$$

$$M_w = \frac{M \times w_M}{R \times w_R + F \times w_F + M \times w_M} \quad (9)$$

2.3.2 手肘法选择最优 K

K-means 算法第一步需要确定 K 的取值,主观选择 K 值会对聚类效果产生一定的影响,未必是最优的 K 值,此次研究通过手肘法来确定最优的 K 值。

手肘法是通过计算每个样本点与其所在簇内质心的误差平方和 SSE,来评估样本聚类效果的好坏。随着 K 值的增加,每个簇内的样本点逐渐减少,簇内的聚合程度也相应提高,SSE 随之递减。但是在 K 增大的过程中,SSE 的下降幅度逐渐减弱,某一时刻 SSE 的下降幅度骤减,然后逐渐趋于平缓。整个过程中,K 值与 SSE 形成的曲线类似于手肘,最优的 K 值即为手肘位置。由图 1 结果显示,最优 K 值为 5。

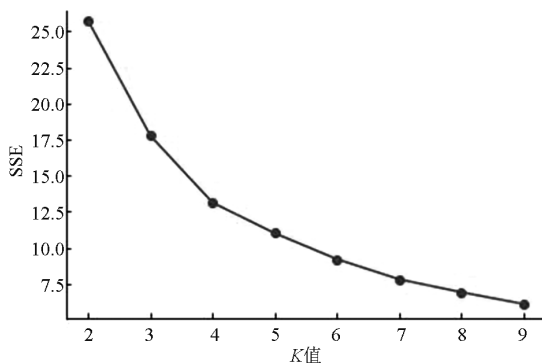


图 1 SSE 与聚类个数 K 之间的关系

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (10)$$

式中: C_i 是第 i 个簇; p 是 C_i 中的样本点; m_i 是 C_i 的质心。

2.3.3 聚类结果分析

将 793 名客户划分为 5 类,每一类的客户数及 R 、 F 、 M 这 3 个指标的质心结果如表 5 所示,3 个指标聚类中心分布情况如图 2 所示。

表 5 RFM 模型聚类结果

类别	R_w	F_w	M_w	客户数	占比/%
1	0.91	0.16	0.06	258	32.5
2	0.94	0.34	0.12	301	38.0
3	0.53	0.23	0.07	102	12.9
4	0.93	0.57	0.16	86	10.8
5	0.91	0.35	0.40	46	5.8

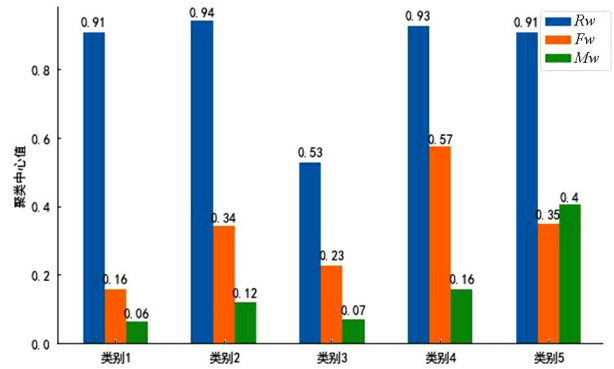


图 2 RFM 模型各指标聚类中心分布

表 5 和图 2 显示,聚类结果的 5 个类别中,类别 5 整体 R_w 、 F_w 、 M_w 值均较高,最近消费时间近、消费频次和消费金额都很高,绝对的优质客户,视为“重要价值客户”,客户占比最低,只有 5.8%;类别 4 整体 R_w 、 F_w 值较高, M_w 值次之,最近消费时间较近、消费频次高,消费金额中等,说明这是个黏性度高的忠诚客户群体且很有潜力,必须重点发展,视为“重要发展客户”,客户占比较低,只有 10.8%;类别 2 整体 R_w 值较高, F_w 、 M_w 值次之,最近消费时间较近,消费金额、消费频次中等,视为“重要保持客户”,这类客户占比最高,达到 38.0%;类别 3 整体 R_w 、 F_w 、 M_w 值处于中等水平,最近消费时间较远、消费频次、消费金额中等,可能是将要流失或者已经要流失的用户,应当基于挽留措施,视为“重要挽留客户”,客户占比为 12.9%;类别 1 整体 R_w 较高,但是 F_w 、 M_w 值较低,最近消费时间较近、消费频次和消费金额较低,属于偶然性消费客户,且消费意向不高,视为“一般客户”,客户占比较高,达到 32.5%。

3 不同价值客户创利情况分析

根据 RFM 模型聚类结果,进一步分析不同层级的客户群体创利情况,不同类型客户创收情况如表 6 和图 3 所示。

表 6 不同类型客户创利情况

客户类型	盈利客户占比/%	亏损客户占比/%	人均消费金额/元	人均利润/元
一般客户	78.29	21.71	1 609.72	173.82
重要价值客户	86.96	13.04	10 139.86	1 871.07
重要保持客户	82.06	17.94	2 982.46	318.92
重要发展客户	84.88	15.12	3 929.63	455.72
重要挽留客户	74.51	25.49	1 762.64	198.96

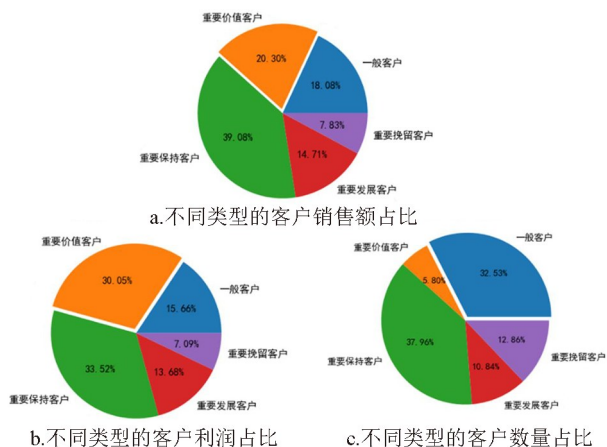


图 3 不同类型客户创利情况分布

“重要价值客户”整体人数占比最低,但是却创造了第 2 的销售额和利润。其盈利客户占比高达 86.96%,人均消费金额和人均利润都是最高的,说明该类客户是价值最高的客户,也是超市创利的中坚力量。

对于“重要发展客户”,盈利客户占比、人均消费金额和人均利润同样较高,但相比于重要价值客户,创利明显降低。整体人数较少,但是销售额和利润却处于中上水平,所以该类客户可以采取一定的营销手段刺激其消费,尽可能将其转换为“重要价值客户”。

“重要保持客户”在人数上具有明显的优势,整体销售额和利润占比最高。虽然人均消费金额和人均利润不如前 2 类客户,但是与重要发展客户相比创利相差不是很显著,该类客户同样是超市创利的主力军。

“重要挽留客户”销售额和利润占比是最低的,其主要原因是客户数占比较低。亏损的客户占比高达 25.49%,人均消费金额和人均利润虽然略高于“一般客户”,但显著低于其他 3 类客户。对于这类客户要时常通过积分换购等优惠方式提醒其消费,防止其流失。

“一般客户”在人数上有一定的优势,但是这类客户消费的偶然性太高,人均利润也不是很高。在营销时可以适当地发放优惠券等,但不要投入过高的成本,不然可能导致投入和产出不成正比。

4 营销策略分析

大数据营销相比于传统营销而言,必然不能对所有客户采取相同的营销策略,针对本文研究得到的结论,主要提出以下营销策略。

策略 1 对于营销成本投入,重要价值客户>重要发展客户>重要保持客户>重要挽留客户>一般客

户,这几类客户的人均利润依次递减,所以对于每类客户的营销成本也依次递减,达到投入与产出成正比的效果。对于价值最大的客户要尽可能地留住,并对有发展潜力的客户进行客户提升活动,使其向上一价值层级转换,对一般客户的投入最少。

策略 2 对于不同价值客户开展不同的营销活动。

1)重要价值客户,这类客户最近消费时间近、消费频次和消费金额都很高,表明其具备一定的经济实力。应将该类客户设置为特别顾客,为其办理特殊会员卡,比普通会员享有更大的折扣。按时将高门槛优惠券发送至用户的会员手机号,或者是在微信公众号中选其为特邀顾客,发送一些价格较高的商品优惠信息,进一步刺激其消费。

2)重要发展客户,属于黏性度高、忠诚但是消费金额中等的客户。表明其经济实力一般,但对该超市的商品较为满意,这类顾客主要是提高其购买金额。对于这类客户,首先可为其办理积分会员卡;其次从历史商品购买信息中,找出其购买频率较低或没有购买的商品,为其提供中等额度的消费券,并给予多倍积分,吸引其购买价格较高的商品,从而提高消费。累计的积分可用于换购商品,进一步提高其忠诚度。

3)重要保持客户,这类客户近期有消费,但是消费次数和金额中等,如果是新客则开展新客促销活动,否则就开展消费次数和金额累计优惠活动,刺激其消费更多;对于这类顾客应增加其消费的频次,提高其忠诚度。可先通过发放调查问卷,找出忠诚度不高的原因,是超市的定位还是商品质量等问题。可根据其经常购买的商品为其推送优惠信息,按时发放一些小金额的优惠券。

4)重要挽留客户,属于即将要流失或已经流失客户,开展老客户召回活动。给客户发送积分换购、老客户特享优惠等优惠方式提醒其消费,防止其彻底流失。

5)一般客户,这类顾客可能属于偶发性消费,不需要对其花费过多的精力。可在超市门口将每日的优惠信息进行展示,可为其办理一般会员卡,享受折扣优惠。

5 结语

对于大型超市而言,采取传统的营销方案进行利润的提升,效果已然不是很显著。在大数据时代,基于超市的客户消费数据进行相关分析,对客户进行价值分层,针对不同价值客户开展不同的营

销活动,将达到事半功倍的效果。对营销效果进行数据跟踪,还能进一步调整相应的营销策略。

本文在传统 RFM 模型的基础上,改进其指标的选取,得到改进的 RFM 模型。基于 RFM 模型理论,通过熵权法对 RFM 模型指标进行科学客观赋权,最后进行 K-means 聚类分析,将超市客户划分为不同价值层级的 5 类:重要价值客户、重要发展客户、重

要保持客户、重要挽留客户和一般客户。然后分别对其进行创利分析,进一步验证聚类分析结果,价值层级越高的客户,其创利情况越佳。最后基于本文分析结果给出营销策略。本文倡导基于“大数据”科学分析的结果,对超市客户进行“差异化”营销,达到投入和产出成正比的效果,使得超市创利步步提升。

参考文献:

- [1] 林盛,肖旭.基于 RFM 的电信客户市场细分方法[J].哈尔滨工业大学学报,2006(5):758-760.
- [2] 徐翔斌,王佳强,涂欢,等.基于改进 RFM 模型的电子商务客户细分[J].计算机应用,2012,32(5):1439-1442.
- [3] 罗彪,闫维维,万亮.基于 ANP 和 K-means 聚类的客户价值分类模型及应用[J].计算机应用,2013,33(10):2954-2959.
- [4] 熊兰,高炳.基于 RFM 多层次客户价值模型的客户细分研究[J].商业经济研究,2017(5):55-57.
- [5] 靖立峰,吴增源.基于改进 K-means 算法的电子商务客户细分研究[J].中国计量大学学报,2020,31(4):482-489.
- [6] 陈方芳,傅仰耿.基于聚类算法识别大型百货商场客户价值的 RFM 模型[J].景德镇学院学报,2021,36(3):68-72.
- [7] 陈东清,叶翀,黄章树.基于熵权法改进 RFM 模型的电商客户价值细分研究[J].西安电子科技大学学报(社会科学版),2020,30(2):39-45.
- [8] 孙琰,马宝龙,李金林.基于 RFM 模型方法的忠诚计划会员顾客价值识别研究[J].数学的实践与认识,2011,41(15):75-79.
- [9] 叶志龙,黄章树.基于新三维客户细分模型的线上会员客户价值研究[J].统计与信息论坛,2016,31(5):96-101.
- [10] 胡少东.客户细分方法探析[J].工业技术经济,2005(7):66-69.
- [11] FATEMEH B, MOHAMMAD J T. Mining customers behavior based on RFM model to improve the customer satisfaction[J]. International Journal of Customer Relationship Marketing and Management, 2011, 2(3): 79-91.
- [12] MEHDI G, SAMIRA M M, SOMAYEH A. Data-mining application for country segmentation based on the RFM model[J]. International Journal of Data Analysis Techniques and Strategies, 2008, 1(2): 126-140.
- [13] 王文贤,金阳,陈道斌.基于 RFM 模型的个人客户忠诚度研究[J].金融论坛,2012,17(3):75-80.
- [14] WU J, SHI L, LIN W P, et al. An empirical study on customer segmentation by purchase behaviors using a RFM model and K-means algorithm[J]. Mathematical Problems in Engineering, 2020(1): 1-7.
- [15] HUGHES, MIDDLETON A. Strategic database marketing: the masterplan for starting and managing a profitable, customer-based marketing program[M]. Chicago: Probus Publishing Company, 1994: 77-79.