

■经济学

基于熵权法改进 RFM 模型的 电商客户价值细分研究

陈东清¹, 叶翀², 黄章树²

(1. 福州大学至诚学院, 福建 福州 350002; 2. 福州大学 经济与管理学院, 福建 福州 350108)

摘 要: 对电商客户价值细分并实施差异化营销策略, 有助于提高电商企业利润, 增强企业竞争力, 推动行业高质量发展。提出基于熵权法改进 RFM 模型的电商客户价值细分方法, 该方法利用熵权法确定 RFM 模型三个得分指标的权重, 计算得分并评价电商客户价值, 采用轮廓系数为标准, 选择最优 K 均值聚类模型对电商客户价值进行聚类分析。以拼多多某商家销售数据进行实证研究, 从所构建模型的客户价值细分结果可知, 每类客户群体的区分度更大, 更加符合经典的客户价值分布特征, 模型效果良好。

关键词: 电商客户价值; 客户细分; RFM 模型; 熵权法; K 均值聚类

中图分类号: F713.365.2

文献标识码: A

文章编号: 1008-472X(2020)02-0039-07

收稿日期: 2020-04-06

基金项目: 国家社科基金项目(19FJYB043); 福建省社会科学规划项目(2014B217)。

作者简介: 陈东清(1986-), 男, 福建晋江人, 福州大学至诚学院, 讲师, 研究方向: 物流信息管理、商务智能与数据挖掘;

叶 翀(1976-), 男, 福建福州人, 福州大学经济与管理学院, 副教授, 研究方向: 流通业态论、流通竞争论。

DOI:10.16348/j.cnki.cn61-1336/c.2020.02.006

一、引言

我国电子商务行业蓬勃发展, 诞生了很多新型商业模式, 重构商品的生产、流通和销售链条, 推动实体经济高质量发展。但是对于中小型电商企业及个人卖家而言, 电商行业竞争越发激烈, 如何科学评价客户价值并精准划分客户群体, 进而指导制定营销策略, 是赢得客户获取利润的重要手段。

目前国内外学者对客户价值进行了深入研究, 客户价值细分是客户关系管理的重要内容, 已经受到广泛重视^[1]。学者王朝晖、钟超研究认为通过对电子商务客户价值细分有助于企业集中资源维护发展高价值客户, 减少或者放弃低价值客户的投入, 提升资源配置效率^[2]。Han 等学者从客户生命周期角度提出客户价值细分模型^[3]。学者 Keränen J、Jalkala A 从供应商的角度提出客户价值评价过程, 主要包含价值识别、评估、实现等过程^[4]。陈静怡研究认为通过对客户价值的挖掘与重塑, 有助于企业获得总体供应链的竞争优势^[5]。周毓萍、黄丽莉采用组合赋权-模糊评价法研究商业银行客户价值, 研究成果有助于银行识别高净值客户, 并提升私人银行服务水平^[6]。Ziafat H 采用聚类分析方法从购买行为方面对客户价值进行细分, 研究成果有利于实施精准的营销策略^[7]。叶志龙、黄章树构建新三维客户细分模型评价线上会员客户价值, 测算每个类型客户得分, 该模型可较好地识别不同价值的客户群体^[8]。

RFM 模型在客户价值细分领域得到广泛应用。Dursun A 等结合数据挖掘技术及 RFM 模型分析酒店客户价值, 建议管理者应依据客户 RFM 的变化采取不同的管理策略^[9]。夏名首认为 RFM 模型对判断客户的长期价值具有较大优势, 可动态展示客户轮廓^[10]。包志强等学者基于客户消费行为视角利用 RFM 模型研究百度外卖客户价值, 建议可对铂金价值的客户群体开展发放优惠券的营销活动^[11]。熊兰、高炳提出 RFM 多层次客户价值模型, 并与经典的 RFM 模型对比发现, 所构建模型不仅可以区分

客户价值水平，还可用于研究客户对不同类型产品的价值^[12]。热依木江·克里木等结合自组织映射神经网络划分的 RFM 模型，以库尔勒香梨网络销售数据为实证研究，认为不同时期客户的消费行为存在异质性^[13]。

从以上的研究成果看出，客户价值细分研究成果丰富，但存在以下改进空间：1.对客户价值评价指标的权重确定当前主要是采用主观赋权法或者层次分析法，主观赋权法存在较强主观性的缺点，而层次分析法需要邀请行业专家并构造判断矩阵，工作量较大；2.传统 RFM 模型的客户价值评价结果应用，主要是根据 R、F、M 三个指标得分的均值，将客户划分为 8 类，该方法存在客户分类模糊的问题。

因此，本文创新性构建基于熵权法的 RFM 模型，将熵权法引入到 RFM 模型中，发挥熵权法客观定权的优势，客观评价客户价值，并采用 K 均值聚类模型对客户价值进行细分，以轮廓系数为标准选择合理的聚类数量，最后以拼多多商家数据作为实证研究验证模型有效性。

二、基于熵权法改进 RFM 的电商客户价值细分模型构建

（一）传统 RFM 模型

传统RFM模型由Hughes A M提出^[14]，该模型包括近度R（Recency）、频率F（Frequency）、购买金额M（Monetary）三个维度。R表示客户最近一次购买时间到某个截止时间的跨度，反映了客户活跃程度；F代表统计时间内客户的购买频率，用于评价客户的忠诚度；M代表统计时间内客户的购买金额，评价客户的购买力。通过对R、F、M三个指标赋予特定权重，即可测算出每个客户价值细分指标得分及综合得分，综合得分越高则认为客户价值越大。

（二）熵权法

熵权法是一种基于信息熵的评价理论^[15]，能够客观确定所评价对象的指标权重，常用于复杂系统问题的评价。本文采用熵权法确定 RFM 模型三个指标得分的权重，具体计算过程如下：

记RFM模型指标的得分数据为 $IS = (is_{ij})_{m \times n}$ ，其中 m 代表样本个数， n 代表指标个数，第 i 个样本的第 j 个指标值记为 is_{ij} （ $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ ）。

计算第 j 个指标下第 i 个样本指标值的比重 p_{ij} ，公式如下：

$$p_{ij} = is_{ij} / \sum_{i=1}^m is_{ij} \quad (1)$$

记第 j 个指标的信息熵为 E_j

$$E_j = -k \sum_{i=1}^m p_{ij} \ln p_{ij}, \text{ 其中 } k = 1/\ln m \quad (2)$$

第 j 个指标的权重记为 w_j

$$\text{则 } w_j = (1 - E_j) / \sum_{j=1}^n (1 - E_j) \quad (3)$$

根据上式得到的指标权重，可计算客户价值综合得分，记为 S

$$S = IS * W \quad (4)$$

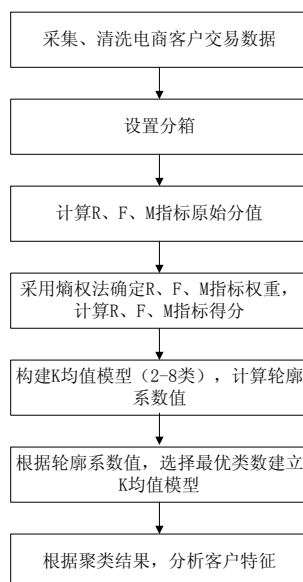
（三）基于熵权法改进 RFM 的电商客户价值细分模型

本文创新性提出基于熵权法改进 RFM 模型的电商客户价值细分方法，模型构建步骤如图 1 所示，该方法包含以下 7 个步骤。

1.采集、清洗电商客户交易数据。从电商交易平台导出相关交易数据，对数据进行脱敏、缺失值、异常值等预处理，确保数据可靠性和准确性。

- 2.设置分箱。一般将 R、F、M 三个指标的交易数据划分为 5 个分箱，每个等级分别赋值 1-5 分。
- 3.计算 R、F、M 指标原始得分。根据分箱结果及各个客户的交易数据，计算每个客户 R、F、M 指标原始得分。
- 4.采用熵权法确定 R、F、M 指标得分的权重，并计算指标得分。以第 3 步骤得到的指标原始得分为基础数据，采用上述介绍的熵权法计算三个指标得分的权重，并分别乘以第 3 步骤得到的原始得分，得到基于熵权法的指标得分。
- 5.构建 K 均值模型，计算轮廓系数值。考虑到 RFM 模型有 3 个指标，每个指标得分存在大于（等于）平均值、小于平均值两种情况，一共有 $2 \times 2 \times 2 = 8$ 种情况，因此最大的聚类数量为 8，最小聚类个数为 2。故构建 7 个 K 均值模型，分别计算每个聚类模型的轮廓系数值。
- 6.选择最大轮廓系数值对应的聚类数建立最优 K 均值聚类模型。比较第 5 步骤得到的轮廓系数值，以最大轮廓系数对应的聚类数作为聚类模型的类别参数，建立 K 均值聚类模型，确定每个样本的归类。
- 7.根据聚类结果分析客户特征。分析不同类别客户的特征，比较不同客户群的价值，提出差异化的营销策略。

图1：模型构建步骤



三、实证研究

（一）数据来源及数据预处理

研究数据选取的是一家从事饼干糕点销售的拼多多商家，时间跨度为 2019 年 8 月 11 日到 2020 年 1 月 19 日。订单数据从拼多多卖家销售管理系统导出，对数据预处理主要包含以下方面。

1.剔除无效样本数据。从导出的交易数据中发现，存在形如“补收差价专用商品”的商品名称，此类商品交易是无效的购买数据；售后状态是“售后处理中”或“退货”状态的样本也属于无效的数据。将这类购买订单数据剔除，最终的交易数据为 32692 条购买记录，共计 29607 个买家。

2.客户识别标志的处理。拼多多销售管理系统对客户手机号码进行了脱敏处理，并且导出后的客户 ID 为空值，难于识别客户标志，因此本文采用物流单地址作为客户的唯一识别标志。

3.购买时间的确定。确认收货时间是衡量购买时间的最佳方式，但由于部分客户未主动收货，而是系统自动确认收货，而拼多多导出的单据中没有记录自动确认收货时间，因此认为客户的订单确认时间即为客户的购买时间。

4.购买金额处理。考虑到拼多多电商平台有优惠券、满减等活动、平台交易费用（信用卡费用、第三方支付费用等），本文的购买金额为扣除优惠活动费用、交易费用后商家获取的实际到账金额。

经过对数据的预处理，确保研究数据准确性，可用于下一阶段的分析与挖掘。

（二）RFM 模型指标值分箱处理

采用分位数方法对 R、F、M 指标值进行分箱处理，详细区间设置如表 1 所示。通过下表的处理，可计算出各个电商客户价值三个指标原始得分。

表 1: RFM 模型指标值分箱处理

分值	R	F	M
1	(116, +∞)	[1, 2)	[4.30, 9.70)
2	(93, 116]	[2, 3)	[9.70, 15.70)
3	(65, 93]	[3, 4)	[15.70, 18.81)
4	(29, 65]	[4, 5)	[18.81, 27.50)
5	[1, 29]	[5, 20]	[27.50, 1459.82]

（三）采用熵权法确定指标权重并计算得分

利用上述介绍的熵权法计算得到 R、F、M 三个指标得分的权重分别为 0.4069、0.1494、0.4437，并分别乘以对应指标原始分值，得到基于熵权法 RFM 的评价结果如表 2 所示，RS、FS、MS 分别代表分箱处理后最近购买得分、购买频率得分、购买金额得分。RS₁、FS₁、MS₁ 分别代表原始得分乘以熵权法权重后各指标得分。S₁ 是电商客户价值综合得分，数值越高代表该客户价值越大。计算公式如下：

$$S_1 = RS_1 + FS_1 + MS_1 \quad (5)$$

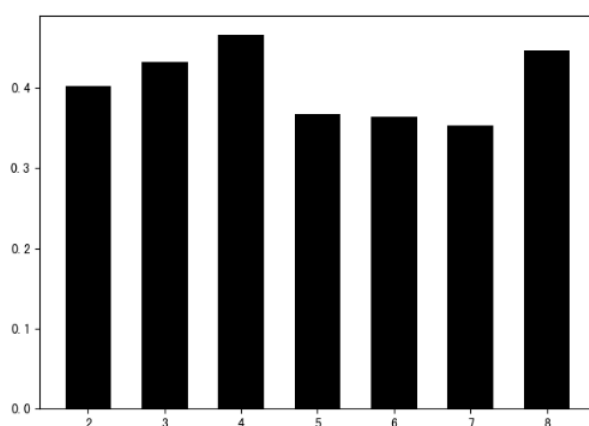
表 2: 基于熵权法 RFM 模型的部分客户价值评价结果

客户编号	R	F	M	RS	FS	MS	RS ₁	FS ₁	MS ₁	S ₁
1	91	1	27.8	3	1	5	1.2207	0.1494	2.2185	3.5886
2	91	1	19.9	3	1	4	1.2207	0.1494	1.7748	3.1449
3	90	1	5.9	3	1	1	1.2207	0.1494	0.4437	1.8138
4	90	1	68.8	3	1	5	1.2207	0.1494	2.2185	3.5886
.....
29603	121	3	79.15	1	3	5	0.4069	0.4482	2.2185	3.0736
29604	108	2	419.9	2	2	5	0.8138	0.2988	2.2185	3.3311
29605	102	2	910.8	2	2	5	0.8138	0.2988	2.2185	3.3311
29606	101	5	799	2	5	5	0.8138	0.7470	2.2185	3.7793
29607	123	10	1459.82	1	5	5	0.4069	0.7470	2.2185	3.3724

（四）电商客户价值细分结果

以熵权法 RFM 模型评价结果（RS₁、FS₁、MS₁）作为聚类分析指标，采用 Python 3.7.4 工具编程实现 K 均值聚类模型运算，聚类数量设置为 2-8 个，以上述指标的均值作为初始化聚类中心，最大迭代次数设置为 100000 次，容忍度设置为 0.00001，计算不同聚类数量模型对应的轮廓系数值如图 2 所示，可直观看出，当聚类数量为 4 时，轮廓系数最大，因此选择最优聚类数量为 4。

图 2: 聚类数量为 2-8 的轮廓系数值



根据 K 均值聚类结果, 可将所研究的拼多多电商客户划分为 4 类, 类中心、样本分布情况如表 3 所示。各个电商客户群价值特征分析及营销策略建议如下。

1. 第 1 类客户 RS_1 中心值排名第三, FS_1 中心值最小, MS_1 中心值最小, 说明这类客户最近无交易, 消费频次最少, 消费金额最小, 定义为低价值客户。针对此类客户, 商家只需一般关注, 如果此类客户在一段时间内各个指标没有明显改善, 可采取放弃策略。

2. 第 2 类客户的 RS_1 中心值最小, FS_1 中心值排名第二, MS_1 中心值排名第二, 说明此类客户最近无交易, 但是消费频次、消费金额较大, 可定义为重要保持客户。需要重点分析此类客户近期无交易原因, 激发客户消费欲望, 促使转化为优质客户。

3. 第 3 类客户的 RS_1 中心值最大, FS_1 中心值排名第三, MS_1 中心值排名第三, 说明此类客户最近有交易, 但是交易频次及交易金额偏低, 可定义为一般发展客户。商家只需对此类客户采取维持现状的营销策略。

4. 第 4 类客户的 RS_1 中心值较大, FS_1 中心值最大, MS_1 中心值排名最大, 说明此类客户最近有交易, 交易频次及交易金额最大, 可定义为重要价值客户。此类客户为商家带来很大利润, 商家需要重点维护, 挖掘客户消费潜力, 提升客户忠诚度。

表 3: 熵权法改进 RFM 模型的聚类结果

类别	RS_1 中心值	FS_1 中心值	MS_1 中心值	样本个数	所占比重(%)	客户定义
1	0.8563	0.1504	0.6587	8045	27.17	低价值客户
2	0.5821	0.1775	1.8034	6799	22.97	重要保持客户
3	1.7472	0.1515	0.9241	8092	27.33	一般发展客户
4	1.6499	0.1845	2.0300	6671	22.53	重要价值客户

（五）结果比较

表 4: 传统 RFM 模型聚类结果

类别	RS_2 中心值	FS_2 中心值	MS_2 中心值	样本个数	所占比重(%)	客户定义
1	0.8952	0.3338	0.3008	11853	40.03	重要保持客户
2	2.4000	0.3197	0.3010	6068	20.49	一般发展客户
3	3.0000	0.3247	0.2980	5521	18.65	重要发展客户
4	2.7843	1.0215	0.4901	242	0.82	重要价值客户
5	1.8000	0.3125	0.2598	5923	20.01	低价值客户

传统 RFM 模型需要对 R、F、M 三个参数设置权重, 计算综合得分再用于客户价值评价, 本文参考徐文瑞关于顾客消费行为与顾客价值预测研究成果^[16], 将其确定为 (0.6、0.3、0.1)。类似上面的分析, RS₂、FS₂、MS₂ 分别代表各指标原始得分乘以参考文献所确定权重得到的分值。采用 K 均值聚类算法进行聚类分析, 发现当聚类数量为 5 时, 轮廓系数最大, 因此选择最优聚类数量为 5。各类客户群体价值特征及定义整理如表 4 所示。

借鉴单因素方差分析思想, 采用组间均方 (MSA) 评价不同类别电商客户综合价值得分 (S) 的组间差异, 计算公式如下:

$$MSA = \frac{\sum_{i=1}^I n_i (\bar{S}_i - \bar{\bar{S}})^2}{I-1} \quad (6)$$

上式中, I 为 K 均值聚类模型的最优聚类个数, n 为样本个数, n_i 是第 i 个类别的样本量,

$\bar{S}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} S_{ij} (i=1,2,\dots,I)$ 是第 i 个类别的样本平均值, $\bar{\bar{S}} = \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^{n_i} S_{ij} (i=1,2,\dots,I)$ 代表所有样本的平均值。

经过计算可知, 熵权法改进 RFM 模型聚类结果中不同客户类别之间的组间均方 MSA₁=5965.4799; 传统 RFM 模型聚类结果中不同类之间的组间均方 MSA₂=5564.6684, 说明熵权法改进 RFM 模型的聚类结果中各个客户群体之间的差异程度更大, 区分度更高, 对电商客户价值的划分效果更好。

另一方面, 从当前的研究成果来看, 大多数客户价值符合二八原则的特征, 即 20% 的客户创造了 80% 的利润, 因此认为 10-20% 左右的客户是高价值客户。从表 3 的结果看出, 可认为 22.53% 是高价值客户, 27.17% 是低价值客户, 接近经典的客户价值分布特征; 而表 4 的结果显示, 高价值客户的比重为 0.82%, 低价值的客户比重为 20.01%, 高价值客户比重明显偏低。原因是, 熵权法确定 M (金额) 的权重为 0.4437, 明显高于参考文献中确定的 0.1 的金额权重。因此, 从经典的客户价值分布特点来看, 所构建模型对电商客户价值细分是有效的。

综合组间均方指标分析, 以及结合经典的客户价值分布特征比较结果, 可认为本文所构建的电商客户价值细分模型具有良好效果。

四、结论

对电商客户价值细分有助于指导商家开展差异化营销策略进而提升利润水平, 本文创新性提出基于熵权法改进的 RFM 模型, 结合 K 均值聚类分析方法, 以拼多多商家的客户价值细分为实证研究, 得到以下结论:

(一) 熵权法能够客观确定 RFM 客户价值评价模型中三个指标的权重, 为 RFM 模型确定指标权重提供新方法。

(二) 相对传统 RFM 模型的电商客户价值评价结果, 熵权法改进 RFM 模型有助于更好评价客户价值, 从基于熵权法改进 RFM 模型的电商客户价值聚类分析结果看出, 各个类别客户之间的差异程度更大, 区分度更高, 客户价值更接近经典的分布特征, 说明本文所提方法对电商客户价值细分是有效的。

不同行业客户价值特征存在较大差异, 本文所构建模型可拓展到其他行业客户价值细分应用, 进一步验证模型的有效性。另一方面, RFM 模型中用购买金额 M (Monetary) 作为评价客户价值指标之一, 但是购买金额与企业最终利润不能简单等同对待, 可考虑将利润指标拓展到传统 RFM 模型中, 进而优化传统 RFM 模型。以上内容可作为后续研究方向。

[参 考 文 献]

- [1] SOLTANI,NAVIMIPOUR.Customer relationship management mechanisms:A systematic review of the state of the art literature and recommendations for future research[J].Computers in Human Behavior,2016(61):667-688.
- [2] 王朝晖,钟超.基于客户价值的服装电商消费者细分研究[J].北京服装学院学报,2016,36(4):41-46.
- [3] SHUI HUA HAN,SHUI XIU LU,STEPHEN C H LEUNG.Segmentation of telecom customers based on customer value by decision tree model[J].Expert Systems with Applications,2012,39(4):3964-3973.
- [4] KERÄNEN J,JALKALA A.Towards a Framework of Customer Value Assessment in B2B Markets:An Exploratory Study[J].Industrial Marketing Management,2013,42(8):1307-1317.
- [5] 陈静怡.O2O 模式下面向客户的供应链协同转型研究[J].西安电子科技大学学报(社会科学版),2019,29(3):1-14.
- [6] 周毓萍,黄丽莉.基于组合赋权-模糊评价法的商业银行高净值客户价值评价研究[J].金融理论与实践,2019(1):60-64.
- [7] ZIAFAT H,SHAKERIM.Using Data Mining Techniques in Customer Segmentation[J].International Journal of Engineering Research & Applications,2014,4(9):70-79.
- [8] 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.
- [9] DURSUN A,CABER M.Using data mining techniques for profiling profitable hotel customers:An application of RFM analysis[J].Tourism Management Perspectives,2016(18):153-160.
- [10] 夏名首. 基于买家在线评价的购后评价特征数据挖掘[J].中国流通经济,2017(7):30-38.
- [11] 包志强,赵媛媛,赵研.基于 RFM 模型和聚类分析的百度外卖客户细分[J].计算机科学,2018,45(11):436-438.
- [12] 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.
- [13] 热依木江·克里木,孙倬,赵红,等.基于含异质性客户数据要素模型的客户终身价值分类研究——以库尔勒香梨网络销售数据为例[J].科技促进发展,2018,14(10):943-950.
- [14] HUGHES A M.Boosting response with RFM[J].Marketing Tools,1996(3):4-8.
- [15] 郑琼,周梅华.基于熵权法的移动社交信任分析[J].软科学,2019,33(5):129-134.
- [16] 徐文瑞. 基于 RFM 模型的顾客消费行为与顾客价值预测研究[J].商业经济研究,2017(19):44-46.

Research on E-commerce Customer Value Segmentation Based on RFM Model Improved by Entropy Weight Method

CHEN DONGQING¹, YE CHONG², HUANG ZHANGSHU²

(1. Fuzhou University Zhicheng College, Fuzhou 350002, China 2. School of Economics and Management Fuzhou University, Fuzhou, 350108, China)

Abstract: The segmentation of e-commerce customer value and the implementation of differentiated marketing strategy will help to improve the profits of e-commerce enterprises and enhance the competitiveness of them, and promote the high-quality development of e-commerce industry. This paper proposes a segmentation method of e-commerce customer value based on RFM model improved by entropy weight method. This method uses the entropy weight method to determine the weights of three scoring indexes of RFM model, calculates the scores to evaluate e-commerce customer value, and selects the optimal K-means clustering model to analyze e-commerce customer value according to the silhouette coefficient. Based on the sales data of a seller in Pinduoduo, it can be seen from the segmentation results of the customer value of the model that each customer group has a greater degree of differentiation, more in line with the classic distribution characteristics of customer value, and the model has a good effect.

Keywords: e-commerce customer value; customer segmentation; RFM model; entropy weight method; K-means clustering

本文推荐专家:

林 翊, 福建师范大学经济学院, 副教授, 研究方向: 电子商务、数据挖掘。

孙秋碧, 福州大学经济与管理学院, 教授, 研究方向: 统计信息管理及其应用、数据挖掘。