

基于改进 RFM 模型的社交电商细分研究

戴炜, 牛少彰

(北京邮电大学计算机学院, 北京 100876)

摘要: 在社交电商高速发展的背景下, 社交电商服务平台也随之兴起。本文根据社交电商的交易数据, 首次提出建立基于 RFM 模型改进的 RCFM 模型对社交电商进行细分。该模型在引入社交电商一定时间内的交易人数属性的基础上, 利用层次分析法优化社交电商价值得分的指标权重, 最后通过优化的 Kmeans 聚类算法实现对社交电商的细分。结果表明基于 RCFM 模型对社交电商的细分能充分体现出社交电商的商业价值, 对社交电商服务平台构建社交电商用户画像, 实现个性化服务具有积极的实践意义。

关键词: 计算机科学与技术; 用户细分; RFM 模型; 层次分析法; Kmeans

中图分类号: TP311

Research on Social E-commerce Segmentation Based on Improved RFM Model

DAI Wei, NIU Shaozhang

(School of Computer, Beijing University of Posts and Telecommunications, Beijing 100089)

Abstract: In the context of rapid development of social e-commerce, social e-commerce service platforms have also emerged. Based on the transaction data of social e-commerce, this paper first proposes to establish an improved RCFM model based on RFM model to subdivide social e-commerce. This model introduces the attribute of the number of traders in social e-commerce within a certain period of time, uses the analytic hierarchy process to optimize the index weight of social e-commerce value score, and finally implements the subdivision of social e-commerce through the optimized Kmeans clustering algorithm. The results show that the segmentation of social e-commerce based on the RCFM model can fully reflect the commercial value of social e-commerce, and it has positive practical significance for the construction of social e-commerce user portraits on the social e-commerce service platform and the realization of personalized services.

Key words: computer science and Technology; user segmentation; RFM model; analytic hierarchy process; Kmeans

0 引言

伴随着移动互联网的发展和手机硬件的更新, 用户在社交网络的时间逐渐增长, 电商行业内将注意力转向了社交平台, 社交电商成为了电商产业的新方向。社交电商作为一个新兴行业, 没有一个成熟的商业模式, 社交电商之间的收入差异较大, 作为给社交电商赋能的平台, 需要为社交电商提供优质化服务, 然而, 对于不同层级的社交电商, 服务平台如何为各个类型的用户做有针对性的服务, 有利于其快速稳定发展。

社交电商是基于人际关系网络, 利用互联网社交工具, 从事商品交易或服务提供的经营活动, 涵盖信息展示, 支付结算以及快递物流等电子商务全过程, 是新型电子商务的重要表

作者简介: 戴炜 (19950803-), 男, 硕士; 牛少彰, 主要研究方向: 大数据与数据挖掘

通信联系人: 牛少彰 (1963-), 男, 博导, 主要研究方向: 网络信息安全、网络攻防技术、信息内容安全、移动通信安全、信息隐藏技术、数字权益管理技术、软件安全以及计算机取证技术等多个方面. E-mail: szniu@bupt.edu.cn

现形式之一^[1]。社交电商作为电子商务一种新的衍生模式，是借助微博、微信等社交平台对商品进行传播、分享，通过社交互动的过程发布产品信息来完成商品的营销，从而诱发消费者的购买欲望，产生购买行为^[2]。2019年社交电商购物用户规模达到5亿，社交电商从业者将近5000万人^[3]，社交电商市场和用户规模不断扩大。

社交电商大多没有成熟的商业模式，目前也面临着许多的发展困境，以微信上的社交电商为例，频繁通过朋友圈等社交途径推销展示商品的方式，极易引起“朋友圈”的厌烦和不满，甚至会被认为是不道德的行为，被屏蔽或拉黑；社交电商在初期的发展，基本是一种“熟人+信任”的经济模式，将商品质量与个人品质挂钩，对人脉关系的依赖程度较高，大多是向商户个人的微信好友来推销商品，销售渠道的扩展受到商户信誉的制约，我国的信用体系尚不完善，很难保证社交电商的信誉问题，同时市场监管机制的不完善，也导致社交电商中存在大量假冒伪劣和三无产品，难以提高用户的忠诚度和信任心理。

在上述的问题中，不同的营销行为对于商户的收入影响很大，如何把握商品推销和潜在消费者数量之间的平衡，实现对潜在消费者的高效转化是社交电商商业模式中急需解决的关键问题。不同交易层级的社交电商其使用社交平台和辅助软件的行为，频次，和时长都不相同，对应的推广模式也有着较大差异。

所以本文利用社交电商的交易数据，基于RFM构建RCFM模型，利用层次分析法定性与定量相结合计算出RCFM模型的指标权重，在利用Kmeans聚类的基础上，利用轮廓系数法优化Kmeans算法的K值选取，Mean Shift结合Kmeans优化选取初始聚类中心，实现基于交易数据的社交电商精准细分，为企业构建社交电商用户画像，实现针对性服务提供基础。

1 研究现状

社交电商行业发展迅猛，但作为一种新的衍生模式，其中必然存在一些问题。社交电商作为一个新兴行业，相关的国内外研究较少，大多数研究或是从交易安全性的角度进行研究，Chuang Wei等^[4]关注社交电子商务的安全性问题，构建社交电子商务声誉模型，解决消费的安全性问题；Zhang,Bo等^[5]针对社交电商与消费者之间交易的安全性问题，提出一种基于在线社交信任度量的混合信任框架，为个人之间的信赖关系提供意见。或是从消费者的角度来进行分析，评估消费者购买意愿的影响因素^[6]；韩晴晴，谢煜^[7]引入感知风险、感知价格和感知乐趣三个变量，基于ATM模型构建社交电商平台用户满意度模型，分析各因素对于用户满意度的影响。目前从社交电商商户角度对社交电商主体进行分析的相关研究比较少。

在将RFM用于用户细分的研究中，严成^[8]引入企业感知价值和客户感知价值构建了RFMS模型来对网约车客户进行细分，采用分段赋权来考虑客户价值，细分结果表明RFMS模型对网约车平台加强客户价值分析等方面有着较好的指导意义。Zhou,J等人^[9]将RFM模型与稀疏K聚类算法结合，处理大型，高维和稀疏的消费者数据，获得市场细分结构，为决策者提供了一种有效的方发确定高价值市场。Yan,C等人^[10]基于RFM模型，增加了评估客户风险的索赔指标对财产保险客户的生命周期价值进行评估，使用hesitant fuzzy理论进行聚类分析，得到四个客户同质组。Song,MN, Zhao,XJ等人^[11]在大型数据集通过RFM的时间序列分段时间间隔对潜在用户进行估值，结合多重对应分析(MCA)通过RFM对聚类

结果进行正则化, 为客户关系管理提供了有效的方法。这些研究都是以客户为主体结合客户的消费数据构建 RFM 模型进行细分, 没有将 RFM 模型应用到商户细分的用例。

80 2 社交电商精准分级问题描述与分析

2.1 基于社交电商交易数据精准分级的问题与分析

对于不同层级的社交电商, 交易层级低的用户更加注重实操层面的痛点, 如终端销售技能、朋友圈打造、引流等。交易层次高的社交电商团队长的痛点主要是团队裂变和团队管理。

85 根据用户的交易数据, 精准界定用户的交易层级, 对于用户的精准画像构建是必须的。各个交易层级用户的各种行为和使用习惯必须是相似的, 同时该层级也会反映出这一类用户的某种特质, 为用户画像的进一步分析提供基础, 对于社交电商日益增长的个性化需求, 可以有针对性的通过差异化服务, 针对性策略对社交电商提供服务。

90 在应用交易数据的用户细分领域中, 大多数都是针对于客户的消费情况对于客户进行细分和价值分析, 在客户细分中, 对于不同的业务场景和客户关注的不同特征, 参考客户的消费行为, 消费水平, 年龄段, 消费偏好等维度, 寻求客户之间的共性特征, 从而实现对客户的细分。

而对于社交电商而言, 从交易数据中进行微商细分需要考虑的是社交电商的收入情况, 类比于消费者的消费情况, 社交电商的收入数据展示了社交电商在商业活动中的收益情况, 这是不同营销行为, 营销策略对于社交电商收入的影响体现, 而对社交电商收入数据的维度和方法选择, 也影响着社交电商细分结果的准确性。

95 仅从社交电商的收入金额维度对其进行层级划分, 考虑社交电商的人数, 留存率, 转化率等随着收入金额变更的变化趋势, 可以认为处于同一收入层级的社交电商的人数, 留存率, 转化率的变化趋势是相近的, 在层级分割点前后的变化趋势有着较大的差异, 通过变化趋势的比对, 可以得到单维收入金额变量的划分点, 以此来确定社交电商基于收入金额的划分级别。但是单维分类难以适合大数据的多样性特征, 就算仅从收入数据来考虑, 仅考虑收入金额的多少不能有效的体现社交电商的一系列特征。

100 RFM 模型简单高效, 而且所需数据也是基于用户的历史交易数据, 所以基于 RFM 模型结合考虑社交电商的交易特点构建适用于社交电商交易数据细分的 RFM 模型体系, 来实现对社交电商价值的体现, 以对接下来的聚类提供基础。

105 2.2 RFM 指标选取与权重问题与分析

RFM 模型能够较为动态的反映用户在其生命周期内价值的变动情况, 在指导服务方进行用户留存、极大化用户价值等方面有着广泛应用。众多研究者结合零售、金融、电子商务等各个行业的特点, 从不同的角度改进优化 RFM 模型, 拓展模型的引用范围, 同时根据 RFM 对用户进行细分, 针对不同类型的细分用户制定合适的营销策略或构建用户画像。但是针对社交电商本身, 应用 RFM 模型对其进行细分相关方面的研究暂无相关研究产出, 而且考虑到针对社交电商的交易数据的细分与针对客户消费数据的细分有一定的区别, 无法直接应用 RFM 模型来对客户进行细分, 需要对模型进行一定的调整, 以此来反映社交电商的价值。

115 传统 RFM 模型中的客户交易频次与交易总金额存在共线性问题, 一般来说客户的交易频次越高, 交易的总金额也就越高, 这两个指标在解释客户价值时存在着一定的重复性, 同样在社交电商交易数据中也会存在这样的问题; 传统 RFM 模型中, 各个指标的重要性被认

120 为是相同的，但在实践中，各个指标的重要程度会在不同行业 and 不同发展阶段有着不一样的体现，比如对于低频高单价的商品或服务来说，行业在对客户细分时往往会比较看重客户的单次交易金额，而对于高频低单价的商品或服务，行业在细分时会比较看重客户的忠诚度即客户的交易频次指标来反映。对于社交电商而言，大多数社交电商出售的商品都属于快消品，消费频率高、使用时限短，且社交电商的订单产生依赖于其推广渠道，推广频度等，那么对于社交电商的 RFM 模型，其指标的重要性就会偏重于社交电商的交易频次与交易人数等相关指标，所以 RFM 模型相关指标的权重也是需要考虑的问题。

125 考虑上述问题，在传统 RFM 模型中以平均每次的交易金额代替总交易金额，消除交易频次与交易总金额的耦合关系，在模型指标中引入社交电商交易人数指标改进 RFM 模型，该指标反映社交电商社交关系网络的密集程度，改进前后的模型指标比较如表 1。

表 1 RFM 与 RCFM 的属性对比

	R(Recency)		M(Monetary)	F(Frequency)
RFM 模型	消费者最近一次消费时间距离分析点的时间		消费者一定时间内消费的总金额	消费者一定时间内的消费的次数
	R(Recency)	C(Clients)	AM(Average Monetary)	F(Frequency)
RCFM 模型	社交电商最近一次交易时间距离分析点的间隔	社交电商一定时间内产生交易的客户数量	社交电商一定时间内平均每次交易的金额即 M/F	社交电商在一定时间内的交易次数

130 有表 1 可以看出，RCFM 模型的各个属性是根据社交电商交易数据生成的新特征，根据这些属性可以较充分的考察社交电商的价值。R 属性体现了社交电商在分析点近期的活跃度；C 属性体现了社交电商的客户深度，从侧面体现出社交电商交易网络的健壮性；AM 属性体现的是社交电商单词交易的收益大小；F 属性为社交电商的交易频次，因为社交电商的交易订单往往和其体现了社交电商的商品推广力度和效率。

135 层次分析法(The analytic hierarchy process)简称 AHP,是一种广泛应用的定性与定量相结合的多属性决策方法，可以根据不同指标的重要性排序并依据重要性标度表赋值得到比较矩阵，从而得到指标权重^[8]。层次分析法可将定性问题量化，决策者可自行判断指标的相对重要性，以此来化为简单的权重进行计算。所以本文考虑使用层次分析法来确定 RCFM 模型中各指标的权重。

层次分析法的基本步骤分为以下几步：

140 1) 建立层次结构模型。在深入分析实际问题的基础上，将有关的各个因素按照不同属性自上而下地分解成若干层次，同一层的各个因素影响其上层因素同时也受到下层因素的影响。最上层为目标层，通常只有 1 个因素，最下层通常为方案或对象层，中间可以有一个或几个层次，通常为准则或指标层。

2) 构造成对比较矩阵，从模型的第二层开始，对影响上一层因素的同层因素用配对比较法和 1-9 的比较尺度构建成对比较矩阵。

145 3) 计算每一个成对比较矩阵的权向量，并进行一致性检验，若通过一致性检验，则矩阵的特征向量即为权向量；若不通过，则需重新构造成对比较矩阵。

3 基于交易数据的社交电商细分

3.1 数据采集与预处理

1) 数据采集

150 由于社交电商的线上交易属性，所以社交电商交易数据较容易获取。本文数据来源于某公司的社交电商辅助平台，获取该平台采集的社交电商交易数据，交易数据主要属性有：社交电商 ID，付款方 ID，交易时间，交易金额等方面的数据。实验包含了 2019 年 2 月 53307 个社交电商当月的交易数据，考虑到平台的运营模式，以每月作为细分时间段，利用 spark 分布式处理框架统计出每个社交电商最近一次交易与分析点（月份最后一天）的时间间隔 (R)，交易次数(F)，平均每次交易金额(AM)，和产生交易的客户数量(C)。抽取的部分数据
155 如表 2 所示：

表 2 社交电商 RCFM 模型指标数据

R	F	AM	C
0	1033	5.247444	123
1	42	114.4976	19
0	151	110.1187	92
0	109	291.7845	69
3	7	113.9514	6
0	30	137.8333	15
12	12	2028.043	10
0	21	137.1519	16
0	103	2.372039	24
24	11	165.6909	8

2) 数据预处理

160 由于 RCFM 模型数据指标的量纲各不相同，数据的取值也存在着较大的差异，当各指标水平相差很大时，如果直接不加处理使用初始指标值，会导致计算出来的结果中存在数值较高的指标在所占权重过大的问题。所以需要对其进行标准化处理从而消除量纲、分布差异较大以及指标数值大小的影响。

本文使用离差标准化，也称为 Min-Max 标准化，是对原始数据的线性变换，使得处理后的结果落在[0,1]区间，相应的转换函数如下：

$$y_i = x_i - \min / \max - \min, (1 \leq i \leq n) \quad (3-1)$$

165 公式中 min 和 max 是数据中样本的最小值和最大值。

考虑到 RCFM 模型中 R 属性对于社交电商的价值存在负相关关系，所以对 R 属性值进行标准化的转换函数为：

$$y_i = \max - x_i / \max - \min, (1 \leq i \leq n) \quad (3-2)$$

对表 2 中的数据进行 Min-Max 标准化后得到的结果如表 3 所示：

表 3 标准化之后的社交电商的 RCFM 指标数据

R	F	AM	C	R1	F1	AM1	C1
1	0.114743	0.00009	0.18209	0	1033	5.247444	123
0.962963	0.004559	0.00194	0.026866	1	42	114.4976	19
1	0.016678	0.00187	0.135821	0	151	110.1187	92
1	0.012008	0.00495	0.101493	0	109	291.7845	69
0.888889	0.000667	0.00194	0.007463	3	7	113.9514	6
1	0.003224	0.00234	0.020896	0	30	137.8333	15
0.555556	0.001223	0.03444	0.013433	12	12	2028.043	10
1	0.002224	0.00233	0.022388	0	21	137.1519	16
1	0.011341	0.00004	0.034328	0	103	2.372039	24
0.111111	0.001112	0.00281	0.010448	24	11	165.6909	8

170 **3.2 RCFM 指标权重的计算**

本文利用层次分析法来确定 RCFM 模型中各指标的权重，利用层次分析法对各指标计算权重的具体流程如下：

a) 建立 RCFM 模型的层次结构模型

构建符合本文研究需求的单层次结构模型如图 1 所示：

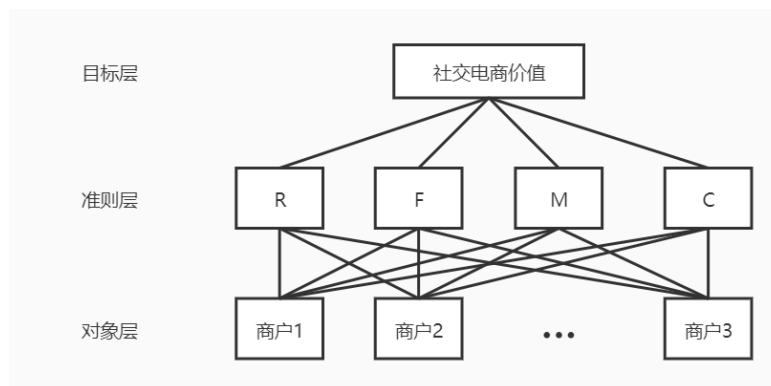


图 1 社交电商价值单层次结构模型

b) 构造判断矩阵并对判断矩阵赋值

根据下层各指标对其上层目标的影响程度确定该层各指标对上层目标所占权重。通过决策者对 R, F, M, C 四个指标进行比较排序，通过对两两指标之间进行比较得出比较结果 a_{ij} （第 i 个指标对第 j 个指标的相对权重）， a_{ij} 在 1-9 及其倒数之间取值，值与指标间相对重要性成正相关。

针对本文研究内容，在参考社交电商运营模式，盈利点的基础上，与某公司社交电商辅助平台的资深运营人员反复研究后确定各指标的相对重要性。首先，对于社交电商而言，如何增加自己的潜在客户群体，扩大自己推广行为的辐射面，提高受推广人群到消费客户的转化率是其所需要考虑的首要问题。大多数社交电商出售的商品都属于快消品，消费频率高、使用时限短，且社交电商的订单产生依赖于其推广渠道，推广频度等，那么对于社交电商的 RFM 模型，其指标的重要性就会偏重于社交电商的交易频次 F 与交易人数 C 等相关指标，

175

180

185

单笔交易金额 M 的重要性不高。

综上所述，社交电商的 RCFM 模型中，指标的重要性排序由高到低依次为 C、F、R、M。最终得出的判断矩阵如下所示：

$$\begin{matrix} & \begin{matrix} R & C & F & M \end{matrix} \\ \begin{matrix} R \\ C \\ F \\ M \end{matrix} & \begin{bmatrix} 1 & 1/5 & 1/4 & 5 \\ 5 & 1 & 2 & 7 \\ 4 & 1/2 & 1 & 6 \\ 1/5 & 1/7 & 1/6 & 1 \end{bmatrix} \end{matrix} \quad (3-3)$$

c) 计算权向量

确定下层各指标对上层目标的影响程度，用指标的权重来体现影响程度。通过求判断矩阵 A 的最大特征值 λ_{max} 即由公式 $A\mathbf{W}=\lambda_{max}\mathbf{W}$ 计算的到特征向量 \mathbf{W} 作为指标的权重向量。

d) 一致性检验

从理论上来说，若判断矩阵 A 是完全一致性矩阵，则满足 $a_{ik} \cdot a_{kj} = a_{ij}$ 。但实际上在构造成对比较矩阵时要求满足上述众多不等式是不可能的。因此只要求成对比较矩阵具有一定的一致性，容忍成对比较矩阵具有一定程度的不一致性。

完全一致的成对比较矩阵，其绝对值最大的特征值等于该矩阵的维数。那么对成对比较矩阵的一致性要求，则为矩阵 A 绝对值最大的特征值和该矩阵的维数相差不大。

计算阶数不小于 2 的判断矩阵 A 不一致程度的指标 CI 的公式为：

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (3-4)$$

N 阶矩阵的平均随机一致性指标 RI 只与矩阵的阶数 n 有关，对于固定的 n, 随即构造成对比较矩阵 A，矩阵的 a_{ij} 从 1、2、...、9、1/2、1/3、...、1/9 中随机抽取，取得足够充分的样本得到 A 的最大特征值的平均值，从而得到 RI 取值表如表 4 所示：

表 4 平均随机一致性指标表

N	1	2	3	4	5	6	7	8	9
RI	0.00	0.00	0.58	0.91	1.12	1.24	1.32	1.41	1.45

根据 CI 和 RI，计算成对比较矩阵 A 的随机一致性比率 $CR=\frac{CI}{RI}$ ，当 $CR<0.1$ 时，则可认为成对比较矩阵 A 的一致性是可以被接受的，而当 $CR>0.1$ 时，不能接受其一致性，需要重新对判断矩阵进行层次排序以及重新赋值。

由 RCFM 的判断矩阵得到特征向量 $\mathbf{W} = (0.12, 0.50, 0.33, 0.05)^T$ ，一致性指标 $CI=0.079$ ，计算一致性比率得到 $CR=0.0881<0.1$ 。因此判断矩阵 A 据有满意的一致性，则 RCFM 模型中的四个指标 R,C,F,M 的权重为 0.12, 0.50, 0.33, 0.05。

3.3 基于 K-means 的社交电商 RCFM 聚类

a) K-means 聚类算法

K-means 聚类算法是在用户细分中常用的聚类方法。其主要思想是随机选择 K 个初始聚类中心，对于数据集中的每一个数据点，根据欧氏距离公式 $D(x_i, x_j) = \sqrt{\sum_{m=1}^d (x_{im} - x_{jm})^2}$ (x_i, x_j 表示两个 d 为数据的对象)，计算其与 K 个初始聚类中心的欧氏距离，将其余最近的聚类中心关联，与同一个聚类中心关联的所有数据点组成一个簇；然后再重新计算每一个簇的聚类中心，重复以上步骤直到簇不再变化或达到最大迭代次数。

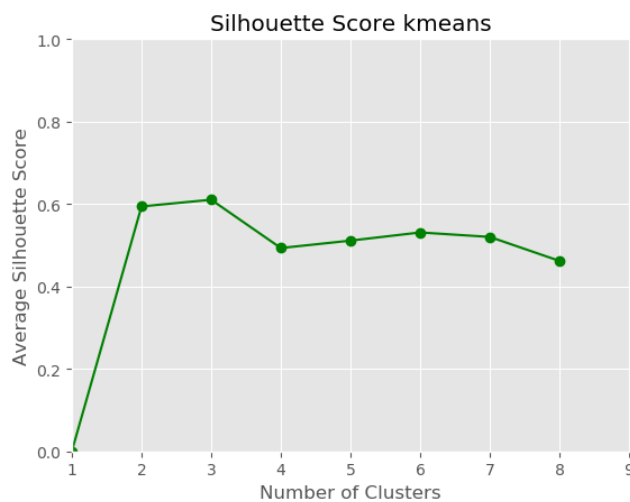
220 Kmeans 聚类算法简单快速，适合对大数据集进行处理，但从其基本思想可以看出，Kmeans 算法中目标聚类数目以及初始聚类中心的确定，对于聚类效果有很大的影响。随机选取聚类中心容易让聚类结果陷入局部最优解，一旦初始聚类中心选择不好，可能无法得到有效的聚类结果。

b) 优化 K 值选取

225 对于一个聚类任务，我们希望得到的聚类结果簇中，簇内尽量紧密，簇间尽量远离，轮廓系数 silhouette coefficient 便是类内密集和类间分散程度的评价指标。轮廓系数计算公式如下：

$$s = \frac{b - a}{\max \{a, b\}} \quad (3-5)$$

其中 a 代表同簇样本到彼此间距离的均值，b 代表样本到除自身所在簇外的所有簇的样本的均值，s 取值在-1 到 1 之间，s 越大则说明样本聚类合理，越小则表明样本更应该被分到其它簇中。所有样本的 s 的均值称为聚类结果的轮廓系数。使用轮廓系数来确定聚类 K 值，选取对应的轮廓系数较大的 K 值。K 取值 2 到 9 分别计算轮廓系数，结果如图 2 所示：



230

图 2 K 取不同值时的轮廓系数

从图中可以看到在 K 取值为 3 时，聚类结果拥有较大的轮廓系数，所以取 K=3 作为聚类算法的目标聚类数目。

c) 优化初始聚类中心

235 本文利用 mean-shift 算法来优化 Kmeans 算法的初始聚类中心选取，mean-shift 算法是一种基于核密度的聚类算法，又称为均值漂移算法，其主要思想是利用点的点的偏移均值，移动该点到其偏移均值，然后以此为新的起始点，继续移动，直到满足一定的条件结束，得到的点即为一个聚类的初始中心点，算法引入核函数使得样本偏移量对均值偏移向量的贡献随着样本与被偏移点的距离不同而不同。该算法的聚类流程会随机确定数据集空间内的一个确定半径的超球体即其球心，计算出超球体的质心，将球心移动至质心处，重复步骤直至球体内密度随球心滑动的变化低于设定阈值。

240

通过 mean-shift 聚类优化初始聚类中心，首先利用 Mean-shift 算法对数据集进行聚类，得到聚类质心，以及质心数量 C，若 C 与目标聚类数量不同，则用 Kmeans 对质心做一次聚类，聚类个数为目标聚类个数，将聚类后得到的质心作为 Kmenas 的初始中心点。

245

4 实验结果及分析

通过层次分析法计算 RCFM 模型权重得到 $WR=0.12$, $WC=0.50$, $WF=0.33$, $WM=0.05$, 加权后的各属性值如表 5 所示:

表 5 加权后 RCFM 示意图

R	C	F	AM
0.12	0.091044776	0.037865243	0.000004
0.115555556	0.013432836	0.001504336	0.000097
0.12	0.067910448	0.005503669	0.000093
0.12	0.050746269	0.003962642	0.000248
0.106666667	0.003731343	0.000220147	0.000097
0.12	0.010447761	0.001064043	0.000117
0.066666667	0.006716418	0.000403602	0.001722
0.12	0.01119403	0.000733823	0.000116
0.12	0.017164179	0.003742495	0.000002
0.013333333	0.005223881	0.000366911	0.000141

250

得到 RCFM 模型后引用 Kmeans 聚类算法以 R, C, F, M 四个变量为聚类属性对数据集进行聚类以实现社交电商的细分, 通过 K 值选取与聚类中心优化, 得到三个聚类中心作为初始聚类中心, RCFM 模型的细分结果如表 6 所示。

表 6 RCFM 模型聚类结果

聚类	AF	APM	AC	AR	RCFM	人数
低值用户	0.000149	0.000257	0.0015	0.031758	0.033663215	16582
中值用户	0.000435	0.000354	0.004496	0.106413	0.111698101	35841
高值用户	0.007505	0.000196	0.06524	0.11734	0.190281312	884

255

根据 RCFM 模型得到的结果, 从社交电商平台获取的 53307 个社交电商中低值社交电商共 16582 个, 占统计人数的 31.1%。这一类社交电商虽然每笔交易的金额较高, 但是交易频率和存在交易的客户数都很少, 最近的一单交易距离分析点远, 这类社交电商极有可能在接下来流失。

260

中值社交电商共 35841 个, 占统计人数的 67.2%。这一类社交电商的交易频率每月在 18 次左右, 存在交易的客户数也仅 10 个上下, 不过他们的每单交易金额是三类社交电商中相对较高的, 可以判断服务平台上的这类社交电商大多是兼职, 这类社交电商基本是一种“熟人+信任”的经济模式, 将商品质量与个人品质挂钩, 对人脉关系的依赖程度较高, 大多是向商户个人的好友来推销商品, 这也和服务平台自身的用户判断相符合。针对这类社交电商, 服务平台可以给他们提供推广资源, 重点培养这类社交电商。

265

高值社交电商共 884 个, 占统计人数的 1.7%。这类社交电商交易频率高, 存在交易的客户数也多, 可以是专业社交电商, 专业社交电商大多数出售的商品都属于快消品, 消费频率高、使用时限短, 所以单笔交易的金额也不会太高。对于高值社交电商, 服务平台可以有针对性的对其进行优质化服务。

5 结论与展望

本文在当今社交电商高速发展的背景下,引入社交电商的交易客户数量改进 RFM 模型,并且利用层次分析法对 RCFM 模型中的四个指标进行权重计算,在考虑到初始 Kmeans 聚类算法的相关缺陷后,利用轮廓系数优化聚类的 K 值选取,通过 Mean-shift 算法实现对 Kmeans 初始聚类中心的优化,使用优化的聚类算法实现基于社交电商交易数据的商户细分。从实验结果的分析可以看出,细分模型能够利用服务平台的社交电商交易数据较为准确的对商户进行细分,有助于服务平台提升服务质量,优化商户体验。

同时模型也存在着一些不足之处,需要在将来工作中加以改进。对于 RCFM 模型中的指标,还可以结合服务平台自身特点进行更加深入的细化,更好地总结出符合社交电商性质的指标。同时,针对 Kmeans 算法初始聚类中心以及 K 值选取的相关优化也有考虑不足的地方,后续实验中可以尝试使用更多的方法来对数据集进行粗聚类,作为 kmeans 聚类的前置步骤。

致谢

我的导师牛少彰老师在我写论文期间给予了我很大的帮助,向导师致以诚挚谢意

[参考文献] (References)

- [1] 中国互联网协会. 社交电商经营规范 [EB/OL].[2018-07-06]<http://images.mofcom.gov.cn/ltfzs/201807/20180706163646991.doc>
- [2] 吉敏, 耿利敏. 分享经济时代下社交电商发展研究[J]. 电子商务, 2019(10):32-34.
- [3] 中国互联网协会. 2019 中国社交电商行业发展报告 [EB/OL].[2019-07-11]https://www.isc.org.cn/editor/attached/file/20190711/20190711170604_91792.pdf
- [4] Sun, Junjie & Wu, Lingyu & Li, Rongbing & Ma, Ziyu & Han, Weiwei & Li, Hongchao. The Influence of Social Network Characteristics on Consumers' Willingness to Buy in Social Business[A].10.1109/ICSSSM[C].2019.8887749.1-5.
- [5] Wei C, Yu ZJ. Research on social e-commerce reputation formation and state-introduced model[J]. KYBERNETES, 2017, 46(6):1021-1038
- [6] Zhang B, Yong RH. A Hybrid Trust Evaluation Framework for E-Commerce in Online Social Network: A Factor Enrichment Perspective[J]. IEEE ACCESS, 2017, 6:7080-7096
- [7] 韩晴晴, 谢煜. 社交电商平台用户满意度影响因素研究--以小红书和拼多多为例[J]. 物流工程与管理, 2020, 42(03):104-106.
- [8] 严成. 基于改进 RFM 模型与聚类算法的网约车客户价值细分研究[D]. 华北电力大学(北京), 2018.
- [9] Zhou J, Zhai L, Pantelous A A. Market Segmentation Using High-dimensional Sparse Consumers Data[J]. Ssrn Electronic Journal, 2019.
- [10] Chun Y, Haitang S, Wei L, et al. An integrated method based on hesitant fuzzy theory and RFM model to insurance customers' segmentation and lifetime value determination[J]. Journal of Intelligent & Fuzzy Systems, 2018:1-11.
- [11] Song M, Zhao X, Haihong E, et al. Statistics-based CRM approach via time series segmenting RFM on large scale data[J]. Knowledge-Based Systems, 2017, 132(sep.15):21-29.