

Meticulous Marketing of Mutual Fund Customers Based on RFMF Model and K-Means Method

Xianlin Huang

School of Information Technology & Management, University of International Business and Economics, Beijing

Email: huangxianlin84@163.com

Received: Oct. 15th, 2019; accepted: Oct. 30th, 2019; published: Nov. 6th, 2019

Abstract

With the rapid development of Internet technology, Internet channels have become the key sales channels for mutual fund companies. Meticulous marketing is critical to reduce costs and improve operational efficiency. Facing customers with huge financial potential growing up with the Internet, how to effectively distinguish them and implement meticulous marketing strategies for different types of customers is an urgent problem for the companies. This paper proposed an improved RFM model, which measures customer value from three dimensions of recency, frequency and management fee, and grouped users through K-Means algorithm. The model was verified by using the historical data of a certain fund company's product, and five clusters were obtained. Based on the characteristics of users in the five clusters, targeted marketing strategies were proposed.

Keywords

RFM Model, Customer Clustering, Meticulous Fund Marketing

基于RFMF模型和K-Means方法的公募基金用户精细化营销

黄先林

对外经济贸易大学信息学院, 北京

Email: huangxianlin84@163.com

收稿日期: 2019年10月15日; 录用日期: 2019年10月30日; 发布日期: 2019年11月6日

摘要

随着互联网技术的飞速发展,互联网渠道已经成为了公募基金公司重点耕耘的销售渠道。精细化的营销对于降低营销成本,提高运营效率至关重要。面对随着互联网成长起来的具有巨大理财潜力的用户,如何对其进行有效区分,并针对不同类型人群实施精细化的营销策略,是当前公募基金公司急需解决的问题。本文提出了改进的RFM模型,从近度、频度、管理费三个维度对用户价值进行衡量,并通过K-Means算法对用户进行聚类,从而实现用户分群。并使用了某基金公司的某产品的申赎历史数据对模型进行了验证,得到了5个聚类结果,并针对5个聚类中用户的特点,提出了具有针对性的营销策略。

关键词

RFM模型, 用户聚类, 基金精细化营销

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

近年来,得益于互联网的进一步普及,中国的互联网金融得到了快速发展。互联网金融的飞速发展,对传统金融机构既是一个巨大的挑战,也是一个转型的机遇。公募基金公司传统销售渠道主要以商业银行、券商为代表,过于依赖银行券商等机构,难以触及随着互联网成长起来的具有巨大理财潜力的用户,无法为其提供个性化的服务。因此,互联网渠道成为了基金公司目前重点耕耘的渠道。精细化的营销对于降低营销成本,提高运营效率至关重要,而用户区分是精细化营销的基础,如何对规模巨大的互联网用户进行有效区分,并针对不同类型人群实施针对性的营销策略,是当前公募基金公司急需解决的问题。

前人对于基金销售策略的研究多从公司层面进行展开,如裴蕾(2017)对基金公司产品、价格、服务、渠道等角度的策略[1]进行了定性的研究,缺乏从用户角度的定量的论据。邹定斌(2005)等人从宏观的角度,阐述了公募基金市场应当如何实施精细化营销,如推进网上销售、改进管理费计提方式、增强同业竞争等[2]。总之,前人研究缺乏从基金公司层面,从用户角度出发的精细化营销的研究。

2. 模型和方法

2.1. RFMF 模型

RFM模型是一种重要的客户响应分析方法[3][4],该模型通过用户的最近购买的时间R、购买频率F以及购买的金额M三项指标来衡量用户的价值水平。RFM模型较为动态地展示了一个客户的基本行为特征轮廓,这对个性化的服务和营销提供了依据。但由于基金产品的特殊性,购买的金额不能较为全面表示用户的价值,因此,本文将购买金额改变为用户购买的金额与持有时长的乘积,得到了RFMF模型。

为了确定R、F、MF三者之间的权重关系,本文使用了美国运筹学家托马斯·塞蒂提出并广泛应用于经济计划和管理、能源政策和分配、行为科学、军事指挥等领域的层次分析法。根据专家们用1~9标注法对各指标相对重要性程度的判断,构造出判断矩阵,通过比较判断矩阵一致性检验后,通过解矩阵最大特征值对应的特征向量便可确定出各特征的权值[5]。

假设事先已知 3 个评价指标的权重向量为： $W = [w_r, w_f, w_m]$ ，比较 R, F 之间的重要性时， $C_{rf} = w_r/w_f$ 是精确比值。同理，任意两个指标之间的相对重要性都可以得到精确比值，从而可得构成的比较判断矩阵：

$$\bar{A} = \begin{bmatrix} w_r/w_r & w_r/w_f & w_r/w_m \\ w_f/w_r & w_f/w_f & w_f/w_m \\ w_m/w_r & w_m/w_f & w_m/w_m \end{bmatrix}$$

是完全精确的比较判断矩阵，并且有：

$$\bar{A} \cdot W = \begin{bmatrix} w_r/w_r & w_r/w_f & w_r/w_m \\ w_f/w_r & w_f/w_f & w_f/w_m \\ w_m/w_r & w_m/w_f & w_m/w_m \end{bmatrix} \cdot \begin{bmatrix} w_r \\ w_f \\ w_m \end{bmatrix} = 3 \begin{bmatrix} w_r \\ w_f \\ w_m \end{bmatrix}$$

其最大特征值为 $n = 3$ ，对应特征向量为 W 。

因此，通过构造判断矩阵，求出矩阵最大特征值所对应的特征向量，即为欲求的权重向量。

判断矩阵构造步骤：

通过比较 R、F、MF 三个指标两两之间的相对重要性程度，比如判断矩阵的第一列元素的确定，根据图 1 中的 1~9 标注法，结合实际业务，专家打分认为相同指标 R 之间的重要性相同，因此判断矩阵元素 C11 为 1，购买频次 F 比购买进度 R 稍微重要，因此，判断矩阵中元素 C21 为 3，带来的管理费 MF 比购买进度 R 远远重要，因此判断矩阵中元素 C31 为 7。同样的方法可以继续构造判断矩阵的其他两列。

通过计算构造好的图 1 中的判断矩阵的最大特征值所对应的特征向量即为所求的权重向量，计算得出 R, F, MF 之间权重为 [0.08, 0.19, 0.73]。

$$RFMF_score = 0.08 * R + 0.19 * F + 0.73 * MF \tag{1}$$

1~9 标度法					
属性值 a_{ij} 赋值	重要性等级				
1	属性 C_i, C_j 同等重要		R	F	M
3	属性 C_i 比 C_j 稍重要				
5	属性 C_i 比 C_j 明显重要	R	1	1/3	1/7
7	属性 C_i 比 C_j 强烈重要				
9	属性 C_i 比 C_j 极端重要	F	3	1	1/5
1/3	属性 C_i 比 C_j 稍不重要				
1/5	属性 C_i 比 C_j 明显不重要	M	7	5	1
1/7	属性 C_i 比 C_j 强烈不重要				
1/9	属性 C_i 比 C_j 极端不重要				

Figure 1. 1-9 Marking and comparing matrix
图 1. 1~9 标注法以及比较判断矩阵

2.2. K 均值聚类方法

“K 均值”这一词最早由麦奎因(1967) [6]提出，但本算法思想可以追溯到 1957 年，斯坦豪斯(1957) [7]、劳埃德(1957) [8]分别在各自的研究领域独立提出来的思想，这一思想在空间聚类分析种应用广泛。简单、高效、成功的应用案例和经验，是其出现数十年后，仍然广为使用的最重要的原因。该算法最常见的形式是采用被称为劳埃德算法(Lloyd algorithm)的迭代式改进探索法。算法首先把输入点分成 k 个初始化分组，可以是随机的或者使用一些启发式数据。然后计算每组的中心点，根据中心点的位置把对象分到离它最近的中心，重新确定分组。继续重复不断地计算中心并重新分组，直到收敛，即对象不再改

变分组。

3. 实验

3.1. 数据集

本文采用的数据集为某基金公司某产品在 2018 年 12 月 11 日上线互联网渠道至 2019 年 8 月 14 日的产品每一笔申购和赎回数据。通过数据处理，得到了每一个用户的 R、F、MF 数据。

3.2. 数据处理

为了去除数据的单位限制,将其转化为无量纲的纯数值,便于不同单位或者量级的指标能够进行和加权,需要对数据进行标准化。由于大部分数据都呈现长尾分布,80%以上都集中在低频低额区间,少数的用户提供了大部分销售,数据的分布十分不均匀,如果直接对用户的 RFMF 数据做线性标准化,会使得 80%的低频低额的用户之间的差别很小,难以做到用户区分,不符合业务场景和需求,因此本文采用分位数的方式,对用户的 RFMF 三维数据进行打分,1~10 分。比如:MF 这一特征,排名前 10%的用户,在 MF 这个指标上给用户打 10 分,前 10%~前 20%的用户打 9 分,前 20%~前 30%的用户打 8 分,依此类推。

3.3. 聚类个数选择

K 均值算法的第一步是确定聚类的个数,生成 K 个质心。确定聚类个数的常用方法包括肘部法则[9][10]和 CH 指标[11]。肘部法则是指:根据数据的潜在模式,当设定的聚类个数不断逼近真实聚类个数时,均方误差呈现快速下降态势,而当设定聚类个数超过真实聚类个数时,均方误差也会继续下降,但下降速度会趋于平缓。通过画出 K-均方误差曲线,找出下降途中的拐点,即可较好的确定 K 值,如图 2,可知 K 最优值为 5。

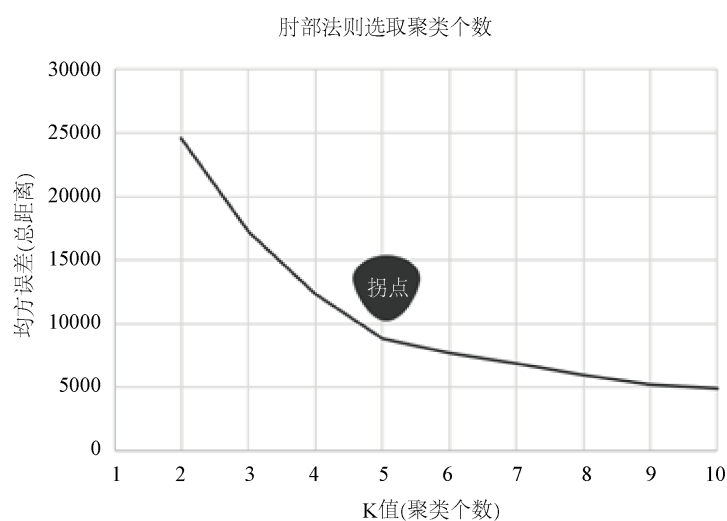


Figure 2. K-SSE curve

图 2. K-均方误差曲线

CH 指标通过计算类中各点与类中心的距离平方和来度量类内的紧密度,通过计算各类中心点与数据集中心点距离平方和来度量数据集的分离度,CH 指标由分离度与紧密度的比值得到,因此,CH 越大代表着类自身越紧密,类与类之间越分散,即更优的聚类结果。画出 K-CH 值曲线,如图 3 可知, K 的最优取值为 5。两种 K 值确定方法都说明 K 取值为 5 为最优情况,因此,确定为 5 个聚类。

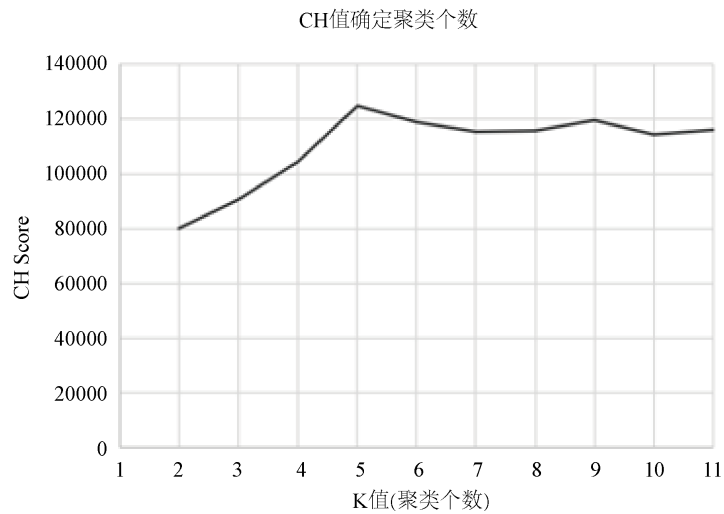


Figure 3. K-CH curve
图 3. K-CH 值曲线

4. 实验结果分析

代码运行结束，得到如表 1 中 5 个聚类，为了更直观地了解各聚类中用户的特点，统计了各聚类人数，用户 R、F、MF 三个指标的平均值和在中位数，RFMF 分值，以及各聚类的用户留存率。

Table 1. RFMF clustering result

表 1. 聚类 RFMF 结果

聚类	人数	留存用户比例	RFMF_score		R (天数)		F (频次)		MF (管理费)	
			平均值	中位数	平均值	中位数	平均值	中位数	平均值	中位数
聚类一	36,948	44.10%	6.3	6.27	211.66	212	1.18	1	13.38	4.7
聚类二	40,589	40.86%	2.53	2.54	214.15	213	1.06	1	0.16	0.05
聚类三	35,017	46.90%	3.14	3.1	138.41	176	1.26	1	0.19	0.1
聚类四	35,396	60.24%	6.73	6.67	150.49	177	1.53	1	14.61	4.84
聚类五	7,338	99.95%	7.61	7.81	38.41	10	18.64	10	29.86	3.56

(1) 聚类一中的用户，总体价值较高，人均累计申购 16,429.22 元，带来的管理费人均 13.38 元，较高，但上次申购距今人均超过 211 天，近期不够活跃。

聚类一中这些客户已经不太活跃了，流失也比较多，首先基金公司应该了解用户的去向，是从本产品转到了公司的其他产品上，还是已经完全流失，成为其他公司的客户，如果已经完全流失的客户，可以通过短信推送的方式，在本产品业绩较好的时期向用户推荐，或者推送公司其他的产品，优先将用户拉回本公司产品线上。

(2) 聚类二中的用户属于一般挽留用户，用户数量最多，但用户价值最低，RFMF 评分仅有 2.53，最近一次申购时间平均超过了 214 天，人均累计申购 580 元，管理费人均 0.16 元，活跃时间也很久远，用户留存率最低，总体来说用户带来的效益很低。

聚类三中的用户为一般发展用户，RFMF 价值评分较低，申购金额、持仓金额、管理费方面都较低，人均累计申购 1690 元，管理费 0.19 元，但活跃时间比较近。总体用户价值仅高于聚类 2 的用户。

聚类二、三中的用户价值水平都比较低，这类用户对基金公司来说，挽留和保持成本较高，但经济收益较低，可适当降低在这些用户群体上的营销成本投入。

(3) 聚类四中的用户是重点发展客户，RFMF 评分高，相对来讲近期比较活跃，最近一次申购行为距今仅 150 天左右，人均管理费方面也达到了 14.61 元，用户留存率超过 60%，但申购频次比较低，具有成为高净值的发展潜力。

针对这类群体，基金公司应当加强投资者教育，多多宣传用户定投的好处，着重引导用户增加申购频次，同时帮助降低用户亏损的风险，将用户高净值潜力变现。

(4) 聚类五中的用户是重要价值用户，用户十分活跃，最近一次申购距今天数距今平均在 38 天左右，而人均购买频次超过 18 次，其中超过一般的用户的申购频次都在 10 次以上。用户带来的管理费方面，人均管理费达到了 29.86 元，是 5 个用户群体中最高的。这类用户群体是基金公司的高价值客户，也是基金公司重点保持的优质客户，为公司带来的经济效益最高，但实际用户比例却很低。

针对这类群体，一方面，基金公司应该注重用户留存，定期监测重要价值客户的流失情况，分析原因，并且推出一些鼓励性的活动，使其对于保持当前状态更有兴趣，例如每月推出专属福利，如申购费用折扣、礼品赠送、电话回访等。另一方面，基金公司可以进一步发展这类客户，如：从中抽取部分用户，提供 VIP 服务，引导其做更为丰富的大类资产配置，形成多元化的资产配置方案，提高风险收益比，增加用户更换基金公司的成本，增加用户粘性。

5. 研究局限与未来展望

5.1. 研究局限

1、限于数据来源，只运用了一只基金产品的申赎数据，对用户价值的概括较为片面。

2、本研究从事后的角度，研究了已经有过购买行为的用户对于基金公司的价值，对于新用户，本方法会有冷启动问题，无法衡量用户价值，并加以区分。

5.2. 未来展望

1、在丰富数据类型角度，一方面，可选择用户购买的所有产品的数据，而不仅限于一只产品，可以更加全面、准确地衡量用户的价值以及用户的偏好，并为其提供相应的营销策略。另一方面，可以增加用户个人信息，如用户所述地区、用户收入水平、用户年龄等，帮助刻画用户特征，事前衡量用户价值，解决冷启动问题。

2、模型优化角度，在本文模型的指标权重确定的方法上，采用的是专家打分和层次分析法，定性和定量的方法相结合，但总体来说难以避免专家个人主观因素的影响，缺乏客观性。未来的研究方向上，可以改进指标权重的确定方式，更加客观地优化各指标权重的赋值方式，增强客观性。

参考文献

- [1] 裴蕾. 新形势下我国开放式基金营销策略研究[J]. 内蒙古科技与经济, 2017(7): 25.
- [2] 邹定斌, 宁娇丽. 我国证券投资基金营销现状与市场精细化营销对策[J]. 湖南科技大学学报(社会科学版), 2005, 8(1): 86-89.
- [3] Cohen, M.D. (2004) Exploiting Response Models—Optimizing Cross-Sell and Up-Sell Opportunities in Banking. *Information Systems*, 29, 327-341. <https://doi.org/10.1016/j.is.2003.08.001>
- [4] Miglautsch, J. (2000) Thoughts on RFM Scoring. *The Journal of Database Marketing*, 8, 35-43. <https://doi.org/10.1055/s-2000-11239>
- [5] 和金生, 赵焕臣, 杜秀珍. 用层次分析法探讨科研成果的综合评价[J]. 系统工程理论与实践, 1985(1): 33-38.
- [6] Macqueen, J. (1965) Some Methods for Classification and Analysis of Multi-Variate Observations. *Proceedings of*

Berkeley Symposium on Mathematical Statistics & Probability, Berkeley, June 21-July 18, 1965, 281-297.

- [7] Steinhaus, H. (1956) Sur la division des corp materiels en parties. *Bulletin L'Académie Polonaise des Science*, **1**, 801.
- [8] Lloyd, S.P. (1957) Least Square Quantization in PCM. Bell Telephone Laboratories Paper. Published in *Journal Much Later*: Lloyd, SP (1982) Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, **28**, 129-137.
- [9] 李永森, 杨善林, 马溪骏, 等. 空间聚类算法中的 K 值优化问题研究[J]. *系统仿真学报*, 2006, 18(3): 573-576.
- [10] 王千, 王成, 冯振元, 叶金凤. K-Means 聚类算法研究综述[J]. *电子设计工程*, 2012, 20(7): 21-24.
- [11] Calinski, T. and Harabasz, J. (1974) A Dendrite Method for Cluster Analysis. *Communications in Statistics*, **3**, 1-27.
<https://doi.org/10.1080/03610927408827101>