单位代码 11799
学　　号 2019610015

重慶工商大學

硕士专业学位论文

基于组合赋权法的 RFMN 模型的信用卡用户价值分析

论文作者：伍维维
所在学院：数学与统计学院
学科专业：应用统计
研究方向：大数据技术与商务应用
指导教师：孙荣
提交论文日期：2021 年 5 月 18 日
论文答辩日期：2021 年 5 月 21 日

中国•重庆
2021 年 5 月

目　　录

基于组合赋权法的 RFMN 模型的信用卡用户价值分析

摘 要

随着移动支付方式成为主流，信用卡风险也与日俱增。现今学者用于研究信用卡风险的模型有很多，在一定程度上可以达到降低信用卡信用风险的目的，但在确定各个影响因素的权重时，学者们仅单独考虑了主观赋权法或客观赋权法，并未将两者相结合。而 RFM 模型作为衡量客户价值的重要工具和手段之一，还未被学者用于研究信用卡风险中。本文创新地将组合赋权法和 RFM 模型相结合，建立了新模型，一定程度上降低了单独使用主观赋权法和客观赋权法在评价信用卡用户价值时所带来的不利影响，同时也达到了评价信用卡用户价值的目的。

在构建基于组合赋权法的 RFMN 模型阶段，由于主观赋权法侧重决策者的意图（即是决策者对不同指标的重视程度），主观性较强，而客观赋权法虽侧重客观，但可能会出现权重和实际相反的情况，本文放弃了单独使用主观赋权法或客观赋权法，选择将集合了主、客观赋权法优势的组合赋权法引入到传统的 RFM 模型中，并结合本文的研究对象，对原始模型的维度进行重新定义和增加新维度的操作，建立了基于组合赋权法的 RFMN 模型，将传统的 RFM 模型改造成适用于测算信用卡用户价值的模型，给出了信用卡用户综合价值得分计算公式。

在模型应用阶段，为了验证本文建立的基于组合赋权法的 RFMN 模型能够得到数据的最优聚类，本文对于组合赋权法、层次分析法、熵值法的 RFMN 模型分别进行了 K-Means 聚类分析，得到了对应的聚类结果，借鉴了单因素方差分析思想，采用组间均方评价不同类别信用卡用户综合价值得分的组间差异，得到了基于组合赋权法的 RFMN 模型的聚类结果对信用卡用户价值的划分效果更好这一结论。将基于组合赋权法的 RFMN 模型的聚类结果作为本文样本数据最优聚类结果的呈现，得到用户分类。对分类用户做 PCA 主成分分析，进行用户特征选择，参考因子得分系数，得到每类用户的特征。综合考虑 K-Means 聚类结果、PCA 主成分分析结果、数据统计分析结论和数据预处理结论，将本文的样本用户分为六类，分别是重要价值用户、重要保持用户、重要发展用户、一般保持用户、低价值用户和风险用户，给出用户画像并提出营销建议。

关键词：信用卡信用风险；RFM 模型；组合赋权法；K-Means 聚类

Analysis of Credit Card User Value Based on RFMN Model of Combination Weighting Method

ABSTRACT

As mobile payment methods become mainstream, credit card risks are increasing day by day. There are many models used by scholars to study credit card risk. To a certain extent, they can reduce credit card credit risk. However, when determining the weight of each influencing factor, scholars only consider the subjective weighting method or the objective weighting method alone. , Did not combine the two. As one of the important tools and methods to measure customer value, the RFM model has not been used by scholars to study credit card risks. This paper innovatively combines the combination weighting method and the RFM model to establish a new model, which reduces to a certain extent the adverse effects of using the subjective weighting method and the objective weighting method alone in evaluating the value of credit card users. Also achieved the purpose of evaluating the value of credit card users.

In the stage of constructing the RFMN model based on the combined weighting method, since the subjective weighting method focuses on the intention of the decision maker (that is, the degree of importance the decision maker attaches to different indicators), it is highly subjective, while the objective weighting method focuses on objectiveness, but There may be situations where the weights are opposite to the actual weights. This article abandons the use of the subjective weighting method or the objective weighting method alone, and chooses to introduce the combined weighting method that combines the advantages of the subjective and objective weighting methods into the traditional RFM model. Combining with the research objects of this article, redefine the dimensions of the original model and add new dimensions, establish an RFMN model based on the combination weighting method, and transform the traditional RFM model into a model suitable for measuring the value of credit card users. The formula for calculating the comprehensive value score of credit card users is introduced.

In the model application stage, in order to verify that the RFMN model based on

the combination weighting method established in this paper can obtain the optimal clustering of the data, this paper conducts K-Means for the RFMN model of the combination weighting method, analytic hierarchy process, and entropy method. Cluster analysis, the corresponding clustering results were obtained, and the idea of single-factor analysis of variance was used for reference. The mean square between the groups was used to evaluate the differences in the comprehensive value scores of different types of credit card users, and the aggregation of the RFMN model based on the combination weighting method was obtained. This conclusion is that the class result has a better effect on the division of credit card user value. The clustering result of the RFMN model based on the combined weighting method is used as the presentation of the optimal clustering result of the sample data in this paper, and the user classification is obtained. Perform PCA principal component analysis on classified users, select user characteristics, and refer to factor score coefficients to obtain the characteristics of each type of user. Considering K-Means clustering results, PCA principal component analysis results, data statistical analysis conclusions and data preprocessing conclusions, the sample users in this article are divided into six categories, namely important value users, important maintenance users, important development users, and general Keep users, low-value users and risk users, give user portraits and make marketing suggestions.

Keywords: credit card credit risk; RFM model; combination weighting method; K-Means clustering

第 1 章 绪论

1.1 研究背景及意义

1.1.1 研究背景

信用卡由商业银行或者信用卡公司发行，是信用合格的消费者的信用证明，具有消费支付、信用贷款、转账结算、存取现金等功能。最先发行的银行信用卡来自于美国加利福尼亚州的富兰克林银行。自 1952 年起，信用卡风潮迅速席卷美国、加拿大和英国等欧美国家，在 20 世纪 70 年代将这股风潮带到了中国香港、中国台湾、新加坡、马来西亚等地区。在信用卡在上述地区发展已趋于成熟时，随着改革开放的进程以及市场经济的发展，信用卡以电子化、现代化的消费工具的姿态进入中国，并在近十年里得到了跨越式的发展。

2020 年第二季度，中国支付清算协会在支付体系运行总体情况的季度报告中提到，截至 2020 年第二季度末，全国的发卡总数量为 86.58 亿张，其中信用卡和借贷合一卡共计 7.56 亿张。数据表明，截至 2020 年第二季度末，我国人均持有信用卡和借贷合一卡 0.54 张。而截至 2020 年第二季度末，我国信用卡和借贷合一卡的授信总额为 17.91 万亿元，人均授信额度 2.37 万元，授信使用率为 41.88%。以上数据均反映出我国信用卡行业仍具有很大的增长空间。

由于信用卡进入中国市场比欧美市场晚了近 20 年，国内大多数消费者的消费观念仍停留在“挣多少，花多少”，这使得信用卡市场还有大片的潜在用户还未得到挖掘，这其中蕴含的巨大经济利润使得国内各个银行争先拓展信用卡业务。但有的银行为了追求业务拓展的速度，对开卡用户的信用评估不到位，增加了坏账风险，使得银行卡的发展脱离正常发展态势。对比 2018、2019、2020 年这三年第二季度末的数据，可以发现，信用卡和借贷合一卡的授信总额逐年增加，但同时信用卡逾期半年未偿信贷总额也在逐年增加。这样的发展态势与拓展信用卡业务的初衷是相违背的。除了个别银行为了追求拓展市场会引起坏账风险之外，由于国内个人征信系统的不完善，也会增加对信用卡风险把控的难度。

如今，移动支付方式占据现代生活的主导地位，这使得信用卡风险日益加重。结合现今信用卡业务的发展来看，信用风险、欺诈风险和操作风险是现存信用卡风险中最主要的三大风险。通俗来讲，信用风险就是持卡个人或单位恶意透支信用卡并不按协定期限归还透支本金及利息的行为所造成的风险；欺诈风险则是由

于信用卡被冒名申请或领取、伪造、盗窃等原因造成的资金损失所形成的风险；操作风险则是由发卡机构在管理、操作上的不当而造成损失所形成的风险。以上风险若频繁出现，尤其是信用风险和欺诈风险不断累积，将会造成不可估量的损失，更有甚者，可能引发金融管理秩序紊乱，造成金融危机。

信用卡的操作风险需要各银行对员工进行更全面且细致的考核才可能降低，而降低欺诈风险则需要以信用风险的降低作为前提，因此，如何降低信用风险是首要考虑的问题。针对这一问题，可以通过对信用卡用户的价值评定来得以实现，银行或信用卡发卡中心会针对用户的价值制定不同的方案来提升（或抑制）用户的信用卡使用频次。但若只使用人工审核来评定用户价值，会造成大量人力物力的损耗，并且人工审核时间周期过长，会降低用户体验，对于用户留存会起反作用，因此，如何利用信息技术对信用卡用户进行价值分析是当前亟待思考和解决的问题。

1.1.2 研究意义

理论意义：

（1）研究 RFM 模型（最近一次消费 R、消费频率 F 和消费金额 M）的改进方法，以适用于信用卡用户价值分析；

（2）研究组合赋权法，将 RFM 模型改进成适用于本文的价值分析模型。

现实意义：

（1）可以降低人工审核所造成的人力物力损耗，缩短审核周期，提升用户体验；

（2）可以有效降低由于信用卡申领个人恶意透支、逾期不还而造成的信用风险，同时为降低欺诈风险奠定基础；

（3）建立的基于组合赋权法 RFMN 模型的用户价值评价体系以及模型应用阶段得到的结论，可以为国内的信用卡用户价值评估提供重大参考价值，得到的用户价值分类和用户画像也可以为国内信用卡用户提供参考，具有现实指导意义。

1.2 国内外相关研究现状综述

1.2.1 信用卡风险

在现存信用卡风险中，信用风险、欺诈风险和操作风险是最主要的三大风险。通俗来讲，信用风险就是由于持卡个人或单位恶意透支信用卡并不按协定期限归

还透支本金及利息的行为所造成的风险；欺诈风险则是由于信用卡被冒名申请或领取、伪造、盗窃等原因造成的资金损失所形成的风险；操作风险则是由发卡机构在管理、操作上的不当而造成损失所形成的风险。以上风险若频繁出现，尤其是信用风险和欺诈风险不断累积，将会造成不可估量的损失，更有甚者，可能引发金融管理秩序紊乱，造成金融危机。因此，如何去降低信用卡风险成为行业内亟待思考和解决的问题，在本文撰写之前，已经有许多学者针对信用卡风险的各类风险进行了研究并取得了很多成果。

有许多学者从理论上对信用卡风险以及风险管理进行了研究。楼芳（2004）、弋涛（2006）、魏鹏（2007）均在自己关于信用卡风险管理的论文中提到了我国现阶段存在的三大信用卡风险：信用风险、欺诈风险和操作风险[1-3]。从这三个方面入手，提出了相应的具有建设性的建议，例如健全发卡机构内部控制体系并推行“客户经理责任制”，对从业人员普及信用卡欺诈教育及培训，推广新防伪技术等。而有的学者则是通过对现实银行的研究，去揭示相应的信用卡风险特征以及防控方法。徐悦，管国强（2013）、马宇盟（2019）、汤虹（2020）分别从商业银行、中国银行、中国建设银行这三个银行去研究，但最终得到的风险特征与仅研究理论的学者所得出的相似，但针对不同类型的银行给出了相应的防控措施[4-6]。

而国内学者鲁齐（2019）、外国学者 Almhaithawi D，Jafar A，Aljnidi M（2020）则是通过对现实数据应用模型和技术方法进行信用卡欺诈风险研究，均成功实现了对信用卡欺诈的检测，但要广泛应用还需要通过长时间的检验[7,8]。

通过以上学者对信用卡风险的研究可以发现，要降低信用卡的操作风险，需要各银行从内部入手，对员工进行更全面的考核，对相关政策进行更细致的研读。而想要降低欺诈风险则需要以信用风险的降低作为前提，因此，如何降低信用风险是首要考虑的问题，这也是本文研究的重点。要降低信用卡信用风险，要从用户入手，建立对用户行为的评分系统，进而达到降低信用风险的目的。而现阶段研究信用卡信用风险的学者，主要是通过建立信用卡用户评分模型来实现对信用卡用户价值的评价，根据评价再提出相应的营销策略建议。

马海英（2008）在研究信用卡评分时，随机抽取 8371 位银行信用卡客户样本数据，利用 R 软件构建 Probit 与 Logistics 模型对其信用卡逾期风险进行评估，最终得到了许多具有现实意义的结论，并针对这些结论给出了相应的对策和建议[9]。而陈为民（2009）是从申请受理与审批和客户关系维护与管理阶段出发，以支持向量机为技术支持，先后建立了个人信用评分模型、拒绝推断模型和客户

行为评分模型，并完成了信用欺诈检测研究[10]，从内容上看是非常完整的，但仍然存在各个模型使用的检验数据不一致的缺陷，这是本文需要注意的地方。方匡南，吴见彬等人（2010）研究了信贷信息不对称下的信用卡信用风险，提出了如何利用统计分析、数据挖掘等技术建立分析模型，去识别和预测信用卡用户行为的风险[11]。把改进后的非参数随机森林分类(RFC)方法应用其中，并与逻辑回归模型和 SVC 模型进行比较，发现新方法的结果总优于传统模型。同样，国内学者闫思羽（2018）、杨怡滨（2020）、廖欣婷，谢磊（2020）和外国学者 Fonseca D P，Wanke P F, Correa H L（2020）以不同的分类算法作为技术支持，对信用卡逾期行为进行研究，都得到了相应的关于信用卡逾期行为的评分模型[12-15]，并且对相应的行为都给出了银行或信用卡发卡中心可以采取的应对手段和营销措施。

上述学者在建立信用卡用户评分模型时用到了 Probit 模型 、逻辑回归模型、支持向量机以及改进后的非参数随机森林分类(RFC)方法等方法，从用户逾期、信用评分、行为判断等方面评价了信用卡用户的价值。这些方法可操作性强，并且在一定程度上可以达到评判用户价值的目标，但同时也存在一定的不足。本文吸取这些学者在研究中的经验和不足之处，创新地提出了基于组合赋权法的 RFM 模型作为信用卡用户的评分模型，从用户的信用卡还款行为入手，评定各行为在用户价值中的权重，得到评价信用卡用户价值的综合评分系统，再进行用户分类，对各类用户的特征进行概括描述，达到评价信用卡用户价值的研究目的。本文选用的 RFM 模型是衡量客户价值和客户创利能力的重要工具和手段之一，但在信用卡信用风险的研究中还未被使用，本文接下来将介绍 RFM 模型的研究领域和所取得的成果。

1.2.2 RFM 模型

本文选用的 RFM 模型是衡量客户价值和客户创利能力的重要工具和手段，是众多客户关系管理的分析模式中应用最为广泛的。它利用最近一次消费（Recency）、消费频率（Frequency）和消费金额（Monetary）这三项指标来测算客户的价值，并将这三项指标的维度细分，最终得到至少 8 类用户：重要价值客户、重要发展客户、重要保留客户、重要挽留客户、一般价值客户、一般发展客户、一般保留客户和一般挽留客户。为了后续研究的简便性以及对人力物力成本的节约，一般将上述 8 类客户总结成 5 类：重要保持用户、重要发展用户、重要挽留客户、一般客户以及低价值客户。

王文贤，金阳（2012）等人在研究银行个人客户忠诚度时，应用了 RFM 模型[16]，将个人客户分成了多类客户群体，并对这几类客户群体给出了相应的处理方案。而 Chun Y，Haitang S，Wei L 等人（2018）在研究财产保险公司客户终身价值分析时，在 RFM 模型的基础上，加入评估客户风险的理赔指标，对财产保险客户的终身价值进行定量评估，利用基于犹豫模糊集的相似性度量理论进行聚类分析[17]，得到四个客户同质群体，计算出其终身价值得分，并分析其特征。再有国外学者 Sahar Tahanisaz（2020）在旅客对服务质量的满意度评估的研究中利用了 RFM 模型[18]，旨在通过考虑预购期望的多样性来找出航空公司如何满足旅客需求。最终得到了旅客分类并给出了合适的营销策略。

王威娜（2019）在对商场会员进行画像时，将 RFM 模型与模糊聚类算法结合[19]，刻画了会员的购买力，并给出了会员的生命周期和状态划分，为管理者策划促销活动提供了了便利。除此之外，将 RFM 模型和 K-Means 聚类算法相结合，是学者们常用的描绘用户画像、给出用户价值评分的方法。聂子临（2019）为对商场会员进行精准化营销，建立一种基于改进 RFM 模型的会员分类模型 RFA 模型，对会员价值进行分析[20]。首先采用层次分析法，确定 RFA 模型参数的权重；其次利用 K-Means 聚类算法，对商场会员进行分类；最后，计算每个会员价值得分。最终结果表明，改进的会员分类方法，能够很好地消除原模型指标的共线性。

而国外学者 P. Anitha 在对客户购买行为进行研究时，利用 RFM 与 K-Means 相结合的方法，对客户的购买记录和行为进行分析，最终得到了用户集群并给出建议[21]。再有卓灵（2019）、陈子璐（2020）、鲁燚（2020）、刘小红（2020）分别在对数字集群用户、电子商务用户、商场会员、品牌服装会员的画像进行描绘时都用到了 RFM 与 K-Means 相结合的方法[22-25]。最终都获得了用户集群的划分，并针对各类别的用户给出了相应的营销策略。但这样的方法还是有一定的不足，例如并没有给出用户评分的计算公式，不能明确地看到各类用户可能的得分区间。

陈东清,叶翀等人（2020）在对电商客户价值细分并实施差异化营销策略时，就意识到了上述学者的不足，提出了基于熵权法改进 RFM 模型的电商客户价值细分方法[26]，该方法利用熵权法确定 RFM 模型三个得分指标的权重，给出评分公式，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。

由于 RFM 模型是应用最为广泛的客户关系管理的分析模式，只要是涉及到线下商场、电子商务、理财公司、保险行业、航空行业以及本文将研究的银行信用卡行业等需要进行客户挖掘、维护的行业，都可能会用到 RFM 模型，并且会对其进行改进，以适应本行业的需求。同时，还没有学者在研究信用卡用户价值时，将 RFM 模型和组合赋权法相结合，因此，本文将创新地将两者结合起来。

1.2.3 K-Means 聚类算法

K-Means 聚类算法是一种迭代求解的聚类分析算法，旨在将数据集中在某些方面相似的数据成员进行分类重组，是一种无监督学习。自 K-Means 聚类算法提出以来，已经得到了很多学者在多领域的应用，并对其提出了改进。

在研究汽车忠诚客户细分方法时，任春华，孙林夫，吴奇石（2019）对传统 RFM 模型进行改进[27]，提出了 LRFAT 模型，再使用经过层次 K 近邻密度峰值优化后的 K-Means 聚类对客户进行分类，最终得到 3 类忠诚客户群，并对其提出了相应的营销建议。王一宾，黄志强，程玉胜（2020）运用 K-Means 聚类算法对无法解决标签缺损问题的 GLOCAL 算法进行改进[28]，最终使得获得的聚类中心矩阵可更好地表现出原始标签与潜在标签间的相关性。由于各种随机因素的制约使得采集综合能源数据的结果总是存在异常和缺失的情况，为了解决这一问题，韩帅，孙乐平等人（2020）提出了基于改进 K-Means 聚类和误差反馈的数据清洗方法，对异常数据进行识别和插补，提高了数据的稳定性和可靠性[29]。

K-means 聚类算法原理简单，易于实现和收敛速度快，但易受初始中心点影响且无法事先确定 k 值。粒子群优化算法搜索速度快，易于实现且不易受初始中心点影响，但具有不可避免的数据稀疏性和冷启动问题，且可扩展性差。为了解决两种方法的局限性，汤深伟（2020）提出了基于改进粒子群的 K-means 聚类算法，并引入艾宾浩斯曲线，在实验中得到改进算法具有更准确的推荐结果且结果更加稳定的结论[30]。

在医学领域，K-Means 聚类也有用武之地。范智浩，吕东澔，王晓峰（2020）在研究肝脏肿瘤分割时就应用了 K-Means 聚类算法[31]。他们以腹部 CT 为处理对象，使用了一种基于 K-means 聚类算法的图像分割方法对腹部 CT 中的肝脏肿瘤进行分割，并将结果与传统的区域生长算法的分割结果进行了对比，结果表明新方法具有更高的分割精度，为医务工作者的临床需要提供了新可能。

而在研究以词包围特征集的推文聚类方法时，外国学者 Poomagal S，Malar B 等人（2020）将改进的 K-Means 聚类算法作为研究方法，提出了改进 K 均值的 T_S 模型[32]，并将改进后的模型和传统随机 K-Means 聚类的词向量进行了性能比较，结果表明，在 70%的情况下，改进模型能获得更好的结果。

K-means 聚类算法在多个领域都有涉及，并且学者对其缺点也有相应的修正，本文在进行对信用卡用户的分类时，要随时借鉴这些学者已经修正了的缺点，并且还有注意不要产生新的缺陷。

1.3 论文主要内容与创新之处

1.3.1 主要内容

本文基于国外某银行的信用卡用户数据，对用户价值评估进行了研究，为国内的信用卡用户价值评估提供了重大参考。从结构上，本文分为四个部分：

第一部分：论文的第一章和第二章。分别对研究背景、研究意义以及相关理论和技术进行了阐述。

第二部分：论文的第三章。阐述了基于组合赋权法的 RFMN 模型和特征选择算法的构建流程。

第三部分：论文的第四章。利用国外某银行的信用卡用户数据对第三章中建立的模型进行应用。进行数据统计分析、数据预处理、维度值分箱处理、权重确定和得分计算等工作。

第四部分：论文的第五章和第六章。实现并对比基于主观赋权、客观赋权、组合赋权的 RFMN 模型的聚类结果。针对最优聚类结果进行特征选择，生成各类用户的用户画像。结合第四章中数据统计分析和数据预处理的结论，得到样本用户的最终用户分类，提出针对性营销建议。

1.3.2 创新之处

（1）模型的创新。在现有的识别客户价值的模型中选择应用最为广泛的，但还未使用到信用卡风险研究领域的 RFM 模型进行改进，将组合赋权法、特征选择算法与其相结合，运用到信用卡用户价值评估工作中。

（2）指标体系的改进。针对信用卡用户的信用度影响信用风险这一情况，在更新了模型参数定义的 RFM 模型的基础上，增加了逾期情况指标，构建了 RFMN 模型。

第 2 章 相关理论与技术概述

2.1 用户分层

2.1.1 传统 RFM 模型

传统 RFM 模型用于衡量用户价值和用户创利能力，由 Arthur Hughes 提出。以其可操作性强、可解释性强成为众多客户关系管理（CRM）分析模式中应用最广泛的。

RFM 模型中包括了 3 个构成数据分析最好的指标：最近一次消费 R（Recency）、消费频率 F（Frequency）和消费金额 M（Monetary）。RFM 分析基于这三个行为指标：一是最近一次消费 R（Recency），即是自上次消费后至当前的时间间隔，R 值越小说明用户再次消费的可能性越大；二是消费频率 F（Frequency），即是在确定的周期内产生的消费次数，F 值越大说明用户对产品的忠诚度越高，换言之，用购买频率越高，用户越忠诚；三是消费金额 M（Monetary），即使确定周期内用户的消费总额，M 值越高同样可以说明用户越忠诚。最传统的 RFM 模型如图 2.1 所示：

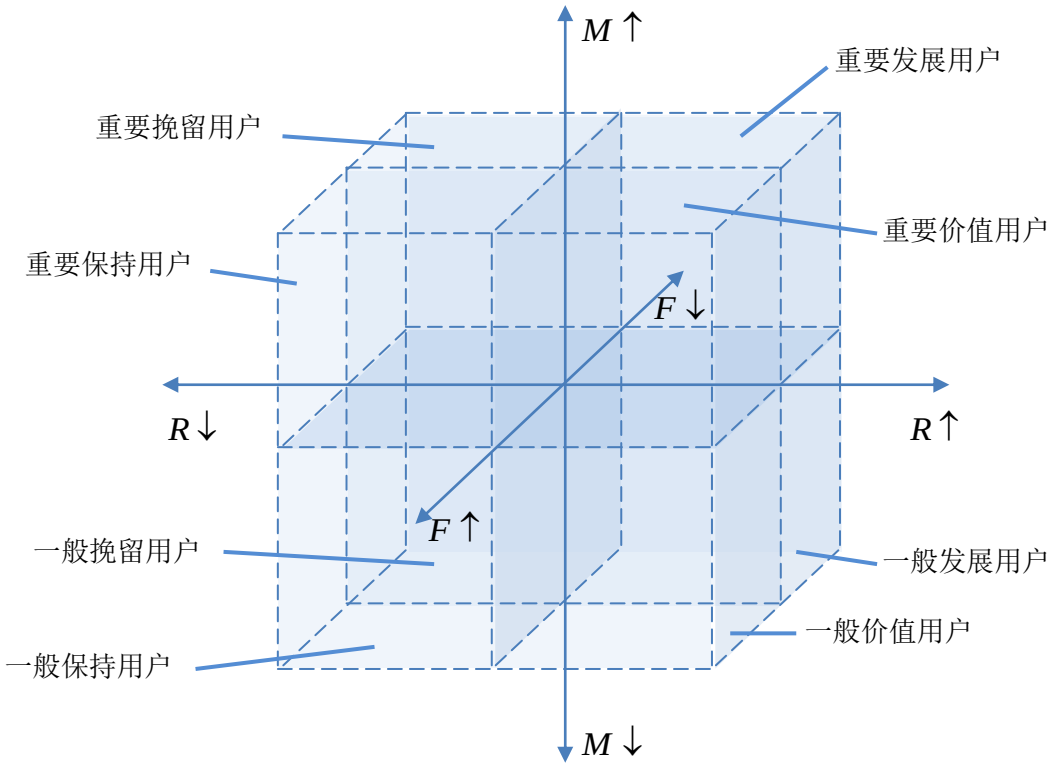

图 2.1 传统 RFM 模型

但由于在评估用户价值时要使用到的数据是长期的、大量的、具有复杂性的，

因此，现在常用的 RFM 模型是将以上三个指标分别进行五等分并用 1-5 分评分后再对用户类型进行细分：

R：上次消费日期越接近当前日期，得分越高，最高 5 分，最低 1 分；

F：消费频率越高，得分越高，最高 5 分，最低 1 分；

M：消费金额越高，得分越高，最高 5 分，最低 1 分。

具体细分方式有两种：

算法 1：$RFM = R \times F \times M$，产生 125 种用户细分；

算法 2：$RFM = R + F + M$，产生 15 种用户细分。

这两种算法都可以产生较多的用户细分，但算法 1 得到的细分更多，更利于评估用户的差异性，以便更精准地策划营销方案。这 125 种用户细分中有的用户可以合并成为一类用户，一般来说，会将这 125 种用户细分按八大类用户进行分层：重要价值用户、重要保持用户、重要发展用户、重要挽留用户、一般价值用户、一般保持用户、一般发展用户和一般挽留用户（如图 2.1 所示）。

在对用户的 R、F、M 三个维度进行评判时，常用到四分法和五分法，即是将 R、F、M 按四分位或五分位分为四或五份，并按照 R、F、M 与用户价值的关系，将数据区间与分数 1、2、3、4（或 1、2、3、4、5）进行匹配，得到基于传统 RFM 模型的用户价值分析结果。

2.1.2 K-Means 聚类算法

K-Means 聚类算法是经典的基于划分的聚类算法之一。算法根据簇类数目和初始聚类中心点，基于某种分配原则进行划分，然后重复执行更新中心点和分配对象集的操作，直到每个对象被划分到最合适的簇类中去。最终要达到的目的是使得类中的点尽可能集中，而类间的距离尽可能大。即是准则函数 F 达到最小，此时得到的聚类结果能够实现上述目的，使得得到的聚类结果达到最好。判断聚类结果优劣的准则函数 F：

$$\min F = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left\| x_i - c_j \right\|^2 \tag{2.1}$$

其中：k 为聚类数目，n_j 为 j 中的数据点总数，x_j 为类 j 中的数据点，c_j 为类 j 的聚类中心。

假设数据集 X 中有 k 个聚类子集 $C_1, C_2, \ldots, C_j$，每个聚类子集的聚类中心分别为 $c_1, c_2, \ldots, c_j$，K-Means 聚类算法的具体步骤如下：

①确定聚类数目 k；

②确定初始聚类中心 $c_1, c_2, \ldots, c_j$；

③依次计算对象 x_i 与每个聚类中心的欧式距离 d，以距离最小为划分标准进行划分；

④计算各新生成类的均值，作为新的聚类中心；

⑤重复③-⑤，直到聚类中心不再改变，终止迭代。

在上述步骤中有以下几个概念需要进行说明。

（1）对象 x_i 与 x_j 间的欧式距离为 $d(x_i, x_j)$，$d(x_i, x_j)=\sqrt{(x_i - x_j)^T (x_i - x_j)}$；

（2）聚类中心 $c_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i$。

2.2 赋权法

赋权法能够确定评价对象的指标权重，从而对复杂的对象做出评价。由于赋权法不依赖于个人主观判断，而是依托于数学方法将原始数据的关系转化成权重，这使得其具有较强的数学理论依据。赋权法分为主观赋权法，客观赋权法和组合赋权法。

2.2.1 主观赋权法

主观赋权法依托于决策者对主观信息的理解以及自身的专业知识、经验判断，常用的主观赋权法有层次分析法、模糊综合评价法、综合指数法等。

层次分析法将于最终决策相关的元素分解成目标层、准则层和方案层，在这基础上进行分析与决策。以其分析方法系统化、决策简洁明了以及所需数据指标少等特点，被学者广泛应用。模糊综合评价法的结果清晰，系统性也强，但主要解决的是难量化的问题，不适用于本文的研究。而综合指数法虽然计算上和结果呈现上占优势，但计算过程中仍然会用到层次分析法和模糊综合评价法去确定各个指标的评价标准值，主观性对结果的影响是叠加的。因此本文组合赋权法中的主观赋权法选择层次分析法。

运用主观赋权法确定权重，能够反映出决策者的意向，但得出的决策或建议主观性极强。在实际应用中，如果只借助主观赋权法，会增加决策失误的风险，使得决策者承担极大压力。

2.2.2 客观赋权法

客观赋权法是根据各个指标间的相关关系或者各指标的变异程度来确定权重系数。常见的客观赋权法有因子分析法、熵权法等。

因子分析是将具有相同本质的变量归为一个因子，得到原始变量中的潜在共性因子的分析方法。将多个变量通过一定规则进行组合，达到减少变量但尽可能不折损信息的目的，适用于初始变量多的情形，并不适用于本文的研究。而熵权法是通过指标的信息熵来判断指标对综合评价的影响（即权重），评价通过熵值判断，不受指标个数影响。因此，本文组合赋权法中的客观赋权法选择熵权法。

比之主观赋权法，客观赋权法遵从了数据本身的特性，更为客观，但同时也增加了权重与实际情况相反的风险。

2.2.3 组合赋权法

根据上述理论可知，主观赋权法侧重决策者的意图（即是决策者对不同指标的重视程度），主观性较强；而客观赋权法侧重客观，可能会出现权重和实际相反的情况。而将两者相结合的组合赋权法平衡了两者的优劣。组合赋权法可以弥补单一赋权带来的主观性/客观性过强的不足，同时又保留了主观随机性和客观公正性，使得权重能够实现主客观统一，评价真实公正。

将由主观赋权法得到的属性权重记为$W^{'}=(W_1^{'},W_2^{'},\ldots,W_n^{'})^T$，而由客观赋权法得到的属性权重记为$W^{''}=(W_1^{''},W_2^{''},\ldots,W_n^{''})^T$。两组权重向量分别满足

$$0\le W_j^{'}\le 1,\sum_{j=1}^{n}W_j^{'}=1;0\le W_j^{''}\le 1,\sum_{j=1}^{n}W_j^{''}=1. \tag{2.2}$$

则组合赋权权重为：

$$W=TW^{'}+UW^{''} \tag{2.3}$$

其中，T、U分别为$W^{'}$、$W^{''}$的重要程度。

为了使得$W=TW^{'}+UW^{''}$中的权重满足

$$0\le W_j\le 1,\sum_{j=1}^{n}W_j=1 \tag{2.4}$$

对T、U进行归一化处理。根据文献[33]对$\overline{T}$、$\overline{U}$进行定义：

$$\overline{T}=\sum_{i=1}^{m}\sum_{j=1}^{n}b_{ij}W_j^{'}/\sum_{i=1}^{m}\sum_{j=1}^{n}b_{ij}(W_j^{'}+W_j^{''}) \tag{2.5}$$

$$\overline{U}=\sum_{i=1}^{m}\sum_{j=1}^{n}b_{ij}W_j^{''}/\sum_{i=1}^{m}\sum_{j=1}^{n}b_{ij}(W_j^{'}+W_j^{''}) \tag{2.6}$$

其中，b_{ij}为规范化后的第i位用户第 j 个指标值，m为用户总数。

2.3 特征选择与用户价值分析

2.3.1 特征选择

特征选择是指从原始特征中选择最有效特征以实现数据集降维的过程，通常又被称作特征子集选择或属性选择。特征选择的过程一般包括产生过程、评价函数、停止准则和验证过程四部分。特征选择的基本框架如图 2.2 所示。

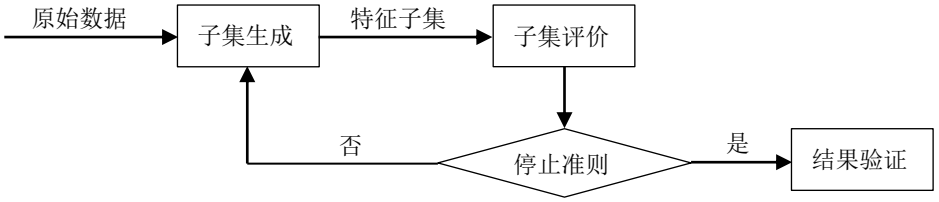

图 2.2 特征选择的基本框架

产生过程是特征子集的搜索过程，是为评价函数提供特征子集的过程；评价函数是评价特征子集好坏程度的准则；停止准则与评价函数相关，当评价函数达到停止准则的阈值时就可停止搜索；验证过程是对特征子集的有效性的验证。

由于信用卡用户相关数据的指标较多，本文将采用 PCA（主成分分析）算法对原始特征进行特征选择，得出信用卡用户数据的最有效特征。PCA 是一种无监督算法，是常见的降维方法之一。理论上第一主成分要尽可能多地解释原始数据，同时要使得变换后的低维变量间的相关性尽可能小，便于后续的分析。

2.3.2 用户价值分析

用户价值分析是根据用户的特征产生的用户画像，对用户的价值进行评估，实现对各类用户的特征的描述，给出对应的营销建议。

用户画像是用户标签的集合体，核心是了解用户。用户画像通过收集用户的基本信息、信用信息和其他维度数据来表现用户的特征属性，是用于分析和挖掘用户潜在价值的工具，也是为用户提供个性化服务的前提。

进行用户画像本质上是进行用户标记。常见的标记用户的方法有统计类标签、规则类标签和机器学习挖掘类标签。本文的用户标签均为统计类标签，例如用户的职业、就业年限、收入、信用等级等。用户画像的产生和应用过程如图 2.3。

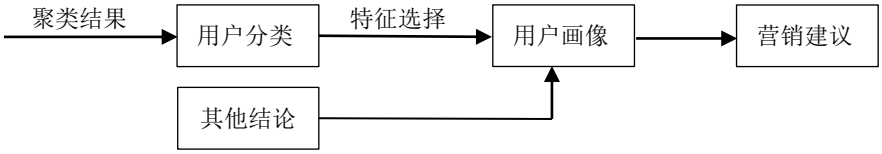

图 2.3 用户画像的产生与应用

2.4 本章小结

本章对传统 RFM 模型、K-Means 聚类算法和赋权法的相关理论进行了阐述。列出了根据传统 RFM 模型得到的八大类用户分类及其特征，为模型应用中，根据 K-Means 聚类算法得到的用户分层描述提供了理论依据。介绍了组合赋权法、特征选择和用户价值分析，为 RFM 模型的改造提供了理论支撑。

第 3 章 基于组合赋权法的 RFMN 模型的构建

3.1 模型构建流程

本文的模型改进是对RFM模型改进，提出了基于组合赋权法的RFMN模型。本文借鉴了陈东清[19]在点上客户价值细分研究中的模型构建流程进行模型的构建。但由于本文在确定 *R*、*F*、*M*、*N* 四个维度的权重时采用的是组合赋权法，并且在完成 K-Means 聚类分析后，为了得到各类别用户更完整的用户画像，本文对各类别用户进行了 PCA 算法的实现，获得了各类别用户的基本属性特征和信用等级特征。因此，在流程上有部分改动，修改后的模型构建步骤如图 3.1 所示，共有 9 个步骤。

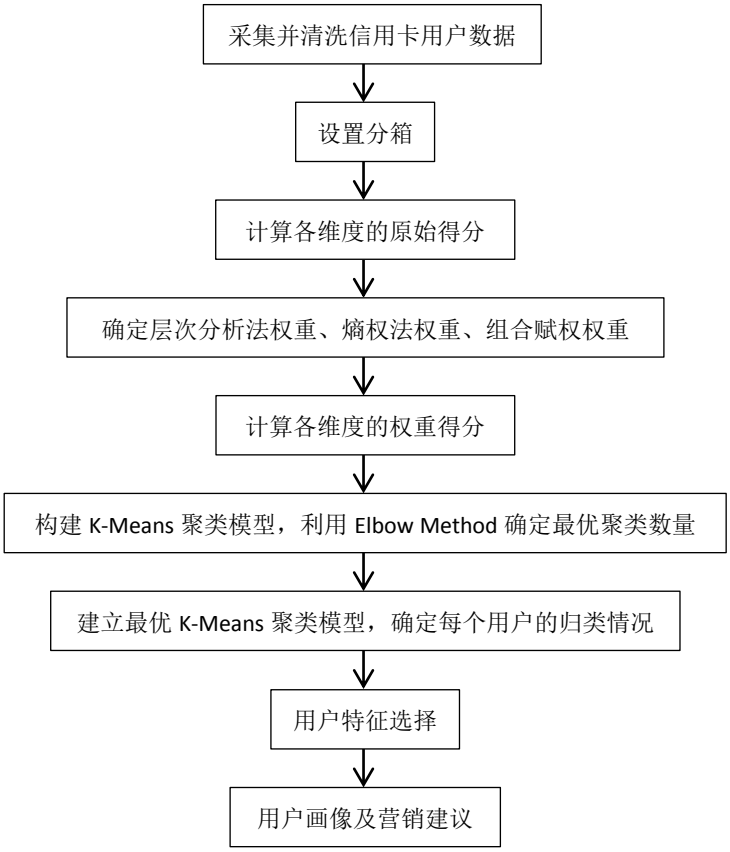

图 3.1 模型构建步骤

步骤 1：采集并清洗信用卡用户数据。由于信用卡数据包含了用户的各类信息（基本信息、信用信息、还款信息等），是涉及隐私的数据，因此，本文展示的数据是处理后的数据。对这些数据进行缺失值、异常值、规范化或归一化等预处理，确保呈现出来的数据和结果不会泄露用户隐私并具有可靠性和准确性。

步骤2：设置分箱。根据RFMN模型的一般操作流程，将R、F、M、N四个维度的数据按照五分法（0.2、0.4、0.6、0.8）进行分箱处理，将分箱处理后的每个等级分别赋值1-5分（赋值需考虑每个维度与用户价值的相关性）。

步骤 3：计算各维度的原始得分。根据分箱结果计算出每个用户的各维度的原始得分。

步骤4：采用层次分析法、熵权法确定权重$W^{'}$、$W^{''}$，根据组合赋权权重公式计算组合赋权权重W。

步骤5：计算各维度的权重得分。将步骤3中的得分作为基础得分，再分别乘以$W^{'}$、$W^{''}$、W，得到基于层次分析法、熵权法、组合赋权法的维度得分。

步骤6：构建K-Means聚类模型，利用Elbow Method确定最优聚类数量。由于本文的RFMN模型有四个维度，而每个维度均存在大于（等于）、小于平均值这两种情况，因此共存在16种情况。基于此，最大聚类值为16，最小为2，共需构建15个K-Means聚类模型。但由于RFM模型对用户的分类一般为八大类，因此本文仅需构建聚类数量为2-8的K-Means聚类模型，共7个，并利用Elbow Method得出最优聚类数量。对基于层次分析法、熵权法、组合赋权法的维度得分均进行此操作。

步骤7：建立最优K-Means聚类模型，确定每个用户的归类情况。借鉴单因素方差分析思想，比较基于层次分析法、熵权法、组合赋权法的维度得分的聚类结果，将最优聚类结果的最优聚类数量作为类别参数，建立最优K-Means聚类模型，确定每个用户的归类情况。

步骤8：用户特征选择。利用PCA算法对步骤7中得到的分类用户进行特征选择，得到各类用户的基本属性特征和信用等级特征。

步骤 9：用户画像及营销建议。结合数据统计分析结论、数据预处理结论和聚类结果，得出最终的样本用户分类。将聚类结果中的还款行为特征和通过PCA算法得到的用户的基本属性特征和信用等级特征相结合，对最终的样本分类用户进行用户画像，给出对应的用户特征描述和营销建议。

以上步骤是对模型构建流程的总体概述，接下来，本文将对模型构建流程中的重要步骤进行详细描述。

3.2 基于组合赋权法的 RFMN 模型

3.2.1 模型指标的改进

（1）模型指标的重新定义

由于 RFM 模型是衡量客户价值和客户创利能力的重要工具和手段之一，有消费才有创利可能，因此它本身的 R、F、M 这三个维度的定义是围绕着客户的消费情况而确定的。但本文的研究对象是信用卡用户，目的是为了评价信用卡用户的价值，研究的出发点是信用卡用户的还款行为，因此，为了使得 RFM 模型能够更好地服务于本文的研究对象，本文对传统 RFM 模型 R、F、M 维度进行了重新定义。

通过研究信用卡用户数据，并结合现实情况，本文将用户对信用卡的使用情况与 R、F、M 进行匹配，最终决定将用户最近一次还款时间点与分析时间点间的间隔、观测时间内的还款次数和观测时间内的还款总金额分别作为本文 RFM 模型的 R、F、M 。

利用上述三个指标含义对传统 RFM 模型的三个维度进行重新定义，使得传统 RFM 模型被修改成适应本文研究的模型，能够从信用卡用户的还款行为中更大程度地体现出该用户的价值。传统 RFM 模型与本文的 RFM 模型的参数对比如表 3.1 所示。

表 3.1 传统 RFM 模型与本文的 RFM 模型的参数对比

传统 RFM 模型	RFM 模型
R(Recency) 用户最近一次消费时间点与 分析时间点间的间隔	R(Recency) 用户最近一次还款时间点与 分析截止时间点间的间隔
F(Frequency) 观测时间内的消费次数	F(Frequency) 观测时间内的还款次数
M(Monetary) 观测时间内的消费总金额	M(Monetary) 观测时间内的还款总金额

（2）模型指标的增加

本文对传统 RFM 模型的参数进行了重新定义，从用户最近还款时间点与分析截止时间点间的间隔、还款频率和还款总金额这三个方面去描述了用户的活跃度和忠诚度，但对于用户的信用度并没有进行描述。在信用卡风险的防控中，信用风险的防控是相当重要的，要实现信用风险的防控，测算用户的信用度是最简

便的。因此，本文对模型的参数进行了改进，新增了一个分析维度：逾期次数 N 。指在观测时间内，用户逾期30天以上次数，N 值越大，说明用户的信用度越低，反之，信用度越高。本文通过上述几个维度来衡量用户价值，如表3.2所示。

表3.2 改进模型指标含义

模型指标	含义
R(Recency)	用户最近一次还款时间点与分析截止时间点间的间隔
F(Frequency)	观测时间内的还款次数
M(Monetary)	观测时间内的还款总金额
N(Number)	观测时间内的逾期30天以上次数

为了使得改进的 RFM 模型能更全面地测算信用卡用户的价值，本文利用上述四个指标建立了改进 RFM 模型——RFMN 模型。对以上四个指标分别进行五等分并用1-5分评分后再对用户类型进行细分，评分遵照如下法则：

R：上次还款日期越接近分析截止日期，得分越高，最高5分，最低1分；

F：观测时间内还款次数越多，得分越高，最高5分，最低1分；

M：观测时间内还款总金额越高，得分越高，最高5分，最低1分；

N：观测时间内逾期次数越少，得分越高，最高5分，最低1分；

在重新定义并新增了维度后，本文提出如下用户分层，给出各类用户的基本描述：

表3.3 用户分层描述

用户分类	R（离当前时间越近越好）	F（次数越多越好）	M（金额越大越好）	N（逾期次数越少越好）	用户描述
重要价值用户	高	高	高	高	优质用户
重要保持用户	高	低	高	高	较优质用户，需促进复购
重要发展用户	低	高	高	高	可能流失用户，需促活
重要挽留用户	低	低	高	高	有价值用户流失，需加强挽留
一般价值用户	高	高	低	低	忠诚用户，需加强复购
一般保持用户	高	低	低	低	新用户，需加强营销
一般发展用户	低	高	低	低	准流失用户，需高频率营销
一般挽留用户	低	低	低	低	已流失用户，需加强挽留

为了对本文模型各个指标的重要程度进行量化，本文将组合赋权法引入到RFMN模型中。接下来，本文将介绍如何建立基于组合赋权法的RFMN模型。

3.2.2 组合赋权法与RFMN模型的结合

如前文所述，主观赋权法和客观赋权法都曾被单独引用到信用卡风险模型中，

用于确定影响因素的权重。主观赋权法侧重决策者的意图（即是决策者对不同指标的重视程度），主观性较强；而客观赋权法侧重客观，可能会出现权重和实际相反的情况。而将两者相结合的组合赋权法平衡了两者的优劣。

组合赋权法可以弥补单一赋权带来的主观性/客观性过强的不足，同时又保留了主观随机性和客观公正性，使得权重能够实现主客观统一，评价真实公正。并且组合赋权法作为主客观统一的确定用户指标权重的方法，却还未被学者用于研究信用卡风险。因此本文使用组合赋权法确定 RFMN 模型中 R、F、M、N 四个维度的权重，这里用到的主观赋权法为层次分析法，客观赋权法为熵权法。计算过程如下：

（1）分别计算出层次分析法和熵权法下的维度权重向量；

（2）根据组合赋权权重计算公式 $W=\overline{T}W^{'}+\overline{U}W^{''}$ 计算出组合赋权权重向量。

接下来本文将详细介绍主观赋权法（层次分析法）和客观赋权法（熵权法）的维度权重计算方法。

3.2.3 模型权重的计算

（1）层次分析法下的维度权重计算

使用层次分析法获取三个维度各自的权重，首先需要按照层次分析法的原则，将相关因素分解成三层，得到基于 RFMN 的层次结构。

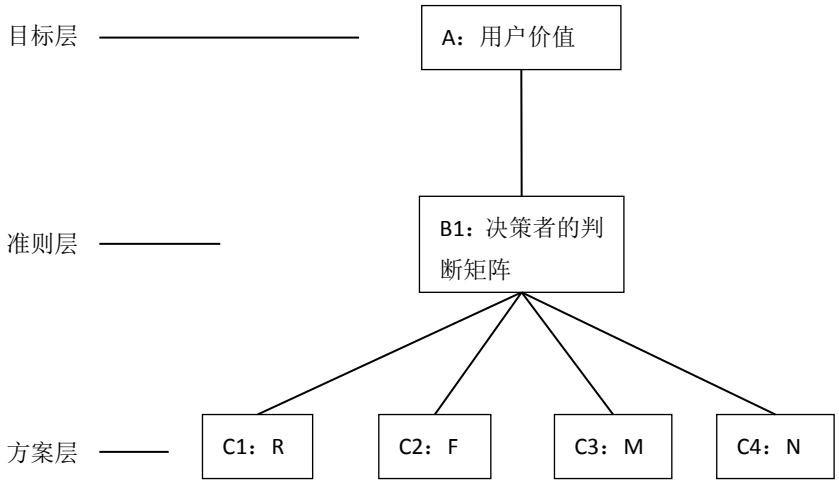

图 3.2 基于 RFMN 的层次结构

其次，根据文献[34]中提到的 1-9 尺度来确定各维度的权重。将专家给出的判断矩阵作为权重的数据来源根据，再通过 SPSSAU 软件进行层次分析，得到各维度的权重 $W_j^{'}$。

（2）熵权法下的维度权重计算

记 RFMN 模型所有维度的原始得分数据集为$OS=(os_{ij})_{m\times n}$，其中m为用户总数，n为维度个数，$os_{ij}(i=1,2,\ldots,m;\ j=1,2,\ldots,n;$本文中$n=4)$为第$i$个用户第 j 个维度原始得分。

计算在第 j 个维度下，第i个用户的维度原始得分所占比重 p_{ij}：

$$p_{ij}=os_{ij}\ /\sum_{i=1}^{m}os_{ij}. \tag{3.1}$$

记第 j 个维度的信息熵为 E_j：

$$E_j=-\alpha\sum_{i=1}^{m}p_{ij}\ln p_{ij},\text{其中}\alpha=\frac{1}{\ln m}. \tag{3.2}$$

因此，第 j 个维度的权重为$W_j^{''}$：

$$W_j^{''}=(1-E_j)/\sum_{j=1}^{n}(1-E_j). \tag{3.3}$$

（3）组合赋权法下的维度权重及用户价值综合得分计算

根据组合赋权法的定义，组合赋权权重的计算公式为$W=TW^{'}+UW^{''}$，在前文中已经给出了层次分析法和熵权法下的维度权重向量$W^{'}$、$W^{''}$，只需确定决策者对$\overline{T}$、$\overline{U}$ 的取值，再根据组合赋权权重的计算公式即可计算得到组合赋权权重向量W。

通过计算得到组合赋权权重向量W后，即可计算出第i个用户第 j 个维度的价值综合得分S_{ij}：

$$S_{ij}=OS_{ij}*W_j. \tag{3.4}$$

而用户的价值综合得分集记为S，$S=OS*W$。

3.3 用户特征选择及画像

如图 3.3 所示，本文的用户画像形成主要分为四个阶段：数据采集、数据分析与预处理、用户特征选择和用户画像成型。本节着重对用户特征选择与用户画像成型进行说明。

（1）用户特征选择

将上一阶段整理好的数据用于这一阶段的基于组合赋权法的 RFMN 模型的

实现、K-Means 聚类的实现以及 PCA 算法的实现。在此阶段，将通过基于组合赋权法的 RFMN 模型得到的各维度的权重得分进行 K-Means 聚类分析，得到最优聚类结果，得到样本用户的分类以及各类别用户的还款行为特征；对各类别用户实现 PCA 算法，得到各类别用户的基本属性和信用等级特征。

（2）用户画像成型

这一阶段是数据预处理与统计分析、用户特征选择这两个阶段结果的整合。最终得到样本用户的分类及特征描述，是提出营销建议的基础。

图 3.3 用户特征选择流程

3.4 本章小结

本章建立了基于组合赋权法的 RFMN 模型。根据本文的研究对象，对传统 RFM 模型的参数进行了合理的重新定义和维度增加，给出了适用于 RFMN 模型的用户分层准则。对如何利用组合赋权法计算用户综合价值得分进行了详细说明，详细阐述了用户特征选择流程，构造出了适合进行信用卡用户价值分析研究的模型，并给出了详细的模型构建步骤。

第 4 章 基于 RFMN 模型的信用卡用户价值分析

4.1 数据预处理

由于信用卡用户数据涉及到用户隐私，国内数据无法通过正规渠道获取到，又考虑到国内外信用卡用户数据的结构相似，本文对国外某银行信用卡用户数据进行研究和分析，得到的结果可以为我国的信用卡用户价值分析提供参考。

4.1.1 数据采集

为了验证信用卡用户价值模型（RFMN 模型），本文采用国外某银行信用卡用户数据（数据来源：https://www.lendingclub.com/info/prospectus.action）进行模型的应用和检验。用户数据在 2016 年 4 月 1 日-2018 年 7 月 1 日期间存在还款，数据导出日期为 2018 年 8 月 1 日，以此为分析截止时间点。由于数据涉及到用户隐私，本文仅展示部分数据和经过处理后的数据。本文随机抽取了上述时间内的 7 万个用户的信用卡使用情况数据进行研究，实现对信用卡用户价值分类、评价分析和营销建议的提出。

4.1.2 数据整理

由于原始数据集中信用卡用户指标过多，且大多数指标对本文的研究无意义，所以对原始数据进行整理。用于本文研究的信用卡用户基本信息、信用信息和还款信息的指标明细见表 4.1。

表 4.1 信用卡用户指标明细

指标	说明	指标	说明
member_id	用户 id	grade	信用等级
emp_title	职业	last_pymnt_d	上次还款时间
emp_length	就业年限	total_ pymnt_n	还款次数
home_ownership	房屋产权	total_ pymnt_amnt	还款总金额
annual_inc	年收入	delinq	逾期 30 天以上次数
loan_amnt	信用额度		

此外，由于本文研究中使用的样本数据的指标中存在用户信息为字符串格式或数据跨度过大的情况，在进行后续研究之前，需要将这些信息转换成分类数据，详细的字符串与分类数值的对应关系在附录 1 中展示，在正文中不再对其赘述。

4.1.3 数据预处理

本文在信用卡用户数据中提取出了建立基于组合赋权法的 RFMN 模型需要的 R、F、M、N 四个维度的数据，这四个维度的数据分别为：

R(Recency)：用户最近一次还款时间点与分析截止时间点间的间隔。

F(Frequency)：观测时间内的还款次数。

M(Monetary)：观测时间内的还款总金额。

N(Number)：观测时间内的逾期 30 天以上次数。

但要将这些数据引入到模型中进行下一步操作，需要先进行数据预处理。基于此，本文对原始数据做了以下预处理工作：

（1）剔除含缺省项的样本数据。剔除样本用户数据中有缺省项的数据，避免实证结果受缺省项的影响。

（2）剔除无效样本数据。根据数据统计分析结果可知，过去 2 年内逾期次数超过 12 次的用户以及现账户中存在逾期欠款且欠款数额较大的用户可以处理成银行的风险用户；根据用户信用等级可知，信用等级为 F、G 时，用户违约风险极高，故将信用等级为 F、G 的用户也直接处理成风险用户。因此，这部分用户不纳入本次需要进行聚类分析的数据中，并且应对其采取特殊的策略进行催收和后续营销。

其次，对于在观测时间内已经偿还完欠款超过 1 年的用户，说明其在偿还完欠款后无再次使用信用卡的情况，以及这类用户可以直接处理成已流失用户（即一般挽留用户或低价值用户），对于此类用户采取的营销方案也需要斟酌考虑，因此，这类用户也不纳入本次需要进行聚类分析的数据中。

（3）数据标准化处理。由于不同维度的变量的单位有所不同，可能造成后续的聚类结果不合理，因此各维度变量需要进行数据标准化处理。本文采用的数据标准化方法是 Z_score 标准化处理。

Z_score 标准化是将原始数据处理成符合标准正态分布的数据，是以原始数据的均值和标准差为切入点的数据标准化方法。Z_score 标准化的公式化表达如下属公式所示，其中，x 为具体数值，μ 为均值，σ 为标准差：

$$Z=(x-\mu)/\sigma$$

Z 为原始数据与该维度均值间的距离，若原始数据小于均值，Z 值为负，反之亦然。经过 Z_score 标准化后的数据展示如表 4.2 所示。

表 4.2 Z_score 标准化后的用户 RFMN 数据

member_id	R_Z	F_Z	M_Z	N_Z
U00001	-0.96	-2.73	-0.28	2.82
U00007	-0.96	-0.35	-0.42	-0.38
U00008	-0.10	-0.35	-0.03	-0.38
U00011	-0.10	-0.35	-0.71	-0.38
U00012	3.34	-0.35	0.12	-0.38

以上数据预处理保证了实证数据的准确性，可以进行下一阶段的研究分析。

4.2 数据描述统计

在将数据引入 RFMN 模型中之前，本文先进行了统计分析，并在这个阶段给出部分评判用户价值的意见。

4.2.1 用户基本信息

（1）用户就业年限

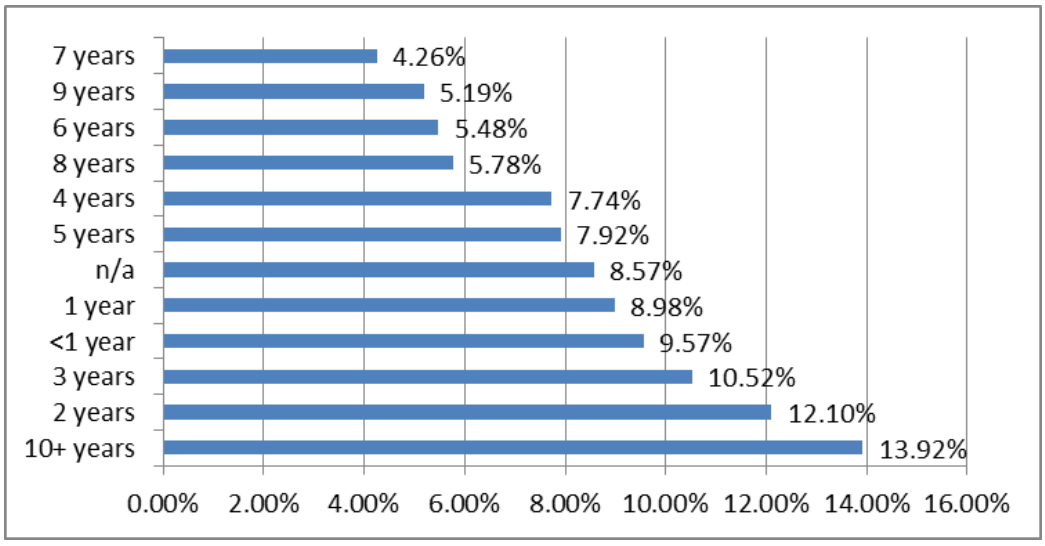

图 4.1 用户就业年限

如图 4.1 所示，还未就业、工作年限小于 1 年、工作 1-3 年以及工作年限超过 10 年的用户更容易产生信用卡账单。这样的局面出现，可能是因为未就业人群和刚就业人群对自我资金管理不到位，就业 10 年以上人群增加了车贷、房贷、子女教育支出等情况。这类人群大体上是属于银行的重要用户的。但是由于随着刚就业人群的工作年限的增加，他们对于信用卡的依赖程度会降低，这是银行在制定服务方案时需要考虑到的不可抗因素之一。

（2）用户房屋产权情况

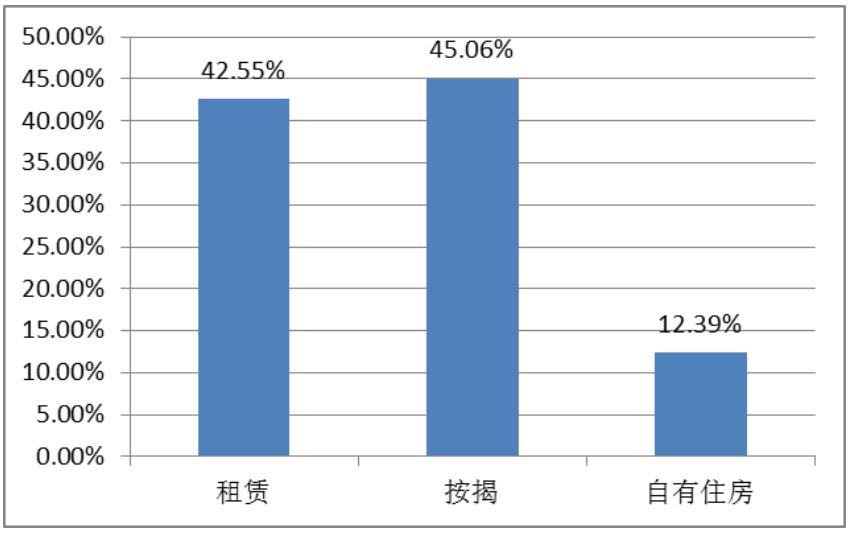

图 4.2 用户房屋产权情况

根据图 4.2 可知，这 7 万个存在信用卡使用的用户中，因为住房产生了财务支出的占了总样本的 87.61%（其中租赁 42.55%，按揭 45.06%）。这一组数据表明，住房产生的财务支出，使得个人或家庭的财政出现紧缺，因此用户会选择使用信用卡来维持日常开销或住房支出。这样的用户的信用卡支出款项大多是刚性消费，并且持续时间较长，属于重要用户类型，银行针对这一类型的用户应制定长期性的服务方案。

4.2.2 信用卡使用情况

（1）过去 2 年内用户逾期 30 天以上次数

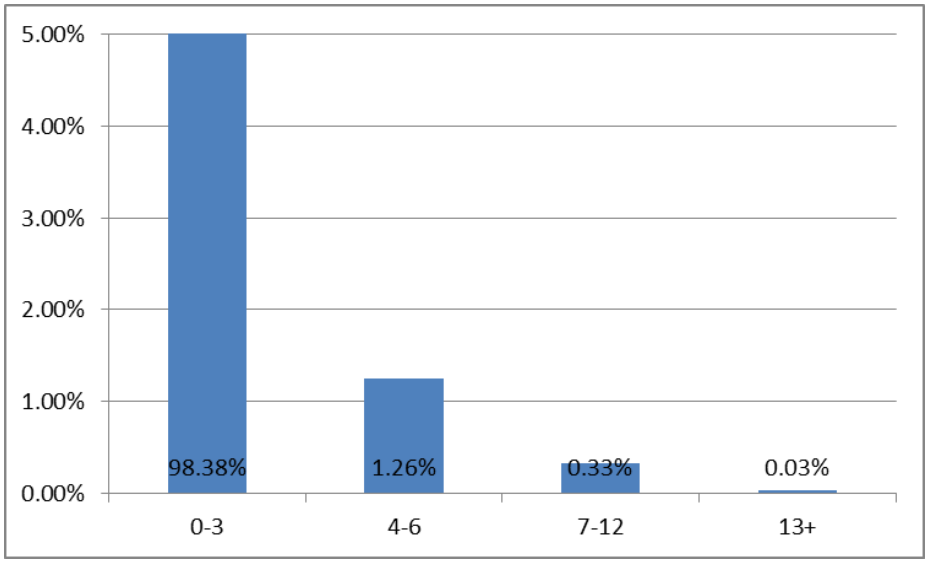

图 4.3 过去 2 年内用户逾期 30 天以上次数

如图4.3所示，7万样本用户中信用程度较高的（逾期次数≤3次）所占比例高达98.38%，说明绝大部分用户拥有自我约束力，同时也具备偿还欠款的能力，不会长期因为突发状况导致无法偿还欠款。另一方面，仍然存在小部分用户信用程度极差（逾期次数>12次）的情况，这部分用户长期无法偿还欠款，承担引发信用卡信用风险的主要责任，这部分用户可以处理成风险用户，不对其进行进一步的营销，以免造成更大的风险。

（2）用户现存逾期欠款情况

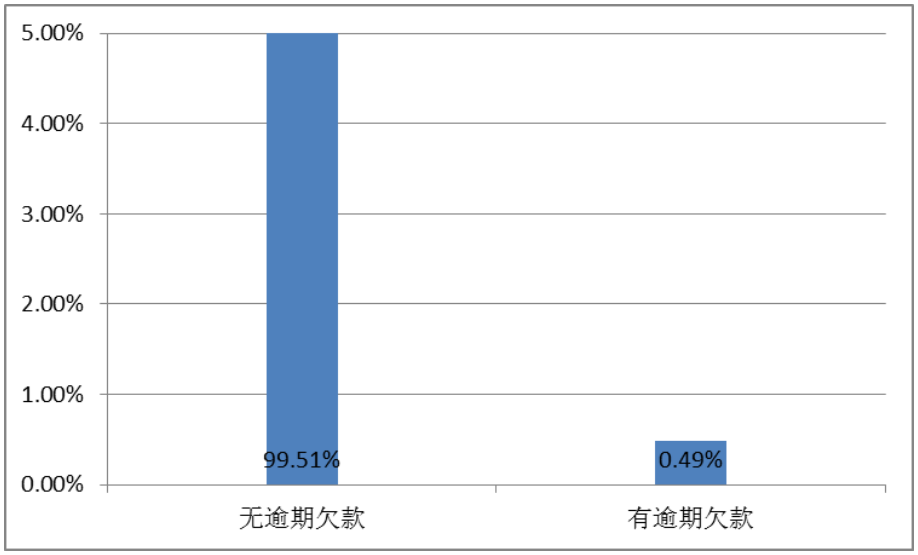

图4.4 用户现存逾期欠款情况

根据图4.4可知，样本用户中绝大多数用户（99.51%）不存在预期欠款，这部分用户的信用程度是极高的，属于银行的重要用户，但属于重要用户中的哪一类用户，还要结合其他信用卡使用情况进行判断。而0.49%的用户存在预期欠款的情况，通过查询原始数据，这部分用户中又存在逾期欠款数额较大的情况，因此，对于这部分用户，建议银行处理成风险用户，不对其进行进一步的营销。

4.2.3 评判用户价值

以上是对原始数据进行的数据统计分析，由于信用卡用户数据的款项较多，本文选取了具有代表性的四个方面进行统计分析，根据以上统计分析结果，本文建议将以下两种情况的用户处理成风险用户，不对其进行进一步的信用卡营销方案实施，以免造成更大风险。

（1）过去2年内逾期次数超过12次的用户；

（2）现账户中存在逾期欠款且欠款数额较大的用户。

4.3 维度权重确定与得分计算

4.3.1 模型维度值的分箱处理

本文采用五分法（0.2、0.4、0.6、0.8）对经过预处理后 R、F、M、N 数据进行分箱处理，区间设置如表 4.3 所示。通过表 4.3 的处理，即可计算得到各信用卡用户的 R、F、M、N 的原始得分，但需要按以下规则进行原始得分设置：

R：Z 值越小，上次还款距离截止时间点越近，得分越高，最高区间 5 分，最低区间 1 分。

F：Z 值越大，观测时间内还款次数越多，得分越高，最高区间 5 分，最低区间 1 分。

M：Z 值越大，观测时间内还款总额越大，得分越高，最高区间 5 分，最低区间 1 分。

N：Z 值越小，观测时间内的逾期 30 天以上次数越少，得分越高，最高区间 5 分，最低区间 1 分。

表 4.3 RFMN 模型维度分箱处理办法

分值	R	F	M	N
1	[2.48,3.44]	[-9.87,-7.65）	[-1.55,-0.33）	[9.87,12.43]
2	[1.62,2.48）	[-7.65,-5.43）	[-0.33,0.90）	[7.31,9.87）
3	[0.76,1.62）	[-5.43,-3.21）	[0.90,2.13）	[4.75,7.31）
4	[-0.10,0.76）	[-3.21,-0.99）	[2.13,3.35）	[2.18,4.75）
5	[-0.96,-0.10）	[-0.99,1.23]	[3.35,4.58]	[-0.38,2.18）

4.3.2 维度权重计算

为了验证本文建立的基于组合赋权法的 RFMN 模型能够得到数据的最优聚类，将对基于组合赋权法、层次分析法、熵权法的 RFMN 模型分别进行 K-Means 聚类分析。基于此，本文给出了层次分析法、熵权法、组合赋权法的维度权重计算结果和基于层次分析法、熵权法、组合赋权法 RFMN 模型的用户价值评价结果。

（1）层次分析法计算维度权重

假定决策者对维度进行两两比较，得到以判断矩阵，将判断矩阵录入到 SPSSAU 分析软件中进行层次分析法分析，并进行一致性检验分析，若未通过一致性检验，需要对决策者给出的判断矩阵进行调整，直到通过一致性检验为止。在通过检验后还需检查判断矩阵是否存在逻辑问题，均无问题后即可得出信用卡

用户的R、F、M、N四个维度通过层次分析法所确定的权重向量$W^{'}$。通过一致性检验、逻辑无问题的决策者判断矩阵如下：

$$\begin{bmatrix} 1 & 1 & 3 & 1/3 \\ 1 & 1 & 3 & 1/3 \\ 1/3 & 1/3 & 1 & 1/3 \\ 3 & 3 & 3 & 1 \end{bmatrix}$$

将这个判断矩阵输入到SPSSAU软件中得到如下结果：

表4.4 层次分析结果

项	特征向量	权重值	最大特征值
R	0.842	0.211	4.155
F	0.842	0.211	
M	0.392	0.098	
N	1.925	0.480	
CI值	RI值	CR值	一致性检验结果
0.052	0.890	0.058	通过

其中，CI=(最大特征根-n)/(n-1)，RI值为n阶对应的随机一致性RI取值，CR=CI/RI。CR值小于0.1，则判断矩阵通过一致性检验，反之则不通过。根据上述层次分析结果得到权重向量：

$$W^{'}=(0.211,0.211,0.098,0.480)。$$

（2）熵权法计算维度权重

根据3.2节中提到的熵权法计算得到R、F、M、N四个维度原始得分的权重向量$W^{''}$。首先要根据公式确定每个维度的信息熵E_j，再根据权重计算公式：

$$W_j^{''}=(1-E_j)/\sum_{j=1}^{n}(1-E_j)$$

计算得到R、F、M、N四个维度的原始得分通过熵权法计算得到的维度权重向量：

$$W^{''}=(0.278,0.024,0.685,0.013)。$$

（3）组合赋权法计算维度权重及得分

层次分析法和熵权法的重要程度取值$\overline{T}$、$\overline{U}$可以根据2.2节中提到的公式：

$$\overline{T}=\sum_{i=1}^{m}\sum_{j=1}^{n}b_{ij}W_j^{'}/\sum_{i=1}^{m}\sum_{j=1}^{n}b_{ij}(W_j^{'}+W_j^{''})$$

$$\overline{U}=\sum_{i=1}^{m}\sum_{j=1}^{n}b_{ij}W_j^{''}/\sum_{i=1}^{m}\sum_{j=1}^{n}b_{ij}(W_j^{'}+W_j^{''})$$

计算得到。通过计算，样本数据的$\overline{T}$、$\overline{U}$取值为$\overline{T}$=0.56，$\overline{U}$=0.44。

再根据组合赋权法的权重计算公式$W=TW^{'}+UW^{''}$，计算出组合赋权权重向量W：

$$W=(0.241,0.128,0.357,0.274)。$$

用权重向量乘以对应维度的原始得分，得到基于组合赋权法RFMN得分，得分结果展示如表4.5。

表4.5 基于组合赋权法RFMN模型的用户价值评价结果（部分）

member_id	R_S	F_S	M_S	N_S	RS	FS	MS	NS	S
U00001	5	4	2	4	1.207	0.514	0.714	1.095	3.531
U00007	5	5	1	5	1.207	0.642	0.357	1.369	3.576
U00008	4	5	2	5	0.966	0.642	0.714	1.369	3.691
……	……	……	……	……	……	……	……	……	……
U69997	5	5	1	5	1.207	0.642	0.357	1.369	3.576
U69998	4	5	1	5	0.966	0.642	0.357	1.369	3.334

R_S、F_S、M_S、N_S分别表示经过分箱处理后的还款间隔得分、还款次数得分、还款总金额得分和逾期30天以上次数得分；RS、FS、MS、NS为各维度原始得分与组合权重的乘积；而S为信用卡用户价值综合得分，计算方法遵从以下公式：

$$S=RS+FS+MS+NS.$$

本文还将经过分箱处理后的还款间隔得分、还款次数得分、还款总金额得分和逾期30天以上次数得分R_S、F_S、M_S、N_S分别与通过计算得到的层次分析法权重和熵权法权重进行乘积运算，得到了$RS1$、$FS1$、$MS1$、$NS1$和$RS2$、$FS2$、$MS2$、$NS2$，以及基于层次分析法、熵权法的信用卡用户价值综合得分$S1$、$S2$。得分结果展示如表4.6。

表4.6 基于层次分析法、熵权法RFMN模型的用户价值评价结果（部分）

member_id	RS1	FS1	MS1	NS1	S1	RS2	FS2	MS2	MS2	S2
U00001	1.055	0.844	0.196	1.920	4.015	1.400	0.096	1.370	0.052	2.918
U00007	1.055	1.055	0.098	2.400	4.608	1.400	0.120	0.685	0.065	2.270
U00008	0.844	1.055	0.196	2.400	4.495	1.120	0.120	1.370	0.065	2.675
……	……	……	……	……	……	……	……	……	……	……
U69997	1.055	1.055	0.098	2.400	4.608	1.400	0.120	0.685	0.065	2.270
U69998	0.844	1.055	0.098	2.400	4.397	1.120	0.120	0.685	0.065	1.990

4.4 本章小结

本章是模型的应用部分。在本章节中，完成了用户信息与分类数据间的转换，为PCA算法实现用户特征选择提供了数据支撑。对原始数据进行了数据统计分析，得到了用户为风险用户的两种情况。完成了数据清洗、数据预处理工作，同样得到了相关结论。制定了维度值分箱办法，得出了用户各维度原始得分。确定了层次分析法、熵值法、组合赋权法的权重向量$W^{'}$、$W^{''}$、W，计算出基于各权重的用户权重得分并给出了信用卡用户价值综合得分。为K-Means聚类分析提供了数据支持。

第 5 章 用户分类结果与价值分析

5.1 基于 K-Means 聚类算法的用户价值分层结果

为了验证本文建立的基于组合赋权法的 RFMN 模型能够得到数据的最优聚类，本文将基于组合赋权法、层次分析法、熵权法的 RFMN 模型的三组用户价值评价结果（*RS*、*FS*、*MS*、*NS*、*RS*1、*FS*1、*MS*1、*NS*1 和 *RS*2、*FS*2、*MS*2、*NS*2）分别作为聚类分析指标，采用 R 语言进行代码编写（详细代码见附录 2），并使用 R 软件实现 K-Means 聚类模型。通过 Elbow Method 得到各聚类数量群集内的总平方和曲线，观察曲线即可获得最优的聚类数量。根据最优聚类数量得到最优 K-Means 聚类模型，得出分类，并对分类用户进行定义。

Elbow Method 得到最优聚类数量的核心思想是：聚类数量 k 越大，样本划分越细，每一个类的聚合程度也会越高，随之而来的是误差平方和（SSE）越小。当聚类数量 k 小于最优聚类数量时，随着 k 的增大，类的聚合程度会陡然增大，SSE 的下降幅度也极大，而当聚类数量 k 大于最优聚类数量时，随着 k 的增大，类的聚合程度会趋于平缓，SSE 的下降幅度也逐渐平缓。这样的变化趋势形似手肘，曲线的拐点（肘部）即为最优聚类数量。

5.1.1 基于组合权重的用户价值分层结果

运行 R 语言代码，通过 Elbow Method 得到各聚类数量群集内的总平方和曲线如图 5.1 所示。曲线中拐点（肘部）的位置为 5，因此，当聚类数量为 5 时，K-Means 聚类模型为最优聚类模型。

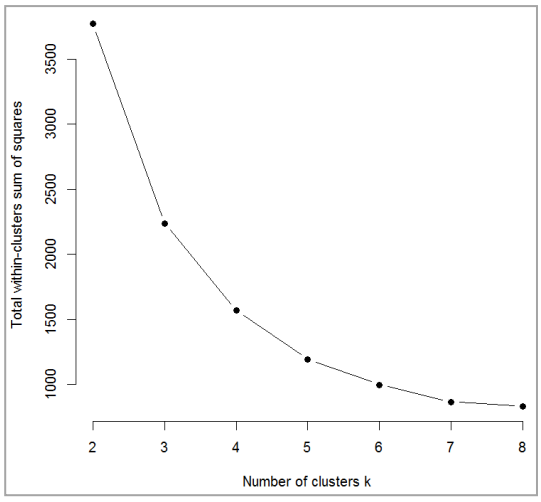

图 5.1 群集内总平方和曲线

根据 K-Means 聚类结果展示，样本信用卡用户可划分成 5 类。表 5.1 展示了类中心、样本分布情况。给出了该分类结果下的各类用户还款情况特征。

表 5.1 聚类结果

类别	中心值								样本个数	所占比重	用户定义
	RS	次序	FS	次序	MS	次序	NS	次序			
1	0.947	4	0.632	1	1.497	1	1.365	1	1578	4.36%	重要发展用户
2	1.025	2	0.625	3	0.714	3	1.359	3	12011	33.21%	重要保持用户
3	0.307	5	0.624	4	0.556	4	1.356	4	3594	9.94%	一般挽留用户
4	1.026	1	0.620	5	0.357	5	1.352	5	14814	40.97%	一般保持用户
5	1.004	3	0.628	2	1.072	2	1.362	2	4165	11.52%	重要价值用户

（1）第 1 类用户的 *RS* 中心值排名第四，*FS* 中心值最大，*MS* 中心值最大，*NS* 中心值最大，说明此类用户最近无还款行为，但历史还款次数多，历史还款总金额高，历史逾期次数少，可将其定义为重要发展客户。

（2）第 2 类用户的 *RS* 中心值排名第二，*FS* 中心值排名第三，*MS* 中心值排名第三，*NS* 中心值排名第三，说明这类用户最近有还款行为，历史还款次数较多，但历史还款总金额较高，历史逾期次数较少，可将其定义为重要保持用户。

（3）第 3 类用户的 *RS* 中心值最小，*FS* 中心值排名第四，*MS* 中心值排名第四，*NS* 中心值排名第四，说明此类用户最近无还款行为，历史还款次数少，但历史还款总金额低，历史逾期次数多，可将其定义为一般挽留用户（低价值用户）。

（4）第 4 类用户的 *RS* 中心值最大，*FS* 中心值最小，*MS* 中心值最小，*NS* 中心值最小，说明此类用户最近有还款行为，但历史还款次数少，历史还款总金额低，历史逾期次数少，可将其定义为一般保持用户。

（5）第 5 类用户的 *RS* 中心值排名第三，*FS* 中心值排名第二，*MS* 中心值排名第二，*NS* 中心值排名第二，说明此类用户最近有还款行为，历史还款次数多，历史还款总金额高，历史逾期次数少，可将其定义为重要价值用户。

5.1.2 基于层次分析法权重的用户价值分层结果

运行 R 语言代码，通过 Elbow Method 得到各聚类数量群集内的总平方和曲线如图 5.2 所示。曲线中拐点（膝盖）的位置为 6，因此，当聚类数量为 6 时，K-Means 聚类模型为最优聚类模型。

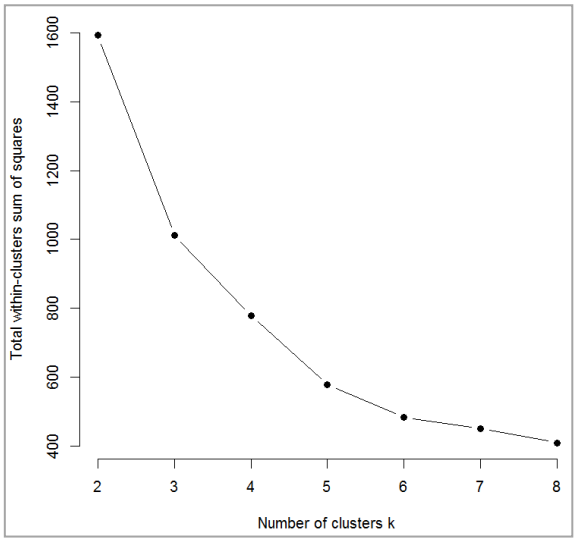

图 5.2 群集内总平方和曲线

根据 K-Means 聚类结果展示，样本信用卡用户可划分成 6 类。表 5.2 展示了类中心、样本分布情况。给出了该分类结果下的各类用户还款情况特征。

表 5.2 聚类结果

类别	中心值								样本个数	所占比重	用户定义
	RS1	次序	FS1	次序	MS1	次序	NS1	次序			
1	0.856	3	0.638	5	0.142	5	1.440	4	239	0.66%	一般保持用户
2	0.880	2	0.812	4	0.157	4	1.920	3	784	2.17%	重要挽留用户
3	0.833	4	1.032	2	0.173	1	2.400	1	22370	61.86%	重要发展用户
4	0.816	5	0.396	6	0.121	6	0.779	5	98	0.27%	一般挽留用户
5	0.274	6	1.031	3	0.170	3	2.393	2	3848	1064%	重要保持用户
6	1.055	1	1.039	1	0.172	2	2.400	1	8823	24.40%	重要价值用户

（1）第 1 类用户的 *RS*1 中心值排名第三，*FS*1 中心值排名第五，*MS*1 中心值排名第五，*NS*1 中心值排名第四，说明此类用户最近有还款行为，但历史还款次数少，历史还款总金额低，历史逾期次数较多，可将其定义为一般保持用户。

（2）第 2 类用户的 *RS*1 中心值排名第二，*FS*1 中心值排名第四，*MS*1 中心值排名第四，*NS*1 中心值排名第三，说明此类用户最近有还款行为，虽然历史还款次数较少，历史还款总金额较低，但历史逾期次数较少，可定义为重要挽留用户。

（3）第 3 类用用户的 *RS*1 中心值排名第四，*FS*1 中心值排名第二，*MS*1 中心值最大，*NS*1 中心值最大，说明此类用户最近无还款行为，但历史还款次数较多，历史还款总金额高，历史逾期次数少，可将其定义为重要发展用户。

（4）第 4 类用户的 *RS*1 中心值排名第五，*FS*1 中心值最小，*MS*1 中心值最小，*NS*1 中心值排名第五，说明这类用户最近无还款行为，历史还款次数少，历史还

款总金额低，历史逾期次数多，可将其定义为一般挽留用户（低价值用户）。

（5）第 5 类用户的 *RS*1 中心值最小，*FS*1 中心值排名第三，*MS*1 中心值排名第三，*NS*1 中心值排名第二，说明此类用户最近无还款行为，但历史还款次数较多，历史还款总金额较高，历史逾期次数少，可将其定义为重要保持用户。

（6）第 6 类用户的 *RS*1 中心值最大，*FS*1 中心值最大，*MS*1 中心值排名第二，*NS*1 中心值最大，说明此类用户最近有还款行为，且历史还款次数多，历史还款总金额高，历史逾期次数少，可将其定义为重要价值用户。

5.1.3 基于熵权法权重的用户价值分层结果

运行 R 语言代码，通过 Elbow Method 得到各聚类数量群集内的总平方和曲线如图 5.3 所示。曲线中拐点（膝盖）的位置为 5，因此，当聚类数量为 5 时，K-Means 聚类模型为最优聚类模型。

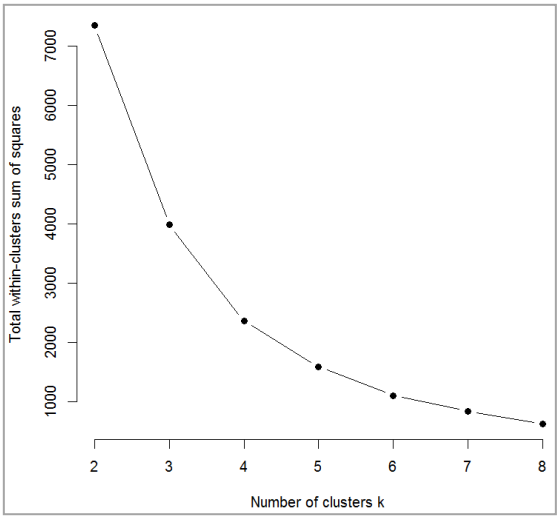

图 5.3 群集内总平方和曲线

根据 K-Means 聚类结果展示，样本信用卡用户可划分成 5 类。表 5.3 展示了类中心、样本分布情况。给出了该分类结果下的各类用户还款情况特征。

表 5.3 聚类结果

类别	中心值								样本个数	所占比重	用户定义
	RS2	次序	FS2	次序	MS2	次序	NS2	次序			
1	1.188	2	0.117	3	1.370	3	0.0645	3	12011	33.21%	重要价值用户
2	1.189	1	0.116	4	0.685	5	0.0642	5	14814	40.97%	一般保持用户
3	0.363	5	0.117	3	0.982	4	0.0644	4	3307	9.14%	一般发展用户
4	1.098	4	0.119	1	2.870	1	0.0648	1	1578	4.36%	重要发展用户
5	1.107	3	0.118	2	2.055	2	0.0646	2	4452	12.31%	重要保持用户

（1）第 1 类用户的 *RS*2 中心值排名第二，*FS*2 中心值排名第三，*MS*2 中心

值排名第三，$NS2$ 中心值排名第三，说明此类用户最近有还款行为，历史还款次数较多，历史还款总金额较高，历史逾期次数较少，可将其定义为重要价值用户。

（2）第 2 类用户的 $RS2$ 中心值最大，$FS2$ 中心值排名第四，$MS2$ 中心值最小，$NS2$ 中心值最小，说明此类用户最近有还款行为，但历史还款次数较少，历史还款总金额低，历史逾期次数少，可将其定义为一般保持用户。

（3）第 3 类用户的 $RS2$ 中心值最小，$FS2$ 中心值排名第三，$MS2$ 中心值排名第四，$NS2$ 中心值排名第四，说明这类用户最近无还款行为，虽然历史还款次数较多，但历史还款总金额较低，历史逾期次数较多，可将其定义为一般发展用户。

（4）第 4 类用户的 $RS2$ 中心值排名第四，$FS2$ 中心值最大，$MS2$ 中心值最大，$NS2$ 中心值最大，说明此类用户最近无还款行为，但历史还款次数多，历史还款总金额高，历史逾期次数少，可将其定义为重要发展用户。

（5）第 5 类用户的 $RS2$ 中心值排名第三，$FS2$ 中心值排名第二，$MS2$ 中心值排名第二，$NS2$ 中心值排名第二，说明此类用户最近有还款行为，历史还款次数较多，但历史还款总金额较高，历史逾期次数较少，可将其定义为重要保持用户。

5.2 结果比较

本文将基于组合赋权法、层次分析法、熵权法的 RFMN 模型的三组用户价值评价结果（RS、FS、MS、NS、$RS1$、$FS1$、$MS1$、$NS1$ 和 $RS2$、$FS2$、$MS2$、$NS2$）分别作为聚类分析指标进行了 K-Means 聚类分析，现对这三组聚类结果进行比较。

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

$$MSA=\frac{\sum_{i=1}^{I}n_i(\overline{S_i}-\overline{\overline{S}})^2}{I-1};\ \overline{S_i}=\frac{1}{n_i}\sum_{j=1}^{n_i}S_{ij}(i=1,2,\ldots,I);\ \overline{\overline{S}}=\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}S_{ij}(i=1,2,\ldots,I)$$

其中，I 为 K-Means 聚类模型的最优聚类数量，n 为样本个数，n_i 是第 i 个类别的样本量，$\overline{S_i}$ 是第 i 个类别的样本平均值，$\overline{\overline{S}}$ 代表所有样本的平均值。5.1 节中建立的三个最优聚类模型的 MSA 计算过程值如表 5.4 所示。

表 5.4 三个聚类模型 *MSA* 的计算结果

	I	$\bar{\bar{S}}$	$\sum n_i(\bar{S}_i-\bar{\bar{S}})^2$	*MSA*
组合赋权法	5	3.556	128588.90	32147.23
层次分析法	6	4.404	159243.40	31848.68
熵权法	5	2.484	89829.13	22457.28

经过计算可知，组合赋权法的聚类结果中不同用户类型之间的组间均方更大，说明基于组合赋权法的 RFMN 模型的聚类结果中各个用户群体间的差异程度更大，区分度更高，对信用卡用户价值的划分效果更好。最终选择基于组合赋权法的 RFMN 模型的聚类结果作为最优聚类结果。聚类结果及用户还款情况特征如表 5.5 所示。

表 5.5 最优聚类结果

类别	样本个数	所占比重	用户定义	还款情况特征
1	1578	4.36%	重要发展用户	最近无还款行为，历史还款次数多，历史还款总金额高，历史逾期次数少
2	12011	33.21%	重要保持用户	最近有还款行为，历史还款次数多，历史还款总金额高，历史逾期次数少
3	3594	9.94%	低价值用户	最近无还款行为，历史还款次数少，历史还款总金额低，历史逾期次数多
4	14814	40.97%	一般保持用户	最近有还款行为，历史还款次数少，历史还款总金额低，历史逾期次数少
5	4165	11.52%	重要价值用户	最近有还款行为，历史还款次数多，历史还款总金额高，历史逾期次数少

5.3 分层价值用户的特征选择

在上一节中得到了用户价值分层，在此基础上，利用 PCA 算法对五个类别的用户分别进行特征选择，此过程通过 SPSS 软件实现。研究发现前三个主成分能够解释该类用户 60%-70%的原始数据。结合因子得分系数，得到该类用户的特征。

表 5.6 因子得分系数

用户类别	职业	就业年限	房屋所有权	年收入	信用额度	信用等级
第一类	0.466	0.161	0.385	0.161	0.485	0.573
第二类	0.276	0.128	0.309	0.066	0.379	0.793
第三类	0.117	0.052	0.200	0.026	0.333	0.902
第四类	0.267	0.189	0.230	0.043	0.364	0.828
第五类	0.23	0.208	0.158	0.037	0.353	0.872

表 5.6 给出了各类用户的因子得分系数，本文将因子得分系数大于等于 0.2 的指标作为该类用户的特征，总结各类用户特征如下：

第一类用户（重要发展用户）：职业、房屋所有权、信用额度、信用等级。

第二类用户（重要保持用户）：职业、房屋所有权、信用额度、信用等级。

第三类用户（低价值用户）：房屋所有权、信用额度、信用等级。

第四类用户（一般保持用户）：职业、房屋所有权、信用额度、信用等级。

第五类用户（重要价值用户）：职业、就业年限、信用额度、信用等级。

接下来，将结合聚类结果、PCA 主成分分析结果、数据统计分析论和数据预处理结论，得到本文样本用户的价值分类和具体用户画像，并给出营销建议。

5.4 用户价值分类与用户画像

结合本文基于组合赋权法的 RFMN 模型的聚类结果、PCA 主成分分析结果、数据统计分析论和数据预处理结论，最终将本文的样本用户分为以下六类，给出每类用户的用户画像并提出营销建议。

第一类，重要价值用户。用户主要特征为：

（1）职业多为企业普通职员；

（2）工作年限在 1-6 年居多；

（3）信用额度在 5000-20000 居多；

（4）信用等级较好，以 B、C 等级为主；

（5）用户的还款情况大致为最近有还款行为，历史还款次数多，历史还款总金额高，历史逾期次数少。

针对这类用户的营销建议是：需要重点维护这类用户，挖掘其主要消费领域，提出针对性的营销方案，提升用户的忠诚度。

第二类，重要保持用户。用户主要特征为：

（1）职业多为企业普通职员；

（2）绝大多数用户无全款住房，为按揭房或租赁房；

（3）信用额度在 5000-20000 居多；

（4）信用等级较好，以 B、C 等级为主；

（5）用户的还款情况大致为最近有还款行为，历史还款次数多，历史还款总金额高，历史逾期次数少。

针对这类用户的营销建议是：这类用户的消费水平较高，但消费行为不够稳

定，需要激发用户消费欲望，促使转化为优质用户。

第三类，重要发展用户。用户主要特征为：

（1）职业多为企业普通职员；

（2）绝大多数用户无全款住房，为按揭房或租赁房；

（3）信用额度在 5000-30000 居多；

（4）信用等级较好，以 B、C 等级为主；

（5）用户的还款情况大致为最近无还款行为，历史还款次数多，历史还款总金额高，历史逾期次数少。

针对这类用户的营销建议是：这类用户存在失活可能，银行需对此类用户采取促活的营销策略，激发用户活跃度和消费欲望。

第四类，一般保持用户。用户主要特征为：

（1）职业多为企业普通职员或私人经营；

（2）绝大多数用户无全款住房，为按揭房或租赁房；

（3）信用额度在 5000-30000 居多；

（4）信用等级较好，以 B、C 等级为主；

（5）用户的还款情况大致为最近有还款行为，历史还款次数少，历史还款总金额低，历史逾期次数少。

针对这类用户的营销建议是：对此类用户需加强营销，但同时还应关注其还款情况，提防其发展为风险用户。

第五类，低价值用户。分为两种情况：

情况一，用户主要特征为：

（1）职业多为企业普通职员或私人经营；

（2）绝大多数用户无全款住房，为按揭房或租赁房；

（3）信用额度在 5000-20000 居多；

（4）信用等级较好，以 B、C 等级为主；

（5）用户的还款情况大致为最近无还款行为，历史还款次数少，历史还款总金额低，历史逾期次数多。

情况二，用户特征为在观测时间内已经偿还完欠款超过 1 年并再无使用。

针对这类用户的营销建议是：挖掘其主要消费领域，提出针对性的营销方案，激发这类用户的消费欲望，促进用户的信用卡使用。

第六类，风险用户。分三种情况：

情况一，过去 2 年内逾期次数超过 12 次。

情况二，现存账户中存在逾期欠款且欠款数额较大。

情况三，用户信用等级为 F、G。

针对这类用户的营销建议是:停止营销手段的实施,并采取必要的催收手段。在结束催收后将其信用等级划入最低等级，10 年（或更长时间）内不再对其开放信用卡服务。

5.5 本章小结

本章模型应用的结论部分。得到了基于组合赋权法、层次分析法、熵权法的 RFMN 模型的 K-Means 聚类结果，借鉴单因素方差分析思想，比较了三种聚类结果的组间均方 MSA，证实了基于组合赋权法的 RFMN 模型能够获得更优的聚类结果，并得到了用户分类。对分类用户做 PCA 主成分分析，参考因子得分系数，得到每类用户的特征，将 K-Means 聚类结果、PCA 主成分分析结果、数据统计分析论和数据预处理结论相结合，得到样本用户最终分类，给出每类用户的用户画像并提出营销建议。

第 6 章 结论与展望

6.1 论文总结

本文在比较了主观赋权法和客观赋权法的优劣势之后，放弃单独使用主观赋权法或客观赋权法，选择将集合了主、客观赋权法优势的组合赋权法引入到传统的 RFM 模型中，并对原始模型的维度进行重新定义和增加新维度的操作，建立了基于组合赋权法的 RFMN 模型，将传统的 RFM 模型改造成适用于测算信用卡用户价值的模型。

为了验证本文建立的基于组合赋权法的 RFMN 模型能够得到数据的最优聚类，本文对于组合赋权法、层次分析法、熵值法的 RFMN 模型分别进行了 K-Means 聚类分析，得到了对应的聚类结果，借鉴了单因素方差分析思想，采用组间均方评价不同类别信用卡用户综合价值得分的组间差异，最终得到两个模型的聚类结果中不同用户类型之间的组间均方分别为 32147.23、31848.68、22457.28，说明基于组合赋权法的 RFMN 模型的聚类结果中各个用户群体间的差异程度更大，区分度更高，对信用卡用户价值的划分效果更好。将基于组合赋权法的 RFMN 模型的聚类结果作为本文样本数据最优聚类结果的呈现，得到用户分类。对分类用户做 PCA 主成分分析，参考因子得分系数，得到每类用户的特征。综合考虑 K-Means 聚类结果、PCA 主成分分析结果、数据统计分析结论和数据预处理结论，将样本用户分为六类，分别是重要价值用户、重要保持用户、重要发展用户、一般保持用户、低价值用户和风险用户，给出用户画像并提出营销建议。

6.2 研究展望

本文的信用卡用户数据样本不够大、设备等级不高是本文的缺点。由于本文使用的数据涉及到用户隐私，笔者无法获取到我国的信用卡用户数据，这是本文的遗憾所在。

但在模型的应用阶段，本文所建立的基于组合赋权法的 RFMN 模型能够获得更优的聚类结果得到了证实。并且，本文建立的基于组合赋权法 RFMN 模型的用户价值评价体系以及模型应用阶段得到的结论，可以为国内的信用卡用户价值评估提供重大参考价值，得到的用户价值分类和用户画像也可以为国内信用卡用户提供参考。因此，本文的研究具有现实指导意义。同时，笔者认为，如果能够获取到我国的信用卡用户数据，本文的用户分类和用户画像将更具有代表性。

参考文献

[1] 楼芳.我国银行信用卡业务的风险管理[J].上海金融,2004(2):60-61.

[2] 弋涛.信用卡风险管理研究[D].西南财经大学,2006.

[3] 魏鹏.当前信用卡风险管理中存在的主要问题与对策[J].西南金融,2007(6):61-62.

[4] 徐悦,管国强.我国商业银行信用卡风险成因及其风险管理措施分析[J].中国物价,2013(4):70-73.

[5] 马宇盟.中国银行 S 分行信用卡风险管理优化研究[D].山东大学,2019.

[6] 汤虹.浅议信用卡业务风险特征及防控—以中国建设银行为例[J].现代企业,2020(5):83-84.

[7] 鲁齐.基于改进的 LLM 模型的信用卡欺诈交易识别研究[D].南京财经大学,2019.

[8] Almhaithawi D, Jafar A, Aljnidi M. Example-dependent cost-sensitive credit cards fraud detection using SMOTE and Bayes minimum risk[J]. SN Applied ences, 2020, 2(9).

[9] 马海英.基于神经网络及 Logistic 回归的混合信用卡评分模型[J].华东理工大学学报(社会科学版),2008(2):49-52.

[10] 陈为民.基于支持向量机的信用卡信用风险管理模型与技术研究[D].湖南大学,2009.

[11] 方匡南,吴见彬,朱建平,谢邦昌.信贷信息不对称下的信用卡信用风险研究[J].经济研究,2010,45(S1):97-107.

[12] 闫思羽.基于随机森林分类的信用卡逾期行为研究[D].云南财经大学,2018.

[13] 杨怡滨.基于 TPOT 的信用卡逾期识别算法[J].企业科技与发展,2020(3):92-94+97.

[14] 廖欣婷,谢磊.基于 Probit 与 Logistics 模型对比的信用卡逾期风险评估实证研究[J].市场论坛,2020(6):73-77.

[15] Fonseca D P, Wanke P F, Correa H L. A two-stage fuzzy neural approach for credit risk assessment in a Brazilian credit card company[J]. Applied Soft Computing, 2020, 92:106329.

[16] 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.

[17] 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 and Fuzzy Systems, 2018, 35:1-11.

[18] Sahar Tahanisaz, Sajjad shokuhyar. Evaluation of passenger satisfaction with service quality: A consecutive method applied to the airline industry[J]. Journal of Air Transport Management, 2020.

[19] 王威娜.基于模糊聚类和 RFM 模型的商场会员画像描绘方法[J].吉林化工学院学

报,2019,36(11):71-73.

[20] 聂子临.基于改进 RFM 模型的 D 公司理财产品用户挖掘研究[D].北京交通大学,2019.

[21] P. Anitha; Malini M. Patil RFM model for customer purchase behavior using K-Means algorithm[J]. Journal of King Saud University: Computer and Information Sciences, 2019.

[22] 卓灵,孙昕.一种基于改进 RFM 模型的数字集群用户分类方法[J/OL].计算机应用研究,2019.

[23] 鲁燚,张国发,刘茂.RFM 改进模型下的商场会员价值分析[J].广西质量监督导报,2020(6):226-227.

[24] 陈子璐. 基于 RFM 模型的电子商务客户细分[J]. 市场周刊, 2020(4):56-58.

[25] 刘小红, 郑茵妮. 基于 RFM 模型的品牌服装会员价格容忍度和忠诚度的关联[J]. 服装学报, 2020, 5(4):372-376.

[26] 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.

[27] 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.

[28] 王一宾,黄志强,程玉胜.基于 K-means 的 GLOCAL 改进算法[J].安庆师范大学学报(自然科学版),2020,26(2):55-62.

[29] 韩帅,孙乐平,杨艺云,吴宛潞,郭小璇,戴承承.基于改进 K-Means 聚类和误差反馈的数据清洗方法[J].电网与清洁能源,2020,36(7):9-15.

[30] 汤深伟.基于改进粒子群的 K-means 聚类算法及其在推荐系统中的应用[D].安徽大学,2020.

[31] 范智浩,吕东澔,王晓峰.基于 K-means 聚类算法的肝脏肿瘤分割[J].电脑知识与技术,2020,16(19):165-167.

[32] Poomagal S, Malar B, Hassan J I, et al. A novel Tag Score(T_S) model with improved K-means for clustering tweets[J]. Sadhana, 2020, 45(1).

[33] 樊治平,赵萱.多属性决策中权重确定的主客观赋权法[J].决策与决策支持系统,1997(04):89-93.

[34] Karayalcin I. The analytic hierarchy process: Planning, priority setting, resource allocation : Thomas L. SAATY McGraw-Hill, New York, 1980, xiii + 287 pages, 15.65[J]. European Journal of Operational Research, 1982, 9(1):97-98.

附 录 1

需要进行处理的国外某银行信用卡用户部分数据指标及选项为：职业{企业普通职员，企业管理层，私人经营，农牧渔劳动者，自由职业，在校学生，其他}；就业年限（<1 年，1 年，2 年，3 年，4 年，5 年，6 年，7 年，8 年，9 年，10 年及以上）；房屋所有权(自有住房，按揭，租赁)；年收入（5000 及以上）；信用额度（1000-40000）；信用等级（A、B、C、D、E、F、G，其中 F、G 两个等级的用户在预处理中处理成风险用户）。

对国外某银行信用卡用户部分数据做如下处理：职业{企业普通职员，企业管理层，私人经营，农牧渔劳动者，自由职业，在校学生，其他}；就业年限{<1 年，1-3 年，4-6 年，7-9 年，10 年及以上}；房屋所有权{自有住房，按揭，租赁}；年收入{5000-10000，10000-50000，50000-100000，100000-150000，150000 及以上}；信用额度{1000-5000，5000-10000，10000-20000，20000-30000，30000-40000}；信用等级{A，B，C，D，E，F，G}（F、G 两个等级的用户在预处理中处理成风险用户，附表中不再出现）。由于指标信息中存在用户信息为字符串格式或数据跨度过大，需要将这些用户信息转换成分类数据，详细的字符串与分类数值的对应关系如附表 1.1 所示。

附表 1.1 用户指标的分类匹配

职业	企业普通职员	企业管理层	私人经营	农牧渔劳动者	自由职业	在校学生	其他
类别	1	2	3	4	5	6	7
就业年限	<1 年	1-3 年	4-6 年	7-9 年	10 年及以上		
类别	1	2	3	4	5		
房屋所有权	自有住房	按揭	租赁				
类别	1	2	3				
年收入	5000-10000	10000-50000	50000-100000	100000-150000	150000 及以上		
类别	1	2	3	4	5		
信用额度	1000-5000	5000-10000	10000-20000	20000-30000	30000-40000		
类别	1	2	3	4	5		
信用等级	A	B	C	D	E		
类别	1	2	3	4	5		

附 录 2

```
library(tidyverse)
library(cluster)
library(factoextra)
df<-read.table("C:/Users/Administrator/Desktop/2.txt",header=T)
head(df)
#  不同的聚类数目对比分析
k2 <- kmeans(df, centers = 2, nstart = 25)
k3 <- kmeans(df, centers = 3, nstart = 25)
k4 <- kmeans(df, centers = 4, nstart = 25)
k5 <- kmeans(df, centers = 5, nstart = 25)
k6 <- kmeans(df, centers = 6, nstart = 25)
k7 <- kmeans(df, centers = 7, nstart = 25)
k8 <- kmeans(df, centers = 8, nstart = 25)

# Elbow Method 确定最佳的 K 数量
set.seed(123)
wss <- function(k) {
   kmeans(df, k, nstart = 25)$tot.withinss
}
k.values <- 2:8
wss_values <- map_dbl(k.values, wss)
plot(k.values, wss_values,
        type="b", pch = 19, frame = FALSE,
        xlab="Number of clusters k",
        ylab="Total within-clusters sum of squares")

#  选择最佳 K 值后重新实施 K 均值算法
set.seed(0)
final <- kmeans(df, 5, nstart = 25)
print(final)
data <- final$cluster
data <- as.matrix(data)
dim(data) <- c(36162*1,1)
write(data,file="C:/Users/Administrator/Desktop/2.csv")
```

致 谢

时光飞逝，两年的研究生学习生涯即将结束。在这段值得纪念的日子里，有许多人在不同的方面给予了我不同的帮助，让我的研究生生活充实而有意义。首先，将我最诚挚的感谢献给我的导师 XX 老师。

从论文选题、开题、论文的写作到论文定稿，一路以来 XX 老师给予了我无限的关心与帮助。在我为研究角度发愁时，XX 老师为我指明了方向；在我写代码卡壳时，XX 老师点出了我的错误之处。除此以外，XX 老师还时常关注着论文的进程。每当我遇到凭借一己之力难以解决的问题时，XX 老师总会为我提供帮助。在这两年的学习生活中，我从 XX 老师身上学到了高调做事，低调做人的人生态度，学到了对待学习和研究要抱有谦虚且谨慎的态度。感谢 XX 老师对我学业和生活的负责与关心，有这样的导师陪伴我度过研究生生涯的最后一课是我的荣幸。

其次，感谢重庆工商大学数学与统计学院的所有老师。我的研究生生涯因为有了各位老师的悉心教导，才能在专业上学有所成。

再次，我要感谢我的室友。在我论文的写作过程中，当我骨子里的惰性开始叫嚣时，是她们的督促使得我没有被懒惰吞噬；当我论文卡壳时，有她们陪我对我的论文内容抽丝拨茧，直到找到有问题的地方为止。两年来，无论是学习上的交流还是生活中的交往，我在她们身上学到了许多自己所欠缺的东西。有这样的室友陪伴我度过研究生的两年，何其有幸。

最后，我要感谢我的父母。感谢他们二十四年以来不求回报的付出，感谢努力工作的他们让我在求知路上无需为经济发愁，感谢他们将他们宝贵的爱毫无保留地献给我。在身边的同学选择大学毕业直接就业的时候，我的父母支持我继续求学的梦想，没有他们的支持，也不会有现在的我。在未来的生活中，我会继续保有学习的热情，努力工作，不辜负他们对我的付出与期望。