

硕士学位论文

基于聚类分析 K-means 算法的房地产客户 细分研究

CLUSTERING ANALYSIS BASED ON K - MEANS ALGORITHM OF REAL ESTATE CUSTOMER SEGMENTATION RESEARCH

于阳

哈尔滨工业大学

2017 年 6 月

国内图书分类号: F 293.31
国际图书分类号: 332.8

学校代码: 10213
密级: 公开

管理学硕士学位论文

基于聚类分析 K-means 算法的房地产客户 细分研究

硕 士 研 究 生: 于 阳

导 师: 芦金锋副教授

申 请 学 位: 管理学硕士

学 科: 管理科学与工程

所 在 单 位: 管理学院

答 辩 日 期: 2017 年 6 月

授予学位单位: 哈尔滨工业大学

Classified Index: F 293.31

U.D.C: 332.8

Dissertation for the Master Degree in Management

**CLUSTERING ANALYSIS BASED ON K -
MEANS ALGORITHM OF REAL ESTATE
CUSTOMER SEGMENTATION RESEARCH**

Candidate:	Yu Yang
Supervisor:	Asso. Prof. Lu Jinfeng
Academic Degree Applied for:	Master of Management
Speciality:	Management Science and Engineering
Affiliation:	School of Management
Date of Defense:	June, 2017
Degree-Conferring-Institution:	Harbin Institute of Technology

摘 要

企业由以产品为中心的管理理念，逐步转变为“以客户为中心，尽可能满足客户的需要”的企业管理理念，行业竞争也日益加剧，企业已经认识到客户才是稀缺资本及利润的根本，忠诚的客户更是企业难以替代的重要资本。因此，客户细分是对房地产企业管理不可或缺的步骤，同时，随着聚类分析技术的发展，在商贸领域中，聚类技术能够辅助管理者对客户的消费数据有效性进行分析，使消费模式通过聚类的帮助对消费者进行概括，有效区分消费群体，做出有价值的决策。因此，本文研究目的就是为了建立房地产客户细分模型，利用聚类分析技术对房地产客户数据进行有效分析，区分客户的类别做出针对性的客户关系管理。

本论文在房地产客户数据信息多而复杂的背景下，以房地产客户为研究对象，结合相关理论，构建了房地产客户细分指标体系，并且建立了房地产客户细分模型，对房地产客户的当前价值、潜在价值和忠诚度三个方面进行了分析，结合聚类分析技术 K-means 算法进行划分聚类，根据细分结果分析客户群，从而对房地产开发商提出合理的建议。

本论文明确了客户细分指标的构建原则以及步骤，阐述了客户细分理论和聚类分析理论，分析了房地产这一特殊行业客户细分指标，从而构建了房地产客户细分指标体系以及建立房地产客户细分模型。其中，对于客户细分指标体系，通过定性和定量的分析选取了影响因素指标变量，该指标体系包括 3 个一级指标，6 个二级指标以及 14 个三级指标，并据此设计出调查问卷，利用问卷星所提供的技术支持，将获取的有效调查问卷数据在 SPSS 软件中运行，结合之前聚类分析技术理论进行数据预处理，消除信息的噪音和冗余，接着进行客户的 k 均值聚类分析，对比原始数据进行类别的划分，区分出每个类别客户的特点，根据分类结果制定出企业的管理策略。

关键词：客户细分；聚类分析；k-means 算法；客户价值；客户忠诚度

Abstract

Enterprise by the product centered management idea, gradually shift to "take the customer as the center, as far as possible to meet customer needs" of the enterprise management concept, industry competition is increasingly intensified, enterprises have realized that the customer is the root of scarce capital and profit, loyal clients is important capital enterprise hard to replace. Therefore, customer segmentation is an indispensable step of real estate enterprise management, at the same time, with the development of the cluster analysis technology, in the field of commerce, clustering technology can assist managers to analyze customer consumption data validity, the consumption pattern of consumers through the help of clustering, to generalize effectively distinguish between consumer groups, make valuable decisions. Therefore, this article research goal is to establish the real estate customer segmentation model, using the cluster analysis technology on the analysis of real estate customer data effectively, to distinguish the customer to specific categories of customer relationship management (CRM).

This thesis more than in the real estate customer data information and complex background, the real estate customer as the research object, combined with related theory, constructs the index system of real estate customer segmentation, and the real estate customer segmentation model was established, the real estate customer current value, potential value and loyalty three aspects are analyzed, combining with clustering analysis technology partition clustering K - means algorithm, according to the result of segmentation analysis customer base, and put forward reasonable Suggestions to real estate developers.

This thesis has been clear about the principles and steps of the construction of customer segmentation index, this paper expounds the theory of customer segmentation and clustering analysis theory, analyzes the real estate customer segmentation index, the special industry to build the index system of real estate customer segmentation as well as the real estate customer segmentation model is set up. Among them, the index system of customer segmentation, through qualitative and quantitative analysis to select the factors affecting the index variable, the index system includes three first-level indicators, 6 secondary indexes

and 14 tertiary indicators, and designed a questionnaire, using the questionnaire star, provided technical support, will obtain the effective questionnaire data in SPSS software, combining with clustering analysis theory for data preprocessing before, to eliminate the noise and redundancy of information, customer k-means clustering analysis, then compared the original data for the category division, distinguish between the characteristics of each category customers, according to the classification results, the enterprise management strategy.

Keywords: Customer Segmentation, Clustering analysis, K-means Algorithm, Customer Value, Customer loyalty

目 录

摘 要.....	I
Abstract.....	II
目 录.....	IV
第 1 章 绪 论.....	1
1.1 研究的背景.....	1
1.2 研究目的和意义.....	2
1.2.1 研究目的.....	2
1.2.2 研究的意义.....	2
1.3 国内外相关研究现状.....	3
1.3.1 国内研究现状.....	3
1.3.2 国外研究现状.....	5
1.3.3 国内外研究现状评述.....	5
1.4 研究内容与方法.....	6
1.4.1 研究内容.....	6
1.4.2 研究方法与技术路线.....	7
第 2 章 客户细分和聚类理论分析.....	9
2.1 客户细分的概念原则以及指标.....	9
2.1.1 客户细分概念.....	9
2.1.2 客户细分原则及指标.....	10
2.2 客户细分的数据挖掘技术的选择.....	13
2.2.1 统计分析方法.....	13
2.2.2 决策树方法.....	14
2.2.3 神经网络算法.....	14
2.2.4 聚类分析.....	14
2.3 聚类分析技术的选择.....	15
2.3.1 聚类分析技术概述.....	15
2.3.2 聚类分析算法.....	16
2.3.3 K-means 算法.....	18
2.4 本章小节.....	20
第 3 章 房地产客户细分模型的构建.....	21

3.1 指标体系构建原则及步骤.....	21
3.1.1 指标体系构建原则.....	21
3.1.2 指标体系构建步骤.....	21
3.2 房地产客户细分指标分析.....	22
3.2.1 基于客户价值细分指标体系设计.....	22
3.2.2 基于客户忠诚度细分指标体系设计.....	23
3.2.3 基于客户价值和忠诚度构建房地产客户细分指标体系.....	24
3.3 客户细分模型设计.....	27
3.3.1 数据预处理.....	27
3.3.2 主成分分析.....	28
3.3.3 房地产客户细分模型.....	30
3.4 本章小结.....	32
第 4 章 房地产客户细分实证分析.....	33
4.1 宝宇地产的企业现状评述.....	33
4.1.1 宝宇地产的企业现状.....	33
4.1.2 宝宇地产 SWOT 分析.....	33
4.2 数据预处理.....	35
4.2.1 数据准备.....	35
4.2.2 主成分分析.....	36
4.3 基于 K-means 算法的客户细分.....	45
4.4 结果分析及建议.....	49
4.4.1 结果分析.....	49
4.4.2 对宝宇地产实施客户细分建议.....	52
4.5 本章小节.....	53
结 论.....	54
参考文献.....	56
附 录.....	60
哈尔滨工业大学学位论文原创性声明和使用权限.....	61
致 谢.....	62

第 1 章 绪 论

1.1 研究的背景

按照国务院有关部署，自 2016 年 3 月开始调整房屋交易环节契税、营业税等优惠政策，以及 2 月 26 日央行行长周小川说“住房贷款应该有大力发展的空间”。随着中国房地产的宏观调控政策频出，房地产企业间的竞争越发激烈。要想在这种市场竞争中谋求生存以及发展，须改变传统的营销模式。另一方面，由于网络通讯技术的迅猛发展，消费者能够轻易的获取信息，并且凭借信息可以决定选择谁、如何选择、何时选择自己偏好的产品和服务，从而使得消费模式变得越发复杂化和信息化，故而，企业和客户间的关系面临十分严峻的考验。

房地产企业的产品与其他产品不同之处是其地域性，满足某特定区位、特定购房需求者的要求。房地产企业重要的一项工作就是进行精准的市场细分以及市场定位，因为在特定区域的市场，客户的需求也是不尽相同的，只有满足客户的需求使客户满意，客户才会购买房地产企业销售的产品。根据马特莱法则——企业利润 80%是由 20%有价值的客户带来的，所以只需把握好这 20%客户就可以获得大部分收益，客户不完全都是上帝，企业的资源不是无限的，唯有抛弃那些对企业没有价值的客户找到价值高的客户，进行针对性的管理才能获得最大利润，于是客户的地位得到了前所未有的提升。随着房产供不应求的时期逐步过去，就房地产企业而言，企业建立客户关系管理已经迫在眉睫。客户细分是客户关系管理中的重要内容，客户细分是指公司在明确的业务模式和特定的市场中，根据客户的某一种或是几种因素对客户进行细分，并根据细分结果提供与其相对应的产品或是服务^[1]。从客户生命周期价值角度进行客户细分，需要从各个方面了解客户，必须对以后的可能性也进来考量。据权威统计显示，企业受客户盈利能力及保持率影响，客户保持率每增长 5%，其平均利润率就会增长 25%-50%，即开发新用户的成本远大于维护原有客户成本，也就是说原有客户的重复购置房产和推荐新客户购置房产以及交叉购置房产是房地产企业的重要获利来源。

在房地产企业发展的过程中都累积了海量的客户信息数据，并且客户数据信息量大且杂，然而很多信息隐藏于数据之中不能被发掘和使用，这样浪费客户信息数据。数据挖掘技术是面向应用的，是多种理论与技术知识的综合，这种技术的客户细分与传统的客户细分相对比能够发现客户关系中潜在的有利信息^[2]。然而在众多的数据挖掘分析技术中，聚类分析技术在客户细分领域内

基于划分的算法应用极为广泛，其优点是：聚类分析处理时间快，操作简单、易于理解，是被普遍应用的数据挖掘算法之一，以 k-means 聚类算法最能代表。聚类分析技术目前已经应用的领域有：在解决遥感成像、模式辨别、统计、生物学、金融商业、地理学等领域，但是在房地产领域应用相对较少，本文通过房地产企业采取聚类算法来细分客户，获得具有差异特征的客户，从而辅助房地产企业制定出相应的购房者的生产和服务的策略，对购房者进行针对性管理，发掘购房者的潜能，提升购房者的价值贡献，合理利用房地产企业资源，使企业获取最大化的利润。

1.2 研究目的和意义

1.2.1 研究目的

企业由以产品为中心的管理理念，逐步转变为“以客户为中心，尽可能满足客户的需要”的企业经营理念，让企业与客户的关系受到了前所未有的重视。2015 年年中以来，中国房地产市场逐步进入回升周期，房价总体水平逐步攀升，房地产投资同比增速也在 2016 年年初开始有了明显回升。2016 年房地产的主基调就是去库存、稳市场，其市场形势已经发生变化，并且向买方市场逐渐转变，而且消费者消费行为更加理性，行业竞争也日益加剧，企业已经认识到客户才是稀缺资本及利润的根本，忠诚的客户更是企业难以替代的重要资本，唯有把握住、吸引住客户才可以使房地产企业获得更多的利益，使企业更具有竞争力。客户细分相关理论的研究在很多领域已经取得很高的成果，并应用于客户管理以及营销实践中，但对于房地产行业研究甚少。

本文针对宝宇地产开发企业现状，以哈尔滨宝宇天邑实际客户数据为例，通过客户细分理论和聚类分析抽取客户样本进行细分，首先根据收集样本数据进行预处理，提取主成分，后续对变量进行聚类分析，根据聚类分析研究，构建客户细分模型，为宝宇房地产公司提出意见，希望能够利用细分结果帮助处理实际问题，保留客户和合理分配企业资源。

1.2.2 研究的意义

目前，随着市场竞争的日益加剧，资源有限的客户转化成为房地产企业最重要的利润来源。无论是宏观还是微观意义的企业，开展客户科学合理的细分这是具有重大的理论和现实意义。本文的研究意义有以下二点：

(1) **理论意义** 1) 客户细分是基于客户属性的集合，是客户关系管理中主要的理论构成成分，也是主要管理工具之一，可以对客户获取、保持及增

值等关系管理过程提供全面支持,提升客户满意和忠诚。2)在地产业的客户细分中引入聚类分析技术,不断扩大的客户细分研究的范畴,开发维护客户、营销战略和客户保持策略,同时,为其他研究学者提供借鉴。

(2) 实践意义 1) 客户是成功的关键,得客户者得天下。客户资源变得稀缺而且客户变得越发理性,企业需要全力以赴地把握住他们,赢得他们的信赖。2) 根据对购房者的分析,尊重购房者,构建重视购房者的客户导向型模式,将客户进行细分,并辨识出客户之间的异同,帮助企业找到优质的客户,使房地产公司可以合理配置资源,为公司创造尽可能多的利润。

因此,本文采用聚类分析技术对房地产企业客户进行分类,从客户价值和客户忠诚度的两个角度建立客户细分指标体系,并本着“以客户为中心”的理念贯穿整个开发,销售,以及后期服务的整个过程,以客户需求为管理决策、计划的出发点和立足点,提高房地产企业的效率。

1.3 国内外相关研究现状

1.3.1 国内研究现状

在客户细分及聚类分析算法研究方面,国内的许多学者在各自领域都有自己的建树,研究的内容主要是客户细分和聚类算法的方法、模型、算法等角度。

(1) 客户细分国内研究现状 基于客户细分方法研究,王婷(2016)认为按照评价体系指标的设计原则,并且增加让渡价值和行为维度的部分指标,为了评估客户价值,利用散点图并根据客户价值评价进行客户细分,把客户分为高价值,潜在价值和低价值^[3]。褚格林(2014)认为客户价值应该包含当前价值和潜在价值,客户细分基于客户的价值把每类客户不同的价值对客户进行相应分类^[4]。于辉、廖小红(2014)认为基于客户价值的细分最重要的是要对客户当前价值和潜在价值进行详细评估,再通过其价值量多少对其进行聚类,以此为基础划分消费群,按照消费者细分群体的差异特点采取并制定针对性的资源分配和客户管理策略,从而增加企业效益^[5]。张焕国(2011)认为运用C均值技术对电话网络用户细分研究能够产生很好的效果^[6]。季越江、吕佳(2009)认为客户细分的主要内容应该分为消费者的价值、客户特点以及人口特征三个角度^[7]。通过学者们不同的研究方法,已将科学的客户细分方法应用很多行业的客户管理以及营销实践之中。

基于意义研究,金雷法(2010)认为精准地进行客户细分,以消费者为中心提升消费者满意度不仅是实施客户保持策略的基础,更是企业竞争优势的主要来源,只有对有价值的客户有效地利用企业资源,才可以提高企业获得持续

利润的能力^[8]。范志国、付波（2010）认为企业取得成功最重要的原因之一就是客户保持，准确的细分是落实客户保持的第一要务，是为公司采取有效的策略提供重要的保障^[9]。陈文沛（2010）从客户创新开始，以客户创新共享客户细分^[10]。通过学者们的研究，客户细分有利于企业发展新的客源、保持住之前的老客户、提高客户的忠诚度，进而提升企业效益，增强竞争实力。对于企业发展乃至经济发展都有很大的积极意义。

基于客户细分模型与实证研究，叶志龙、黄章树(2016)认为从客户的当前价值、客户潜在价值、客户忠诚度着手，构建客户评价体系用来适应线上会员评价^[11]。纂欣德（2013）发现目前的细分方法细分都是从单一维度进行的，构建静态的模型且大多没有预测功能，认为客户细分方法的新的三个角度：多维度、动态性、预测性，即细分模型不但可以呈现出结果，而且能够预测客户以后的行为概率^[12]。徐翔斌（2012）认为在对电商公司客户细分时将用户引入总利润因素并改进了 RFM 模式，提出了 RFP (P: Profit, 总利润)模型^[13]。刘潇潇（2013）以数据挖掘为基础，结合了电力消费分类理论与客户价值评价指标，建立了消费行为的细分模型^[14]。随着客户细分研究的深入，相关细分技术模型和方法也越来越多，并且在各个行业的实际应用都已经趋于成熟。

（2）聚类分析的国内研究现状 由于客户的需求趋向多样化，细分指标越发细微，细分维度越来越大，在其相关理论研究的过程中，传统的算法无法适应繁多而杂乱的信息量，因此，聚类分析技术的诞生很好的解决了这一问题。

在聚类分析技术方面，王勇、唐靖（2014）提出使用数据层来获得搜索组数的上限，并建议基于聚类评价指标，获得最佳聚类数^[15]。马儒宁、王秀丽（2014）提出了新的凝聚聚类算法，这种基于多层聚类中的算法被称为 MULCA 算法^[16]。韩凌波（2015）提出如果有多个参数值密度最大时，怎样选择采样点作为起始中心的聚类方法，其结果是一致的，而且提升了效率，但该方法不足之处是需要很长的时间运算^[17]。

在聚类分析中 K-means 算法方面，何佳知（2015）按照噪声阈值和平均密度方式，将数据集分成组，并设置噪声点集合的高密度集合，并且根据情况分别选择聚类用来消除噪音点引起的 k 均值问题^[18]。许茂增（2015）按照遗传算法随机设计的突变概率和云服务的稳定性趋势，在发展进程中引进 K 均值解决收敛速度慢的问题^[19]。邢长征（2014）提出在比平均密度更高的采样点中选择初始聚类中心^[20]。

在 K-means 算法实际应用上，樊宁（2011）认为该模型能够加快客户细

分的速度和细分的准确性，从而可以改善我国银行“海量数据，缺乏信息”的尴尬状况^[21]。彭凯（2011）研究了短信业务的客户细分，他为了提高构造模型的质量和效率，首先对原始数据进行因子分析，能够减少冗余字段，然后利用 k 均值算法进行客户细分^[22]。赵铭（2013）利用 k 均值算法对银行的往来客户进行了聚类，并且利用判别函数进行检验，分析分类结果，有效地识别了基金理财客户^[23]。

1.3.2 国外研究现状

（1）客户细分国外研究现状 Bain Company（2009）公布的报告表明，客户细分目前已经是全世界十大管理工具之一^[24]。Gremler 和 Brown（2001）认为把忠诚可以分为行为、意识以及情感忠诚，强调对客户行为趋势的分析，表达出客户与企业的产品或服务之间的感情^[25]。Poel（2007）将客户忠诚度指标引入客户细分中，并从行为、态度忠诚二个角度对客户忠诚度进行表述^[26]。J. Albrecht（2007）认为由于有效的客户细分，有益于满足客户的需求，以及有益于企业的产品和服务能够满足客户，更有益于资源的更有效、合理地配置^[27]。Bottcher M（2009）提出解决客户细分问题最好的技术方式就是聚类技术分析^[28]。

（2）聚类分析的国外研究现状 Yedla M（2010）概括聚类分析内涵，要求能合理地按各自的特征来划分类别，在没有先验知识的情况下将数据划分到不同的类或者簇这样的一个过程，因此同一个类样本中其各个样本相似性强，而不同类样本中样本相异性强^[29]。Dehariya V K（2010）指出聚类不同于分类，分类具有先验知识条件是参考学习，对于分类模型而言，样本数据的所属的类标号是已知的^[30]。Aldahdooh R T(2013)认为聚类的效果对初始中心点选取有所不同，并且其选择中心点的合理性会直接反应到算法的聚类效果，如选择不合理会使目标函数收敛过缓，甚至会使算法迭代次数增加，其结果不稳定^[31]。Jose（2011）认为将 Hopfield 递归方法与 K 均值方法相融合取长补短，将会获得类似的全局最优解^[32]。

1.3.3 国内外研究现状评述

通过国内外文献现状分析可以看出，在研究方法上，国内外学者们目前主要研究均是非传统的统计分析方法，并且相应的细分技术方法种类繁多，但对客户细分的运用主要是利用数据挖掘技术，在其众多技术中，应用最广泛并且最重要的就是聚类算法^[28]。此方法问世至今，已取得相当丰富的经验和成果，

并研究出了很多的算法，大致分类为划分法、层次法、密度法、网格法以及其他聚类方法^[33]。并且其算法的具体应用已经趋于成熟，并且应用广泛。

对于企业实证而言，由于客户的消费行为繁杂，就决定了客户细分方式没有绝对有效，也就是说没有统一的细分方式，细分方式必须要与细分目标完全相同，并且消费者行为的善变性也表明了细分的方式也需要相应变化的，利用不同的细分方式分析同一例子，分析的结果有时也会不同，因此企业根据本企业的实际情况以及具体的环境和细分目的，选取相适应的细分方法，构建面向该企业的消费者细分模型，得到精确的细分结果。

在应用方面，国内外学者研究主要面向在一些特定的行业与领域之中，这些领域不但包含传统的实体生产行业，也应用在新兴虚拟的市场之中。目前国内房地产市场已经趋于成熟，但是房地产客户消费与其他客户消费有所不同，不可能直接利用其他行业的细分指标和方式，应根据房地产企业特点寻找适合自己身的一套细分指标和方法。

1.4 研究内容与方法

1.4.1 研究内容

目前，关于房地产企业客户细分的方法还处于传统的经验描述法和统计分析法，这二种方式的细分仅做到粗略的客户细分^[34]。本文的研究内容便是融合客户细分理论与聚类分析理论，摸索出可以应用到房地产企业的客户细分模型，然后利用聚类分析技术，结合收集到的宝宇房地产企业的客户调查问卷数据进行细分研究，以此为基础找到宝宇天邑项目的目标客户并根据聚类的结果，全面实施项目客户细分，取得实际的销售业绩，区别对待不同的聚类结果的客户，根据聚类结果落实相应的客户管理策略。本文主要以以下几个方面研究：

（1）客户细分指标体系的构建 当前国内外的研究人员已对客户细分指标研究的很深入了，可是对于房地产这个特殊的行业，之前学者偏向研究如何完善房地产购房客户售后服务水平，以及客户购买房产后的满意度研究，即客户价值和满意度这二个维度，却很少研究侧重于房地产客户的忠诚度，故而本人将结合房地产行业的特殊性，探索适用于房地产企业的客户细分指标，并将此作为客户细分的依据。

（2）获取数据的约简 将获取的房地产客户数据信息预处理，在客户细分前不出现信息冗余，提高细分的准确性，本文根据客户细分特点运用主成分分析方法将获取的房地产客户数据信息进行约简。

(3) **聚类分析技术的选择** 由于房地产企业自身的特点和实际客户数据需求的精度及可操作性等,用 K-means 聚类算法中对客户数据进行细分, K 均值聚类结果会分为不同的客户群,接着根据不同组群的特点,最后针对分析的数据结果提出相应的市场策略。

围绕上述内容,本文分为四章:

第一章:阐明论文选题的研究背景,在对国内外研究现状进行论述与分析的基础上,明确论文研究目的及意义,并提出论文的主要研究内容、研究方法。

第二章:阐明了客户细分以及聚类分析相应的理论。首先是客户细分的理论发展,介绍了客户细分指标以及客户细分方法,之后比较分析各种方法的优缺点,选取聚类分析技术方法,评述算法并且介绍聚类分析算法中的 K-means 算法中相关内容。

第三章:首先,建立客户细分指标体系。根据地产行业和客户细分特征加之客户价值和忠诚度建立房地产企业消费指标体系,在前述部分以聚类分析方法为基础,选取 k 均值算法对客户进行细分,为客户细分前不出现信息冗余将获取的房地产客户数据信息进行预处理,最后,构建出利用聚类分析的房地产企业客户细分模型。

第四章:将宝宇房地产客户回收的调查问卷信息,运用 SPSS 软件进行主成分分析,消除信息的噪音和冗余,接着进行客户的 k 均值聚类分析,对比原始数据进行类别的划分,区分出每个类别的特点,根据分类结果制定出企业的管理策略。

最后提出论文的结论,总结全文的内容,并提出论文的创新之处和存在的不足。

1.4.2 研究方法与技术路线

本文的研究主要采用了以下方法:

(1) **文献研究的方法** 通过对客户细分和聚类分析的研究和知识梳理,以归纳文献作为论文的理论基础。借助查阅文献资料,分析和比较不同的客户群和细分方法的差异,并在研究中选取适应于房地产行业的分项指标。

(2) **定性与定量研究的方法** 将选取的细分指标变量做定性的表述。再利用调查问卷回收的数据利用 SPSS 软件进行定量的运算,验证定性表述的准确性。

具体的研究技术路线如图 1-1 所示。

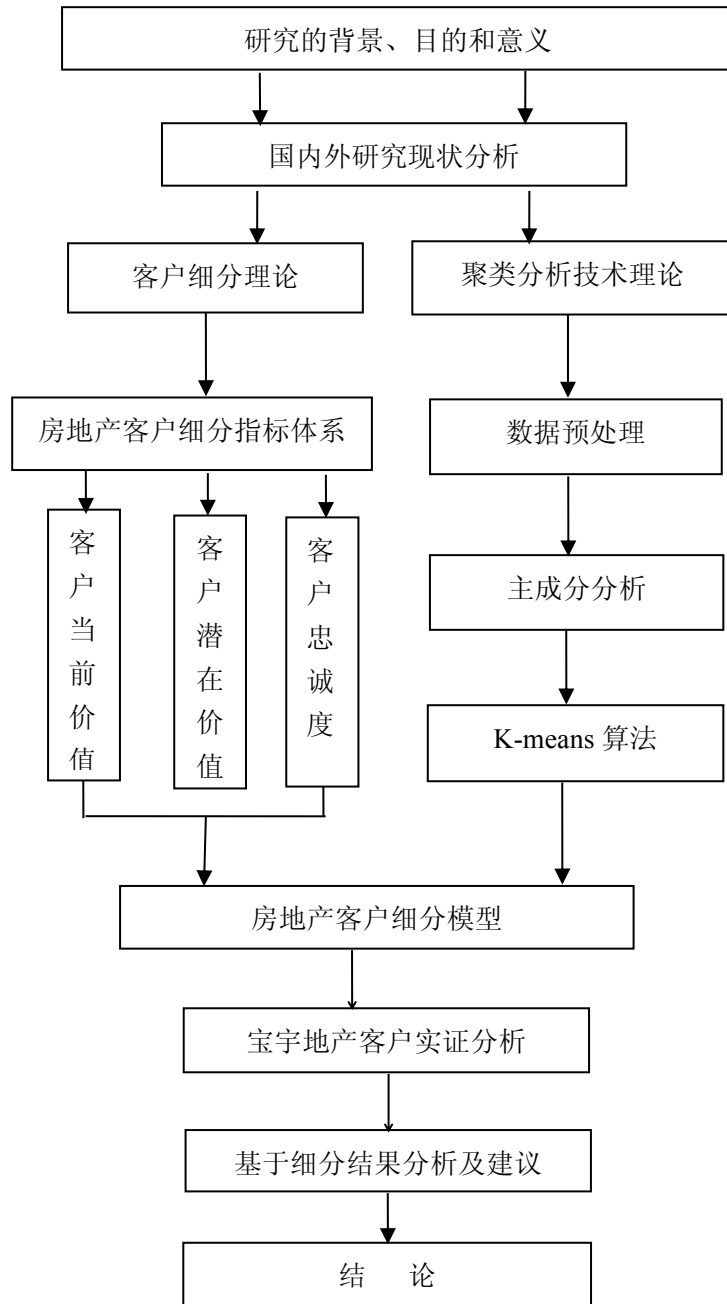


图 1-1 论文研究技术路线图

第2章 客户细分和聚类理论分析

2.1 客户细分的概念原则以及指标

客户（又称消费者、顾客）是指去商店购物的人或去服务行业寻求服务的人，包括已经来或是将要来购买你的产品或服务的个人或单位都可以算是客户。在现实生活中企业的各种产品和服务最终面向的就是与其相关联的客户。没有客户企业的产品便无从销售，进而在市场竞争的浪潮中也无法生存，客户是企业经济利益的根本来源，如何有效的处理企业与客户的关系、企业如何有效的划分自身的产品从而针对市场中不同客户的需求而有的放矢，客户细分显得尤为重要。

2.1.1 客户细分概念

自二十世纪五十年代由美国学者温德尔史密斯先生基于消费需求的差异性及在公司资源利益最大化基础上，首次提出了客户细分的概念，这一概念的提出备受关注，它一直吸引着许多专家学者对它进行研究。概念提出至今，客户细分主要分为两派：一派是以产品为导向，另一派是以客户为导向^[35]。本文探讨的是第二种即客户导向性的细分。客户细分概念是指在一个明确的策略，商业模式和具体的市场环境中，根据不同企业的客户固有的特点，需求差异，消费习惯，偏好等属性，提供了客户分类处理的产品定位或客户服务模式^[5]。

客户细分是基于客户属性的集合，是客户关系管理的重要理论构成成分，其意义主要表现以下四点：

（1）**合理有效的配置企业资源** 客户资源对于企业而言是弥足珍贵。根据马特莱法则——企业利润五分之四是由五分之一有价值的客户带来的，所以只需把握好这五分之一客户就可以获得大部分收益，客户不完全都是上帝，故而采用均衡分配资源的方式显然不够正确，唯有放弃低价值的客户寻找到高价值的客户，对高价值的客户特殊的对待，通过有效的客户细分，企业能够针对不同价值层次的客户差异性来配置资源，在最大化利用有限资源的同时也给企业创造更多的价值。

（2）**拓展新的客源** 当前大多数产品已经是供大于求，买方已经主导市场，消费者可以任意的选择满意的公司的产品或服务，客户是企业的生存基础，不断发展的企业更需要新鲜血液来消费其生产的产品和服务，基于客户细

分,既要挖掘已有客户关系中潜在的有利之处,又要拓展潜在消费的客户,制定相适应的保持策略,引导拓展新的客户市场。

(3) **维系原有客户** 据权威统计显示,企业受客户盈利能力及保持率影响,客户保持率每减少 5%,行业的平均利润率就会减少 25%-50%之间,即开发新用户的成本远大于维护原有客户成本,也就是说原有客户的重复购置房产和推荐新客户购置房产以及交叉购置房产是房地产企业的重要获利来源。因此客户细分可以通过对客户不同消费习性行为的分析,及时关注原有客户的动向,以达到维系客户数量和客户质量的目的。

(4) **提高竞争力,扩大市场占有率** 全球一体化进程的加快,任何企业要想在这种竞争十分激烈的市场环境中生存和发展,处理好客户关系显得尤为重要,企业需要及时了解客户需求的变化、更新产品和服务、调整营销战略,做到有的放矢,满足不同的消费需求,提升竞争力,扩大市场占有率。

2.1.2 客户细分原则及指标

2.1.2.1 客户细分原则

客户细分的价值和作用主要就是支撑企业市场营销、客户关系管理等战略和策略的制定,因此客户细分保持以下几个原则,才可以实现其应有的价值和作用。

(1) **客户群体特征的可识别性原则** 客户细分的最主要目地就是能得到不同特征的客户群体,对其结果能简单的进行识别。因此企业要对既有客户信息的细分维度进行分析,确定细分的可度量性,确保结果是每个细分类别有各自的相同的特征,但是不同群体具有明显的差异的特征。如果无法满足这个基本的原则,客户细分也就没有意义了。

(2) **细分指标的有效性原则** 客户细分的意义在于,合理有效的配置企业资源,最大化利用有限资源的同时也给企业创造更多的价值。因此在客户细分中,一定要选取对客户关系管理工作有意义的指标,避免无用的指标浪费企业资源。例如:对客户价值进行细分时,需要选择能真正有效的能代表客户价值的指标,如购买金额、总额,交叉购买意愿、利润等。只有指标选取上确保有效性,才能确保之后的销售策略和管理策略等措施的有效性,这样才可以对应准确的目标客户,实施起来也不浪费企业的有限资源,可以发挥客户细分真正的效用。

(3) **细分结果的合理性原则** 客户细分如果方法选择不妥,会使得有一部分目标客户的数量极少,企业没有必要制定针对性管理策略,这样不合理

的结果使客户细分也没有意义的。因此，客户细分需要有合适的客户数量和经济范围，为了获得最终合理有价值的客户细分结果，在选择客户细分方法上，对可能产生的客户群体做出初步的评估，在其预判的基础上选择正确细分方法和客户范围，确保细分结果的合理性，并适用于客户管理策略的实施。

（4）**客户细分的时效性原则** 企业营销策略是有时效性的，因此支持企业营销策略制定的客户细分群体的细分指标也应具有时效性，所以应当选择具有相对稳定性的细分指标，并且尽量保证客户数据信息及时提取和版本的动态更新，来确保企业制定针对性策略。因此客户细分应该不断的改进和更新规避信息滞后，其时效性是保证营销策略实施的有效性的基本保障。

2.1.2.2 客户细分的指标

企业往往根据客户细分的实际需求不同，使得客户细分模型是不固定的。常用的方法有定性和定量两种：定性的客户细分方法主要凭借决策者主观判断，在分析过程凭借经验会出现偏差，容易造成判断失误；定量客户细分方法是借助相应的客户细分变量(统计特征、客户心理、消费行为、客户生命周期等)为基础，利用定量分析技术进行客户细分。客户细分变量指标通常分为以下五类变量：

第一类是基于客户的统计学特征。该细分技术主要通常依赖于客户的自然属性，一般针对自身条件状况而言，具体包括性别、年龄、家庭背景、职业、收入水平等人口统计指标属性，例如女性衣物，根据不同年龄阶段老、中、轻推出不同的产品与服务。这一类客户的统计学特征的细分方法其数据信息收集很容易，但是此类方法只是反映消费者的自然属性信息，并不能表现出客户需求的差异，且在划分和预测时很容易出现误差，因此，对于企业参考价值较低。

第二类是基于客户心理偏好。客户心理偏好，是指客户对企业某产品或某项服务的喜好程度^[36]。认知心理学研究表明，偏好是自身经验及多重感知而产生的综合性心理意象情感^[37]。相应的指标包含客户个性方面的特征，如需求、个性、动机等。这些特征会对用户购买倾向有一定的影响作用，并且具有一定积极市场意义。可是在现实的应用中，消费行为受多重因素相互影响，很难预测客户的消费变化。

第三类是基于客户的消费行为。消费行为是指客户为了获取产品或服务而实施的不同的决策行动过程^[38]。按消费的行为属性来划分不同的客户群，不但可以高质量的反应出客户内在的消费规律和消费变化，而且可以挖掘出潜在的有价值的客户。这一类细分指标包含购买时机、次数、总额等变量参数，差异的反应了客户种类的行为变化，通常这种划分方式可以很好的反映出差异客

户带给企业的区别的价值比,但是没能表达出客户是否对产品或服务的满意和忠诚程度。例如 RFM (Recency Frequency Monetary)分析模型^[13],该模型是按照消费时间段、消费频率、消费额度三个方面按比例进行划分的,即基于客户的行为属性来对消费者进行划分,因此该模型使用广泛,划分结果较优,缺点就是计算复杂,容易增加分析成本。

第四类是基于客户的生命周期。客户生命周期对于不同的行业其详细的定义有所不同。例如就房地产行业而言,指的就是从接触楼盘,与其他楼盘进行比较,进而购买,或是由于某种原因购买其他地产公司楼盘的全过程^[39]。根据地产行业销售特点,一般可以分为四个阶段:阶段 1,建立期;阶段 2,成长期;阶段 3,成熟期;阶段 4,退化期。这一类细分指标变量可以区分购房者所在阶段,可是无法区分客户间的差别。

第五类是基于客户价值相关指标。由于企业营销的目的是为了实现利润最大化,所以决策者会尽可能多地获取和保持更多有价值的客户,但是客户的价值是有差别的。根据马特莱法则——企业利润 80%是由 20%有价值的客户带来的,所以只需把握好这 20%客户就可以获得绝大部分利益,客户不完全都是上帝,故而采用均衡分配资源的方式显然不够正确,唯有放弃低价值的客户寻找到高价值的客户,对高价值的客户特殊的对待,才能充分利用企业有限资源,利润和价值逐渐驱动着企业进行客户细分,因此,出现了以客户价值维度作为细分的方法。客户价值包括当前和潜在价值,客户价值虽然可以反映出客户为企业带来利润的多少,但是不能看出客户的忠诚程度。因此目前学者也在寻求和完善相关指标,摸索更好的测算客户价值的方法。

分析以上的客户细分指标变量,总结表 2-1 所示。

本文主要探索的是房地产企业的客户细分,其目标就是辨识出消费者价值的差异性,寻找高价值的消费者,利用有限的企业资源对其进行针对性的管理。另外以往的学者在房地产行业很少注意到客户的忠诚度这一重要指标,这一指标经常在互联网经济强调口碑和用户评价,忠诚指标是目前能找到的最优的承载客户导向化的载体,它既代表消费者的信赖与认可,又是从客户端反馈公司业务短板的一面镜子,持续推动组织自我修复,并持续改善提升维持企业竞争力。例如哈尔滨宝宇地产在天邑项目和环球港项目中,通过忠诚客户高满意度转换为再购和推荐购买的一种可持续的增长模式,充分利用已有客户信息,对其加强宣传,很多客户都购买了新的宝宇楼盘或是推荐亲朋好友购买宝宇新楼盘,即维护老客户带给企业的价值远大于新发展客户所带来的价值,并且需要的资源要少的多,加之忠诚是企业业务管理举足轻重的决策刻度可以持续

提升品牌信赖度。因此,本文选取以上二个指标进行表述,但是这二个指标难以定量计算,故采用间接方式进行表述,把客户价值和客户忠诚度拆解成若干个能够量化的变量指标。具体的指标选择将在第三章进行讨论。

表 2-1 各类客户细分指标的比较

类型	依据	细分变量	特点	应用局限性
1	基于客户自然属性和人口统计学	性别、年龄、家庭情况、工作类型等	数据容易收集,细分过程简单	适用于了解客户结构,但难以预测消费行为
2	基于客户心理偏好	个性、需求、动机感知、态度、学习	能反映客户对产品或服务不同需求	客户行为影响因素较多,此方法预测会出现偏差
3	基于客户行为	购买时机、次数、总额、追求利益	可以反映不同类型客户消费行为变化	不能反映客户对企业满意度等方面特征
4	基于客户生命周期	客户与企业关系的时间	从长远考虑,对客户与企业关系变化进行分析	不能识别处在同一阶段的不同客户间差异
5	基于客户价值	客户的直接价值和间接价值	从客户创造利润角度分析,全面反映客户特征	客户价值的测算比较复杂,不能反映客户忠诚度

2.2 客户细分的数据挖掘技术的选择

随着网络通信技术和数据挖掘技术的迅速发展,企业的产品和服务日益的信息化和数字化,并且客户信息量迅速的膨胀,人们意识到传统的取决于决策者的主观经验定性客户细分的方法,受主观影响的会产生偏差不能为人们提供精确的决策信息,因此众多学者着手客户细分定量分析的研讨,提出的算法不胜枚举,这些算法各自差异的适用于某个领域或是某种情况,每个算法都有自己的优缺点和适用的条件,为了确保细分结果的准确性有时候更是几个算法结合起来使用,还有的随着具体问题改进相应的算法。这些算法较为广泛使用的技术有以下几种:

2.2.1 统计分析方法

统计分析是将大量难以识别的隐含数据信息转化为简单的可以解释的并且能方便运用的信息。其优点是软件好学易懂,算法精简,建模和运算速度快,

并且可以快速的得到结果。但是也有一定局限性，如线性回归无法处理复杂的大量的信息，细分变量和属性值间会出现显著的线性关系；非线性回归具有操作复杂，表述问题的实质很难并且有很强的不确定性等问题；主成分分析在指标多且每个指标间影响较平均的情况下无法发挥细分作用。因此，统计分析方法只适用于变量少的、关系简单、数据量小的情况，并且获得的结果还无法保证精确性。

2.2.2 决策树方法

决策树方法将决策集合用分类回归树表示，从而形成一定的规则。因此决策树在解决细分问题中是主流算法，利用决策树方法的优点是决策树的规模不依赖于数据库的规模，并且效率高，易理解，易扩展，比较适合处理大型数据；其缺点是该方法不能够考虑到属性的相关性，并且需要完整的数据信息，如有缺失数据极大的影响处理效果。

2.2.3 神经网络算法

神经网络就是一对连接 I/O 单元，并且一个连接将同某个权重项关联，其本质是一种非线性预测模型，和生物的神经网络结果相类似，其过程就是通过各种迭代算法使得样本能够很好的拟合。神经网络算法的优点是处理噪音数据时有较强的稳定性和容错性，自适应能力强并且结果准确度高；其不足之处在于难操作并且有时其结果不易理解。

2.2.4 聚类分析

聚类分析是一种寻找统计规律的方法。该分析方法应用及其广泛，例如计算机、生物、商务管理等。聚类的优点是处理时间快，操作简易、容易解释；缺点是使用时需要设置参数，有时结果可能会出现局部最优的情况。聚类分析算法是本文主要的使用的方法。将在下一节具体评述。

以上评述的算法各有各的特点，各种方法的对比结果表 2-2 所示。

根据表 2-2 各类算法比较可以得出结论：统计分析方法简单并且可以快速处理数据，但是很难处理复杂数据并且精度低；决策树方法过于依赖参数并且运算时间较长，需要完整的数据信息，如缺失数据极大的影响处理效果；神经网络优点是处理噪音数据时有较强的稳定性和容错性，不足在于操作复杂、成本高；聚类分析处理时间快，操作简单、易于理解，缺点是对参数依赖，精度一般。每种方法都有各自的优缺点，因此对客户细分方法选取时要根据具体的

情况和细分目的选取相应的算法。本文是研究房地产客户细分研究的，因此选取聚类分析能够符合细分准确性目标，加之此这类分析技术具有操作简易、结果容易理解等优点本文选取聚类分析技术。

表 2-2 各类算法优劣比较分析表

方法	精度	操作难度	运行时间	处理模糊情况	结果可解释性	参数依赖
统计分析	低	简单	短	否	可解释	自学习
决策树方法	一般	一般	长	否	可解释	依赖
神经网络算法	高	较难	长	否	不可解释	自学习
聚类算法	一般	简单	短	是	可解释	依赖

2.3 聚类分析技术的选择

当数据信息量少时，能够看出数据及数据特征的区别和联系。但当给定一个大数据集时，就很难发现数据之间的潜在联系，很难将数据属性相似划为一类，或是区分差异的数据属性。于是出现了聚类分析技术，这项技术应用领域广泛，应用于商业数据分析、分析医疗病患数据、对图像进行分析、生物基因特征、对地貌特征分类以及天文研究等。例如：在生命工程中，聚类分析技术在基因数据中寻找相似功能的基因，或是在生物学中能够辅助动植物分类；在地理中，用聚类分析技术找到相似用途的区域用来辅助进行矿产勘查、资源预测等；在商贸领域中，聚类技术能够辅助管理者对客户的消费数据有效性进行分析，使消费模式通过聚类的帮助对消费者进行概况，有效区分消费群体，做出有价值的决策。

2.3.1 聚类分析技术概述

聚类分析是将样本或指标分类的多元统计分析，面对数目庞大的样本，根据每个样本的特点合理的聚类，在没有先验知识的情况下进行的把信息聚类至不同的类的过程，在一个类中的样本相似性很大，而在不在一个类中的样本没有相似性^[29]。根据对象的不同可以分为样本分类和变量聚类，即分别是对样

本进行分类和对变量进行分类。聚类不同于分类，分类具有先验知识条件是参考学习，对于分类模型而言，样本信息的所属的类标号是已知的^[30]。

2.3.2 聚类分析算法

聚类算法类型众多，选取时应结合具体的数据模型、分类目的等要素确定选取哪种算法。常用的算法大致可以分为层次法、划分法、密度法、网格法以及模型法等聚类算法：

(1) **基于层次的方法** 基于层次的方法是通过构造一个结构对已知数据集进行分解，这一方法是聚类分析在现实中被广泛的应用。按照分类的不同可分为两种类型：凝聚式和分裂式，是按照自下而上的或自上而下的形式区分。凝聚法是自下向上的方式，根据某种规则，将所以记录记做一簇，相近的合并直至满足预期效果，一般的层次聚类方法归属于此类。分裂式是自上向下的方式，将全部数据集记为一个整体，不断的迭代细分为若干个子类，直至满足预期效果求得最终的结果。

层次的聚类算法缺点是：聚类过程中计算复杂度高，难以确定聚类的类簇数目，并且类簇分裂和合并的过程是不可逆转的。应用中一般把这类算法与其他算法结合使用，用来减少聚类的计算复杂性和迭代次数。常见的经典算法有 BIRCH、CURE、ROCK 等聚类算法^[40]。

(2) **基于划分的方法** 划分法首先把一个包括 n 个样本点的数据集初始划分成 k 个簇，然后按照特定的评判规则以及循环迭代的方式，不断定位每个簇中的样本点。最后的结果是每个样本点有且仅有唯一的一个归属簇，且每个簇中样本点个数在一个以上。

划分的聚类方法的优点是：在大型数据集上效果较好，能够发现球形和类球形结构的类簇，并且收敛速度比较快；其缺点是：需要凭借用户经验确定 k 个聚类数，孤立点或噪声点对结果会产生很大干扰。常见的划分技术方法有 K-means, K-medoids 等聚类方式^[41]。它们之间的差异是目的准则函数的含义以及划分的方式不同。之后的文段将对 K-means 算法有所介绍。

(3) **基于密度的方法** 基于密度方法就是通过分析样本的密度来识别簇，原理是判定各个相邻样本的数量是否大于假设的阈值，要是等于或是大于阈值就被认定此样本的旁边有足够的密度。

基于密度的聚类方法的优点是：能辨别任何形状布局的数据分布，能够发现全部的非球状的簇，还能够用来处理聚类受噪声数据、孤独点影响的缺点；其缺点是：计算复杂，很难处理离散数据集聚类，并且，对于阈值假设需要很

高的技巧。常用的算法有 DBSCAN, OPTICS, DENCLUE 等聚类算法^[42]。

(4) **基于网格的方法** 基于网格的聚类方法本质上以空间单元作为基础，其思想是把给定的数据集进行量化，根据空间的网格结构特性进行分解，形成有限数目的网格单元，并以此实现聚类的操作。

基于网格聚类方法优点是：能够有效的解决大规模以及分布繁琐的数据集并且解决的速度快，可以解决多维度的数据并且其结果不受孤立点干扰而且需要录入的参数少。其缺点是：精度难以保证。常用的方法有 STING 算法和 CLIQUE 算法等^[43]。

(5) **基于模型的方法** 基于模型的聚类方法，先是为各个类假设对应一个模型，接下来就是选取最佳数据拟合这一模型的过程。常用的方法包括 COBWEB 算法^[44]和 SOM 算法等。COBWEB 算法优点是：不需要手动输入聚类数就能够自动调整划分类的数目，缺点是不使用大规模数据进行聚类；SOM 算法优点是：抗噪性能强，可以辨别出向量空间中最优特征。其缺点是：联系权初值选择会直接决定结果的敛散性。聚类分析技术的一般分类如图 2-1 所示：

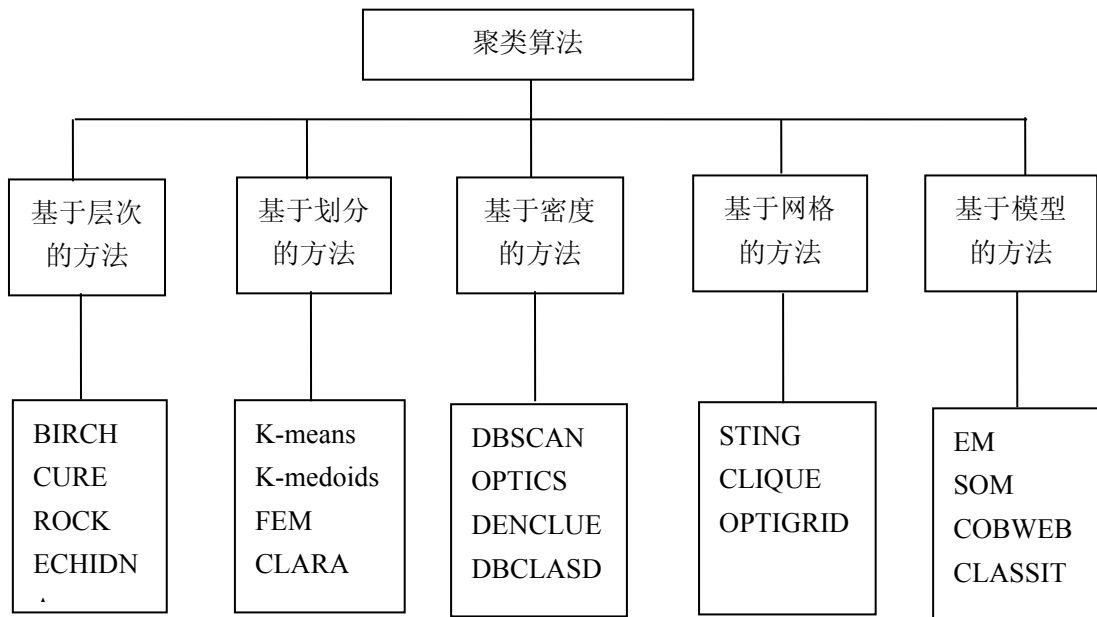


图 2-1 聚类算法分类

目前聚类分析技术通常会与不同领域交织应用，根据具体的问题具体分析，结合每种算法的优点交叉融合，规避不足更加准确的得到聚类结果，更好的解决实际问题。

2.3.3 K-means 算法

K-means 算法是根据距离的划分聚类算法,该方法是 1967 年由 MacQueen 提出并描述的,根据距离设定相似性的评价变量,即距离近的样本,其相似性就越大。这一算法已经趋于成熟,并适用于商贸领域、生物领域、图像技术、医疗数据等方面。

(1) K-means 算法的基本原理 K-means 算法是由用户指定类别数的大样本资料的逐步聚类分析。把样本资料视作 k 维空间上的点,以距离作为判断数据间远近关系的指标,并以浪费多个解为代价换取较高的执行率。K-means 算法的基本思想是:把每个样品聚集到其最近形心(均值)类中去,即先对数据进行初始分类,然后逐步调整,得到最终的分类结果。

(2) K-means 算法实现的步骤评述

1) 需要进行聚类的数据集 $S = (x_1, x_2, \dots, x_n)$, k 个聚类中心为 (c_1, c_1, \dots, c_k) 。

K-means 算法的两个体 (X_i, X_j) 欧式距离是两个体 K 个变量值之差的平方和的平方根算式为:

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (2-1)$$

其中 $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, $x_j = (x_{j1}, x_{j2}, \dots, x_{jp})$, 表示 2 个 P 维属性的数据对象。

2) 全部样品点平均距离的算式如下:

$$Meandist(S) = \frac{2}{n(n-1)} \times \sum_{i \neq j, i, j=1}^n d(x_i, x_j) \quad (2-2)$$

其中 n 为数据集样本对象总数, $d(x_i, x_j)$ 为样本点 x_i 和 x_j 的欧氏距离。

3) 目标函数为平方误差准则函数算式如下:

$$E = \sum_{i=1}^k \sum_{j \in N_i} \|x_j - c_i\|^2 \quad (2-3)$$

N_i 代表是第 i 个簇集合, c_i 代表是第 i 个簇的中心。

K-means 算法的步骤可以简单描述如下:

第一步,选取 k 个样本作为初始中心点,或是将所有的样本分为 K 个初始类,接着把这 k 个均值当作初始的中心点。

第二步,除去中心点以外的其它样本进行归类,把各个样本归纳到中心点距离最近的那个类中,此类的中心点会被更新,将成为当前的均值,直到全部样本都归类完毕。

第三步，重复步骤第二步，直至满足终止条件为止。

终止条件有两个一个是根据迭代次数；另一个是根据类中心点偏移程度。两个终止条件满足一个就可以终止迭代。

经典的 K-means 算法流程图如下图 2-2 所示：

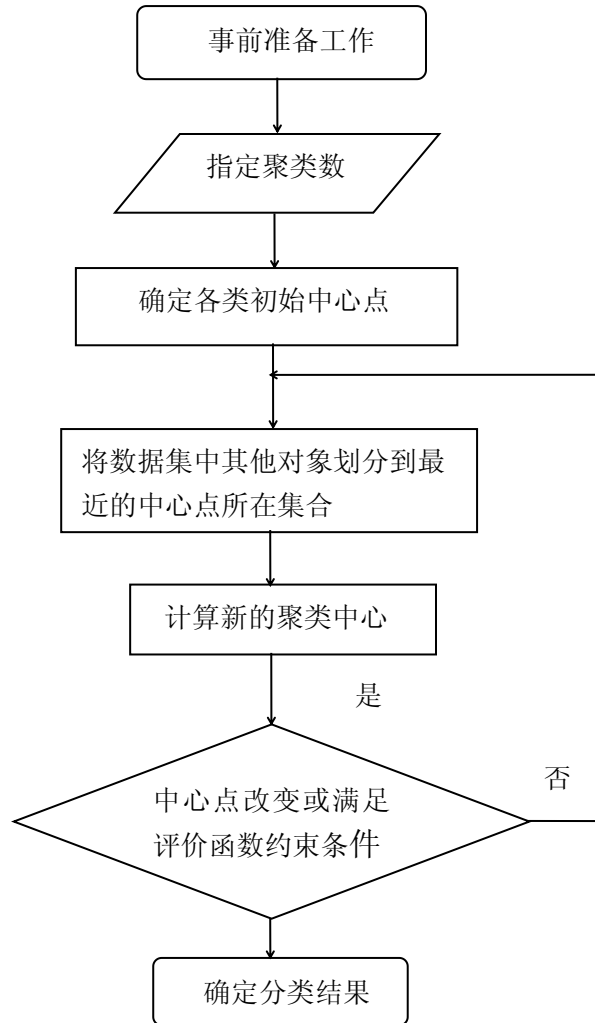


图 2-2 经典 K-means 算法流程图

(3) K-means 算法的适用性 K-means 算法的优点如下：

1) 具有应用广泛性

现实世界都是多维的，用某一种指标不能够对其分类，而 K-means 算法就是分析多维指标的，利用相似性度量转化成样本间的相似性，从而进行聚类，该算法广泛适用于商贸领域、生物领域、图像技术、医疗数据、统计学等很多领域。

2) 算法简单、高效

K-means 算法结构简单, 解决大数据集时可伸缩性和有效性较好, 并且结果容易理解, 算法对数值型数据的聚类效果好, 因为一般多选取欧式距离作为其相似性度量, 所有所得结果的几何意义清楚明确。

3) 算法客观、现实

K-means 算法处理数值型数据集时, 对数据集的范畴没有特殊要求, 同样是按照样本间的相似性实现分类, 而不是根据所谓经验划分选择 k 值, 并且所得结果和数据的录入次序没有关系。

K-means 算法的缺点如下:

1) 由于对分析的领域具体信息把握不足, 不能选取最为合理的 K 个初始凝聚点, 故而在执行 K-means 算法时输出结果会产生偏差。

2) K-means 算法对于独立的样本使得凝聚重心产生偏移, 因此受到噪声、边缘点以及孤立点影响, 对输出结果的稳定性和精准度会有所干扰。

3) 迭代次数等于指定的迭代次数终止聚类, 不确定是否能得到合适的收敛性, 迭代所形成的类重心点的最大偏移量小于指定量, 输出的结果有时不满足终止条件, 样本类会不断调整, 导致结果不是最优解。

2.4 本章小节

本章是对相关概念进行分析梳理, 为后续章节的讨论提供理论基础。首先介绍了客户细分的概念, 客户细分原则总结了客户细分指标, 并且针对不同的指标特点以及局限性进行分析, 根据本文所要研究的房地产行业的特殊性, 选取客户价值指标变量以及客户忠诚度指标变量, 来表述房地产客户细分。接着介绍了客户细分技术方法原理, 并对应用情况以及优缺点进行评述, 通过对客户细分技术分析比较, 结合本文研究需求的精度以及此类算法具有操作简单、速度快、结果容易解释等优点选取聚类分析技术进行应用, 最后详细介绍聚类分析技术原理以及使用条件, 最后选取 K-means 算法作为后续解决实际问题的方法。

第3章 房地产客户细分模型的构建

对于客户细分，首先要确定细分目的，其次才可以根据具体的业务确定一套完整的指标，然后选择相对应的变量进行数据准备，形成一个分析变量表，最后才可以根据指标变量进行聚类分析，因此客户细分指标体系的设计对聚类结果显得尤为重要。

3.1 指标体系构建原则及步骤

3.1.1 指标体系构建原则

构建指标体系时，目的不同选取的指标体系也不一样，在构建中不仅要考虑到单个指标的科学、真实、正确、可行，也要确保指标体系各个指标相互协调，以及全部指标的必要和齐备。因此选取指标时本文按照如下原则：

（1）**合理性原则** 选取的指标不仅有代表性的全面、真实的反映客户所有的特点，而且必须保证指标之间没有重叠、交叉、冗余等，故而，合理的选取完整和不交叉指标是评价指标体系构建的主要原则。

（2）**可获取性原则** 由于客户细分指标需要相应数据的支撑，如果有些指标的相关资料很难获得甚至无法获得，则该指标是没有意义的。因此，在选择指标的时，要保证数据是可够获取的。

（3）**可操作性原则** 应有现实用处的细分指标才可以被企业所利用，因此在设置指标体系时，不仅需要避免指标过于复杂，而且还要考虑到其可操作性，能否量化设定的指标，尽可能的选择简单易操作的指标。

（4）**定性与定量相结合原则** 选取指标时要尽量概况的选取定性指标和定量指标，二者相结合是非常有必要的。定性指标很难量化，但是在客户细分时不可以或缺，而且不可避免的描述性指标。具体定量指标难以量化时，可以选取近似替代处理。

3.1.2 指标体系构建步骤

能否得到满意的客户细分结果，不仅在于选择合适的分类方法上，而且与细分变量的选择和体系的构建密不可分。构建指标体系具体有如下几步：

（1）**明确研究问题** 构建指标体系前，必须清楚的明确研究的问题，了解与其相关的知识和理清相关概念。

(2) **筛选指标** 构建指标是一个逐步深化的过程，首先尽量充分的收集指标，概况研究问题，但是这些指标有可能重叠或是相关性高，过多的重叠会导致失真，因此可以适当的修正并筛选有代表性、可操作的指标，例如主成分分析法就是常用的筛选方法之一。

(3) **优化应用指标体系** 在整体思考指标体系的基础之上，把分散的指标组成相异的类别用来代表整体指标的相异特征。优化过后的体系需要通过实践检验指标体系是否合理，并且对不合理的指标进行修正。

3.2 房地产客户细分指标分析

与其他产品相比，房地产有其特有的性质如：消费区域性、长期使用性、投资额度大、具有风险性、增值性等。这些特性与其他消费产品不同，尤其长期使用性和消费区域的特性，购买频数低且交易时间长等特点决定了房地产客户细分指标与其他行业的细分指标不同。

之前章节已经介绍了，本文通过聚类分析算法对房地产企业客户进行细分，基于客户价值与客户忠诚度两个角度，构建客户细分指标体系，在整个营造房地产过程落实“以客户为中心”的管理理念，以客户需求作为决策、计划的出发点，立足点，提升房地产企业的口碑。

3.2.1 基于客户价值细分指标体系设计

客户被定义为：一个你满足某一需求且在需求方面你能使其感觉到愉悦的人^[45]。客户价值被定义为消费者通过购买商品得到的收益和花费的代价的差值，用来衡量客户对企业绩效的综合性指标，企业对客户价值的考察可以从潜在、知觉、实际实现等层面价值分析。事实已经印证了，基于客户价值的营销工作，使得销售成果的提升和完善起着举足轻重的作用^[46]。

本文针对企业为主体，客户为客体的研究方法评述，即站在企业的立场上分析客户的价值，企业为价值感受主体，客户为客体。客户的价值有助于进行精准的用户评价，利用其结果细分客户群，用来合理的利用企业的资源。本文设计客户价值指标体系，综合两个方面考虑：

(1) **客户当前价值** 客户的当前价值从直接计算的角度上主要计算实际的净利润大小，主要是基于成本分析基础上的进行的盈利核算，可以计算但是过程复杂操作比较难；从间接角度上评述，需要提出间接变量来表述，很多学者都探索各自的间接指标变量，归纳文献材料列举如下几种指标：购买频率、购买量、净利润、服务成本等。间接指标简单容易理解，但是需要结合实际问

题以及研究目的科学的选取适当的指标。

(2) **客户潜在价值** 客户潜在价值主要计算是选取一些消费者的特殊表述变量,预测以后阶段的价值潜在变化。很多学者都寻求各自的间接指标变量,本文通过对相关文献资料总结归纳,列举如下几种常用的指标:客户成长、还款及时度、推荐亲朋购买、信用等级等。变量的选取根据指标选取的规则,结合实际问题科学的选取。

3.2.2 基于客户忠诚度细分指标体系设计

目前伴随着房地产市场激烈的竞争,客户消费日趋理性,要想在同质化的市场获取一定的席位,赢取客户并且提高客户的忠诚度,已经是每个房地产企业追求的目标,并且客户需求的多样化以及个性化,使得住宅消费呈现了多样化、个性化的发展趋势,其原因是忠诚客户给企业带来的价值有着给企业降低成本、提高溢价、广泛的传播等益处,进而赢得更持久的收益^[47]。

客户忠诚度,是指客户热衷于消费某一品牌或是某一类产品,反复的购买这种品牌的产品以至于具有“依附性”。忠诚客户体现为:首先是通过重复消费、推荐亲朋购买、交叉购买,为企业创造更多利润;有了客忠诚度,客户口碑明显,客户对价格敏感度就会降低并且可以牢牢的把握住客户,减少了客户流失,另外有良好口碑的企业,客户对此类企业的错误会报以宽容,降低了服务的成本。客户忠诚度可以体现在行为和情感二个维度上。

(1) **行为忠诚度** 行为忠诚度体现的是客户多次重复购买、交叉购买、推荐该企业产品或服务给自己的亲朋行为,例如,根据 2015 年的统计数据表明,购房业主有 54.4%在购房时受到亲朋好友推荐,有 10.6%重复购买,并贡献了 60 亿左右的销售额。

有学者认为行为忠诚度是:通过客户的重复购买体现的。本文通过相关文献资料的归纳总结,得到以下常用指标变量描述:客户保留、推荐购买、向上购买、购买次数、购买倾向、购买份额、重复购买等。

(2) **情感忠诚度** 情感忠诚度体现的是对企业产品或是服务的偏好以及双方关系的投入程度,具体表现为价格的敏感度,例如,每次 IPHONE 新发布手机时无论价钱多贵,都有忠实的“果粉”在苹果商店排队等候购买。

有学者认为情感忠诚度主要通过价格承受度来度量,并且从定性和定量两方面研究情感忠诚度的描述,希望可以表达行为忠诚度指标体系,本文通过相关文献资料的归纳总结,得到以下常用指标变量描述:满意度、信任程度、客户关系程度、转换成本、价格的敏感度等。

3.2.3 基于客户价值和忠诚度构建房地产客户细分指标体系

客户忠诚度是对客户价值能够实现的保证，同时扩充了主观感知指标变量。只有高价值的消费者才会提升到忠诚客户这一层面，才能够推荐亲朋购买，也只有这样客户的潜在价值才能够转化成为企业的利润指标。上一节介绍的客户价值与客户忠诚度指标变量中有几个指标变量部分出现重叠或是交叉，因此在本文构建指标体系时候将重叠的指标加以融合。此外，构建房地产客户细分指标体系时根据购房客户与其他消费客户不同的针对性指标，例如，由于房地产消费是刚性需求产品并且高额消费的商品，根据消费者的生命周期理论要评估自己未来的购买能力，考虑自己能否完成购买；又例如购房客户是按揭贷款的客户，需要考虑客户的银行信用等级以及评价其还款能力。房地产作为特殊的商品有其特殊性但同一般商品相比也存在同样的属性，例如购买金额、总额、次数、利润等指标。

本文参考国内外学者的细分指标体系基本框架并且根据购房客户消费特点以及之前章节归纳分析的指标变量，本文构建了适用于房地产企业的客户细分指标体系，包括 3 个一级指标，6 个二级指标以及 14 个三级指标。如下图 3-1 所示：

由于单纯采取定性或定量方式来评估房地产客户细分指标会有一定的局限性，因此，本文构建的客户细分指标体系包括定性与定量两种方法，下面对各个指标具体解释如下：

3.2.3.1 客户当前价值

（1）交易额指标 本文设定的交易额指标包含购买金额和总额 2 个变量。

购买金额是消费者最后一次交易的购房金额，即从消费者带来的销售收入的直接效益。

购买总额指的是客户到目前为止在房地产企业购房花销的总金额。

这二个指标是用过去的消费情况来预测未来消费能力，并且假设过去交易额高，未来的交易额也会高^[48]。

（2）利润指标 企业是以盈利为目的的，需要获取利润才能够生存和发展，因此利润指标是最为重要的指标之一。本文利润指标包括毛利润和关系成本 2 个变量。

毛利润是总利润扣除所有税费后企业的剩余利润，数值越高，客户的当前的价值就越高，企业都希望追求高额的利润。

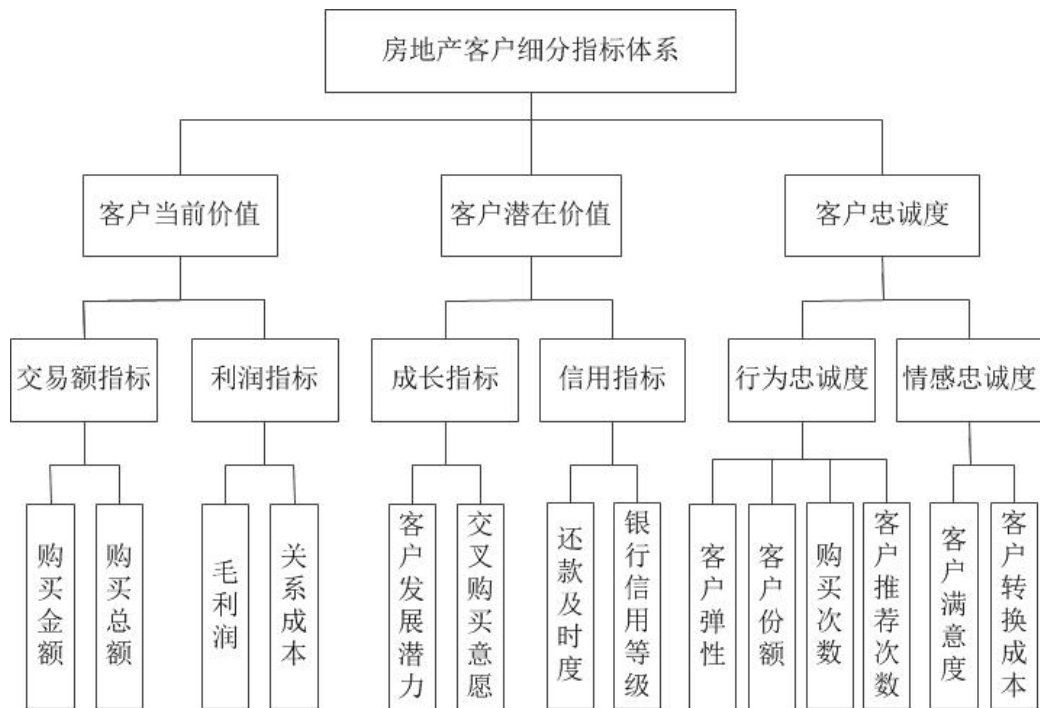


图 3-1 房地产客户细分指标体系

关系成本是指为了获取、维持、发展客户所消耗的成本。关系成本是由交易成本细化而来。关系成本的高低与客户当前价值成反比，和企业获取的净利润也成反比。

3.2.3.2 客户潜在价值

之前章节介绍过房地产作为特殊的商品，与其产品不同，由于现阶段房价高，还有相当多一部分人没有能力购买房产，重复购买的更是难上加难，目前购房主要是刚性需求为主，也有部分客户处于投资或是改善住房等目的的二次或多次购买。

(1) 成长指标 本文成长指标包括客户发展潜力、交叉购买意愿 2 个指标变量。

客户发展潜力是指随着时间的推移，客户自身的发展会积累一定的经济实力，随之而来会有更多的资金购买房地产企业的产品或服务。客户发展潜力高低与其潜在价值影响成正比。

交叉购买是指已经购房的业主有投资该企业商铺的想法，或者是租赁商户有选购住宅的想法等。即商业地产（购物）与住宅地产（购房）的客户互动。

(2) 信用指标 房地产作为高消费的商品，大多数购买房产客户都是

通过商业贷款或是公积金贷款等按揭贷款的方式支付房款,因此在银行信用指标这一方面考虑客户的潜在价值尤为重要。

本文信用指标包括银行的信用等级和还款及时度 2 个指标变量。

银行信用等级从偿还、收益能力,经营状况、履行合同约定情况与未来潜力四个方面进行评价,本文主要考虑的客户银行不良信用记录多少,银行信用等级低,客户潜在价值就低。

随着我国个人征信体系不断铺开和成熟,个人征信体系不仅可以维持金融平稳规避金融风险,也能够推动全社会的信用体系的建立。还款及时度是指客户按揭还款或是使用信用卡的情况下,客户未及时还款的次数,会降低个人征信的评定,会影响购房贷款,客户的潜在价值就低。

3.2.3.3 客户忠诚度

客户的忠诚度是指消费者重复购买同一品牌产品,这种忠诚是非理性的,按照程度的深浅分为情感忠诚和行为忠诚。情感忠诚是一直使用该产品并且始终保持满意对此类产品所产生的偏好。行为忠诚是指用户对于使用的产品愿意克服阻挠去消费的行为。

(1) **情感忠诚度** 本文情感忠诚度包括客户满意度、客户转换成本 2 个指标变量。

客户满意度是客户通过使用产品后的感受与使用前的期望程度的对比指数。主要通过企业的满意度调查以及投诉建议等方式。一般对企业满意度高,对企业的投诉以及建议就会减少,反映出客户的忠诚度就会高。

客户转换成本是迈克·波特所提出的,是指当客户从之前商品供应者转到其他的商品供应者时所产生的成本^[49]。房企利用广告宣传其楼盘的与众不同,使购房消费者意识到转换购买其他楼盘单位所消耗的转换成本会很高,就会放弃购买本公司的楼盘的代价就大,反而客户的忠诚度就高。

(2) **行为忠诚度** 本文行为忠诚度包括客户弹性、客户份额、购买次数、客户推荐次数 4 个指标变量。

客户的弹性是指购房业主对该房企品牌和产品的推荐宣传以及对该房企所给出的合理化建议,这是业主对房企的无形的贡献。客户的弹性与客户的行为忠诚度成正比。

客户份额是指购房业主购买该房企住宅以及商业地产的数量占据同类性地产总数的比值,其中客户的份额与客户忠诚度成正比。

购买次数是指业主购买该房企楼盘物业的总次数。购买的次数多少与客户的忠诚度成正比。

客户推荐次数是指具有良好的口碑房企,业主对该企业非常满意会将亲朋好友等潜在的客户推荐购买该房企的楼盘,推荐次数越多,给企业带来的新客户越多,被推荐的客户满意企业的产品会继续推荐客户,因此客户忠诚度就越高。

3.3 客户细分模型设计

进入互联网时代后,产品和服务的日益信息化和数字化,各种数据信息指数型增长,在处理实际问题时候,我们得到的大量的数据,知识提取技术帮助我们分析海量的数据,并且可以将数据中相关联的和隐藏的信息转换成为易于观察的简单数据,借此做到成功的决策。一般情况下我们收到的数据都是有噪音的、不一致的、重复冗余的甚至错误的数据,这些数据会降低我们算法的执行效率,甚至导致任务的失败,因此我们在需要对原始数据进行处理,在不损失数据信息的情况下,改进数据的质量,进行噪音消除,减少不相关或是具有误导性的数据,降低算法运行时间。

3.3.1 数据预处理

大多数资料都是把数据预处理分为数据的清理、集成、变换和约简四个部分^[50]。数据预处理目的,主要是清除原始数据的缺损、噪音等问题,进而把分散的数据源划分到统一的集合里,这种形式称之为数据的集成,在数据集成过程中会产生模式不一致、冲突等问题,这时候需要对数据进行变换。最后数据集大小直接影响后期任务的进行,所以最好寻找一个够用的较小的数据集来大幅度加快效率。

(1) **数据清理** 数据清理就是清除数据中的错误,处理信息不一致的问题,比如预测填充缺失数据值,平滑噪音,寻找并剔除孤立点等^[51]。数据清理任务是解决缺失值、噪音和反复记录三个问题。

1) 我们的数据中难免会出现一些数据的缺失现象,缺失值是指属性上的空缺记录。主要有三个原因:自身所在属性完全独立其他完整的属性的完全随机缺失、与完整属性有关联可以通过其他属性预测的随机缺失、无法通过其他属性预测只与缺失属性本身有关的非随机缺失。

2) 噪音数据可能与处理任务不相关或是没有意义的字符,也可能是数据集中出现的错误。现实的数据往往都是有噪音的,根据文献总结常用的消除方法有均值光滑、中值光滑、边界值光滑、回归以及聚类等方法。

3) 由于数据记录在结构上和词法上有着不同的形式,因此同一个数据在

其他的数据源上可能会有不同的表达方式。这会加大我们的匹配难度和精准性。而重复记录检测是从数据源中识别多个形式相异，但现实中表示同一属性的记录。

(2) 数据集成 数据集成是将多个数据源中的数据结合起来，存放一致的数据存储中的过程，其目的就是为了给用户提供一个统一的视图，将不同的数据源中的数据进行整合。数据集成方法有二种：1) 物理集成，将提取不同的数据源中信息，合并成为统一数据源；2) 逻辑集成，在需要时候抽取，不改变数据的物理位置，知识提供给一个虚拟的全局视图。

(3) 数据变化 数据源中的信息可能存在类型的不同或是格式的差异，为了满足算法的要求而进行的数据变换。数据变换主要方法有以下几种：1) 数据规范化，通过比例缩放的方式将数据值映射到一个宽带较小的空间中的转换过程；2) 数据聚集，为了使数据表示的粒度与实际的需求相符合，我们需要对数据进行聚集；3) 属性构造，是这样一种操作，当我们需要一个新属性时，在数据源中这个属性已经有某个属性生成；4) 离散化，有些分析技术只能解决离散类型数据，因此我们需要离散化算法将其转换。

(4) 数据约简 数据集大小直接影响后期任务的进行，所以最好寻找一个够用的较小的数据集来大幅度加快效率，这个步骤就是数据约简，包括以下几个方法：1) 实例约简，选取满足要求相适应的实例子集，用来约简数据量，例如随机抽样；2) 属性简约，从数据集中选取和任务相关的数据，减少信息量；3) 属性值约简，选取小于原始数据量的数据进行替代，例如聚类、回归和下节要介绍的本文需要用到的主成分分析法等。

3.3.2 主成分分析

主成分分析法是一种降维的线性特征约简的方法。以失去少量信息为基础，用几个综合性指标代表原始指标变量的多元统计方法^[52]。把综合指标称为主成分，并且主成分均是原始指标变量的线性组合，且主成分间没有相关性，这就说明主成分比原始数据变量具有很多优良的特征。因此在处理复杂数据信息时，就能够只用几个主成分信息代表所有的原始信息而不损失过多的原始信息。从而可以把握主要的矛盾，使问题变得简单，提升运算效率。

主成分分析法的具体步骤如下：

(1) 将原始数据标准化 原始数据集有 n 个样本， p 个指标，则样本矩阵为：

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix} \quad (3-1)$$

$$\text{标准化公式为: } Z_{ij} = \frac{X_{ij} - \overline{X_j}}{S_j} \quad (i=1,2,\dots,n; j=1,2,\dots,p) \quad (3-2)$$

其中, X_{ij} 表示第 i 个样本的第 j 个指标值; $\overline{X_j}$ 表示 n 个样本的第 j 个指标的平均值; S_j 表示为第 j 个指标的标准差; Z_{ij} 表示处理完之后第 i 个样本的第 j 个指标值。

(2) 标准化处理完, p 个变量的相关系数矩阵计算:

$$R = (r_{ij})_{p \times p}, \quad r_{ij} = \frac{T_{ij}}{\sqrt{T_{ii} \times T_{jj}}}, \quad (i,j=1,2,\dots,p) \quad (3-3)$$

式中, T_{ij} 被称为 X_i 与 X_j 的协方差。

(3) 相关矩阵 R 的特征值、向量的计算 其中, 特征值记为

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$$

特征向量记为 $\mu_1, \mu_2, \dots, \mu_p$, 且 $\mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{ip})$, $i=1,2,\dots,p$ (3-4)

(4) 各个主成分的方差贡献率 α_k 和累计方差贡献率 $\alpha(k)$ 的计算 第 k 个主成分 F_k 的方差贡献率为

$$\alpha = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i}$$

$$\text{累计方差贡献率为: } \alpha_k = \frac{\sum_{j=1}^k \lambda_j}{\sum_{i=1}^p \lambda_i} \quad (3-5)$$

表示前 k 个主成分累计提取的信息占全部信息的百分比。因为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, 说明第一个主成分贡献率是大于其他主成分贡献率的, 依次递减直到最后一个。通常情况下累计贡献率大于 85%, 就可以说明前 k 个主成分提供了原始变量的大部分信息。

(5) 主成分意义 因为各个主成分均是原数据的线性组合, 故而现实

意义不够明确，需要结合专业知识对其赋予相应的名称，并且需要将蕴涵的信息给予适当的解释。

3.3.3 房地产客户细分模型

本文 2.3 节已经评述了常用的客户细分算法，对比并结合房地产行业的特殊性，本文选取聚类分析算法，该算法能够适应房地产客户细分准确性的要求，加之此类算法易操作、运行快、结果易理解等特点。之后 2.4 节详细评述了常用的聚类分析算法，综合考虑各种算法的恰当性和操作性等方面，本文选取了聚类分析技术中划分聚类中的 K-means 算法，并且在 2.4.3 节对该算法进行评述。结合以上章节，本文构建的房地产客户细分模型如图 3-2 所示

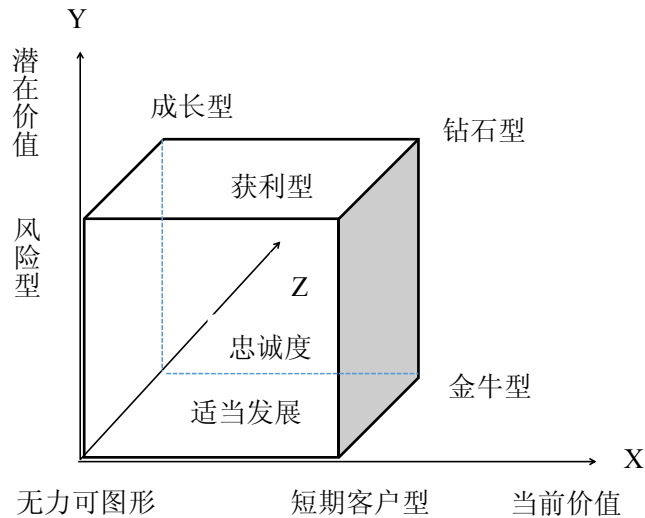


图 3-2 房地产客户细分模型

本文建立的房地产客户细分模型是三维的客户细分模型，不同于其他单一的基于生命周期或是客户价值的房地产模型，X轴表示客户当前价值，Y轴表示客户潜在价值，Z轴表示客户忠诚度，其中高X值、高Y值、高Z值的客户被定义为钻石型客户；高X值、低Y值、高Z值的客户被定义为金牛型客户；低X值、高Y值、高Z值的客户被定义为成长型客户；高X值、高Y值、低Z值客户被定义为获利型客户；低X值、低Y值、高Z值的客户被定义为适当发展型客户；高X值、低Y值、低Z值的客户被定义为短期客户型客户；低X值、高Y值、低Z值的客户被定义为风险型客户；低X值、低Y值、低Z值的客户被定义为无利可图型客户。可以根据客户细分模型划分客户属于哪

一类，进行针对性资源分配和客户管理策略。

根据本章前面所述的细分指标以及获取数据后的预处理等方面内容，本文研究的房地产客户细分流程图如图 3-3 所示。

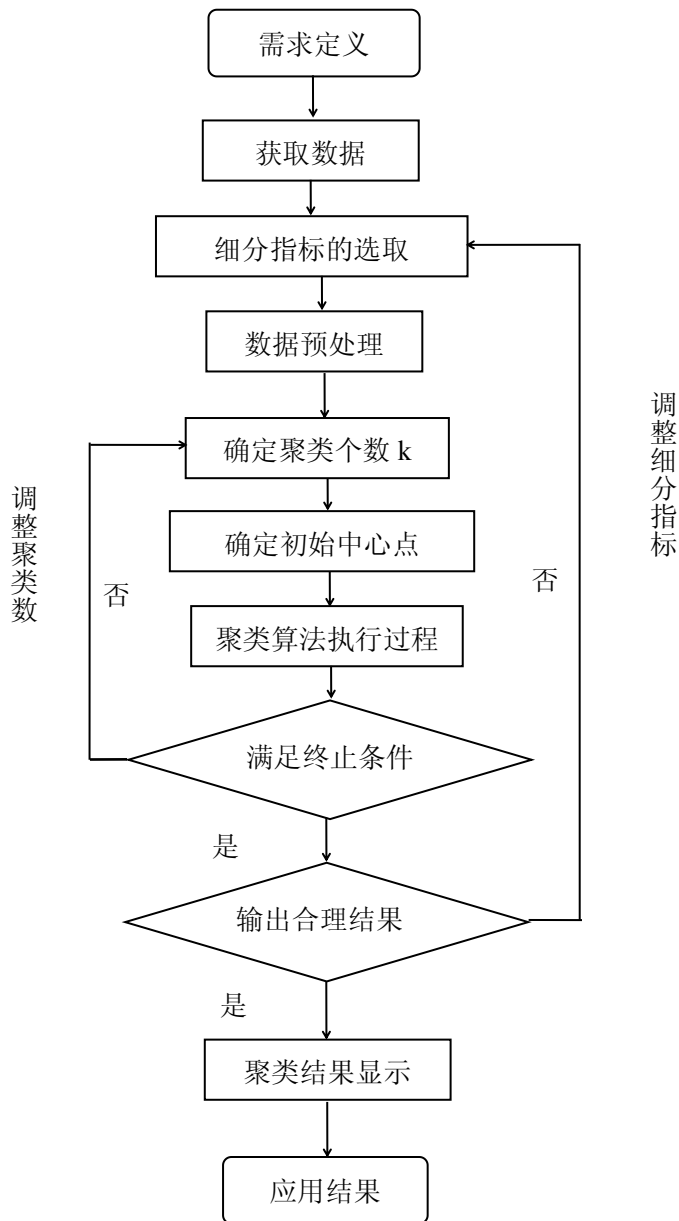


图 3-3 房地产客户细分流程图

需求定义是根据不同的业务需求，结合实际问题具体问题具体分析；获取数据是根据实际任务需求在数据库选取相关的指标；根据指标选取原则并且结合实际需求选取合适的指标；数据预处理以及算法等都是之前章节介绍过；应用结果是根据 K-means 算法得出的结果进行具体的执行营销策略、客户保留

策略、资源配置策略等。从上述客户细分流程图可以看出，根据聚类分析的 K-means 算法的客户细分是一个多次迭代不断优化的过程，根据实际情况调整变量和聚类数，反复迭代，直至满足聚类分析终止条件获得最优的能够指导实践的结果。

3.4 本章小结

本章主要是构建房地产客户细分模型。首先介绍了指标体系的选取原则和指标体系构建步骤，接着分析了房地产客户细分的指标，选取了客户当前价值、客户潜在价值、客户忠诚度三个指标进行评述，然后评述了数据预处理的各种方法和主成分分析法，给出了房地产客户细分的流程图。最后，构建房地产客户细分模型。

第 4 章 房地产客户细分实证分析

本章主要是通过第三章所提出的房地产客户细分模型为基础,将使用到宝宇地产客户细分实证应用中,用来验证第三章设计模型的可行性和有效性。在实际操作中,选择 SPSS (统计产品与服务解决方案)统计软件处理获取的客户数据信息以及执行 K-means 算法,SPSS 软件易操作、易编程、多模块组合、能够读取以及能够输出多种格式的文件等特点应用十分广泛,与 SAS 和 BMDP 一起被称为三大统计软件,目前已经更新到 IBM SPSS Statistics 24.0 版本。

4.1 宝宇地产的企业现状评述

4.1.1 宝宇地产的企业现状

成立于 2001 年的黑龙江宝宇房地产开发(集团)公司,凭借本地市场资源和行政资源优势,打造了“荣耀上城”、“荣耀天地”、“宝宇天邑”和即将开盘销售“宝宇环球港”等多个高端地产项目,铸就了宝宇地产的品牌,在冰城有着良好的口碑,并在 2014 年成为黑龙江唯一入围中国地产 500 强企业。

在哈尔滨购买宝宇地产楼盘的客户大多属于中高端客户群,宝宇地产楼盘基本都是沿着松花江江边而建,一般属于观江房,因此与同期同区域所售楼盘相比宝宇地产楼盘通常都比竞争对手贵,购买宝宇地产的客户大多都是冲着宝宇地产口碑以及感动式物业服务而来。

目前宝宇地产客户细分培训时主要是按照客户购买心理分析:①理智型,对楼盘、开发商、宣传推广方面比较了解,购房者有主见,销售人员需要耐心做到“百问不厌”;②疑虑型,挑选商品时候小心谨慎,“三思后行”,销售人员介绍房屋时候要耐心、细致,切记热情过度;③习惯型,对有实力公司或宝宇地产楼盘有较高热衷性购买,销售人员应特别注意、尊重、想方设法满足他们的习惯要求;④价格型,分为低价偏好和高价偏好两类,实力地产的客户基本归为后者,他们重视楼盘内在质量和外观、建筑和周边环境的协调美。但是实际操作时主要是按照职业或是客户来源途径主观的进行分类。

4.1.2 宝宇地产 SWOT 分析

使用 SWOT 分析法可以明确企业自身的 strength(竞争优势)、weaknesses(竞争劣势)、opportunities(机会)和 threats(威胁),是把企业战略和企业内在资源与外在环境结合的战略分析方法。运用 SWOT 分析法可以帮助公

司认识到目前形势，抓住机遇，锻炼和提升公司的核心竞争力^[53]。本文采用该方法分析宝宇地产营销市场进如图 4-1 所示。

<p>竞争优势 S:</p> <ol style="list-style-type: none"> 1、本地市场资源优势 2、行政资源优势 3、本地口碑优势 4、融资渠道顺畅 5、高素质的领导者 	<p>竞争劣势 W:</p> <ol style="list-style-type: none"> 1、公司起步晚 2、人才缺失、薪酬待遇落后 3、企业文化缺失 4、培训观念错位
<p>机会 O:</p> <ol style="list-style-type: none"> 1、城市化款式发展 2、二三线城市地产市场前景好 3、异地项目顺利进行 4、企业的多元化发展 	<p>威胁 T:</p> <ol style="list-style-type: none"> 1、外来房企的注入 2、城市内土地资源稀缺 3、消费需求多样化

图 4-1 宝宇地产公司 SWOT 分析

受到外来房企的竞争以及哈尔滨市区内土地资源稀缺的威胁，加之企业没有开拓省外市场，以及缺失正规的企业管理以及工程人才，对企业的发展会产生一定的影响。总结宝宇地产营销管理中的问题有：

（1）缺乏有效的客户细分 宝宇地产从公司成立至今已经拥有了庞大的客户数量，积累众多的客户信息，并且在哈尔滨地区有良好的声誉，但是如何从这些客户信息中挖掘出对宝宇地产有价值的信息呢？如何保留住老业主以及发展新业主呢？这个问题对于营销部门仍然很难回答，原因就在于一直没有充分利用这些信息对客户进行有效的分类。目前在宝宇地产售楼处接待客户的置业顾问仅是凭借客户的工作单位、客户学历、客户薪资水平等属性粗略的把客户的区分为尊贵客户、一般客户、意向客户等，这种客户分类方法在一定程度上可以区分客户的消费水平和购买能力，但是准确性不高，依据这种方法制定的营销策略时往往对销售人员带来误判，例如，被认为是尊贵客户的有钱人却一直没有购买过宝宇的房产，反而消耗了公司的资源，也浪费了置业顾问的精力。

（2）缺少对客户流失的分析 随着房地产行业间的竞争，客户可选择的楼盘很多，所以房企都注重客户的满意度和忠诚度，希望通过产品和服务的提升，改善企业形象维系更多的客户，对于高消费的房地产商品，购房者都非

常谨慎，对于房产的质量、价格以及后期的维护和物业服务都有很高的要求，一旦某个环节服务不到位，客户虽然不能立即换房，但是会直接影响到购房业主的口碑，会对即将要购买的业主一个负面评价，将会流失部分客户，据权威统计显示，企业受客户盈利能力及保持率影响，客户保持率每增长 5%，行业的平均利润率就会增加 25%-50%之间，即开发新用户的成本远大于维护原有客户成本，但是对于房地产客户流失现象，宝宇地产一直以来也没有采用有效的技术手段来分析客户流失的正真原因、流失的行为特征、流失去向以及客户结构。因此对于客户的全面分析，针对不同的客户投其所好，制定相应的营销策略才是宝宇地产预防客户流失的手段。

(3) **差异性服务不到位** 根据马特莱法则——企业利润 80%是由 20%有价值的客户带来的，购买 200 平米房产和购买 30 平米房产对于企业现金回流和利润率一定是不同的，因此宝宇地产应该注意优质客户的积累，根据细分的客户群提供差异性服务。虽然，在售楼处有 VIP 接待室，由销售经理直接接待高价值的客户，但是售楼处缺乏明显标识，销售人员的引导和推荐力度不够，对于高端客户没有体现到差异性服务，

本人主要是研究房地产客户细分，对其他方面内容不加以评述，并且建立一套完整体系，用来收集、存储客户数据信息，但是该企业对于客户分类仅根据购房者的购买金额、购买总额、购买次数等指标粗略的把客户的区分为尊贵客户、一般客户、意向客户，这种分类的方法只是分析了客户的当前价值，没有深入的察觉其潜在价值以及忠诚度。因此本文利用在 3.2.3 节介绍的房地产客户细分指标，通过将宝宇地产实际客户数据带入模型中进行客户细分，希望达到以下目的：1) 分析细分结果基础上，用来检验房地产客户细分指标体系是否实用以及高效；2) 按照客户细分的类型不同，制定差异的资源分配和客户管理策略。

4.2 数据预处理

4.2.1 数据准备

本文在宝宇房地产客户群中利用问卷星提供技术支持收集的有效问卷中的数据进行描述性统计分析，抽取出 40 个客户信息作为原始的数据，结合 3.2.3 节所选取的细分指标，从而生成一个 40×14 的数据矩阵，矩阵其中的属性字段说明见表 4-1 所示。

表 4-1 属性字段说明

字段	说明	字段	说明
C11	购买金额（万元）	C42	还款及时度
C12	购买总额（万元）	C34	推荐客户次数
C21	毛利润（万元）	D11	客户弹性
C22	关系成本	D12	购买次数
C31	客户发展潜力	D13	购买份额（%）
C33	交叉购买	D21	客户满意度
C41	银行信用	D23	转换成本

4.2.2 主成分分析

之前 3.3.2 节已经介绍了主成分分析法是一种线性特征约简的方法，该方法够将有用的信息提取出来，将冗余的信息去除，从而达到降维的目的。本文使用统计学软件 SPSS19.0 版本对样本进行主成分分析。具体步骤如下：

（1）准备工作 在 SPSS 19.0 中打开数据文件，通过选择文件→打开→数据命令将数据调入到 SPSS 19.0 的工作文件窗口上，依次选择分析→降维→因子分析命令，打开因子分析主对话框。如图 4-2 所示。



图 4-2 变量选择界面

（2）指定参与分析的变量 在因子分析主对话框中，从左侧的变量表中选取参与分析的变量，添加到右边的变量框中。本文选取 14 个变量作为参与分析变量。

（3）运行主成分分析过程 首先，单击描述按钮，打开描述统计选择

相关矩阵下面的系数选项和 KMO 和 Bartlett 球型度检验,如图 4-3 所示;其次,单击抽取按钮,设置方法为主成分,选取输出选项碎石图,如图 4-4 所示。



图 4-3 描述统计



图 4-4 抽取-输出

(4) 主成分分析结果 结果是 14 个变量的相关系数矩阵,可看出 C11 变量与 C12、C21、C22、C31 相关性很高。相关矩阵如表 4-5 所示。

1) 结果 4-6 输出的是 KMO 检验和 Bartlett 球度检验结果。本文的 KMO 值是 0.708 大于 0.5,因此运行主成分分析是没有问题的,同时可以看出,Bartlett 球度检验统计量的观察值为 575.383,相应的 Sig 为 0.000.小于显著性水平 0.05,拒绝 Bartlett 球度检验的零假设,认为相关系数矩阵与单位矩阵有显著差异。KMO 和 Bartlett 检验如表 4-6 所示。

表 4-5 相关矩阵

		C11	C12	C22	C21	C31	C33	C41	C42	D11	C34	D12	D13	D21	D22
相 关	C11	1.000	0.882	-0.857	0.995	-0.857	-0.019	0.003	0.038	-0.050	-0.129	0.203	-0.209	0.225	-0.360
	C12	0.882	1.000	-0.723	0.879	-0.783	-0.006	-0.028	0.040	0.057	0.005	0.526	-0.517	0.510	-0.336
	C22	-0.857	-0.723	1.000	-0.850	0.807	0.161	0.022	-0.039	0.035	0.203	-0.185	0.148	-0.154	0.303
	C21	0.995	0.879	-0.850	1.000	-0.847	0.002	-0.015	0.030	-0.079	-0.110	0.204	-0.209	0.219	-0.360
	C31	-0.857	-0.783	0.807	-0.847	1.000	0.160	-0.086	-0.080	.0064	0.057	-0.271	0.244	-0.254	0.162
	C33	-0.019	-0.006	0.161	0.002	0.160	1.000	0.193	-0.023	-0.368	0.062	-0.045	-0.035	-0.176	-0.034
	C41	0.003	-0.028	0.022	-0.015	-0.086	0.193	1.000	0.479	-0.276	-0.187	-0.133	0.180	-0.171	-0.130
	C42	0.038	0.040	-0.039	0.030	-0.080	-0.023	0.479	1.000	-0.252	-0.267	-0.174	0.253	-0.078	-0.093
	D11	-0.050	0.057	0.035	-0.079	0.064	-0.368	-0.276	-0.252	1.000	0.366	0.428	-0.427	0.372	0.251
	C34	-0.129	0.005	0.203	-0.110	0.057	0.062	-0.187	-0.267	0.366	1.000	0.460	-0.476	0.320	0.562
	D12	0.203	0.526	-0.185	0.204	-0.271	-0.045	-0.133	-0.174	0.428	0.460	1.000	-0.972	0.723	0.067
	D13	-0.209	-0.517	0.148	-0.209	0.244	-0.035	0.180	0.253	-0.427	-0.476	-0.972	1.000	-0.728	-0.112
	D21	0.225	0.510	-0.154	0.219	-0.254	-0.176	-0.171	-0.078	0.372	0.320	0.723	-0.728	1.000	0.041
	D22	-0.360	-0.336	0.303	-0.360	0.162	-0.034	-0.130	-0.093	0.251	0.562	0.067	-0.112	0.041	1.000

表 4-6 KMO 和 Bartlett 检验

KMO 和 Bartlett 的检验		
取样足够度的 Kaiser-Meyer-Olkin 度量。		0.708
Bartlett 的球形度检验	近似卡方	575.382
	df	91
	Sig.	0.000

2) 结果 4-7 给出了从每个原始变量中提取了多少信息。结果如图 4-7 所示。

表 4-7 公因子方差

	初始	提取
购买金额 (C11)	1.000	0.965
购买总额(C12)	1.000	0.933
关系成本(C22)	1.000	0.852
毛利润(C21)	1.000	0.963
客户发展潜力(C31)	1.000	0.885
交叉购买(C33)	1.000	0.890
银行信用(C41)	1.000	0.734
还款及时度(C42)	1.000	0.784
客户弹性(D11)	1.000	0.658
推荐次数(C34)	1.000	0.809
购买次数(D12)	1.000	0.924
客户份额(D13)	1.000	0.941
客户满意度(D21)	1.000	0.760
转换成本(D22)	1.000	0.879

3) 结果 4-8 是代表着各主成解释原始变量总方差。可以看出本文数据保留了五个主成分，第一个主成分对应的特征值 λ_1 为 4.995，即第一个主成分描述了原始变量总变差 14 中的 4.995，标准化后的原始变量总的方差是

$$\sum_{i=1}^{14} \lambda_i = 4.995 + 3.347 + \dots + 0.003 = 14, \text{ 第一个主成分的贡献了}$$

$$\text{计算是 } \frac{\lambda_1}{\sum_{i=1}^{14} \lambda_i} \times 100\% = \frac{4.995}{14} \times 100\% = 35.679\%, \text{ 即第一个主成解释了 14 个原始}$$

始变量信息的 35.679%，同理可知其他四个主成解释了 14 个原始变量信息

的贡献率分别是 23.91%、10.092%、8.497%、7.371%。前五个主成分的累计

贡献率是 $\frac{\lambda_1 + \lambda_2 + \dots + \lambda_5}{\sum_{i=1}^{14} \lambda_i} \times 100\% = \frac{4.995 + \dots + 1.032}{14} = 85.548\%$ 。累计贡献率已

经达到 85.548% 大于 85% 的要求，表示这五个主成分可以表达原始样本中指标体系中绝大部分信息。结果如表 4-8 所示。

表 4-8 解释总方差

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	4.995	35.679	35.679	4.995	35.679	35.679
2	3.347	23.910	59.588	3.347	23.910	59.588
3	1.413	10.092	69.680	1.413	10.092	69.680
4	1.190	8.497	78.177	1.190	8.497	78.177
5	1.032	7.371	85.548	1.032	7.371	85.548
6	.563	4.018	89.567			
7	.454	3.243	92.809			
8	.342	2.441	95.251			
9	.308	2.198	97.449			
10	.174	1.243	98.691			
11	.115	.822	99.514			
12	.050	.356	99.870			
13	.015	.109	99.979			
14	.003	.021	100.000			

4) 结果 4-9 输出的是主成分的碎石图，根据特征值的转换趋势从陡峭开始到趋于平缓的转折点，根据转折点分析本文数据保留了 5 个主成分为宜。结果如图 4-9 所示。

5) 结果 4-10 成分矩阵输出是第一至第五个主成分与原始变量的关系。如表 4-10 所示。而五个主成分对应的特征根 $\lambda_1 \dots \lambda_5$ 分别为 4.995、3.347、1.413、1.190 和 1.032。可以得到五个特根的平方根 $\sqrt{\lambda_1} \dots \sqrt{\lambda_5}$ 分别为 2.235、1.829、1.189、1.091 和 1.016。用 $\sqrt{\lambda_1} \dots \sqrt{\lambda_5}$ 分别除以结果 4-10 第二至六列可以算出主成分表达式的系数（即每个特征值对应的单位特征向量），因此可以写成

各个主成分的表达式：

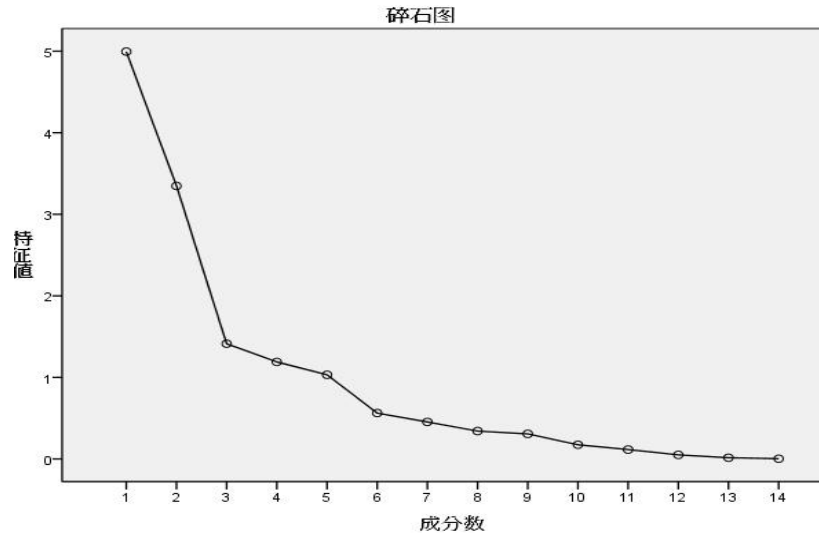


图 4-9 碎石图

表 4-10 成分矩阵

	1	2	3	4	5
购买金额 (C11)	0.907	-0.329	-0.081	-0.106	0.125
购买总额(C12)	0.956	-0.049	0.107	-0.035	-0.060
关系成本(C22)	-0.823	0.335	0.204	-0.009	-0.143
毛利润(C21)	0.903	-0.328	-0.074	-0.139	0.127
客户发展潜力(C31)	-0.857	0.244	0.038	-0.058	-0.293
交叉购买(C33)	-0.111	-0.125	0.602	-0.706	-0.027
银行信用(C41)	-0.073	-0.404	0.669	0.323	0.118
还款及时度(C42)	-0.032	-0.429	0.470	0.600	0.135
客户弹性(D11)	0.135	0.662	0.312	0.322	-0.010
推荐次数(C34)	0.041	0.736	0.161	-0.177	0.458
购买次数(D12)	0.564	0.693	0.288	0.062	-0.195
客户份额(D13)	-0.551	-0.722	-0.281	0.044	0.186
客户满意度(D21)	0.533	0.584	0.173	0.226	-0.232
转换成本(D22)	-0.292	0.521	0.017	0.018	0.722

$$y_1 = 0.406 \times C_{11} + 0.428 \times C_{12} - 0.368 \times C_{22} + 0.404 \times C_{21} - 0.383 \times C_{31} - 0.05 \times C_{33} - 0.033 \times C_{41} - 0.016 \times C_{42} + 0.060 \times D_{11} + 0.018 \times C_{34} + 0.252 \times D_{12} - 0.247 \times D_{13} + 0.238 \times D_{21} - 0.13 \times D_{22}$$

$$y_2 = -0.180 \times C_{11} - 0.027 \times C_{12} + 0.183 \times C_{22} - 0.179 \times C_{21} + 0.133 \times C_{31} - 0.068 \times C_{33} - 0.220 \times C_{41} - 0.235 \times C_{42} + 0.362 \times D_{11} + 0.402 \times C_{34} + 0.379 \times D_{12} - 0.395 \times D_{13} + 0.319 \times D_{21} + 0.285 \times D_{22}$$

$$y_3 = -0.068 \times C_{11} + 0.090 \times C_{12} + 0.172 \times C_{22} - 0.062 \times C_{21} + 0.032 \times C_{31} + 0.506 \times C_{33} + 0.563 \times C_{41} + 0.395 \times C_{42} - 0.262 \times D_{11} + 0.135 \times C_{34} + 0.242 \times D_{12} - 0.236 \times D_{13} + 0.242 \times D_{21} + 0.014 \times D_{22}$$

$$y_4 = -0.097 \times C_{11} - 0.032 \times C_{12} - 0.008 \times C_{22} + 0.127 \times C_{21} - 0.053 \times C_{31} - 0.647 \times C_{33} + 0.296 \times C_{41} + 0.550 \times C_{42} + 0.295 \times D_{11} - 0.162 \times C_{34} + 0.057 \times D_{12} + 0.040 \times D_{13} + 0.207 \times D_{21} + 0.016 \times D_{22}$$

$$y_5 = 0.123 \times C_{11} - 0.060 \times C_{12} - 0.141 \times C_{22} + 0.125 \times C_{21} - 0.289 \times C_{31} - 0.027 \times C_{33} + 0.116 \times C_{41} + 0.133 \times C_{42} - 0.010 \times D_{11} + 0.451 \times C_{34} - 0.192 \times D_{12} + 0.183 \times D_{13} - 0.228 \times D_{21} + 0.71 \times D_{22}$$

6) 将原始数据标准化处理。选择分析→描述统计→描述命令打开描述性对话框，将左端的对话框中依次选择变量 C11,C12,.....,D22 使之移到对话框右端的变量框中。勾选将标准化得分另存为变量，保存标准化后的结果，如图 4-11 所示。



图 4-11 标准化数据描述

标准化的变量依次为新的变量名 ZC11,ZC12,.....,ZD22 保存在数据编辑窗口，如图 4-12 所示。

其次，计算各个主成分得分。依次选择转换→计算变量，在计算变量对话

框的目标变量输入 y1（第一主成分得分），并且在数据表达式框内输入 y1 的表达式：

$$y1=0.406\times ZC11+0.428\times ZC12-0.368\times ZC22+0.404\times ZC21-0.383\times ZC31-0.05\times ZC33-0.033\times ZC41-0.016\times ZC42+0.060\times ZD11+0.018\times ZC34+0.252\times ZD12-0.247\times ZD13+0.238\times ZD21-0.13\times ZD22$$

户弹性D11	Z推荐次数C34	Z购买次数D12	Z客户份额D13	Z转换成本D22	Z客户满意度D21
.89316	.11984	.81184	-1.14519	-.05028	.27516
-.29772	-.67910	-.66423	.71389	-.72068	-.51101
-1.48859	-1.07856	-.66423	.71389	-.05028	-.51101
.89316	.11984	-.66423	.71389	-.72068	1.06133
-2.67947	-1.87750	-.66423	.71389	-1.39108	-1.29718
.89316	.11984	.81184	-1.14519	.62012	1.06133
-.29772	.91878	-.66423	.71389	-.05028	-.51101
.89316	-.27963	-.66423	.71389	.62012	.27516
-1.48859	.11984	.81184	-1.14519	-1.39108	1.06133
.89316	-1.07856	2.28790	-1.77728	-2.06148	1.84750
.89316	-.67910	-.66423	.71389	-.72068	-.51101
-.29772	-1.07856	-.66423	.71389	-.05028	.27516
-1.48859	.91878	.81184	-1.14519	-.72068	1.06133

图 4-12 新标准化后的新变量

在 SPSS 数据编辑窗口将会出现一个名为 y1 的变量，即为第一主成分得分。如图 4-13 所示。

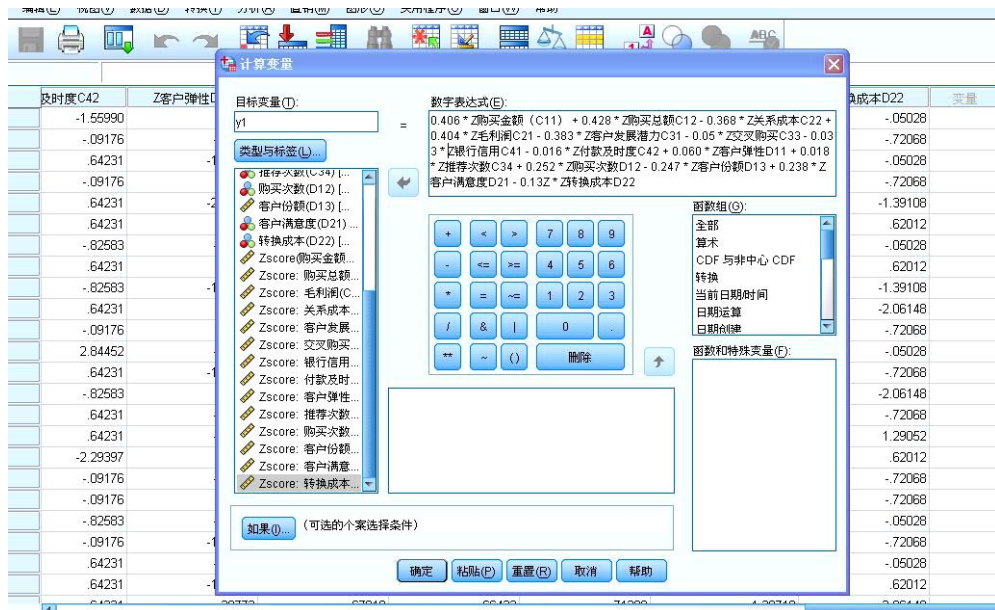


图 4-13 计算变量

同理，在数字表达式中 y_2, \dots, y_5 。可以在 SPSS 数据编辑窗口中生成其他主成分得分。

最后，计算综合结果得分，按照下式：

$$y = 0.35679 \times y_1 + 0.2391 \times y_2 + 0.10092 \times y_3 + 0.08497 \times y_4 + 0.07371 \times y_5$$

其中 y_1 至 y_5 是第一至第五主成分得分，仍可以在计算变量对话框中完成，在 SPSS 数据编辑窗口中生成一个 y 的变量，即主成分综合得分，如图 4-14 所示。



图 4-14 主成分的综合得分

将主成分 y_1 至 y_5 以及主成分综合得分 y 计算进行整理，即得到数据约简的结果。如图 4-15 所示。

本文设置提取主成分的条件是特征值大于 1 并且累计贡献率大于 85%，本文累计贡献率 85.548% 已经大于 85% 的要求，说明这五个主成分可以表达原始样本中绝大部分信息，并且可信度较高。因此可以选取五个主成分替代原有的 14 个指标是合理的，即将有用的信息提取出来，将冗余的信息去除，从而达到降维的目的。图 4-15 中的数据将会进行后续的分析工作。

客户	y1	y2	y3	y4	y5	y
1	-.33	1.79	.45	-.89	-.90	.21
2	.59	-1.74	-1.12	.48	.05	-.27
3	-1.69	-1.53	.33	.23	-.29	-.94
4	-.54	.31	-.93	1.71	-.62	-.11
5	-1.34	-3.33	.72	.06	-1.16	-1.28
6	-.18	2.16	.27	1.34	-.51	.56
7	-1.21	-.30	.21	-.36	.60	-.47
8	-.70	-.06	-.52	1.48	.48	-.16
9	2.16	-.15	1.27	-1.29	-1.32	.66
10	4.90	.37	.79	1.61	-2.32	1.88
11	-.18	-1.02	-.74	.22	-.40	-.39
12	-3.01	-1.01	2.39	1.80	-.69	-.97
13	6.01	-1.21	1.48	-1.98	.68	1.89
14	2.68	-2.43	-1.42	-.28	-1.02	.13
15	-.36	-2.17	-.04	-.25	-.08	-.68
16	.94	.72	2.56	-.60	1.28	.81
17	-2.11	3.91	-.84	-1.08	-1.27	-.09
18	2.60	-2.32	-.88	-1.27	.40	.20
19	3.84	.95	-.52	-.22	-.67	1.47
20	-2.48	1.09	-2.53	.24	-1.11	-.94
21	.54	-2.50	-.59	-1.37	.02	-.58

图 4-15 数据约简结果

4.3 基于 K-means 算法的客户细分

本节是把上节经过预处理后的样本数据基于 K-means 算法进行聚类细分。具体步骤如下：

(1) **建立数据文件** 通过选择文件→打开命令将上节五个主成分数据文件调入 SPSS19.0 工作文件窗口，在依次打开分析→分类→K-均值聚类命令，打开 K-means 对话框，接着将参与 K-means 聚类的变量放入变量框内。如图 4-16 所示。

(2) **文件设置** 在 K-均值聚类分析对话框中打开迭代对话框，其中最多迭代次数指的是迭代该次数时候即使没有满足迭代的收敛标准，也将终止聚类分析过程。在 SPSS 中系统默认为 10 次。在选项框内勾选初始聚类中心、ANOVA 表、每个聚类信息。将聚类数改成 8，点击确定运行结果。

(3) K-均值聚类分析的结果分析

1) 结果 4-17 是 SPSS 指定的初始类中心点，由于需要快速聚类成 8 类，因此指定了 8 个类中心的初始位置。对照原始数据可以知道八个类别分别使用了第 13 个,23 个,32 个,17 个,10 个,29 个样本数据作为初始位置如表 4-17 所示。

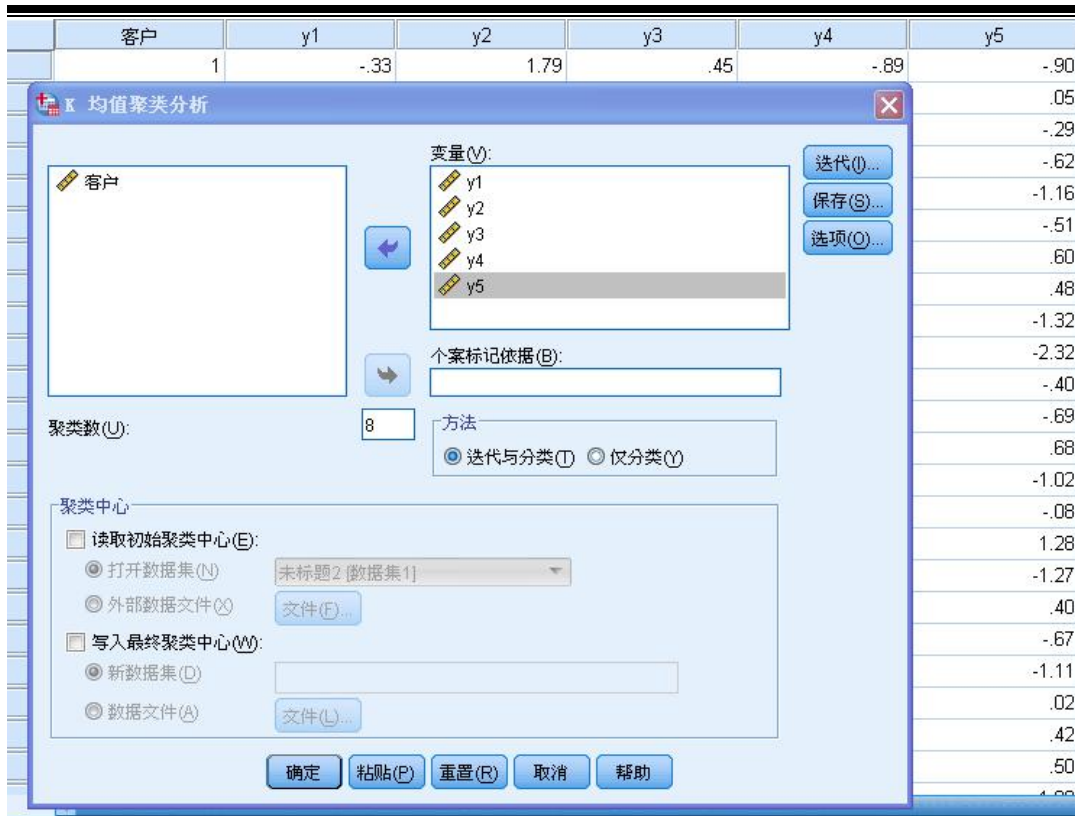


图 4-16 K-均值聚类

表 4-17 初始聚类中心

	1	2	3	4	5	6	7	8
y1	6.01	2.68	-3.27	4.35	-1.34	-2.11	4.90	-0.95
y2	-1.21	-2.43	-0.08	2.06	-3.33	3.91	0.37	0.88
y3	1.48	-1.42	1.79	1.01	0.72	-0.84	0.79	-2.05
y4	-1.98	-0.28	-1.12	1.96	0.06	-1.08	1.61	0.21
y5	0.68	-1.02	0.50	2.32	-1.16	-1.27	-2.32	1.46

2) 结果 4-18 是 K-均值聚类分析的迭代历史过程表。可以看出总共进行了四次迭代后收敛。第一次迭代后形成的类中心和初始类中心点的距离分别是：0.00,1.642,1.685,2.522,1.555,1.750,1.522,1.780。第二次迭代后类中心点的距离分别是：0.00,0.00,0.223,0.00,0.271,0.00,0.00,0.333。第三次迭代后类中心点的距离分别是：0.00,0.00,0.00,0.00,0.0193,0.00,0.00,0.188。第四次迭代后类中心点没变化或是变化非常小，距离均为 0.00。如表 4-18 所示。

表 4-18 迭代历史记录

	1	2	3	4	5	6	7	8
1	0.000	1.642	1.685	2.522	1.555	1.750	1.522	1.780
2	0.000	0.000	0.223	0.000	0.271	0.000	0.000	0.333
3	0.000	0.000	0.000	0.000	0.193	0.000	0.000	0.188
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

3) 结果 4-19 给出的是 K-均值聚类分析最终的类中心点位置, 此结果和结果 4-17 比较, 中心点位置有一定的变动, 表明了迭代过程中中心点的位置发生了转移。如表 4-19 所示。

表 4-19 最终聚类中心

	1	2	3	4	5	6	7	8
y1	6.01	1.71	-1.82	3.04	-1.07	-1.16	4.37	-0.53
y2	-1.21	-1.83	-0.10	2.28	-1.77	3.10	0.66	0.53
y3	1.48	-0.55	1.26	0.10	-0.08	0.37	0.13	-1.14
y4	-1.98	-0.75	-0.11	0.51	0.08	-1.07	0.69	0.82
y5	0.68	-0.37	0.75	1.02	-0.44	-1.05	-1.49	0.31

4) 结果 4-20 是最终的类中心之间的欧式距离。从中可以看出来第一类类中心点和第六类类中心点距离最大, 为 8.663; 而第三类类中心点和第五类类中心点距离最小, 为 2.570。如表 4-20 所示。

表 4-20 最终距离

聚类	1	2	3	4	5	6	7	8
1		5.060	8.129	5.411	7.638	8.663	4.456	7.785
2	5.060		4.515	4.751	2.937	5.824	4.126	3.721
3	8.129	4.515		5.570	2.570	3.960	6.768	2.982
4	5.411	4.751	5.570		5.964	5.016	3.282	4.228
5	7.638	2.937	2.570	5.964		5.063	6.081	2.795
6	8.663	5.824	3.960	5.016	5.063		6.313	3.833
7	4.456	4.126	6.768	3.282	6.081	6.313		5.374
8	7.785	3.721	2.982	4.228	2.795	3.833	5.374	

5) 结果 4-21 为 K-均值聚类分析后形成的各类样本之间的单因素方法分析 (ANOVA) 结果, 用于检验 K-均值聚类分析的 5 个变量间的差异性, 表格

每一行对应相应变量的分析结果。若显著水平 $\text{Sig.} < 0.05$ ，就表明在 0.05 的显著水平下各类均值存在明显的差异。

表 4-21 ANOVA

	聚类		误差		F	Sig.
	均方	df	均方	df		
y1	22.855	7	1.083	32	21.102	0.000
y2	14.393	7	.931	32	15.462	0.000
y3	4.111	7	.880	32	4.672	0.001
y4	2.743	7	.849	32	3.232	0.001
y5	2.880	7	.627	32	4.591	0.001

首先是 y1 变量，平均组间平方和为 22.855，平均组内平方和为 1.083，F 统计量为 21.102，F 统计量的相伴概率 $\text{Sig.} = 0.000$ 。相伴概率 Sig. 小于显著水平 0.05，认为对于变量 y1 而言，8 个类包括数据间具有明显差异。其余 4 个变量可以做变量类似分析。如表 4-21 所示。

因此，由上述结果可知，8 个类别的 5 个变量都在 0.05 的显著水平下高度显著，8 类指标的差异显著，5 个变量都对分类贡献显著，换句话说就是所分 8 类是合理的，聚类分析是成功的并且效果理想。

6) 结果 4-22 给出了 K-均值聚类分析的分类结果，以及包含样本的情况。表格最后一列表示该样本离类中心点的距离。其中第 16 号客户离第三类的中心点距离最远为 3.234，第 13 号客户离第一类中心点距离最近为 0.000。如表 4-22 所示。结果表明：

一类别包括 1 个客户，为编号 13 号客户。

二类别包括 5 个客户，为编号 2、9、14、18、21 号客户。

三类别包括 8 个客户，为编号 7、12、16、23、28、30、36、39 号客户。

四类别包括 3 个客户，为编号 32、34、38 号客户。

五类别包括 8 个客户，为编号 3、5、11、15、22、24、25、37 号客户。

六类别包括 4 个客户，为编号 1、17、26、31 号客户。

七类别包括 2 个客户，为编号 10、19 号客户。

八类别包括 9 个客户，为编号 4、6、8、20、27、29、33、35、40 号客户。

表 4-22 聚类成员

案例号	聚类	距离	案例号	聚类	距离	案例号	聚类	距离
1	6	1.572	15	5	0.943	29	8	1.679
2	2	1.815	16	3	3.234	30	3	1.563
3	5	0.808	17	6	1.750	31	6	2.981
4	8	1.328	18	2	1.419	32	4	2.522
5	5	1.916	19	7	1.522	33	8	1.458
6	8	2.390	20	8	2.898	34	4	1.389
7	3	1.269	21	2	1.544	35	8	2.154
8	8	1.112	22	5	1.573	36	3	1.486
9	2	2.744	23	3	1.858	37	5	0.983
10	7	1.522	24	5	0.973	38	4	1.764
11	5	1.348	25	5	1.292	39	3	1.681
12	3	3.046	26	6	2.515	40	8	1.711
13	1	0.000	27	8	2.394			
14	2	1.642	28	3	2.105			

7) 结果 4-23 表示了 8 个类中的所包含的样本数,即结论 4-22 中各类别所包括的客户数。其中显示缺失值是 0,总有效数为 40,与样本值一样,表示所有样本均匀归属其中的某一类。如表 4-23 所示。

表 4-23 案例数

案例数				
聚类	1	1.000	5	8.000
	2	5.000	6	4.000
	3	8.000	7	2.000
	4	3.000	8	9.000
有效	40.000			
缺失	0.000			

4.4 结果分析及建议

4.4.1 结果分析

通过上一节所获得的聚类结果可知,将 40 个样本客户分为 8 类,下面结合原始数据进行对比分析,可以得到最终的客户细分结果如下 4-24 所示。

因此,分析客户数据可以得到一下结论:

第一类（钻石型客户），为编号 13 号客户。

第二类（金牛型客户），为编号 4 号、6 号、8 号、20 号、27 号、23 号、33 号、35 号和 40 号客户。

第三类（成长型客户），为编号 3 号、5 号、11 号、15 号、22 号、24 号、25 号和 37 号客户。

第四类（获利型客户），为编号 10 号和 19 号客户。

第五类（适当发展型客户），为编号 32 号、34 号和 38 号客户。

第六类（短期客户型客户），为编号 2 号、9 号、14 号、18 号和 21 号客户。

第七类（风险型客户），为编号 7 号、12 号、16 号、23 号、28 号、30 号、36 号和 39 号客户。

第八类（无利可图型客户），为编号 1 号、17 号、26 号和 31 号客户

得到细分结果后，结合各类客户的特点，可以将其相关类型的客户进行有针对性的资源配置策客户管理策略，分析如下：

（1）**第一类** 钻石型客户。此类型的业户已经给房地产公司带来了高额利润，而且以后也会为房地产公司带来利润，并且此类业户对本公司的忠诚度高，对公司非常的信任，因此对于这一类的客户企业要保持长久稳定的联系，充分了解客户的需求，重点投入企业资源并制定重点的保持策略，让业主一直保持对本公司满意和忠诚，或是积极的为公司推荐新的客户，使公司有持续获利的保障。

表 4-24 八种客户细分类型

类别	当前价值	潜在价值	忠诚度	客户类型
1	高	高	高	第一类（钻石型）
2	高	低	低	第六类（短期客户型）
3	低	高	低	第七类（风险型）
4	低	低	高	第五类（适当发展型）
5	低	高	高	第三类（成长型）
6	低	低	低	第八类（无利可图型）
7	高	高	低	第四类（获利型）
8	高	低	高	第二类（金牛型）

（2）**第二类** 金牛型客户。此类型的业户已经为公司带来了高额利润，并且对公司有很高的忠诚度，但是对公司潜在价值低，如果该类业户是刚刚与

本公司建立的客户关系,那么公司有必要为这类业户投入资源,对客户进行深入的分析,提供必要的需求,挖掘并提高客户的潜在的价值,在未来为公司带来越来越多的利润,若是该类业户早已与本公司保持关系,那么此类业户的潜在价值被完全利用殆尽,那么企业没有必要花大力气投入资源,仅投入少量资源保持好客户关系就可以。

(3) **第三类 成长型客户。**此类型业户对本公司的忠诚度高,对公司产品及服务比较满意,并且将来有继续交叉购买的能力和愿望,但是由于某种原因目前没有与企业产生大规模的交易,一旦时机成熟这类客户会增大企业的消费量。因此,公司需要主动的与这类业户交流,为业户提供好服务,尽量实现客户的合理需求,甚至个性化解解决方案,稳住客户对企业的忠诚,重点投入资源,保留住客户,使客户继续推荐朋友购买本企业产品为企业持续创造价值,争取让这类客户转向成为钻石型客户。

(4) **第四类 获利型客户。**此类型业户已经给公司带来了高额的利润,将来也会有很高发展潜力,并且有交叉购买的想法,也就是说这类业户对公司目前和将来的贡献都很大,但是企业的忠诚度低,因此企业应该加大投入资源,寻找此类客户忠诚度不高的原由,有针对性的做好服务,甚至提供超值服务挽救信任从而提高客户的忠诚度,避免客户流失,加大对此类业户的培养,争取转向成为钻石型客户。

(5) **第五类 适当发展型客户。**此类型业主当前为公司带来的利润比较少,同时在将来为公司带来利润的能力也不高,但是对本公司十分的忠诚,这些高度信任本公司的客户,虽然创造的价值低,但是推荐客户,为本公司品牌的宣传也是一种无形的贡献,有利用公司减少新业户的吸引成本,因此对于这类客户能力有限,有可能流失,被视为适当发展型客户,公司只要投入适量的资源,维系好关系,保持对本公司的忠诚,既不挽留又不鼓励离开。

(6) **第六类 短期客户型客户。**此类型业户目前为公司创造利润比较高,但是在将来为公司带来的利润能力比较低,没有进步消费的意愿,并且对本公司忠诚度不高,他们随时有可能转向成为竞争对手的客户,因此这类型的客户有二种可能性,一种可以通过企业投入大量资源,提供优质的服务能够激发业户的交叉购买欲望;另一种可能就是客户对本企业的产品或是服务要求有些苛刻,即使投入了过多的资源,最后难以获得相应的回报,故而投入少量资源,在其流失前获得最高的利润。因此这类型业户被视为短期客户类型,公司应当对业户潜在价值进行分析,选择该类型客户合适的投入资源和管理策略。

(7) **第七类 风险型客户**。此类型业户目前为公司带来的利润比较低, 对公司忠诚度也不高, 但是在将来为公司带来的利润会很高, 并且有进步消费的意愿, 对于这类高潜力目前消费份额小的业户, 有可能是关系不稳定的新客户, 如果不增加投入, 这类型客户有可能转向成为竞争对手的客户, 但是如果投入过多资源, 在潜在价值增值之前, 为企业获得短期的较低的利润, 故而, 此类型客户被视为风险型客户, 因此, 从长远来看, 企业应该投入适量的资源, 加强客户关系的培养, 为企业持续创造利润。

(8) **第八类 无利可图型客户**。此类型业户当前和将来为公司带来的利润都很低, 并且对公司忠诚度也不高, 对于此类型客户公司没有必要投入过多的资源维系客户之间的关系, 应当把节约下来的资源分配到其他有价值的业户身上获取更高的利润。与此同时, 企业也应当认真分析此类客户对本企业产品不活跃的原因, 是本企业的产品问题还是客户自身问题, 企业应当花些精力避免这类客户的负观点传递给其他客户, 故而此类型客户被视为无利可图型客户, 只要此类业户不对公司产生负能量, 公司无需投入资源。

综上所述, 本文的房地产客户细分结果能够用于实际的管理工作, 并且能够进行相应的解释和预测, 为企业制定合适的资源配置策略和保持策略有一定实际指导意义。

4.4.2 对宝宇地产实施客户细分建议

本文在 4.1.2 节已经对宝宇地产企业进行了 SWOT 分析, 并总结宝宇地产营销管理中存在问题, 下面结合宝宇地产客户细分研究的结论, 为宝宇地产提出几点建议, 希望能利用细分结果辅助解决实际的客户营销问题。

(1) **建立完备的客户信息数据库** 宝宇地产从公司成立至今已经拥有了庞大的客户数量, 积累众多的客户信息, 但是在记录客户信息时出现记录的信息不规范, 不全面等问题, 因此, 宝宇地产应当构建完备的客户信息数据库, 重视客户关系管理, 调整已有的客户信息, 并对全部的数据信息重新进行分类, 构成为统一、全面的集合, 方便以后需要时候对数据信息直接高效提取利用, 为客户细分提供数据基础。

(2) **及时更新客户信息数据库** 客户是随着生命周期的而逐渐成长的, 企业不断的发展, 客户也不断的变化和进步的, 比如原本低价值的客户可能继续交叉购买本企业的产品为企业创造更高的价值成为高价值的客户, 或是低忠诚度的客户改变最初的对宝宇地产态度, 积极的为身边友人推荐宝宇地产的楼

盘，进而转变成为高忠诚度的客户等等。因此宝宇地产应当实时的更新客户的数据，并且根据客户的成长变化情况对客户进行动态的调整，确保客户细分持续有效。

（3）**优化客户关系管理系统** 客户关系管理是用不同的方式区别对待有价值差异的客户，据统计，三分之二的客户离开企业是因为企业对他们关注不够；93%的老总均认为客户细化管理是公司最具有竞争力的元素；客户细分既是客户关系管理的重要理论组成成分，又是急待解决问题之一，可以对客户获取、保持及增值等过程提供全面支持，提升客户满意度和忠诚度。因此，借助客户细分结果研究出客户交叉购买模型、客户流失模型等模型，体现客户细分价值，优化客户细化管理系统，用于之后指导实际的销售工作和客户保持工作。

4.5 本章小节

本节首先，进行了数据预处理，利用统计学软件 SPSS19.0 版本对原始数据进行了主成分分析，并提取 5 个主成分；其次，把提取的主成分进行了 K-均值聚类，将客户细分成为 8 类；再次，结合原始数据对比分析每个类的客户，针对各类型客户特点，对每类型客户提出相应的分配资源和客户管理策略；最后，对宝宇地产实施客户细分给出三点意见。

结 论

本文以哈尔滨宝宇地产客户为研究对象,首先提出房地产客户细分的背景和意义,对客户细分的方式和指标进行了系统的分析和梳理,其次阐述了聚类分析技术的方法选取 K-均值技术用来分析并结合客户价值以及客户的忠诚度建立了房地产客户细分的模型,同时通过问卷调查的方式收集了宝宇地产天客户相关的指标数据,进行了验证了模型的可行性和有效性。最后,对宝宇房地产开发企业提出了几点意见。

(1) 论文的研究成果如下

1) 根据房地产客户的当前价值、潜在价值以及忠诚度构建了三维的客户细分模型,不同于其他单一的根据生命周期或是客户价值的房地产模型。本文总结之前学者的和其他领域的研究经验,在结合房地产客户特有的消费特征基础上,总结归纳出房地产客户细分的指标体系,比其他房地产细分模型多一个忠诚度维度,解决了传统的只针对客户生命周期或是客户价值的单一维度细分的局限性,能够全方位的、科学的、动态的反应客户的特征,有利于企业准确的进行客户细分管理。

2) 将聚类分析引入了客户细分之中,能够有效的对大量的客户信息进行处理,本文分析了聚类分析各种方法选取了 K-means 算法对客户进行细分。在执行该算法前,选择主成分分析的方法对原始数据进行预处理。最后,构建了根据聚类分析的房地产企业客户细分模型。

3) 利于宝宇地产实际客户信息检验模型的可行性和实用性。从宝宇房地产客户调查问卷获得的实际客户数据信息,利用构建的客户细分模型对收集到的客户信息进行聚类,对聚类的结果对比原始客户数据进行分析,并根据结果制定公司的资源配置策略和客户管理策略。

(2) 论文的不足之处如下

1) 影响房地产客户细分指标,即客户价值和忠诚度因素众多,既有主观性的因素又有客观性的因素,也存在一些难以量化的指标,在本论文中,只是总结相关客户细分文献中选取了具有代表性的指标,如果可以囊括和完善更多的参考性的和有实际价值的指标并进行调整,则可以更加精准和详细的了解分析客户细分各方面特征。

2) 本文选用聚类分析中的 K-均值算法,该方法的特点是简单、高效、客观、可解释性强、现实应用广泛性等优点,适合房地产客户细分的要求。但该算法也存在着一些不足,例如该算法对非线性的数据聚类识别效果差,受到

噪声、边界点影响，对聚类结果的稳定性和准确性有一定的干扰，并且对于聚类值 k 的选取主观性强，可能对结果产生误差。因此，在以后的研究中可以对该算法进行改进，规避算法的不足之处，能够更加准确的反应聚类的结果。

3) 问卷发放和回收的数据有限，如果能更多的发放问卷，尤其在没有任何购买宝宇地产项目的客户发放问卷收集没有购买宝宇地产的原因，分析这类客户的需求可否成为潜在客户，将会更加全面的验证模型的有效性，并且问卷回收数据有限并且主观性强，故而，在以后的研究中要加大问卷的投放量增加数据信息的数量并且可以探索对定性指标进行量化的方法，减少主观性。

参考文献

- [1] 王扶东, 马玉芳. 基于数据挖掘的客户细分方法的研究[J]. 计算机工程与应用, 2011, 47(4):215-218.
- [2] Data clustering : 50 years beyond K2means [J] . Pat2 tern Recognition Letters ,2010 ,31 (8) :651 - 666.
- [3] 王婷. 基于客户价值的服务评价研究——以 A 公司为例[J]. 物流技术, 2016, 35(3):61-65.
- [4] 褚格林. 基于聚类模型的电信客户细分研究[J]. 统计与决策, 2014(8):176-179.
- [5] 于辉, 廖小红. 客户细分方法综述[J]. 中小企业管理与科技旬刊, 2014(33):17-18.
- [6] 张焕国, 吕莎, 李玮. C 均值算法的电信客户细分研究[J]. 计算机仿真, 2011, 28(6):185-188.
- [7] 季越江, 吕佳. 基于聚类分析的客户细分研究[J]. 办公自动化:综合版, 2009(8):36-39.
- [8] 金雷法. 基于客户价值的客户细分及保持策略分析[J]. 江苏商论, 2010(15):108-108.
- [9] 范志国, 付波. 基于客户利润贡献度、贡献增长率和关系可靠性的客户细分研究[J]. 华北电力大学学报(社会科学版), 2010, 2010(1):30-35.
- [10] 陈文沛. 基于消费者创新性的市场细分研究[J]. 商业时代, 2010(15):17-18.
- [11] 叶志龙, 黄章树, YEZhi-long,等. 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.
- [12] 綦欣德. 客户细分方法新视角[J]. 商业时代, 2013(26):31-33.
- [13] 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.
- [14] 刘潇潇, 郭馨泽, 刘金,等. 基于电力客户行为的客户细分与价值评价研究[J]. 电子世界, 2013(24):245-248.
- [15] 王勇, 唐靖, 饶勤菲,等. 高效率的 K-means 最佳聚类数确定算法[J]. 计算机应用, 2014, 34(5):1331-1335.
- [16] 马儒宁, 王秀丽, 丁军娣. 多层核心集凝聚算法[J]. 软件学报, 2013(3):490-506.
- [17] 韩凌波, HANLingbo. 基于密度的 K-means 初始聚类中心选取算法[J]. 电

- 子科技, 2015, 28(7):105-107.
- [18] 何佳知, 谢颖华. 基于密度优化初始聚类中等的 K-means 算法[J]. 微型机与应用, 2015, 37(3):17-19.
- [19] 许茂增, 余国印. 基于吉自适应遗传算法的 K-means 聚类分析[J]. 数学的实践与认识, 2015, 17: 6.
- [20] 邢长征, 谷浩. 基于平均密度优化初始聚类中心的 K-means 算法[J]. 计算机工程与应用, 2014, 50(20):135-138.
- [21] 樊宁. K 均值聚类算法在银行客户细分中的研究[J]. 计算机仿真, 2011, 3(28):369-372.
- [22] 彭凯. 应用因子分析和 K-means 聚类的客户分群建模[J]. 计算机科学, 2011(5):154-159.
- [23] 赵铭, 李雪, 李秀婷,等. 基于聚类分析的商业银行基金客户的分类研究[J]. 管理评论, 2013, 25(7):38-44.
- [24] Darrell Rigby, Barbara Bilodeau .Management Tools and Trends 2009 [J]. BAIN COMPANY Inc Research Report , 2009: 1-10.
- [25] Gremler Dwayne D, Stephen W. Brown. Customer Loyalty and Satisfaction: What Resonates in Service Contexts[J]. Journal of Marketing , 2001, (8):207-220.
- [26] Buckinx W, Verstraeten G, Poel D V D. Predicting customer loyalty using the internal transactional database[J]. Expert Systems with Applications, 2007, 32(1):125-134.
- [27] J. Albrecht, T- Sollner. Market Segmentation and the Structure of Competition Applicability of the Strategic Group Concept for an Improved Market segmentation on, Industrial Markets. Journal of Business Research , 2007, 51(1):25-36.
- [28] Bottcher M, Spott M Nauck D, Kruse R. Mining changing customer segments in dynamic markets[J]. Expert Systems with Applications, 2009, (36):155-164.
- [29] Yedla M, Pathakota S R, Srinivasa. T M. Enhancing K-means Clustering Algorithm with Improved Initial Center[J]. International Journal of Computer Science & Information Technologies, 2010, 1(2):121-125.
- [30] Dehariya V K, Shrivastava S K, Jain R C, Clustering of image data set using k-means and fuzzy k-means algorithms[C] Computational Intelligence and Communication Networks (CICN), 2010 International on.IEEE , 2010: 386-391.

- [31] Aldahdooh R T , Ashour W.DIMK-means" Distance-based Initialization Method for K-means Clustering Algorithm"[J]. International Journal of Intelligent Systems and Applications, 2013, 5(2):41.
- [32] Jose J. Lopez, Jose A.Aguado,F.Martin,F.Munoz , A.Rodriguez , Jose E.Ruiz.Hopfield-K-Means clustering algorithm : A proposal for the segmentation of electricity customers[J]. Electric Power Systems Research, 2011,(81):716-724.
- [33] 黄红伟, 黄天民. 基于网格和图论的初始聚类中心确定算法[J]. 计算机应用与软件, 2015(2):228-231.
- [34] Zhang GZ, Gao J. The customer segmentation study based on Data Mining of CRM[J]. Journal of Marketing, 2005: 24-25.
- [35] 吴小红. 综述客户细分的方法与技术[J]. 科技信息, 2012(1):110-111.
- [36] LICHTENSTEIN S , SLOVIC P The construction of preference[M]. New York: Cambridge University Press, 2006.
- [37] 罗仕鉴, 李文杰, 傅业焄. 消费者偏好驱动的 SUV 产品族侧面外形基因设计[J]. 机械工程学报, 2016, 52(2):173-181..
- [38] 刘英. 可持续消费行为研究的新视角:基于行为阶段变化理论[J]. 消费经济, 2016(3):57-61.
- [39] 潘晓靖. 大数据时代下房地产行业的客户生命周期管理[J]. 财经界:学术版, 2015(3):125-125.
- [40] 席景科, 谭海樵. 空间聚类分析及评价方法[J]. 计算机工程与设计, 2009, 30(7):1712-1715.
- [41] 谢娟英, 屈亚楠. 密度峰值优化初始中心的 K-medoids 聚类算法[J]. 计算机科学与探索, 2016, 10(2):230-247.
- [42] 戴阳阳, 李朝锋, 徐华,等. 初始点优化与参数自适应的密度聚类算法[J]. 计算机工程, 2016, 42(1):203-209.
- [43] 赵慧, 刘希玉, 崔海青. 网格聚类算法[J]. 计算机技术展, 2010, 20 (9): 83-85.
- [44] 苟光磊, 崔贯勋, 王柯. 基于属性重要性的 Cobweb 算法[J]. 重庆理工大学学报自然科学版, 2012, 26 (12) : 70-73.
- [45] 杨晓岑. 客户价值评价指标体系的构建[J]. 现代经济信息, 2011, 4 (14): 51-52.
- [46] 杨璐熙. 基于客户价值的企业营销策略[J]. 经营者, 2016, 30(10): 34-36.
- [47] 王栖. 浅论如何提高客户的忠诚度[J]. 企业改革与管理, 2016 (15) :

- 221-221.
- [48] 苑世梅, 陈来. 民营物流企业客户价值评价指标体系研究[J]. 物流科技, 2013, (10): 25-27.
- [49] 柳晨. 购物中顾客转换成本的影响因素研究[J]. 经贸实践, 2015 (8): 43-44.
- [50] Han, J, and Kambert, M. Data mining: Concepts and techniques second edition, Morgan Kaufmann, San Francisco, 2006.
- [51] Ahmed, I. and Aziz, A. Dynamic approach for data scrubbing process. International Journal on Computer Science and Engineering (IJCSE), Volume 02, Number 02, 2010, 3: 416-423.
- [52] Rajashree Dash, Rasmita Dash, Debahuti Mishra. A Hybridized Rough-PCA Approach of Attribute Reduction for High Dimensional Data Set [J]. European Journal of Scientific Research, 2010, 44 (1): 29-38.
- [53] 饶静. 基于 SWOT 分析的企业核心竞争力培育研究—以中小型房地产企业为例[J]. 河南财政税务学报, 2015, (4): 34-36.

附 录

宝宇业主客户细分研究调查问卷

您好，我是哈尔滨工业大学管理学院的硕士研究生，并且也是宝宇一期 F 栋 1 单元 25-1 的业主，为保证毕业论文的顺利完成，需要获取一定的数据。希望占用各位邻居较短的时间，了解各位邻居的对宝宇地产的一些消费行为以及意见，请您选出您认为最恰当的评分。本次调查所有数据和结果仅用于论文研究，我们会严格保密，决不会泄露您的个人信息，非常感谢您的参与。

- 1、您的性别： A 男 B 女
- 2、您的年龄范围： A 20-30 B 31-40 C 41-50 D 50 以上
- 3、您的家庭收入： A 3-8 万 B 8-12 万
 C 12-20 万 D 20 万以上
- 4、最近一次购房的金额 万元
- 5、截止目前为止在宝宇地产购房所花费金额 万元
- 6、在宝宇地产购房的次数 次
- 7、推荐亲朋好友购买宝宇地产产品的次数 次
- 8、是否有过银行不良记录，如果有是 次
- 9、是否有过由于某种原因未能及时还款，如果有是 次
- 10、对宝宇地产房屋质量以及物业服务总体满意情况评分
非常不满意 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 非常满意
- 11、对宝宇品牌产品和服务推广宣传以及合理化建议评分
从未推荐 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 积极推荐
- 12、交叉购买宝宇地产意愿评分
不强烈 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 非常强烈
- 13、自我评定未来的您发展潜力
停滞不前 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 巨大发展
- 14、如有机会再次选择您会选择其他楼盘还是继续选择宝宇地产评分
其他楼盘 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 选择宝宇
- 15、宝宇客服人员是否与您建立一对一关系，并且关系是否密切评分
没有建立关系 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 十分密切

哈尔滨工业大学学位论文原创性声明和使用权限

学位论文原创性声明

本人郑重声明：此处所提交的学位论文《基于聚类分析 K-means 算法房地产客户细分研究》，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名：

于阳

日期：2017 年 6 月 25 日

学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

(1) 学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；(2) 学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；(3) 研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名单位为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：

于阳

日期：2017 年 6 月 25 日

导师签名：

苏金峰

日期：2017 年 6 月 26 日

致 谢

随着硕士学位论文的完成，我也即将毕业，走出校门。在工大，我受教于教学态度认真、教学水平高超的老师；在工大，我感受着“规格严格，功夫到家”这一文化软实力；在工大，我结交了积极进取、德才兼备的同学。有太多美好回忆值得抒写，有太多人和事值得感谢。

首先，感谢的是哈尔滨工业大学能给我提供一个宝贵的学习机会，作为一名已经工作多年的学生来说，能再次回到母校攻读研究生，我倍感珍惜，感谢辛勤给我们讲授课程知识的专业课和公共学位课的老师，是你们诲人不倦的精神给我们教授科学研究的方法、学科前沿的知识，为研究生阶段的学习和科研奠定了良好的基础，他们辛勤的培养和教育使我顺利完成学业，并且使我的学习水平和工作能力都有所提升。

其次，感谢的是我的导师芦金锋副教授，攻读硕士学位的两年中，芦老师对我进行了悉心的指导，小到生活状况、课程学习，大到科学探索、论文撰写，芦老师的指导让我受益匪浅。撰写学术论文时，芦老师认真严谨、一丝不苟地耐心的帮助指导我进行论文的修改，并提出宝贵的意见和建议；撰写学位论文时，芦老师在百忙之中抽出大量时间与我进行论文选题的探讨，并对论文的撰写与修改提供了很多有价值的指导和建议，使得我能够顺利完成学位论文工作。

然后，感谢参与论文答辩的老师，他们从专业的角度，积极指出了论文不合理的地方，并提出了改进的意见，给予了我许多写作思路。接着要感谢的是我的同学们在读研期间给予我的大量帮助，在工大，我结交到了优秀的同学，并积极的向他们学习以补自己之短，同时，在平时的生活中，当我遇到困难之时，同学们都积极的伸出援助之手，让我感受到了工管这个大家庭的温暖。

最后要感谢的是我的父母和妻儿一如既往的支持我求学，他们始终是我前进的动力和坚强的后盾，无论是在物质上还是精神上都给予的最大的保障，使我能够安心学业，去追求自己的梦想。

研究生生涯行将结束，我也将离开校园，但我相信，我在哈工大读研期间所学习到的知识、能力以及视野的开阔将会使我受用终生。