

学校编码: 10384

分类号_____密级_____

学号: 15420101151909

UDC _____

厦门大学

硕士学位论文

银行产品营销的数据挖掘方法及其应用

Application of Data Mining to Products Marketing
in Banking

孙宁

指导教师姓名: 方匡南 副教授

专业名称: 统计学

论文提交日期: 2013年4月

论文答辩日期: 2013年 月

学位授予日期: 2013年 月

答辩委员会主席: _____

评阅人: _____

2013年4月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

伴随着经济全球化和高技术的快速发展,现代商业社会已进入到一个全新的知识经济时代,移动互联和大数据等创新应用推动知识以更快的加速度普及扩张,客户消费行为趋向短期化、碎片化、个性化甚至过度追求用户体验,这使得企业开拓新客户和保留既有价值客户的成本越来越高,难度逐步增大;现代企业尤其银行业如需提高竞争力,还需从既有的信息资源和未来的数据储备中去发掘和创造独有的商业优势。数据资产已成为银行企业取得竞争优势的主要武器,包括银行自有的、内部长期积累的历史数据,外部环境中的交换数据以及互联网所蕴藏的海量数据信息,对这些数据进行深入的挖掘有助于分析已知的事实,揭示和预测未知的结果,分析出发展业务所需的关键因素以及发现企业实现跨越式发展的关键点,向数据要策略、向分析要规划已成为经营管理中最重要的手段。

本文以某银行企业的经营数据为背景研究了统计方法结合数据挖掘技术在银行产品精准营销过程中的应用,并着重在银行产品定价规则、客户贡献度和客户分类营销等方面进行了研究。针对银行产品的特点提出精准营销与产品定价之间的相互作用因素;同时通过 K-Means 聚类、Logit 回归、非参数支持向量回归以及 RF 随机森林检验对银行客户群进行分类分析,为银行企业提供价值客户群体的特征描述,并对目标客户群体的营销策略和优惠方案提出建议;然后运用文本检索对银行的外部数据信息进行分析判断,验证了银行潜在客户的分析挖掘过程,并针对此类潜在客户的可推广产品做出了竞争力对比分析,最后实证给出银行产品的精准营销设计思路。

本文通过基于统计学的原理方法对银行产品的精准营销支持研究,推导出以下几个结论:

(1) 客户贡献度与其所处的行业与企业规模关联密切,主要是由于企业对资金和银行产品服务的需求存在行业经济周期性。(2) 占客户总量份额较大的中小型企业特定银行产品上贡献度优于大型客户,比较适合于特定银行产品的营销(3) 外部数据如行业联盟的网站信息或商户所在的虚拟社区数据可以很好地补充银行对其客户的认知,提高潜在客户识别率和精准营销成功率。

关键词：精准营销; SVM; RF; 文本检索

厦门大学博硕士

Abstract

With the rapid expanding of new technology and the deeper globalization tendency around world, Enterprises have introduced more and more innovative application such as mobile internet, big data, into their daily operation. It will push information expand much faster than before. as the result of that ,Consumers have more expectation to the services and products which enterprises will provide, while have lower satisfaction and less loyalty to enterprise at the same time. Consumer's behaviour are more and more rationality ,personality and fragmental even heading to over-expertising.

Enterprises ,especially banking corporation must utilize the data and information accumulated and mine the huge business potential value in it to strengthen enterprise's competitive power .

In this paper, one bank plans to launch a marketing campaign to understand customers entirely as long as sale products and services precisely. Firstly, with statistics model, such as K-Means classify methods and Support Vector Machines (SVM), and Random Forest, based on bank's historical data, we concluded that Random Forest method has the better classify result among all models. On the other hand, we utilized text searching by R language to collect the website information to improve perspective of current customer and potential customer, above it, we realized data mining aiming at products and services marketing.

As the conclusion, the performance of banking customer has correlation with the scale of total customer number; the small medium corporation who has bigger asset volume got better performance than the bigger company. While outer data source from internet or virtual social network will help banks to understand their customer more precisely, avoid more risk ,push effectively marketing campaign as well.

Key Words: Precision Marketing; SVM; RF; Text Searching

目 录

第 1 章 绪论	1
1.1 课题背景及意义	1
1.2 国内外文献综述	5
1.2.1 国外研究现状	5
1.2.2 国内研究现状	6
1.3 研究内容与方法	7
1.3.1 主要研究内容	8
1.3.2 创新点	10
第 2 章 银行产品与客户细分	11
2.1 银行产品	11
2.1.1 利率	14
2.1.2 费率	16
2.2 银行客户模型	17
2.2.1 客户信息模型	18
2.2.2 客户营销	20
第 3 章 数据挖掘方法与模型	23
3.1 数据挖掘	23
3.2 数据预处理	25
3.2.1 数据清洗	25
3.2.2 数据规整	28
3.2.3 衍生变量	28
3.3 聚类分析	29
3.3.1 度量	29
3.3.2 方法	30
3.4 非参数支持向量算法 (SVM)	32
3.4.1 支持向量分类	34
3.5 随机森林组合算法	35
第 4 章 基于内部数据的客户与营销分析	38
4.1 营销产品	39
4.2 数据处理	39
4.2.1 数据清洗	40
4.2.2 衍生变量	42
4.2.3 孤点数据	43
4.2.4 二次分类	44
4.2.5 数据标准化	45

4.3	客户分类模型	48
4.4	客户贡献度模型	51
4.5	初步分类	52
4.5.1	聚类分析	53
4.5.2	检验	54
4.5.3	Logit 模型检验	55
4.5.4	SVM 检验	56
4.5.5	RF 检验	58
4.6	初步分类推论	60
第 5 章 基于外部数据的客户与营销分析		63
5.1	数据定义	63
5.1.1	客户数据	64
5.1.2	产品数据	66
5.2	文本检索	67
5.2.1	文本挖掘	68
5.2.2	分词算法	68
5.2.3	向量空间模型	70
5.2.4	事务相似性	70
5.3	客户发掘与营销	71
5.3.1	数据准备	71
5.3.2	替代产品	75
5.4	推论	77
第 6 章 结论		78
参考文献		79
附件		82
A.1	编码表	82
A.2	R 语言爬虫代码	82
A.3	R 语言文本检索代码	83
A.4	R 语言随机森林代码	83

Table of Contents

CHAPTER ONE: PREFACE	1
1.1 BACKGROUND	1
1.2 LITERATURE REVIEW	5
1.2.1 The Current Situation of Precision Marketing in Worldwide	5
1.2.2 The Current Situation of Precision Marketing in China.....	6
1.3 METHODS AND COMMENTS	7
1.3.1 Methods and Focusing	8
1.3.2 Outlook	10
CHAPTER TWO BANKING PRODUCTS AND CUSTOMER	
SEGMENT	11
2.1 BANKING PRODUCTS	11
2.1.1 Interest Rate.....	14
2.1.2 Service Charge Rate	16
2.2 CUSTOMER INFORMATION MODEL	17
2.2.1 Customer Information Facility	18
2.2.2 Customer Marketing	20
CHAPTER THREE DATA MINING WITH MODELS	23
3.1 DATA MINING	23
3.2 DATA PREPARING	25
3.2.1 Data Checking	25
3.2.2 Data Specification	28
3.2.3 Derive Variables	28
CLASSIFICATION ANALYSIS	29
3.3.1 Variables Measure.....	29
3.3.2 Classification Methods	30
3.4 SUPPORT VECTOR MACHINE MODEL	32
3.4.1 SVM Classification	34
3.5 RANDOM FORESTS	35
CHAPTER FOUR CUSTOMER AND MARKETING ANALYSIS	
BASED ON THE INNER DATA	38
4.1 MARKETING CAMPAIGN	39

4.2 DATA PREPARING	39
4.2.1 Data Clearing	40
4.2.2 Derive Variables	42
4.2.3 Isolating Variables	43
4.2.4 Derive Classification	44
4.2.5 Data Standardlization	45
4.3 CUSTOMER MODELS	48
4.4 CUSTOMER PERFORMANCE MODELS.....	51
4.5 DATA CLUSTERING	52
4.5.1 K-Means Clustering	53
4.5.2 Logit Regression.....	55
4.5.3 SVM	56
4.5.4 RF	58
4.6 CONCLUSION ONE.....	60
CHAPTER FIVE CUSTOMER AND MARKETING ANALYSIS	
BASED ON THE INTERNET DATA	63
5.1 DATA DEFINITION	63
5.1.1 Customer Data.....	64
5.1.2 Product Data.....	66
5.2 TEXT SEARCHING	67
5.2.1 Text Mining	68
5.2.2 Parsing	68
5.2.3 Vector Space Model	70
5.2.4 Corralation.....	70
5.3 CUSTOMER AND MARKETING MINING	71
5.3.1 Data Preparing	71
5.2.3 Competitive Products Analysis	75
5.4 CONCLUSION ON STEP 2.....	77
CHAPTER SIX SUMMARIES AND OUTLOOKS.....	78
REFERENCE.....	79
APPENDIX	82
A.1 PARAMETERS	82
A.2 R CODE OF WEB CRAWLER.....	82
A.3 R CODE OF TEXT SEARCHING	83
A.4 R CODE OF RANDOM FORESTS	83

第 1 章 绪论

1.1 课题背景及意义

中国经济结构和增长方式正在悄然发生变化,中国银行业的高速增长态势正在趋缓,过去三十年来依靠政府投资拉动、低成本劳动力及出口扩张推动的经济发展速度已逐步放慢;同时,金融行业的国际化进程从资本压力、经营竞争、商业模式调整等多方面带给各家商业银行全方位的挑战,虽然自加入世贸以来尤其是过去的 5 年中国银行业有了长足的发展,五大国有银行屡屡出现在世界二十大银行的名单中,但资产规模靠前并不代表我国的商业银行具备了与之相匹配的整体实力和竞争力的优势,相反,全面的经营管理能力不足、缺乏开拓创新力、科技水平与管理理念错位等问题的存在,已日渐成为国内银行企业做强做精的主要障碍。以有续经营为例,国内商业银行的固有盈利模式,是通过简单粗放的资本和市场扩张来增加资本规模与客户,从存款利率和贷款利率固有的利差空间中获取稳定的收益,同时,很多国内银行尤其是地方商业银行承担着各级政府的财务供给和债务支持的责任,因而享有来自政府天然的规模负债来源和资产业务担保,造成对大型企业级客户尤其是当地支柱企业的严重依赖,客户的结构性风险和单一性风险较大,甚至于忽视对整体资产质量的监控;而对其他客户尤其是占比较多的中小型企业与个人客户则多以简单归类的方式,如按照统计 AUM(平均资产负债)的方式来分门别类,对潜在价值客户的管理维持在提供低水平服务和原始银行产品的层面上,或仅能提供缺乏创新性的同质化产品,使得这部分具有潜力的价值客户很难得到挖掘增长。

随着宏观经济及外部环境的持续变化,中国银行业几十年来所拥有的制度保护伞行将消失,2012 年 6 月 8 日和 7 月 6 日,人民银行宣布国内金融机构人民币存贷款基准利率及其浮动区间经历了两次调整,截至目前存款利率和贷款利率的浮动区间分别为基准利率的 1.1 倍和 0.7 倍。这是 2004 年人民银行对存贷款上下限的调整以来利率市场化的第二步调整,存款利率的上浮和存款利率的下调将极大冲击银行的盈利能力,迫使银行必须以更有竞争力的服务来争取客户、争夺

资金、赢得市场。

利率市场化改革带来的另一个冲击是金融脱媒，2012 年企业的债券融资已经成为仅次于银行的第二大融资渠道，基本上所有省市当地核心战略客户都成为具备发债资格的企业。这些企业不再需要通过银行借钱来实现融资，加之金融市场的其他行业如信托业的快速膨胀，严重压缩了商业银行传统的融资业务生存空间。这种情况下要想在传统业务中保持既有优势和竞争力，银行就必须要在深度认识客户和挖掘银行衍生业务当中加大投入，贴近企业客户的经营环节和生产供应链，跟踪零售客户的日常财务消费支付路径。简而言之，更准确、更透彻地了解客户才能够创造出匹配度更好、性价比更高的银行产品及服务。这些对银行的营销战略和执行手段提出了更高的要求。

国内银行面对的另外一个冲击是以互联网金融为代表的科技脱媒，汇丰银行早在 2010 年就将最大潜在竞争对手定为纯粹的互联网科技巨头，谷歌公司 (Google.Co.ltd)，认为未来银行最大的竞争性威胁并非来自金融行业本身或者周边衍生关联行业，而是信息管理和沟通渠道领域的霸主企业，特别在网络信息领域。自 2011 年始，传统商业活动的电子商务运用在短短几月内迅速蔓延至各个行业，网络采购与支付已成为各个行业争相应对的主战场。而商业模式的改变更深远的影响是社会消费模式的调整，例如凭借搜索引擎、社交网络和云计算等新技术，以 16-46 岁为主力的消费群体中在全球范围内出现了几个重要趋势：(1) 移动支付替代传统支付业务(如信用卡、银行汇款)。例如移动支付运营商 Square 公司自创立起年均增长率为 1200%，2012 年支付总金额超过 100 亿美元，而全球移动支付总金额 2016 年将达到 6169 亿美元，这完全来自于银行的传统市场份额。(2) 互联网撮合直接借贷替代传统存贷款业务。比如 2007 年成立的美国 Lending Club 公司，到 2012 年年中已经促成会员间贷款 6.9 亿美元，利息收入约 0.6 亿美元。国内同步也涌现出很多网络信贷公司如人人贷、宜信融资、拍拍贷、安心贷等公司虽然在监管上仍受一定的限制，但业务规模增长远远超过预期。(3) 众筹融资(crowd funding)，即通过互联网为投资项目募集股本金的融资模式，将替代传统的证券业务。2012 年 4 月，美国通过 JOBS 法案(Jumpstart Our Business Startups Act)，允许美国的小企业通过基于互联网的众筹融资来获得股权资本。可以看到在 B2C (对零售客户) 的领域传统业务的缩减已经不可逆转，而在 B2B

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

廈門大學博碩