

学校编码: 10384

分类号_____密级_____

学号: X2013230725

UDC _____

厦 门 大 学

工 程 硕 士 学 位 论 文

数据挖掘技术在保险业客户细分和购物篮分析的应用研究

Study on Application of Data Mining Technology for Insurance

Customer Segmentation and Basket Analysis

郑秀香

指导教师: 王备战教授

专业名称: 软件工程

论文提交日期: 2015年4月

论文答辩日期: 2015年5月

学位授予日期: 年 月

指导教师: _____

答辩委员会主席: _____

2015 年 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

近年来，随着我国保险行业的不断发展、快速壮大，保险企业已经积累了大量的销售数据，因此如何准确地识别目标客户和最大限度的开发现有客户的潜在价值成为企业的重要研究难题。此时，数据挖掘技术应运而生，其在银行、通讯业、零售业的初步应用受到了广泛关注，并开始快速发展，这为保险企业的研究提供了很好的途径。数据挖掘技术可以帮助保险公司识别出能够给企业带来利润的目标客户，帮助企业完成市场细分，制定决策，使得业务人员能更好地驾驭保险市场。

本文对保险企业在客户识别、客户开发和险种捆绑销售活动中应用数据挖掘技术的相关理论和实施进行探讨。首先简要介绍了本文研究所涉及到的数据挖掘相关理论知识。接着，对核心业务系统的相关数据进行分析探索，设计出客户细分和购物篮分析的数据挖掘模型，然后从现有大量的、模糊的数据中选择并抽取与建模相关的信息，完成数据的理解、预处理、转换等数据准备工作，经过循环往复的数据探索和模型调优，成功建立了客户细分和购物篮分析的数据挖掘模型，最后完成了目标客户识别、客户开发和险种捆绑销售的挖掘任务，从实验上验证了保险行业应用数据挖掘技术的可行性。最后，总结了整个工作研究过程中所得出的主要结论和不足，指出了未来工作和研究的方向。

关键词：数据挖掘；客户细分；购物篮分析

Abstract

In recent years, with the continuous development and the rapid growth of the insurance industry in our country, insurance companies have accumulated a lot of sales data. So, how to accurately identify the target customers and maximally develop the potential value of existing customers has become an important problem of the enterprise. At this time, the data mining technique emerges as the times require. Its preliminary application in banking, communications industry and the retail industry has been widely concerned, and started to grow rapidly, which provides a good way for insurance enterprises to study the problem. Data mining technology can help the insurance company to identify the target customers that can bring enterprise high profit; help the insurance company to complete the market segmentation and decision making; enable the sales staff to better control the insurance market.

This dissertation is the study of the application of related mining technology theory and its implementation for insurance enterprise in customer identification, customer development and insurance binding sales activities. Firstly, this article briefly introduces the related theory of data mining knowledge. Then, the data mining models of customer segmentation and basket analysis have been designed by the analysis and exploration of the relevant data of the core business system. By the select and extract relevant information about data modeling from the large number fuzzy data, the data preparation such as data understanding, pre-treatment, conversion have been completed. The data mining models of customer segmentation and basket analysis have been built after repeatedly data exploration and model optimization. Finally, the dissertation achieves the mining tasks of the target customer identification, customer development and product bundling sales tasks. It validates the feasibility of application of data mining technology in the insurance companies from the experiment. In the last of the dissertation, it summarizes the main conclusions and defects that learned from the whole research process and points out the future work and the research direction.

Keywords: Data Mining; Customer Segmentation; Basket Analysis

厦门大学博硕士学位论文摘要库

目 录

第一章 绪论	1
1.1 研究背景和意义	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 国内外研究现状	3
1.2.1 国外发展现状.....	4
1.2.2 国内发展现状.....	5
1.3 论文的研究内容	5
1.4 论文的组织结构	6
第二章 数据挖掘相关技术	8
2.1 数据挖掘的定义	8
2.2 数据挖掘的对象	8
2.3 数据挖掘的功能	8
2.4 数据挖掘的过程	9
2.4.1 商业理解.....	10
2.4.2 数据理解.....	10
2.4.3 数据准备.....	10
2.4.4 建立模型.....	11
2.4.5 模型评估.....	11
2.4.6 结果部署.....	11
2.5 数据挖掘的算法	12
2.5.1 Apriori 算法.....	12
2.5.2 K-means 算法	14
2.6 数据挖掘的工具	16
2.7 本章小结	16
第三章 需求分析与模型设计	17

3.1	需求分析	17
3.1.1	现有系统功能分析.....	17
3.1.2	数据挖掘功能需求分析.....	17
3.2	可行性分析	19
3.2.1	数据可行性.....	19
3.2.2	技术可行性.....	20
3.3	数据挖掘对象的设计	20
3.3.1	客户细分模型.....	20
3.3.2	购物篮分析模型.....	23
3.4	数据挖掘过程设计	24
3.5	本章小结	26
第四章	数据挖掘模型的实现	27
4.1	客户细分模型的实现	27
4.1.1	数据准备.....	27
4.1.2	建立模型.....	37
4.1.3	模型解读.....	40
4.1.4	业务建议.....	45
4.2	购物篮分析模型的实现	47
4.2.1	数据准备.....	48
4.2.2	建立模型.....	50
4.2.3	模型解读.....	52
4.2.4	模型评估.....	52
4.2.5	业务建议.....	53
4.3	本章小结	53
第五章	总结与展望	54
5.1	总结	54
5.2	展望	55
	参考文献	56

致 谢.....58

厦门大学博硕士论文摘要库

Contents

Chapter 1 Introduction.....	1
1.1 Research Background and Significance	1
1.1.1 Research Background	1
1.1.2 Research Significance	2
1.2 Domestic and Foreign Research Profile	3
1.2.1 The Development of Foreign	4
1.2.2 The Development of Domestic	5
1.3 Contents of Dissertation.....	5
1.4 Structure of the Dissertation	6
Chapter 2 Data Mining Technology	8
2.1 The Definition of Data Mining	8
2.2 The Object of Data Mining.....	8
2.3 The Function of Data Mining	8
2.4 The Process of Data Mining.....	9
2.4.1 Business Understanding	10
2.4.2 Data Understanding	10
2.4.3 Data Preparation.....	10
2.4.4 Modeling	11
2.4.5 Evaluation	11
2.4.6 Deployment.....	12
2.5 The Algorithms of Data Mining	12
2.5.1 Apriori Algorithms	13
2.5.2 K-means Algorithms	14
2.6 The Tool of Data Mining	16
2.7 Summary	16
Chapter 3 Requirements Analysis and Model Design	17

3.1 Requirement Analysis	17
3.1.1 The Functional Analysis of Current System	17
3.1.2 The Requirement Analysis of Data Mining	17
3.2 Feasibility Analysis	19
3.2.1 The Feasibility Analysis of Data.....	19
3.2.2 The Feasibility Analysis of Technology.....	20
3.3 The Design of the Object of Data Mining.....	20
3.3.1 Customer Segmentation Mode.....	20
3.3.2 Basket Analysis Model	23
3.4 The Process Design of Data Mining	24
3.5 Summary	26
Chapter 4 Data Mining Model Implementation	27
4.1 Implementation of Customer Segmentation Model	27
4.1.1 Data Preparation.....	27
4.1.2 Modeling	37
4.1.3 Model Explanation.....	40
4.1.4 Business Advice	45
4.2 Implementation of Basket Analysis Model.....	47
4.2.1 Data Preparation.....	48
4.2.2 Modeling	50
4.2.3 Model Explanation.....	52
4.2.4 Model Evaluation.....	52
4.2.5 Business Advice	53
4.3 Summary	53
Chaper 5 Conclusions and Outlook	54
5.1 Conclusions	54
5.2 Outlook	55
References.....	56

Acknowledgements58

厦门大学博硕士学位论文摘要库

第一章 绪论

1.1 研究背景和意义

1.1.1 研究背景

近年来，我国保险企业在迅速发展的同时，也面临着更加严峻的考验。

1. 保险业发展迅速

在 2015 年的保险监管工作会议上，已经明确指出在党中央国务院的正确领导下，我国保监会继续坚持把释放保险企业的发展潜力作为监管工作的重要着力点，通过改革推动、需求拉动、政策驱动三管齐下，坚持稳中求进、稳中有为，2014 年的保险市场实现强势发展，目前保险行业的发展已经驶入快车道。

(1) 业务增速持续提高。2014 年保险业总资产已经超过 10 万亿元，全国的签单保费总收入突破 2 万亿元，保险业增速达到 17.5%。

(2) 结构调整走向深入。不管是保证保险还是年金保险等都实现了大幅增长。其中增长最快的是年金保险，其同比增长了 77.2%。

(3) 行业经营效益显著提升。保险资金运用实现收益同比增长了 46.5%；资金投资收益率达到 6.3%，同比增长了 1.3%；综合收益率达到 9.2%，同比增长了 5.1%。

(4) 行业资本实力明显增强。到 2014 年底，全行业净资产超过 1.2 万亿元，同比增长了 56.4%^[1]。

2. 保险业市场竞争激烈

据中国保险年鉴统计如下，

(1) 从资本结构属性看，中资保险公司有 89 家，外资保险公司有 54 家；人寿保险公司有 71 家，财产保险公司有 64 家，再保险公司有 8 家。

(2) 从公司规模看，我国共有 170 家一级分公司，73819 家二级及二级以下营业机构，保险从业人员有 37.42 万人^[2]。

快速发展的保险市场，使我国的保险企业面临着国内市场和国际市场的双重竞争。保险公司对客户资源的争夺战愈演愈烈，但新市场的开发却变得越来越难。

3. 保险企业发展需要

(1) 随着我国保险业的快速发展，很多保险公司都被快速增长的数据信息包围着，这些数据包括了承保系统、理赔系统、客服系统以及电销系统等积累的大量客户数据、销售记录以及语音和图片等。

(2) 随着社会经济的发展，现在客户的需求和消费观念等都发生了潜移默化的变化，越来越多的客户日益不再满足标准化、面向大众的共性产品，而更期待是个性化的产品。

(3) 保险企业生存并且能够发展的基础是客户，真正带来高利润的客户更是各家保险公司之间相互竞争的对象。因此，从慢慢积累下来的大数据中，挖掘出客户的真正需求，才能够帮助决策者有制定有针对性的销售策略，最终辅助企业实现最大的利润。可见，数据挖掘技术在保险业的应用已成为必然趋势，其对保险企业的发展有着重要的推动作用。

1.1.2 研究意义

根据一家调查机构的数据显示，挽留一位老客户的成本不到开发一位新客户的成本的三分之一；向老客户销售新商品能达到 50% 的成功率，而向新客户销售新商品却只有 15% 的成功率。如果每年将客户的流失率降低 5%，企业的利润将达 25%-85%；而且企业 60% 的新客户是来自老客户推荐或者介绍的。这些数据都说明现有老客户价值要引起高度重视。根据二八原理，保险公司 80% 的利润来自于 20% 的客户。因此，如果想要增加企业利润，就必须识别出这 20% 的黄金客户。由此可知，对客户进行识别和细分、研究客户的行为模式是至关重要的，直接关系到企业的利润收入。

国内保险企业经营积累多年，获取了海量的客户的业务数据信息、自身经营和外部获取的基本客户资料情况、客户对产品和服务的反馈信息。目前，保险公司对于数据信息的使用主要都处于基础阶段，集中体现在对于数据原始信息的查询和简单的统计，而对于数据的认识程度和利用程度都不够充分，无法在客户消费行为的预测、在客户对投保产品的取向和要求等环节上得到更充分的数据进行支持。所以，如何基于数据挖掘建立模型，运用聚类技术开发和识别不同类型的客户群组，运用关联规则技术针对客户购买产品的组合进行分析，利用神经网络

技术预测和评估客户可能的消费行为等有着极大的应用前景。具体表现有以下三点：

1. 识别客户群特征

随着经济的发展以及时代的进步，人们对保险产品的要求越来越多样化、越来越个性化，而保险企业又不可能付出巨大的成本，一一对每个客户的需求进行分析统计。所以，对客户进行细分是符合保险公司需要的，传统主观的细分方法已经不再适用时代发展了，而数据挖掘是以大量的客户及其交易数据为基础，通过合适的算法构建具体的模型，得到更加合理准确的客户细分结果。保险公司就可以根据细分模型结果，向不同的客户提供差异化的服务，最终实现客户满意度的提高，公司利润的增长。

2. 预防欺诈行为

如果保险企业是先根据投保客户年龄的年龄段、所从事的职业等计算分析客户发生各类意外事故的概率，然后来制定合理的保费和保额等，从原则上来讲，保险公司的赔付率就不会很高，更不会亏损，但是难免有极少数个别客户会有欺诈行为。为了从源头上制止这类欺诈客户投保或者对已投保的这类客户进行及时预警，这些都需要数据挖掘的偏差检测等技术分析客户欺诈产生的可能变量等，帮助保险公司来控制这类欺诈行为的产生。

3. 整体把握产品和客户的好坏

企业只有了解公司的盈利险种和亏损险种，“正能量”客户和“负能量”客户，才能获得足够好的经营效益。而数据挖掘能帮助保险公司完成险种和客户的具体细分，回收高风险产品和减少高赔付率客户，降低企业经营风险。

综上所述，从保险企业发展的各个方面来看，本文利用数据挖掘技术对目标客户细分和购物篮分析及其在保险行业的应用具有现实意义^[3]。

1.2 国内外研究现状

数据挖掘出现于 20 世纪 80 年代末，最开始是在数据库领域发展起来的。如今数据挖掘的应用领域已经非常广泛了，只要一个企业，它有分析价值与需求的数据库，就可以利用数据挖掘技术及其工具进行符合企业实际需要的挖掘分析。目前零售业、制造业、银行业和通讯行业已有不少成功应用案例，而且保险企业

对于数据挖掘的应用也开始越来越多。

1.2.1 国外发展现状

在国外，数据挖掘技术发展快速，它已被广泛的应用于各个领域。在国外的保险业中，数据挖掘也已有数例成功案例，如表 1.1 所示。

表 1.1: 国外保险公司利用数据挖掘的成功案例

国家	保险公司	数据挖掘应用
美国	All State Financial HCSC Empire Blue Cross	为了保证有在最能给公司创造收益的客户集中地区来运营，通过数据挖掘技术，分析不同目标客户的消费特征，制定销售策略和交叉销售策略 通过数据挖掘技术提高了识别欺诈性的健康医疗索赔的准确性和效率，不仅大大降低了投保人的保险费用，还帮助公司每年减少几百万美元的赔付 利用数据挖掘技术，1997 年为公司共计节省了 3850 万美元的理赔支出，同时也根据数据挖掘的模型成功告发了一些不实开立医疗凭据的医生
法国	Groupama Coface	通过数据挖掘技术，挖掘目标客户行为和价值费率及理赔策略的制定前会先根据数据挖掘出的客户风险等级，进行评估
澳大利亚	Allianz Elementer HIC	对客户忠诚度和客户流失率进行数据挖掘，提高客户的保持度 利用统计，分类，关联规则以及回归等挖掘方法预测医疗保险欺诈
土耳其	AXA OYAK	不仅利用数据挖掘技术提高公司的市场份额和销售利润，还通过数据挖掘分析降低赔付风险
西班牙	AXA Seguros Inversiones	利用数据挖掘技术，分析市场机会，分析政策变动对消费者行为的影响，不仅减小政策风险，甚至为客户提供适合他们的保险产品

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts 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.