**湖 南 信 息 学 院**

**数据挖掘课程论文**

|  |  |
| --- | --- |
| **论文题目** | **数据挖掘在保险业客户识别中的应用** |
| **学生姓名** | **谭雅华** |
| **学号** | **201501420332** |
| **二级学院** | **电子信息学院** |
| **专业名称** | **软件工程** |
| **班级** | **三班** |
| **指导教师** | **张钰莎** |

二〇 一八年 十月二十二日

**摘 要**

近 10 年来保险行业不断发展和壮大，经过长时间运营的保险公司已经积累了大量的历史销售数据。与此同时，数据挖掘技术在近十年间也受到广泛关注而得到了快速的发展。保险公司大量的历史数据在数据挖掘技术的支持下正好能得到最大的利用。数据挖掘技术可以帮助保险公司识别出能为保险公司带来高利润的目标客户，帮助保险公司完成市场细分，帮助制定决策，使得保险销售人员能更好地驾驭市场。

本文针对保险业的特点，对保险客户销售与一般客户销售的不同性进行对比，对保险行业在客户识别方面进行数据挖掘技术的相关理论的研究。

**关键词**：数据挖掘；保险；客户识别

**ABSTRACT**

The insurance industry has been rapid development in the past 10 years. Long time's operation makes the insurance companies accumulated a large number of historical sales data.

At the same time, data mining has been rapid development in the past decade too. A lot of historical data in Insurance companies can support the work of data mining, and let the historical data in Insurance companies to get the maximum use. Data mining technology can help the insurance company complete market division, recognize the high value customers, and make the insurance staff have a better marketing control.

Based on the characteristics of the insurance industry, this paper compares the differences between the sales of insurance customers and that of general customers, and conducts the research and practical exploration on the relevant theories of data mining technology in the insurance industry in the aspects of customer identification and bundled sales activities of insurance types.

**Key words**:Data mining；insurance；Customer identification

# 目 录

**[摘 要 II](#_Toc31876)**

[ABSTRACT III](#_Toc27558)

**[第一章 绪 论 1](#_Toc804)**

[1.1 研究背景与研究意义 1](#_Toc31578)

[1.2 国内外研究现状 1](#_Toc1848)

**[第二章 数据挖掘相关知识介绍 2](#_Toc3594)**

[2.1 数据挖掘的定义 2](#_Toc8832)

[2.2 数据挖掘的流程 2](#_Toc12748)

[2.3 数据挖掘的主要算法 3](#_Toc1932)

[2.3.1 决策树算法 3](#_Toc32338)

[2.3.2 Apriori算法 3](#_Toc26915)

[2.3.3 K-means算法 3](#_Toc21941)

**[第三章 数据挖掘在保险业客户分析中的应用 4](#_Toc31137)**

[3.1 客户识别 4](#_Toc16100)

[3.1.1 客户细分 4](#_Toc21043)

[3.1.2 目标客户分析 6](#_Toc19565)

[3.3 本章小结 7](#_Toc26876)

**[结 论 8](#_Toc14984)**

**[参考文献 9](#_Toc13677)**

# 第一章 绪 论

## 1.1 研究背景与研究意义

在经济全球化的带动下，人们收入水平不断提高，人们的保障意识也在不断加强，购买保险产品已经成为人们日常一种较为稳健的并能为生活提供有效保障的投资理财方式，所以，近 10 年来保险行业不断壮大并且具有很广阔的发展空间。

近年来，随着数据挖掘技术的兴起和蓬勃发展，数据挖掘技术已经在在银行业与零售业等多个行业都取得了成功的应用。保险公司也开始意识到原来自己企业积累了多年的保单销售数据原来是一块大矿石，只要能够正确运用，对于企业来说，将会是一笔不可多得的宝贵资源。因此，保险公司纷纷开始大力研究数据挖掘技术，希望能通过对保险业务数据的分析、挖掘，能寻找出隐藏在数据背后的深层次信息，辅助公司找到新的商机，达到增占市场份额、实现利润最大化的根本目标。

保险公司作为一所企业，实现利润最大化是其经营的最终目标。保险公司渴望花费最少的资源寻找到最有价值的目标客户，而数据挖掘技术的出现和发展，使得保险公司的愿望终于有实现的机会。数据挖掘技术能从保险公司数据库内大量的保单销售历史数据中提取出有用的信息和规律，让保险公司从数据库的“奴隶”变身为数据库的“主人”。通过挖掘保险销售系统数据库中的数据来分析客户的特征，如性别、年龄、职业、婚姻状况等，可以帮助保险公司识别具有最大利润值的的目标客户，完成市场细分，使销售人员可以更好地掌握市场。

## 1.2 国内外研究现状

数据挖掘技术首次出现在1989 年，刚开始时被称为数据库中的知识发现（KDD，Knowledge Discovery in Database），在数据库中的知识中发现并在数据库领域中很好的发展。发展到现在，研究重点逐渐发现部分转向系统的应用程序，以及不同学科之间的相互渗透的方法理论，其中包括与软件工程相结合。一个优秀的数据挖掘流程会综合利用各种技术手段，收集不同数据市集的数据进行挖掘，提炼出知识，为决策人员提出决策支持，有利于促进业务的发展。

数据挖掘技术在分析大量数据时优势尤其明显，在国外，数据挖掘的分析技术在金融、保险、证券、银行、电信、零售等领域都得到了广泛的应用。利用数据挖掘技术识别客户的购买行为模式、预测银行或商业上的诈骗行为，都是数据挖掘方面成功的经典的应用例子。

在国内，数据仓库和数据挖掘也开始兴起，不过相对国外的情况，国内的对数据挖掘的应用还处于一个起步阶段，特别是在保险业的应用相对较少。现在国内保险业更多只是停留在事务处理软件的开发上，分析层面上最多也只会生成日常的统计报表。还没有充分利用到数据挖掘技术。现有的文献也主要研究数据挖掘于保险业的保额欺骗和保单流失方面，对目标客户的识别还没有非常突出的例子。

数据挖掘应用在不同领域的已有很多成功的例子，其中成功的经典应用包括：客户群体划分（Customer Segmentation Classification），客户信用计分（Credit Scoring），客

户流失性分析（Churn Analysis），交叉销售（Cross-sell），欺诈发现（Fraud Detection）

等，保险行业也可以借镜这些例子去利用数据挖掘技术解决保险领域中的商业问

题。

# 数据挖掘相关知识介绍

## 2.1 数据挖掘的定义

数据挖掘（Data Mining，DM），又被称为数据库中的知识发现（Knowledge Discovery in Database，KDD）。按照知识发现最基本的定义，数据挖掘只是知识发现的整个过程中其中的一个重要的核心步骤，但由于目前“数据挖掘”一词已经在各行业，媒体和专业的数据库的研究领域都广泛使用并被普遍接受，因此“数据挖掘”一词也广义地被用来表示整个知识挖掘过程，即数据挖掘就是一个从数据库、数据仓库或其它信息资源库的大量数据中发掘出有趣的知识。

## 2.2 数据挖掘的流程

由几个公司组成的联盟开发的与具体产品无关的数据挖掘模型——“跨行业的数据挖掘标准过程”是较为公认的数据挖掘指导模型，其定义一个数据挖掘项目的生存周期应包含六个阶段。数据挖掘的流程如图 2-1 所示。

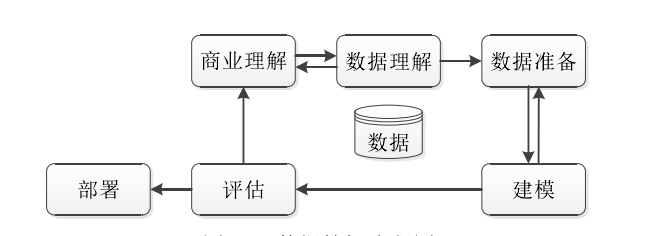


图2-1 数据挖掘流程图

图2-1显示了CRISP-DM 的基本流程，从图可看出数据挖掘一个是以数据为基础、围绕数据进行的不断验证、不断重复的过程，直至得出满意的数据挖掘模型为止。

## 2.3 数据挖掘的主要算法

### 2.3.1 决策树算法

决策树算法主要用于解决分类的问题，可能是最受欢迎的数据挖掘技术；决策树算法是针对因变量与预测变量之间的关系，将因变量进行分层的分析，把最有可能发生的事情预测出来。

### 2.3.2 Apriori算法

Apriori算法 是一种最有影响的挖掘布尔关联规则频繁项集的算法。其核心是基于两阶段频集思想的递推算法。该关联规则在分类上属于单维、单层、布尔关联规则。在这里，所有支持度大于最小支持度的项集称为频繁项集，简称频集。

### 2.3.3 K-means算法

K-means算法是很典型的基于距离的[聚类](https://baike.baidu.com/item/%E8%81%9A%E7%B1%BB" \t "https://baike.baidu.com/item/_blank)算法，采用距离作为相似性的评价指标，即认为两个对象的距离越近，其相似度就越大。该算法认为簇是由距离靠近的对象组成的，因此把得到紧凑且独立的簇作为最终目标。

# 第三章 数据挖掘在保险业客户分析中的应用

## 3.1 客户识别

### 3.1.1 客户细分

客户细分是客户识别的第一步，客户细分是指在已划分好的特定市场中，根据客户的基本特征，以及交易记录将客户细分为具有不同险种消费需求的群体。保险企业若能准确对客户进行细分，不仅能清楚地分辨出高价值客户和潜在客户，还能区别出哪些客户忠诚度较高不易流失，而哪些客户流失倾向严重，这样就可以有针对性地提供保险产品。

而数据挖掘的出现恰好迎合了企业对这方面的需求，特别是决策树分类和聚类这两项技术不仅能够对初步细分后的客户群的客户消费信息进行分析和处理，还能把不同的客户群再划分为更具体的的类，最大化类别之间的差异性，最大化类内的相似度，从而帮助保险企业更加精准定位目标客户并对目标客户的未来消费行为进行预测，实现低成本的客户识别。

本文在客户细分时考虑了客户年收入和年缴保费这两个变量，将其发展为2×2 的客户细分矩阵（见图 3-1）。通过这个细分矩阵，把客户群分为四组，CRM的管理者针对不同的客户细分群体，可以采取以下不同的客户策略。

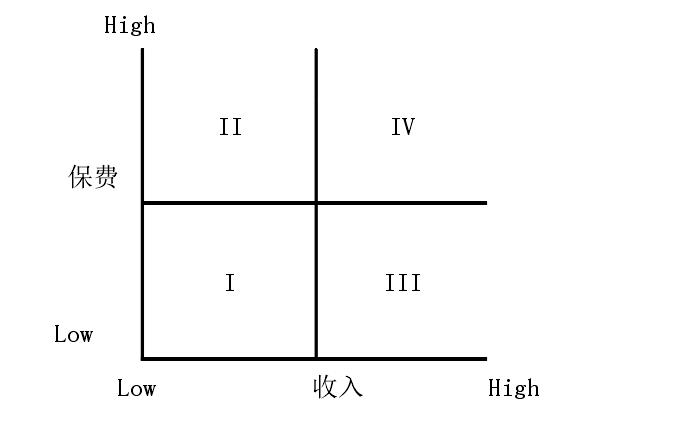


图3-1 基于收入和保费的客户细分矩阵

（1）Ⅰ类客户：该类客户的收入和保费都相对较低，对企业未来的利润增长贡献不大，因此该类客户对保险企业并不具吸引力。

（2）Ⅱ类客户：该类客户具有低收入、高保费特点，由于收入较低，购买保险的能力受限，所以在保险企业的增量销售、交叉销售方面该类客户已经没有多少潜力可挖掘。但同时，该类客户具有比较高的当前价值，是保险企业现有利润的一个重要来源，其利润贡献仅次于第Ⅳ类客户。

（3）Ⅲ类客户：此类客户收入很高，经济实力较好，可能由于消费观念、缺乏保险意识等原因并未大额购买保险，保险企业并没有得到其潜在价值的大份额。为了将该类客户变成企业利润的最大贡献者，保险企业应投入更多的资源以提高对该类客户的服务水平，努力提升客户忠诚和客户关系，进而增加客户的增量购买和交叉购买。

（4）Ⅳ类客户：该类客户是保险企业最具有价值的一类客户，不仅购买保费高，而且还具高收入的特点。从客户生命周期的角度看，保险企业与这类客户的关系保持的最好，客户的忠诚度较高，可能已进入稳定期。与第Ⅲ类客户不同的是，Ⅳ类客户本身就具有极大的发展潜力，他们的消费总量仍在不断增长。

针对上述四种不同的客户群类型，保险企业应采取不同的营销策略，分别为：

1. 低保费低收入客户群，保险企业应少投入 ，甚至不投入，令其关系自动解除，以降低客户的管理成本；
2. 低收入高保费客户群，保险企业应适当投入，对其关系进行再造；
3. 高收入低保费的客户群，保险企业应当重点投入，进行高水平客户关系保留，挖掘其潜在价值；

（4）高收入高保费的客户群应是资源投入的重中之重，作为黄金客户，保险企业应竭尽全力地保留此类客户，努力维系好与该类客户的关系。

客户细分只是进行精准交叉销售的前提，要想得到更加准确的数据挖掘结果，必须对保险企业的所要达到的目标和战略规划进行全方位的理解和分析，这样做就可以尽量避免因分类方案过于复杂导致的方案难以理解、实施等困难。

### 3.1.2 目标客户分析

保险企业的目标客户就是指保险产品或服务的提供对象，包含潜在客户和现有客户两部分。

（1）潜在客户分析

在保险业这个特定领域，那些保险产品和服务的可能购买者可能就是潜在客户，在对潜在客户进行分析时，由于数据量受限，要想利用数据挖掘，可以采取两种解决方案。

第一种方案：对于这些潜在客户，保险企业对他们的了解较少，掌握的客户信息数据有限，因此，难以利用数据挖掘技术进行建模分析。针对此种情况，保险企业应采取某些必要措施来获取潜在客户信息数据。在掌握相关数据后，企业可以根据实际情况进行一定规模的实验，然后再根据反馈信息建立数据挖掘预测模型，寻找公司保险产品的优势，实现成功销售。

第二种方案：从节约时间和成本角度考虑，保险企业可以通过对现有客户基本信息数据的挖掘，找到已购买某类保险的客户具备的基本特征，然后以发现的这些特征为依据，找出那些具有同样特征的人群，这些人群可能就是企业的潜在客户，然后将这些潜在客户转化为实际客户。

（2）现有客户分析

对现有客户的分析其实就是有价值客户的分析，有价值客户就是指那些能给保险企业带来利润增长的客户。企业要想知道某些客户是否能带来盈利，就必须进行客户的盈利率分析。通过盈利率分析，就可以找到企业的“有价值”客户。这些“有价值”客户包括三类：一类是目前给企业带来最大利润的客户，也称之为“高价值”客户，即那些企业投入较少成本就可以获得高收益的客户，保险企业可以针对这些高价值客户进行更深层、更细致的客户关系管理；另一类是现在给企业带来丰厚利润，但未来可能成为最大利润贡献的客户；还有一类则是现在能给企业做出利润贡献，但正逐渐在失去价值的客户。

## 3.3 本章小结

本章立足保险业客户识别现状以及对数据挖掘的各类模型和算法进行研究的基础上，针对保险公司客户关系管理过程中面临的问题，提出引入数据挖掘技术进行客户分析。

通常，客户细分是按照客户的潜在价值和当前价值这两个维度划分的，客户的当前价值比较容易计算，但是潜在价值的计算需要考虑很多因素，在实际中难以操作，数值难以确定。为了便于实施，本文并未按照这一传统方法进行保险客户的细分，而是从客户的年缴保费额度和年收入这两个维度建立了细分矩阵，在某种程度上，这也是本文的一个创新点。

# 结 论

在已经达到白热化竞争程度的保险行业中，各类保险企业若想得到跨越式的发展，就不得不把目光集中于现有客户身上。不断加大向现有客户推广新的保险产品或服务的力度，进一步提升现有客户对保险企业利润增长的贡献价值。

通过运用数据挖掘技术，对保险公司的运作会产生越来越大的影响，随着数据挖掘技术应用的不断深入，它必将在保险公司的各个领域发挥重要的作用，不但能提高风险把控能力，而且对提高公司服务品质、提高市场竞争力、发掘潜在客户等方面也是很有意义的。

# 参考文献

[1] 王保义.客户关系管理中客户细分的数据挖掘研究[D].西安电子科技大学硕士论文,2009,06.

[2]吴昊.面向寿险领域的数据挖掘研究及其应用[D].北京邮电大学硕士学位论文,2010,12.

[3]唐 淑 君 . 保 险 业 客 户 流 失 预 测 模 型 的 建 立 与 实 现 [J]. 软 件 导 刊 ，2011(10):62-64.

[4]Ahmed, S. R.. Applications of data mining in retail business[J]. Information Technology: Coding and Computing, 2004（2）：455–459.

[5]Chen, Y. L., Tang, K., Shen, R. J., & Hu, Y. H. (2005). Market basket analysis in

a multiple store environment.Decision Support Systems, 40 , 339–354.

[6]陈明亮.客户关系管理基础理论体系框架探讨[J]管理工程学报,2006 (4):36-41.

[7]中国保险年鉴编委会.2012 中国保险年鉴[M].中国保险年鉴社, 2012,9.

[8]Ling, R., & Yen, D. C. (2001). Customer relationship management: An analysis framework and implementation strategies. Journal of Computer Information Systems, 41, 82–97.