**湖 南 信 息 学 院**

**数据挖掘课程论文**

|  |  |
| --- | --- |
| **论文题目** | **数据挖掘在电信业的应用** |
| **学生姓名** | **雷利鲜** |
| **学号** | **201501420314** |
| **二级学院** | **电子信息学院** |
| **专业名称** | **软件工程** |
| **班级** | **15软件工程三班** |
| **指导教师** | **张钰莎** |

二〇 一八 年 十 月 十一 日

**摘 要**

数据挖掘就是从大量的、不完全的、有噪音的、模糊的、随机的实际应用数据中，提取蕴含在其中人们事先不知道的、但又是潜在有用的信息和知识的过程。大数据时代，利用数据分析和数据挖掘手段获取企业竞争优势已经越来越常见。一方面企业因为精确管理精确营销的运营管理要求，希望无论营销销售还是服务各层面各环节决策都能以用户/客户为中心，用数据说话；另一方面，由于数据挖掘存在专业性强、建模效率低等问题，而导致对精确营销精确管理运营全流程支撑的力不从心，因此，急需建立相应数据挖掘建模平台，通过集约数据、集约建模、以及模型在不同分公司的复用来提高挖掘建模工作和模型推广应用的效率。

**关键词**：数据挖掘、数据挖掘平台、客户分群、聚类分析

**ABSTRACT**

Data mining is a process of extracting potentially useful information and knowledge from a large number of incomplete, noisy, fuzzy and random data. In the era of big data, it is becoming more and more common to use data analysis and data mining to gain the competitive advantage of enterprises. On the one hand, because of the requirement of precise management and operation management of precise marketing, enterprises hope that all links of marketing, sales and service decisions can be user/customer-centered and speak with data; on the other hand, due to the problems of strong specialization and low modeling efficiency in data mining, precise marketing management is caused. Therefore, it is urgent to establish the corresponding data mining modeling platform, through intensive data, intensive modeling, and model reuse in different branches to improve the efficiency of mining modeling and model promotion and application.

**Keywords:** Data mining, data mining platform, customer clustering and clustering analysis

**目录**

**[第一章 前 言](#_Toc25856_WPSOffice_Level1)**  **[1](#_Toc25856_WPSOffice_Level1)**

**[第二章 客户分析](#_Toc1791_WPSOffice_Level1)** **[2](#_Toc1791_WPSOffice_Level1)**

[2.1 客户行为分析系统架构](#_Toc1791_WPSOffice_Level2)  [2](#_Toc1791_WPSOffice_Level2)

[2.2 客户行为分析系统应用](#_Toc24479_WPSOffice_Level2) [2](#_Toc24479_WPSOffice_Level2)

[2.3 交叉/ 提升销售模型](#_Toc29362_WPSOffice_Level2) [3](#_Toc29362_WPSOffice_Level2)

**[第三章 电信领域的数据挖掘模型](#_Toc24479_WPSOffice_Level1)**  **[4](#_Toc24479_WPSOffice_Level1)**

[3.1 客户分群](#_Toc2898_WPSOffice_Level2) [4](#_Toc2898_WPSOffice_Level2)

[3.2 客户消费模式分析](#_Toc4567_WPSOffice_Level2) [4](#_Toc4567_WPSOffice_Level2)

[3.3 客户市场推广分析](#_Toc31560_WPSOffice_Level2) [4](#_Toc31560_WPSOffice_Level2)

[3.4 客户欠费分析和动态防欺诈](#_Toc15721_WPSOffice_Level2) [4](#_Toc15721_WPSOffice_Level2)

[3.5 客户流失分析](#_Toc23384_WPSOffice_Level2) [4](#_Toc23384_WPSOffice_Level2)

[3.6 大客户的识别](#_Toc30082_WPSOffice_Level2) [4](#_Toc30082_WPSOffice_Level2)

[3.7 客户利润分析](#_Toc10323_WPSOffice_Level2) [5](#_Toc10323_WPSOffice_Level2)

**[结 论](#_Toc29362_WPSOffice_Level1)** **[6](#_Toc29362_WPSOffice_Level1)**

**[参考文献](#_Toc2898_WPSOffice_Level1)** **[7](#_Toc2898_WPSOffice_Level1)**

第一章 前 言

数据挖掘技术主要是基于人工智能、机器学习、统计学等技术，高度自动化地分析企业原有的数据，做出归纳性的推理，从中挖掘出潜在的模式，预测客户的行为，帮助企业的决策者调整市场策略，减少风险，做出正确的决策。

　数据挖掘技术有如下特点：①数据挖掘技术适合于对海量数据进行分析处理。②数据挖掘技术方法多、实用性强，可以有针对性地提高客户管理中的很多决策问题。数据挖掘任务一般可以分两类：描述和预测。描述性挖掘任务刻划数据库中数据的一般特性；预测性挖掘任务在当前数据上进行推断，以进行预测。

这样，在“数据过剩”和“信息爆炸”的同时，人们又感到“信息贫乏”和“数据关在牢笼中”，人类正被信息淹没，却饥渴于有用信息的提取。面临浩渺无际的数据海洋，人们迫切需要一种自动地和智能地将待处理的数据转化为有用的信息和知识的方法， 从而达到为决策服务的目的。

第二章 客户分析

2.1 **客户行为分析系统架构**

包含数据挖掘分析功能层、数据挖掘分析流程与模型管理层和客户行为分析应用层３个层次。数据挖掘分析功能层为客户行为分析应用层提供基础挖掘分析功能， 包括数据预处理、分类及预测、聚类分析、社会网络分析、关联挖掘、序列挖掘和图挖掘等。数据挖掘分析功能层的实现与具体应用无关， 每种具体应用（例如客户细分）都可以通过调用由数据挖掘分析功能层提供的基础挖掘功能来实现。数据挖掘分析流程与模型管理层则定义了数据挖掘分析的流程， 从而形成客户行为分析应用， 并对分析产生的模型进行存储访问管理。客户行为分析应用层包括各种针对客户行为的分析应用，例如客户生命周期价值分析、客户细分模型、客户离网分析、重入网分析、交叉/ 提升销售模型、客户投诉分析、客户欠费预测、欺诈检测、客户信用评分模型、竞争对手分析模型和客户消费行为分析等。

**2.2 客户行为分析系统应用**

客户行为分析系统可以支持多种分析应用，下面重点介绍客户生命周期价值分析、客户细分模型、客户离网分析、重入网分析和交叉/ 提升销售模型。数据挖掘技术可以帮助企业很好地完成上述各种客户行为分析任务。

在电信行业中， 客户生命周期是指电信客户从成为电信公司的客户并产生业务消费开始，到消费成长、消费稳定、消费下降， 直到最后离网的过程。客户生命周期模型全面考虑了客户的当前价值、潜在价值和忠诚度，为企业维护好客户关系，发现高价值客户和挽留高价值客户提供了依据。该模型使用聚类的方法帮助企业发现高价值客户的普遍特征，为发现和开发新的高价值客户提供参考。

客户生命周期价值分析包含两个阶段： 计算客户生命周期价值的３个维度（即客户当前价值、客户潜在价值和客户忠诚度）和基于这３个维度的聚类分析。在第二个阶段可以选用合适的聚类算法来对３个维度进行聚类分析。选择聚类算法的依据是算法速度要快，能够处理大规模的数据集合，并且可以处理孤立点问题，避免孤立点对结果造成很大影响。由于客户生命周期价值的３个维度都是数值类型，所以不用考虑

非数值类型数据的处理。综合上面的考虑， 推荐使用的聚类算法包括DBSCAN 算法、K-Means 算法和层次划分算法等。

客户生命周期价值分析的评估指标分为两个部分， 第一部分是对客户生命周期价值的３个维度计算结果的评估，第二部分是在得到３个维度之后对聚类效果的评估。维度计算结果的评估指标包括在计算客户潜在价值和客户忠诚度时， 使用分类算法进行预测的准确率、召回率等。聚类效果的评估包括聚类结果的准确率

（即标准聚类结果与实际聚类结果的对比）和聚类结果的内在效果

（即簇内部元素差异最小，簇间元素差异最大）。

在客户生命周期价值计算的每一步，都可以通过图形化的方法对该步骤的结果进行展现。计算客户当前价值以后， 可以通过曲线图展现客户当前价值人数的分布曲线， 由此可以清晰地看出客户的当前价值的分布规律。计算客户潜在价值和客户忠诚度后， 可以通过饼状图或柱状图等形式展现预测准确率等信息。

在对客户生命周期价值的３ 个维度进行聚类之后， 可以得到客户聚类结果的图形化展示。例如， 可以使用二维数据得到客户在一个二维平面上的分布情况， 其中不同的簇可以用不同的颜色来表示； 或者可以使用三维数据得到客户在一个三维立体空间上的分布情况， 不同的簇同样可以使用不同的颜色来表示，并且可以方便地通过旋转立方体来得到关于客户分布的感性认识。

**2.3 交叉/ 提升销售模型**

交叉销售和提升销售在电信领域的典型应用场景是数据业务的推广，其目的是向顾客销售更多的产品或服务。交叉销售是根据顾客早先的购买信息，发现顾客的多种潜在需求，并推荐销售多种相关的产品( 或服务) 。提升销售是指销售人员通过对客户需求的深度挖掘和对客户工作环境的深度理解， 向客户推荐价值更高的产品， 从而更好地满足客户的需求， 以求达到双赢的目的。

交叉/ 提升销售模型所采用的分析方法主要是数据挖掘中的频繁模式挖掘方法。首先，使用频繁模式挖掘算法从事务集中挖掘出频繁项集， 即那些被用户同时购买的、频率超过一定阈值的项集； 然后，从产生的频繁项集中生成置信度超过一定阈值的关联规则，比如a b → c ，这条关联规则表示同时购买产品a 和b 的用户很可能（概率超过最小置信度）还会购买产品 c 。因此， 可以利用关联规则的方法进行交叉 / 提升销售的分析。分析方法中用到的关联规则挖掘和频繁模式挖掘算法主要包括Apriori 算法、FP-growth 算法和FPclose 算法等。

交叉/ 提升销售模型的分析评估指标包括可解释性和执行效率。可解释性是指算法挖掘结果的可解释性。执行效率是指算法运行的时间效率和空间效率。

挖掘结果的展现可以分成两个部分。第一个部分展示关联规则的分布曲线图和频繁项集个数的分布曲线图，从而帮助了解挖掘出的关联规则和频繁项集的特征。第二个部分展示挖掘出来的关联规则列表。

1. 电信领域的数据挖掘模型

电信的数据挖掘应用开发, 主要进行客户分群、客户流失分析、客户发展分析、客户行为分析等数据挖掘专题分析, 逐步实现业务预测和信息挖掘等功能。

**3.1 客户分群**

客户分群将一个大消费群体划分为一个个细分群体, 同属一个分群的消费者相似, 而隶属于不同分群的消费者被视为不同的群体。企业根据客户提出的要求和实际所做的不断地改善产品和服务, 从而不断提高该客户群的满意度。

**3.2 客户消费模式分析**

从消费能力、消费习惯、消费周期等诸方面对客户的话费行为进行分析和预测, 从而为相关经营决策提供依据。

**3.3 客户市场推广分析**

利用数据挖掘技术实现优惠策略的仿真, 根据数据挖掘模型进行模拟计费和模拟出账, 其仿真结果可以揭示优惠策略中存在的问题, 并进行相应的调整优化, 以达到优惠促销活动的收益最大化。

**3.4 客户欠费分析和动态防欺诈**

总结各种骗费、欠费行为的内在规律, 并建立欺诈和欠费行为的规则库。当客户的话费行为与库中规则吻合时, 系统可以提示运营商相关部门采取措施, 从而降低运营商的损失风险。

**3.5 客户流失分析**

根据已有的客户数据, 建立客户属性、服务属性、客户消费情况等数据与客户流失概率相关联的数学模型, 然后根据此模型来监控客户流失的可能性, 如果客户流失的可能性过高, 则通过促销等手段来提高客户忠诚度, 防止客户流失的发生。

**3.6 大客户的识别**

企业的大客户群体往往是利润的主要来源。识别出大客户, 为他们制定有针对性的措施, 提高大客户的忠诚度, 是电信运营商继续保持领先的关键所在。此外数据挖掘工具不仅能够根据现有的消费量的多少来判断用户是否为大客户, 还应该可以根据现有大客户的资料, 提取大客户的特征, 并发现潜在的大客户。

**3.7 客户利润分析**

分析客户的利润分布情况, 得到带来不同利润的客户的具体特征。系统将客户利润分为高、中、低三种类型。用户可以对高、中、低的具体情况进行设置。

**结 论**

电信行业是一个与经济发展密切相关的行业，一个长期增长的行业， 一个不断变化的行业， 同时更是一个竞争日趋激烈的行业。电信企业要在复杂多变的市场环境中获得更好的生存和发展，一方面，必须不断改进技术， 创新产品， 提升企业硬实力， 另一方面， 也必须以客户为中心，了解客户需求，提升客服质量，增强客户体验，提升企业软实力。

客户行为分析作为一种建立在大量客户数据基础之上并广泛融合各种先进成熟的数据挖掘技术的应用， 可以成为企业提高软实力的得力助手。随着电信企业应用需求的不断增加， 各种新兴的前沿数据挖掘技术将被逐步引入到客户行为分析系统中，而由此为企业带来的效益也必将逐渐呈现。

在如今的生活中，4G时代将会被淘汰，5G时代即将来领，我认为电信行业需要去占据市场，而此时，利用好数据挖掘这门技术，将这门技术运用到此行业各个领域，各个经营业务中去。

总之，电信企业一直以来有将营销销售运营管理从产品导向向客户导向转型，实现精确管理、精确营销/销售/服务的要求，这就要求企业能在现有以产品分析为主要核心思想的数据仓库基础上建立一个以客户为核心组织相关信息的客户统一视图，并在此基础上建设数据挖掘平台，使得今后所有的营销销售活动都基于数据挖掘平台提供的数据来进行，这样既能因更了解客户需求而使销售营销活动更精准，也使全企业营销销售服务等运营管理工作能够真正实现统筹，给客户提供更好体验的同时，提升全企业现有业务运营管理效率；另外，通过使企业分散各处的数据资源能因整合真正体现数据关联性、完整性和持续更新性的巨大价值，才能真正发挥电信较其他行业所不具备的特有的数据资产的核心竞争优势。

**参考文献**

[1]黄叔武,刘建新.计算机网络教程[M].北京: 清华大学出版社,2004.

[2]戴红,王海泉,黄坚.计算机网络安全[M].北京: 电子工业出版社,2004.[3]陶露菁.基于数据挖掘的电信客户分群设计和实现[D].南京大学硕士学位论文,2005.06.

[4]邓晓梅.基于数据挖掘的电信客户细分模型研究[D].大连理工大学硕士学位论文,2006.10.

[5]韩家炜, Kamber M,著. 数据挖掘: 概念与技术[M].范明译.北京: 机械工业出版社,2001.

[6]刘文.凯森(KXEN)商业数据挖掘[EB/OL].www.datom- i.com.

[7]杜冰.数据挖掘在电信行业中的应用[J].科技和产业,2005,11.

[8]贾琳,李明.基于数据挖掘的电信客户流失模型的建立与实现[J].计算机工程与应用,2004,04.