学校编码：10384　　分类号＿＿＿密级＿＿

学号：X2013232007　　UDC ＿

厦门大学

工　程　硕　士　学　位　论　文

数据挖掘技术
在宽带客户流失预警中的应用

The Application of Data Mining Techniques in Early Warning for Loss of Broadband Customer

陈晓龙

指　导　教　师：　刘昆宏副教授

专　业　名　称：　软　件　工　程

论文提交日期：　2016　年　3　月

论文答辩日期：　2016　年　5　月

学位授予日期：　2016　年　6　月

指　导　教　师：＿＿＿＿＿＿

答辩委员会主席：＿＿＿＿＿＿

2016 年 3 月

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

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

摘要

随着三大运营商全业务经营和民营企业迅猛发展，中国宽带市场越发呈现价格战和服务战的竞争格局，电信宽带用户流失的形势日益严峻，用户保有工作已成为宽带业务发展的重中之重。本文旨在研究如何将数据挖掘技术与宽带客户流失预警相结合，利用数据仓库系统以及数据挖掘工具，进行数据挖掘，建立宽带客户流失预警模型，为挽留客户以及针对性营销提供有力支撑。

本文首先简要介绍数据挖掘、数据仓库的相关理论知识，以及数据挖掘工具的选择。然后，应用 CRISP-DM 过程模型，即业务理解、数据理解、数据准备、建立模型、模型评价、模型实施的数据挖掘步骤，构建文章的框架，并对宽带流失的客户进行数据挖掘。同时，将电信 CRM 系统的客户档案，计费系统的用户消费信息以及大数据平台清洗后的用户上网行为信息，加入到数据挖掘的数据源中，利用 SPSS 软件进行数据挖掘的模型建设以及模型评估。最后，应用决策树算法对电信宽带用户数据进行挖掘分析，从中获取影响宽带用户流失的主要因素以及主要规则，为业务部门进行客户挽留提供决策支持信息，同时应用宽带客户流失预警模型从最近一个月的宽带用户数据中挖掘出流失可能性高的用户，为一线营销人员提供针对性营销的目标用户清单，从而有效保留客户。

论文研究的宽带客户流失预警模型已应用于企业针对性营销平台上，有效降低存量用户的离网率，对电信企业宽带市场战略实施具有现实的指导意义，进而有助于提升企业的核心竞争力并持续保持竞争优势。

关键词：数据挖掘；CRISP-DM；客户流失

Abstract

With the rapid development of the three operators of the whole business and private enterprises, broadband market presents as the competition pattern of the price war and service war. The loss of telecom broadband customer is increasingly grim, so how to hold the customer has become a top priority. The purpose of this dissertation is to research how to combine the data mining technology and the broadband loss warning, and use the data warehouse and data mining tools to build the model of broadband churn prediction to provide strong support for retaining customers and targeted marketing.

This dissertation briefly introduces the relevant theoretical knowledge of data mining, data warehouse and the choice of data mining tools firstly. Then, building the framework of this article and digging the broadband customer by the CRISP-DM process model, which includes data mining step such as business understanding, data understanding, data preparation, model building, model evaluation, model implementation steps. At the same time, the dissertation combines the customer file of the telecom CRM system, the information of user consumption of billing system, and the customer access to the Internet which is cleaned by the large data platform, to the data sources for data mining, which build and evaluate the model of data mining construction by the SPSS software. Finally, in order to get the main factors affecting the loss of broadband customer and the main rules, this dissertation apply the decision tree algorithm of telecom broadband user data for data analysis, which provide decision support information for the department of business to retain customers. Apart from that, mining the user data with high possibility of churn from broadband user data at last month by the broadband loss warning model, can provide the target user list of marketing for the front-line sales staff, so as to retain customers effectively.

The prediction model of broadband churn in this research, which can reduce the churn rate of the customer in stock effectively, has been applied to the marketing platform in enterprises. It has practical guiding significance to the telecommunication enterprise wide band market strategy implementation, and then enhances and sustains competitive advantage of the core competitiveness of enterprises.

Key Words: Data Mining；CRISP-DM；Loss of Customer

目录

Contents

第一章　绪论

1.1 电信宽带市场发展现状

在电信运营商改革转型以及三网融合的大背景之下，基础电信运营商也呈现出向全业务运营商转型的趋势, 向全业务运营商转型的关键在于有效整合现有电信业务资源, 并在此基础上开展互联网化的增值服务。在这过程中宽带业务及宽带用户的发展程度对于电信运营商发展的意义不言而喻，与此同时在我国战略型新兴产业转型发展的今天，宽带业务对拉动我国国民经济发展也具有至关重要的推进作用。

从 2011 年国家启动“宽带中国·光网城市”的战略以来，中国电信致力于打造覆盖每一个居民区域、人人可触及的一体化宽带网络。从宏观上讲，中国电信旨在用近三年时间，致力打造出一个全方位、全覆盖、在神州大地每一寸土地都触手可及的一体化宽带网络，为中国宽带长远战略的实施垫定坚实基础。新一轮宽带网络建设、宽带业务以及宽带用户发展正全面展开,而电信运营商正面临着国内、国际日趋多元化、复杂化的竞争态势，用户对商家提供服务的内容、方式以及质量的要求越来越高。因此各电信运营商在市场竞争中不断创造更高价值的同时，也必须更加注重调整经营理念和策略，从而不断提高维系客户关系的管理水平。

当前通信运营商在经营中所面临的一个突出问题是客户流失，这也成为影响运营商经营收入的一个掣肘。客户流失一方面必将造成运营商的收入份额下降、营销成本增加、市场占有率下降等问题；而另一方面，恶意流失会造成客户恶意欠费，带来一些不必要的损失[1]。有关调查表明, 每增加百分之五的“用户保有率”，就可以给电信运营商增加百分之八十五的收入；而挽留老用户的成本比发展新用户的成本低百分之八十；此外就推荐新产品的成功率来看，向老用户推荐也比向新用户推荐的成功机率高很多。由此可见,电信运营商当务之急是要进一步采取措施维持存量宽带客户的关系，唯有加大提升客户挽留力度，并同时防止高价值用户的流失，才能使运营商在激烈的市场竞争处于不败之地，从而为企业带来超额利润。新形势下，市场竞争更加激烈，竞争压力与日俱增，迫切需要提升技术支撑能力，为精细化管理和营销等提供有力的信息支撑。

1.2 数据挖掘技术在电信运营商中的应用

电信行业随着电信体制改革的进一步深化，其竞争也呈现出日趋激烈的趋势。电信行业与其他行业相比，拥有更多有关用户消费习惯、地理位置、个人偏好的相关数据信息。要想在商业竞争中处于优势，电信行业就必须正确分析数据，并最终达到向用户提供更多更好信息服务的目的。电信企业一方面必须要保存用户呼叫数据，以计算资费，另一方面也要对用户数据加以分析，进而得出规律以优化网络。由此可知，数据挖掘分析在电信业的应用价值不言而喻。

作为典型的数据密集行业，电信业务数据量相当庞大，业务系统众多，如果利用手工报表方式等传统的信息获取手段，信息数据在提供的速度、质量、范围等方面都远远滞后于业务决策需求。然而“从电信运营企业庞大的业务处理系统要随时随地地获取信息难度较大，因此一定要引入新的技术，来支持企业业务对信息的需要。”因此，将数据挖掘技术顺利引入电信行业中就显得尤其必要[2]。

基于数据挖掘技术在电信 CRM、EDW 系统的重要性，目前电信运营商大都建立了自己的一套专用的电信业务综合管理信息系统。用户通过运用电信网络综合管理系统、业务处理和业务查询系统，可以十分轻易地获得庞大的客户数据，接着再在电信 CRM 系统的基础上把所有与客户有关的数据信息进行整合，最终形成面向主题的数据仓库（EDW）。最后运用数据挖掘工具可以获得所需要的信息和模式，为经营管理决策提供参考。在电信系统中，数据挖掘技术主要应用于以下几种情形：

一是客户获得。社会经济和科技日新月异的发展带动电信产品种类的日益增多，国内电信市场也打破之前一家独大的垄断格局，因此可供用户选择的空间也愈来愈大，与此同时电信企业对客户资源的争夺也愈演愈烈。电信行业历来是得客户者得天下，在电信市场竞争异常激烈的情况下，发展电信新客户对电信行业可持续发展的重要性不言而喻。电信新客户应当既包括没听说过也曾经并不需要电信服务的人，也包括经营商竞争对手的原有的客户。而数据挖掘的运用可以通过数据挖掘分析辨别出潜在的企业客户群，从而进一步提高市场活动的准确性。

二是交叉销售。目前电信市场竞争呈现出愈演愈烈的态势，电信企业和客户之间的关系变动十分频繁，而电信运营商竞争的关键就是要尽全力保持住原有的客户关系。最佳的客户关系应当包括最长时间保持、最频繁与客户交流的同时应当确保交易获得最大的利润。由此可以看出电信运营商要想获得利益最大化，就必需采取交叉销售的方式对

现有客户资源进行优化。交叉销售是追求企业和用户之间的双赢，用户可以从中享受到运营商所提供的优质服务，而反过来看企业也因客户的重复选择带动销售增长而获益。交叉销售数据挖掘分析可以让企业在用户从前的购买行为中得出更多切实有效的客户信息，进而进一步推定出左右客户决定下一次是否购买的关键所在。

三是客户保持。保持原有的客户成为电信运营商赢得竞争的重中之重。一般可以把企业客户分为以下三种类型：一是根本无价值客户；二是有价值的客户，这种客户群体不容易流失掉；三是价值追求型的客户，这种客户群体不断寻求更多更好的服务。前两种客户群体是旧市场管理理论开展活动的主要针对对象，在现代管理理论中却坚持保持第三类客户才是现代市场活动成败的关键所在。而通过数据挖掘分析可以及时掌握客户最新动态，找出容易流失的客户群体，提高客户挽留的针对性，以积极有效的客户关系管理机制保持住原有的客户群体，这样才能保持企业经济效益的稳步增长。

四是一对一营销。CRM 系统依据客户的具体属性把众多的客户资源分成了不同的种类。因此，运营商可以充分掌握大量及时、有效、准确的数据信息，并以数据仓库为基本单位组建起一系列直接隶属垂直化管理的、包含各项业务指标的完善的数据挖掘分析系统。利用建立的数据挖掘分析系统，为了提高客户满意度，企业可以给不同类别的客户群提供更加多元化且针对性的服务内容，因此企业对客户进行分类并进一步制定出因人而异的经营策略就显得尤为必要[3]。数据挖掘技术能为电信行业的经营管理人员和一线业务人员提供业务咨询与分析、营销策略等服务，从而进一步提升电信的服务质量。

五是分析用户行为。电信企业的主要业务活动（包括建设与维护、市场营销、网络规划，用户注册与放号、计费及用户服务、创造新业务等）产生大量数据，这些数据为奠基起诸如用户信息、呼叫、账单等事务型数据库。有效整合并合理利用这些事务性数据，并通过对用户网络使用行为和使用效率的数据挖掘，进一步推断出用户使用网络频率高的时段、区域及特点分布，并通过数据分析得出用户的使用偏好，进而对网络管理进行相应的调整，以优化资源的分配。

六是欺诈行为分析和异常模式识别。在通信行业的市场竞争日益白热化，客户的需求越加多样化的当下，运用数据挖掘技术可以便捷地分析出具有欺诈行为的用户并同时预防市场欺诈行为。通过数据挖掘技术识别异常模式，可以通过数据挖掘，识别出潜在的盗用他人网络资源的客户以及他们的异常使用方式，通过离群点分析、多维分析、聚类分析等方式，可以发现许多存在异常消费行为的用户，建立异常识别模型，进而保护企业的形象以及企业的资源不受侵害[3]。

1.3 论文的组织结构

第一章绪论部分主要介绍本次论文的研究背景、研究意义以及文章的组织结构。

第二章数据挖掘相关理论及工具部分主要介绍了数据挖掘相关的理论、数据仓库及ETL、数据挖掘工具的选择，数据挖掘相关理论中包括数据挖掘的定义、基本挖掘过程及经典算法。

第三章模型总体设计方案部分主要介绍宽带客户流失预警模型建立前的业务理解以及数据理解，阐述宽带客户流失预警模型的方案设计。

第四章数据准备部分阐述了宽带客户流失预警模型的数据处理过程，利用数据仓库系统的平台数据处理能力，详细描述涉及建模的数据内容与字段的选取。

第五章预警模型建设与评估部分，主要介绍宽带客户流失预警模型的建设，模型的算法选择，应用选择的算法进行数据挖掘后的评价，最后选择最适合的模型，并应用到测试数据中，分析挖掘出影响宽带用户离网的主要因素以及主要规则，从而为业务部门提供决策依据。

第六章总结与展望部分主要总结了论文研究的价值所在，介绍宽带客户流失预警模型建立的全过程，进一步阐述宽带客户流失预警模型对于企业在精确化营销、存量客户保有中工作的积极意义。此章节还对研究所关注的问题进行更为深入地探讨，从而提出优化解决方案。