

分 类 号: TP311

研究生学号: 201053R236

单位代码: 10183

密 级: 公 开



吉 林 大 学

硕士学位论文

电信企业客户忠诚度预测系统的设计与实现

**Design and Implementation of Telecom Enterprise
Customer Royalty Prediction System**

作者姓名: 严浩

专 业: 软件工程

研究方向: 计算机应用

指导教师: 陈思国 副教授

培养单位: 计算机科学与技术学院

2014 年 12 月

未经本论文作者的书面授权，依法收存和保管本论文书面版本、电子版本的任何单位和个人，均不得对本论文的全部或部分内容进行任何形式的复制、修改、发行、出租、改编等有碍作者著作权的商业性使用（但纯学术性使用不在此限）。否则，应承担侵权的法律责任。

吉林大学博士(或硕士)学位论文原创性声明

本人郑重声明：所呈交学位论文，是本人在指导教师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

严浩

日期：2014 年 12 月 2 日

电信企业客户忠诚度预测系统的设计与实现

**Design and Implementation of Telecom Enterprise
Customer Royalty Prediction System**

作者姓名：

专业名称：软件工程

指导教师：陈思国 副教授

学位类别：

答辩日期：2014年11月22日

摘 要

电信企业客户忠诚度预测系统的设计与实现

电信市场的变革，即运营商的合并联盟以及新技术的发展（3G）意味着电信客户享有更多样的选择以及更便捷的服务，从而导致电信运营商之间的竞争越来越激烈。对于所有电信运营商而言，如何在激烈的市场竞争中保持稳定并获得新的发展，是企业管理者所必须思考的问题，顾客关系管理这一议题愈来愈受到企业的重视。因此客户对于企业而言不仅仅是销售和服务的对象，而且变成了企业的一项重大战略资源，是企业在竞争中立于不败的关键因素之一。基于此，本课题选择研究电信企业客户忠诚度预测系统。

电信企业客户忠诚度预测系统基于 J2EE 平台进行构建，为此本课题主要做了如下几项工作：

(1)本文使用 J2EE 实现系统功能，以 SQL Server2005 数据库存储数据信息。为了在线处理客户忠诚度预测业务，系统基于 B/S 模式进行构建。电信企业客户忠诚度预测系统覆盖了客户忠诚度预测管理涉及的方方面面，其由安全管理、客户管理、客户挽留、消费资料管理以及忠诚度预测等功能组成。

(2)本文以时序图的方式对电信企业客户忠诚度预测系统的模块功能进行了详细设计，并为系统构建客户忠诚度预测模型，该模型可为电信企业提供低忠诚度客户名单，以便企业采取针对性的挽留措施。

(3)在设计方案的指导下，完成了电信企业客户忠诚度预测系统的实现，并将其部署到了电信企业。

目前系统已成功应用到某电信企业，经过一段时间的试用，取得了不错的效果，企业的客户流失率得到了降低，同时也降低了企业的客户维系成本。

关键词：

电信企业，忠诚度预测， J2EE 平台， SQL Server 2005 数据库

Abstract

T Design and Implementation of Telecom Enterprise Customer Royalty Prediction System

Telecom market has changed, the development that operators merge Union and new technology (3G) means the telecommunications customers enjoy more choices and more convenient service, which leads to the competition between telecommunication operators is becoming more and more fierce. For all the telecom operators, how to maintain stability in the fierce competition in the market and gain a new development, is the enterprise managers must consider the problem of customer relationship management, the issue is getting more and more attention by enterprises. Thus the customer to the enterprise is not only the object of sales and service, and become an important strategic resource of enterprises, is one of the key factors of the enterprise stand unbeaten in the competition. Based on this, this topic selected research telecom enterprise customer loyalty forecast system.

The customer loyalty of telecom enterprises prediction system based on J2EE platform to build, this paper mainly do the following work:

(1) this paper uses J2EE to realize the function of the system, using SQL Server2005 database to store data information. In order to predict customer loyalty online processing business, the system based on B/S mode for construction. The customer loyalty of telecom enterprises prediction system covering predict customer loyalty management involves in all aspects, which is composed of safety management, customer management, customer management, consumption data and retain loyalty prediction function composition.

(2) taking of the sequence diagram of customer loyalty of telecom enterprises prediction function module of the system carried out a detailed design, and system building customer loyalty forecast model, this model can provide low loyalty customer list for the telecom enterprises, so that enterprises take targeted measures to retain.

(3) in the design of guidance, completed the system to achieve customer loyalty of telecom enterprises prediction, and deploy it to the telecom enterprises.

The present system has been successfully applied to a telecom enterprise, after a period of time of trial, and achieved good results, the enterprise customer churn rate has been reduced, but also reduces the cost of enterprise customer retention

Keywords:

telecom enterprises, loyalty prediction, J2EE Platform, SQL Server 2005

目 录

摘 要	I
Abstract.....	I
目 录	I
第 1 章 绪 论	1
1.1 课题研究背景.....	1
1.2 研究目的与动机.....	2
1.3 国内外研究现状.....	3
1.4 本文主要研究内容.....	4
1.5 论文的章节安排.....	5
第 2 章 系统相关的理论与技术.....	6
2.1 数据挖掘技术.....	6
2.1.1 数据挖掘的概念.....	6
2.1.2 数据挖掘工作流程.....	6
2.1.3 算法选择.....	7
2.2 开发技术	8
2.2.1 J2EE 架构	8
2.2.2 Myeclipse 平台	9
2.3 数据存储平台的选择.....	10
2.4 B/S 架构	11
2.5 本章小结	12
第 3 章 系统需求分析.....	13
3.1 需求概述	13
3.2 功能需求分析.....	13
3.2.1 安全防护.....	13
3.2.2 客户管理.....	15
3.2.3 客户挽留.....	16
3.2.4 消费资料管理.....	17
3.2.5 忠诚度预测管理.....	18
3.3 用例分析	19
3.5 非功能性需求分析.....	20
3.6 本章小结	21
第 4 章 系统详细设计.....	22
4.1 设计原则	22
4.2 系统概要设计.....	23
4.2.1 技术架构设计.....	23
4.2.2 拓扑结构设计.....	24
4.2.3 逻辑结构设计.....	25
4.3 功能模块详细设计.....	26
4.3.1 安全防护模块.....	26
4.3.2 客户管理模块.....	28

4.3.3 客户挽留模块.....	30
4.3.4 消费资料管理模块.....	31
4.3.5 忠诚度预测模块.....	32
4.4 忠诚度预测模型构建.....	33
4.4.1 构建流程.....	33
4.4.2 商业理解.....	34
4.4.3 数据理解.....	34
4.4.4 数据准备.....	36
4.4.5 模型构建.....	39
4.4.6 模型验证.....	42
4.5 安全性设计	42
4.6 本章小结	43
第 5 章 系统实现	44
5.1 实现环境	44
5.2 系统实现	44
5.2.1 导入功能的实现.....	44
5.2.2 短信挽留功能的实现.....	45
5.2.3 数据维护功能的实现.....	47
5.2.4 贝叶斯网络算法实现.....	47
5.3 系统应用	49
5.4 本章小结	51
第 6 章 总结和展望	52
6.1 本文总结	52
6.2 进一步展望	52
参考文献	54
作者简介及在学期间所取得的科研成果.....	56
致 谢	57

第 1 章 绪 论

1.1 课题研究背景

电信市场的变革，即运营商的合并联盟以及新技术的发展（3G）意味着电信客户享有更多样的选择以及更便捷的服务，从而导致电信运营商之间的竞争越来越激烈。对于所有电信运营商而言，如何在激烈的市场竞争中保持稳定并获得新的发展，是企业管理者所必须思考的问题，顾客关系管理这一议题愈来愈受到企业的重视。因此客户对于企业而言不仅仅是销售和服务的对象，而且变成了企业的一项重大战略资源，是企业竞争中立于不败的关键因素之一。

随着市场的不断的扩大，发展新客户变得越来越困难，但是现有客户流失却日趋严重。根据最新的调查指出，某家电信运营商在近 5 年内流失掉 20% 的客户^[1]，这些客户中有许多是对企业具有高度贡献的重要顾客，然而当他们转换服务时，甚至运营商都没有丝毫反应。Frederick E. Reichheld 指出^[2]，企业吸引一名新客户所耗费的代价是保留一名老客户的 5 到 10 倍。同时，一个高忠诚度的老客户对企业的贡献要远远高于刚发展的新客户。因为老顾客更容易挽留，每年耗费额也比较高，同时服务成本比新客户要低，甚至有的客户能为公司免费宣传，介绍新客户。Peppers & Rogers 指出^[3]，企业若可以将客户流失率减少 5%，利润就可以提升 20% 至 30%。在企业不停投资开发新客户的同时，若无法获得更大的利润空间时，企业确实应该重新审视与客户之间的关系：如何维系既有客户，增进顾客的利润贡献度，挽留有流失倾向的客户。这就要求企业从海量的客户资料中找到有价值的信息，找出这些信息之间的关联，从而得到客户的忠诚度指数，而企业可据此作出相关决策挽留客户，可是这是一个非常艰巨的任务。数据挖掘技术正是为解决这样的任务而产生的，它能从海量的数据中找出某些因素潜在的某些联系，以便作出最优决策，预测客户转换服务的倾向。

数据挖掘是一门综合性较强的学科，其融合了统计学、数据库、模式识别以及人工智能等多学科知识^[4]，正由于融合多种先进方法，数据挖掘才得到广泛应用，并受到专家学者的重视，各领域都在使用挖掘技术。目前数据挖掘有许多不同类型的算法，比如说贝叶斯网络以及遗传算法等，各种算法有不同的应用场景和适用范围。贝叶斯网络的优势在于易于理解，可以结合样本数据信息和先验信息，充分利用已有的知识，进行定量和定性的分析。它能够将影响问题结果的因素用图清晰的表示出来，便于理解，具有分

类、聚类以及预测等功能，并且预测效果好，能够处理海量数据。

正是在这样的背景下，本文选择重点研究贝叶斯网络技术，通过对现有电信客户资料的分析，抽取出客户的注册资料、缴费资源以及通话记录，运用数据挖掘技术，经过数据预处理以及转换等，得出影响客户忠诚度的关键因素，从而建立客户忠诚度预测模型。该模型可以根据现有信息得出客户的忠诚度指数，当客户类别为低忠诚度时，可采取措施进行维护，以便降低客户流失率，提升利润。上述模型需要应用到客户忠诚度预测系统中。

1.2 研究目的与动机

目前国内电信业的竞争日趋激烈，电信企业为了吸引客户，不断推出各类高成本的促销手段，如积分等，企图以优惠的条件来拉拢更多的客户使用本企业的服务，然而这些手段均以获取新客户为基本目标，往往就忽略了已经存在的客户，也就是所谓的“旋转门”效应，即在获得新客户的同时，已有的老客户却流失了。事实上这种面向大众的全面性的行销方式，不仅耗费的代价过高，而能起到的作用却非常有限。因此若能将某些方案直接面向忠诚度较低的客户，适时为这些客户提供优惠服务，则更易于挽留客户。对于电信企业而言，若能将有限的资源用在最需要关注的用户群体上，所获得的效益将非常巨大。

然而，电信企业如何找出那些忠诚度不高或可能流失的顾客？传统的工作方式是定期从已流失的客户中选取样本，再以市场调查研究方法，搜集并分析流失的原因、流失的客户特征。这种方法是把流失的客户看成一群集，用统计方法，尝试将流失客户群的特性，与留下来的客户作比较，找寻其差别以及流失原因，然后加以补救。此模式虽然可以找出客户流失的原因并作长期挽救，但在目前分秒必争的市场环境中，在客户流失后几个月才进行挽留，早已失去时效。

现在，电信可以胡利用客户的个人基本资料、缴费资料或者通话资料，比如说年龄、服务时长、消费额以及投诉次数等，将这些因素组合成影响因子，并使用数据挖掘技术，由计算机寻找这些数据之间的联系，容易发掘隐藏的导致客户流失的关键因素。针对客户流失问题，国外银行通常利用这种方式，以自己现有或搜集到的客户基本资料以及通话资料等次级资料用于分析，但是由于运营模式以及环境的区别，直接照搬国外的分析方式应用于国内市场是极其困难的。针对国内电信企业所掌握资料现况，利用现有的运营资料，从事客户忠诚度数据挖掘研究，期望经过对客户大量的交易记录的分析，探索客户转换服务前行为的变化来发现有效的预测模式。

依据实践的经验可知，客户转换服务不是一朝一夕的事情。这些客户在使用企业服务的行为上会发生变化，且有别于正常使用客户。利用数据挖掘技术，如侦查到这个客户的使用情况和之前的使用习惯有所变化，透过分析出其转换服务的征兆（例如卡内余额呈递减趋势，甚至为 0），藉此找出符合这些征兆的客户，从而进行挽留客户的活动，可以有效的减少客户流失。而这些征兆隐含在客户的交易记录中，如何对这些资料进行挖掘并加以运用，是本论文数据挖掘所欲达成的目标。

由上可知，本论文的研究目标在于运用贝叶斯网络技术，建立客户忠诚度预测模型，协助电信企业解决客户流失问题，让电信企业可直接面对低忠诚度、具备高流失倾向的客户群体，并作出相应的挽留措施，以免客户流失。

1.3 国内外研究现状

国外学者很早就开始对客户忠诚度问题进行研究，并形成了一定的理论体系。早在 1990 年哈佛大学的教授就出版了书籍《忠诚效应》^[5]，该书专门研究客户忠诚问题，并明确指出了客户忠诚对企业的重要性，企业高忠诚度客户越多，可获得的利润也就越多。基于此，许多专家学者开始投入精力研究忠诚度问题，目前客户关系管理已成为每个销售类企业关注的焦点。

由于客户忠诚度对企业有非常重要的影响，经过几十年的研究，目前出现了大量的研究成果。Wei^[6]等人从客户的基础数据以及消费行为数据中抽取了对客户忠诚度存在影响的指标，在此基础上通过决策树方法来获得客户的忠诚度指数，该方法存在较高的准确率。Rosset^[7]等人分析了客户价值与忠诚度之间的关联，并使用 logistic 技术构建了忠诚度预测模型，与之前的研究成果相比，其准确率更高。Kim 等人重点关注了电信行业的客户忠诚度问题，其所使用的技术为 logistic，其从电信行业的特点出发构建客户流失 预警模型，该模型的本质通过预测客户的流失倾向指数来决定该客户是否有必要进行挽留。Torsten^[8]等人从德国电信市场的需求出发，利用线性方法对忠诚、满意度以及客户维持三各属性之间的相互作用做了深入的研究，并取得了相应的结论，即提高客户满意度是提升客户忠诚度的唯一途径，只有提供优质的服务才能提升客户忠诚度，才能维系着客户，让客户持续为企业创造利润，客户的流失率也会大大降低。Piew^[9]等人对 CHAMP 算法进行了详细的介绍，该算法使用的是神经网络法，经过一段时间的应用，CHAMP 算法产生了很好的效果，结果表明神经网络法是有效的。Mozer 等人综合了 ANN 技术和收益计算两种方法的优点，构建了忠诚度预测模型，该模型与其他模型相比，准确率更高。从上述分析可以知道，国外在客户忠诚度的方面研究成果特别多，也非常成

熟，并且应用特别广泛，大部分企业都构建了客户关系管理系统用于提升客户忠诚度。

而国内较之国外起步晚了很多，但依然取得了许多不错的成果。邱义堂^[10]很早就开始研究客户忠诚度问题，其融合了决策树以及专家知识等两种技术用于构建预测模型，该模型的最大优点在于预测完成后，可直接将存在流失倾向的客户名单提供给电信公司，从而公司能根据客户名单采取响应措施来挽留客户、维系客户。盛昭瀚等人在深入研究粗糙集方法和 ID3 决策树方法的基础上，构建了预测模型，该模型同样是以预测客户度指数为目标，经过测试集的测试，结果表明应用效果较好。王维佳^[11]等人从客户资料中抽取出了忠诚度指标体系，在这个过程中应用了准确率较高的 RBF 预测工具。姚敏等人在构建模型的过程中使用 ANN 方法和决策树方法，两种方法在模型中的作用各不相同，前者主要对客户数据进行处理，以便转换为模型所需要的数据，而决策树则是进行忠诚度预测。在模型构建完成后，利用测试集对模型进行了测试，测试结果表明模型准确率较高，可满足企业的应用需求。由上述分析可以知道，国内也取得了许多的研究成果，但是与国外相比，差距较大，还需要投入更多的精力来研究客户忠诚度问题，这是提升企业竞争力的有效措施。

1.4 本文主要研究内容

本课题在详细分析客户忠诚度预测需求的基础上，构建了一个客户忠诚度预测系统。下面对本课题的研究内容进行说明，具体包括如下几个方面：

(1)本文首先分析了电信企业引入客户忠诚度预测系统的意义，然后在对比分析现有信息系统解决方案的基础上，决定使用J2EE+SQ; Server2205+B/S的技术组合，基于B/S模式构建系统是为了满足电信企业在线预测客户忠诚度的需求。除此以外，对系统潜在用户进行了研究分析，明确了系统的功能需求，其需包含安全防护、客户资料管理、查询统计模块、消费资料管理以及忠诚度预测管理等功能。

(2)客户忠诚度预测系统的设计和实现。在明确客户忠诚度预测系统需求的基础上，需完成系统的设计工作。在设计阶段，需完成系统架构设计、数据库设计以及安全设计，并以流程图的方式完成功能设计。在设计文档的指导下，完成了客户忠诚度预测系统的实现。

(3)客户忠诚度预测系统的测试。在客户忠诚度预测系统被部署到电信企业之前，必须进行严谨的测试。在测试阶段，需完成功能测试和性能测试两项工作，防止系统部署到实际环境后出现重大问题，需要耗费大量精力来维护系统。

客户忠诚度预测系统存在很多的创新点，具体表现在如下几个方面：

(1)利用数据挖掘技术构建客户忠诚度预测模型

利用客户的开卡资料 and 消费资料来预测客户忠诚度指数，并根据客户忠诚度指数来得到低忠诚度客户名单，由此可采取针对性的营销策略，以便维系老客户，吸引新客户。

(2)实现客户维系功能

当电信客户的忠诚度较低时，工作人员可利用维系功能中短信维系和邮件维系功能向相应的客户发送优惠信息以及节日祝福，这种营销策略更有针对性。

(3)构建完整的备份方案

系统中存储着电信客户信息，这些都是电信企业的核心。为了提高系统数据的可靠性，本课题引入了完整的数据备份方案。当遇到数据库故障时，可利用备份文件快速处理故障。

1.5 论文的章节安排

本论文分六个部分来阐述电信企业客户忠诚度预测系统的设计实现过程，首先在分析现有电信企业现有模式的不足的基础上，阐述了推进信息化建设对电信企业的意义和重要性，以便找到课题的工作方向，最后根据软件工程思想分需求分析、系统设计实现以及测试等几部分来进行总结。

第一章：绪论。本章需重点分析电信企业客户忠诚度预测业务信息化建设的必要性，除此以外，还需明确本论文的组织结构。

第二章：相关技术说明。选择合适的技术组合有助于提升系统的开发效率，本章阐述了电信企业客户忠诚度预测系统选择J2EE + SQL Server2005 +B/S的技术组合的原因，并对上述技术进行了详细介绍。

第三章：需求分析。本章重点完成三项工作，明确功能需求，确定性能指标，并为系统构建清晰的权限体系。上述三项工作可确保系统具备良好的可用性和易用性。

第四章：系统设计。本章以时序图对系统各功能进行了设计，通过时序图可明确功能的工作流程以及各步骤中所使用的方法，可降低系统实现的复杂度。除此以外，还构建了客户信用风险评估模型。

第五章：系统实现与测试。本章阐述各模块功能的实现，并完成了系统的功能测试和性能测试。

第六章：总结与展望。本章首先总结了在推进客户管理业务信息化建设过程中所获得的经验，同时对系统存在的不足也进行了详细阐述，以便指导未来工作。

第2章 系统相关的理论与技术

2.1 数据挖掘技术

本节重点阐述数据挖掘技术的工作理念，同时阐述了使用贝叶斯技术构建客户忠诚度预测模块的原因。

2.1.1 数据挖掘的概念

数据挖掘的定义最早由Fayyad提出^[12]，其目的在于从海量的维修资料中找出新的规则。Berry & Linoff在1997年对数据挖掘进行了重新定义，其认为数据挖掘就是从海量数据中归纳出有价值的规则。具体而言，企业在日常运营中会积累大量的数据，事实上这些数据中隐藏着大量对企业运营有价值的规则，但是这些规则必须在综合分析的基础上才可获得，而数据挖掘技术就是为发现隐藏规则而提出的，其不仅可以利用被闲置的资料，同时也能为企业运营提供决策依据，是非常有意义的。

由上可知，数据挖掘不仅仅是算法，更强调其发现的过程，所以数据挖掘的定义具体可表述为数据挖掘就是从企业运营资料中抽取出有价值的属性，在对属性进行转换和清理的基础上，从中找出有价值的规则，以便提供给企业运用，用于提升企业竞争力和利润。

2.1.2 数据挖掘工作流程

数据挖掘的整个工作流程包含六个步骤^[13]，比如说数据选择、数据预处理、数据转换、数据挖掘以及模式评估等，其中数据挖掘是工作核心。整个数据挖掘的过程并非线性化的过程，中间包含许多回馈的过程，下面简单的介绍数据挖掘的工作流程，具体如图2.1所示：

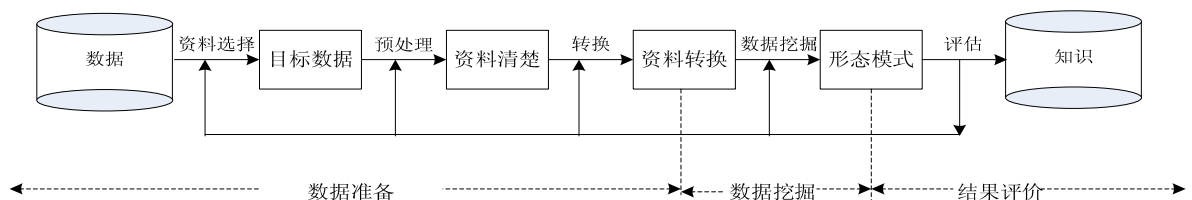


图 2.1 数据挖掘流程

下面对这六个步骤进行简单的介绍^[14]：

(1)数据选择：数据挖掘的第一步就是选择挖掘资料，将无价值的资料清除，留下对构建模型有意义的数据集。

(2)数据清除：选择的资料未经过处理，原始数据不一定能满足模型构建需求，需进行清除。面对不同的情况有不同的处理方法，如直接删除或者利用平均数据代替缺失数据等，这些都是处理数据缺失情况的有效措施。

(3)资料转换：若想从海量的数据中发现有价值的信息，就必须缩减数据量。在缩减数据量时，可根据数据挖掘的目的来采取相对应的方法，以便能被转换后的数据能被模型使用。

(4)选择数据挖掘算法：目前存在多种类型的数据挖掘算法，每种算法有不同的适用范围，需根据实际需要选择适用的算法，有助于提高模型准确率。

(5)数据挖掘：该步骤是整个流程的核心，前面所有步骤都是为数据挖掘服务，在该过程中需完成如下几项任务，第一、完成模型构建，第二、完成模型验证，在两者完成后，模型才能被应用到业务处理中。

(6)结果呈现：数据挖掘阶段构建了模型，在模型应用后，需将模型的效果展示给用户，以便给管理者提供决策依据。

2.1.3 算法选择

目前存在许多不同类型的数据挖掘算法^[15]，如遗传算法、贝叶斯网络算法以及决策树算法等，每种算法均有各自不同的适用领域以及各自的优缺点，本节在对比分析各种数据挖掘算法的基础上，决策使用贝叶斯网络技术作为客户分类模型的构建技术。贝叶斯网络又称为信度网络，是Bayes方法的扩展，是目前不确定知识表达和推理领域最有效的算法之一。贝叶斯网络理论结合了概率学、统计学以及图形学等诸多领域的学科成果，使得其在数据处理方面相比其他数据挖掘技术有更多的优势，具体体现在如下几个方面^[16]：

(1)贝叶斯网络用图形的方式表达节点之间的依赖关系，能清晰地表达这些节点之间的相互影响因子的大小，属于不确定性的因果关系模型，它与其他模型相比优势在于可以综合利用概率论以及相关的推理知识，易于理解，语义明了^[17]。

(2)贝叶斯网络善于处理不确定性问题，甚至对数据的完备性都不予要求。传统算法必须输入完备的数据，若缺少其中一项，则模型的预测结果就可能出现偏差，而贝叶斯网络则可以改变这种状况，它能在数据不完备以及问题不确定性的条件下进行学习和推力，不会因为某一项的缺失而出现较大的预测偏差^[18]。

(3)贝叶斯网络与概率统计学的结合能充分利用先验知识和样本数据的信息，避

避免因只利用先验知识而导致所设计的贝叶斯网络客户忠诚度预测模型不符合实际的情况，同时用条件概率分布来描述变量之间依赖关系的强弱，更接近对现实的模拟，因此利用贝叶斯网络技术构建客户分类模型效果更好。

2.2 开发技术

本节在对比分析不同技术实现方案的基础上，决策使用J2EE+SQL Server2005+B/S的技术组合，下面将对系统开发过程中使用的J2EE、MyEclipse平台进行介绍。

2.2.1 J2EE 架构

J2EE是构建客户忠诚度预测系统的核心，系统的核心业务由J2EE进行处理，具体而言，整个平台可由客户层、Web层、业务层以及EIS层等四部分组成，各部分在客户忠诚度预测系统中承担的任务存在区别。四层的关系如图2.2所示^[19]。

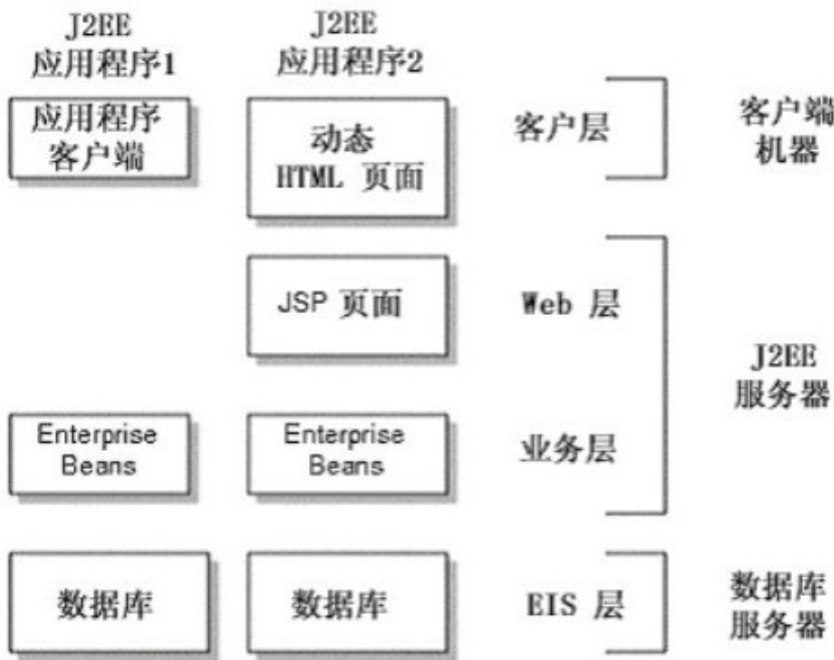


图 2.2 J2EE 的四层结构图

本节将对各层的作用进行详细介绍。

(1)客户层

客户层是针对C/S结构而言的，J2EE平台不仅支持开发B/S模式的信息系统，也支持开发C/S模式的系统。在开发不同模式的系统时，所使用的技术的是不相同的，C/S模式下的应用的技术为Swing组件，而B/S模式下应用的技术组合为HTML语言和JAVA代码。当Struts框架提出后^[20]，通常使用HTML语言和JSP标签的技术组合。

(2)Web层

Web层是针对B/S模式下的系统而言的，Web层所使用的技术经历了多种变化，最开始则使用HTML和JAVA，这就意味着web层需要处理系统业务^[21]，随着MVC模式的提出，Web层的作用越来越单一化，只承担展示功能，也就是HTML语言中不再需要嵌入JAVA代码，核心业务的处理由业务层承担，由此web层的负担则大大减少，同时页面开发者不再需要深入理解业务需求，可专心开发页面，有助于提升效率。

(3)业务层

业务层在J2EE平台中起着极其重要的作用^[22]，任意信息系统的核心任务都由业务层负责，Web层或者客户层只需将用户的请求或者提交的信息传输到业务层，业务层会根据请求进行处理，在处理完成后，会根据Web层或者客户层的要求进行封装，Web层或者客户层只需展示。业务层所使用的技术也在不断变化，从最开始的servlet技术、bean技术，到现在可引入各种框架技术，如Spring框架技术等，业务层的处理效率也在不断提升。

(4)EIS层

EIS层在信息系统的最后一个环节，通常指用于存储系统信息的数据库，在当前应用较多的数据库包括Oracle数据库、Mysql数据库以及SQL Server2005数据库等，J2EE平台支持上述任意数据库。

2.2.2 Myeclipse 平台

用于开发JAVA代码的工具有很多，应用较为广泛的包括NetBeans、JBuilder、eclipse以及MyEclipse等^[23]，本课题选择MyEclipse作为开发平台的原因在于其能很好的支持Web系统的开发，其他开发工具对web系统的开发支持不够，同时MyEclipse平台拥有许多的插件可降低客户忠诚度预测系统的开发难度。

MyEclipse的基础是eclipse平台，其是在eclipse平台上添加插件形成的，可应用于开发Java、Java EE以及移动应用程序，并且能很好的兼容各种开源工具。MyEclipse平台支持Java Servlet, AJAX, JSP, JSF以及SSH框架等各种技术，其是一个兼容度很高，并且易用性很高的JAVA开发工具，在目前JAVA开发中，应用比例最高。

MyEclipse平台是Sun公司为了推广JAVA技术而提出的开发工具^[24]，其支持开发基于B/S模式的信息系统，在该平台中，融合了EJB以及Tomcat等用于开发web系统的插件。MyEclipse平台自从提出以来，经过不断的升级，形成了多个不同的版本，截

至到目前，已经MyEclipse 2014 版本，该版本整合了Hibernate、Struts2和Spring等最新版本内容，解决了以前版本存在的缺陷，除此以外，MyEclipse平台直接安装了Tomcat插件，在开发web系统时，可直接进行调试，不需要单独安装Tomcat服务器来进行调试，能有效降低系统的开发难度，让整个开发过程更有效率。

2.3 数据存储平台的选择

在信息系统的开发过程中，最被常用的数据库无非就是Oracle数据库、MYSQL数据库以及SQL Server数据库，但是上述数据库有各自的优缺点和适用范围，必须按照实际需求进行选择^[25]，如Oracle数据库的特点是存储性能出色，可满足百万级的访问数据，缺陷在于对数据库服务器要求较低，并且应用成本较高，而Mysql数据库的优点是应用成本低、占用内存少，但是缺陷在于无法满足大中型信息系统的处理需求，SQL Server2005数据库的优点为占用内存少、应用成本较低，同时能满足大型信息系统的处理需求，但是性能与Oracle数据库相比相对较低，在综合考虑客户忠诚度预测系统的需求的基础上，决策选择SQL Server2005数据库作为存储平台。

与Oracle以及MYSQL等数据库相比，SQL Server 2005在很多方面都表现出色，下面进行详细分析，具体如下^[26]：

(1)SQL Server2005数据库占用的内存较少，但是性能表现出色，可满足大中型信息系统的处理需求。

(2)SQL Server2005数据库支持与当前主流的开发技术进行交互，如C++技术、.NET技术以及Java技术等

(3)SQL Server2005数据库兼容当前主流的操作系统，如Windows以及Linux等，可操作性相对其他数据库更强。

(4)SQL Server2005数据库支持应用到B/S和C/S模式下的信息系统，网络技术的发展让SQL Server2005数据库能更好的发挥相关性能。

(5) SQL Server2005数据库引入了数据挖掘以及数据仓库等全新的技术，从而有效扩大了数据库的应用范围。

(6)管理员可根据需求为信息系统设置不同权限的实例用户，这样有助于提高系统的数据安全性。

(7)SQL Server2005之前的数据库是无法存储图片数据的，这一缺陷在SQL Server2005中得到了解决。

(8)SQL Server2005数据库引入了全新的标记语言，最新的标记语言能降低SQL语句的编写难度，简单易用。

2.4 B/S 架构

B/S是Browser/Server的英文缩写，又称为浏览器和服务模式，与该模式相对的是C/S结构，即客户端与服务器模式。B/S模式和C/S模式存在较大区别，在B/S模式下，浏览器充当C/S模式中的客户端实现与服务器的交互，由此服务器将承担所有的压力，而在C/S模式下，服务端与客户端各自承担一部分系统任务，但是其需要每个PC都安装客户端，否则无法实现两者的交互。具体而言，B/S结构的示意图如图2.3所示。

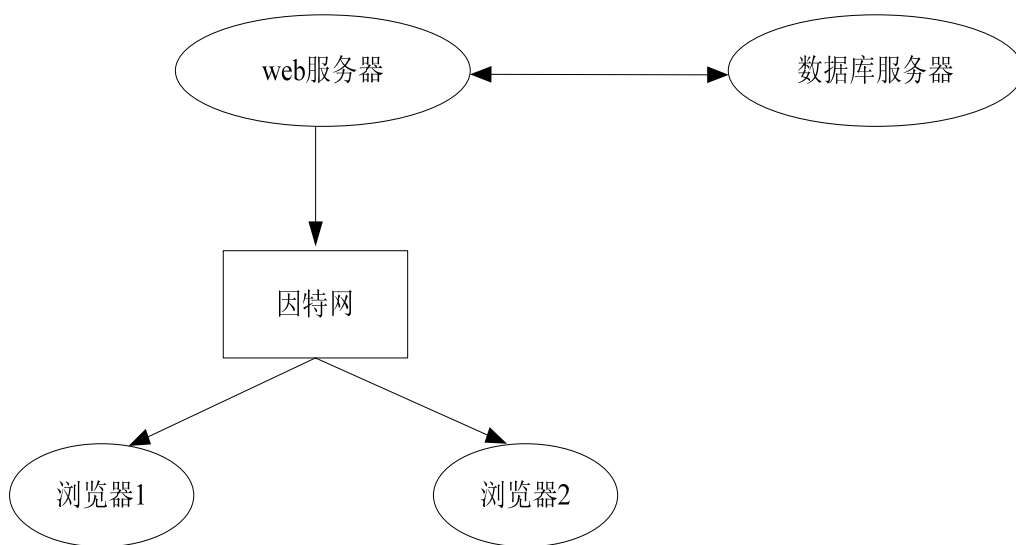


图 2.3 B/S 结构图

由图2.3可以知道，在B/S模式下，浏览器需要通过因特网来实现与服务器端的交互，web服务器支持并发访问，并且不需要安装客户端。对于系统用户而言，显然简单易用。B/S模式与C/S模式相比，拥有许多的优势，具体表现在如下几个方面：

(1)B/S模式适用范围广，简单易用，C/S下构建的系统通常只适合应用在局域网范围内，而B/S却不受到这种限制，只要求服务器和PC是可连的就可以实现访问。除此以外，B/S下只需要安装浏览器，而C/S下需要安装客户端，相对而言，B/S模式更简单易用。

(2)维护成本较低，C/S模式下，若需要对系统进行更新，则需要对每个客户端进行单独处理，维护成本较高，而B/S模式下，则只需要对服务器进行更新处理，维护成本相对较低。

正由于B/S模式拥有上述优势，并且从电信企业的实际情况出发，客户忠诚度预测系统选择基于B/S模式进行构建，以便满足系统的应用需求。

2.5 本章小结

本章对构建客户忠诚度预测系统的关键技术进行了详细阐述,首先对构建客户忠诚度预测模型的数据挖掘技术金项链详细阐述,其次对实现客户忠诚度预测系统的技术组合J2EE+SQL Server2005+B/S进行了详细介绍。

第3章 系统需求分析

3.1 需求概述

电信企业引入客户忠诚度预测系统的目的在于提升客户忠诚度和客户黏性,如此可防止客户流失,有助于提升电信企业的利润。在详细分析客户忠诚度预测业务流程的基础上,对实现客户忠诚度预测系统提出了新的要求,具体包括如下几个方面:

(1)系统需构建一个客户忠诚度预测模型,该模型可预测客户的忠诚度指数,从而可根据客户的忠诚度指数来分析客户是否易流失客户,若是则需要采取针对性营销策略,这样有助于提升客户忠诚度。

(2)系统需提供客户挽留功能,工作人员可以通过短信和邮件的方式向客户发送挽留信息,如节日祝福以及优惠信息等,这种措施有助于提升客户忠诚度。

(3)系统需实现权限控制,需约定不同用户的权限体系,否则权限混乱会给系统带来安全威胁。

(4)电信企业客户忠诚度预测系统中存储着客户信息以及客户忠诚度信息,若出现数据库故障导致数据丢失,则会给电信企业带来巨大的损失,为此系统需引入数据备份功能,定时备份系统数据,以便降低意外故障带来的损失。

(5)电信企业客户忠诚度预测系统要求易学易会、操作简捷,界面明了,方便系统各类用户使用。

3.2 功能需求分析

本节将对系统的功需求进行详细分析,其需囊括安全防护、客户资料管理、查询统计模块、用户管理以及忠诚度预测管理等功能,下面对各个部分进行详细阐述。

3.2.1 安全防护

安全防护与电信企业客户忠诚度预测系统的核心功能无关,但是与系统的安全息息相关,在系统中起着至关重要的作用。下面对安全防护模块进行详细阐述。

(1)身份识别

身份识别功能是系统的安全屏障,其会对用户进行身份认证,认证结果决定系统用户能否登录。只有在登录成功的基础上才会拥有相对应的操作权限,否则无法操作。

(2)注销

基于 B/S 模式构建的系统通常会引入注销功能，原因在于用户登录系统时，浏览器会记录用户信息，若用户未主动清除，就存在信息泄露的风险。

(3)密码修改

系统用户若长时间不更改密码，则密码容易泄露给其他用户，会给系统带来安全威胁。为了改变这种缺陷，系统强制要求用户每个 4 个月就更改一次密码。

(4)数据维护

数据维护是提升系统数据可靠性的有效措施，在备份时将数据保存在 SQL 文件中，当需要恢复时，可以将 SQL 文件中的数据重新读入到客户忠诚度预测系统中。

(5)日志记录

日志记录的是系统用户的详细操作行为，但并不是所有的行为都会记录日志，如查询操作。当系统中的数据出现错误时，可利用日志直接定位操作者。

(6)用户管理

系统包含管理员以及工作人员等两种类型的用户，上述四种类型的用户需要根据需求进行动态变化的，如工作人员离职与入职等，由此可知用户信息并不是固定的，需实现动态管理。

系统管理员拥有安全防护模块所有功能的操作权限，工作人员拥有身份识别、注销以及密码修改等功能的操作权限，据此可得到安全防护功能的用例图，具体如图 3.1 所示。

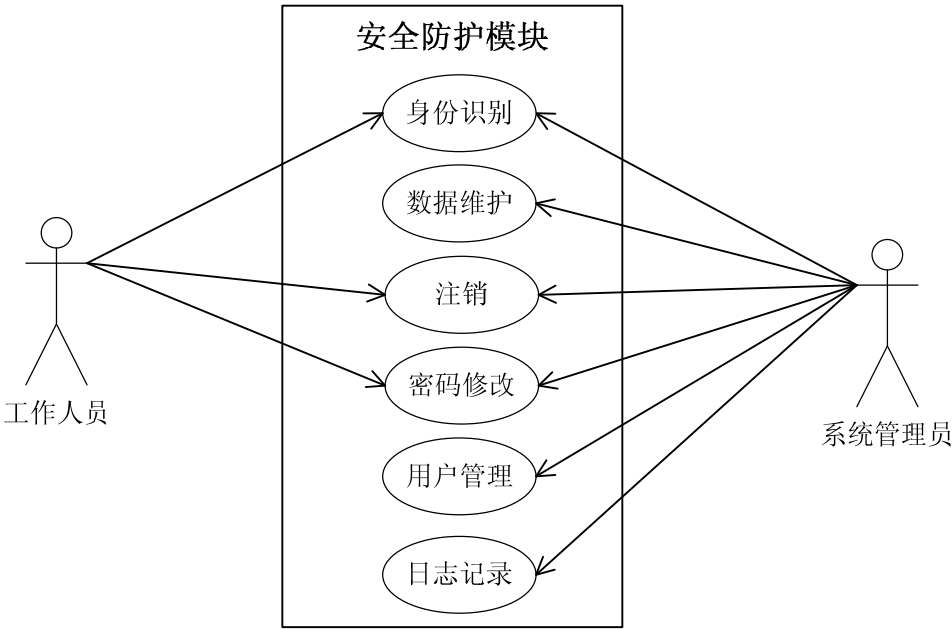


图 3.1 安全防护模块用例图

3.2.2 客户管理

客户是电信企业的核心资源，其可以为电信企业带来巨大的经济效益，上述资料也是客户忠诚度预测模型的基础。本节将阐述客户管理模块中各部分组成功能在系统中的作用。

(1)客户信息添加

客户信息主要指的是电信客户的开卡资料，即基本属性资料，如年龄、姓名、性别以及手机号码等，当有用户开卡时，则需要将相关资料添加到客户忠诚度预测系统的数据库中。

(2)客户信息删除

当工作人员由于误操作为客户忠诚度预测系统添加了不真实的客户信息时，可利用客户删除功能将客户信息从数据库中删除。删除成功后，工作人员则无法在查看已删除的客户信息。

(3)客户信息更新

若已录入的客户信息存在错误或者出现变化时，工作人员可将客户信息进行更新，防止错误的客户信息影响客户忠诚度预测系统的准确度。

(4)客户信息搜索

工作人员可根据客户关键字来精确定位客户忠诚度预测系统中存在的客户信息，如客户名称等，如此可节省工作人员时间，若从客户列表中逐一查找，则会浪费大量时间。

(5)客户信息导入

客户信息导入功能是将存储在 excel 中的客户信息导入到客户忠诚度预测系统数据库中，其能实现客户信息的批量添加，与客户信息添加功能相比，导入功能的工作效率更高。

(6)客户信息导出

客户信息导出功能是将符合条件的客户信息从数据库查询出来，并保存到 excel 文件中，以便工作人员进行查看。

工作人员拥有客户管理模块所有功能的操作权限，由此可得到客户管理模块的用例图，具体如图 3.2 所示。

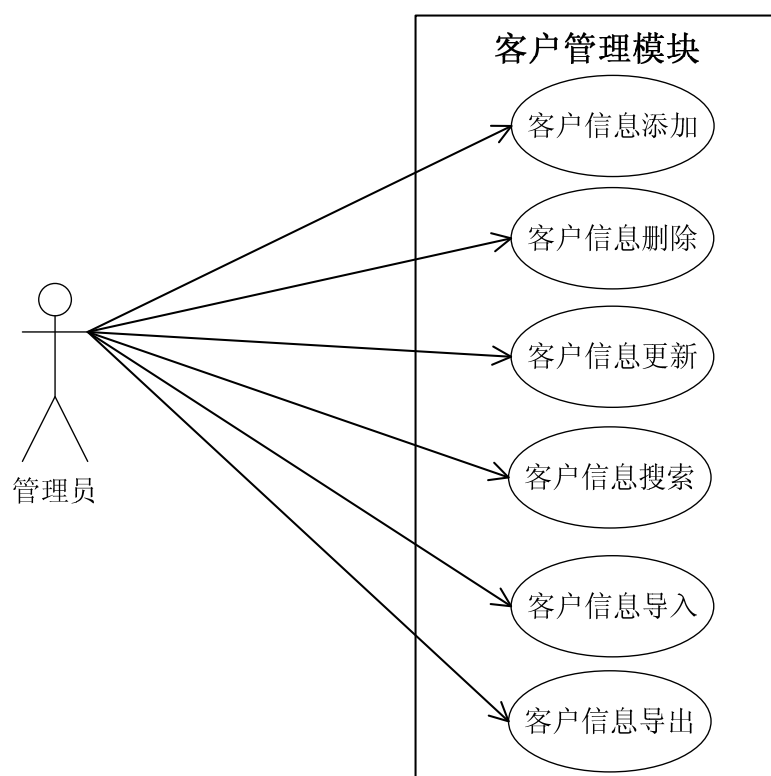


图 3.2 客户管理模块用例图

3.2.3 客户挽留

客户挽留功能的主要作用在于采取措施挽留低忠诚度客户，包括短信挽留和邮件挽留，其通过短信和邮件的方式将积分信息或者节假日祝福发送给客户，有助于挽留客户、吸引客户。下面对客户挽留模块的功能组成进行详细阐述。

(1)短信挽留

工作人员可通过短信的方式向电信客户发送积分信息以及节假日祝福，尤其是低忠诚度客户，通过上述方式有助于提升客户忠诚度以及企业利润。

(2)邮件挽留

工作人员可通过邮件的方式向电信客户发送积分信息以及节假日祝福，尤其是低忠诚度，通过上述方式有助于提升客户忠诚度以及企业利润。

(3)短信服务器配置

客户忠诚度预测系统使用短信猫技术进行短信发送，在进行短信发送之前，必须配置短信服务器，指定短信猫的波特率以及端口标识等，只有上述配置正确的情况下，才能进行短信发送。

(4)邮件服务器配置

客户忠诚度预测系统使用 J2EE 平台的 mail 插件进行邮件发送，在进行邮件发送

之前，必须配置邮件服务器，指定服务器用户名和密码等，只有上述配置正确的情况下，才能进行邮件发送。

工作人员拥有短信挽留以及邮件挽留等功能的操作权限，系统管理员拥有短信服务器配置和邮件服务器配置等功能的操作权限，根据上述分析可得到客户挽留、模块的用例图，具体如图 3.3 所示。

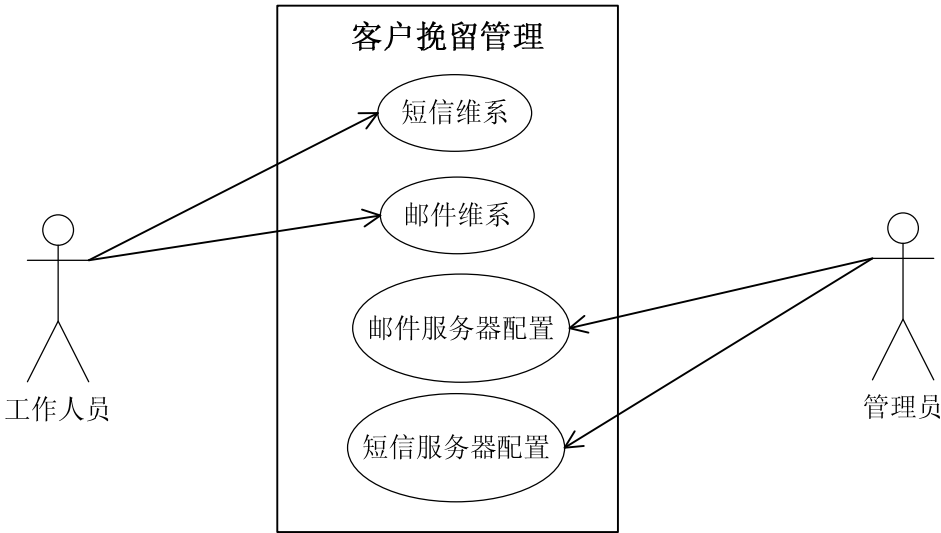


图 3.3 会员维系管理模块用例图

3.2.4 消费资料管理

消费行为资料与客户资料都是客户忠诚度预测系统的基础资料，模型需要从消费行为资料中抽取属性值用于预测忠诚度指数，下面对消费行为资料管理模块的功能组成进行详细阐述。

(1) 缴费信息管理

缴费信息管理功能主要用于管理电信客户的缴费信息，缴费记录在系统中存在两种来源，第一种为远程调用客户缴费系统提供的访问接口，第二种为通过 excel 文件将缴费记录数据导入到系统。

(2) 通话记录管理

通话记录管理功能主要用于管理电信客户的通话记录，通话记录在系统中存在两种来源，第一种为远程调用客户管理系统提供的访问接口，第二种为通过 excel 文件将通话记录数据导入到系统。

(3) 投诉行为管理

投诉行为管理功能主要用于管理电信客户的投诉信息，投诉信息在系统中存在两种来源，第一种为远程调用客户管理系统提供的访问接口，第二种为通过 excel 文件

将投诉信息导入到系统。

工作人员拥有缴费信息管理、通话记录管理以及投诉行为管理等功能的操作权限，由此可得到工作人员的用例图，具体如图 3.4 所示。

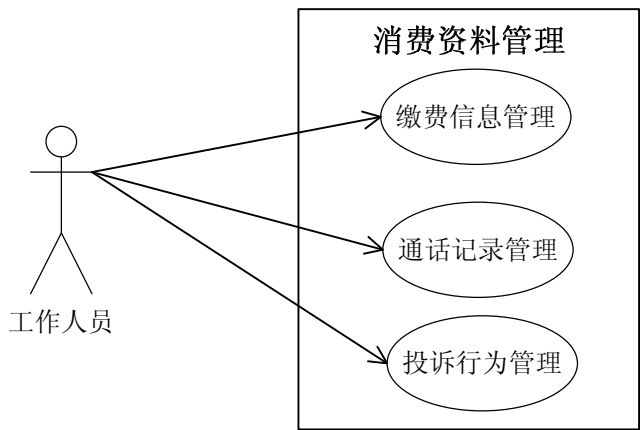


图 3.4 消费资料管理模块用例图

3.2.5 忠诚度预测管理

忠诚度预测管理模块是系统的核心，其主要预测客户的忠诚度指数，并为电信企业提供低忠诚度客户名单以便采取挽留措施，本节将阐述忠诚度预测管理模块中各部分组成功能在系统中的作用。

(1)客户忠诚度预测

客户忠诚度预测功能是根据客户信息的已知值(对应着网络模型中节点的状态)，用概率推理的方式推导出客户信息所反应出来的忠诚度指数。若客户的忠诚度指数低于 20%., 则定义低度忠诚客户，若高于 20%但低于 50%。则定义为中度忠诚客户、指数，若高于 50%，则定义为高度忠诚客户。

(2)忠诚度查询

工作人员可利用客户名称等关键字来查询电信客户的忠诚度指数，其是系统易用性的体现。

(3)名单导出

工作人员可以将低忠诚度的名单导出到 excel 文件中，以便查看或打印，从而可采取措施来维系客户。

工作人员拥有忠诚度预测管理模块所有功能的操作权限，由此可得到工作人员的用例图，具体如图 3.5 所示。

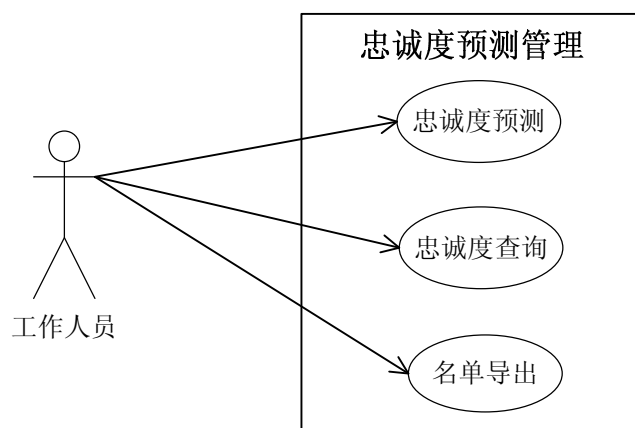


图 3.5 忠诚度预测管理模块用例图

3.3 用例分析

本节将以用例图的方式来为系统构建权限体系，主要作用在于明确各用户类型的权限组成。

(1) 管理员

管理员主要用于保证系统的稳定运行，其拥有身份识别、注销、密码修改、数据维护、日志记录、用户管理、短信服务器配置以及邮件服务器配置等功能的权限，由此可得到管理员的用例图，具体如图 3.6 所示。

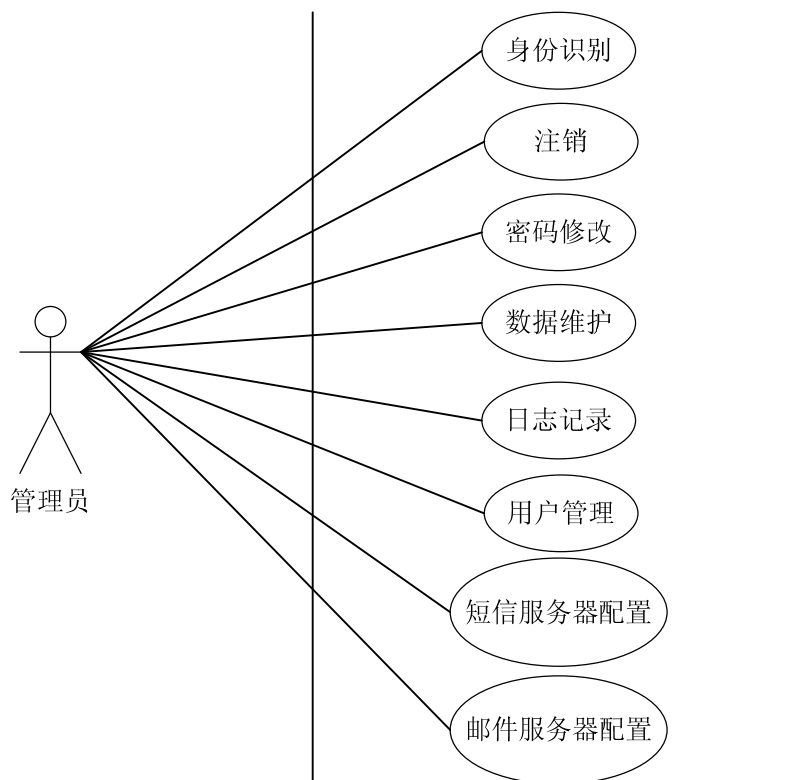


图 3.6 管理员用例图

(2) 工作人员

工作人员是系统的核心用户，其拥有忠诚度预测系统大部分功能的权限，本节将以用例图的方式来为系统构建权限体系，具体如图 3.7 所示。

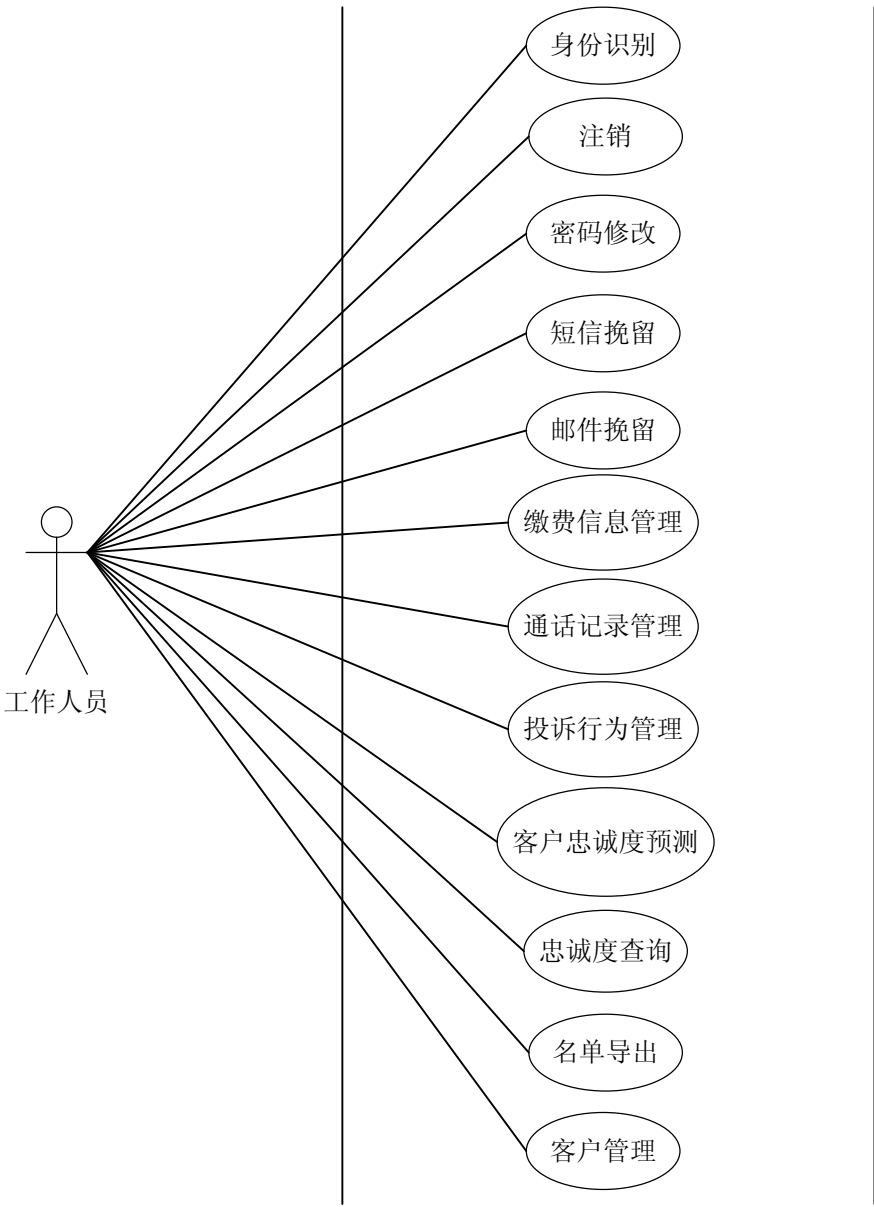


图 3.7 工作人员用例图

3.5 非功能性需求分析

非功能性需求分析在需求分析过程中占据极其重要的位置，若构建完成的系统在部署到实际环境后，性能不达标，则容易出现系统崩溃等现象，尤其是并发用户指标以及响应速度等几方面。本节将对电信企业客户忠诚度预测系统的非功能性需求进行详细分析。

(1)并发用户数

在构建系统时，需重点考虑并发用户指标，若实际应用为 500，构建的系统的系统只支持 400 人，那么当 500 人同时上线时，则可能发送服务器崩溃或者响应速度变慢等现象，这些都是系统用户无法接受的。从该电信公司的实际情况出发，拥有系统操作权限的工作人员 400 人，考虑到电信业务处于不断增长，防止人员增加时，系统无法满足需求，本课题构建的系统支持 500 人并发访问。

(2)响应速度

响应速度是系统的重要指标，若用户提交操作请求，需要长时间才能处理请求，展示处理结果，则会降低用户使用系统的兴趣。本课题设定系统响应速度需低于 3 秒，在并发用户数达到 500 人时，系统的响应速度需低于 5 秒，当然这不包括客户信息导入等功能，该项功能的响应速度与导入的数量呈正比关系。

(3)数据一致性

在操作数据库的过程中，需引入事务保护机制，防止数据插入过程中发生故障造成数据的不一致，从而查询出错误的设备信息。

(4)高效性

由于电信公司的人员基数较大，其业务量较多，需要经常操作数据库。为了提高访问系统，系统引入了 JDBC 等数据库连接技术，同时使用了存储过程及优化查询语言等方法，使得访问数据库的性能得到了很大的提高，从而减轻服务器的负担，提高了系统的整体性能。

3.6 本章小结

本章首先对系统的业务需求进行了描述，然后从功能需求和非功能性需求等两方面对电信企业客户忠诚度预测系统的需求进行了分析，上述需求分析为系统的构建奠定了基础。

第4章 系统详细设计

4.1 设计原则

在构建系统之前，必须遵循一定的原则，一份完整、详细的设计方案能大大降低系统开发的复杂度，否则开发人员在构建系统前需要耗费更多的时间和精力。本节将从先进性、实用性、可用性以及安全性等对电信企业客户忠诚度预测系统的设计原则进行详细阐述。

(1) 实用性

在设计系统时，第一要素就是实用，其需符合电信公司的客户忠诚度预测需求，否则按照设计方案构建的系统与实际业务不相符，也就无法部署到电信公司用于处理相关业务。

(2) 先进性

构建的系统必须能部署在电信公司现有的网络环境中，若有特殊需求，也要尽可能的降低部署成本。当需要进行扩容时，也要兼容现有的软硬件设备，否则会加大应用成本。

(3) 安全性

电信企业客户忠诚度预测系统中存储着电信公司的关键数据，若这些数据被篡改或丢失，则可能引发安全问题，给电信公司带来损失，这也是信息化建设时必须考虑的问题，因此安全问题不容忽视，可引入权限管理、数据维护以及身份校验等多种方式。

(4) 稳定性

电信公司工作人员时刻需要访问电信企业客户忠诚度预测系统，因此系统必须保持稳定，长时间为电信公司服务，若系统在工作时间崩溃，则会降低工作效率，与推进信息化的目的正好相反，这是用户无法接受的。

(5) 易用性

电信企业客户忠诚度预测系统包含许多的页面，所有页面都必须保持风格统一，同时系统操作提示都需保持友好。

4.2 系统概要设计

本节将对电信企业客户忠诚度预测系统进行概要设计，以便了解其技术架构、功能架构以及部署场景。

4.2.1 技术架构设计

电信企业客户忠诚度预测系统基于 J2EE 平台进行构建，因此其架构层次与 J2EE 平台保持一致，划分为四层，即客户层、页面层、业务层以及 EIS 层，各层在电信企业客户忠诚度预测系统中承担着不同的任务，其技术架构如图 4.1 所示。

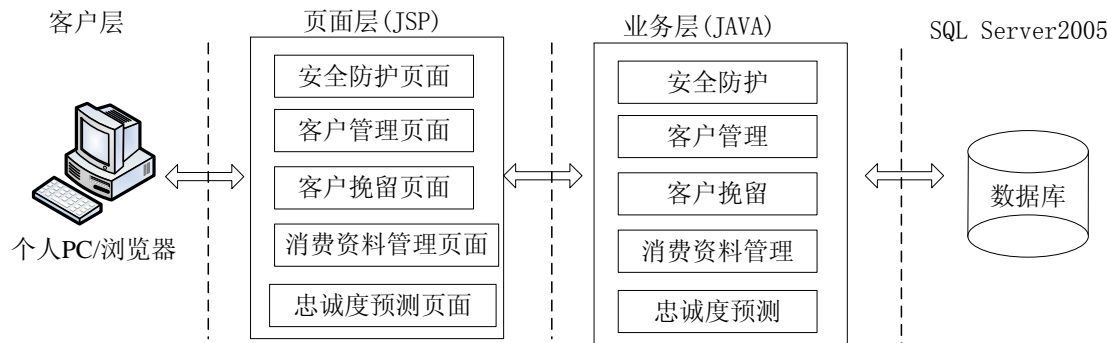


图 4.1 技术架构图

(1)客户层：电信企业客户忠诚度预测系统是基于 B/S 模式进行构建，在 B/S 系统中，浏览器承担着 C/S 架构中客户层的作用。电信企业客户忠诚度预测系统支持当前主流的浏览器，如 IE、火狐以及 360 等，用户在有网络的前提下，可通过浏览器完成与电信企业客户忠诚度预测系统的交互。

(2)页面层：页面主要接受来自客户层的请求，并转发到业务层。最早的开发模式中该层使用 HTML+JAVA 的技术组合，即页面层也承担业务处理，本系统在该层引入了 Struts 框架，从而不需要在 HTML 中嵌入 JAVA 代码，逻辑相对更加简单，后续维护难度也会降低。

(3)业务层：该层承担着电信企业客户忠诚度预测系统的核心任务，即处理各项业务，包括安全防护、客户管理、客户挽留、消费行为资料管理以及忠诚度预测管理等，其主要使用 JAVA 代码来实现业务处理。

(4) EIS 层：该层存储着电信企业客户忠诚度预测系统涉及的数据信息。本系统在对分析各类数据库的基础上，决策使用 SQL Server2005 来存储相关数据，应用成本较低，操作更简单。

4.2.2 拓扑结构设计

电信企业客户忠诚度预测系统基于 B/S 模式进行构建，支持系统用户在能连接到电信企业客户忠诚度预测系统 web 服务器的条件下处理相关业务，根据实际应用情况，电信企业工作人员可在内网和外网同时访问电信企业客户忠诚度预测系统，但是在外网访问时，由于存在不确定性容易给内网带来网络威胁，为了保证系统的安全，防止被网络攻击，在外网的入口设置了防火墙，防火墙的不是可大大提高网络安全。而内网为局域网，安全威胁相对较低，公司内部网络为总线型，在连入到服务器时，通过路由器，在路由器上也可以设置简单的安全防护措施，可满足内部网络的防护需求。根据上述分析可得到电信企业客户忠诚度预测系统的网络拓扑结构图，具体如图 4.2 所示。

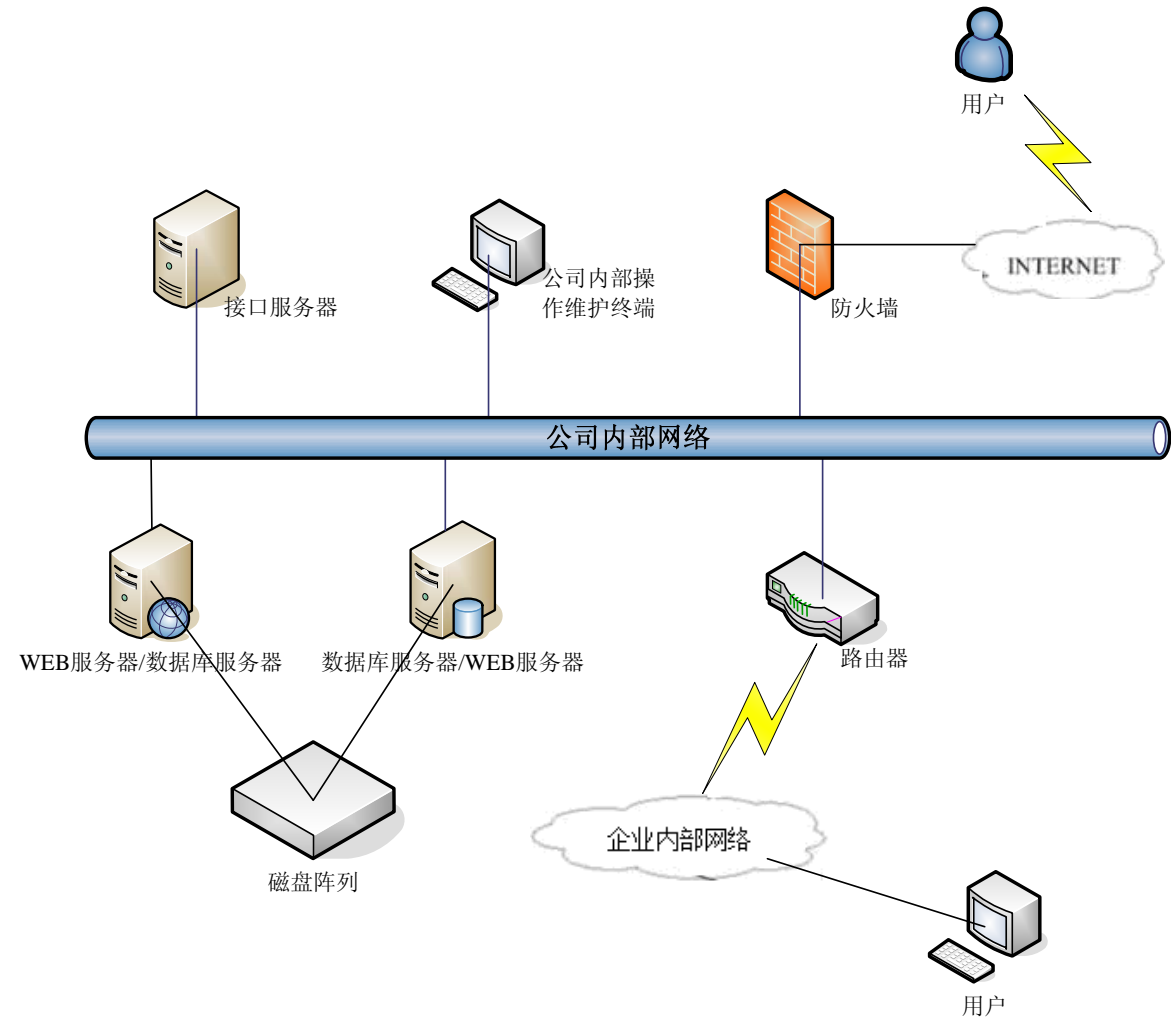


图 4.2 拓扑结构图

4.2.3 逻辑结构设计

在构建系统前，必须将系统功能模块化，并根据模块来进行组织代码结构，否则系统开发会混乱。从模块化的角度，电信企业客户忠诚度预测系统可划分为安全防护、客户管理、客户挽留、消费行为资料管理以及忠诚度预测管理等五个部分。下面对系统模块进行详细阐述。

电信企业客户忠诚度预测系统需提供如下基本功能：

(1)安全管理

安全管理的主要作用在于保证电信企业客户忠诚度预测系统的安全和稳定，可从多角度维护系统的安全，其由身份识别、注销、密码修改、数据维护、日志记录以及用户管理等功能组成。

(2)客户管理

客户是电信企业客户忠诚度预测系统的核心，是电信企业经济效益的来源，也是客户忠诚度预测模型进行预测前提，其中的某些指数需要被使用，客户信息越完整，模型的预测就越正确，同时由于电信客户是不断调整的，必须实现客户信息的动态管理。

(3)客户挽留

客户挽留功能的主要作用在于通过短信或邮件的方式挽留客户，其由短信挽留、邮件挽留、短信服务器配置以及邮件服务器配置等功能者组成。

(4)消费行为资料

消费行为资料是客户忠诚度预测模型进行预测的前提，从消费行为资料抽取的属性是模型的核心因素，其由缴费信息管理、通话记录管理以及投诉行为管理等功能组成。

(5)忠诚度预测管理

忠诚度预测管理是系统的核心，其会综合客户的资料来预测客户的忠诚度，并为电信企业提供低忠诚度的客户名单，以便采取措施进行挽留。

由此可得到电信企业客户忠诚度预测系统的功能架构图，如图 4.3 所示。



图 4.3 功能结构图

4.3 功能模块详细设计

本节将以顺序图来描述电信客户忠诚度预测系统的各个模块功能的工作流程，以便了解各功能的实现细节。

4.3.1 安全防护模块

安全防护模块主要从代码角度维护系统安全，保障系统数据不被篡改，本节将以顺序图的方式对登录功能以及数据管理中数据恢复功能进行介绍。

(1)登录

登录是系统安全的第一步，其属于系统的门槛，不提供合法的凭证无法进行业务处理，其工作流程包括如下几个方面：

①用户进入电信客户忠诚度预测系统的登录页面，即 login.jsp 页面。

②在 login.jsp 页面提交包含 userName 以及 psw 的登录凭证。在 login.jsp 会调用 check 方法进行格式校验。

③将 userName 以及 psw 传递到 UserManager 类，其会调用 UserDao 类中的 login 方法。login 方法会以 userName 和 psw 为条件进行查询，并返回结果。

根据上述分析可得到登录功能的顺序图，具体如图 4.4 所示。

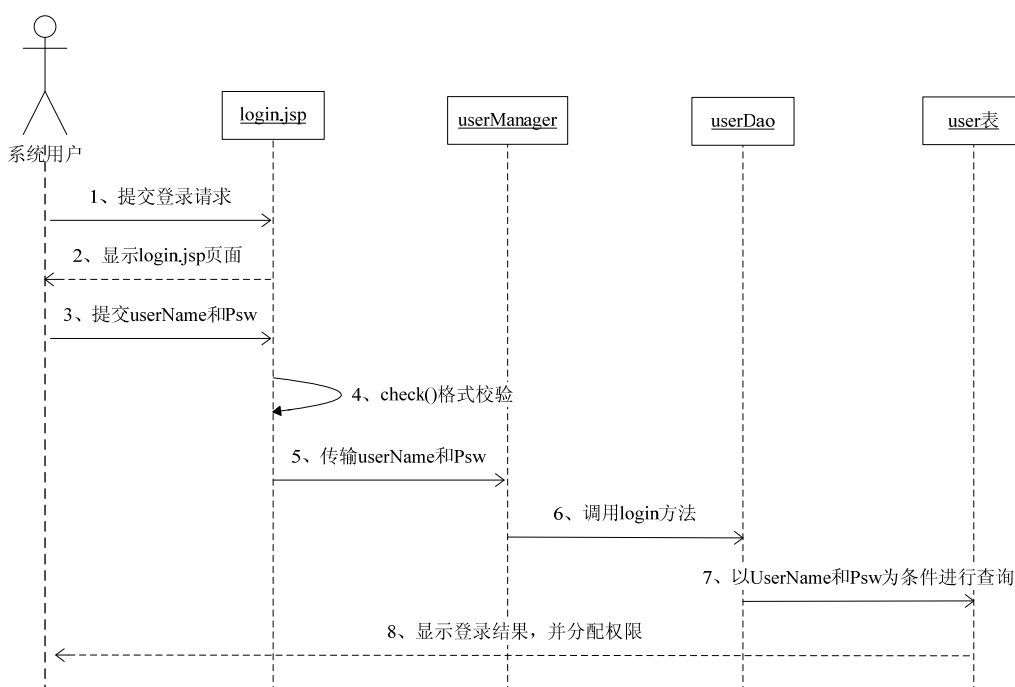


图 4.4 登录功能顺序图

(2)数据恢复

数据恢复功能是提升数据可靠性的关键措施，当出现数据丢失时，管理员可利用备份文件和数据库日志来恢复全部数据，其工作流程包括如下几个方面：

①用户进入系统的数据恢复页面，即 dataRestore.jsp 页面。

②在 dataRestore.jsp 页面中选择备份文件的地址，并确定恢复。

③dataRestore.jsp 页面会调用 fileIsNull 方法进行空校验，然后调用 typeCheck 方法校验备份文件的类型。

④将文件地址对象传递到 DataManager 类，其会调用 DataDao 类中的 dataRestore

方法将备份文件中的数据保存到数据库中。

根据上述分析可得到安全防护模块中的数据恢复功能的顺序图，具体如图 4.5 所示。

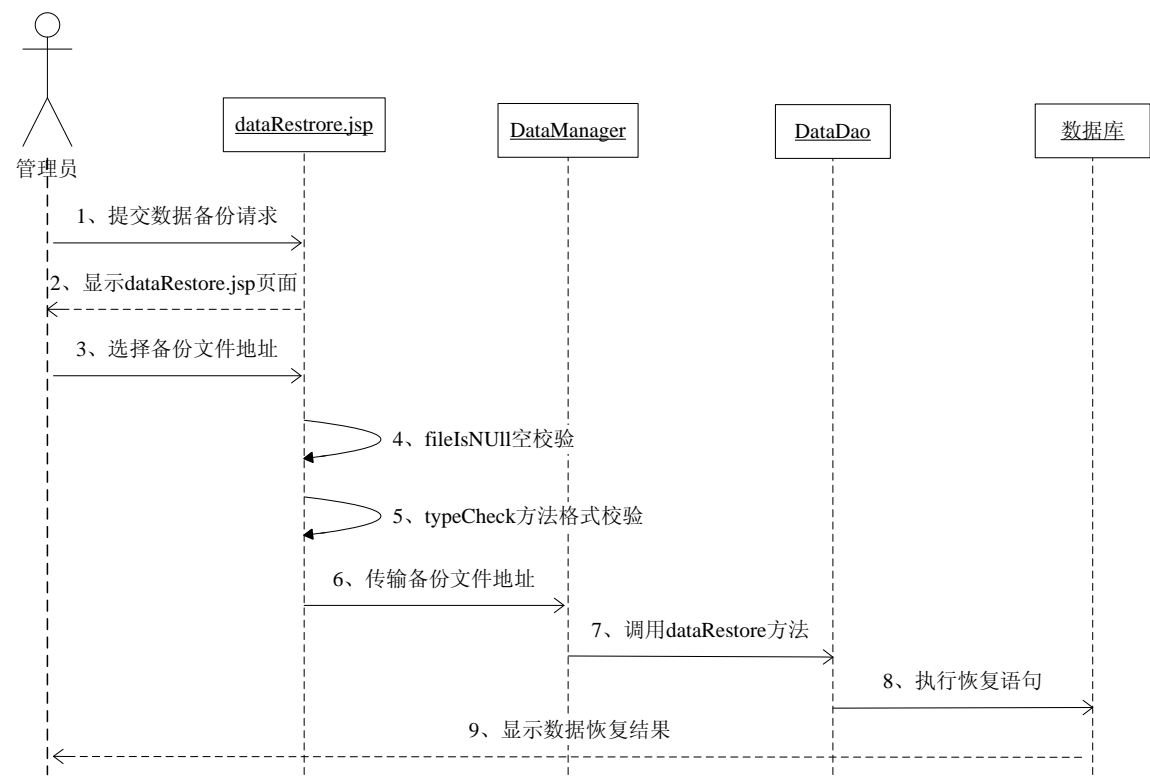


图 4.5 数据恢复功能顺序图

4.3.2 客户管理模块

客户资料是系统的核心资源，为电信企业创造利润，其基本属性是模型节点集的一部分，本节将以顺序图的方式对客户信息导入功能以及客户信息导出功能进行介绍。

(1) 客户信息导入

客户信息导入功能就是将保存在 excel 文件中的客户信息添加到数据库中，其工作流程包括以下几个部分：

- ①工作人员进入系统的客户信息导入页面，即 customerImport.jsp 页面。
- ②在 customerImport.jsp 页面中选择存储客户信息的文件的地址，并导入。
- ③customerImport.jsp 页面会调用 typeCheck 方法校验文件类型。
- ④将文件地址对象传递到 customerManager 类，从文件中读取客户信息，并以 list 进行存

储。然后调用 `customerSave` 方法将 `list` 中的客户信息存储到数据库中。

根据上述分析可得到客户管理模块中的客户信息导入功能的顺序图，具体如图 4.6 所示。

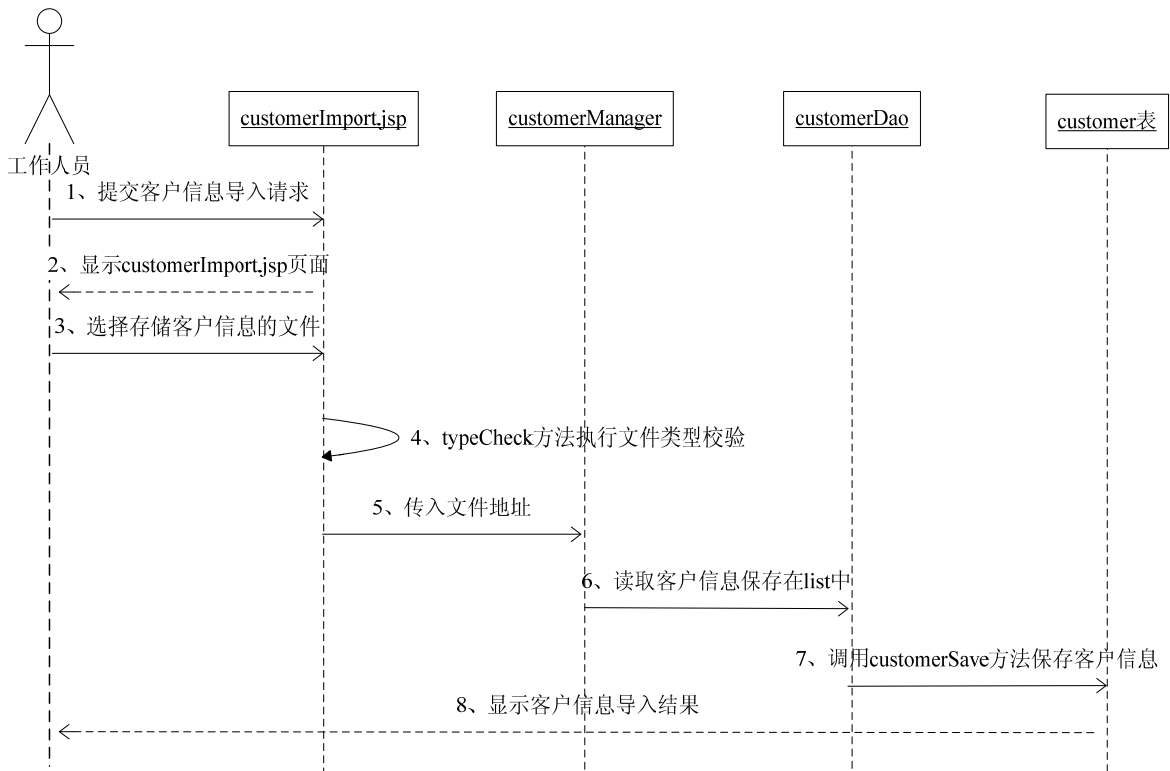


图 4.6 客户信息导入功能顺序图

(2)客户信息导出

客户信息导出功能就是将符合条件到数据导出到 `excel` 文件中，以便查看或打印，其工作流程包括如下几个部分：

- ①工作人员进入系统的客户信息导出页面，即 `customerExport.jsp` 页面。
- ②在 `customerExport.jsp` 页面中，输入客户信息关键字，然后执行 `searchCustomerInfo` 方法。
- ③系统会将客户查询结果展示给工作人员，并以 `list` 存储查询结果，然后点导出，调用 `customerImport` 方法将 `list` 中数据写入到 `excel` 文件中。

根据上述分析可得到客户管理模块中的客户信息导出功能的顺序图，具体如图 4.7 所示。

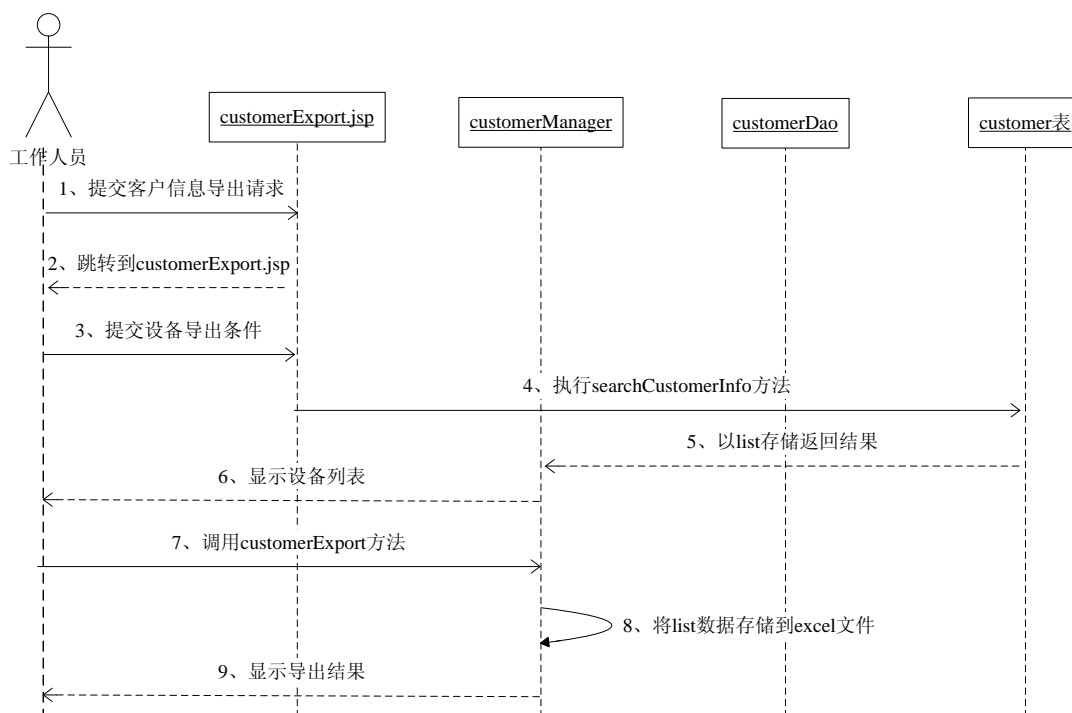


图 4.7 客户信息导出功能顺序图

4.3.3 客户挽留模块

客户挽留模块的主要作用在于通过短信或邮件的方式来挽留低忠诚度客户，其中邮件服务器配置和短信服务器配置是客户挽留模块进行工作的前提。在短信服务器未配置的情况下，短信是无法发送的。本节将以顺序图的方式对短信服务器配置功能进行介绍。

短信服务器配置功能的工作流程包括以下几个部分：

- (1)管理员进入短信服务器配置页面，即 smsServer.jsp 页面。
- (2)管理员在 smsServer.jsp 页面提交短信服务器参数，并提交保存。
- (3)系统调用 check 方法来校验短信服务器参数的正确性，在短信服务器参数正确的情况下，调用 test 方法测试短信服务器参数是否能发送短信。
- (4)若现有参数能发送短信，则将调用 serverSave 方法将短信服务器参数保存到数据库表中。

根据上述分析可得到客户挽留模块中的短信服务器配置功能的顺序图，具体如图 4.8 所示。

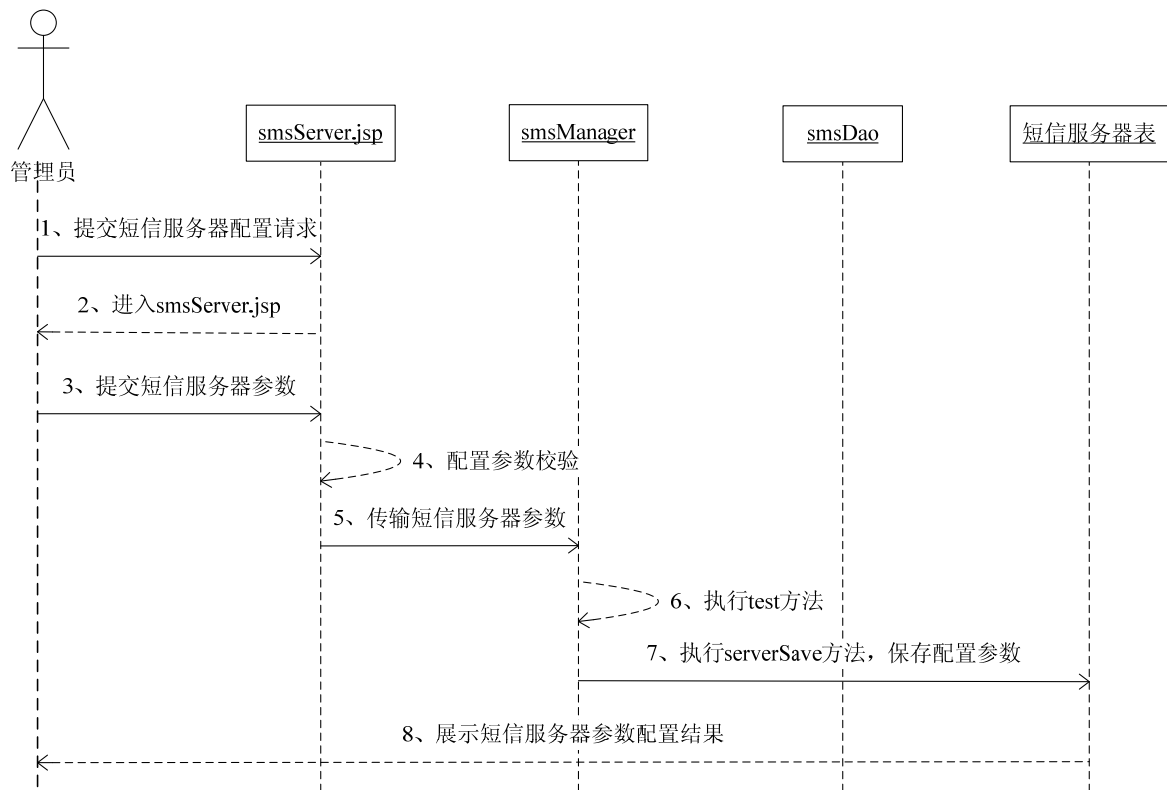


图 4.8 短信服务器参数配置功能顺序图

4.3.4 消费资料管理模块

消费资料管理模块主要管理客户的缴费信息、通话记录以及投诉信息，其包括各项资料信息的添加、删除、导入以及导出等功能。本节将以顺序图的方式对通话记录删除功能进行介绍。

通话记录删除功能的作用在于将通话记录从电信客户忠诚度预测系统数据库中进行清除，防止错误的数据保存在电信客户忠诚度预测系统中。通话记录删除功能的工作流程如下：

(1)工作人员进入通话记录管理页面，系统会自动执行 `getAllRecord()`方法，该方法主要用于获得通话记录列表。

(2)在通话记录列表中，选择待删除的通话记录，将通话记录组装成一个 `list`，不妨以 `recordList` 表示，将 `recordList` 传入到数据层。

(3)在后台针对 `recordList` 中每个 `id` 执行删除 SQL 语句，即 `delete from tbl_recordInfo where recordID = recordList.get(i)`。

(4)系统会自动执行上述语句，并返回执行结果，提示各个通话记录的删除结果，以便出现删除失败等意外情况时，用于定位原因。

根据上述分析可得到通话记录删除功能的时序图，具体如图 4.9 所示。

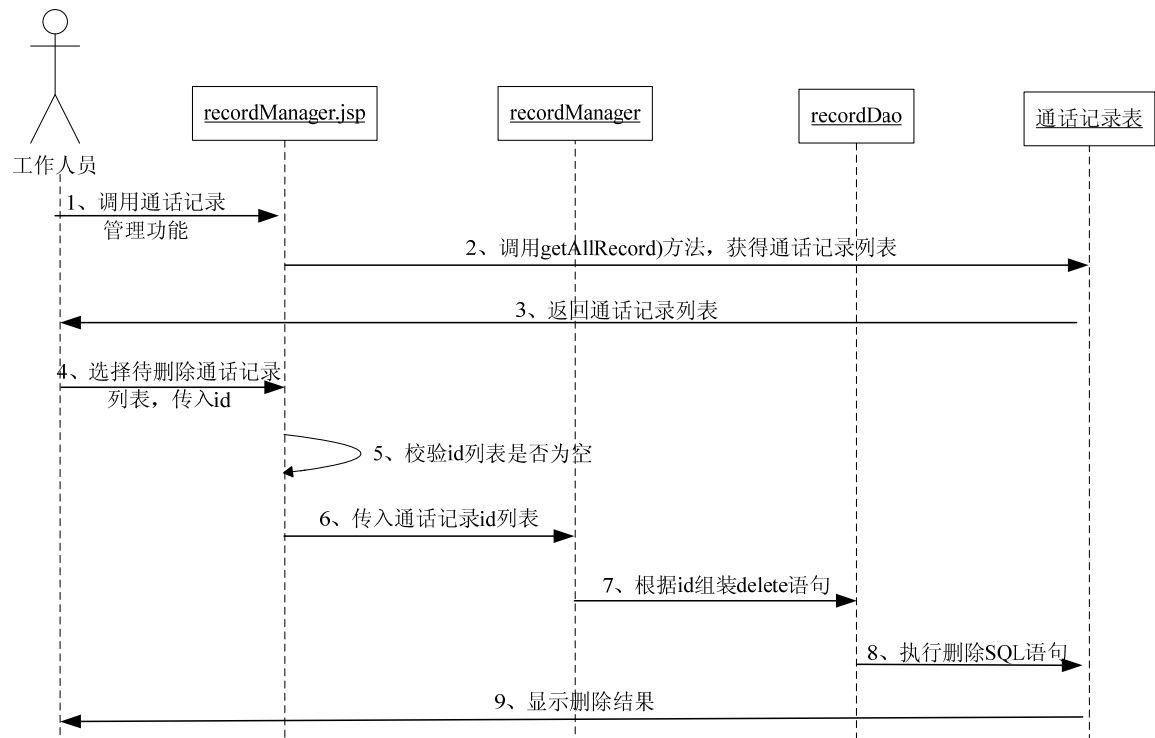


图 4.9 通话记录删除时序图

4.3.5 忠诚度预测模块

忠诚度预测模式处理着系统的核心业务，其需要根据客户的基本资料来获得客户的忠诚度指数，同时工作人员也可查询客户的忠诚度指数。本节将以顺序图的方式对忠诚度查询功能进行介绍。

忠诚度查询功能的主要作用在于快速定位系统中客户的忠诚度指数，以便查看，其工作流程包括如下几个部分：

(1)工作人员进入忠诚度预测页面，即 `loyaltyManager.jsp`。

(2)在 `loyaltyManager.jsp` 页面，存在忠诚度查询功能，工作人员需提交客户忠诚度查询关键字。

(3)系统会调用 `searchLoyalty` 方法查询系统中客户的忠诚度指数。系统会将查询到的结果以列表形式展示给工作人员。

根据上述分析可得到忠诚度查询功能的顺序图，具体如图 4.10 所示。

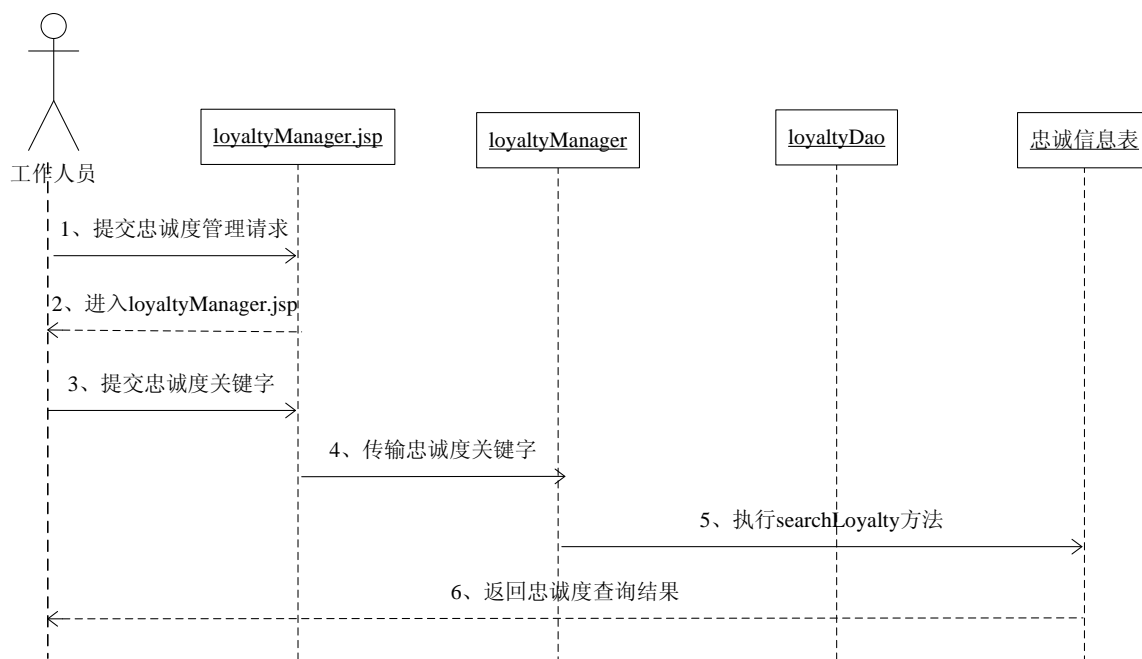


图 4.10 忠诚度查询功能顺序图

4.4 忠诚度预测模型构建

在电信客户忠诚度预测系统中，有一项重要功能为忠诚度预测，预测功能的主要目的在于获得客户的忠诚度指数，以便根据忠诚度指数给出低忠诚度名单。本节将对客户忠诚度预测模型的构建过程进行详细阐述。

4.4.1 构建流程

客户忠诚度预测模型是贝叶斯网络在实际生活中的具体应用，因此客户忠诚度预测模型的建立流程跟贝叶斯网络的主要工作流程是一致的，一般都包括如下五个阶段^[27]：

(1)商业理解

本课题的主要目标是利用数据挖掘技术来预测电信客户的忠诚度指数，以此得到低忠诚度客户名单。

(2)数据理解

熟悉当前可利用的数据，包括通话记录、缴费数据以及投诉数据等。

(3)数据准备

从现有数据中获得构建基于贝叶斯网络的客户忠诚度预测模型的节点集。

(4)建立模型

利用节点集确定各节点的关系，并以此确定客户忠诚度预测模型的网络拓扑结

构，除此以外还需要明确其条件概率分布表。

(5)模型评估

待模型确定后，则需要对模型进行评估，即检验模型的准确率。

本节将从上述五个部分来阐述客户忠诚度预测模型的构建过程。

4.4.2 商业理解

本课题的主要目的在于构建客户忠诚度预测模型，对于电信企业而言，客户忠诚度的定义为客户继续使用电信公司服务的意愿或者客户转换服务商的倾向程度。客户忠诚度存在许多的影响因子，根据因子来源存在两种类型，第一部分来自企业，电信企业的服务质量与客户忠诚度息息相关；第二部分来自客户本身，如工作需要或者卡损坏等。显然客户原因是不可控，无法改变的。电信企业所能做的就是通过改善服务质量提升客户忠诚度。通过上述分析可知道电信行业客户忠诚度所代表的含义，由此当获得低忠诚度客户名单时，可通过积分计划或者优惠计划等方式来挽留客户。

上述分析完整的阐述了电信行业构建客户忠诚度预测模型的商业目，客户忠诚度预测模型就是获得电信客户的忠诚度指数，从而将其中低忠诚度客户名单提供给电信工作人员，以便采取措施进行挽留。

4.4.3 数据理解

在利用数据开展数据挖掘工作之前，最重要的工作就是分析电信企业所能掌握的客户资料，从而能知道每类资料所存储的数据以及对企业的意义，在理解数据的基础上，才能分析资料、抽取影响因子。

当前电信企业所能掌握的资料包括如下几部分：

(1)客户信息表：主要代表客户的注册资料，其包含的信息如表 4.1 所示。

表 4.1 客户信息基本属性

序号	字段名	字段意义	字段类型
1	user_id	用户编号	int
2	tel_number	电话号码	varchar
3	user_name	用户名	varchar
4	certification_type	证件类型	varchar
5	certification_number	证件号码	varchar
6	user_flag	客户状态	int
7	city	归属市县	varchar

序号	字段名	字段意义	字段类型
8	birthday	出生日期	datetime
9	age	年龄	int
10	job	职业	varchar
11	user_type	客户类型	int
12	count_scheme	计费方案	varchar
13	number	联系人数量	int
14	update	更新时间	datetime

(2)通信记录表：主要用于存储客户的通话记录，其包含的信息如表 4.2 所示。

表 4.2 通话记录基本属性

序号	字段名	字段意义	字段类型
1	user_id	用户编号	int
2	billing_data_number	帐期编码（区号）	int
3	a_area_code	主叫区域	varchar
4	a_tel_code	主叫号码	varchar
5	p_area_code	被叫区域（区号）	varchar
6	p_tel_code	被叫号码	varchar
7	begin_time	开始时间	datetime
8	end time	结束时间	datetime
9	duration	时长	int
10	acct_item_type_id	账目类型编码	int
11	call_fare	费用	money
12	area_id	区域标识	int

(3)消费行为表：消费行为表由通信记录表进行统计而得，其包含的信息如表 4.3 所示。

表 3.3 客户通话行为基本属性

序号	字段名	字段意义	字段类型
1	local_fee	本地话费	float
2	network_fee	上网通信费	float
3	msg_fee	短信费用	float
4	long_fee	长途费用	float
5	favor_fee	优惠费用	float
6	owed_fee	欠费费用	float
7	month_rent	月租	float
8	other_fee	其他业务费用	float
9	internet_fee	网间费用	float
10	many_fee	漫游费用	float

(4)投诉信息表：其主要用于记录客户的投诉信息，其包含的信息如表 4.4 所示。

表 4.4 客户投诉行为基本属性

序号	字段名	字段意义	字段类型
1	user_id	投诉用户	int
2	user_number	投诉号码	varchar
3	complaint_reason	投诉原因	varchar
4	complaint_time	投诉时间	datetime
5	remark	备注	varchar

根据上述的一些客户数据，可以推导出一些的数据变量，主要描述客户使用特定消费的比例，当前使用较多的衍生指标主要如下：

- (1)本地话费比例 = 本地话费/总消费额；
- (2)长话比例 = 长途话费/总消费额；
- (3)漫游比例 = 漫游话费/总消费额；
- (4)短信比例 = 短信费用/总消费额；
- (5)网间比例 = 网间费用/总消费额；
- (6)投诉次数：对同一投诉用户的记录条数求和

4.4.4 数据准备

本节在详细分析现有资料，选择了进入客户忠诚度预测模型中的节点集合，本节将对节点进行详细介绍。

(1)欠费状态：欠费状态在一定程度上直接预示着客户低忠诚度，一个用户若欠费超过一个月以上，那么该用户的离网倾向就比较高，因此需高度注意客户的欠费状态。

(2)投诉次数：若一个用户对某企业投诉非常频繁，则应引起注意，因为对于企业服务质量的不满会导致客户的直接流失，因此当客户离网倾向非常高的时候，需要综合考虑客户投诉的原因，以便作出相应决策进行挽留。

(3)年龄因素：一个人的年龄会影响其选择的通信服务类型，比如说年龄在时间段 16 到 24 的客户选择通信企业时可能更看重资费、年龄大于 40 以上则倾向于选择服务质量好、网络覆盖率高的通信企业，由此年龄的变化对客户的忠诚度存在影响。

(4)地区因素：中国电信业存在许多家企业，在每个地区所提供的服务是不一样，

比如中国联通的网络覆盖率比较低，在中国偏远地区甚至没有信号，经常在此地工作的客户担心会因网络问题而未接到重要来电，这样会影响客户的忠诚度。

(5)职业：不同职业的客户对通信服务的要求是不一致的，学生用户可能会选择资费便宜的校园套餐，从事服务业的客户则可能会选择接电话不需要资费的通信服务以及从事销售业的客户则会选择通信质量好的通信服务，因此职业变化会导致客户的流失。

(6)漫游时间：若一个用户的漫游时间过长，则预示着该用户可能工作的地点发生了改变，这种类型的用户也极有可能流失，选择漫游地区企业提供的通信服务，因此漫游时间过长会导致客户的流失。

除此以外，还包括其他的一些因素。通过对类似这样的分析以及对相关资料的研究分析，本文选择如下一些影响客户流失的因素作为贝叶斯网络模型的节点集，具体如下：

- (1)地区；
- (2)年龄；
- (3)职业；
- (4)欠费状态；
- (5)漫游时间；
- (6)网间通话比例；
- (7)市话比例；
- (8)长话比例；
- (9)短信比例；
- (10)总消费额；
- (11)投诉次数；
- (12)流失状态；

影响因素(1)至(3)来源于对客户属性资料，影响因素(4)至(10)来源于客户消费行为资料，影响因素(11)来源于客户投诉行为资料，影响因素(12)是根据先验知识选择的，事实上其他的因素均为预测客户忠诚度指数而服务的，这样基于贝叶斯网络的客户忠诚度预测模型的属性选择就完成了（以 A, B, C, \dots, L 分别表示），因此对于贝叶斯网络的学习均基于以上选择的节点集。

由于各个属性的取值不一，且有的属性的取值区间非常大，因此为了处理方便，需要对这些数据的取值进行标准化处理。又因为贝叶斯网络技术要求所输入的数据必须是离散值，因此也需要对属性取值是连续型的进行转换，将其转换成离散型，确定节点的状态空间。结点的状态空间可以由专家根据先验知识获得，同样也可以根据结点数据本身的特点进行分析获得。

为了降低贝叶斯网络模型概率推理的计算复杂度，本文将所有因素的状态离散成三个级别，即高、中以及低三种，在数据库中用数值 1、2 以及 3 分别对应表示。下面具体分析各自表示的含义：

(1)地区指数：若该号码所在地区为国内一、二线城市，比如说省会城市、沿海城市以及经济较发达的城市，则在数据库中用 1 代替存储，若该号码属于地级市区和县区，则在数据库用 2 代替存储，城镇地区以及农村地区用 3 代替存储。

(2)年龄指数：根据年龄分为三类，则 0~25、25~45 以及 45 以上，在数据库中分别用 1、2 以及 3 进行存储。

(3)职业指数：职业指数无法根据职业类型进行统一描述，因此只能根据该职业用户的收入水平分为三类，即收入水平从 0-4000、4000-10000 以及 10000 及以上，在数据中分别用 1、2 以及 3 代替存储。

(4)欠费状态：欠费状态根据是否欠费以及欠费时间长度将欠费分为三个级别，即未欠费、欠费但未超过 1 个月以及欠费两个月，在数据库中用 1、2 以及 3 进行存储。

(5)漫游时间：根据漫游时间的长度将漫游时间指数分为三部分，即短期漫游、中期漫游以及长期漫游，对应的时间长度分别为 0-1 个月，1-3 个月以及 3 个月以上，在数据库中分别用 1、2 以及 3 表示。

(6)网间通话比例、市话比例、长话比例以及短信比例：这四者都是表示各自的消费额所占的消费总额的比例，因此可以根据所占的比例的大小划分为三个等级，即比例低于 10%、比例在 10%至 30%以及比例超过 30%，在数据库中分别用 1、2 以及 3 表示。

(7)总消费额：该消费额表示的是月消费总额，根据消费的额度将其分为三个等级，即低额消费、中额消费以及高额消费，分别对应 0-100、100-500 以及 500 以上，在数据库中分别用 1、2 以及 3 表示。

(8)投诉次数：投诉次数对于客户忠诚度状态的预测极其重要，根据次数将其分

为三个等级，即低频率投诉、中频率投诉以及高频率投诉，分别对应的次数为 0-5、5-10 以及 10 次以上，在数据库中分别用 1，2 以及 3 表示。

(10)忠诚度状态：同样也有高、中以及低三种表示，其由证据变量推导出来。

4.4.5 模型构建

在客户忠诚度预测模型节点集确定的基础上，需完成两个任务，第一个为确定拓扑结构，第二个确定条件概率分布表。下面从上述两方面进行分析。

(1)拓扑结构确定

在客户忠诚度预测模型中，首先，我们认为“年龄”与“网间通话比例”、“欠费状态”、“短信息比例”以及“投诉”之间存在关系。“年龄”大的客户的联系人数量比较多，联系人中使用的通信服务不一，会导致“网间通话比例”增大。“年龄”小的客户由于没有赚钱能力因此容易欠费停机。“年龄”小的客户喜欢发短信，因此“短信比例”相对“年龄”大的客户要大一点。同时“年龄”小的客户对于通信服务过程中遇到的不满意的情况，则会进行“投诉”，而“年龄”稍长的客户可能会通过其他方式进行交涉。

其次，“职业”与“总消费金额”、“欠费状态”、“市话比例”以及“漫游时间”存在联系。“职业”对“消费总额”的影响在于高收入“职业”的客户选择的更有可能是高价套餐，打电话的频率也会更高，同样低收入“职业”的客户选择通信服务的企业更可能考虑的是价格。同样高收入“职业”的客户出差的机会比较多，漫游时间也相对起来会比较多，“市话比例”会相应的降低，同时为了随时是实现通信，因此他们一般不会使自己的手机“欠费”。

“地区”与“职业”以及“欠费状态”存在依赖关系。处于一线城市的“地区”的收入相对地级市“地区”而言要高出很多，而“职业”是按照收入分类的，因此存在一定的联系。营业网点多的一二线城市“地区”，客户交费比较方便，客户停机时能及时续交话费，而城镇以及农村“地区”营业网点比较少，缴费不方便，则可能造成客户长时间处于“欠费状态”。

经过以上的分析，基本就确定了客户忠诚度预测模型中网络结点之间的相互关系，从而可以描绘出其客户忠诚度预测模型的贝叶斯网络。具体的图形如图 4.11 所示。

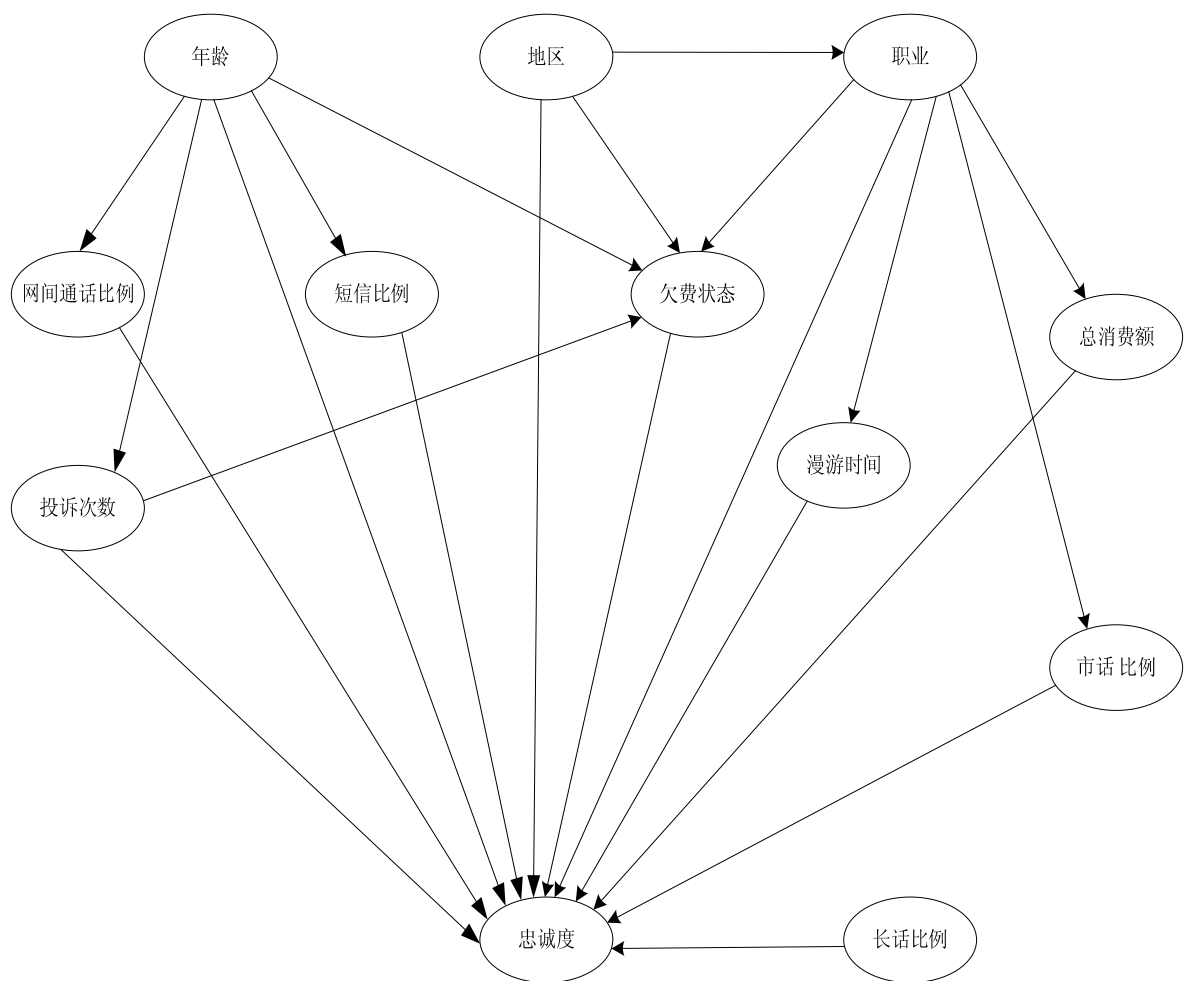


图 4.11 贝叶斯网络图

(2) 拓扑结构确定

在模型拓扑结构已经确定的基础上，需确定模型的条件概率表，因此模型预测与条件概率表息息相关。没有条件概率表，模型也就无法发挥作用。条件概率表的确定流程如下：首先明确各节点的概率分布表，其通常使用先验知识 β 进行确定。在上述基础上，需利用样本数据来对先验知识确定的概率表进行修正，修正方式为找到 $P(S|B, \beta)$ 最大值所对应的网络结构。本课题的先验知识 β 取值了 $1/3$ ，代表每个状态的发生概率都是一样的。

显然先验知识所确定的概率表具有很强的主观性，在此概率表下进行客户忠诚度预测，准确率会比较低，还需要深入分析样本数据，从样本数据找寻到因子之间的影响大小，进而对概率表进行更新。本课题采用了极大似然原理，以最大概率为基本准则，从而可实现问题的转换，将其转换为最优化问题。

不妨设在对历史样本数据进行分析之后，获得了客户分类模型中的某些节点观察值集合 B ，不妨以 α 代表贝叶斯网络结构的任意性，根据最大概率原则，实际上就是求使得 $P(B|\alpha)$ 的取值最大的 α 的值。根据前章对贝叶斯网络参数学习的介绍可以知道 $L(\alpha) = P(B|\alpha)$ 或 $L(\alpha) = \ln(P(B|\alpha))$ 就是似然函数。求 $L(\alpha)$ 的最大值即为极大似然估计。

设 B 为贝叶斯网络节点集的一个子集，并且已知 B 种的元素满足如下条件：即 B_1, B_2, \dots, B_n 之间相互独立。根据极大似然原则，对于 B_h 而言，目标在于寻找一个网络结构使其在在 B 的联合概率密度上取得最大值，不妨用 $P_w(B)$ 表示节点集的联合概率密度，同时令 W 为网络结构的联结概率： W_{ijk} 为第 i 个节点在其父节点取第 j 个状态时第 k 个状态的概率，而函数 $\ln(P_w(B))$ 是 $P_w(B)$ 的严格递增函数，因此求 W 使 $P_w(B)$ 达到极大与求 W 使 $\ln(P_w(B))$ 达到极大两者是等价的。本文采用梯度下降的方法进行计算，首先计算 $\ln(P_w(B))$ 对 W_{ijk} 的偏导数，再求和。

$$\frac{\partial \ln(P_w(B))}{\partial w_{ijk}} = \sum_{h=1}^m \frac{\partial P_w(B_h)}{\partial w_{ijk}} \times \frac{1}{P_w(B_h)} \quad (4.1)$$

$Z_{i1} = k$ 表示事件的第 i 个节点处在第 k 个状态， $Z_{i2} = j$ 表示事件的第 i 个节点的父节点还处在第 j 个状态， $Z_i = (k, j)$ 表示事件的第 i 个节点在第 k 个状态且第 i 个节点的父节点在第 j 个状态。

$$\frac{\partial \ln(P_w(B))}{\partial w_{ijk}} = \sum_{h=1}^m \frac{P_w(Z_i = (k, j) | B_h)}{w_{ijk}} \quad (4.2)$$

$P_w(Z_i = (k, j) | B_h)$ 指在已知条件 B_h 时，第 i 个点在父节点取第 j 个状态且第 j 个节点在第 k 个状态的联合概率。从而得到数据分析后的客户忠诚度预测模型的的的结点条件概率表。

(3)模型运用

在模型构建成功的基础上，需要使用根据模型来获得客户的忠诚度状态，在实际使用时，需要利用概率知识。对于一个电信客户而言，若已明确除忠诚度状态之外的子节点的值，则可以得到忠诚度指数，不妨假设存在一个完整的数据集，

首先给出一个完整的数据集，推导客户的“忠诚度状态”，“地区”为农村，即 $A=3$ ，“年龄”位于 25 至 45 之间，即 $B=2$ ，客户的“职业”所带来的收入为 5000，即 $C=2$ ，“欠费状态”为未欠费，即 $D=1$ ，“漫游时间”超过 3 个月，即 $E=3$ ，“网间通话比例”为 8%，即 $F=1$ ，“市话比例”为 50%，即 $G=3$ ，“长话比例”为 20%，即 $H=2$ ，“短信比例”为 4%，即 $I=1$ ，“总消费额”为 300，即 $J=2$ ，“投诉次数”为 5，即 $K=2$ ，根据条件概率公式，我们可以求出客户“忠诚度状态”为 1、2 以及 3 的概率，用数学公式表示如下：

$$P(L=1|A=3,B=2,C=2,D=1,E=3,F=1,G=3,H=2,I=1,J=2,K=2) \quad (3.3)$$

$$P(L=2|A=3,B=2,C=2,D=1,E=3,F=1,G=3,H=2,I=1,J=2,K=2) \quad (3.4)$$

$$P(L=3|A=3,B=2,C=2,D=1,E=3,F=1,G=3,H=2,I=1,J=2,K=2) \quad (3.5)$$

在对比上述三项的结果后，对其进行比较，概率较大的则为对应的忠诚度指数，若高忠诚度的概率较大，则其为忠诚客户，依次为稳定客户和低忠诚度客户。

4.4.6 模型验证

通过上述步骤，客户忠诚度预测模型已经构建完成，但是构建完成并不代表模型能应用到客户忠诚度预测系统中，需对模型的准确率进行验证，准确率过低反而会给企业带来损失，这是不可接受的，为此本节将对模型进行验证，以便了解模型的准确率。

在验证过程中重点关注两项指标，第一项为准确度，第二项为错误度，准确度的含义是实际流失的名单与预测为低忠诚度的客户数量的比值，错误度的含义是低忠诚度客户被预测为高忠诚度的概率，前者出现错误对企业不会造成损失，若后者出现错误，则可能错失采取措施挽留客户的机会，只能放任客户流失，这是电信企业不愿意看到的。

利用测试集对模型进行验证，得到了上述两个指标的值，准确度为 94%，后者为 5%，从结果可知，模型准确率较高，系统可以部署到电信企业用于预测客户忠诚度。

4.5 安全性设计

系统安全的目标在于综合各种安全措施来防止电信客户忠诚度预测系统免遭攻击，当发生网络威胁等事件时^[28]，需将威胁事件控制在最小范围内，在设备管理系统中主要采取如下几项措施：第一项为代码部分，第二项引入防病毒技术，下面

对各个部分进行阐述。

(1) 防火墙安全方案

从保障电信客户忠诚度预测系统的安全出发^[29]，本系统在外网边界部署了一套防火墙，根据实际情况在防火墙上设置有效的过滤原则，如协议过滤以及端口过来等，同时对于服务器的访问只开放与系统相关的服务，其他服务全部关闭，防止不法分子利用 TCP/IP 协议的漏洞来攻击服务器。

(2) 代码安全

在代码层面有许多方面可采取措施来提升系统的安全性和可靠性，本系统在安全方面做了如下工作^[30]。

第一、引入身份识别功能防止非法用户进入电信客户忠诚度预测系统，保证进行业务操作的都是合法用户^[31]。

第二、引入数据备份恢复措施。数据备份和数据恢复措施的引入可处理一般的数据库故障，防止出现数据丢失等意外情况，可提升数据可靠性^[32]。

第三、强制用户定时修改密码，在用户登录时，会检测用户多长时间未修改密码，若超过四个月未更新用户密码，则会自动进入密码修改界面，要求用户更新密码。

4.6 本章小结

本章的主要目的在于对电信客户忠诚度预测系统进行详细设计，首先对系统的软件框架、网络拓扑结构以及逻辑功能框架进行了阐述，然后以时序图的方式对各模块重点功能的工作流程进行了介绍，最后利用贝叶斯网络技术构建了客户忠诚度预测模型，以便从中查找出低忠诚度客户名单，上述工作为系统的实现奠定了基础。

第 5 章 系统实现

5.1 实现环境

系统能部署在何种环境中是部署时需重点考虑的事情，否则最终开发无法应用到实际的环境中，或者系统性能会受到影响，这些都会影响系统的可用性。本章将对电信客户忠诚度预测系统的实现环境进行介绍，具体如表 5-1 所示。

表 5-1 实现环境

软件环境	
操作系统	Microsoft Windows XP
系统开发环境	MyEclipse 平台、JDK
数据库	服务器：Microsoft SQL Server 2005
浏览器	IE 系列或者火狐
硬件环境	
服务器	包括应用服务器和数据库服务器两部分，两者的配置要求一致，都要求 CPU 为 4 核，同时内存大于 4G，硬盘大于 500G
客户端设备	客户端计算机：内存需大于 2G，硬盘需大于 350G

5.2 系统实现

本节将对电信客户忠诚度预测系统的重点功能的代码实现进行介绍。

5.2.1 导入功能的实现

在电信客户忠诚度预测系统中有许多场景需要使用导入、导出功能，如客户信息管理以及消费资料管理等，为此将导出功能的代码抽取出来成为公共代码，从而可提升开发效率。导出功能的核心代码如下所示：

```

public bool DataSetToExcel(DataSet dataSet, bool isShowExcle)
{
    DataTable dataTable = dataSet.Tables[0];
    int rowNumber = dataTable.Rows.Count;
    int columnNumber = dataTable.Columns.Count;
    if (rowNumber == 0)
    {
        return false;
    }

    //建立 Excel 对象
    Microsoft.Office.Interop.Excel.Application excel = new Microsoft.Office.Interop.
    excel.Application();
    excel.Application.Workbooks.Add(true);
    excel.Visible = isShowExcle; //是否打开该 Excel 文件
    //填充数据
    for (int c = 0; c < rowNumber; c++){
        for (int j = 0; j < columnNumber; j++)
        { excel.Cells[c + 1, j + 1] = dataTable.Rows[c].ItemArray[j]; }
    }
    return true;
}

```

5.2.2 短信挽留功能的实现

短信挽留功能就是利用短信或者邮件的方式向低忠诚度客户发送信息，包括优惠政策等，以便达到挽留客户的目的。在实现短信发送时，本系统所使用的技术为短信猫，短信猫只是短信发送流程中的一部分，下面对整个短信发送流程进行阐述。

(1)构建一个短信发送线程，以此来发送短信，在短信发送线程中，需要使用 BlockingQueue 队列。

(2)BlockingQueue 队列的作用在于存储待发送的短信任务，BlockingQueue 队列的特点在于包含阻塞机制，当队列中存在任务时，则发送，不存在短信任务时，则阻塞，处于等待状态，直到队列中有短信任务加入。

(3)当系统发送短信告警时，系统需要构建一个短信任务实例用于存储待发送的短信内容，并将其添加到 BlockingQueue 队列中。

下面对短信发送流程的核心代码进行阐述。

```

public SmsSendThread extends Thread//定义一个短信发送线程
{

```

public BlockingQueue<SmsContent> smsQueue = new ArrayBlockingQueue<SmsContent>(1000);//定义一个容量为 1000 的 BlockingQueue 队列，以此来存储短信任务。

```
public void run()//定义一个不死线程，处于不断的运行中
{
    //take 函数包含阻塞机制，存在任务这发送，否则阻塞
    SmsContent smsInfo = smsQueue.take();
    SmsServer server = getSmsSever();
    if(server!=null)//判断是否配置短信服务器，
    {
        //有配置，则发送短信，否则记录失败原因。
        sendInformation(server,smsInfo);}
    else
    {
        //记录短信发送失败原因}
    }
}
```

sendInformation 函数主要利用短信猫来发送短信，本文所使用的为某企业的专业短信猫，其提供完整的 API 来支持短信发送，下面对具体的短信实现进行阐述。

短信发送功能的实现代码如下所示：

```
public String sendInformation(SmsServer server,SmsContent smsInfo) {
    //传入的参数主要用于通过串口连接短信猫，在连接成功的基础上才能发送短信
    boolean sendflag = false;//表示短信猫的连接状态
    String return_num="";//存储短信发送结果
    sendflag =this.getConnectionModem(server.getPort(),server.
    getRate(),server.getSn); //在短信猫连接成功的情况下，sendflag 为 true，否则为 false
    if(sendflag){//在该语句判断为真的情况下，可进行短信发送 byte[]send_con
    = smssendinfo.getUNIByteArray(smsInfo.GetContent ()); //短信发送 API 要求将短信内容转换为 Unicode，否则短信猫无法识别，会发送乱码
    String[] telephone_num= smsInfo.GetTelphonenum.split(",");
    for(int i=0;i< telephone_num.length;i++){//实现群发功能
    if(!smssendinfo.GSMModemSMSsend(null,8, send_con, telephone_num [i],
    false))
    {
        //代表短信发送未成功，需要记录原因，以便定位}
    }
    else
```

```

    { //打印发送失败原因 }
    if(return_num.equals("")){
        //代表短信发送成功
        return_num = "ok";
    }
    closeConnection(); //释放掉建立的连接，以便节省资源
    return return_num;
}

```

5.2.3 数据维护功能的实现

数据维护由数据备份和数据恢复两部分组成，本节将对其中的数据恢复功能进行详细阐述。数据恢复功能的实质是将备份文件中的 SQL 执行，将保存的数据再依次保存到数据库中，在这个过程中需要使用 SQL Server2005 提供的 restore 方法，其可以实现针对备份文件的处理。在恢复过程中，只需要将备份文件的路径传递到 restore 方法中即可。数据恢复功能的实现代码如下所示：

```

    以 dbname 存储数据库名称
    String dbname="datatest";
    try{File restorefile = new File(restoreFilePath);
        // 数据库名
        cs.setString(1, dbname);
        // 已备份数据库所在路径
        cs.setString(2, restoreFile_Path);
        // 还原数据库
        cs.execute();
        // 恢复数据库连接
        ps.execute();
        flag=true;
    }
    //用于捕获错误，以便定位故障原因
    catch (Exception e)
    {
        //打印错误
    }
}

```

5.2.4 贝叶斯网络算法实现

由前文的分析可以知道，贝叶斯网络的拓扑结构以及条件概率分布表都已经

确定。为了实现贝叶斯网络算法，本课题使用了基于 MATLAB 的贝叶斯网络工具箱 BNT，BNT 中提供了许多贝叶斯网络学习的底层基础数据库，支持精确推理以及近似推理。本节将对具体的实现进行介绍。

(1) 贝叶斯网络的表示方式

为了清晰的表示客户忠诚度预测模型中各个结点之间的关系，本课题使用矩阵的方式来表示贝叶斯网络的结构，若两个结点之间存在关联，则矩阵的对应位置的值设置为 1.由此可明确各结点之间的关联。

(2) 贝叶斯网络学习算法函数

在 BNT 工具包中，存在许多学习算法，具体描述如下：

第一、扩展贝叶斯网络结构的 TANC 算法，即 `learn_struct_tan()`;

第二、数据完整条件下的参数学习算法，包括 K2 算法、贪婪算法以及爬山算法。

第三、数据缺失条件下的参数学习算法，包括最大期望 EM 算法以及马尔可夫链蒙特卡罗算法。

(3) 贝叶斯网络的推理

为了提高运算速度，使各种推理算法能够有效应用,BNT 工具箱采用了引擎机制,不同的引擎根据不同的算法来完成模型转换、细化和求解。整个推理过程如图 5.1 所示。

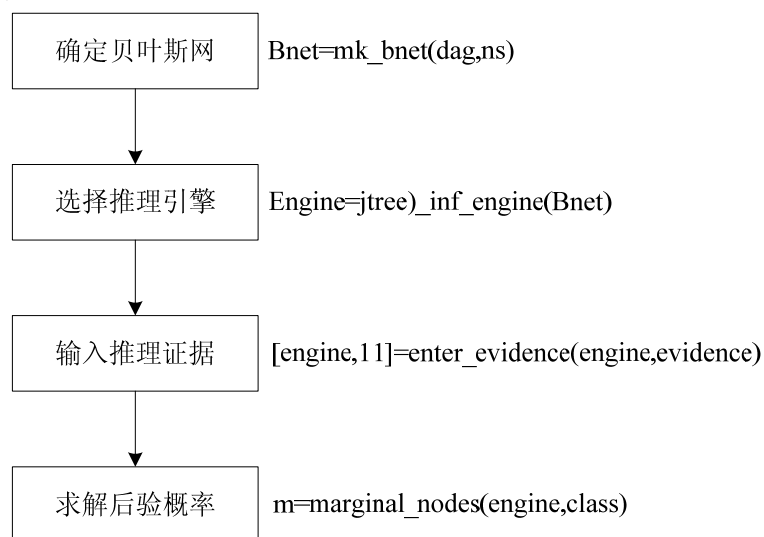


图 5.1 推理过程

(4) 贝叶斯网络算法的实现

由前文的分析可以知道，整个模型包含 10 个结点，各个节点存在三个状态值，为了将客户忠诚区别开来，以 class 进行表示。具体的 matlab 编码如下所示：

```

n=10;
ns=[3,3,2,2,2,2,3,3,3,2];

```



```

names={ A,B,C,D,E,F,G,H,I,Class};
A=1;B=2;C=3;D=4;E=5;F=6;G=7;H=8;I=9;Class=10;
order=[4,5,6,7,8,9,10,3,2,1]; %指定节点次序
max_fan_in=3;
result_matrix=zeros(ns(Class),ns(Class));
%读入训练数据集
fn='bdata_0313.txt' ;
load(fn); %装入测试例文件
data_train1=' bdata_0313' ;
[num_attribnum_cases]=size(data_train1);
data_train=zeros(num_attrib,num_cases);
%根据实例进行网络结构学习
dag_gbn=zeros(n,n);
dag_gbn=learn_struct_K2(data_train,ns,order,'max_fan_in',max_fan_in);
bnet2=mk_bnet(dag_gbn,ns);
%对生成的结构进行参数学习
priors=1;seed=0;rand('state',seed);
fori=1:n
bnet2.CPD{i}=tabular_CPD(bnet2,i,'CPT','unif','prior_type','dirichlet',
'dirichlet_type','BDeu','dirichlet_weight',priors);
End
bnet4 = bayes_update_params(bnet2,data_train);
CPT3=cell(1,n);
for(I = 1:n)
s = struct(bnet4,CPD[I]) ;
CPT3[i]=s.CPT;
END

```

5.3 系统应用

本节将对电信客户忠诚度预测系统的应用情况进行说明。

(1)登录

登录功能在系统安全中占据极其重要的位置，任意用户都必须经过登录功能的身份校验才能获得相应的操作权限，否则无法使用系统处理客户忠诚度预测业务。在登录时。需要根据用户提供的登录凭证到用户信息表进行匹配，若找到对

应的凭证，则识别成果， 否则失败。登录功能的实现界面如图 5.2 所示。



图 5.2 登录功能实现

(2)客户信息管理

客户信息管理模块主要实现对客户信息的动态管理，包括客户信息的添加以及删除等功能， 客户信息添加功能的实现界面如图 5.3 所示。

添加客户属性资料

用户姓名:	<input type="text"/>	归属市县:	沈阳市 <input type="button" value="v"/>
证件类型:	<input type="text"/>	证件号码:	<input type="text"/>
职业收入:	<input type="text"/>	客户状态:	<input type="text"/>
出生日期:	<input type="text"/>	电话号码:	<input type="text"/>
客户类型:	<input type="text"/>	计费方案:	iPhone 合约 <input type="button" value="v"/>
联系人数量:	<input type="text"/>	入网时间:	<input type="text"/>

添加

清空

返回

图 5.3 客户信息添加功能

(3)用户管理

用户管理用于实现对用户信息的动态管理，包括用户添加、删除以及修改等，系统包括工作人员以及管理员等两种用户类型，其主要实现对上述类型用户的管理。用户管理的界面如图 5.4 所示。



图 5.4 用户管理功能

(4) 忠诚度查询

工作人员可利用忠诚度查询功能获得各个客户的忠诚度指数，由此可得到低忠诚度客户名单，红色警报代表客户易流失，需采取挽留措施。忠诚度查询功能是系统易用性的体现，否则要工作人员从客户名单逐一查找低忠诚度客户的名单，则需要耗费工作人员大量的时间，会降低整体的工作效率。忠诚度查询功能的界面如图 5.4 所示：

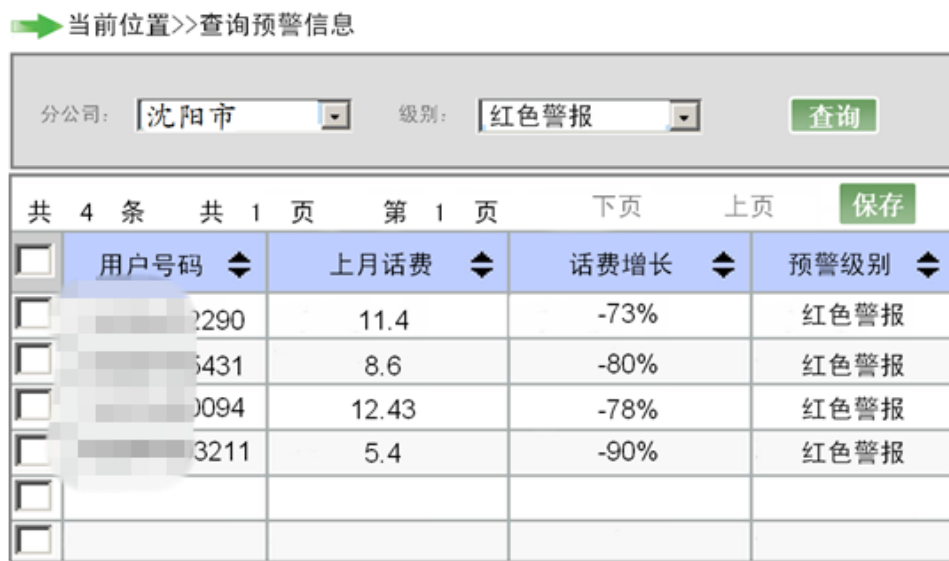


图 5.5 忠诚度查询

5.4 本章小结

本章从实现环境、代码实现以及系统应用等三个方面对系统的构建过程进行了详细说明，通过应用可以知道客户忠诚度预测系统可满足电信公司的应用需求，可部署到电信公司进行试用。

第6章 总结和展望

6.1 本文总结

电信市场的变革，即运营商的合并联盟以及新技术的发展（3G）意味着电信客户享有更多样的选择以及更便捷的服务，从而导致电信运营商之间的竞争越来越激烈。对于所有电信运营商而言，如何在激烈的市场竞争中保持稳定并获得新的发展，是企业管理者所必须思考的问题，顾客关系管理这一议题愈来愈受到企业的重视。因此客户对于企业而言不仅仅是销售和服务的对象，而且变成了企业的一项重大战略资源，是企业在竞争中立于不败的关键因素之一。基于此，本课题选择研究电信客户忠诚度预测系统。

本文从三个方面对本课题的工作内容进行总结归纳：

(1) 电信企业客户忠诚度预测系统的需求分析：分析阶段最重要的工作就是对系统进行功能分析，确定系统的功能架构。本课题所构建的电信客户忠诚度预测系统由安全防护、客户管理、客户挽留、消费资料管理以及忠诚度预测等功能组成。

(2) 电信企业客户忠诚度预测系统的设计。在详细设计过程中，模块设计的方法很多，本课题以时序图方式对电信企业客户忠诚度预测系统的各模块进行了详细设计，据此可了解各个模块的实现细节，由此可降低开发复杂度。为了提高采取措施进行有针对性的客户挽留，本系统构建了客户忠诚度预测模型，其可以为电信企业提供低忠诚度客户名单。

(3) 电信企业客户忠诚度预测系统的实现。在设计方案的指导下，实现了电信企业客户忠诚度预测系统。

6.2 进一步展望

目前系统已被部署到电信公司用于进行客户忠诚度预测，以便提升企业的竞争力，经过一段时间的试用，系统产生了很好的效果，企业的客户流失率得到了降低。但是在使用过程中还存在一些缺陷需要改进，具体表现在如下几个方面：

(1) 本课题所设计的系统并未实现对电信企业其他信息系统的对接，由此许多的数据都必须通过人工导入的方式，这导致效率过低，在未来可进行对接，系

统可定时读取相关系统的资料，有助于提升效率。

(2)本系统的重点是客户忠诚度预测，事实上模型可应用于反向逆推，即获取客户忠诚度低的原因，是由于价格原因或者服务原因，从而电信企业能采取更精确的维系策略，而不单单是优惠策略。

参考文献

- [1]刘晨, 孙秀杰, 李响. 电信企业客户维系挽留应用数据挖掘探讨[J], 商场现代化, 2008, 549: 29
- [2]朱明编著. 数据挖掘[M], 中国科学技术出版社, .2002.5.129-163
- [3]朱爱群. 客户关系管理与数据挖掘[M], 北京:中国财政经济出版社, 2001.8
- [4]刘元元. 基于数据挖掘的客户关系管理[M], 南京:南京理工大学, 2004
- [5]邵峰晶, 于忠清. 数据挖掘原理与算法[M], 中国水利水电出版社, 2003.8.1-11
- [6]朱明编. 数据挖掘[M], 北京:中国科学技术出版社, 2002, 129-163.
- [7]王辉. 用于预测的贝叶斯网络[J], 东北师大学报(自然科学版), 2002, (1):9-14.
- [8]王双成, 苑森淼, 王辉. 基于类约束的贝叶斯网络分类器学习[J], 小型微型计算机系统, 2004, 25(6):968-972.
- [9]茆诗松, 王静龙, 濮晓龙. 高等数理统计[M], 北京: 高等教育出版社, 1998.
- [10]邵峰晶, 于忠清. 数据挖掘原理与算法[M], 中国水利水电出版社, 2003, 1-11.
- [11]林士敏, 田凤占, 王双成. 贝叶斯网络的建造及其在数据采掘中的应用[J], 清华大学学报(自科版), 2001, 41(1):49-52.
- [12]C.K Chow and C.N Liu.Approximating discrete probability distributions with dependence trees[J],IEEE Transactions on Information Theory,1968,14:462-467.
- [13]Cooper G F.A simple constrain-based algorithm for efficiently mining observational databases for causal relationships[J],Data Mining and Knowledge Discovery 1997,1(1):203—224.
- [14]Jiawei Han 著, 范明, 孟小峰等译. 数据挖掘:概念与技术[M], 北京: 机械工业出版社, 2001.
- [15]王军, 周伟达. 贝叶斯网络的研究与进展[J], 电子科技, 1999, 11(8): 6-7.
- [16]王双成, 林士敏. 贝叶斯网络结构学习分析[J], 计算机科学, 2000, 27(10): 77-79.
- [17]李振, 肖书, 唐胜群.实用软件体系结构研究[J].计算机工程与应用, 2000, 36(8):169-171.
- [18]Heckerman D,Geiger D,Chickering D M.Learning Bayesian networks:the combination of knowledge and statistical data[J],Machine Learning,1995,20:197-243.
- [19]Wong M L,Lam W,Leung K S.Using evolutionary programming and minimum description length principle for data mining of Bayesian networks[J],IEEE Transactions on Pattern Analysis and Machine Intelligence,1999,21(2),174-178.

- [20]MIHELIS G, GRIGOROU D IS E.Customer satisfaction measurement in the private bank sector[J].European Journal of Operation Research2001,130(2):347-360
- [21]Holland J H.Adaptation in Nature and Artificial Systems.The University of Michigan Press,1975,MIT Press,2002
- [22]Zhang Yanhua,Chen Yuming,Chen Guijing.Adaptive Nearest Neighbour Prediction[J].安徽大学学报.2002,23(2):9-16,24
- [23]Hong S J, Weiss S M.Advances in Predictive models for data mining[J],Pattern Recognition Letters,2001,22(1):55-61.
- [24]Chickering, D M.Learning equivalence classes of Bayesian net structures[J],Machine Learning,2002,2:445-498.
- [25]Agrawal R, Imicliniski T, Swami A.Mining Association Rules Between Sets of Items in Large Databases[A],Proceedings of the 1993 ACM SIGMOD Conference[C],Washington, DC, 1993,5:207-216.
- [26]De Campos L M, Huete J F.A new approach for learning belief networks using independence criteria[J],International Journal of Approximate Reasoning,2000,24:11-37.
- [27]Marco Ramoni,Paola Sebastiani.Parameter Estimation in Bayesian Networks from Incomplete Databases[J],Intelligent Data Analysis,1998,2:139-160.
- [28]赵宏波, 孟雅玲. 数据挖掘在客户关系管理中的应用[J], 电信技术, 2001, 12: 9-11.
- [29]Sergios Theodoridis,Konstantinos Koutroumbas. 模式识别[J], 北京:电子工业出版社, 2004
- [30]高禹, 毕振波.软件开发过程模型的发展[J], 计算机技术与发展, 2008,7(18):83-86.
- [31]张桂元, 贾燕枫. Structs 开发入门与项目实践[M], 人民邮电出版社, 2000:35-96.
- [32]康斯坦丁(Constantine, L.L.)著, 谢超等译.人性化的软件开发[M], 北京: 人民邮电出版社, 2004.2: 42-43.

作者简介及在学期间所取得的科研成果

作者严浩，男，汉族，1979 年 4 月 5 日出生，辽宁沈阳人，大连理工大学电子信息工程毕业，工学学士学位。

致 谢

感谢陈思国教授三年来对我的关心和指导，感谢邴静高级工程师对我工作及论文上的帮助，感谢我的公司未提供的继续深造的机会，感谢我的家人及同事的理解和支持。谢谢!!