

# > SPSS China 电信行业应用手册

## 通过数据挖掘 增进盈利能力

### 目录

I. 前言.....	2
II. 客户流失.....	6
III. 营销响应.....	13
IV. 客户细分.....	21
V. 客户满意度.....	27
VI. CRISP-DM简介.....	34
VII. 数据挖掘经验谈.....	36
VIII. 数据挖掘部署策略.....	40
IX. 成功案例.....	41
国外成功案例.....	41
国内成功案例.....	46

# 1.前言

随着电信业务的发展和体制改革地不断深化,国内电信行业内部各大运营商之间的竞争日趋激烈,这一趋势在国内移动通信业内表现的尤为突出。移动通信运营商为了获取更多的客户资源和占有更大的市场份额往往采取“简单”的价格竞争和“此起彼伏”的广告宣传战,其弊端显而易见。这就要求运营商要采取以客户为中心的策略,根据客户的实际需求提供多样化、层次化、个性化的服务解决方案。因此,客户关系管理(CRM)成了电信运营商增加收入和利润,提高客户满意度、忠诚度的有效工具。在客户关系管理的流程中,为了准确、及时地进行经营决策,必须充分获取并利用相关的数据信息对决策过程进行辅助支持。近几年迅速发展起来的数据挖掘技术就是实现这一目标的重要手段。

从电信业务层面来讲,电信业务已从单纯的提供市话和长话服务演变为提供综合电信服务,如宽带、移动电话、语音、传真、图像、电子邮件、计算机和 Web 数据传输,以及其他数据通信服务。电信、计算机网络、因特网和各种其他方式的通信和计算机的融合是目前的大势所趋。而且随着许多国家对电信业的开放和新兴计算与通信技术的发展,电信市场正在迅速扩张并越发竞争激烈。因此,为了提高企业竞争力利用数据挖掘技术来挖掘现有电信业务能力,提高商业效率具有重要意义。

## 数据挖掘的概念

数据挖掘是根据企业的既定业务目标和存在的问题,对大量的业务数据进行探索,揭示其中隐藏的规律,并将其模型化,指导并应用于企业的实际经营。

数据挖掘是建立在数据仓库基础上的高层应用,但数据挖掘跟数据仓库的其它一些应用如 OLAP 分析、预定义报表和即席查询等有很大的区别。后三者通常是用户根据已知的情況对所关心的业务指标进行分析;而前者则是在业务问题和目标明确但考察的问题不清楚时,对数据进行探索,揭示隐藏其中的规律性,进而将其模型化。

电信运营商拥有许多成熟的数据库应用系统,如网管系统、财务系统、计费账务系统、112 障碍管理系统、缴费销账系统等,并产生了大量的业务处理数据。如果针对客户关系管理相关决策分析的需求,对这些数据进行重组整合,就能充分利用这些宝贵的数据,体现信息的真正价值。

## 数据挖掘技术在电信行业客户关系管理的主要应用领域如下:

### (1) 客户消费模式分析

客户消费模式分析(如固话话费行为分析)是对客户历年来长话、市话、信息台的大量详单、数据以及客户档案资料等相关数据进行关联分析,结合客户的分类,可以从消费能力、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,从而为固话运营商的相关经营决策提供依据。

### (2) 客户市场推广分析

客户市场推广分析(如优惠策略预测仿真)是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）客户欠费分析和动态防欺诈

通过数据挖掘，总结各种骗费、欠费行为的内在规律，并建立一套欺诈和欠费行为的规则库。当客户的话费行为与该库中规则吻合时，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。

### （4）客户流失分析

根据已有的客户流失数据，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，找出这些数据之间的关系，并给出明确的数学公式。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。这就彻底改变了以往电信运营商在成功获得客户以后无法监控客户流失、无法有效实现客户关怀的状况。

### 数据挖掘成功方法论

按照 CRISP-DM（跨行业数据挖掘方法论），数据挖掘可以划分为以下六个步骤：



图 1.1：数据挖掘项目实施流程示意图

#### 商业理解

在这个阶段要明确面临的商业问题和数据挖掘想要达到的目的，完成商业问题到挖掘问题的定义过程。这阶段将产生的主要文档及提交物包括：

- 确定商业目标。包括电信行业背景，需要实现的商业目标和成功标准等；
- 形势评估。包括电信运营商拥有的资源、需求、假定和限制、风险偶然性、专业术语和成本收益等；
- 在上述基础上确定数据挖掘的目标和成功标准；
- 制定的项目计划和阶段性里程碑的设定说明。

明确挖掘的目的。在进行数据挖掘前，一定要明确挖掘的目标。因为电信业务变化快，不同的时期，挖掘的对象并不一样，挖掘目标也千差万别，而挖掘目的不一样，决定了挖掘使用的模型和算法完全不一样。例如：电信开通新视通业务后，根据不同的类型用户对新视通业务的使用情况，预测 2-3 个月内可能使用该业务的用户类型，使用机率和话务量。

明确影响挖掘目的的可能因素和结果。影响用户消费行为的因素是多种多样的，要比较的指标也是多种多样的，所有这些因素均应在模型中考虑。比如，要考虑 IP 业务对用户的影响，不仅要分析用户性质，还要分析用户消费额度；不仅要分析已经使用 IP 业务的用户性质和消费额度分布，而且要分析已使用 IP 业务的用户在未使用该业务前在原来的未开通 IP 业务用户中的分布和比例，还要分析使用 IP 业务的用户使用前和使用后消费行为的变化。

## 数据理解与数据准备

数据理解和数据准备在真实的数据挖掘项目中会占很大的比重，搜集所有需要的数据和数据质量的保证是数据挖掘结果正确与否的关键。

需要确定数据来源、确定数据整理方法。如对 IP 业务开通对用户行为的影响，原始数据来自用户资料、用户详单。根据上述确定的因素，定义原始指标和衍生指标。

## 建立模型

针对不同的数据挖掘目标 and 数据特性，应采用不同的挖掘算法建立模型，并对筛选出的每种候选模型进行数据处理并产生结果。目前在电信行业常用的算法有：C5.0 决策树、CART 决策树、神经元网络和 Logistic 回归等算法。

## 模型检验

对产生的模型结果需要进行比对验证、准确度验证、支持度验证等检验以确定模型的价值。在这个阶段需要引入更多层面和背景的用户进行测试与验证，通过对几种模型的综合比较，产生最后的优化模型。

## 模型发布与应用

建立和检验模型并不是数据挖掘的目的，只有把模型发布到相关决策者手中，才能使我们通过数据挖掘提高企业利润或降低企业成本。例如在研究客户流失这个案例中，模型建立好后除了提交一份完整的报告和测试结果外，还需要通过发布工具将模型嵌入到用户的应用系统中，使决策者和相关管理者可以及时作出决策，实施相应的市场措施。

模型发布后并不意味一个数据挖掘项目的结束，数据挖掘系统与业务系统间存在着作用与反作用的交互关系，随着时间的推移和数据的变化，这个闭环作用的系统中的很多关键参数需要及时调整，才能保证挖掘结果的质量并延长其有效的生命周期。所以有时我们说“数据挖掘是一门科学，同时也是一门艺术”。

## 国内数据挖掘应用中存在的问题

### 数据质量和完备性

国内电信运营商现有的、面向事务的数据在质量、完整性和一致性上存在许多问题，必须投入大量的精力去进行数据的抽取、净化和处理。此外，业务问题的相关数据有时难以全面收集。例如客户信用是客户价值评估中的关键因素，但由于国内未建立完善的信用体系，无法根据现有客户数据建立优质的信用评价模型，从而导致客户价值模型有效性的降低。



## 相应的人员素质

在数据挖掘应用过程的多个环节中，人的主观辨识和控制是应用成败的关键，这就对系统使用人员提出了很高的要求。如果没有具备相应素质的使用和维护人员，必将导致分析系统与现实脱钩，无法达到预期效果。

## 应用周期

数据挖掘存在一个较长的应用周期。技术本身不能给使用者解决任何问题，只能从数据中把一些潜在的情况呈现到使用者面前，由使用者采取相应措施。数据挖掘应用的有效方法是：从一个较小的、关键的问题出发，建立起相对有效的模型，并通过应用实践不断检验和完善模型，逐步替使用者解决问题。

## 电信数据挖掘项目实施的意见建议

### 数据挖掘理解上的误区：

- a) 数据挖掘系统将会直接告诉你有关商业问题的答案；
- b) 数据挖掘系统将在工作流程中替代专业业务管理人员的角色；
- c) 数据挖掘系统的建设过程中无视内部的知识 and 专家经验；
- d) 数据挖掘系统是理论性的、以研究学习为导向的方法。

### 数据挖掘项目的建议：

结合项目的投资确定数据挖掘项目的目标和回报周期，保证项目的可实施性。

- a) 数据挖掘项目整个实施流程的管理和控制是非常重要的；
- b) 数据挖掘项目需要和用户的专业业务人员紧密配合，共同寻找答案；
- c) 培训业务人员，传递产品技能，提供挖掘分析的技术支持，让业务专家而不是技术人员成为项目的主导者；
- d) 需要提供外部观点和一般经验，同时与内部特色相结合；
- e) 确保实际的，以结果为导向的项目原则。

本手册将从客户流失、客户细分、客户满意度和营销响应四个大方面，使用 Clementine 具体讲述数据挖掘在电信行业的具体应用。

## 2.客户流失

随着电信体制的改革和不断深化，国内电信行业内部各大运营商之间的竞争日趋激烈。电信运营商为了获取更多的客户资源和占有更大的市场份额，往往采取名目繁多的促销活动和层出不穷的广告宣传来吸引新客户。然而，统计发现，发展一个新客户比保持一个老客户的费用要高 7 倍，而如果“用户保持率”增加 5%，将有望为运营商带来 85% 的利润增长。因此，对老客户的保留直接关系到运营商的利益，无论是客户流失还是话务量流失都将对运营商的经营产生深远的影响。针对这一问题，目前国外电信界应用最为广泛的解决方案之一，是应用数据挖掘技术，研究流失客户的特征，从而对流失进行预测、并对流失的后果进行评估，采取客户保留措施，防止因客户流失而引发的经营危机，提升公司的竞争力。

具体说来，客户流失是指客户终止与企业的服务合同或转向其它公司提供的服务。客户流失分析是以客户的历史通话行为数据、客户的基础信息、客户拥有的产品信息为基础，通过适当的数据挖掘手段，综合考虑流失的特点和与之相关的多种因素，从中发现与流失密切相关的特征，在此基础上建立可以在一定时间范围内预测用户流失倾向的预测模型，为相关业务部门提供有流失倾向的用户名单和这些用户的行为特征，以便相关部门制定恰当的营销策略，采取针对性措施，开展客户挽留工作。

### 客户流失需要解决的问题

#### 1) 哪些现有客户可能流失？

客户流失的可能性预测。主要对每一个客户流失倾向性的大小进行预测。

#### 2) 现有客户可能在何时流失？

如果某一客户可能流失，他会在多长时间内流失。

#### 3) 客户为什么流失？

哪些因素造成了客户的流失，客户流失的重要原因是什么。主要对引起客户流失的诸因素进行预测和分析

#### 4) 客户流失的影响？

客户流失对客户自身会造成什么影响？

客户流失对电信公司的影响如何？对可能流失客户进行价值评估，该客户的价值影响了运营商将要付出多大的成本去保留该客户。

#### 5) 客户保留措施？

针对电信公司需要保留的客户，制定客户和执行保留措施。

### 电信客户流失的类型

为了避免由客户流失造成的损失，必须找出那些有流失危险和最有价值的客户，并开展客户保留活动。电信行业的客户流失现象可以分为以下三种情况：

1) 公司内客户转移：客户转移至本电信公司的不同网络或不同业务。主要是电信公司增加新业务，或者资费调整引发的业务转移，例如从普通的固定电话转至 IP 电话。这种情况下，虽然就某个业务单独统计来看存在客户流失，并且会影响到公司的收入，但对公司整

体而言客户没有流失。

2) 客户被动流失：表现为电信运营商由于客户欺诈或恶意欠费等行为而主动终止客户使用网络和业务。这是由于电信运营商在客户开发的过程中忽视了客户质量造成的。

3) 客户主动流失：客户主动流失可分为两种情况。一种是客户不再使用任何一家电信运营商的电信业务；另一种是客户选择了另一家运营商，即所谓的“客户跳网”。“客户跳网”的原因主要是客户认为公司不能提供他所期待的价值，即公司为客户提供的服务价值低于另一家电信运营商。这可能是客户对电信公司的业务和服务不满意，也可能是客户仅仅想尝试一下别家公司提供而本公司未提供的新业务。这种客户流失形式是研究的主要内容。

### 如何进行客户流失分析？

对于客户流失行为预测来说，需要针对客户流失的不同种类分别定义预测目标，即明确定义何为流失，进而区别处理。预测目标的准确定义对于预测模型的建立是非常重要的，它是建立在对运营商的商业规则和业务流程的准确把握的基础之上。在客户流失分析中有两个核心变量：财务原因 / 非财务原因，主动流失 / 被动流失。对不同的流失客户按该原则加以区分，进而制定不同的流失标准。例如，非财务原因主动流失的客户往往是高价值的客户，他们会正常支付服务费用并容易对市场活动有所响应，这种客户是电信企业真正需要保留的客户。而对于非财务原因被动流失的客户，对于预测其行为的意义不大。

研究哪些客户即将流失时，是一个分类问题。将现有客户分为流失和不流失两类，选择适量的历史上流失客户和未流失客户的属性数据组成训练数据集，包括：客户的历史通话行为数据、客户的基础信息、客户拥有的产品信息等。Clementine 提供人工神经网络，决策树，logistic 回归等模型用于建立客户流失的分类模型。

关于流失用户特征的分析，是一个属性约减和规则发现问题。Clementine 提供关联分析方法，可以发现怎样的规则导致客户流失。也可以利用 Clementine 的决策树方法，发现与目标变量——是否流失——关系最为紧密的用户属性。由于不同类型的客户可能具有不同的流失特征，因此，在进行深入的客户流失分析时，需要先进行客户细分，再对细分之后的客户群分别进行挖掘。

在预测客户流失时一个很重要的问题是流失的时间问题，即一个客户即将要流失，那么它可能什么时候会流失。生存分析可以解决这类问题。生存分析不仅可以告诉分析人员在某种情况下，客户可能流失，而且还可以告诉分析人员，在这种情况下，客户在何时会流失。生存分析以客户流失的时间为响应变量进行建模，以客户的人口统计学特征和行为特征为自变量，对每个客户计算出初始生存率，随着时间和客户行为的变化，客户的生存率也发生变化，当生存率达到一定的阈值后，客户就可能流失。

分析客户流失对客户自身的影响时，主要可以考虑客户的流失成本和客户流失的受益分析。客户流失成本可以考虑流失带来的人际关系损失等因素，通过归纳客户的通话特征来表征。减少客户流失的一个手段就是增加客户的流失成本。客户流失的受益分析就是判断客户流失的动机，是价格因素还是为了追求更好的服务等。这方面内容丰富，需作具体分析。

分析客户流失对电信公司的影响时，不仅要着眼与对收入的影响，而且要考虑其他方面的影响。单个的客户流失对电信公司的影响可能是微不足道的，此时需要研究流失客户群对电信公司收入或业务的影响。这时候可能需要对流失客户进行聚类分析和关联分析，归纳客户流失的原因，有针对性的制定防止客户流失的措施。

在预测出有较大流失可能性的客户后，分析该客户流失对电信公司的影响。评估保留客户后的收益和保留客户的成本。如果收益大于成本，客户是高价值客户，则采取措施对其进行保留。至于低价值客户，不妨任其流失甚至劝其流失。

总之在利用数据挖掘研究客户流失问题时，需要明确并深入理解业务目标，在明确的业务目标的基础上准备数据、建模，模型评估，最后将模型部署到企业中。

## 案例分析

为了举例说明，我们设想为一个虚构的电信公司 ABCTelcom 的客户流失建模。ABCTelcom 原本是市场的领先者，但目前正受到来自其它电信公司日益激烈的竞争。由于竞争对手接连推出了一系列新产品，并进行了大量的促销活动。最近半年来，ABCTelcom 的客户流失较为严重。为了保持其战略性市场主导地位，ABCTelcom 公司计划开展客户保留活动。在活动进行之前，为了尽可能提高活动收益，ABCTelcom 需要对现有的客户的数据进行分析，从众多客户中找出流失可能性高的优质客户并针对其开展活动。而对于流失可能性低或者保留成本大于收益的客户，则可以不展开活动。此外，客户流失预测也能帮助 ABCTelcom 发现那些申请服务后不久就欠费停机的客户，从而减少这类客户带来的损失。

### 1. 商业理解

预测现有客户在未来六个月内流失的概率并分析保留客户活动的成本和收益，针对收益大于成本的客户开展客户保留活动，以减少公司优质客户的流失。

### 2. 数据理解

#### a) 数据说明

选取一定数量的客户信息（包括流失的和未流失的），选择相关的客户属性（包括客户资料、通话行为特征、消费属性、客服信息等）。利用直方图、分布图等数据分析工具初步确定哪些因素可能影响客户流失。所选取的数据说明如下：

1. 用户编号
2. 长途通话时间
3. 国际通话时间
4. 本地通话
5. 掉线次数
6. 付款方式
7. 本地话单类型
8. 长途话单类型
9. 年龄
10. 性别
11. 婚姻状态
12. 孩子数目



13. 估计收入

14. 是否有车

15. 是否流失

其中客户流失状态有三种属性：

1. 被动流失

2. 主动流失，这是分析中特别关注的一类客户

3. 未流失，仍然是 ABCTelcom 的客户

在分析中，我们主要关注的是，主动流失的客户。被动流失对电信公司来说是意义最小的，因为被动流失通常是客户发生欺诈，欠费等行为后不再继续使用该公司服务。主动流失指的是客户停止在 ABCTelcom 的业务，转向了其竞争对手，这通常是因为别的公司能够提供更切合客户需求的产品服务，是该行关注的焦点。在具体的应用之前，ABCTelcom 将所有现有的客户归到上述的三个类别中。同时，所有的人口统计信息（也就是从年龄到是否有车）每六个月更新一次，而交易信息（从长途通话时间到长途话单类型）则是每个月实时更新。为了让预测模型能预先进行指示以便采取补救措施，在目标变量（因变量）和输入变量（自变量）之间设定了 6 个月的延迟。比如，人口统计信息和交易信息都是 2007 年 1 月收集的，而是否流失则是 2007 年 7 月的数据。也就是说，输入变量采集六个月后再将客户流失状态分类；因此该模型提早 6 个月预测客户流失。

#### b) 数据描述和图形

Clementine 提供了一系列数据描述和可视化的工具，可以帮助分析人员对数据分布以及数据之间的关系进行基本探索。这些工具包括：数据审核，统计分析，网络图，直方图，散点图，两步聚类等等。

图 2.1 描述了客户的性别分布。可见，主动流失的客户以女性为主。而被动流失的客户男女比例基本持平。在三种客户状态中，被动流失的比例是最小的。

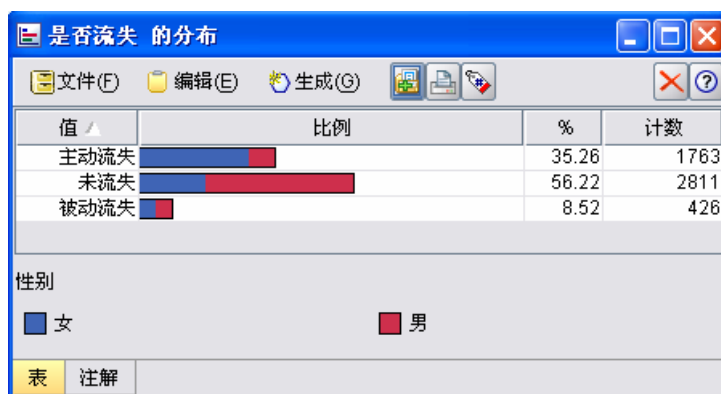


图 2.1：用户流失与性别的关系

### 3. 数据准备

在数据准备阶段，分析人员根据数据理解的结果准备建模用的数据，包括数据选择、新属性的派生，数据合并等。

在本例中，由于长途通话时间，国际通话时间和本地通话时间都是反映用户的对电话的使用情况，因此，将三者合并，得到新的通话总时间变量：

通话总时间 = 长途通话时间 + 国际通话时间 + 本地通话时间

由于用户编号是一个系统自动生成的属性，不具有实际意义，因此，在分析中将用户编号排除到模型之外。

#### 4. 建立模型

将准备的数据划分为训练集和检验集，首先利用 C5.0 决策树模型进行属性约减，然后以约减后的属性为自变量，以是否流失为因变量，训练神经网络模型，得到相应的客户流失预测模型。对检验集应用该模型，并根据预测结果的准确性评价模型。下图是模型建立和评估的数据流图。

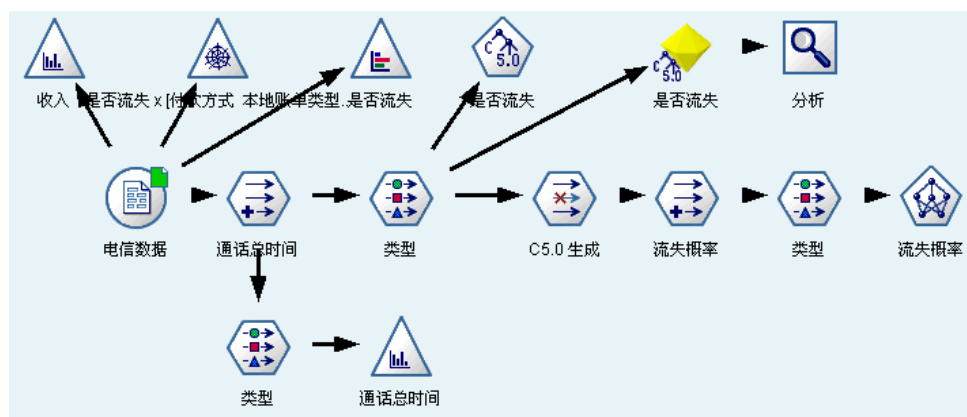


图 2.2: Clementine 中对客户流失建模的数据流图

执行数据流，C5.0 算法得到的规则集如下图所示。可见其中年龄，收入，国际通话时间，本地通话时间和性别是最重要的五个变量。C5.0 得到的规则集里，有三类客户很容易主动流失。但是这只是一个初始的分析结果，还需要使用神经网络对这五个变量进行进一步建模。

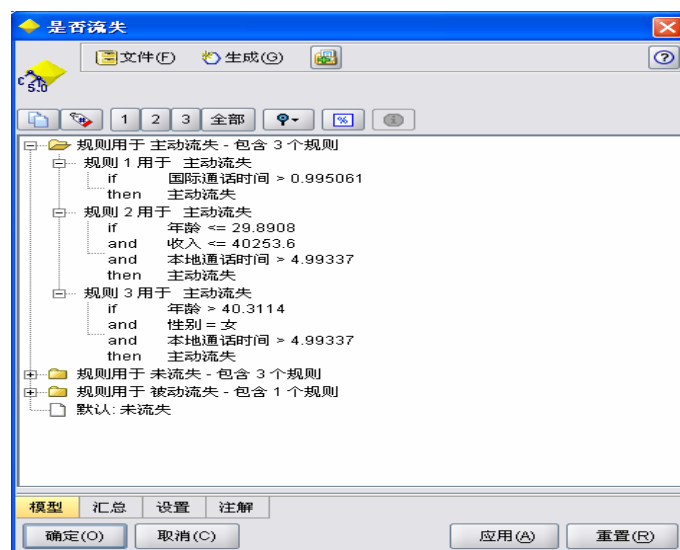


图 2.3: Clementine 中对客户流失进行属性约减得到的结果

由于我们的关心是客户主动流失，因此派生一个新变量流失概率，如果是是否流失属性为主动流失，则取值为 1，否则取值为 0。采用年龄，收入，国际通话时间，本地通话时间和性别为输入变量，流失概率为目标变量，根据数据集训练得到一个 5-1-1 结构的神经网络模型信息如下。

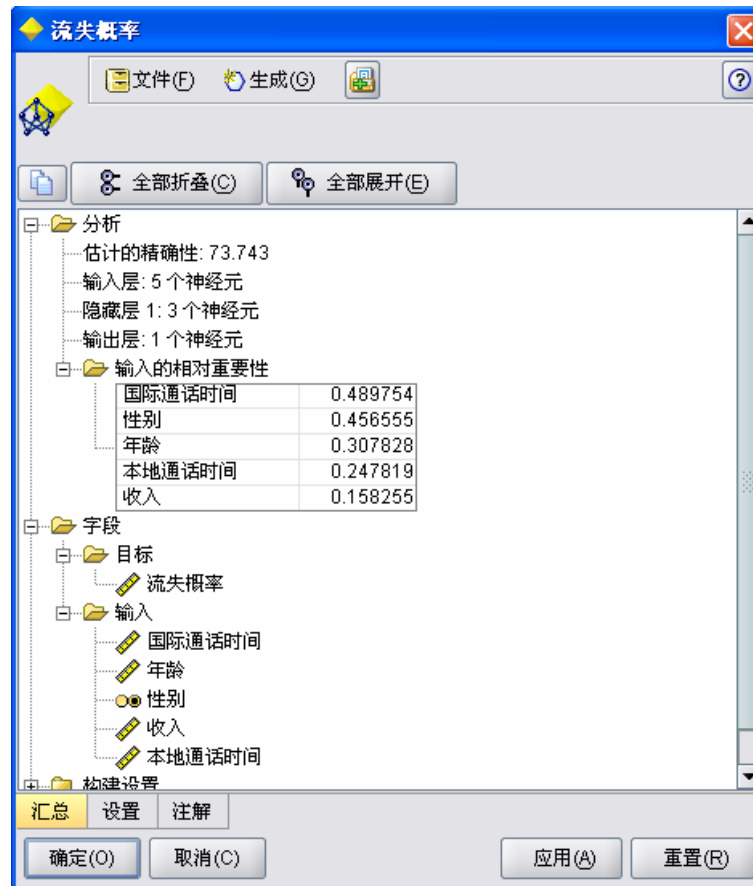


图 2.4: Clementine 中采用神经网络对客户流失建模得到的结果

## 5. 模型评估和部署

对训练集应用神经网络模型，可以对每个客户流失的可能性打分。将客户按照流失概率由大到小排序，然后根据公式：价值=长途通话时间\*2+国际通话时间\*5+本地通话时间\*1 派生出每个客户的价值大小。分析客户价值和流失概率之间的关系，对高价值高流失概率的客户进行采取措施，如给与一定的优惠，进行挽留，对低价值的客户可以任其流失。

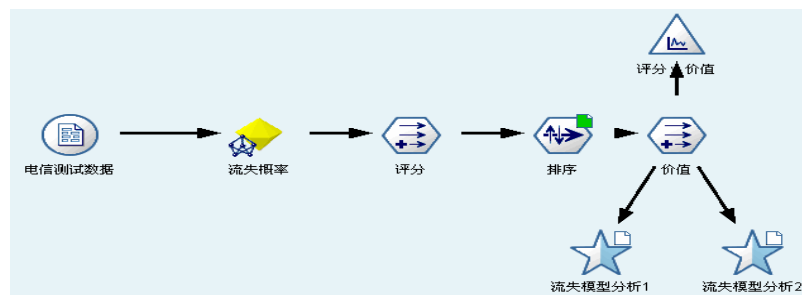


图 2.5: 模型评估和部署的数据流图

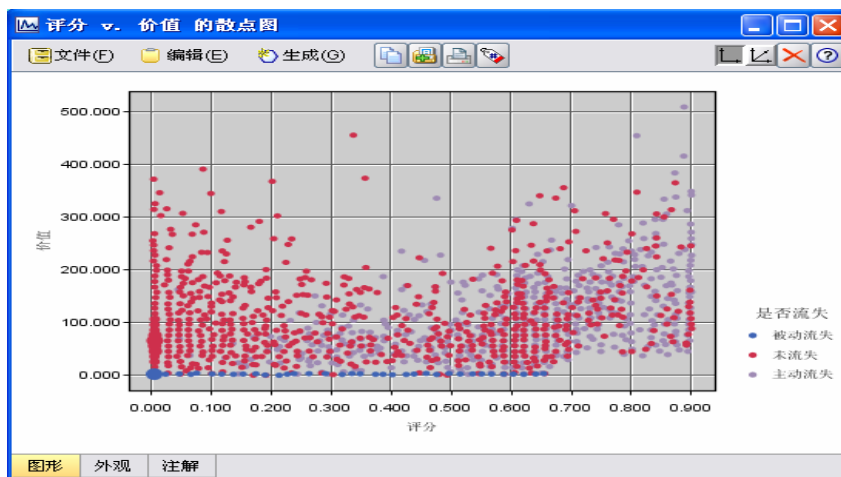


图 2.6: 评分和价值的散点图

对高价值高流失概率的客户采取营销活动进行挽留的成本和收益如下图所示。可见，通过及时地发现要流失的客户并根据客户价值及时采取挽留措施，可以避免因客户流失而带来的损失。在验证模型后，可以利用 Clementine Publisher 发布模型，使得模型可以在多种环境中使用。

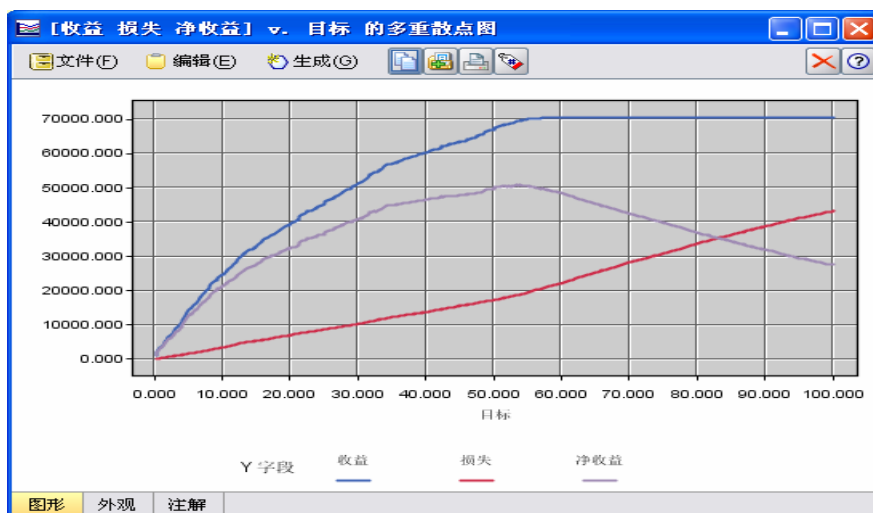


图 2.7: 营销活动的成本和收益

## 3.营销响应

### 为什么要进行营销响应分析？

为了发展新客户和推广新产品，电信公司通常会针对潜在客户推出各种直接营销活动。然而，如果目标客户的选择不明确，营销活动往往花费巨大而取得的实际效益不佳，甚至可能遭遇由于活动响应率太低而无法收回成本的境况。事实上，一方面，在当今竞争激烈的电信市场上，客户每天通过短信，电话，邮件，电子邮件，网站广告等方式会接触到大量的电信业务广告。缺乏针对性和足够吸引力的营销活动往往会被客户直接忽略。另一方面，电信用户越来越看重个性化服务，对新的电信业务具有较大的需求。

为了更好的满足客户需求，许多公司采用了促销活动管理系统来帮助执行促销活动。这些管理系统帮助增加了公司采取的促销活动的数量，却并不一定能改善促销活动的效率。事实上，不合适的促销活动和过多的促销活动只会导致用户对公司的不满意程度增加。

所以，有效促销活动不在于数量的多少，而在于要在恰当的时机，通过恰当的方式，向恰当的用户推销恰当的产品。也就是说，有效的促销活动，不在于涉及客户的数量多少，而在于针对的都是具有高响应概率的目标人群。这不仅可以提升客户的满意度，增强客户对公司的忠诚度，而且可以降低客户获取费用，增加营销活动投资回报率，直接带来公司效益的增加。

提高营销响应率的成果是显然的。假设有 100000 个潜在客户，每客户期望平均收益为 ¥100 元，若我们将响应率提高 0.5%，达到 1%，则花费同样的市场营销费用，每月能多增加收入 ¥50000 元。

### 什么是营销响应？

营销响应模型是电信企业使用最频繁的一种预测模型。目标变量是预测谁会对某种产品或服务的宣传进行响应，因变量是客户及其行为的各种属性，如：客户年龄，客户收入，客户最近一次购买产品的时间，客户最近一个月的购买频率等。利用响应模型来预测哪些客户最有可能对营销活动进行响应，这样，当以后有类似的活动时，利用响应模型来预测出最有可能的响应者，从而对他们采取相应的营销活动。而对响应度不高的客户就不用对他们进行营销活动，从而减少活动成本，提高投资回报率。

### 如何提高营销响应率

电信公司应当在深入了解客户需求和客户特征的基础上，制定营销策略，从而达到增加营业收入和客户满意度的双重目标。我们提倡的不是针对最佳的客户群展开营销活动，而是针对每一个客户开展活动。所谓“知己知彼，百战不殆”，建立在对客户需求良好把握基础之上极具针对性的营销将极大地提高营销活动的成功率。要开展这样的营销活动，首先需要回答以下几个问题：

- 对谁开展营销活动？
- 多长时间开展一次营销活动？
- 何时开展营销活动？
- 如何开展营销活动？



电信公司在数据挖掘技术的帮助下，针对客户数据建立营销响应模型，在合适的时间，通过合适的渠道，以一种合适的接触频率，对合适的客户开展活动，从而提高营销活动的响应率和投资回报率。



图 3.1: 营销活动的四个要素

#### 1) 选择合适的客户

电信公司对以往的营销数据进行分析，采用决策树分类等数据挖掘方法，识别出具有高响应率特征的客户作为营销活动的目标客户。通过选择合适的客户，可以排除对促销活动反应不积极的客户，将目标客户的数量大大减小，从而在实现更有针对性地营销的同时减小营销成本。据统计，通过减小目标客户的数量，通常可以节省 25-40% 的营销费用，同时增加营销响应率。

#### 2) 选择合适的渠道

第二步是要针对用户选择合适的营销渠道，也就是和客户接触的方式。通过使用每个用户偏爱的方式与之接触，也有利于提升客户响应率。在确定促销渠道时，要考虑客户对渠道的偏爱，渠道成本，期望的响应率，其他营销限制条件等。

#### 3) 选择合适的时间

在当今竞争激烈的社会里，客户有很多满足自己需求的机会和选择。因此，一旦发现了客户尚未被满足的需求或者出现遗失客户风险时，一定要及时和客户接触。这种事件驱动的促销方式，通常也可以取得较高的响应率。

#### 4) 选择合适的活动频率

此外，并不是促销活动越多，效果越好。活动计划者需要根据实际情况，针对具体的客户，选择一个最优的活动次数，既使得客户的各种需求得到较好的满足，又避免因为过于频繁的接触而导致客户的反感。同时，过多的营销活动，也会增加营销成本。需要在提高的成本和提高的响应率带来的收益之间寻找一个最优点。客户自身的偏好对于营销活动的频率确定也至关重要，比如，对于不喜欢过于频繁的接到促销电话的客户，就要适量降低电话促销的频率。通过选择更有针对性地选择客户和根据客户的需求和偏好来推广促销活动，可以将促销活动的投资收益率提高 25%—50%。

表 3.1: 营销活动过程

阶段	1. 合适的客户	2. 合适的渠道	3. 合适的时间	4. 合适的营销频率
目标	为营销活动选择最佳客户	为目标客户选择最优营销方式	在合适的时间对目标客户开展营销	确定最适合客户的营销频率
方法	预测分析	渠道优化	事件营销	促销优化
策略	预测谁是最有可能响应营销并且能给营销活动带来收益的客户	在客户偏爱的方式和公司的成本与能力之间选取最优结合点	变小的, 经常性的营销活动为事件触发的营销活动	在客户响应率和公司收益之间选择最佳结合点
好处	降低 25%-40% 的营销成本	降低接触客户的成本	响应率的提高至少达到两倍	提高 25%-50% 的收益

### 案例分析

为提升业绩, 某电信公司要决定执行直接营销活动。为此, 分析人员需要识别出可能回应直接营销活动的客户, 从而展开有针对性的直接营销。

#### 1. 商业理解

识别出可能回应直接营销活动的客户, 提高营销活动的响应率。

#### 2. 数据理解

收集部分以往的营销活动数据 (包括对活动响应的客户数据和未对活动响应的客户数据), 选择客户属性, 包括客户人口统计学特征和账户信息等。利用直方图、分布图来初步确定哪些因素可能影响客户响应。所选取的数据属性包括: 编号, 年龄, 收入, 孩子数目, 是否有汽车, 是否抵押贷款, 居住区域, 性别, 婚姻状况, 在该银行是否有储蓄帐户, 在该银行是否有活期账户, 是否对促销活动响应等 12 个字段。

其中是否响应是预测的因变量, 共有两个属性:

否: 客户未响应营销活动

是: 客户响应营销活动

	用户编号	年龄	收入	孩子数目	是否有车	是否抵押	居住区域	性别
1	ID12001	24	48630	3	是	是	城郊	男
2	ID12004	32	34277	2	是	否	市区	女
3	ID12006	73	28351	1	否	是	农村	女
4	ID12009	47	60740	2	是	否	农村	女
5	ID12010	38	65901	1	是	是	城镇	男
6	ID12014	47	56279	1	是	是	城郊	男
7	ID12015	48	39822	1	否	否	市区	男
8	ID12027	29	48937	0	是	是	农村	女
9	ID12029	44	42020	1	是	是	城郊	男
10	ID12030	48	66882	3	是	是	农村	男
11	ID12031	33	26697	2	否	是	城镇	男
12	ID12032	20	15443	1	是	是	农村	男
13	ID12036	21	50116	1	是	是	城镇	男
14	ID12037	53	26380	1	是	否	城镇	男
15	ID12039	44	48361	2	是	是	城郊	女
16	ID12040	63	41384	1	是	是	城郊	男
17	ID12045	52	65890	2	是	是	城镇	男
18	ID12052	65	31047	0	是	否	农村	男
19	ID12053	47	29664	2	否	否	市区	男
20	ID12054	52	76439	3	是	否	城镇	男

图 3.2: 数据

首先采用直方图，散点图等工具对数据之间的关系进行初步探索。下图是按照响应与否察看收入与孩子数目之间的散点图。可见，如果只考虑“孩子数目”属性，发现随着孩子数目增加，响应的客户比率降低。同时考虑“孩子数目”和“收入”属性，发现响应比率与“收入”和“孩子数目”的比值相关，这个比值通常被成为“相对收入”。

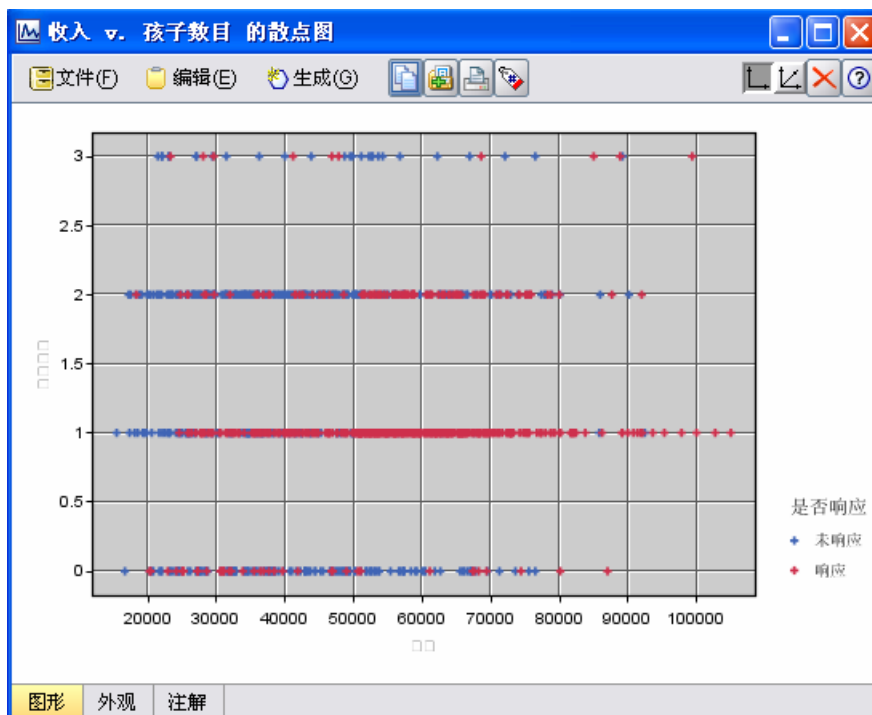


图 3.3: 收入与孩子数目的散点图

下图是孩子数目的分布图。有一个孩子的客户数目最多，占到了 44.3%。而在这些有一个孩子的客户中，大部分是对直接营销活动进行响应的客户。总的说来，随着孩子数目增加，客户响应率降低。

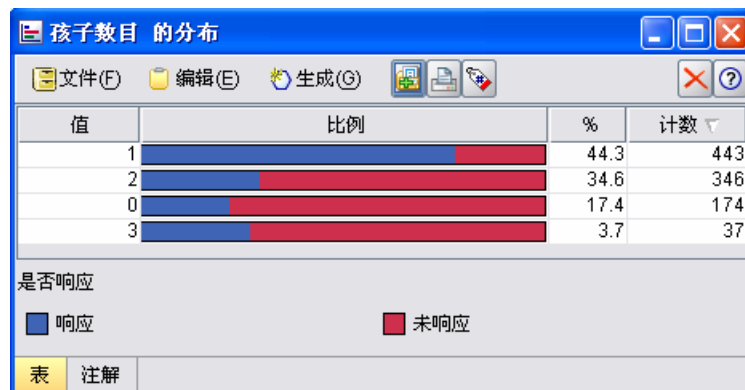


图 3.4: 孩子属性分布图

### 3. 数据准备

根据数据理解的结果准备建模用的数据，包括数据选择、新属性的派生，数据合并等。

由于用户编号是一个系统自动生成的属性，不具有实际意义，因此，在分析中将用户编号排除到模型之外。

此外，在数据理解中发现，是否响应与“收入”和“孩子个数”的比率有关，因此，派生出“相对收入”属性，定义为：

如果“孩子个数”为 0，则“相对收入” = “收入”；

否则，“相对收入” = “收入” / “孩子个数”。

在下边的分析中，收入和孩子个数将被排除到模型之外，而代之以“相对收入”进行分析。

### 4. 建立模型及评估

对数据进行预处理之后，分别使用 C5.0 决策树分类模型，神经网络模型，C&RT 决策树分类模型，以客户属性为输入变量，以客户是否响应为目标变量进行分类。然后对测试集分别应用这三个模型，选取效果最好的模型部署到企业中。

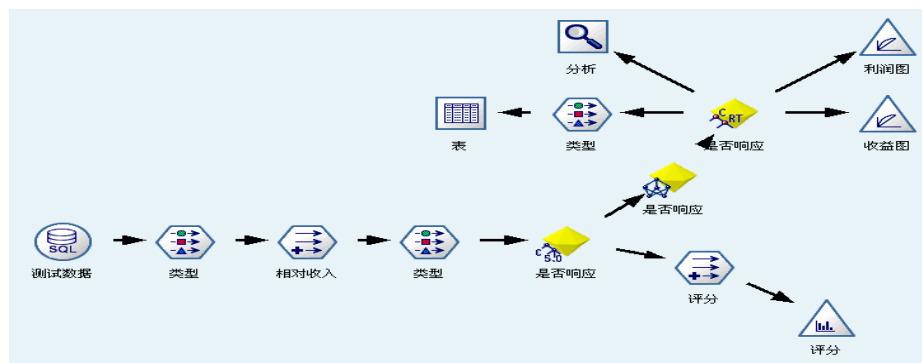


图 3.5: 部分数据流程图

使用 C5.0 决策树对是否响应建模，得到的部分规则集如下图所示。

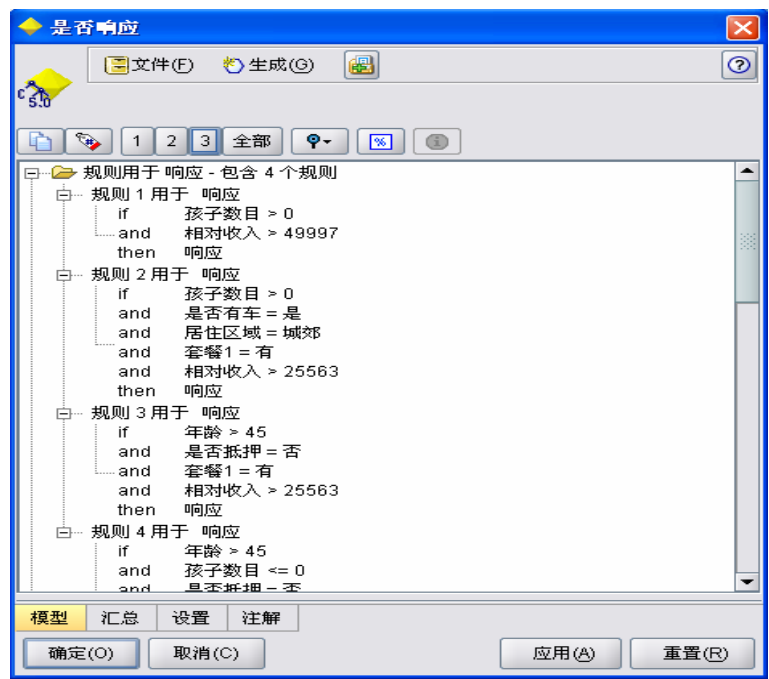


图 3.6: C5.0 决策树分类结果

使用神经网络模型对是否响应进行分类，得到的结果如下。通过训练，得到一个 20-3-2 结构的神经元网络模型，其估计精度达到了 87.77%。其中最重要的字段包括：相对收入，孩子数目，收入等。

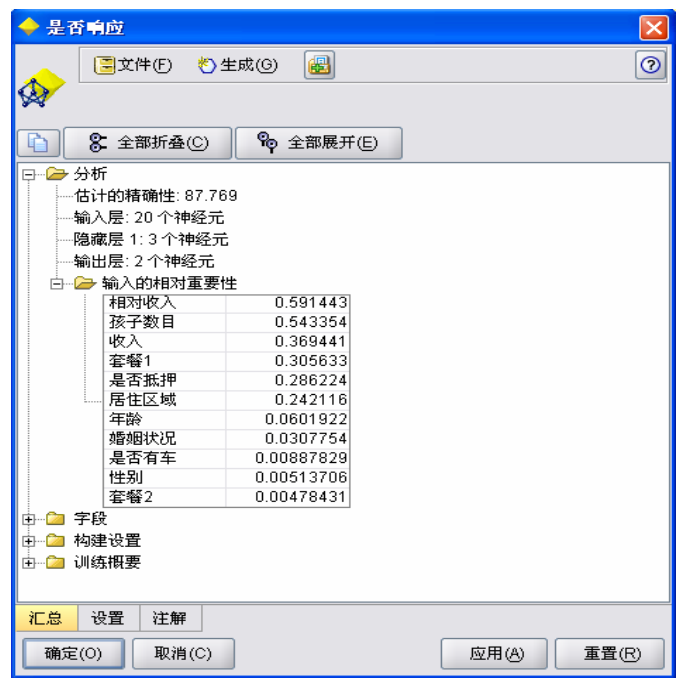


图 3.7: 神经网络分类结果

使用C&RT决策树对是否响应建模，得到的分类树如下图所示。可见，当相对收入小于



25564.5 时，客户倾向于不响应；当收入大于 25564.5，孩子数目小于等于 0.5（需要根据实际情况进行分析），没有抵押贷款，且年龄小于等于 45 时，倾向于不响应；当收入大于 25564.5，孩子数目小于等于 0.5，没有抵押贷款，且年龄大于 45 时，这样的客户响应率高。

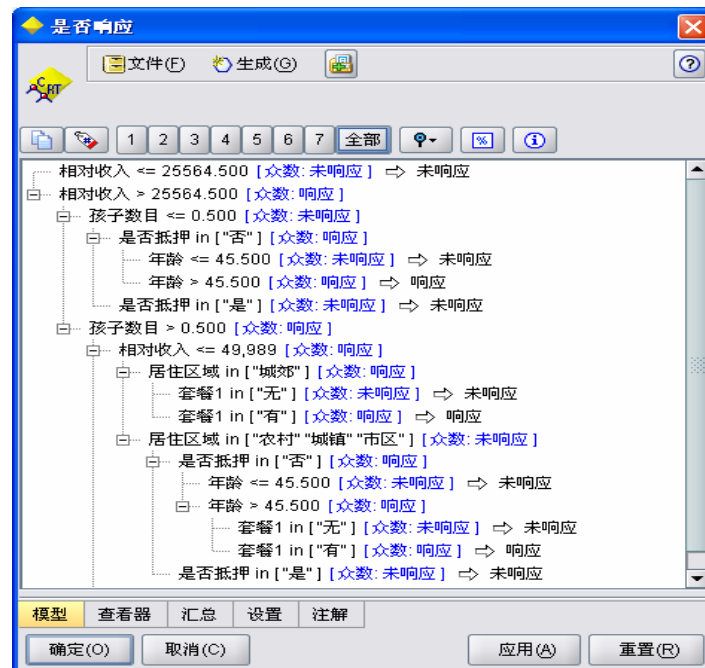


图 3.8: C&RT 分类结果

之后使用测试集评估不同模型的表现。下图给出了模型评估的结果。其中是否响应表示目标变量的真实值，\$C-是否响应代表使用C5.0 模型得到的目标变量预测值，\$N-是否响应代表使用神经网络模型得到的目标变量预测值，\$R-是否响应代表使用C&RT模型得到的目标变量预测值，可见，这三个模型中，C5.0 的预测精度是最高的，达到了 95.29% 以上。最后，还可以查看不同模型预测结果的一致性。

输出字段: 是否响应的结果			
个别模型			
比较 \$C-是否响应 与 是否响应			
正确	11,435	95.29%	
错误	565	4.71%	
总计	12,000		
\$C-是否响应的同步矩阵 (行表示实际值)			
	响应	未响应	
响应	2,090	367	
未响应	198	9,345	
比较 \$N-是否响应 与 是否响应			
正确	11,008	91.73%	
错误	992	8.27%	
总计	12,000		
\$N-是否响应的同步矩阵 (行表示实际值)			
	响应	未响应	
响应	2,057	400	
未响应	592	8,951	
比较 \$R-是否响应 与 是否响应			
正确	11,332	94.43%	
错误	668	5.57%	
总计	12,000		
\$R-是否响应的同步矩阵 (行表示实际值)			
	响应	未响应	
响应	2,081	376	
未响应	292	9,251	
\$C-是否响应 \$N-是否响应 \$R-是否响应 之间的一致性			

图 3.9: 模型评估

5. 模型部署

通过建模和评估后，选择预测精度最高的 C5.0 模型部署到企业中。部署的数据流图如下图所示。新的用户数据在经过 C5.0 模型评分后，按照流失评分的高低排序，通过 Clementine Publisher 发布。

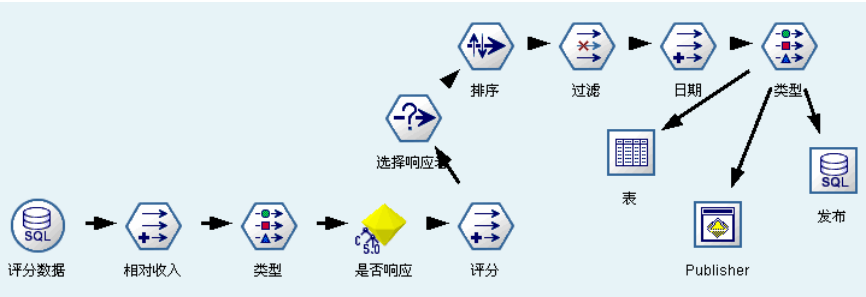


图 3.10: 模型部署数据流图

最后发布的结果如下表所示。评分就是某个客户流失的概率大小。

table (3 个字段, 2,320 条记录)			
	用户编号	评分	日期
1	ID19987	0.993	13/10/07
2	ID19985	0.993	13/10/07
3	ID19803	0.993	13/10/07
4	ID19187	0.993	13/10/07
5	ID16115	0.993	13/10/07
6	ID19261	0.993	13/10/07
7	ID14859	0.993	13/10/07
8	ID19677	0.993	13/10/07
9	ID19909	0.993	13/10/07
10	ID19871	0.993	13/10/07
11	ID19863	0.993	13/10/07
12	ID19607	0.993	13/10/07
13	ID19805	0.993	13/10/07
14	ID19803	0.993	13/10/07
15	ID19589	0.993	13/10/07
16	ID19529	0.993	13/10/07
17	ID19783	0.993	13/10/07
18	ID19475	0.993	13/10/07
19	ID16763	0.993	13/10/07
20	ID19421	0.993	13/10/07

图 3.11: 对新数据评分

## 4. 客户细分

### 客户细分的背景

随着电信体制改革的深化，中国加入 WTO，我国电信运营业的竞争也日趋激烈。现今，“个性化”营销、“一站式”服务、沟通“零距离”、“用户至上，用心服务”等词层出不穷。综合来说，这些与营销和服务相关的术语、理念，对象都是电信客户。服务源于良好的营销，因此，客户关系最终决定了电信市场营销。

我国电信业发展到今天，不仅初步形成六大运营商同台竞技的市场格局，而且从“卖方市场”转向“买方市场”，决定了客户关系的“大一统”必然向细分客户关系和“个性化”营销和服务的转变。因此，细分客户关系是电信营销的关键所在，客户细分可以对客户获取、客户保持、客户增值等提供全面的支持。恰当的细分标准必须得到有效的细分方法的辅佐，细分方法的选取应该取决于细分的目的和主题，当目的明确时可以选取若干细分标准而后进行多维细分，但当运营商着眼于更广泛的消费者特性分析时，这种主观定义的细分方法就显得力不从心，数据挖掘方法相比于基于统计的客户细分方法的功效就强大得多，运营商对数据挖掘的重视程度也日益加深。基于数据挖掘技术的分类方法，为更深入的客户细分提供有效的手段。与其他行业相比，中国电信运营业拥有更多有关用户的数据。谁能正确地挖掘与分析隐含这些数据中的知识，谁就能更好地向用户提供产品与服务，能够发现更多的商机，从而在竞争中获胜。

### 客户细分的概念

客户细分的概念是美国市场学家温德尔·史密斯(Wendell R. Smith)于 20 世纪 50 年代中期提出来的。

客户细分(Customer Segmentation)是指按照一定的标准将企业的现有客户划分为不同的客户群。客户细分是客户关系管理的核心概念之一，是实施客户关系管理重要的工具和环节。Suzanne Donner 认为：正确的客户细分能够有效地降低成本，同时获得更强、更有利可图的市场渗透。通过客户细分，企业可以更好地识别不同客户群体对企业的价值及其需求，以此指导企业的客户关系管理，达到吸引合适客户，保持客户，建立客户忠诚的目的。

所谓客户细分主要指企业在明确的战略、业务模式下和专注的市场条件下，根据客户的价值、需求和偏好等综合因素对客户进行分类，分属于同一客户群的消费者具备一定程度的相似性，而不同的细分客户群间存在明显的差异性。客户细分的理论依据主要有：

(1) 客户需求的异质性。影响消费者购买决策因素的差异决定了消费者的需求、消费者的消费行为必然存在区别。因此可以根据这种差异来区分不同的客户，客户需求的异质性是进行客户细分的内在依据。

(2) 消费档次假说。随着经济的发展和消费者收入水平的提高，消费量会随之增加。但消费量的增加并非线性增长，而是呈现出区间性台阶式的变化形式，一旦消费者达到某种消费层次之后，消费变化的趋势将变得非常平缓。根据消费档次假说，消费者的消费档次或消费习惯在一段时期内是相对稳定的，这就为通过消费行为来划分消费群体提供了理论前提和基础。

(3) 企业资源的有限性和有效市场竞争的目的性。资源总是稀缺的，由于缺乏足够的资源去应对整个客户群体，因此必须有选择地分配资源。为了充分发挥资源的最大效用，企业

必须区分不同的客户群，对不同的客户制定不同的服务策略，集中资源服务好重点客户。

(4) 稳定性。有效的客户细分还必须具有相对的稳定性，足以实现在此基础上进行的实际应用，如果变化太快，应用方案还未来得及实施，群体就已面目全非，这样的细分方法就显得毫无意义。

客户细分模型

客户群细分的目的是为了选择适合企业发展目标和资源条件的目标市场。客户细分模型是指选择一定的细分变量，按照一定的划分标准对客户进行分类的方法。一个好的细分模型，首先是要满足细分深度的要求，不同的使用者对客户细分的深度也有不同的要求，这就要求模型划分的结果能满足不同使用者的需要。其次是对数据的处理能力和容错能力，现代数据库的存储容量越来越大，数据结构也趋于多样性，误差数据也会随之增多，这就要求模型能适应数据在量和样上的膨胀，对误差数据能做出判别和处理。最后是模型要有很强的适用能力，变化是绝对的，而稳定只是相对的，无论是个人消费者还是消费群体，他们的消费行为都是在变化的，这就要求模型对客户的细分标准要随新的情况而不断更新。在对客户进行细分的方法中，除了传统的按照客户基本属性进行分类的方法以外，还有其他多种客户细分模型，如基于客户价值贡献度的细分模型、基于不同需求偏好的细分模型和基于消费行为的细分模型。基于消费者消费行为的客户细分模型研究，主要是以消费者的购买频率、消费金额等为细分变量，如 RFM 模型和客户价值矩阵模型。

(1) RFM 模型。RFM 细分模型是根据消费者消费的间隔、频率和金额三个变量来识别重点客户的细分模型。R—recency 指客户上次消费行为发生至今的间隔，间隔越短则 R 越大；F—frequency 指在一段时期内消费行为的频率；M—monetary 指在某一时期内消费的金额。研究发现，R 值越大、F 值越大的客户越有可能与企业达成新的交易，M 越大的客户越有可能再次响应企业的产品和服务。

(2) 客户价值矩阵模型。客户价值矩阵模型是在对传统的 RFM 模型修正的基础上提出的改进模型。用购买次数 F 和平均购买额 A 构成客户价值矩阵，用平均购买额替代了 RFM 模型中存在多重共线性的两个变量，消除了 RFM 模型中购买次数和总购买额的多重共线性的影响。在客户价值矩阵中，确定购买次数 F 和平均购买额 A 的基准是各自的平均值，一旦确定了坐标轴的划分，客户就被定位在客户价值矩阵的某一象限区间内。依据客户购买次数的高低和平均购买额的多少，客户价值矩阵将客户划分成四种类型，即乐于消费型客户、优质型客户、经常型客户和不确定客户，如图 4.1 所示。



图4.1：客户价值矩阵

客户细分并没有统一的模式，企业往往根据自身的需要进行客户细分，研究目的不同，用于客户细分的方法也不同。总的来讲，客户细分的方法主要有四类，一、基于客户统计学特征的客户细分；二、基于客户行为的客户细分；三、基于客户生命周期的客户细分；四、基于客户价值相关指标的客户细分。

### 客户细分模型的基本流程

客户细分包括六个基本流程：

**第一步：理解业务需求。**在未来的业务中，知道谁是客户是个非常好的起始点，以了解瞬息万变的市场环境。清楚地了解客户也是对每个客户组采取有针对性措施的基础。客户细分就是根据其特征将相似的客户归组到一起，这是了解客户和针对特定客户组进行市场定向所不可缺少的。客户细分可根据许多不同条件而进行。这些条件可由简单的年龄、性别、地理位置或这些变量的组合来构成。当这些条件变得越来越复杂时，数据挖掘技术就应运而生了。决定使用哪些条件取决于客户细分的目的和应用方法。在使用数据挖掘开发客户细分时，最重要的部分是其结果应当在业务远景中意义深远，并且能够在实际业务环境中进一步得到应用。需要记住的一点是：由于市场环境是动态变化的，细分建模过程应当是重复性的，且模型应随着市场的变化而不断革新。

**第二步：选择市场细分变量。**由于变量选择的优劣对细分结果质量的影响非常显著，所以变量选择应该建立在理解业务需求的基础之上，以需求为前提，在消费者行为和心理的基础上，根据需求选择变量。此外，变量的选择还应该有一定的数量，多了不好，少了也不好。

**第三步：所需数据及其预处理。**为创建数据模型，必须使用收集到的原始数据，并将其转换成数据模型所支持的格式。我们称这个过程为初始化和预处理。在运营商中进行客户行为细分通常需要行为数据和人口统计数据等类型的数据。行为数据是客户行为，可通过运营商的事务数据(账单数据、详单数据、CDR 数据等)而捕获。人口统计数据是当客户首次启用其通话服务时，他们会将客户人口统计数据(如年龄、性别、工作等)提供给运营商，这在识别或描述客户组的特征时很有用。

**第四步：选择细分技术。**目前，通常采用聚类技术来进行客户细分。常用的聚类算法有 K-means、两步聚类、Kohonen 网络等，可以根据不同的数据情况和需要选择不同聚类算法来进行客户细分。

**第五步：评估结果。**在对用户群进行细分之后，会得到多个细分的客户群体，但是，并不是得到的每个细分都是有效的。细分的结果应该通过下面几条规则来测试：与业务目标相关的程度；可理解性和是否容易特征化；基数是否足够大，以便保证一个特别的宣传活动；是否容易开发独特的宣传活动等。

**第六步：应用细分模型。**根据客户细分的结果，市场部门制定合适的营销活动，进行有针对性的营销。

总之，客户细分是电信运营商与用户二者实现双赢的重要举措。目前用户需求呈现多样化、个性化的趋势，只有通过深入分析用户消费行为，精确识别、细分用户市场，开发出针对不同层次用户的服务品牌进行服务营销，方能使得各方价值发挥到最大，实现共赢。不同级别的客户对服务的需求以及“赢”的概念是不同的，正是因为为不同的客户提供不同的服务才能使客户都达到满意，从而在市场上占据有利地位。



## 细分方法介绍

在数据挖掘中，往往通过聚类分析的方法来实现细分。聚类分析方法至少有以下几类：

1. **K-Means 聚类法**。使用者需要首先确定数据分为 **K** 群，该方法会自动确定 **K** 个群的中心位置，继而计算每条记录距离这 **K** 个中心位置的距离，按照距离最近的原则把各个记录都加入到 **K** 个群，重新计算 **K** 个群的中心位置，再次计算每条记录距离这 **K** 个中心位置的距离，并把所有记录重新归类，再次调整中心位置，依次类推……，达到一定标准时，结束上述步骤。这种方法运算速度快，适合于大数据量。

2. **两步聚类法**：这种方法首先需要确定一个最大群数（比如说 **n**），并把数据按照一定的规则分为 **n** 个群，这是该方法的第一步。接着按照一定的规则把 **n** 个群中最接近的群进行归并，当达到一定的标准时，这种归并停止，这就是该种方法最终确定的聚类群数（比如说 **m**），这是第二步。两步聚类法的一个显著优点是可以不指定聚类群数，它可以根据结构本身自动确定应该把数据分为多少群。

3. **Kohonen 网络聚类法**：是运用神经网络的方法对数据进行细分的数据挖掘方法。

## 客户细分实例

假设某电信运营商拥有以下数据：帐户明细综合帐，存放的是帐户级费用和 **SI** 级的费用，包括应收帐、已欠帐、调帐。时间为 2006 年 11 月到 2007 年 3 月。字段信息如下表所示：

1. 费用标识
2. 费用类型
3. 已销帐金额
4. 当前欠费
5. 帐户标识
6. 本地网标识
7. 表示费用产生时费用所属营业点
8. 转入总费用
9. 使用费原始费用
10. 总发生费
11. 总调帐费用
12. 原始新产生费用
13. 转出总费用
14. 补退费金额
15. 账目类型
16. 客户标识

费用类型说明表，包括以下字段：

1. 费用标识
2. 费用名称
3. ARPU状态
4. 费用类型

假设该电信运营商希望更深入地了解小灵通客户的情况，建立更为有效的市场营销战略来给小灵通客户推销其电信产品。

首先根据维度表费用标识字段，选择小灵通客户的信息，包括上网通信费、区间通话费、区内通话费、国内长途、国际长途、电话信息、来电显示、短信，根据费用说明表中的 ARPU 值以及费用类型字段区分上述几种业务情况。

我们将帐户综合明细帐，和费用类型说明中的数据根据客户标识进行合并，得到小灵通客户的费用和费用类型信息。数据流如下：

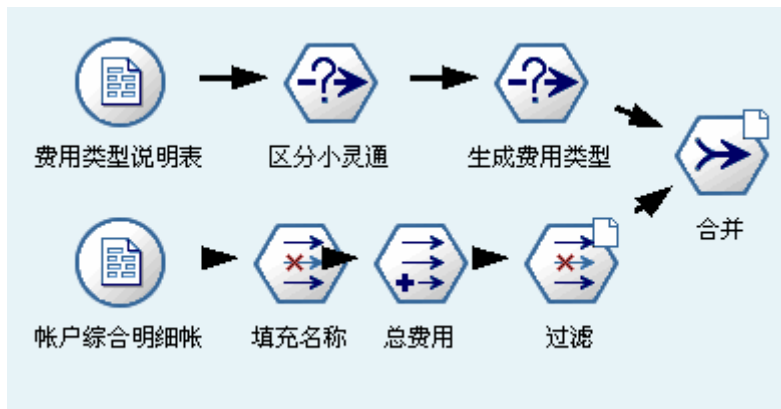


图4.2：数据理解数据流

重新结构化数据，并对每个客户每种类型的费用进行汇总，超级节点分别派生出本地话费、增值业务、国内长途、国际长途的数据，得到建模需要的数据：

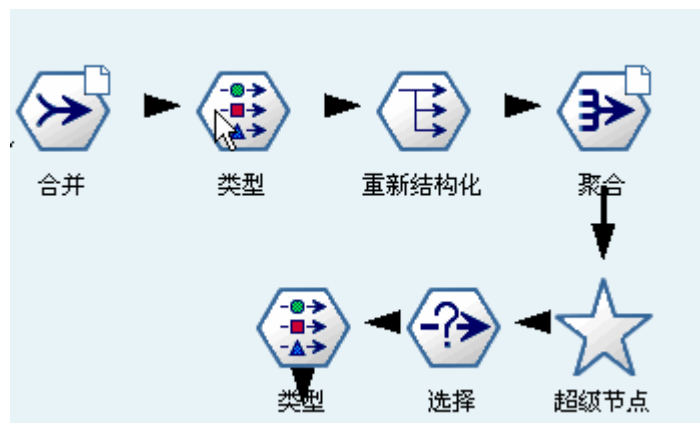


图 4.3：数据准备数据流

利用客户的交费记录，使用聚类算法，自动生成多个类别（如 5 档）。这样不仅对小灵通客户有初步的认识，也能够为后续分析，如进行客户金字塔分析建立初步的分析基础。

根据上述数据，对小灵通客户五种不同类型业务的消费情况进行聚类，找出具有相似消费倾向的客户。

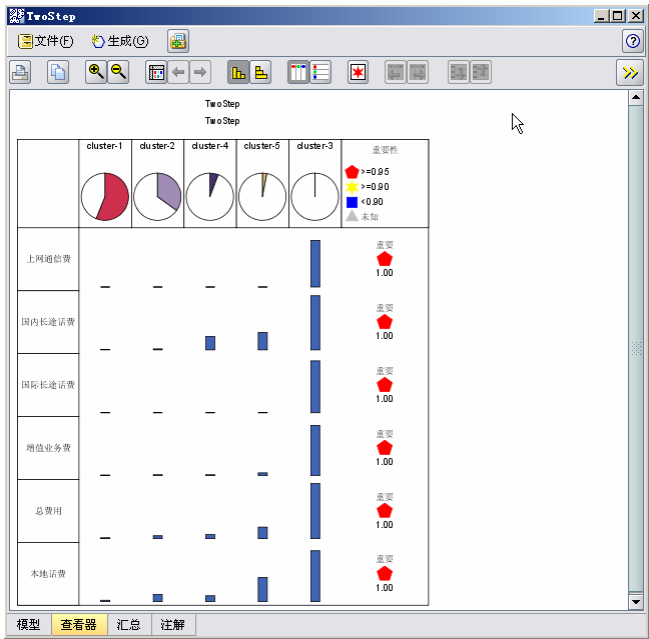


图 4.4：两步聚类结果

从分析的结果可以看出：Cluster-1 和 Cluster-2 中的客户基本上都是低价值的客户，Cluster-3 中的客户在每个类型的服务中消费都比较高，即高价值客户。为了更好的发现高价值客户的特征，我们对 Cluster-3 中的客户进一步聚类，结果如下：

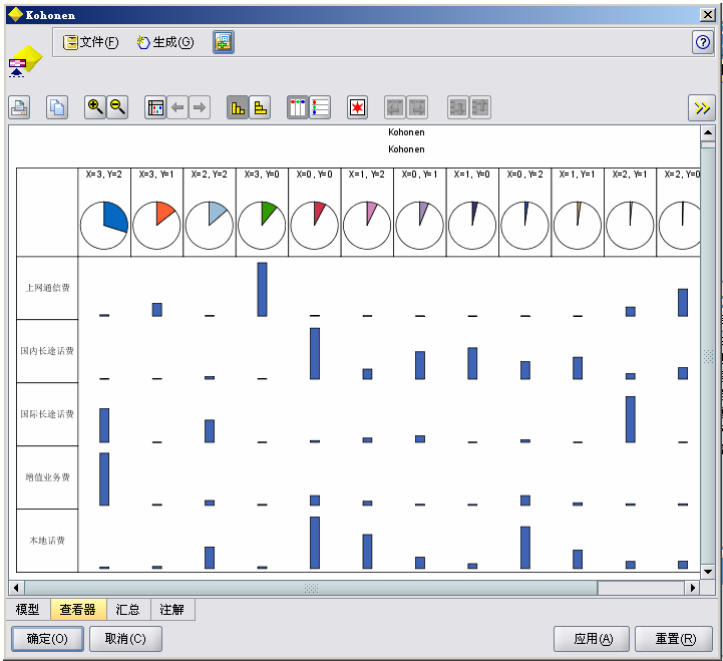


图4. 5：高价值客户聚类

从结果看出，X=0,Y=0 的客户，增值业务消费较高，X=0,Y=2 的客户，上网通信消费较高，而 X=3,Y=2 的客户使用小灵通的目的主要是打电话。

## 5.客户满意度

### 结构方程模型用于客户满意度测评中的应用

由于客户满意度研究对企业的重大指导作用,科学高效的研究方法和手段将成为满意度研究机构的核心竞争力,其研究成果将为提高整个社会的整体满意水平,指导企业更加合理地配置资源提供科学决策依据。而结构方程模型越来越被业界认可为“满意度研究的科学高效的研究方法和手段”,成为进行满意度研究必不可少的工具。目前,美国用户满意指数(ACSI)、瑞典用户满意指数、欧洲用户满意指数、中国用户满意指数(CCSI)等国家级用户满意度研究都是采用结构方程模型构建关系。

一个有效实用的客户满意度研究结构方程模型的构建,需要对客户需求和感知的深入研究,通过大量的前期工作,比如客户焦点小组访谈、客户需求分解、客户预调查、行业专家拜访、购买消费现场观察等多种手段,才能构建起一个基础模型。然后通过消费者试调查,采集到一定的数据后,对数据进行多种统计处理、分析和检验,根据相应的结果对模型进行必要的调整,然后才能应用到实际的客户满意度分析当中去。否则的话,随意构建的模型只能成为使得数字游戏显得高深莫测的工具。下图是运用结构方程模型来研究满意度模型的整个过程和大致思路:

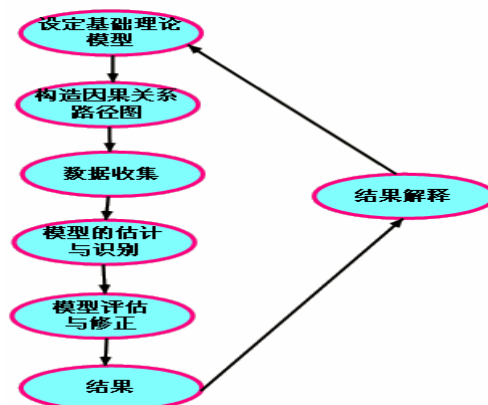


图 5.1: 利用结构方程模型进行满意度研究过程和思路

在满意度研究中常用的结构方程模型示例

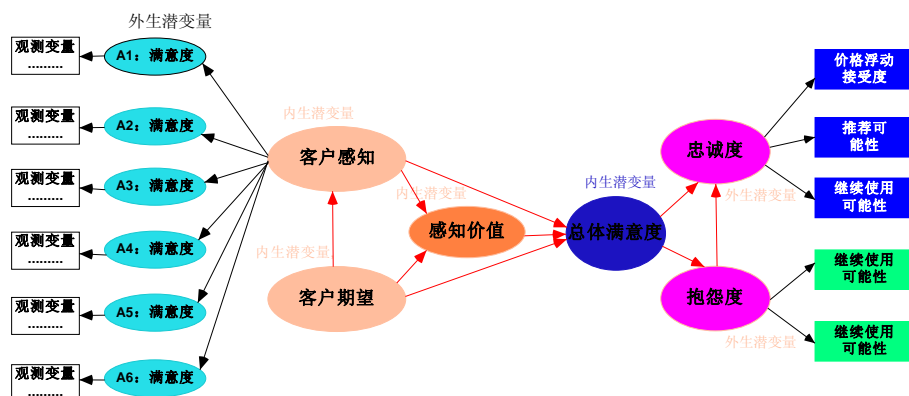


图 5.2: 满意度模型框架示例

## 满意度研究在电信行业中的应用

中国加入 WTO 梦已成真，各行各业的市场竞争更加激烈，特别是电信行业，将面临与日俱增的市场压力、更精明的竞争对手和更苛刻的消费者。

中国通信行业正从快速增长阶段迈入平稳成长阶段，服务质量已经成为企业的核心竞争力。因而电信运营商市场营销工作的重点不仅仅是吸引新用户，突出价格策略的重要性，而是要将如何维护老用户作为市场营销工作新的重点，从而服务策略的重要性逐渐突现。目前市场上的价格竞争将逐渐过渡到服务竞争。移动与联通已经形成了较为激烈的竞争格局，随着电信、网通等国内运营商及国外运营商进入移动通信市场，这种竞争愈演愈烈。电信运营商能否在如此严峻的经营环境下顺利发展，关键因素在于能否站在消费者的角度考虑产品和服务，消费者是否满意其提供的产品或服务。“客户满意”越来越成为众多电信运营商已经意识和正在努力提高的经营指标，成为运营商工作的重点。

下面借某电信运营商在某省的满意度研究案例，介绍如何应用结构方程模型进行满意度研究。

### 研究目标

该电信运营商期望通过客户满意度研究了解：不同品牌客户和集团客户对该公司的整体服务工作的满意度评价，以及不同品牌客户对本公司各商业流程环节上的服务感受及满意度水平，并了解不同品牌客户对本公司的忠诚度。同时判断当前业务、服务工作中存在的主要问题。重点围绕各个商业过程，有针对性地发现问题，从而改善服务短板。通过满意度指数，比较各地市分公司的服务工作差异，以便为省公司的考核提供依据。进行与竞争对手 CSI 测评的比较分析，确定该省电信运营商客户服务工作中有待改善的地方，并以绩优区域为标杆，不断提高和推动该省该电信运营商的用户满意度。预测今后的业务、服务竞争趋势，制定应对竞争的一系列策略和方案。

### 研究过程

#### 本研究中的结构方程的满意度模型

在本研究中，各方面满意度的研究（如不同品牌客户和集团客户对整体服务工作的满意度、不同品牌客户对各商业流程环节上的满意度等等）都是基于图 5.3 所示的满意度框架模型，只是在不同的满意度分析时具体的观测指标有些变动（如不同品牌相应的商业过程的指标）。

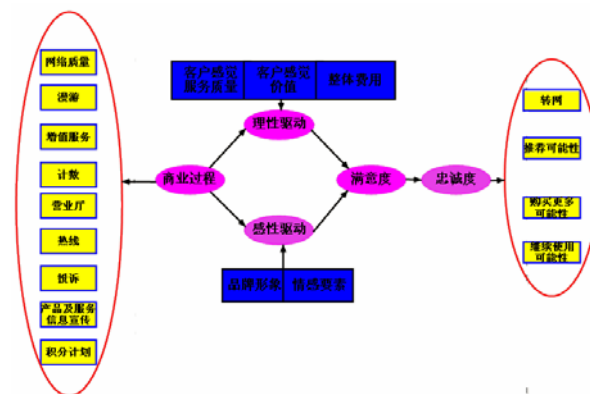


图 5.3: 某省电信运营商的满意度研究的框架模型满意度模型框架示例



## 满意度研究的结果分析

根据图 5.3 所示满意度理论模型及潜在变量对应的观测变量，分别建立不同品牌满意度模型。在不同品牌的客户满意度理论模型得到验证后，就可利用调查数据计算得出客户的总体满意程度及各分项满意程度。

## 结论

最终本次满意度研究获得了圆满成功，给电信运营商提出了有建设性和可执行的满意改进和提升客户忠诚度的改进计划。具体包括：

### 客户总体满意度

图 5.4 运营商及其竞争对手各品牌客户的总体满意度指数

- 图 5.4 给出了基于图 5.3 所示的满意度模型及调查结果得出的该运营商各品牌及其竞争对手的总体满意度指数。
- 图 5.5 给出了基于图 5.3 所示的满意度模型及调查结果得出的该运营商及其竞争对手各地区各品牌的总体满意度指数。

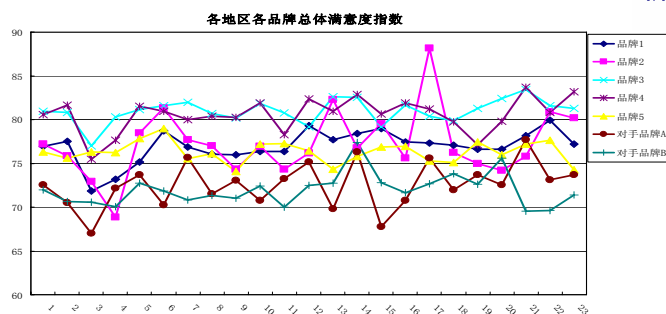


图 5.5 运营商及其竞争对手各地区各品牌客户的总体满意度指数

- 从图 5.4，可以看在该省电信运营商的各品牌及客户群中，品牌 1-4 客户的满意度处于高水平并比较接近，与去年相比，该电信运营商各品牌的服务水平比较稳定。只有品牌 5 客户满意度水平显得较低。
- 与竞争对手比较，该运营商继续保持较高的服务领先优势。品牌 3、4 的满意度明显高于竞争对手的品牌 A、B；客户忠诚度比竞争对手高 15 个左右百分点；但对手品牌 B 对品牌 2、小灵通对品牌 1 的离网风险比去年有较大增加。小灵通对品牌 1、2 客户的蚕食现象也显著增加。
- 从图 5.5 可以看出该省部分市区的客户总体满意度仍然较低，个别地区尤其明显。

### 各品牌主要商业过程满意度及服务改进策略

## ■ 各品牌主要商业过程总体满意度（以下以品牌 1、2 说明）

品牌 1:

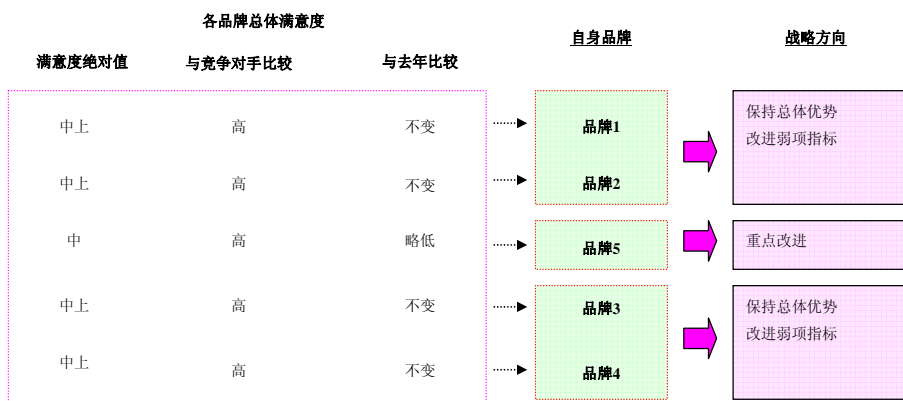
- ✓品牌1客户满意度较高的是网络质量(84)与缴费方便程度(83)
- ✓品牌1客户对梦网服务的满意度比较低, 得分为64.6分;
- ✓除梦网外, 品牌1客户对热线方面满意度也较低。

品牌 2:

- ✓品牌2满意度较高的是整体网络质量与缴费方便程度。
- ✓品牌2对梦网服务的满意度较低。
- ✓品牌2对特色服务满意度中等。

## 全省服务改进方向

根据研究结果, 我们为该运营商提出全省的服务改进建议, 如下:



## 各品牌主要商业过程的具体满意度及改进策略

研究中主要围绕网络服务、服务厅服务、热线服务、缴费充值服务、梦网服务、优惠活动、网站服务满意度这些主要商业过程的提出了满意度的具体改进策略, 如对于网络服务的满意度和具体改进策略, 有如下结果

### 网络服务满意度水平及改进策略

- ➡ 总体上, 客户网络方面总体满意度较高, 其中最高的是品牌 3、品牌 5 相对偏低;
- ➡ 与竞争对手比较, 网络服务总体满意度具有明显优势。
- ➡ 在通话质量方面明显高于竞争对手;
- ➡ 总体上在绝大部分场合的网络信号都处于中高水平, 但在山区的信号满意度较低;
- ➡ 对于不同场合拨打电话无法接通率, 总体上室内的无法接通率最高, 其中品牌 2 的室内无法接通率较高, 超过 50%; 品牌 3-5 的无法接通率相对较低。
- ➡ 竞争对手的无法接通率与品牌 2 较为接近, 对手品牌 A 的无法接通率大于对手品牌 B。
- ➡ 对于网内网间通话无法接通率方面, 自身各品牌客户打竞争对手品牌的客户手机时无法接通率较高, 品牌 2 尤其较高; 而竞争对手客户打市话或该运营

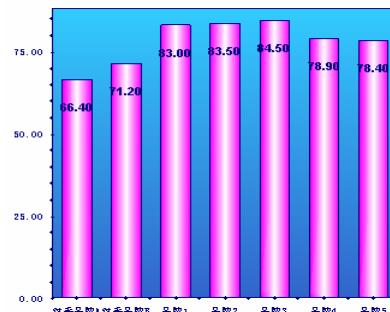


图 5.6 客户对网络方面总体满意度

商时无法接通率较高。

- ➡ 对于掉话率，品牌 1、2 的掉话率高于品牌 3-5（品牌 2 更高），室内的掉话率高于室外，室内掉话率较高的有品牌 2(51%)，竞争对手的掉话率也达 40% 以上。
- ➡ 对于短信故障率方面，短信发送不成功发生率较高（品牌 5 与品牌 1 最为明显），其次是短信发送成功但对方较长时间才收到；品牌 5 在短信故障率方面发生率最高。
- ➡ 各品牌客户对网络质量不满意的地方主要是信号差，其次是网络覆盖不广；
- ➡ 品牌 2 不满的地方还表现在通话不稳定/易断线，品牌 1 则是难接通/接通率低，品牌 5 则是信息接收不好/慢；竞争对手在各方面的不满的比例更高。

## 忠诚度

### 忠诚度水平与性质

#### 各品牌忠诚度

- ❖ 自身各品牌忠诚度都较高，各品牌中忠诚度最高的是品牌 2 的客户（89.8%），最低的是品牌 5 的客户（83.7%）；
- ❖ 总体上，自身品牌客户的忠诚度高于竞争对手的品牌；竞争对手两品牌的忠诚度都为 75%。

#### 忠诚度细分

模型说明：根据忠诚度和满意度的不同，我们将客户共分成四大类型：安全羊、劝服者、异动者、流动者。

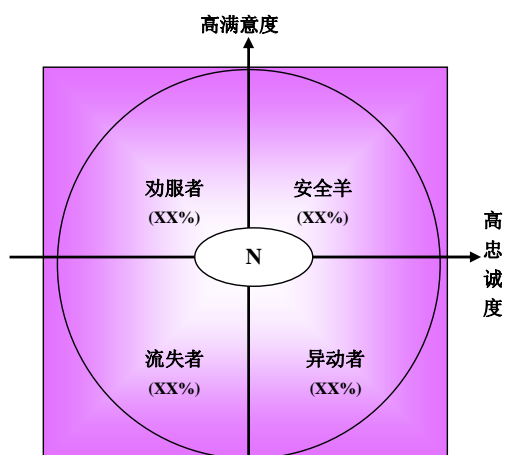


图 5.7 忠诚度细分：模型说明

#### 1) 安全羊（高满意度高忠诚度）：

当前品牌的重要客户群。他们对该品牌产生了服务固化的态度。

#### 2) 劝服者（高满意度低忠诚度）：

有一定现存风险的客户，对服务有较高的满意程度，但没有强化他们的忠诚度，可能与他们喜欢尝新事物的兴趣、竞争对手的吸引等有关。

#### 3) 异动者（低满意度高忠诚度）：

有一定潜在风险的客户，对服务不满意，但仍表示将继续使用服务，可能与目前市场环境如运营商的选择数量有限有关。如果市场环境没有改变，这类客户不会流失。

#### 4) 流失者（低满意度低忠诚度）：

这类客户属于高现存风险的客户，他们对所属运营商的服务不满意，并表示不想继续使用服务。

- ❖ 自身品牌的安全羊比例明显高于竞争对手；自身品牌中，品牌 5 客户的流失者与异动者较高。
- ❖ 与去年相比，今年品牌 3-5 的客户的异动者（无奈的忠诚）比例有明显增加，一旦市场出现新的运营商或网络品牌，这部分不稳定客户将释放较大的离网风险。
- ❖ 不同忠诚度类型客户的满意度：  
将不同忠诚度的四种客户进行分析发现：
  - 各品牌安全羊的满意度都达 85 分以上，满意度较高的是品牌 3、4；

- 各品牌劝服者的满意度都达 84 分以上；
- 各品牌异动者与流失者的满意度都低于 60 分；大客户的流失者满意度更低。

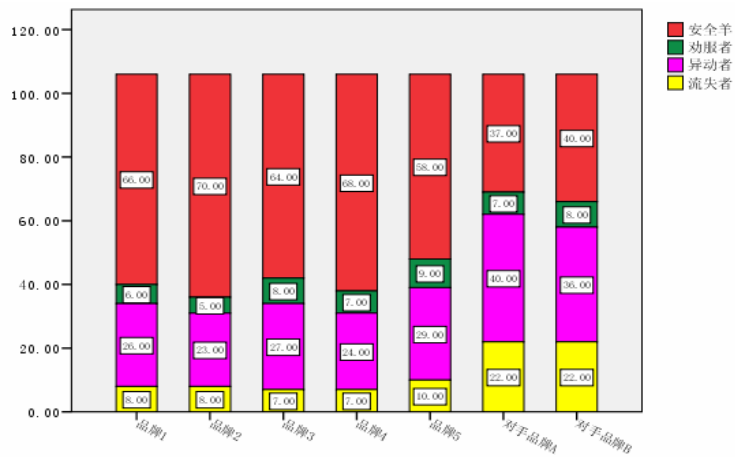


图 5.8 各品牌忠诚度类型

满意度得分	安全羊	劝服者	异动者	流失者
品牌2	87.4	86.1	57.9	50.3
品牌1	86.7	84.7	57.6	54.6
品牌5	85.1	84.3	57.5	54.5
品牌4	88.7	84.4	58.5	55.6
品牌3	88.4	85.5	58.3	57.1

表 5.1：不同忠诚度类型客户的满意度

### 离网与蚕食风险

#### ❖ 短期离网风险

- 品牌 1、2 客户的主要流失方向是竞争对手与小灵通；品牌 3-5 客户的主要流失方向为内部流动，流向自身的其它品牌；竞争对手客户的主要流失方向是品牌 1、4。
- 品牌 2，转向竞争对手的比例较高；品牌 1，转向对手品牌 A 与自身品牌 4 的比例较高；品牌 5，转向品牌 4 的比例最高；品牌 4，主要转向自身的其它品牌；品牌 3，转向品牌 4 的比例最高。

#### ❖ 与去年比较离网风险比较

与去年比较，潜在流失客户有变化，转向品牌 1 的客户在减少，转向品牌 4 的客户在增加，具体如下：

- 品牌 2：转向对手品牌 2 的比例上升；

- 品牌 1：转向小灵通的比例上升；
  - 品牌 5：转向品牌 1 的比例上升；
  - 品牌 4：转向品牌 1 的下降，转向品牌 3 比例的上升；
  - 品牌 3：转向品牌 1 的下降，转向品牌 4 比例的上升；
  - 对手品牌：转向品牌 1 的下降，转向品牌 4 比例的上升
- ❖ 蚕食风险
- 自身及竞争对手的品牌都面临被小灵通蚕食的危险，品牌 1 受小灵通品牌蚕食最大，品牌 3-5 客户受小灵通蚕食较小；
  - 品牌 1、2 主要被竞争品牌蚕食，品牌 3-5 主要表现为“网内”蚕食。
- ❖ 与去年比较被蚕食的风险比较
- 自身及竞争对手的品牌被小灵通蚕食的比例都在上升，被小灵通蚕食最为严重的是品牌 1、1；
  - 自身品牌中，品牌 4、5 被品牌 3 蚕食的比例上升；品牌 1、5 被品牌 4 蚕食的比例上升。

## 提升策略

基于以上的忠诚度分析结果，研究中对各个品牌分别针对不同的忠诚度类型的客户提出了具体的服务改进策略，以提高忠诚度。下面以品牌 1 为例说明。

品牌 1

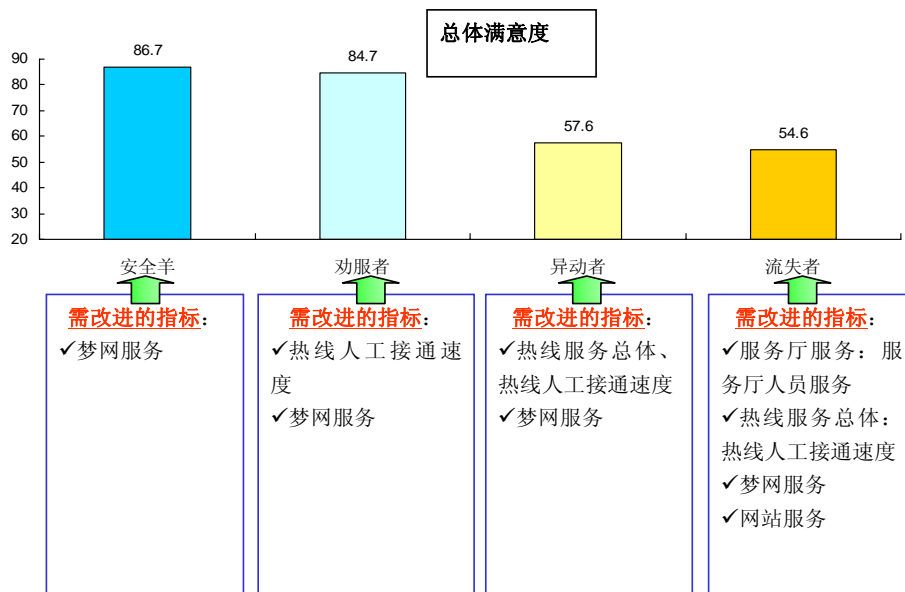


图 5.9 品牌 1 忠诚度提升策略

## 6.CRISP-DM 简介

在 1996 年，当时数据挖掘市场是年轻而不成熟的，但是这个市场显示了具有爆炸式增长的潜力。三个在这方面经验丰富的公司 DaimlerChrysler、SPSS、NCR 发起建立一个社团，目的建立数据挖掘方法和过程的标准。在获得了 EC（European Commission）的资助后，他们开始实现他们的目标。为了征集业界广泛的意见共享知识，他们创建了 CRISP-DM Special Interest Group（简称为 SIG）。

大概在 1999 年，SIG（CRISP-DM Special Interest Group）组织开发并提炼出 CRISP-DM（Cross-Industry Standard Process for Data Mining），同时在 Mercedes-Benz 和 OHRA（保险领域）企业进行了大规模数据挖掘项目的实际试用。SIG 还将 CRISP-DM 和商业数据挖掘工具集成起来。SIG 组织目前在伦敦、纽约、布鲁塞尔已经发展到 200 多个成员。SIG 组织已经发布了 CRISP-DM Version 1.0 Process Guide and User Manual 的免费使用电子版。当前 CRISP-DM 提供了一个数据挖掘生命周期的全面评述。它包括项目的相应周期，其中各自任务和这些任务之间的关系。当前这个数据挖掘的程序模型，为数据挖掘项目的生命周期提供了一个综合的描绘。它包括了一个数据挖掘项目所要经历的各个阶段，各阶段的任务以及这些任务之间的相互关系。从描绘的层面来看，是不可能鉴别出所有这些任务之间的关系的。但本质上看，这些任务之间是否存在关系，取决于使用者的目的，背景及其利益所在，与此同时，更重的还在于数据。

数据挖掘项目的生命周期由六个阶段组成。如图展示了这一数据挖掘过程的各个阶段，这些阶段之间的顺序并不固定，在不同阶段之间来回反复往往是非常有必要的。究竟下一步要执行哪个阶段或者哪一个特定的任务，都取决于每一个阶段的结果。图中的箭头表明了阶段之间最重要和最频繁的依赖关系。图中最外层的这个循环表明了数据挖掘本身的循环性质。经过一个具体的数据挖掘项目得到了某项解决措施或方法并加以展开，并不代表数据挖掘本身已经结束。从这一数据挖掘过程以及解决措施展开的过程中所吸取的经验、教训，又引发了新的，通常是更加焦点的商业问题。接下来的数据挖掘过程将会从过去的项目经验中获利。在接下来的内容中，我们将简要的勾勒一下每个阶段的轮廓：

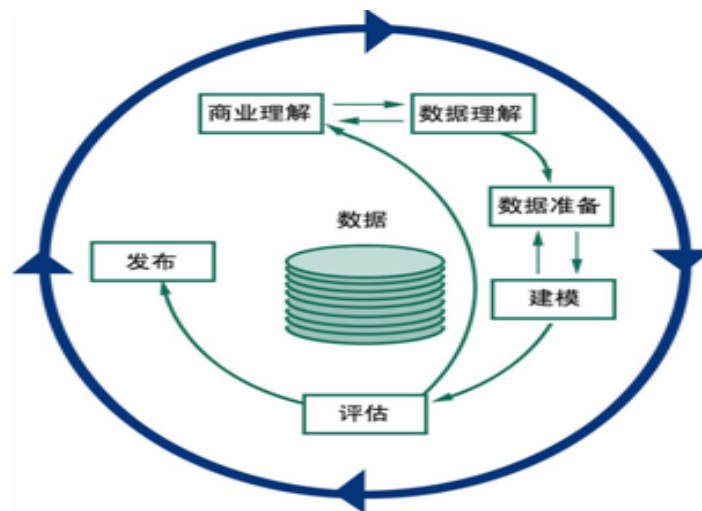


图 6.1: 数据挖掘方法论流程示意图



## 商业理解

这一初始阶段主要集中在对项目目标的理解,以及从商业角度考虑,对客户需求的理解。进而把这些理解转化为一个数据挖掘的定义和为了达到目标的初步方案。

## 数据理解

数据理解阶段开始于数据的收集工作。接下来就是熟悉数据的工作,具体如:检测数据的质量,对数据有初步的理解,探测数据中比较有趣的数据子集,进而形成对潜在信息的假设。

## 数据准备

数据准备阶段涵盖了从原始粗糙数据中构建最终数据集(将作为建模工具的分析对象)的全部工作。数据准备工作有可能被实施多次,而且其实施顺序并不是预先规定好的。这一阶段的任务主要包括:制表,记录,数据变量的选择和转换,以及为适应建模工具而进行的数据清理等等。

## 建模

在这一阶段,各种各样的建模方法将被加以选择和使用,其参数将被校准为最为理想的值。比较典型的是,对于同一个数据挖掘的问题类型,可以有多种方法选择使用。一些建模方法对数据的形式有具体的要求,因此,在这一阶段,重新回到数据准备阶段执行某些任务有时是非常必要的。

## 评估

从数据分析的角度考虑,在这一阶段中,您已经建立了一个或多个高质量的模型。但在进行最终的模型部署之前,更加彻底的评估模型,回顾在构建模型过程中所执行的每一个步骤,是非常重要的,这样可以确保这些模型是否达到了企业的目标。一个关键的评价指标就是看,是否仍然有一些重要的企业问题还没有被充分地加以注意和考虑。在这一阶段结束之时,有关数据挖掘结果的使用应达成一致的决策。

## 部署

模型的创建并不是项目的最终目的。尽管建模是为了增加更多有关于数据的信息,但这些信息仍然需要以一种客户能够使用的方式被组织和呈现。这经常涉及到一个组织在处理某些决策过程中,如在决定有关网页的实时人员或者营销数据库的重复得分时,拥用一个能够即时更新的模型。然而,根据需求的不同,部署阶段可以是仅仅像写一份报告那样简单,也可以像在企业中进行可重复的数据挖掘程序那样复杂。在许多案例中,往往是客户而不是数据分析师来执行部署阶段。然而,尽管数据分析师不需要处理部署阶段的工作,对于客户而言,预先了解需要执行的活动从而正确的使用已构建的模型是非常重要的。

## 7.数据挖掘经验谈

### 采用 CRISP-DM 方法论

采用 CRISP-DM 方法论作为数据挖掘的指导能帮助确保获得成功的商业结果。在现实中对于一个数据挖掘项目来说,最初设定的商业目标很容易淹没在复杂数据挖掘技术和海量数据中,所以以一个已经被验证方法论为指导是很关键的。

### 以终为始

为了能在项目终结时得到期望的 ROI (Return On Investment, 投资回报率),你应该在项目启动前已经确定了如何评估最终的结果的标准(例如:使用什么样的商业考核指标,它们是被如何计算或派生的)。例如你是不是想在 20% 的客户中找寻潜在流失者? 基于客户保留计划的消费和营销反应程度,你如何将这信息转换成商业收入增长期望值? 或者你知不知道如果确定出额外的十件欺诈信息能节约多少开支?

### 设定期望值

确保项目投资者明白数据挖掘不是解决商业问题的魔术棒。数据挖掘是借助计算机技术辅助解决商业问题的一种方法。就像任何商业问题,投资者需要首先提出可解决的问题,然后找寻方案。例如你计划为公司市场部做客户细分,那么应该与市场部的同事一起明确什么样的结果是最终希望得到的(例如:“我们使用产品信息和人口统计数据,所以希望得到基于客户的收入、年龄等信息的细分,这样能显示不同层次客户对产品的喜好”)

### 限定最初项目范围

以现实可行的目标和日程表为开始,当你获得成功后,再转向更复杂的项目。例如与其试图立刻提高新客户的获取值,还不是集中精力在小的更实际的目标如对某一区域进行交叉销售,客户保留项目。

### 确保团队合作

数据挖掘项目是一个团队工作。数据挖掘需要商业使用者理解实际问题 and 数据,也需要数据分析家提供分析解决方案,以及数据库管理者提供权限。例如,你可能在项目中需要数据挖掘专家、数据库专家和市场经理。因为他们来自不同的部门,可能在数据挖掘项目合作中会出现问题,所以找到可行的合作方式是很重要的。

### 避免陷入数据垃圾

在项目进行中,始终明确需解决的商业问题,确保项目结果的最终完成。如果你只是在没有项目计划的情况下简单的开始分析一堆数据,你将会很容易迷失在数据里而且浪费时间。不要让项目被大量数据单纯驱动,集中精力在商业目标上。你可能不需要使用系统中的所有数据,仅仅使用和项目相关的数据就可以了。你甚至可能会发现现有的数据不能足以解决现实的商业问题。即使海量数据也不能保证你就拥有准确的用于建模的数据。例如,使用最新的信息进行预测客户行为往往比用大量的历史数据准确。

## 8. 数据挖掘部署策略

数据挖掘的结果发布可以很简单，例如只是生成一个规则集，对具体某个商业问题给出一个参考建议；也可能很复杂，如需要实时嵌入到客户的决策支持系统，为决策者提供前瞻性决定提供依据。以下阐述四种优化策略帮助部署高级分析结果，以及为获得最大投资回报设定的预测分析解决方案。这些策略是通过概括现实中使用 Clementine 数据挖掘平台的众多部署案例得到的，具有普遍应用性。

- **快速更新批处理方式：**使用快速高效的批处理功能部署数据挖掘，为数据简单快速的打分
- **海量数据批处理方式：**策略性应用代码部署，注意代码开销平衡，集中在为海量数据高速打分
- **实时封装方式：**将数据挖掘部署封装应用并将集成风险最小化，应用在用户定制的高速、实时为数据打分上
- **实时定制方式：**将数据挖掘部署到为客户量身定做的应用产品上，在企业组织结构下实现不同功能的实时打分或者不能以实时封装方式部署的特殊商业目标

### 策略 1 --快速更新批处理方式

快速更新批处理打分是应用最广泛的方式之一，适应于不需要实时响应的系统。例如，许多公司使用这种方式对客户数据库保持更新，通过客户关系管理应用为决策者提供最新预测分析。这种方式以及相关部署应用为与各式操作系统环境提供灵活集成。

Clementine's Batch 和 Clementine Solution Publisher 帮助你高效部署数据挖掘结果和迅速更新数据。在不需要 Clementine 客户端界面情况下，后台部署 Clementine 数据挖掘流。Clementine 批处理模式在命令行执行，而 Clementine Solution Publisher 是一个灵活的打分组件，可以嵌入到应用中。这种部署方式可以在数据库内按计划执行，例如一个预测应用需要每月、每周、每天甚至每小时使用最新数据。因为这些运行方式都是执行整个数据流，更新打分时只需重新运行 Clementine 挖掘流。而且这种使用 Clementine Batch 或 Clementine Solution Publisher 的运行方式可以完全被 SPSS Predictive Enterprise Service 中的 Predictive Enterprise Manager Module 自动使用。

不象其他的数据挖掘解决方案，Clementine 能够部署整个数据挖掘过程，包括关键的数据准备，建模以及应用模型打分任务，可以在 IBM® DB2®，Oracle® Database, and Microsoft® SQL Server™内使用。这些 Clementine 部署操作因为使用三层体系架构，利用数据库的检索、优化和数据库内挖掘功能，确保高效数据库内挖掘。通常，大部分的公司只使用 Clementine 客户端运行所需要的批处理打分，而使用 Clementine Batch 或 Clementine Solution Publisher 进行周期性打分。Clementine 客户端使用数据流描述语言（Stream Description Language）将需要执行的数据挖掘任务发送到 Clementine 服务器端。Clementine 服务器分析挖掘任务决定那些可以在数据库端执行，将数据转移量降至最低。在数据库内运行完这些操作后，将剩余的和已整合的数据传递给 Clementine 服务器。

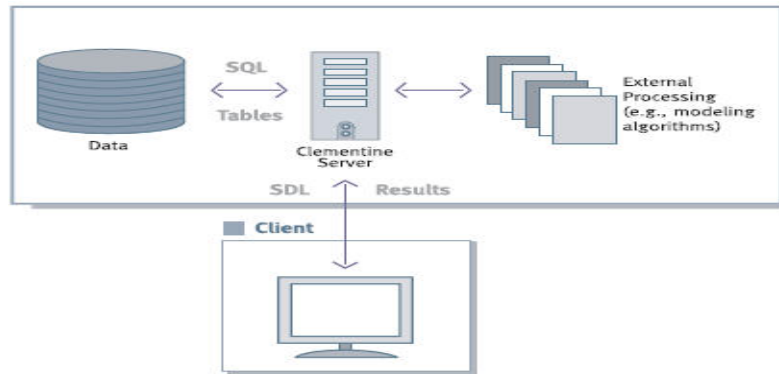


图 8.1: Clementine 三层体系架构

## 策略 2 --海量数据批处理方式

许多 Clementine 实施应用将 Clementine 挖掘流部署在一个可解释的打分引擎上，这样就需要评估数据挖掘的实际时间以及所需的最小代码开发量。数据挖掘的过程可以被描述在一个可以被打分组件或应用程序所解析的文件。而对于海量数据批量打分，其应用在至少部分数据挖掘过程已经转化成某种编程语言，并且编译成计算机可识别代码的基础上。从定义上讲，已编译模式比可解释模式要快。

因为在海量数据打分环境下运行瓶颈经常发生在数据准备阶段而不是对新数据打分阶段，所以 SPSS 高速打分操作经常使用混合模式。Clementine 使用 PMML（Predictive Model Markup Language）将预测模型输出给高速打分引擎，PMML 是一种描述数据挖掘模型的行业标准的 XML 标识性语言（PMML 标准是一个用于模型交换的中间方式，这个标准是由一个由主流数据挖掘和数据库厂商组建的独立数据挖掘组织制定的）。PMML 模型部署是在已将数据准备过程代码化后，将模型描述储存在 PMML 文件里，然后根据打分数据的要求将整个过程编译。SmartScore® 就是 SPSS 基于以上应用的 PMML 打分软件开发包（SDK）。Clementine PMML 模型也可以实施在 IBM DB2 数据库内挖掘，不需要将数据 DB2 数据库中迁入牵出，提高了性能。

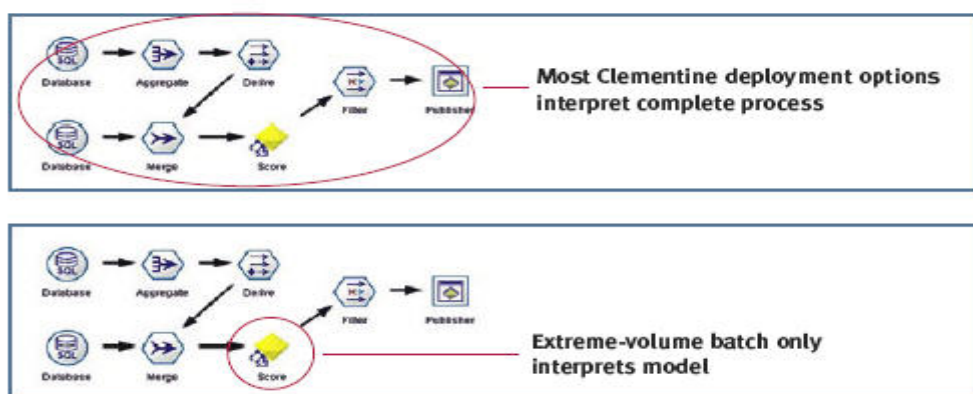


图 8.2: 两种典型部署方式

## 策略 3--实时封装方式

预测型分析应用在特定商业目标上时，通常需要实时、高速的对大数据集打分，SPSS 预测分析应用例如 PredictiveCallCenter™ and PredictiveWebSite™ 就是为已存在的客户交互

功能软件和CRM系统而设计的实时高速产品。PredictiveCallCenter™与呼叫中心系统集成，提供如提升销售、交叉销售、客户保留等的即时推荐和决定支持。采用被已证明可行性商业规则结合实时预测分析技术，PredictiveCallCenter™自动提供推荐内容并结合销售意见和其他代理需要知道的信息将其推荐内容显示在代办人员的屏幕上。

Clementine 可以将 Clementine 预测模型直接发布到 SPSS 预测分析应用产品上如 PredictiveCallCenter。这样就允许将你自己训练的复杂模型使用 PredictiveCallCenter 实施并精炼成实时推荐应用发布出来。模型可以被开发成于多种类数据如网页、文本以及有关态度的数据相结合，并且具有特定商业目标的应用如提高交叉销售。多种预测模型可以被合并，例如，将交叉销售和欺诈检测模型相结合，可以确保销售人员不用将精力放在具有明显欺诈风险的客户身上。

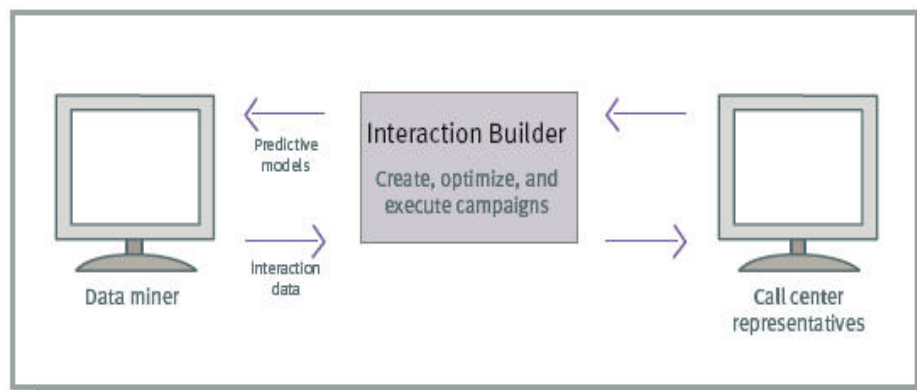


图 8.3: 呼叫中心推荐信息产生流程

#### 策略 4--实时定制方式

对于在企业决策支持系统中决策优化目标需要不同功能的实时打分应用，或者针对某特定的商业实际需求，客户化的界面是十分关键和必要的。Cleo™ 和SPSS Predictive Analytic Framework™ 就是为适应广泛商业应用需求而设计的客户化实时打分产品。

Cleo是创建基于网页打分应用的软件平台，它能简单快速的创建客户化网络应用。企业级用户使用Cleo，基于网络的数据挖掘实施工具，给予同一公司的多用户使用Clementine模型和根据特定需求的数据打分的权限。采用Clementine创建的预测模型，通过Cleo d部署向导就可以很轻松的在线进行模型发布。在Cleo server能即时产生网络应用，当决策者需要借助Clementine生成的预测流程支持他们的决策时，通过网络应用很轻松的实时为数据打分。不象其他一些基于网络应用的分析工具需要安装桌面软件或插件，Cleo应用是真正的瘦客户端，所有使用者只需使用网页浏览器就可以获得权限。



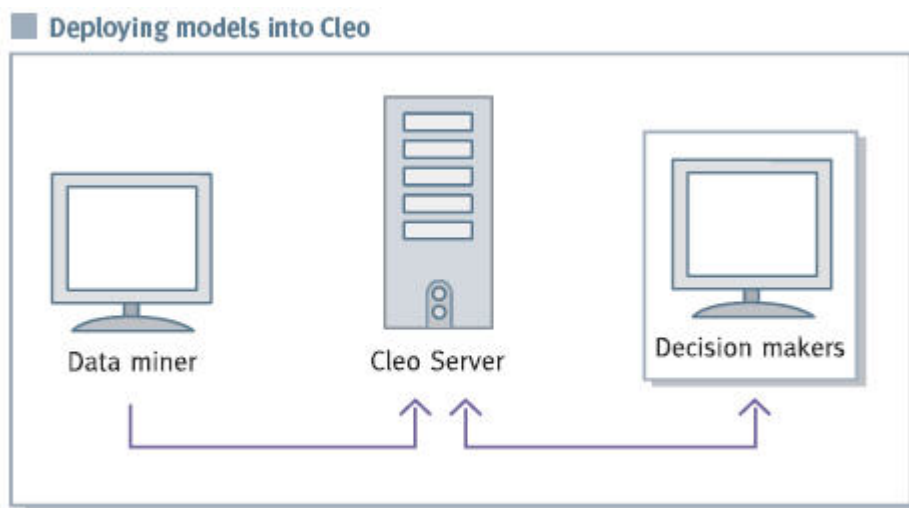


图 8.4: Cleo 实时打分

当企业需要更多高级功能，特别是通过与SPSS系统集成开发出更复杂的网络应用时，通常会使用Predictive Analytic Framework. 这个平台和Cleo很类似，是一个可供多种客户使用的，网络分布的瘦客户端打分环境。它还加入了一些额外的功能如为了方便商业用户更新模型的简单易用界面，监控预测分析性能而自动生成的增益曲线等等。通过Predictive Analytic Framework向导，模型可以很轻松的发布到网络平台上。

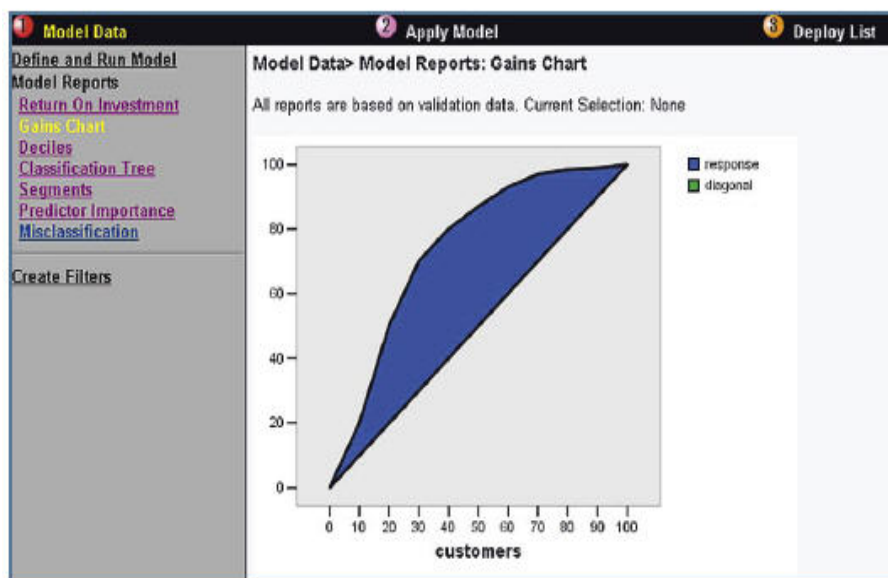


图 8.5: SPSS 定制开发应用示例



## 9. 成功案例

### 国外成功案例

#### Southwestern Bell（西南贝尔）

易于使用和强大的统计分析功能使西南贝尔公司的地区经理和居民质量咨询师 Mike Gallagher 对 SPSS for Windows 爱不释手。在 1993 年刚开始接触分析软件的时候，Gallagher 使用的是另一个非 Windows 界面的产品，但很快他听说了 SPSS。如今，Gallagher 和他的同事们都在使用 SPSS 来为高级公司决策者作数据分析工作。

**“SPSS 使我可以专注于统计分析，而不用在了解软件上拖延时间。”**

“SPSS 使我可以专注于统计分析，而不必在了解软件的使用上拖延时间。”

Gallagher 几乎每天都使用 SPSS 来分析可能会对位于五个州的近一万名员工产生影响的问题。经过数据分析，Gallagher 提出了各种可以减低费用、增加总体客户满意度、提升西南贝尔市场地位的工作改进方案。

一个例子是有关“重复维修”的问题：维修人员在第一次维修后十天内必须再次进行维修。Gallagher 使用 SPSS 做了两个步骤的分析。首先，他分析了以前在西南贝尔的大型机系统收集和保存的大约 85 万条记录，这使他基本了解了哪些因素对重复维修有最大影响。然后，他用交叉表和回归技术对大约 12000 条记录进行了分析，以确定在维修过程中需要改进的地方。最后，重复维修流程的改进显著地减少了成本，增加了客户满意度。西南贝尔每年减少了大约 15000 次维修派遣。

Gallagher 在讲到他们成功的关键时说

“关键在于能够快速灵活地处理大量数据。从这方面来说，SPSS 正是我需要的、能告诉我要在什么地方集中注意力的工具。”

#### CallCounter

##### 公司简介：

CallCounter, Inc., 是英格兰一家有近几年成立的新公司，主要业务是为电信公司高效地收集和分析用户呼叫数据，从而为电信公司增加收入。

##### 面临问题：

要为电信公司采集准确的呼叫数据，并识别出收入损失的地方并进行趋势预测，CallCounter 必须做到：

- 每月对数百万消费者的呼叫数据进行分析，避免数据丢失
- 区分电话呼叫类别，避免少计费
- 实施市场研究，帮助电信公司更好地了解他们的客户

## 解决方案:

CallCounter 用 SPSS 快速、方便地对呼叫数据进行分析, 发现收入损失的地方, 并提出解决建议。为其客户节约了时间和费用。

## 结果:

帮助客户减少收入损失 每年为客户节约 15 万美元以上 对任何电信公司来说, 保证收入都是最基本的。由于电信公司的收入来源越来越复杂, 保证收入的问题也变得更加严峻。不幸的是, 很多主管人员忽略了它, 因为他们认为收入损失并不重要, 只不过是简单的业务成本而已。但是, 这方面的专家 Deloitte 和 Touche 最新的研究证明: 收入损失超过了业务成本。根据他的研究, 电信公司由于错误的呼叫数据一般损失 11% 的收入。这些损失是由于不正确地采集和标记数据、错误的账单和糟糕的订单录入造成的。

Charlie Seymour 在 1996 年成立 CallCounter 的动机就是防止收入减少。他希望同时为 Cellular 和 Land-line 电话公司解决他们关心的保证收入和数据分析问题。Seymour 在电信行业已经工作 25 年, 他的主要工作是开发账单系统。他相信天衣无缝的数据收集系统是不存在的, 而且没有明确的方法让电信公司分析他们的呼叫数据来回答严峻的问题, 如“所有的呼叫都记账了吗?”, “客户正在打国内长途或本地电话吗?”。

他的方案是: 帮助电信公司更准确地分析数据, 使他们了解收入是怎么减少的, 如何最终减少损失。CallCounter 依赖 SPSS 进行这些分析。Seymour 用 SPSS 向电信公司提供可靠的分析数据, 简化并加速了他们的市场研究和数据查询过程。更重要的是, 帮助这些电信公司减少了收入损失。CallCounter 如何做到这些呢? 根据 Seymour 的说法, CallCounter 是惟一个直接从交换机一拨号音的起源地收集数据的公司。一般来说 CallCounter 首先分析电话公司一个月的数据, 在月底用 SPSS 的分析结果协调电话公司的账单。为了让电话公司图形化地对照比较, CallCounter 用易于理解的饼状图来表示分析结果。

## 帮助客户减少收入损失

即便是只有几个员工的小型电话公司每月也可能有两百万次电话呼叫。如此大量的数据可能会使电话公司错误地分析数据, 导致收入损失。“如果在月底不能对数据进行分析, 他们就不能向长途电话公司和用户提供正确的账单, 从而引起收入损失。电话公司无法确定账单是对还是错。更糟糕的是, 很多电话公司不相信他们有这种问题。幸运的是, SPSS 帮助这些

公司有效地分析数据, 避免了收入损失。” CallCounter 市场部经理 Andrew Desjardins 说。位于缅因州的 Saco River 电话公司是 CallCounter 早期的客户之一。为帮助 Saco 避免收入损失, CallCounter 赞助了它的数据收集软件。这个软件会在异常情况下触发警告, 如设定时间内异常的低音量呼叫。CallCounter 的软件使用 25 个报警器来分析和监控收集的实时数据。通过纠正异常报警, CallCounter 保证了进入记账系统的数据的准确性, 从而避免了收入损失。仅在第一年, CallCounter 就发现了 5 个呼叫数据处理错误, 为用户避免了 10 万元的潜在收入损失。

**“SPSS 使我们只需较少的工作量  
就能为电话公司增加收入。”**

**— Charlie Seymour  
CallCounter 总裁**

“SPSS 使我们只需较少的工作量就能为电话公司增加收入。” Seymour 说。Land-line 电话公司向运营商支付账单，如 Sprint、AT&T 和 MCI，因为使用了他们的当地设备。这是一个大额的月度账单，轻易就达到地方电话公司收入的一半。

“市话通讯公司报告的市内和国内长途电话数量，决定了需要向运营商支付的费用。” Seymour 说，“有些电话呼叫难以确定是从哪里打出的。这种情况下，你要么听信运营商的，要么进行检查数据。”

大多数电话公司非常希望量化电话呼叫，但很多公司没有这方面的能力。CallCounter 可以用 SPSS 分析这些呼叫并提供准确的结果。根据 Seymour 的解释，通过更认真准确地分析市内和国内长途电话呼叫，市话通讯公司不仅避免了收入损失，还节约了 10% 的接入成本。

### 每年为客户节约 15 万美元以上

每一个电话公司都要求进行数据分析。这种要求同时来自公司外部和内部。如，市场部需要知道什么样的方案才能获得和保持满意的客户，工程部门需要分析一天内不同时间的呼叫通讯量，以准确地进行需求预测，管理部门需要进行数据分析以确保和运营商签订的合同最大限度地保护了公司利益。Seymour 发现大多数公司在电话呼叫数据的分析上遇到过问题。

“在我们的帮助下，仅仅通过准确的数据采集、标定和分类，每年就轻易地为小型电话公司节约了 15 万美元。”

## Rural Cellular Corporation

### 背景介绍

Rural Cellular Corporation (RCC) 为美国中西部、东北、西北和南部地区提供无线通讯服务。RCC 服务于邻近大都市的郊区或旅游业、农业和小规模商业盛行的高速公路附近。自从 1990 年以来，RCC 发现在乡村通讯市场存在的机会并逐渐发展成为主要的区域性无线通讯提供商。通过提出创新性的解决方案，与运营效果好的公司进行合作和提供优异服务的承诺，RCC 占领了市场。

### 面临的问题

为了保持公司业务和经济效益的持续增长，RCC 必须提出独特的解决方案以最大程度满足客户的需求。如果有一个较好的市场研究方法，能够明确客户需求，将有利于做出决策。

### 解决方案

RCC 公司使用了 SPSS 统计分析软件来进行它的市场研究工作，其中包括客户满意度分析、品牌研究和其它研究以定位市场和特征等。通过连续的收集和分析市场信息，RCC 能够较好的决定产品、改善品牌及分配资源。

### 结果

使用 SPSS 产品后，RCC 公司已经体会到从大量数据中得到支持商业决策信息的重要性。“我们不能想象如果没有统计数据的支持该如何做生意，” RCC 公司市场研究部项目经理 Noel Roos 说道，“一些公司的决策是来自于领导核心的直观判断，这是非常危险的，因为人总是趋向于相信你愿意相信的事情。令人惊异的是，市场数据能改变人的思想。”

## 较好的产品配置和服务方案

大多数 RCC 的研究集中在改善产品特性上。“我们做了大量的研究以理解如何改善产品和服务或发展哪种类型的新的服务，” Roos 说，“这实际上是告诉我们客户想要什么和我们需要用什么来满足他们的需求。”

例如，几年以前，公司的一个竞争对手提出了 Single-rate Plan，许多公司也推出了相似的方案，而这些方案的实施都是非常昂贵的。RCC 则采用 SPSS 估计了客户的需求。

“我们需要知道我们的客户需要什么而不是盲从于市场，” Roos 说，“根据对竞争对手的方案进行研究发现，Single-rate Plan 是非常昂贵的。同时，我们也看到方案的某些特性是客户确实需要的，例如无漫游费和更大的覆盖面积。”

“基于市场数据的决策对底线有直接的影响。你拥有的信息越多，你的效率越高。利用 SPSS 来做出的决策能够影响我们做的每一件事情” —Noel Roos, RCC 市场研究部项目经理

在投资前，RCC 调查了他们的客户想要什么样的服务，他们想用这些服务来做什么和他们愿意为这些服务付多少钱。为了推出一项新的服务，RCC 请客户讨论关键利益。“理解客户对我们的服务已经了解的程度能帮助我们决定突出那些明显的或隐蔽的利益，” Roos 解释道。

## 改进资源分配能力

利用 SPSS，RCC 分析了市场数据以决定如何分配资源。据 Roos 所说，“基于市场数据的决策对底线有直接的影响。你拥有的信息越多，你的效率越高。利用 SPSS 来做出的决策能够影响我们做的每一件事情

例如，通过分析数据，有可能发现用户拥有电话的数量与用户年龄等级之间的相关。这样，就可利用这个信息，根据满足不同目标市场的需求来分配资源。

“SPSS 能够帮助我们更好的理解将钱花在什么地方，”Roos 说，“例如，我们是将钱花在扩大通讯覆盖范围还是推广新的服务如语音拨号上？没有 SPSS 来解答这些问题，我们可能会在投资上有失误，例如，当客户需要高质量的服务时，我们却将钱花在推广新的服务内容上。”

## 增加品牌形象的意识程度

RCC 每年进行品牌调查以评估它们给客户的形象是否就是它想要给客户的形象。调查中，请客户回忆 RCC 的广告并鉴别他们对品牌的印象。RCC 也收集客户对 RCC 的品牌形象与竞争对手的品牌形象对比的数据。

利用 SPSS，RCC 能够比较满意的评估数据以决定客户是否能将它的品牌从竞争对手的品牌区分出来。“研究品牌调查的数据能够帮助我们衡量广告的有效性和评估我们是否走在正确的轨道，” Roos 说，“我们是在建立我们想要的名声吗？我们的品牌是突出的还是与别的品牌混杂在一起？如果我们的客户难于将我们从竞争对手中分辨出来，那们这将是一个价格问题而且是我们不想要的结果。”

“我们做了大量的研究以帮助我们理解如何定位服务…这告诉了客户想要的和我们需要什么来满足他们的需求。” —Noel Roos, RCC市场研究部项目经理

## 英国电信公司

### 面临的问题:

为了从市场营销预算中获得最大的价值,英国电信需要建立模型来预测潜在客户的购买倾向和他们变为用户之后可能的价值。在对客户类别作了准确描述之后,英国电信公司希望针对某个客户群体设计一种新产品,以期达到以下结果:更高的市场活动回应率,增加产品收入,更高的市场占有率。

### 解决方案:

英国电信选用了 SPSS 的数据挖掘产品 Clementine,来为其“Business Highway”活动分析数据和建立探索模型,“Business Highway”是为小型商业客户设计的一款产品。如今,英国电信公司已经对这些客户及其电信消费行为有了一个更好的理解。在未来,该公司将继续依赖 Clementine 可视化的快速建模环境为其未来的营销活动指引道路。

### 结果:

一向太平的电信行业现在已充满了火药味。前垄断者—英国电信公司依然是英国本地电话、国内长途、国际长途和数据服务提供商的领袖。该公司虽然每年约有 290 亿元的销售额,但来自英国其它约 160 个电信公司的竞争,使其不得不去了解哪些人会购买它的产品和服务,以保留现有客户、获取新客户、从每个客户身上获得最大价值。

该公司特地成立了一个由高级咨询顾问 Stephen O'Brien 带领的客户和市场分析小组,以更好地了解客户。该小组的首次任务就是为新产品—“Business Highway”的市场活动建立客户特征描述。“Business Highway”是一款为小商业用户提供一条线三个号码的产品。1998 年 9 月进行了一次包括大型直邮和国家媒体广告的市场活动。

### 数据: 收集信息

相关的客户、产品、帐单和历史数据分散在公司的各个部门。因此,分析小组首先集中数据,形成一个新的数据集市,以便于数据库营销。由于必须从各不同系统中抽取数据,并进行匹配,清洗工作,所以准备工作要花费大量时间。只有在做好这些数据准备工作之后,才能开始作数据分析工作。数据准备和数据分析均可利用 Clementine 工具进行。

### 工具包: 用数据挖掘发现隐含的模式

为对样本数据进行挖掘并从中发现潜在的模式和趋势,英国电信公司选择了 SPSS 公司的快速建模环境—Clementine。O'Brien 之所以选择 Clementine 为其数据挖掘工具,是因为 Clementine 为客户和市场分析小组提供了多种分析方法—包括聚类、神经网络、关联规则和决策树。它也能容易地处理常规数据问题,如异常值、缺失值和低价值数据。”

#### 分析数据和建模过程

分析小组用 Clementine 作为其主要的数据分析和建模工具。

在数据分析过程中,首先,该部门员工利用 Clementine 进行了数据质量主题分析,熟悉数据及其分布,排除了与购买“Business Highway”产品不太相关的数据属性。然后,对客户购买产品的倾向相关的单个数据属性的预示力度进行测量。例如:两位地区数字代码、地理指示器,与回应和购买属性明显相关。



在这些分析之后，分析小组利用 Clementine 提供的决策树算法构建并检验了一系列探索性模型。O'Brien 称该工具的最吸引人的地方是：“你不会在数据挖掘项目中迷惑，Clementine 让你快速试验各种想法并排除一些错误想法，你可以在几天内构建多个探索性模型，从而降低了数据挖掘项目的失败率。”

向销售人员和营销活动提供了“最佳潜在客户”清单

“Clementine 的主要输出结果是对数据的洞察力—那是数据挖掘的全部所在—和这些洞察结果的可视化展示。”O'Brien 说。

“我们提供给销售人员和营销活动的是客户清单和图形，并表明了为什么他们应该向这些客户推荐 'Business Highway' 产品。”

直邮活动回应率提高了 100%

“Business Highway 项目的成功说明了如何在商业活动中通过数据挖掘获利。有了 Clementine，先前所作的探索性数据分析和可视化结果，使我们能够制订出满意的客户选择标准。甚至在完成最终模型之前，我们就已经可以超越原始目标，并使活动的回应率提高了 100%。”O'Brien 说。

还有更多的事情要做。下一步，分析小组计划使用 Clementine 识别有最大潜在利润的客户和垃圾客户。未来他们或许会尝试确定客户欺诈模式是否与客户流失模式一致。

### **建模：投资回报**

成功地客户特征描述需要商业知识、合适的数据和合适的产品。英国电信新的建模方案使之可以在产品和活动的整个生命周期中定位目标客户，确定不断变化的市场的发展趋势并加强在不同市场领域的渗透。SPSS 的数据挖掘工作平台—Clementine 将为营销活动提供快速、可靠的统计分析。而回报呢？就像英国电信的“Business Highway”案例所表明的：更好的客户和更高的销售额。

## **国内成功案例**

### **某省级电信公司**

#### **面临的问题**

小灵通用户是某省电信企业利润的主要来源之一，也是市场竞争的焦点。在目前的市场形势下，发展新客户的成本远远大于留住已有客户的成本。但在其电信经营分析与决策支持系统中无法做到提前预警，挽留用户。某省级电信公司希望借助 Clementine 数据挖掘技术发现离网小灵通用户的潜在模式，提前进行预测，从而变被动服务为主动服务。

#### **解决方案**

选用了 SPSS 的数据挖掘产品 Clementine，使用企业级数据挖掘平台 Clementine，按照 CRISP-DM 数据挖掘标准流程，以全省某特定时间离网的后付费小灵通用户为研究对象，同时选取同期在网的后付费小灵通用户作为对照研究对象；在用户产品订购数据、用户消费行为特征数据和用户其他行为特征数据的基础上，分别建立神经网络模型、C5.0 模型和 Logistic 回归模型，最终得到预测的具体离网用户名单以及离网概率和离网用户的特征描述信息。



## 结果

通过应用 Clementine 数据挖掘结果,发现了离网小灵通用户的潜在模式特征,为下一步的客户挽留提供可靠依据

## 某省级移动公司

### 面临的问题

公司希望借助数据挖掘软件识别客户特征,通过了解客户的概貌指导公司制定合适的营销策略从而达到保持、拓展现有的客户价值。

### 解决方案

使用企业级数据挖掘平台 Clementine,按照 CRISP-DM 数据挖掘标准流程,结合客户资料数据、客户服务数据以及计费帐务数据,使用 Clementine 中的各种聚类算法生成客户分群模型,分析不同算法得到的分群结果,基于行业应用评估标准对已生成的模型评估,最后使用决策树算法得到各客户群特征。

## 结果

Clementine 良好的稳定性和卓越的性能大大提高数据挖掘生产效率,可视化编程模式让业务人员和行业专家能够参与到数据挖掘项目过程中,业务经验和数据挖掘技术的完美结合,提高了数据挖掘模型的质量。通过数据挖掘专家和行业专家的合作,对移动客户业务方面提供了如针对性差异化套餐设计、差异化客户服务以及设计促销活动等多方面建议

## 部分国内外电信客户的名单(排名不分先后)

AT&T——美国

Westel——匈牙利

BPL移动公司——印度

中国电信总公司

英国电信——英国

武汉电信

西班牙电信——西班牙

新疆电信

Etisalat——阿联酋

中国移动总公司

KPM——荷兰

江苏移动

NTT——日本

内蒙古移动

Orange——瑞士

辽宁移动

Panafon——希腊

吉林移动

SingTel——新加坡

Telecel——葡萄牙

Telemig Cellular——巴西

Verizon——美国

VIAG Interkom——德国

Vodafone——澳大利亚

## 公司简介

SPSS公司成立于1975年,总部在芝加哥。在纳斯达克上的股票代码为SPSS。在全球100多个国家和地区有分支机构或合作伙伴,直系员工超过2000人。全球约有25万家产品用户,它们分布于通讯、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等多个领域和行业,全球500强中有80%的公司使用SPSS,而在市场研究和市场调查领域有超过80%的市场占有率,是世界上应用最广泛的专业统计软件之一!

SPSS公司致力于提供高效、易用的统计分析软件和数据挖掘解决方案,解决数据获得和数据分析问题,从而使数据分析广泛地应用于决策制定中。同时结合您的数据和商业知识创造并实施最佳预测模型。

作为统计预测分析的先驱,SPSS在财政金融、政府机构、教育机构、电信、市场研究、零售、电子商务等分析方案方面已有超过30年的经验,为您提供从数据输入整理、探索分析、分析报告、建立模型、预测分析到结果发布的完整解决方案,使您能够更好地预测未来,把握先机。

2006年5月18日在中国成立了SPSS China (胜策软件(上海)有限公司),至此SPSS产品正式迈入中国软件市场。SPSS China (胜策软件(上海)有限公司)将在中国为客户提供产品、渠道等相关问题的咨询及支持。

[www.spss-china.com](http://www.spss-china.com)

The SPSS logo is displayed in white text on a red square background.

胜策软件

SPSS China / 胜策软件(上海)有限公司  
地址:上海市淮海路93号时代广场2901-02室(200021)  
电话:+86 021-61415258  
传真:+86 021-61415268

SPSS China 胜策软件北京代表处  
地址:北京市西城区南礼士路66号建威大厦1913室(100045)  
电话:+86 010-68080383  
传真:+86 010-68080331

更多信息,请访问[www.spss-china.com](http://www.spss-china.com)  
Email:[info@spss-china.com](mailto:info@spss-china.com)