

# 客户细分在电信营销中的应用研究

## Research On Customer Segmentation In Telecom Sales

马子斌 杨鸿宾 (北京邮电大学经济管理学院 北京 100876)

**摘要:**今天的电信行业面临着如何更好的理解客户信息将其转变为有用的知识,如何维护客户关系和发展有价值客户并提供个性化的服务的难题。数据挖掘能从大量客户数据中发现潜在和有价值的知识,这无容置疑的为电信营销提供的极大的支持。进行客户分群分析的目标是通过综合客户消费行为,帐单行为以及人口统计信息(主要是客户行业)等因素,按照价值和行为两个维度进行战略分群,对客户进行细分。为了重点研究客户分群分析在电信营销中的应用;另外也针对基于数据挖掘模型的应用给出了一些市场战略和个性化营销方法。

**关键词:**数据挖掘 客户细分 客户聚类 个性化营销

客户细分(Customer Segmentation)是市场研究的重要分析工具。它依据与企业成功相关的属性,从客户的角度去划分市场结构,从而丰富企业的市场营销手段并提升企业对客户产品使用行为的理解,进而强化客户管理与沟通。本文运用数据挖掘的方法从定义业务问题范围、数据准备(选取和抽样)、探索性数据分析、建模、模型评估、模型发布和实施六个步骤展开论述。

### 1 数据来源讨论及业务理解

挖掘分析所需的数据全部来源于统一的决策分析系统数据仓库。尽管数据仓库中已经对原始的业务数据进行了清洗、转换和集成,仍然需要按照挖掘任务的要求生成挖掘专用的数据集,并将数据从数据仓库抽取过来,并建立数据仓库源数据到数据挖掘集市的数据映射关系。同时数据挖掘专家将对业务进行理解,明确某些变量的计算方法,并初步评估源数据的质量。探索性数据分析的目的在于:

- (1) 数据质量检查、整理和理解;
- (2) 分析待选自变量和目标变量之间的关系;
- (3) 最终选择建立模型的变量和记录。

客户范围是基础客户群的限定,即针对哪些客户进行分析。由于计费帐务系统中客户资料数据不全,

因此我们无法准确确定目标客户的范围,只能通过取用户表中的该品牌用户,将拥有同一客户名称或者客户ID的用户归并得到目标客户。经过数据情况的初步分析确定分群目标用户群的范围是:某品牌用户且用户状态正常且用户类型为非特殊用户,至少拥有一种语音、数据、新业务等产品的用户。

数据来源讨论过程也是对业务数据的理解过程,在此过程中数据挖掘专家将和业务人员一起搜寻并检查业务数据,创建源数据映射关系描述,将客户数据与建模相关的各个属性对应起来。数据能被整合到一个适当的程度,省略不适当的记录、不完整的数据记录等等。在挖掘过程中,还会不断遇到问题,不断反馈,需和业务人员进行交流,对所选变量进行不断调整。

对于变量选择过程,数据挖掘小组设计了变量选择模板。该模板提供方便的新增、修改、维护建模所需变量的功能,并且可以根据选择的变量自动生成建立挖掘分析宽表的SQL语句。

### 2 电信客户分群分析

#### 2.1 变量选择和标准化

在数据探索的基础之上,通过 Value/Behavior Segmentation(价值行为细分),就可以进行V(Value)、B(Behavior)聚类分析分析量收不平衡现象。建模的

流程包括变量选择生成 V、B 分析基表, 标准化, 评分/聚类结果回写三个大的流程。具体见图 1 所示。

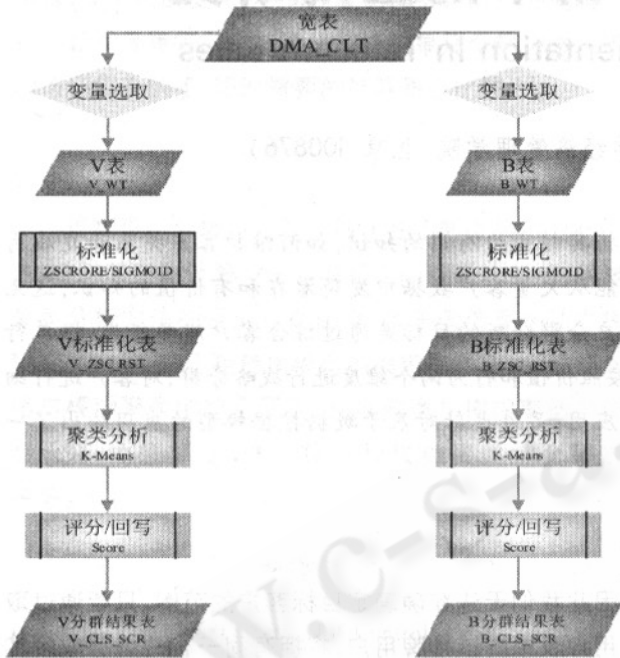


图 1 客户分群建模流程图

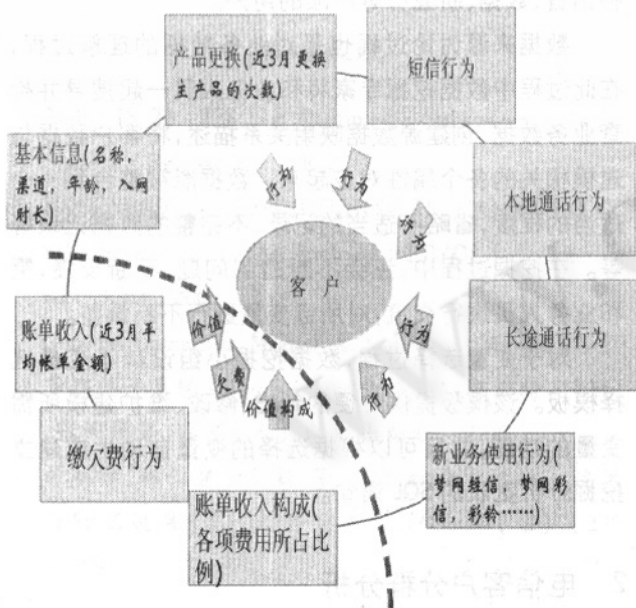


图 2 价值、行为挖掘变量选择

建立宽表之后, 需要分别得到参予 V、B 分群的变量, 具体处理方法是把这 V、B 变量分别抽取出来另建

新表。可以提高效率, 规范流程。

对于聚类分析来说, 输入的变量量纲不同, 会对结果造成很大影响, 因此在聚类分析之前需要对所有参予聚类的变量进行标准化处理。这里着重介绍本次分析中用到的两种方法: Zscore 和 Sigmoid。

Zscore 方法的公式如下, 变量的每个值减去变量均值再除以标准差, 得到的结果其实是标志一个值距离中心(均值)的距离是标准差的几倍。这是统计学中非常经典的标准化方法, 可以在很多分析中应用。

$$f(x) = \frac{x - \frac{1}{n} \sum_{i=1}^n x_i}{\sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - (\frac{1}{n} \sum_{i=1}^n x_i)^2}}$$

Sigmoid 的公式如下, 通过 Sigmoid 变换, 会对变量的最大值和最小值进行比较大的压缩, 而中间的数据压缩率不会很高。经过 Sigmoid 变换会消除一些奇异值的影响, 当然也可能丢失一些信息。所以我们最好不要直接对原始变量进行 Sigmoid 变换, 丢失的信息太多。本次分析采用了先 Zscore 再进行 Sigmoid 变换的方式。

$$f(x) = \frac{1}{1 + e^{-x}}$$

经过标准化的表的命名规则为 X\_ZSC (SIG)\_XXXX, 后面加上日期, 以便区别。如果进行了两次标准化, 将两次标准化先后写在表名中。例如 B\_ZSC\_SIG\_0330。

## 2.2 抽样和模型评分

为了提高模型的运行效率, 可以进行抽样的方式建立样本集。通过对样本的分析, 可以得到对总体的统计理论保证下的了解。样本表的命名规则为 X\_ZSC (SIG)\_SPL\_XXXX, 后面加上日期, 以便区别。模型的参数需要根据实际的情况进行选择。在本次项目中, 我们选择的分群个数一般是 5-9 个, 最大迭代次数一般为 100 步, 算法选择的是传统的 K-Means。具体方法参见图 3。

模型运行收敛后, 就可以对模型每个客户进行评分, 判定该客户属于某一组。评分结果表的命名规则为 X\_ZSC (SIG)\_SCR\_XXXX, 后面加上日期, 以便区别。得到了每个客户属于某一个组, 需要将评分结果回写

到初始表中。结果表的命名规则为 X\_ZSC(SIG)\_RST\_

果回写到初始表中的实现通过 SQL 脚本来实现。

### 2.3 模型评估暨客户群特征刻画

聚类模型的评估原则为符合业务规则,对业务有用的分群就是好分群。因此,模型的评估和对各个分群的特征刻画紧密的结合在一起。特征刻画 1 主要是计算各个分群每个指标的最大值,最小值,均值,标准差。通过对每个分群特定变量上均值的比较,可以抓住一个群最大的特征,可以对总体以及每一个分群的特征进行刻画。。特征刻画 1 可以提供对每个分群的主要特征的把握,但是对于更细致的数据分布情况,特征刻画 1 就显得单薄了。于是引入直方图分析手段进行特征刻画 2 的分析。

### 3 分群结果在电信营销中的应用

模型的作用就是要将其转化为效益,对于客户分群分析来说,模型的发布就是提交各分群客户名单列表并将特征刻画的结果展现出来,供分析人员应用。

根据分群的结果可以针对不同的客户群设计不同的营销计划,设计营销计划的大致原则如下:找到目前最需要关注的客户群,例如大量使用竞争对手 IP 电话的用户。针对这些客户的特征,并给予了恰当的 Offers; 依据客户的需求,捆绑多种不同业务,以增加客户产品拥有率;经验说明高产品拥有率与客户的忠诚度有直接的正相关性;制定区分高价值和低价值客户的优惠力度;选择合理的渠道组合,包括电话,直邮,客户经理等手段。

### 4 结论

利用适当的聚类分析模型和算法对电信系统的客户进行分群研究,可以得到更精细的客户分布情况,有利于发现优质客户和目标客户,为制定更加精细的营销计划提供支持,对于提供运营商的盈利能力和减低运营成本大有益处。

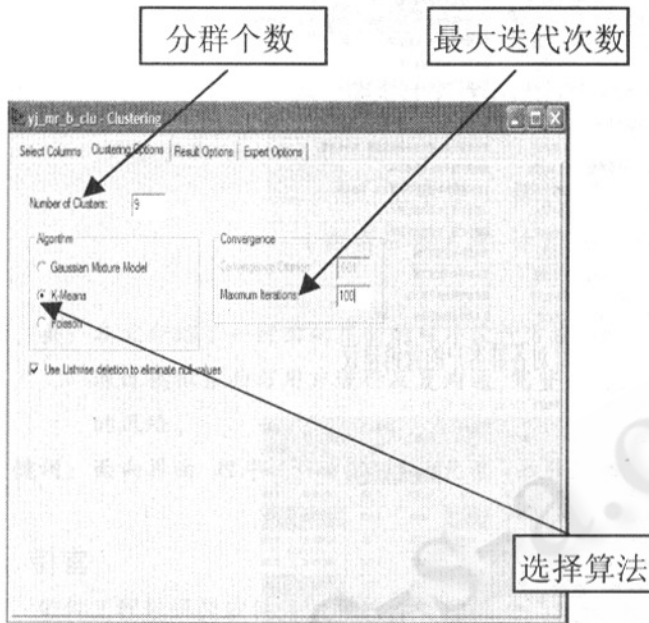


图 3 客户分群模型设置

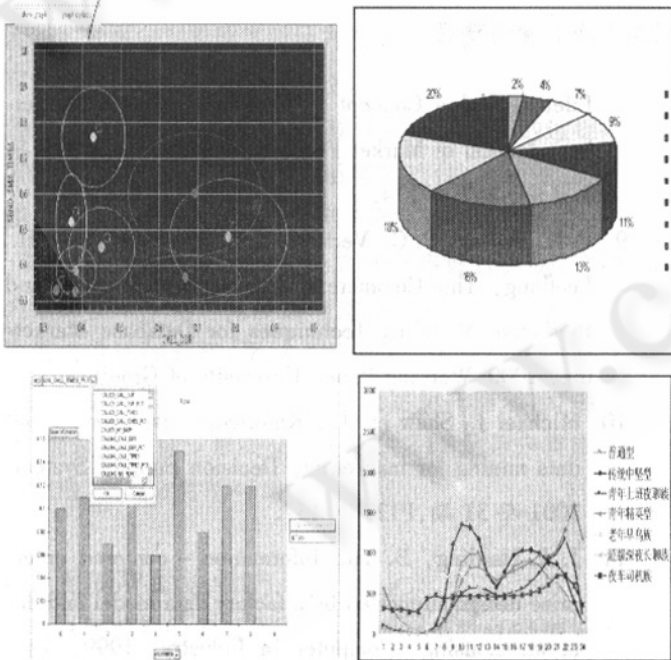


图 4 客户分群模型结果和客户群特征

XXXX,后面加上日期,以便区别。具体将模型评分结

