

# 基于数据挖掘的移动客户流失分析系统

## Analysis System of the Customer Churn in Mobile Company Based on Data Mining

叶 进 张向利 张润莲 (桂林电子工业学院 通信与信息工程系 541004)

**摘要:**客户流失是竞争日益激烈的市场中移动公司面临的一大难题。“移动客户流失分析系统”在数据挖掘的基础上,实现了客户流失模型的管理应用,其中关键的环节是通过流失客户的预测和分析,辅助市场经营人员制订相应的策略。文中介绍了决策树和聚类两种数据挖掘算法在该过程中的应用,并对得到的结果进行了说明和分析。实践证明,该系统提供的信息能够科学地帮助移动公司解决客户流失的问题。

**关键词:**数据挖掘 决策树 聚类分析 客户流失预测

### 1 引言

移动通信用户的客户流失是一个长期以来困扰全球移动电话运营商的难题。在欧洲,每年有 35 ~ 50% 的客户流失;而获取一个新客户的平均成本超过了 \$ 700,这几乎相当于一个客户 5 年内给公司带来的净利润。这种情况直接导致客户回报率的下降。从全球的范围来看,移动通信运营商每年因为客户流失所带来的损失在 100 亿左右<sup>[1,6]</sup>。根据流失的客户和没有流失的客户性质和消费行为,进行挖掘分析,建立客户流失预测模型,分析哪些客户的流失概率较大,流失客户的消费行为如何,正在成为移动公司面临的重要课题。

“移动客户流失分析系统”是一个基于数据挖掘的智能信息分析系统,它从用户资料、帐单、详单等业务数据中提取相关的信息,进行流失客户的预测和分析,辅助市场经营人员制订相应的销售策略,同时为挽留客户提供决策依据,该过程中主要采用了决策树和聚类算法两种数据挖掘算法。

### 2 基于数据挖掘的系统设计

数据挖掘又称数据库中的知识发现 ( Knowledge Discovery in Database, KDD ), 是一个利用各种分析工具在海量数据中发现模型和数据间关系的过程,这些模型和关系可以用来做出预测。数据挖掘实施的步骤一般包括三个部分:数据的准备、模型的建模模型的验证

和评估。

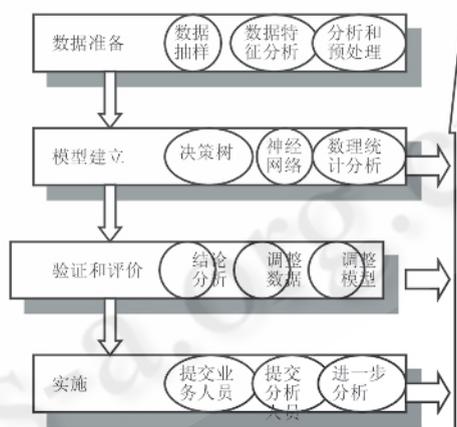
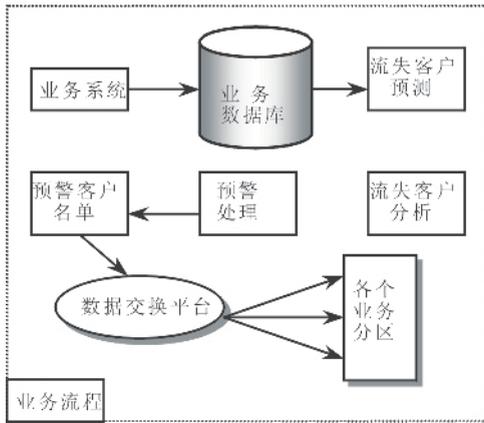


图 1 数据挖掘的步骤

图 1 所示的各个步骤并不是线性的,要取得好的结果就要不断重复这些步骤。在“移动客户流失分析系统”中,数据挖掘只是一小部分,却是关键的一个部分,系统中把数据挖掘得到的知识和市场的经验、客服的信息结合起来,应用于数据库中的数据进行流失客户的预测、分析,对确认有流失倾向的客户根据不同的情况进行预警处理,包括套餐资费调整、服务方式更改、竞争对手调查等,同时将预警客户名单通过公司的数据交换平台下发给各个业务分区,进行摸底跟踪,实施关怀工程。整个系统的业务流程如图 2 所示,其中流失客户的预测和分析是两个关键的环节,需要选择

合适的数据挖掘算法,获取有用的模型和知识应用于系统中,才能进行科学的辅助决策。



2 “移动客户流失分析”系统业务流程

### 3 数据挖掘在系统中的应用

流失又分为被动流失和主动流失,被动流失是因为客户欠费或不履行责任而被移动停机。主动流失有不同的原因,一类是客观原因限制所引起的异动,如搬迁等,另一类是客户主动放弃,如因为竞争对手的优惠政策,对目前的服务不满意等。

在这些群体中,那些主动流失的客户是我们分析的对象,在这类群体中,用户价值和信用度高的可能流失客户是我们最关心的对象,如果能够成功预测出这类将要流失的客户,并且能够成功的挽留这类大客户,将给移动公司带来可观的利润。

#### 3.1 数据准备

数据准备是数据挖掘过程的最重要的组成部分。客户的历史行为数据中隐含着大量与流失相关的行为模式,数据必须围绕移动市场分析得到的与流失相关的信息来组建。客户流失分析用到的数据包括客户的基本信息,客户的合同信息,客户消费数据(每月帐单),通话行为特征(如通话频度统计、被叫方资料分析、语音和数据的通信特点)等,数据来自有 BOSS 系统、大客户管理系统、客户服务热线、帐务管理系统以及手工报表等其他方式获得的数据。

数据库中的数据包含着大量的属性,每月帐单数据库中的费用属性就有近百个,并不是每个属性都是可用的,如果将不相关或相关性很小的属性用于建模,

将会成倍增加计算的代价。另外,如果区分客户流失的能力很差的属性被选择,用它所建立的流失模型不准确,就会影响预测的效果。如何选择与流失相关的属性呢?

属性选取的目标是:在属性组成的向量空间内,使得类之间的距离尽量的大,类内部的距离尽量的小,同时,尽量减少属性的个数。为了达到这个目标,需要选择有用的属性并抽取、变换组合出新的属性向量,用这种方法力求在减少属性数目的同时,保留甚至提高属性之间有用的关联信息。

结合移动市场得到的先验知识,我们选取和组合的流失变量见表 1。

表 1 客户流失建模的变量定义

A( fee_rate1)	本月话费下降比率
B( fee_rate2)	上月话费下降比率
C( call_remove)	是否呼叫转移到网外
D( call_rate_union)	与联通用户通话比例
E( call_1001)	是否拨打 1001( 联通服务热线电话)
R( is_churn)	是否离网“1”表示在用用户, “0”表示离网用户

由移动部门提供的 10000 条抽样数据经过清洗、处理异常等工作,获取了 1250 条符合要求的候选数据,数据集的格式见表 2。

将这些多维数据导入 SQL SERVER2000 后,使用 Analysis Servers 中的工具可以获得所需的数据挖掘模型<sup>[5]</sup>。

#### 3.2 用决策树方法建立移动客户流失模型

决策树是应用最广泛的预测技术,它的基本组成部分为决策节点、分支和叶子,建立决策树的过程即树的生长过程,利用信息论中的信息增益理论寻找数据集中具有最大信息量的字段,建立决策树的一个结点,再根据字段的不同取值建立树的分支,在每个分支子集中重复建树的下层结点和分支,这样不断归纳学习,就可生成反映一定规则的决策树。使用决策树方法得到的客户流失分析模型如图 3 所示。

该决策树可以区分不同的流失客户的群组及每一群组的潜在流失因素。例如,标识为‘2’的叶子节点

表示,在该群体中客户流失的概率只有 17.65%,他们包括以下特征:

D (call\_rate\_union) < =0.11;与联通用户通话比例小于话单总数的 11% ;

B (fee\_rate2 <0.4;上月话费下降比率低于 40% ;

A (fee\_rate1) < 0.6;本月话费下降比率低于

60%。

标识为‘3’的叶子节点表示,在该群体中客户流失的概率达到 88.24%,他们的特征与标识为‘2’的群体的区别在于:

A (fee\_rate1) > =0.6;本月话费下降比率高于

60%。

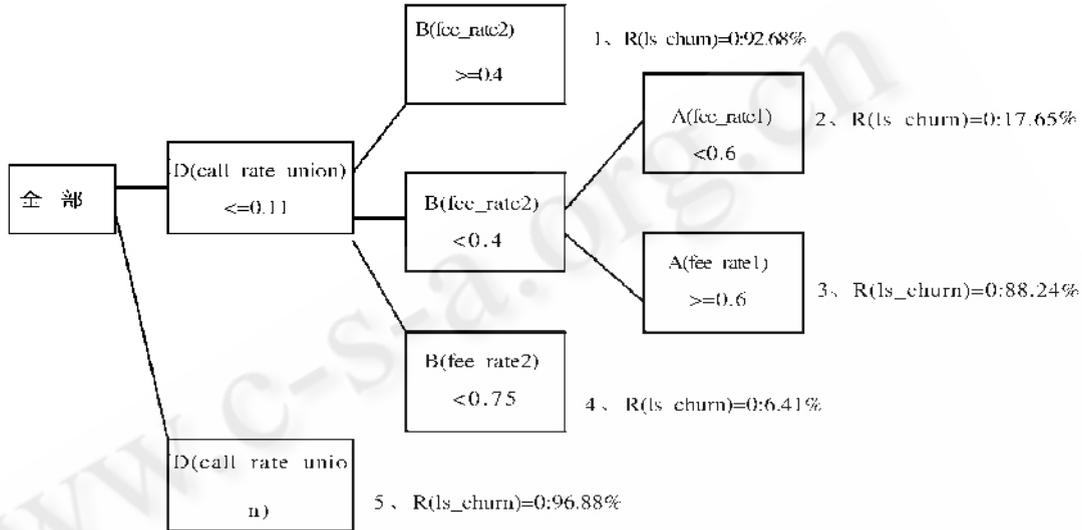


图 3 使用决策树方法得到的客户流失分析模型

表 2 用于数据挖掘的多维数据集格式

A (fee_rate1)	B (fee_rate2)	C (call_remove)	D (call_rate_union)	E (call_1001)	R (is_churn)
0.43	0.97	0	0.09	0	1
0.23	0.92	0	0.07	0	1
0.04	0.42	0	0.07	0	1

表 3 使用聚类方法得到的客户细分结果表

客户群编号	占总体客户的比例	与联通常话比例	上月话费下降比率	前月话费下降比率	是否呼叫转移到网外	是否拨打联通服务热线
1	2.1%	中等	正常	正常	NO	YES
2	35.1%	偏高	正常	正常	NO	NO
3	24.2%	中等	较高	较高	NO	NO
4	1%	中等	较高	正常	YES	NO

从该决策树分析,与客户流失关系最密切的因素是与联通常话的比例,其次是上月话费下降比率和本月话费下降。而呼叫转移到网外和拨打联通服务热线

1001 两个因素由于出现的概率非常少,在以信息熵为划分标准的决策树分类中无法体现。而由经验知识可知,这两个特征一旦出现,将很快导致客户流失,因此

有必要对不同类型的客户采取不同的方法进行流失分析,从而实现效率更高的一对一营销。

### 3.3 使用聚类分析对流失客户进行细分

客户群体的细分是一对一营销的基础,聚类分析是通过分析数据库中的记录数据,根据一定的分类规则,合理地划分记录集合,确定每个记录所在类别。通过进行聚类分析,能够有效地把数据划分到不同的组中,组之间的差别尽可能大,组内的差别尽可能小。对前面预测的流失客户进行聚类分析,选定  $K=9$  (作为聚类算法的参数),表 3 列出得到的其中 4 种不同的典型客户群。

第一种流失客户的主要行为是拨打联通服务热线,需要进一步调查他是否拨打移动服务热线进行咨询或投诉,这类客户流失的原因很可能是心理状态;

第二种流失客户的主要行为是与联通用户通话比例较高,需要加强与对手的竞争力度来减少流失;

第三种流失客户的主要行为是话费连续下降,需要了解他们的消费习惯和特点,制定符合他们的套餐业务来刺激消费;

第四种流失客户的主要行为是呼叫转移到网外,需要跟踪呼转电话,主动介绍和推荐公司的新业务、新服务,通过关怀措施及时进行挽留。

这些聚类分析的结果对市场营销人员制定正确的市场策略、开展准确的促销活动是非常有价值的。

## 4 结束语

“移动客户流失分析系统”在数据挖掘的基础上,实现了客户流失模型的管理应用,包括模型所需数据的导入、模型的自动运行、流失客户名单的导出以及流

失客户总体情况的联机分析处理 (OLAP)。系统自 2003 年 2 月正式开发以来,经过用户环境下的数据初始化、现场测试、三个月的试运行后,已经顺利投入使用,实践证明,该系统提供的信息能够科学地帮助移动公司解决客户流失的问题。

在越来越多的行业中都面临着客户流失的问题,只有不断在数据挖掘的过程中进行知识学习,才能够降低客户流失率、提高市场竞争力。

## 参考文献

- 1 段云峰、吴唯宁,中国的移动经营分析系统介绍,电信技术,2001 年 11 月。
- 2 Methmed Kantardzic,数据挖掘——概念、模型、方法和算法[M],清华大学出版社,2003,8.
- 3 李军,利用数据挖掘实现电信行业客户流失分析.数据仓库之路,www.dwway.com,2001 年 7 月。
- 4 朱浩刚等,基于数据挖掘的移动通信业客户流失管理,计算机工程与应用[J],2004,1:215-219.
- 5 <http://www.microsoft.com/sql>
- 6 Anderson Consulting, Battling churn to Increase Shareholder Value: Wireless Challenge for the future, Anderson Consulting Research Report,2000.
- 7 Hyun\_Moon Shin, An application of data mining for marketing in telecommunication[J]. Management of Engineering and Technology, 2001:247.
- 8 Lian Yan, Predicting customer behavior in telecommunications[J]. Intelligent Systems, IEEE, Vol 19, Issue 2, March - April 2004, 50 - 58.