Hans Journal of Data Mining 数据挖掘, 2016, 6(1), 28-36
Published Online January 2016 in Hans. http://www.hanspub.org/journal/hjdm
http://dx.doi.org/10.12677/hjdm.2016.61004

Hans汉斯

Research on Telecom Customer Churn Prediction Based on Customer Value Classification in 3G Environment

Lin Xu, Zhiguo Zhu*, Huilu Li, Min Li

School of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian Liaoning
Email: *zhuzg0628@126.com

Received: Dec. 27th, 2015; accepted: Jan. 11th, 2016; published: Jan. 14th, 2016

Abstract

Telecom operators are facing an urgent problem of telecom customer churn that should be solved as soon as possible. This paper, according to the three-month average customer consumption, divides the levels of customer value, comprehensively uses decision tree algorithm and clustering algorithm modeling of data mining, introduces confusion matrix model for model evaluation, and uses the model output rules set for targeted customers' maintaining marketing, so as to reduce customer churn, improve the efficiency of marketing, and enhance the core competitiveness of telecom operators in 3G environment.

Keywords

Customer Churn, Data Mining, Decision Tree, Confusion Matrix

3G环境下基于客户价值分类的电信客户流失预测研究

徐　麟，朱志国*，李会录，李　敏

东北财经大学管理科学与工程学院，辽宁　大连
Email: *zhuzg0628@126.com

*通讯作者。

文章引用: 徐麟, 朱志国, 李会录, 李敏. 3G 环境下基于客户价值分类的电信客户流失预测研究[J]. 数据挖掘, 2016, 6(1): 28-36. http://dx.doi.org/10.12677/hjdm.2016.61004

收稿日期：2015年12月27日；录用日期：2016年1月11日；发布日期：2016年1月14日

摘　要

电信客户流失问题是电信运营商面临的迫切需要解决的问题。本文针对3G环境下，根据客户三个月平均消费水平进行客户价值划分，综合运用数据挖掘中决策树算法和聚类算法进行建模，引入混淆矩阵对模型进行评估，利用模型输出的规则集有针对性的进行流失客户维系营销，从而达到降低客户流失，提高营销效率，提升电信运营商核心竞争力的目的。

关键词

客户流失，数据挖掘，决策树，混淆矩阵

1. 引言

随着科技的进步，电信行业发展的步伐日益加快，自 2012 年，中国电信行业开始加大了对 3G 的投入，到 2014 年全面进入 3G 时代，电信客户数据迅速增长，同时也面临客户大量流失，出现“增量不增收”的情况[1]。因此，客户流失预测在电信企业增加收入、提高客户保有、降低营销成本等多方面起着至关重要的作用。

本论文通过对电信客户流失实际情况的分析，结合当前 3G 环境下客户广泛使用手机流量这一特征，利用 SPSS 公司的 Clementine12.0 数据挖掘软件[2]，采用 CRISP-DM 的数据挖掘建模标准，提取了某县电信公司 2014 年 9 月至 12 月业务支撑系统和客户数据库里的客户数据，运用 C5.0 决策树和 CART 决策树分别建立了客户流失预测模型，利用 K-Means 聚类算法建立了流失客户聚类模型，引入混淆矩阵对模型进行评估，最后根据模型输出的规则集制定相对应的客户维系策略。

2. 相关研究综述

随着电信行业发展的不断加速，数据挖掘技术在电信行业应用也不断深入，国内外研究者在这方面的研究也取得了一定的成果[3]。T. Sato 通过研究发现，利用主成份分析方法建立的流失客户模型较之 C5.0 决策树建模得出的规则集能获得更大的收益，并且首次将该方法应用到电信流失客户问题的研究中[4]。Louis 对决策树和逻辑回归判别分析建立的客户流失预测模型进行了对比分析[5]。Rosset 等人对客户价值进行分类，利用逻辑回归建立了基于客户价值分类的流失预测模型，使得模型的针对性更强，准确率得到了大幅度提高[6]。Piotr 在客户流失分析方法的研究中提出将 K-Means 算法与传统分类算法相结合，最终的研究结果表明该算法应用在客户流失预测的准确率高于传统的分类预测算法[7]。Cardeln 等人运用决策树建模的方法以美国某公司的客户数据为研究对象进行流失预测分析，最终不仅取得了较高的准确率，还获得了更有价值的客户流失规则。Mozer 等人在对美国某公司的客户进行流失预测的研究中，不仅对数据进行了抽样分析，还将 ANN 技术和引入收益计算相结合，最终获得了较好的研究结果[8]。因此，本文认为好的数据挖掘模型是需要建立在充分了解行业知识的基础上，灵活运用算法，才能够得出有价值的结果，没有最好的模型，只有最适合的算法。

3. 实验模型的建立

建模过程参照 CRISP-DM 的数据挖掘建模标准，通过商业理解、数据理解、数据准备、建立模型步骤建立实验模型[9]。

3.1. 实验数据

原始数据来源于某县电信运营商的业务数据库，其中需要用到的数据表如表 1 所示。

通过数据提取，最后汇总成实验数据宽表，如图 1 所示。

下面，进一步对客户价值进行分类：

根据客户 3 个月的平均消费得出客户价值属性，*ARPU* 代表月消费，C_h 和 C_l 代表高价值客户和低价值客户，如公示(1)和(2)所示：

$$C_h = \frac{ARPU_9 + ARPU_{10} + ARPU_{11}}{3} \geq 50 \tag{1}$$

$$C_l = \frac{ARPU_9 + ARPU_{10} + ARPU_{11}}{3} < 50 \tag{2}$$

分类后的实验数据宽表，如图 2 所示。

3.2. 实验模型

首先利用清理好的客户数据宽表，建立客户流失预测模型，其目的是为了发现客户流失的一些基本特征[10]，然后筛选出流失客户数据，通过聚类方法建立流失客户聚类模型，根据生成的聚类规则，分别针对每一类的流失客户制定相对应的挽留策略。

客户流失预测模型的具体建模步骤是首先使用数据挖掘中的 Apriori 关联规则算法，计算出宽表中所有客户属性特征和客户是否流失之间的关联强度[11]，将关联性较弱的属性剔除，目的是为了提高挖掘效率，然后筛选出具有以上特征属性的样本数据，使用 C5.0 决策树和 CART 决策树算法[12]，分别对在网客户和流失客户进行挖掘，找出流失客户的具体消费行为规则，然后对比两种决策树建模得出的规则集的优劣，选择出最佳的建模算法，最后根据选定的模型得出的规则集筛选出符合要求的数据，即有流失趋势的客户。类别倾斜是数据挖掘中常出现的一个问题，即因选取数据比例失衡导致模型将大量的数据对象都划分到占比大的一方去，容易产生空树。本文在数据准备阶段是按照 1:1 的比例来提取流失客户和在网客户的，从而有效的避免了类别倾斜的问题[13]。

流失客户聚类模型主要使用到聚类算法中的 K-Means 算法，对离网客户进行划分，根据最终输出的不同聚类规则，找出相应的客户离网原因，实施挽留措施，使下一步的客户维系工作目的性更强，提高客服人员的工作效率，节省客户维系成本。

原始样本数据中的流失客户数据亦可用作为流失客户聚类模型的样本数据，为了提高实验效率，现

Table 1. Database table
表 1. 使用到的数据库表

表名	备注
客户信息表	包含性别、身份证、家庭住址、职业等信息
用户表	包含手机号码、入网时间、状态变更时间、用户状态、用户类型
号码月账单表	账户编号、账户类型、出账收入、入账收入
预存表	包含手机号码、预存类型、预存金额、办理时间
欠费信息表	包含账户编码、欠费时间、欠费金额、敏感客户属性
9~11 月账单汇总表	包含 2014 年 9 月至 11 月客户消费情况
9~11 月流量汇总表	包含 2014 年 9 月至 11 月客户流量使用情况

P	Q	R	S	T	U	V	W	X	Y
入网时间	状态变更时	服务计划名称	是否欠费	是否办理预存	是否敏感客户	9月消费值	10月消费值	11月消费值	11月流量值
2003/8/29 11:05	########	CMNET交换机接入专线套餐				0	0	0	
2004/3/31 16:56	########	CMNET交换机接入专线套餐				0	0	0	
2004/3/31 17:03	########	CMNET交换机接入专线套餐				0	0	0	
2004/3/31 17:04	########	CMNET交换机接入专线套餐				0	0	0	
2004/5/25 11:45	########	广域网互连业务套餐				0	0	0	
2004/6/30 18:11	########	CMNET交换机接入专线套餐				0	0	0	
2004/9/29 15:40	########	CMNET交换机接入专线套餐				0	0	0	
2004/9/29 16:05	########	广域网互连业务套餐				0	0	0	
2005/3/31 10:06	########	CMNET交换机接入专线套餐				0	0	0	
2002/12/24 8:45	########	神州行分县卡二(上饶)	yes			48.2	29.06		
2002/12/24 15:11	########	全球通商旅套餐58	yes	yes		91.3	93.21	83.3	
2002/12/25 14:27	########	动感地带音乐套餐		yes	yes	60.44	38.91	87.89	71.09375
2003/1/2 19:31	########	神州行分县卡二(上饶)	yes			18.5	21.76	28	
2003/1/2 21:40	########	全球通商务98	yes						
2003/1/3 10:55	########	新神州行家园卡(全县一张网)(上饶)	yes	yes		39	39	39	0.008789
2003/1/3 14:54	########	全球通公务68	yes			80	80		
2003/1/28 9:50	########	神州行分县卡二(上饶)	yes			88			
2003/2/10 19:07	########	全球通商旅套餐88	yes			115.7	114.59		
2003/2/12 15:11	########	神州行家园卡(原田园卡)	yes			11.6	11.56	10.8	0.006836
2003/2/21 16:21	########	全球通商旅套餐88-语音主产品	yes		yes	106	106		
2003/3/9 9:49	########	神州行分县卡二(上饶)		yes		20.5	21.41	19	
2003/3/9 10:11	########	神州行大众卡15(非签约),可转签约	yes	yes		68	26.8	23.62	
2003/4/1 12:27	########	神州行大众卡15(非签约),可转签约	yes	yes		74.52	121.61	194.38	1077.287
2003/4/1 16:26	########	神州行大众卡畅听版		yes		29.95	32.1	22.28	0.314453
2003/4/7 11:14	########	神州行分县卡二(上饶)	yes	yes				80	
2003/4/18 15:03	########	神州行大众卡畅听版				5	5	5	

Figure 1. Summary data
图 1. 汇总数据宽表示例

X	Y	Z	AA	AB	AC	AD	AE	AF
是否欠费	是否预存	是否敏感客户	9月消费值	10月消费值	11月消费值	3月平均消费值	话费波动	客户价值
否	是	是	60.44	38.91	87.89	62.4	0.45	高价值客户
是	是	否	39	39	39	39.0	0.00	低价值客户
是	是	否	74.52	121.61	194.38	130.2	1.61	高价值客户
否	是	否	29.95	32.1	22.28	28.1	-0.26	低价值客户
是	是	否	42	42	42	42.0	0.00	低价值客户
否	是	否	13.1	13.9	13.8	13.6	0.05	低价值客户
否	是	否	64.7	76.46	69.57	70.2	0.08	高价值客户
是	是	否	26.45	32.81	29.3	29.5	0.11	低价值客户
是	是	否	19.5	25.84	20.9	22.1	0.07	低价值客户
是	是	否	22.02	69.06	82.38	57.8	2.74	高价值客户
否	是	否	49.22	58.72	22.81	43.6	-0.54	低价值客户
是	是	否	50.85	19.5	14.2	28.2	-0.72	低价值客户
是	是	否	39	39	39	39.0	0.00	低价值客户
是	是	否	49.5	45.49	61	52.0	0.23	高价值客户
是	是	是	173.8	181.29	173.91	176.3	0.00	高价值客户
否	是	否	55.8	19.14	24.6	33.2	-0.56	低价值客户
否	否	否	13.4	33.98	29.58	25.7	1.21	低价值客户
是	是	否	22	22.3	39	27.8	0.77	低价值客户
是	否	否	31.5	24.46	44.5	33.5	0.41	低价值客户
是	是	否	42.3	66	42.3	50.2	0.00	高价值客户
否	是	否	20.9	15	131.8	55.9	5.31	高价值客户
是	是	否	100.6	111.1	101.1	104.3	0.00	高价值客户
是	是	否	30.8	18	21	23.3	-0.32	低价值客户
是	是	否	42.8	44	37.3	41.4	-0.13	低价值客户
是	是	否	29.8	5.4	52.2	29.1	0.75	低价值客户

Figure 2. Data table classification
图 2. 汇总数据宽表示例

将流失客户聚类模型和客户流失预测模型的样本数据放置在同一张数据宽表中，即在 SPSS Clementine 同一个数据挖掘流中同时建立起这两个模型。

利用 SPSS Clementine 中决策树建模模块[14]分别使用 C5.0 决策树和 CART 决策树对高价值客户和低价值客户建立流失预测模型和流失客户聚类模型，如图 3 所示。

4. 实验模型的评估

在模型的评估上，本文引入了混淆矩阵。混淆矩阵可以用来作为分类规则特征的表示，它包括了每一类的正确分类样本数和错误分类样本数。

对于 n 类的分类问题，误差可能有 n^2-n 类，如果仅有两类(正样本和负样本，用 T 和 F 来象征性地代表)，就只有两类误差，期望为 T，但分类为 F，称为假负，期望为 F，但分类为 T，称为假正。另外，期望为 T，但分类为 T，称为真正，期望为 F，但分类为 F，称为真负。将把它们汇总在表 2 正负样本的混淆矩阵中，如表 2 所示。

考虑这样一个分类问题：所有样本都必须用一个可能的类进行标记。为此引入 5 个参数：敏感性(Sensitivity)、特异性(Specificity)、精度(Precision)、错误正例(False positives)和错误负例(False Negatives) [15]。这些度量定义为：

$$\text{敏感性} = \frac{t\_pos}{t\_pos + f\_pos} \tag{3}$$

$$\text{特异性} = \frac{t\_neg}{t\_neg + f\_neg} \tag{4}$$

$$\text{精度} = \frac{t\_pos}{t\_pos + t\_neg} \tag{5}$$

$$\text{错误正例} = \frac{f\_neg}{t\_neg + f\_neg} \tag{6}$$

$$\text{错误负例} = \frac{f\_pos}{t\_pos + f\_pos} \tag{7}$$

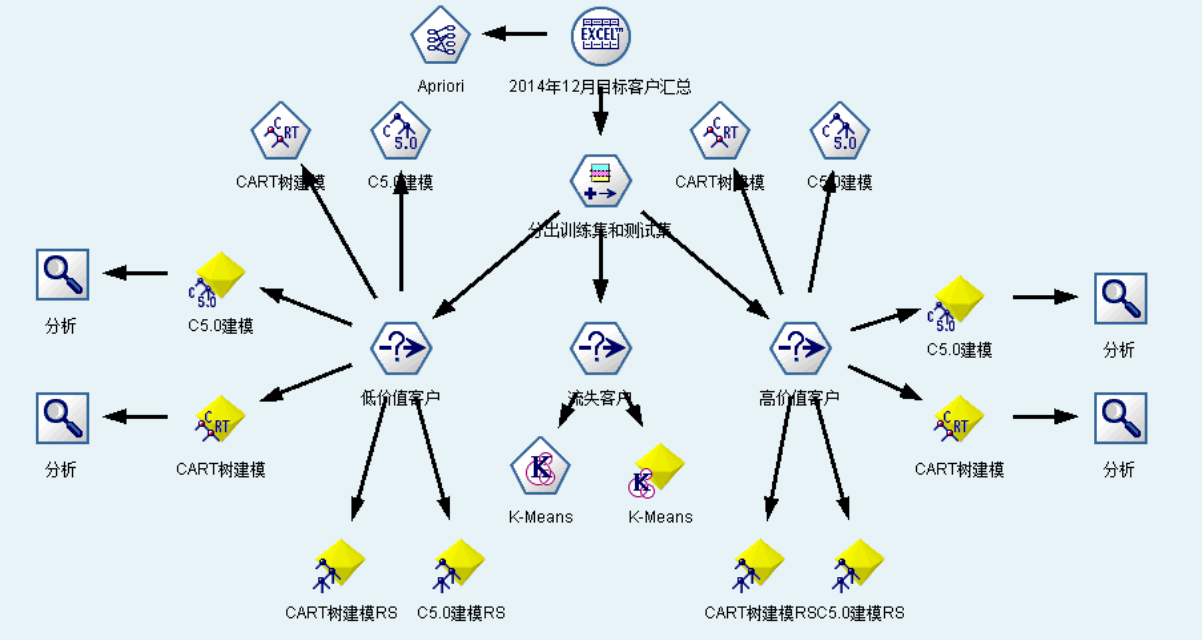

Figure 3. Customer churn prediction model and churn customer clustering model flow chart
图 3. 客户流失预测模型和流失客户聚类模型流程图

Table 2. Confusion matrix
表 2. 混淆矩阵

	T 预测数	F 预测数
T 实际数	真正	假正
F 实际数	假负	真负

其中，敏感性和特异性对分类器识别正负样本的能力做出评估[16]，*t_pos* 和 *f_pos* 分别是真正和假正样本个数，*t_neg* 和 *f_neg* 指的是真负和假负样本个数，最终准确率可定义为：

$$A = \frac{t\_pos + t\_neg}{t\_pos + f\_pos + t\_neg + f\_neg} \tag{8}$$

通过模型输出结果，如表 3 至表 6 所示。

由公式(3)、(4)、(5)、(6)、(7)、(8)可以计算出低/高价值客户流失预测的敏感性、特异性、精度、错误正例、错误负例，如表 7，表 8 所示。

Table 3. Forecast confusion matrix of C5.0-low value customers
表 3. C5.0-低价值客户预测混淆矩阵

低价值客户	预测流失客户	预测非流失客户
实际流失客户	1218	255
实际非流失客户	229	1984

Table 4. Forecast confusion matrix of C5.0-high value customers
表 4. C5.0-高价值客户预测混淆矩阵

高价值客户	预测流失客户	预测非流失客户
实际流失客户	2141	170
实际非流失客户	174	1983

Table 5. Forecast confusion matrix of CART-low value customers
表 5. CART-低价值客户预测混淆矩阵

低价值客户	预测流失客户	预测非流失客户
实际流失客户	1161	312
实际非流失客户	283	1930

Table 6. Forecast confusion matrix of CART-high value customers
表 6. CART-高价值客户预测混淆矩阵

高价值客户	预测流失客户	预测非流失客户
实际流失客户	2145	166
实际非流失客户	246	1911

Table 7. Churn prediction assessment of C5.0-low/high value customers
表 7. C5.0-低/高价值客户流失预测评估

度量值	敏感度	特异性	精度	错误正例	错误负例	准确率
低价值客户	82.68%	89.65%	38.03%	10.34%	17.31%	86.83%
高价值客户	92.64%	91.93%	51.91%	8.06%	7.35%	92.30%

Table 8. Churn prediction assessment of CART-low/high value customers
表 8. CART-低/高价值客户流失预测评估

度量值	敏感度	特异性	精度	错误正例	错误负例	准确率
低价值客户	78.81%	87.21%	37.56%	12.78%	21.18%	83.85%
高价值客户	92.81%	88.59%	52.88%	11.40%	7.18%	90.98%

从上表可以看出，C5.0 决策树的预测准确性分别为低价值客户 86.83%，高价值客户 92.30%；CART 决策树的预测准确性分别为低价值客户 83.85%，高价值客户 90.98%。

由此可以看出，C5.0 决策树更加适合该县电信运营商的客户流失预测。而两种决策树对低价值客户的预测准确率都在 85%左右，不算太高，可能和选取的客户数据完整性有关，电信运营商对低价值客户资料的登记存在不完善之处，因此电信运营商需要对低价值客户资料的录入进行完善和重视。

5. 实验结果分析

5.1. 客户流失预测模型规则集分析

低价值客户的流失规则集，描述如下：

规则 1：如果客户在网时间小于 1 年，并且是 3G 客户，并且有欠费，则客户可能会流失；

规则 2：如果客户在网时间小于 1 年，并且是 3G 客户，无欠费，第三个月流量值小于 7.8 M，并且有预存，话费波动大于 0.28，则客户可能会离网；

规则 3：如果客户在网时间小于 1 年，不是 3G 客户，有欠费，无预存，话费波动小于 0.39，则客户可能流失；

规则 4：如果客户在网时间大于 1 年，不是 3G 客户，有欠费，是敏感客户，并且流量波动小于 0.67，则客户可能会流失；

规则 5：如果客户在网时间大于 1 年，不是 3G 客户，无欠费，第三月流量值大于 92 M，流量波动大于 16.28，则客户可能流失。

高价值客户的流失规则集，描述如下：

规则 1：如果客户在网时间小于 1 年，并且是 3G 客户，并且有欠费，并且第三个月流量值小于 104 M，话费波动小于 0.47，则客户可能会流失；

规则 2：如果客户在网时间小于 1 年，不是 3G 客户，则客户可能流失；

规则 3：如果客户在网时间大于 1 年，无预存，不是 3G 客户，有欠费，第三个月消费小于 121 元，流量波动小于 0.404，则客户可能流失；

规则 4：如果客户在网时间大于 1 年，不是 3G 客户，有欠费，话费波动小于 0.533，流量波动大于 9.7，则客户可能流失；

规则 5：如果客户在网时间大于 1 年，是 3G 客户，有欠费，流量波动小于 0.404，则客户可能流失。

分析两类客户的离网判定规则可以得出流失客户的一些共同属性：在网时间不足 1 年，是 3G 客户，流量值较低的客户可能流失；在网时间大于 1 年，不是 3G 客户，流量值波动较大，有欠费的客户可能流失。

5.2. 流失客户聚类模型规则集分析

K-Means 聚类算法将不同规则下的流失客户分成了四类，如表 9 所示。

根据表 9 的数据可以得出每个聚类的特点如下：

聚类 1：为高价值客户，在网时间大于 1 年，话费波动为 0.879，流量波动为 233.005，非 3G 客户，非敏感客户，有欠费，有预存；

聚类 2：为低价值客户，在网时间小于 1 年，话费波动为 0.842，流量波动为为 312.481，非 3G 客户，是敏感客户，无欠费，无预存；

聚类 3：为高价值客户，在网时间小于 1 年，话费波动为 1.082，流量波动为 316.937，是 3G 客户，非敏感客户，有欠费，无预存；

Table 9. Clustering statistics of customer loss
表 9. 流失客户各聚类统计

聚类	聚类 1 1061 条记录	聚类 2 946 条记录	聚类 3 6729 条记录	聚类 4 4793 条记录
话费波动	0.879	0.842	1.082	0.173
流量波动	233.005	312.481	316.937	189.802
客户价值	高价值客户	低价值客户	高价值客户	低价值客户
是否 3G 客户	否	否	是	是
是否敏感客户	否	是	否	是
是否欠费	是	否	是	是
是否预存	是	否	否	是
在网时间	大于 1 年	小于 1 年	小于 1 年	大于 1 年

聚类 4：为低价值客户，在网时间大于 1 年，话费波动为 0.173，流量波动为 189.802，是 3G 客户，是敏感客户，有欠费，有预存。

6. 总结

从上文中客户流失预测模型得出的结果可以看出：对于低价值客户，如果有以下特征则容易产生流失：

1) 在网时间小于 1 年，且为 3G 客户，流量值波动明显；

2) 在网时间大于 1 年，不是 3G 客户，有欠费，流量值波动大于 16.2。

对于高价值客户，如果有以下特征则容易产生流失：

1) 在网时间大于 1 年，是 3G 客户，有欠费，流量使用值较小；

2) 在网时间小于 1 年，不是 3G 客户。

因此对于有以上特征的低价值和高价值客户需要客户维系人员重点关注，建立每月的客户数据监测机制，及时开展客户维系工作。

从流失客户聚类模型中，可以得出：

聚类 1：主要是在网时间大于 1 年的高价值客户，非 3G 客户，有欠费，有预存，流量波动较大。该聚类一共 1061 人，属于老客户群体。这类客户可能会因高流量费用导致流失。因此，可以针对这类客户给予一些免费流量的优惠政策；

聚类 2：主要为低价值客户，在网时间小于 1 年，非 3G 客户，有欠费，无预存，流量波动明显。该聚类一共 946 人，分析属于一些外出务工反乡人员。这类客户的流失风险较低，电信运营商可以针对这部分客户群进行一些小额话费预存优惠政策进行客户挽留；

聚类 3：主要是高价值客户，在网时间小于 1 年，是 3G 客户，有欠费，有预存，流量波动巨大。该类一共 6729 人，属于年末重点发展的 3G 促销客户。这类客户一般是收入稳定的工薪阶层，刚刚接触 3G 业务，可能因为一些高收费的 3G 业务导致流失。因此，针对这类客户需要电信运营商做好 3G 业务和流量优惠政策宣传，并敦促这类客户尽快办理 3G 流量套餐包；

聚类 4：主要是在网时间大于 1 年的低价值客户，非 3G 客户，流量波动不明显，话费波动明显，无欠费，无预存。该类一共 4793 人，属于较为稳定的 2G 客户群体，这类客户流失主要原因还是传统的语

音话费问题，针对这类客户，电信运营商可以推广一些优惠的预存赠话费活动来挽留客户。

基金项目

中国博士后科学基金面上一等资助项目(编号：2015M570249)；辽宁省高等学校优秀人才支持计划资助(编号：WJQ2014040)。

参考文献 (References)

[1] 许恺. 基于数据挖掘技术的电信客户流失预测[J]. 电脑知识与技术, 2009, 5(13): 3437-3438.

[2] 熊平. 数据挖掘算法与 Clementine 实践[M]. 北京: 清华大学出版社, 2011.

[3] 夏国恩. 客户流失预测的现状与发展研究[J]. 计算机应用研究, 2010, 27 (2): 413-416.

[4] Sato, T., Huang, B.Q., Huang, Y., Kechadi, M.-T. and Buckley, B. (2010) Using PCA to Predict Customer Churn in Telecommunication Dataset. *Lecture Notes in Computer Science*, **6132**. http://dx.doi.org/10.1007/978-3-642-13217-9

[5] Cox Jr., L.A. (2002) Data Mining and Causal Modeling of Customer Behaviors. *Telecommunication Systems*, **21**, 349-381.

[6] Rosset, S. and Neumann, E. (2003) Integrating Customer Value Considerations into Predictive Modeling. *Proceedings of the* 3*rd IEEE International Conference on Data Mining*, Washington DC, 19-22 November 2003, 283-290.

[7] Wojewnik, P., Kaminski, B., Zawisza, M. and Antosiewicz, M. (2011) Social-Network Influence on Telecommunication Customer Attrition. *Lecture Notes in Computer Science*, **6682**, 64-73.

[8] Mozer, M.C, Wolniewicz, R., Grimes, D.B., Johnson, E. and Kaushansky, H. (2000) Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry. *IEEE Transactions on Neural Networks*, **11**, 690-696.

[9] 郭亮. 用 CRISP-DM 模型来规范企业数据中心建设[J]. 华北科技学院学报, 2008, 5(4): 69-72.

[10] 龙志勇. 数据挖掘在电信行业客户关系管理中的应用[J]. 信系网络, 2003(12): 24-26.

[11] 袁玉波. 数据挖掘与最优化技术及其应用[M]. 北京: 科学出版社, 2007.

[12] 李如平. 数据挖掘中决策树分类算法的研究[J]. 东华理工大学学报(自然科学版), 2010, 33(2): 192-196.

[13] 梁循. 数据挖掘算法与应用[M]. 北京: 北京大学出版社, 2006.

[14] 薛薇. 基于 Clementine 的数据挖掘[M]. 北京: 中国人民大学出版社, 2012.

[15] Dunhan, M.H. 数据挖掘教程[M]. 北京: 清华大学出版社, 2005.

[16] Han, J.W. and Kamber, M., 著. 数据挖掘: 概念与技术[M]. 第二版. 范明, 孟小峰, 译. 北京: 机械工业出版社, 2007: 147-154.