

生存分析在客户流失预测模型中的应用

——以某运营商的联通客户为例

骆 杨

西南大学数学与统计学院, 重庆
Email: 1556847579@qq.com

收稿日期: 2021年1月2日; 录用日期: 2021年1月28日; 发布日期: 2021年2月4日

摘 要

文章通过对某运营商的联通客户历史数据的研究, 试图分析各个因素对客户流失的影响, 并通过逻辑回归和生存分析模型, 进一步探究这些因素与客户流失的相关性。同时利用描述分析和模型分析, 研究了不同因素水平对客户流失的影响, 掌握了客户流失的情况, 从而对如何有效地预防联通客户的流失提出了有针对性的建议。

关键词

客户流失, 生存分析, COX模型, 逻辑回归, 流失预测

Application of Survival Analysis in Customer Churn Prediction Model

—A Case of China Unicom Customer

Yang Luo

School of Mathematics and Statistics, Southwest University, Chongqing
Email: 1556847579@qq.com

Received: Jan. 2nd, 2021; accepted: Jan. 28th, 2021; published: Feb. 4th, 2021

Abstract

Through the research on the historical data of China Unicom's customers, the article attempts to analyze the impact of various factors on customer churn, and further explores the correlation between these factors and customer churn through logistic regression and survival analysis models.

Using description analysis and model analysis, the impact of different levels of each factor on customer churn was studied, and the situation of customer churn was mastered, thus providing targeted recommendations on how to effectively prevent the loss of Unicom customers.

Keywords

Customer Churn, Survival Analysis, Cox Model, Logical Regression, Churn Prediction

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

手机作为人们日常通信的必备工具，正在发挥着越来越大的作用。国内运营商在 2008 年实现重组，构成电信、移动、联通三大运营商。随着国内市场环境逐渐成熟，三大运营商的通信业务和产品营销模式日益趋同，竞争压力越来越大，高新增客户率已经成为过去时，与此同时高客户流失率已成为非常普遍的现象。在移动通信市场，联通面临着移动、电信两方面的强大挑战[1]。如何提高客户满意度，降低客户流失率，有效发展客户，进而提升收入，成为运营商经营管理的重要研究内容。下面给出从 2011 年到 2015 年(近 5 年)中国移动、中国联通和中国电信的移动客户数增长趋势图。

从图 1 中我们可以看到用户增长十分缓慢，市场已经出现饱和状态，中国联通甚至在 2015 年时出现了负增长。如何保留现有客户，避免客户流失成为很多运营商亟待解决的问题。

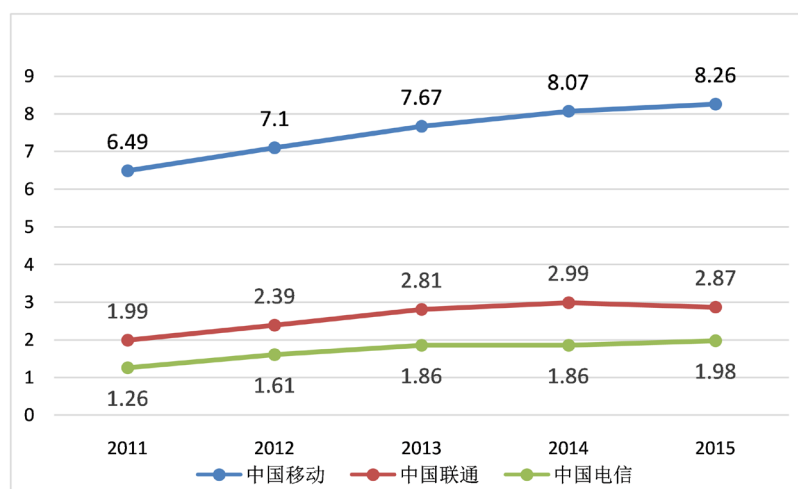


Figure 1. The number of customers of mobile communication company from 2011 to 2015 (unit: 100 million people)

图 1. 2011~2015 年各移动通信公司客户数(单位: 亿人)

事实证明，挽留住老客户和发展、挖掘新客户相比是更重要的。Kotler (1994)研究结果显示，发展一位新客户的成本是挽留一位老客户成本的 5 倍，且挽留老客户给企业带来的利率为发展新客户成本的 16 倍，所以减少客户流失等于用更少的成本减少利润的流失[2]。由此，我们知道建立在客户关系的维护和管理基础上，怎样挽留住客户对联通服务公司的生存和发展是非常重要的。2014 年三家运营商主要运营数据比较。

从表 1 的 2014 年三家运营商的主要运营数据看, 在营收方面, 中国移动、中国联通、中国电信在 2014 年度总营收分别为 6414 亿元、2846.8 亿元、3243.9 亿元, 总和为 12504.7 亿元, 其中移动占比 51.3%。同 2013 年相比, 2014 年三大运营商营收增幅放缓。在客户用户量方面, 2014 年中国移动用户为 8 亿户, 比 2013 年增长 3493 万户, 增长率为 4%; 中国电信用户数达到 1.86 亿户, 比 2013 年增长 4 万户, 增长率为 0%; 中国联通用户数达到 3 亿户, 比 2013 年增长 1810 万户, 增长率为 6.4%。三家运营商用户增长率下降明显。

Table 1. The comparison of main operation data of three operators in 2014
表 1. 三家运营商 2014 年主要运营数据比较

企业名称	总运营收入(亿元)	总运营收入增长率	利润(亿元)	总客户数(亿)	总客户增长率
中国移动	6414	1.8%	1094	8	4%
中国电信	3243.9	0.9%	177	1.86	0%
中国联通	2846.8	3.5%	121	3	6.4%

数据来源: 《2014 中国移动公司财务报表分析》。

在市场竞争越来越激烈的趋势下, 运营商实现企业利润同步增长, 扩大规模, 客户资源是三家运营商竞争的核心资源。因此, 优秀的运营商一定要懂得如何留住自己的客户。那么如何有效预测出潜在流失客户, 提升客户挽留工作效率成为我们关注的问题。目前对客户流失预测问题的研究十分广泛, 比如决策树算法, 应用最广泛[3]。决策树算法分类准确率较高, 建模也较简单、分类准确率高, 而且它能导出比较简明易懂分类规则, 但是它也有一定的缺点, 此外, 很多专家学者提出用 Logistic 回归[4]和人工神经网络[5]等研究方法建立模型, 但是, 由于客户流失数据集本身的特殊性, 这些方法使用起来都有值得商榷之处。

鉴此, 本文根据逻辑回归和生存分析方法的基本概念, 建立客户流失场景下的逻辑回归模型和生存分析模型, 并结合 2012 年到 2014 年间某运营商的联通客户的相关数据进行了具体分析, 构建了一个客户流失模型。且实践证明, 这个模型对于指导企业进行客户流失管理是有效的。

2. 数据介绍

本文采用某运营商的真实数据, 共有 4975 条有效观测。每条观测来自于一个手机号码的某个年度。

1) 使用月数

这是我们关心的因变量, 它记录的是截止到观测结束(2012.01~2014.01)某个手机客户使用联通服务的时间长短, 以月记。这里并不代表使用月数就是我们所说的生存时间。因为对于那些已经没有使用联通服务的客户来说, 该变量确实代表了他的生存时间; 但是, 对于那些还在继续使用联通服务的客户来说, 无法确定在未来的某个时间他们也会不再继续使用联通服务, 他们的生存时间大于或者等于已经实现了的使用联通服务的时间。

2) 流失用户

这是一个 0-1 变量。“1”代表是, 即在 25 个月的观测期内, 用户已经流失了, 相反“0”代表否, 表示该用户尚未流失。使用月数和流失用户一起构成了我们的因变量。

为了能有针对性地制定营销策略, 采取行动让客户满意, 留住客户, 将最好的客户留住更长的时间, 以提升客户存在期的价值, 最终达到减少客户流失的目的。我们需要探究是什么是导致客户流失的原因, 谁是潜在的流失客户等问题, 从而建立有效的客户流失预测模型。我们将考虑以下几个指标, 并对选择其作为解释性变量的原因进行详细的讨论。

3) 套餐金额

这是一个定性变量，它表示用户购买的月套餐的金额，1 为 96 元以下，2 为 96 到 225 元，3 为 225 元以上。2008 年通信行业完成重组后，三大运营商提供的服务和价格越来越同质化。这样一来，价格将在一定程度上左右客户对运营商的选择，如果客户认为该套餐费与其他营运商的套餐费相比没有优势，将导致客户转投到其它运营商。

4) 改变行为

这是一个定性变量，它表示的是用户是否曾经改变过套餐金额，“1”代表是，“0”代表否。

5) 额外通话时长

这是一个定量变量，记录的是用户的实际通话时长减去套餐内包含的通话时长，得出用户在使用期间的每月额外通话时长，这部分需要用户额外交费。数值是每月的额外通话时长的平均值，以分钟记。用户的通话时长越长，表明该用户对运营商的依赖性越强，所以流失的可能性越小。

6) 额外流量

这是一个定量变量，记录的是用户使用的实际流量减去套餐内包含的流量，得出用户在使用期间的每月额外流量，这部分需要用户额外交费。数值是每月的额外流量的平均值，以兆记。同通话时长类似，用户使用流量越多，用户对运营商的依赖性越强，流失的可能性越小。

7) 服务合约

服务合约是相对于传统的大众促销而提出的崭新模式，是针对特定顾客群体而进行的促销活动，这种营销往往针对性较强，而且对顾客的个性化服务和优惠幅度较大，深受顾客的青睐。运营商为顾客制定合适的服务条约，会产生较少的顾客流失。因此我们选择这一指标作为解释变量，该指标主要包括两个水平 0, 1。“1”代表是，即用户与联通签订过服务合约；“0”代表否，即用户没有与联通签订过服务合约。

8) 关联购买

这是一个定性变量，它表示用户在使用联通移动服务过程中是否还同时办理其他业务(主要是固定电话和宽带业务)，“1”代表同时办理一项其他业务，“2”代表同时办理两项其他业务，“0”代表没有办理其他业务。消费者希望能在同一服务提供商处购买到所需要全部产品和服务。运营商通常会将家庭的固定电话、宽带及移动电话捆绑销售，并且对二者之间的通信提供优惠。这样，大大提高了选择关联产品的顾客的稳定性，客户流失减少。

9) 集团用户

运营商会为批量购买的集团用户给与内部通话方面的优惠，即相比个人业务，集体办理的号码在集团内拨打有一定优惠。该指标包含了两个水平，“1”代表是，即用户办理的是集团业务；“0”代表否，即用户办理的不是集团业务。

3. 描述分析

3.1. 简单描述分析

我们首先对数据做简单的描述分析。本数据的总样本量为 4975，其中 3894 的客户在观测期间内流失，剩下的 1081 位客户被截断了，截断比例为 21.73%。整个样本的生存函数如图 2 所示。从图中可以看出，随着使用月数的增加，客户的流失率也在增加，其中在 12 月份产生较大的流失。我们猜测可能是受客观环境的影响，中国的传统节日春节在 12 月份，我们猜测，在该月，大量使用联通服务的用户纷纷回老家过年，手机使用会暂停，甚至可能取消。春节过后，大家回到工作地点，又重新考虑继续使用或者停止使用。

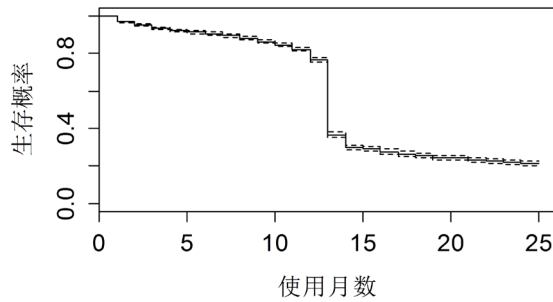


Figure 2. The survival function of all samples
图 2. 全部样本生存函数

下面，我们尝试对各个解释变量分组，重新计算生存函数，并做对比分析。首先对用户购买的月套餐金额和用户是否曾经改变过套餐金额做分析，如图 3 所示。

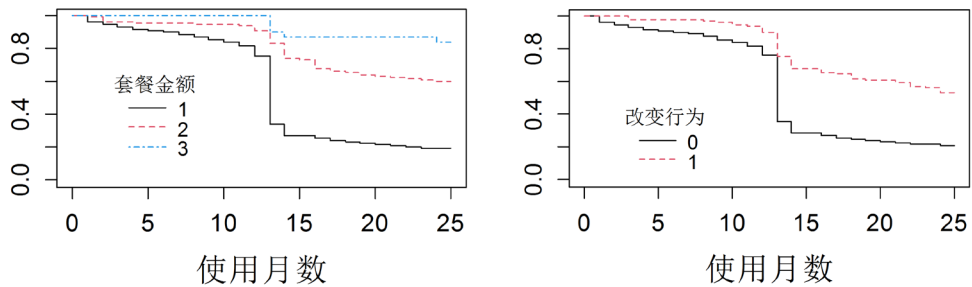


Figure 3. The survival function of package amount and change behavior
图 3. 套餐金额和改变行为生存函数

我们可以看出，在不同的月套餐金额下，客户的流失情况也不同：相比月套餐金额高的客户，月套餐金额低的联通客户流失严重。月套餐高的客户(月套餐金额 = 3)稳定性较好，客户保留程度高。同时可以看出，用户是否曾经改变过套餐金额对于客户保留有明显的影响，改变过套餐行为比未改变过套餐行为的用户有更长的生存时间，未改变过套餐的用户中有 50% 的客户至少使用了 13 个月，而改变过套餐的用户约有 75% 的人至少使用了同样长的时间。这是因为，通过改变套餐，使得客户办理了更为匹配自己需求的套餐，因此客户更为忠诚。

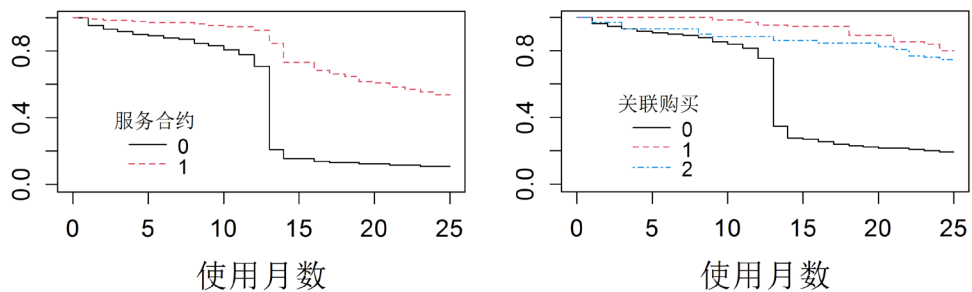


Figure 4. The survival function of service contract and associated purchase
图 4. 服务合约和关联购买生存函数

图 4 表明，在用户是否和联通签订了服务合约和用户在使用联通移动服务过程中是否还同时办理其他业务的影响下，顾客的保留程度也不同。相比没有服务合约的顾客(服务合约 = 0)，与联通签订了服务合约的顾客(服务合约 = 1)的保留情况明显较好。不同的关联购买情况下，顾客流失情况也不同：相比关

联购买多的客户，没有关联购买的客户流失情况严重，而同时办理了固定电话或者宽带其中一项业务的客户(关联购买 = 1)稳定性最好，保留程度高。

图 5 表明，是否办理集团业务的情况下，顾客的流失程度也会不同：相比没有办理集团业务的用户，办理的是集团业务的用户保留率高，流失率低。办理集团业务的用户(集团用户 = 1)稳定性好，保留程度高。

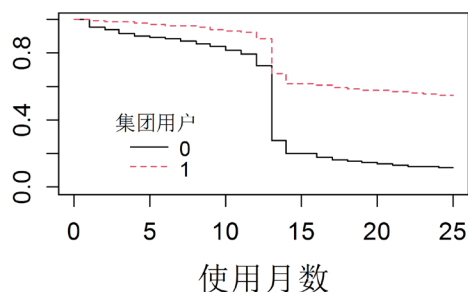


Figure 5. The survival function of group user
图 5. 集团用户生存函数

最后，把所有用户按照额外通话时长分为两组。其中额外通话时长为 1 表示用户的该月的实际通话时长大于套餐内包含的通话时长，额外通话时长为 0 表示客户没有超过套餐内包含的通话时长。同理，额外流量也分为两组，额外流量为 1 表示用户的使用的实际流量大于套餐内包含的流量，额外流量为 0 表示。结果如图 6 所示：额外通话时长和额外流量均超过了套餐包含量的联通客户的稳定性好，保留程度高。这是因为，额外通话时长或者流量的客户与亲朋好友的交流更多，社交需求更大，保持稳定的联系方式对他们来说更为重要，换号对其来说更为不便，因此离网率相对更低。

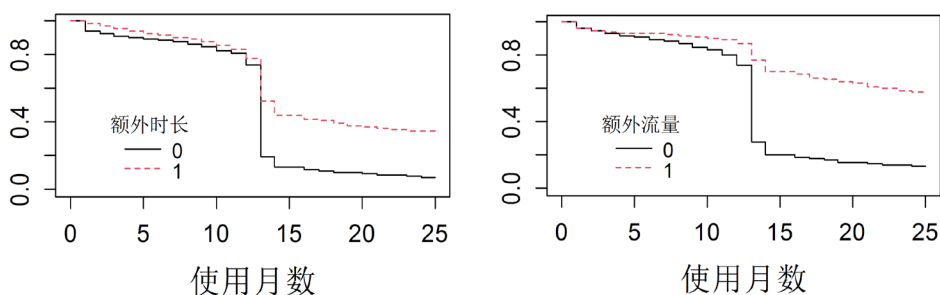


Figure 6. The survival function of extra call duration and extra traffic
图 6. 额外通话时长和额外流量生存函数

3.2. 流失用户特征分析

流失用户特征分析是指在流失用户中分析在网时长、套餐结构、改变行为、服务和约、关联购买、集团购买、额外通话时长和额外流量等不同因素下流失的比例。

1) 流失用户使用月数分析

结论：从表 2，我们可以看出使用月数在 1 年到 2 年之间的用户流失占比较高，占了总流失用户的 70% 左右。

2) 流失用户套餐金额和改变行为结构分析

结论：从表 3，我们可以看出用户流失大部分都集中在套餐金额为 96 元以下的，占了总流失数的 97.56%。且大部分流失用户在使用联通服务过程中没有改变套餐。

Table 2. The structure of months of lost customers

表 2. 流失用户使用月数结构表

使用月数	流失用户数	占比
3 月以下	334	8.58%
3~6 月	139	3.57%
6~12 月	700	17.98%
12~24 月	2721	68.88%
24 月以上	0	0%

Table 3. The analysis of package amount and change behavior structure of lost customers

表 3. 流失客户套餐金额和改变行为结构分析

因素名称	水平	流失用户数	占比
套餐金额	96 元以下(1)	3799	97.56%
	96~225 元(2)	90	2.31%
	225 元以上(3)	5	0.13%
改变行为	改变(1)	49	1.26%
	未改变(0)	3845	98.74%

3) 流失用户关联购买和服务合约分析

Table 4. The analysis of package amount and change behavior structure of lost customers

表 4. 流失客户关联购买和服务合约结构分析

因素名称	水平	流失用户数	占比
关联购买	未办理业务(0)	3859	99.10%
	同时办理一项(1)	15	0.39%
	同时办理两项(2)	20	0.51%
服务合约	签订(1)	554	14.23%
	未签订(0)	3340	85.77%

结论：从表 4，我们可以看出流失客户主要集中在使用联通移动服务过程中，固定电话和宽带业务都没有办理的，占了总流失数的 99.10%。没有签订服务合约的客户占了总流失数的 85.77%。

4) 流失用户集团业务分析

Table 5. The analysis of the business structure of lost customers

表 5. 流失客户集团业务结构分析

集团业务	流失用户数	占比
是(1)	509	13.07%
不是(0)	3385	86.93%

结论：从表 5，我们可以看出用户流失主要集中在办理的不是集团业务的用户，占了总流失数的 86.93%。

5) 流失用户额外通话时长和额外流量分析

我们仍按照上面的分类，把所有用户按照额外通话时长(流量)分为两组。其中额外通话时长(流量)为 1 表示用户的该月的实际通话时长(流量)大于套餐内包含的通话时长(流量)，额外通话时长(流量)为 0 表示客户没有超过套餐内包含的通话时长(流量)。

Table 6. The analysis of extra call duration and extra traffic of lost customers

表 6. 流失用户额外通话时长和额外流量分析

因素名称	水平	流失用户数	占比
额外通话时长	0	2202	56.55%
	1	1692	43.45%
额外流量	0	3514	90.24%
	1	380	9.76%

结论：从表 6，我们可以看出就额外通话时长而言，用户流失在实际通话时长大于或小于套餐内包含的通话时长占比差不多，而就额外流量而言，用户流失主要集中在没有超过套餐内包含的流量的用户。

4. 模型分析

4.1. 逻辑回归分析

首先，顾客使用联通服务到最后流失是一个系列的整体过程。我们关注变量是否对联通服务行业客户的流失产生影响，我们将数据分为测试集和训练集，随机筛选数据的 70%为训练集，剩余 30%为测试集。对训练集利用逻辑回归得出结果，回归的伪 R^2 为 0.335。对训练集建立模型的参数估计如下：

Table 7. The logistic regression analysis of influencing factors of lost customers

表 7. 用户流失的影响因素逻辑回归分析

因素名称	参数估计	标准误差	卡方统计量	p-值
截距项	3.996	0.193	20.666	***
套餐金额	-1.218	0.162	-7.523	***
额外通话时长	-0.0003	0.000	-4.313	***
额外流量	-0.001	0.000	-7.005	***
改变行为	-1.151	0.3048	-3.777	***
服务合约	-1.730	0.1073	-16.121	***
关联购买	-1.972	0.1765	-11.173	***
集团用户	-1.794	0.1072	-16.739	***

注：***为 0.001 显著性水平。

根据表 7 的回归结果，我们可以看出所有解释变量对客户流失的影响均是显著的(P -值 < 0.0001)，而且所有解释变量的极大似然估计量均为负，说明：1) 用户购买的月套餐金额越高，客户流失的可能性越小。2) 联通服务提供的套餐内包含的通话时长和流量与用户该月所需的实际通话时长和流量匹配程度越低，客户流失的概率越大。3) 在联通服务中，用户主动改变套餐金额的行为会负向影响顾客流失，即与不主动去改变套餐金额的顾客相比，修改过套餐金额的客户流失的可能性更低。4) 服务合约对顾客流失

产生负向影响，与没有服务合约的用户相比，有过服务合约的用户流失可能性更低。5) 用户在使用联通移动服务过程中还同时办理其他业务(主要是固定电话和宽带业务)，用户流失的可能性越低。6) 与用户没有办理集团业务相比办理集团业务的用户流失的概率更小。

4.2. 逻辑回归模型评估

基于上述逻辑回归模型，我们将模型应用于测试集，对测试集数据进行预测，预测结果如下：

Table 8. The prediction results of logistic regression analysis

表 8. 逻辑回归预测结果

预测值	0	1	合计
0	262	54	316
1	274	903	1177
合计	536	957	1493

从表 8，我们计算了错判率为 17%，预测效果较好。

对于预测结果，我们绘制了 ROC 曲线如下：

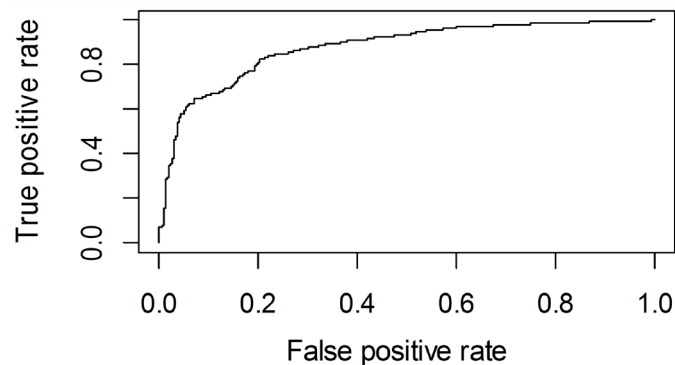


Figure 7. ROC curve

图 7. ROC 曲线

通常理想的 ROC 曲线应紧贴左上角，即当 ROC 曲线下的面积越大，我们的分类越好。由图 7 我们可以看到，该模型具有较好的预测效果。

4.3. COX 模型

我们通 COX 等比例模型对各个因素同使用月数之间的关系做模型分析[6]，结果如表 9 所示。

从表 9 可以看到，所有的预测协变量在显著水平为 0.001 下均为显著的。在显著水平为 0.001 下，套餐金额、额外通话时长、额外流量、改变行为、服务合约、关联购买、集团用户等 7 个因素的回归系数均为负值，与客户流失呈负相关关系。且从表 8 知：月套餐金额为 96 元以下的用户流失的可能性分别是 96 到 225 元和 225 元以上的客户的 1.96 倍和 10.81 倍；客户没有超过套餐内包含的通话时长流失的可能性大于套餐内包含的通话时长的 1.379 倍；用户没有超过套餐内包含的流量流失的可能性大于套餐内包含的流量的 2.66 倍；未改变行为流失的风险是改变行为的 1.779 倍；无服务合约流失的可能性是有服务合约的 3.11 倍；没有关联购买的客户流失的可能性分别是关联购买了 1 项和 2 项的 8.13 倍和 6.31 倍；用户没有办理集团业务流失的概率是办理集团业务的 1.95 倍。

Table 9. The parameter estimation and test results
表 9. 参数估计以及检验结果

因素名称	水平	参数估计(B)	Exp(B)	Exp(-B)	标准误差	P-值
套餐金额	2	-0.673	0.510	1.960	0.109	***
	3	-2.381	0.093	10.810	0.448	***
额外通话时长	1	-0.321	0.725	1.379	0.035	***
额外流量	1	-0.979	0.376	2.661	0.057	***
改变行为	1	-0.576	0.562	1.779	0.144	***
服务合约	1	-1.135	0.321	3.112	0.0496	***
关联购买	1	-2.096	0.123	8.130	0.259	***
	2	-1.842	0.158	6.310	0.225	***
集团用户	1	-0.668	0.513	1.950	0.050	***

注：***为 0.001 显著性水平。

5. 总结讨论

5.1. 结果讨论

本文分析了 4975 个用户使用联通服务的使用月数情况，研究了联通用户的使用月数同月套餐金额、额外通话时长、额外流量、改变行为、服务合约、关联购买、集团用户之间关系。从结果发现，月套餐金额越高的用户使用联通服务的月数更长；用户使用联通服务过程中该月的通话时长或流量超过了月套餐内包含的通话时长或流量，则他们使用联通服务的月数更长；用户有更改过套餐金额的使用联通服务的月数更长；用户没有关联购买联通的固定电话或者宽带的更容易流失；用户与联通签订的服务合约的忠诚度更高；用户办理了集团业务的使用联通的月数更长。

5.2. 管理建议

首先，在对联通客户进行流失预测时，联通服务套餐所提供的通话时长和流量与客户实际所需要的通话时长和流量是否匹配是值得注意的。当联通用户该月所用的通话时长和流量与套餐提供的相比超过了或者还没有用完，都会影响客户流失。套餐提供的服务和顾客所需的匹配度越低，说明联通服务不能很好地满足客户的需求。所以我们应有警示，及时与客户沟通交流，了解他们是否需要改善服务，从而减少联通用户的流失。

其次，主动更改过套餐类型、用户关联购买了联通其他相关业务的顾客保留程度会更高。所以联通公司可以积极地开展一些宣传活动，让用户更加深入了解联通的套餐类型、服务类型和其他一些相关业务的优惠活动，实施一些有针对性的捆绑销售，这样可以增加顾客忠诚度，减少客户流失。

最后，服务合约对用户流失也有影响。联通企业应该提出一些针对性强、个性化多的服务合约，让签了服务合约的用户在合理的价格套餐内能得到更多的优惠，提高顾客保留量。

参考文献

- [1] 付杰. 移动通信行业客户流失预测与挽留研究[D]: [硕士学位论文]. 合肥: 合肥工业大学, 2015.
- [2] 王汉生. 商务数据分析与应用[M]. 北京: 中国人民大学出版社, 2011.
- [3] 盛昭瀚, 柳炳祥. 客户流失危机分析的决策树方法[J]. 管理科学学报, 2005, 8(2): 20-25.

- [4] 石永华. 电信业务流失建模的研究[J]. 广东通信技术, 2003(6): 15-20.
- [5] 管东升. 移动通信业客户流失行为预测技术的研究[D]: [硕士学位论文]. 大连: 大连理工大学, 2005.
- [6] 邓森文, 马溪骏. 基于 Cox 模型的移动通信行业中低端客户流失预测研究[J]. 合肥工业大学学报(自然科学版), 2010, 33(11): 1698-1701.