



自我网络特征 对电信客户流失的影响

周静¹, 周小宇², 王汉生³

1 中国人民大学 统计学院, 北京 100872

2 上海科技大学 创业与管理学院, 上海 201210

3 北京大学 光华管理学院, 北京 100871

摘要:近年来,随着移动通信行业的蓬勃发展,市场饱和度越来越高,企业获取新用户的成本也越来越大。随着中国三大运营商竞争的加剧,产品和服务的同质化程度也越来越高,这使企业在老客户的保留上变得异常困难,客户流失率也在逐年上升,如何识别高风险流失客户并有效防止客户流失已经成为该行业管理者普遍关心的问题之一。

着眼于客户流失影响因素研究,运用社交网络分析方法,通过构造与网络结构相关的变量进行影响因素的探讨,运用逻辑回归方法构建客户流失预警模型。从社交网络的视角出发,利用客户的通话详单数据建立客户之间的通信网络,在自我网络的相关理论框架下,构建个体的度、联系强度、个体的信息熵3个自我网络特征变量。运用中国某移动运营商公司的月度客户数据(包括基础通信数据和通话详单数据),通过逻辑回归构建基于社交网络变量的客户流失预警模型。

研究表明,个体的度、联系的强度和个体的信息熵都对预测客户流失有显著效果。具体的,个体的度越大,联系强度越强,个体的信息熵越大,客户越不容易流失。外样本AUC值平均可以达到0.75以上,模型具有良好的预测精度。

研究结果对企业实践具有重要的意义,合作企业应用客户流失预警模型进行高风险流失客户的识别,预测精度可以达到70%,达到了企业的实践预期。客户流失预警模型可以帮助企业提前识别高风险流失客户,极大地降低企业维系客户的成本。建议企业管理者在未来更加关注与客户社交网络有关的变量,从网络结构的视角理解消费者行为,更好地进行客户关系管理。

关键词:社交网络;度;信息熵;客户流失;自我网络

中图分类号:F490.6

文献标识码:A

doi:10.3969/j.issn.1672-0334.2017.05.003

文章编号:1672-0334(2017)05-0028-10

引言

近年来,随着移动通信行业的快速发展,市场饱

和度越来越高,企业获取新用户的成本也越来越大。

另外,随着中国三大运营商的竞争加剧,产品和服务

收稿日期:2016-11-13 **修返日期:**2017-07-20

基金项目:中国人民大学科学研究基金重大项目《互联网统计学研究》资助

作者简介:周静,管理学博士,中国人民大学统计学院讲师,研究方向为复杂网络建模、社交网络分析和在线用户创造分析等,代表性学术成果为“Estimating spatial autocorrelation with sampled network data”,发表在2017年第1期《Journal of Business & Economic Statistics》,E-mail:zhoujing_89@126.com

周小宇,管理学博士,上海科技大学创业与管理学院助理教授,研究方向为营销战略和创业企业公司治理等,代表性学术成果为“An empirical investigation on firms' proactive and passive motivation for bribery in China”,发表在2013年第3期《Journal of Business Ethics》,E-mail:zhouxy@shanghaitech.edu.cn

王汉生,统计学博士,北京大学光华管理学院教授,研究方向为搜索引擎营销和社交网络分析等,代表性学术成果为“Factor profiled sure independence screening”,发表在2012年第99期《Biometrika》,E-mail:hansheng@pku.edu.cn

的同质化程度也越来越高,客户流失率逐年上升^[1]。企业在开发新客户的同时也要注重对老客户的维系,因此,如何维系老客户并降低流失率已经被重新提上企业的议事日程^[2]。已有研究主要集中在探讨客户流失的原因,如来自对服务质量的感知^[3]和用户特征^[4]的影响,鲜有关注意到与社交数据相关的影响因素^[5-7]。伴随着大数据处理和分析技术的广泛应用,企业可以借助社交网络数据制定客户维系和流失管理的相关市场决策,因此,从社交网络数据视角研究企业的客户流失成为一个新的可供探索的研究方向。

本研究以中国通信行业为背景,该行业市场竞争日趋激烈,市场发展的压力迫使运营商不断推出内容丰富、价格优惠的个性化产品,以期吸引更多的客户,但这些措施仍然无法缓解离网率居高不下、用户平均收益和利润持续走低的严峻形式。各大运营商近几年的用户增长十分缓慢,中国联通甚至在2014年出现了负增长。面对这样的严峻形势,客户保留已经成为企业最关心的问题之一,运营商开始关注哪些因素影响客户的流失,而在这个过程中他们并没有关注到客户自我网络特征的影响。从学术研究的角度,基于移动通信网络的客户数据是典型的自我网络结构化数据,分析和研究客户自我网络结构数据可以帮助探索自我网络的特性如何影响消费者的选择,并且运营商的平台能够为营销实证研究提供良好的田野实验环境,进而有助于将学术研究成果更好地服务于企业管理实践。

1 相关研究评述和研究假设

社交网络数据近年来被广泛应用于市场营销研究^[8-9]。相关研究表明,加入社交网络信息的营销模型比传统模型有更好的解释力^[10-12],并且能够为营销实践活动提供强有力的实地试验研究^[13-16]。本研究对自我网络的相关研究进行梳理评述,并且提出相关的研究假设。

1.1 自我网络的概念

社交网络是由一群个体和他们之间的互动关系构成的复杂网络结构^[17]。而自我网络就是从特定的个体出发,他所有的社会互动关系构成的网络^[18]。在自我网络的概念体系中,已有研究集中关注个体和邻居两个主要内容。个体指代的是网络中的节点(或成员),邻居指代的是与个体具有直接联系的其他个体^[19]。由此,个体和邻居构成了自我网络的主体结构,并且网络中的联系反映了个体与其他个体之间的互动关系。

已有学者对自我网络的研究主要集中在网络效应对个体行为的影响。自我网络不但会对个体的消费行为产生重要影响,而且会对企业的创新行为产生巨大影响^[20]。在个体的消费行为方面,DUBOIS et al.^[21]研究个体所感知的社交网络紧密程度如何影响消费者的口碑宣传,他们发现消费者更加愿意在亲密程度高的群体中传播品牌的负面信息,在亲密程

度低的群体中宣传品牌的正面信息;KATONA et al.^[22]通过对一组互联网上的社群研究发现,消费者对新产品的购买行为受到自我网络的节点度和自我网络的联系紧密程度的影响;RISSELADA et al.^[23]也发现了自我网络对消费者购买未知风险较高的科技产品方面具有重要的影响作用。而在企业的创新行为方面,FANG et al.^[24]发现企业的合作关系网络以及全球价值链网络对企业的新产品研发活动有着重要的促进作用。类似地,WANG et al.^[25]发现企业的高层管理者流动所创造的社会关系网络对企业的知识传播和创新研发有着显著的推动作用。

除此之外,网络结构变化也是自我网络研究的另一个重点。这方面的研究者关注自我网络的结构稳定性是如何影响个体行为的,特别是网络中的个体也存在着脱离旧网络和进入(创建)新网络的行为。这样的自我网络结构变化能够改变个体的社交范围以及相关的社会资本结构^[26]。同样地,个体的自我网络动态变化影响着企业营销实践^[27],一个典型的例子就是企业的客户流失。客户流失是客户关系管理中的重要研究问题,而且在管理实践中客户流失会对企业的当期利润和未来成长造成巨大的损失^[28]。

1.2 研究假设

个体的度是自我网络中的核心概念,它测量的是个体在网络中联系的数量^[17]。已有的实证研究认为,个体的度对网络稳定性具有重要的影响^[19]。从网络转换成本角度看,随着个体在网络中建立的联系数量上升,个体离开网络关系的转换成本也随之上升。这样的个体终结自我网络相关的合作关系的的可能性低,所以个体度越高的客户越不容易流失^[29]。同样地,从社会资本的视角出发,随着个体度的上升,随之增加的还有网络关系中隐含的情感承诺和持续承诺等创造的社会资本^[30-31]。由此,个体度高的客户能够获得更高的影响力和收获更多的利益,他们也更不愿意脱离现有的关系网络^[32-33]。对于一个度很大的移动通信用户,他换号(即客户流失)的成本很高,即一旦转网换号就意味着需要通知很多人新的号码,也许还会因为换号码而失去与一些朋友的联系。由此可以推断,个体的度越大,越不容易换号,即流失的概率越低。因此,本研究提出假设。

H₁ 个体客户在自我网络中的度越大,他流失的可能性越低。

自我网络中另一个重要的变量是个体与邻居之间关系的平均强度^[17]。自我网络数据结构是由二元变量组成,在自我网络分析中研究者往往通过个体与邻居之间的互动次数测量个体的网络关系强度。社交网络的相关研究表明,个体在网络中的关系强度越高,网络结构越稳定。此外,高强度的社会网络关系也增加了个体的社会资本,从而使个体更忠诚于所在的社会网络,产生更多与邻居的互动行为,进而提升网络关系强度^[34]。周涛等^[35]在基于社区用

户的行为研究中发现,增加个体与邻居之间互动能够提升信任关系,从而增加社会资本和社群的归属感;黄敏学等^[36]通过分析消费咨询网络中意见领袖的演化机制,发现网络关系互动会提升个体作为网络意见领袖的可能性,从而使个体更不容易脱离网络。基于此,本研究认为平均关系强度越高的个体越不容易脱离现有的关系网络,并提出假设。

H₂ 个体客户在自我网络中的平均强度越高,他流失的可能性越低。

自我网络结构的稳定性不但反映在个体与全部邻居的关系数量和平均关系强度所代表的整体均值上,而且体现在个体在网络关系中联系强度的分布情况。本研究假设有A和B两个用户,都与10个人通话100分钟,但是A用户的90分钟都是打给同一个用户,剩下的10分钟用于与其他9个用户通话;B用户将100分钟平均分配于10个不同的用户。可以看到,与A用户紧密相连的其实仅有1个用户,对于A用户来说其换号的成本是很低的,因为他只需要把新号码告诉与他联系紧密的人即可。因此本研究推断,在这种情况下,A用户的流失概率要大于B用户的流失概率。为了刻画上述现象,本研究借鉴信息论中信息熵的概念^[37]。信息熵最初用于描述信息源的不确定性,通常一个信息源发出什么样的信号是不确定的,可以根据它出现的概率进行度量,概率大,出现机会多,不确定性就小。

网络结构中的熵值反映了个体在网络关系中的权重分布。个体在网络结构中可以与许多邻居建立关系,但是关系的强弱水平不一,即个体与邻居之间的互动频率存在差异。如果个体在网络中的强关系都集中在一小部分邻居上,这样的网络关系权重分布的方差很大,存在网络稳定性的隐患,即高强度关系的邻居流失可能带动个体脱离网络。而如果个体的网络关系权重分布比较均匀,即个体在网络中与每一个邻居都建立了同等强度的关系,这样的网络结构相对比较稳定,不会因为部分强关系邻居的流失而导致个体的自我网络瓦解。基于这样的推论,本研究认为具有均匀分布关系权重的个体更不容易脱离现有的关系网络。因此,本研究提出假设。

H₃ 个体客户在自我网络中的关系权重分布越均匀,他流失的可能性越低。

2 数据和变量

2.1 数据收集和清理

本研究数据全部来自某移动通信公司在某三线城市分公司的内部经营分析底层数据,随机选取5万个左右VIP(平均每月花费大于80元)客户,以2014年3月至8月共6个月的数据为样本。基于研究需要,继续收集两部分数据,第1部分是按月份统计的客户基础通信数据,这部分数据包括用户的入网时间、当月花费、话费情况等;第2部分是按月份统计的客户点对点通信数据,即客户的通话详单,这部分数据是本研究中量级最大且最重要的数据,因为通过客户

的点对点通信数据可以构建一个用户的社交网络,进而可以测量一些与自我网络相关的变量。从数据量上看,平均每月客户的点对点通信数据在500万条左右。

数据清理主要遵循以下原则。①对于缺失值的处理,由于样本量较大,且缺失值的情况较少,所以对于个别缺失值的情况采取删除该条数据的处理。②对于一些不合乎正常值范围内的数据(如花费金额出现了负数的情况)也采取删除该条数据的处理方式。③对重复观测记录的数据(即同一条记录被记录了多次)采取删除重复数据的处理。此外,关于一些异常值的处理将在数据建模的描述性分析中进行阐述。

2.2 变量生成

在传统的客户关系管理研究中,有众多关于影响客户流失因素分析的研究,但是这些因素基本都只涉及到客户本身的人口统计学特征等信息^[38-39]。而个体并不是独立存在于这个社会,个体会与周围的其他个体交往,从而形成各种各样的社交网络,这样每一个个体在每一个社交网络中其实都被赋予了不同的角色和社会地位。所以,在分析消费者行为时,有必要把这种来自邻居的信息考虑进来。在本研究中,客户的通话详单呈现了一个通信网络,在这个网络中,可以清晰的看到每个用户都与谁通过电话、通过几次电话和每次的通话时间。基于这样的一个数据,本研究可以进一步总结出一些与自我网络相关的变量,并且这些变量对于解释客户流失有很重要的意义。

在具体解释新生成的变量前,先做一些简单的符号定义。在社交网络分析中,通常用节点和边表示网络中的个体和存在的关系,在本研究中,每个用户可以被看作是通信网络中的一个节点,用*i*表示, $i = 1, 2, 3, \dots, N$, N 为样本量。用社交网络分析中的邻接矩阵*A*表示用户之间的网络结构,*A*是一个 $N \times N$ 的0-1方阵, $A = a_{i,j} \in \mathbb{R}^{N \times N}$, $a_{i,j}$ 为矩阵中的元素。假设任意两个用户*i*和*j*,如果发现*i*与*j*通过电话,那么定义 $a_{i,j} = a_{j,i} = 1$,表示*i*与*j*之间存在一条边。

(1) 个体的度^[17]。本研究将*i*的度定义为与*i*有过通话记录(包括呼入和呼出)的不同用户的总数,用 D_i 表示, $D_i = \sum_{j \neq i} a_{i,j}$, i 的度越大,说明与*i*通话的人数越多,反之越少。

(2) 联系强度^[17]。本研究用平均通话时长类比联系的程度。 T_i 为*i*用户总的通话时长, Tie_i 为*i*的联系强度,即平均通话时长, $Tie_i = \frac{T_i}{D_i}$,表示在通话人数一定的前提下,如果平均通话时长越长,用户与他的联系人之间的关系越紧密,反之越疏远。

(3) 个体的信息熵^[40]。 E_i 为*i*的信息熵, $E_i = - \sum_{a_{i,j}=1} p_{i,j} \log(p_{i,j})$, $p_{i,j} = \frac{Comm_{i,j}}{T_i}$, $Comm_{i,j}$ 为*j*与*i*通话的时长, $p_{i,j}$ 为*j*与*i*通话的时长占*i*总通话时长的比例。从直

观上看,该公式刻画的是与*i*通话的所有用户的平均通话时长的分布,如果信息熵越大,说明平均通话时长的分布越分散,信息熵越小,说明平均通话时长的分布越集中。

以上3个变量为本研究重点探讨的变量,在研究模型中还加入了比较重要的指标作为控制变量,分别为入网时长、当月花费、本月与上月相比的花费变化、本月与上月相比个体的度的变化和本网用户占比。具体的变量说明见表1。

3 数据建模

3.1 描述性分析

在进行数据建模前要对所有变量进行描述性分析,为了描述的方便,以2014年8月份的数据为例,表

2给出描述性分析结果,其他月份的描述性分析结果与该月份基本相似。

在后续的建模分析中,本研究以均值加减3倍标准差作为识别异常值的标准,如果取值在这个范围之外,被认为是异常值,在建模分析中予以删除。

本研究的因变量是客户是否流失,是一个典型的1-0变量,1为流失,0为非流失。关于流失的定义,本研究遵循调研企业的做法,认为只要符合以下3条中的1条即被认为是流失,①客户主动申报离网,②当月未出账,③累积3个月延迟缴费。本研究使用的数据每个月的离网率统计见表3。

由表3可知,该公司平均每月的客户流失率基本维持在1.200%左右。在探索流失客户与非流失客户之间的差异时,本研究对所有自变量根据流失/非流

表1 变量说明
Table 1 About the Variables

变量	变量名称	单位	计算方法
个体的度	D_i	人数	$D_i = \sum_{j \neq i} a_{i,j}$
联系强度	Tie_i	分钟/人	$Tie_i = \frac{T_i}{D_i}$
个体的信息熵	E_i		$E_i = - \sum_{a_{i,j}=1} p_{i,j} \log(p_{i,j}), p_{i,j} = \frac{Comm_{i,j}}{T_i}$
入网时长	<i>Tenure</i>	天	当前时间减去入网时间
当月花费	<i>Expense</i>	元	直接获取
花费变化率	<i>Chgexpense</i>	%	$\frac{\text{当月花费} - \text{上月花费}}{\text{上月花费}}$
个体的度变化率	<i>Chgcount</i>	%	$\frac{\text{当月个体的度} - \text{上月个体的度}}{\text{上月个体的度}}$
本网用户占比	<i>Dxprop</i>	%	$\frac{\text{通话人数中本网用户数}}{\text{总人数}}$

表2 2014年8月份数据的描述性分析结果
Table 2 Results for Description Analysis of Data in August 2014

变量	均值	中位数	标准差	最小值	最大值
个体的度	76.215	57	82.081	1	1 763
联系强度	9.883	7.599	29.705	0.035	5 015.065
个体的信息熵	2.982	3.038	0.913	0	7.042
入网时长	1 404.499	1 138	989.292	10	4 603
当月花费	171.734	141.370	114.767	0	4 446
花费变化率	0.037	0	0.927	-1	176.504
个体的度变化率	0.048	-0.015	0.868	-0.994	123.500
本网用户占比	0.869	0.980	0.302	0	1

注:样本量为47 731。

表3 2014年3月至8月每月离网率统计
Table 3 Monthly Attrition Rate Statistics
from March 2014 to August 2014

月份	3月	4月	5月	6月	7月	8月
离网率/%	1.267	1.393	1.150	1.314	1.294	0.578

失进行分组的对比分析。以 D_i 、 Tie_i 、 E_i 、 $Tenure$ 和 $Expense$ 为例, 本研究发现, 非流失客户与流失客户相比, 平均拥有更长的人网时长、更高的花费、与更多的人通话、人均通话时长更长以及更加分散的通话时长分布。在通话人数这个变量上, 这种差异显得更为明显, 说明 D_i 在判断流失与否这个问题上占有重要的地位, 其他4个指标对于判断客户流失与否的问题也具有一定的意义。下面重点阐述模型的建立、估计结果和预测精度。

3.2 模型分析

本研究采取逻辑回归进行分析, 1为流失, 0为非流失。又因为本研究关心的是预测问题, 所以在分析中所有自变量来自当期(即当前月份), 而因变量

是否流失来自下一期(即下一个月份), 为了检验模型结果的稳健性, 重复4次这样的逻辑回归, 即用4月份的因变量对3月份的自变量建模, 用5月份的因变量对4月份的自变量建模, 以此类推。逻辑回归结果见表4。

表4中, April表示用3月份的自变量预测4月份因变量, 以此类推。本研究探索3个模型, 模型1为仅有传统变量的模型, 模型2为仅有自我网络特征变量的模型, 模型3为同时加入传统变量和自我网络特征变量的全模型。通过AUC值可知, 全模型优于传统变量的模型。由表4模型3的回归结果可知, ①入网时长越长, 客户越不容易流失, 该结果除了在July的回归中不显著外, 在其他月份都显著。②花费越多, 客户也越不容易流失, 表现稳定。 D_i 显著为负, 说明通话的人数越多, 客户越不容易流失, 表现稳定, H_1 得到验证。 Tie_i 显著为负, 说明平均通话时长越长, 客户越不容易流失, 表现稳定, H_2 得到验证。 E_i 在除July外的其他月份的结果均显著为负, 说明平均通话时长的分布越分散, 用户的流失概率越低, 表现相对稳定, H_3 得到验证。③关于花费变化率和个体度变化率, 随着通话人数的增加, 客户越不容易流

表4 逻辑回归结果
Table 4 Results for Logistic Regression

	April			May			June			July		
	模型1	模型2	模型3	模型1	模型2	模型3	模型1	模型2	模型3	模型1	模型2	模型3
常数项	3.384** (1.172)	-2.180*** (0.167)	0.543 (1.166)	3.487** (1.167)	-2.229*** (0.158)	1.833 [†] (1.113)	4.038*** (1.088)	-2.193*** (0.157)	1.758 (1.100)	1.158 (1.457)	-2.755*** (0.186)	-2.423 [†] (1.344)
个体的度		-0.019*** (0.003)	-0.015*** (0.003)		-0.014*** (0.002)	-0.009*** (0.002)		-0.013*** (0.002)	-0.010*** (0.002)		-0.033*** (0.004)	-0.031*** (0.004)
联系强度		-0.036*** (0.006)	-0.030*** (0.006)		-0.032*** (0.005)	-0.025*** (0.005)		-0.033*** (0.006)	-0.029*** (0.006)		-0.017** (0.006)	-0.014* (0.006)
个体的信息熵		-0.399*** (0.082)	-0.393*** (0.083)		-0.413*** (0.076)	-0.413*** (0.077)		-0.414*** (0.075)	-0.414*** (0.076)		-0.103 (0.100)	-0.117 (0.102)
入网时长	-0.329*** (0.059)		-0.148* (0.063)	-0.391*** (0.059)		-0.253*** (0.063)	-0.422*** (0.059)		-0.254*** (0.063)	-0.271*** (0.076)		0.022 (0.082)
当月花费	-0.807*** (0.059)		-0.476*** (0.068)	-0.831*** (0.051)		-0.665*** (0.055)	-0.761*** (0.051)		-0.489*** (0.058)	-0.661*** (0.063)		-0.271*** (0.078)
花费变化率	-0.804*** (0.173)		-0.516** (0.177)	-1.149*** (0.171)		-0.910*** (0.178)	-0.053 (0.123)		-0.003 (0.100)	-0.310 (0.183)		-0.243 (0.189)
个体的度变化率		-1.078*** (0.132)	-0.966*** (0.134)		-1.355*** (0.132)	-1.132*** (0.132)		-0.949*** (0.124)	-0.927*** (0.125)		-0.573*** (0.127)	-0.520*** (0.129)
本网用户占比	-2.444 (1.541)		0.483 (1.505)	-1.656 (1.531)		0.728 (1.440)	-2.439 (1.408)		-0.079 (1.403)	-1.064 (1.911)		1.063 (1.671)
AUC值	0.676	0.766	0.775	0.693	0.761	0.785	0.641	0.737	0.739	0.652	0.782	0.783

注: 因变量为是否流失, 1为流失, 0为非流失; 表中数据为参数估计结果, 括号内数据为参数估计的标准误; ***为0.001显著性水平, **为0.010显著性水平, *为0.050显著性水平, [†]为0.100显著性水平。下同。

失;随着花费的增加,客户也越不容易流失。本网用户占比在所有月份中并不显著,说明用户所属的运营商类型对是否流失没有显著的影响。综上,本研究提出的解释变量对预测用户是否流失都是显著的,说明在考虑用户流失的问题上,除一些传统的解释变量(如入网时长、每月花费),还应考察与用户社交圈相关的变量,因为这些变量对预测一个用户是否离网有着重要的作用。

3.3 交互效应分析

本研究探讨两个可能的交互效应。由于网络关系权重分布是决定网络中信息传播的重要因素,它与其他网络结构变量存在多种互动关系。网络关系权重分布越均匀,个体从关系邻居中获得信息的依赖性分布也会越平均,从而降低个体脱离关系网络的成本。虽然某些客户在移动通信网络中具有较高的度,但是个体客户对于单个网络邻居获得的信息量并不大,而且随着度的上升,个体从单个邻居处获得的信息量显著下降。所以,对于度较高的个体客

户而言,随着它对网络邻居的关系权重分布趋于平均,它离开这个网络的转化成本也显著降低。同理,网络关系权重均匀分布也会降低网络中平均强度较高的个体对网络的依赖性。虽然个体客户与网络邻居存在多次的互动次数,但是每次互动中获得的信息量并不多,而且互动数量越多,单次互动所承载的信息量会显著下降,从而降低个体在网络中的社会资本以及对其他邻居的意见的影响。因此,这样均匀分布的网络关系会削弱自我网络的度和平均关系强度对流失率的影响,故本研究提出两个推论,①关系权重均匀分布调节自我网络的度对流失率的影响,表现在均匀分布的关系权重降低高网络关系数量客户的流失成本;②关系权重均匀分布调节自我网络的平均强度对流失率的影响,表现在均匀分布的关系权重降低高平均强度客户的流失成本。

为了验证以上两个推论,本研究进行交互效应的回归,模型4为引入 D_i 与 E_i 的交互项,模型5为引入 D_i 与 Tie_i 的交互项。具体的回归分析结果见表5。

表5 交互效应回归结果
Table 5 Regression Results for Interactive Effect

	April		May		June		July	
	模型4	模型5	模型4	模型5	模型4	模型5	模型4	模型5
常数项	0.429 (1.139)	0.442 (1.169)	1.696 (1.079)	1.603 (1.109)	1.601 (1.076)	1.630 (1.104)	-2.586* (1.313)	-2.468† (1.347)
个体的度	-0.063*** (0.009)	-0.012*** (0.004)	-0.064*** (0.008)	-0.003 (0.003)	-0.062*** (0.008)	-0.006* (0.003)	-0.088*** (0.012)	-0.029*** (0.005)
联系强度	-0.023*** (0.006)	-0.022** (0.008)	-0.018*** (0.005)	-0.009 (0.006)	-0.018** (0.006)	-0.018* (0.008)	-0.006 (0.006)	-0.009 (0.009)
个体的信息熵	-0.329*** (0.084)	-0.402*** (0.084)	-0.288*** (0.080)	-0.418*** (0.077)	-0.303*** (0.078)	-0.423*** (0.076)	-0.068 (0.101)	-0.125 (0.103)
入网时长	-0.153* (0.063)	-0.151* (0.063)	-0.252*** (0.063)	-0.258*** (0.063)	-0.258*** (0.063)	-0.258*** (0.063)	0.018 (0.082)	0.021 (0.082)
当月花费	-0.438*** (0.069)	-0.467*** (0.069)	-0.603*** (0.055)	-0.631*** (0.055)	-0.447*** (0.060)	-0.477*** (0.059)	-0.232** (0.079)	-0.267*** (0.079)
花费变化率	-0.530** (0.178)	-0.521** (0.178)	-0.975*** (0.181)	-0.937*** (0.180)	-0.022 (0.103)	-0.004 (0.101)	-0.265 (0.192)	-0.243 (0.190)
个体的度变化率	-0.824*** (0.131)	-0.963*** (0.133)	-0.920*** (0.129)	-1.094*** (0.131)	-0.777*** (0.121)	-0.924*** (0.125)	-0.409** (0.125)	-0.520*** (0.129)
本网用户占比	0.851 (1.461)	0.539 (1.505)	0.943 (1.384)	0.792 (1.429)	0.292 (1.360)	-0.023 (1.404)	1.566 (1.614)	1.090 (1.672)
个体的度 × 个体的信息熵	0.011*** (0.002)		0.012*** (0.002)		0.012*** (0.002)		0.015*** (0.003)	
个体的度 × 联系强度		0 (0)		-0.001*** (0)		0 (0)		0 (0)
AUC 值	0.779	0.775	0.792	0.788	0.748	0.740	0.789	0.784

由表5的回归结果可知, D_i 与 E_i 交互项的系数显著为正, 说明 E_i 加强了 D_i 的主效应, 这与本研究的推论1不符, 为此本研究检验各个变量的 VIF 值, 发现 D_i 与 $D_i \cdot E_i$ 有很强的多重共线性, VIF 值超过30, 因此可能导致回归模型的估计结果不准。对于 D_i 与 Tie_i 的交互项, 只有在5月份的数据中显著为负, 验证了推论2, 但是在其他月份均未得到验证, 各个变量的 VIF 值表现正常。本研究认为 Tie_i 对 D_i 的调节效应并不是很稳定。

3.4 模型预测精度

在学术领域, ROC 曲线和 AUC 值经常被用来评判一个逻辑回归的预测效果如何^[41], 但是在商业实践中, 这样比较专业的术语很难直观地给出具体的解释。所以本研究除了汇报模型的 AUC 值, 还采取覆盖率-捕获率指标评判模型的预测精度。假设共有100个客户, 其中有20个客户在下个月会流失, 如果不用任何模型, 想要确定这20个流失的客户, 就需要将营销成本花费在100个客户身上, 此时覆盖率为 $\frac{100}{100} \times 100\% = 100\%$, 捕获率为 $\frac{20}{20} \times 100\% = 100\%$ 。如果可以识别出一些流失风险高的客户, 就可以有针对性地去实施营销策略。通过建模可以给出每个客户一个预测的流失概率, 将预测的流失概率值从高到低排序, 排在越前面的客户, 越是高风险流失客户, 这时企业可以有针对性的进行营销, 如只对前40个高风险客户进行营销, 在这40个客户中只识别出15个流失风险高的客户, 此时的覆盖率为 $\frac{40}{100} \times 100\% = 40\%$, 捕获率为 $\frac{15}{20} \times 100\% = 75\%$, 也就是说只需覆盖40%的客户, 就可以抓到75%的真实流失客户, 即企业只要付出40%的成本, 就可以得到75%的收益。这样做虽然不能100%抓住所有要流失的客户, 但可以用较低的成本识别出相对较多的流失客户。

在本研究中, 可以针对每一个逻辑回归绘制覆盖率-捕获率曲线, 以6月自变量预测7月因变量为例进行说明。所有的自变量来自6月份, 因变量是否流失来自7月份, 在此可以看到每一个客户的真实流失状态, 根据模型的估计结果, 可以算出每一个客户的流失概率, 如果这个数值越高, 说明客户流失的可能性越大, 将客户按照流失概率从高到低进行排序, 排在最前面的是最容易流失的客户, 越往后越不容易流失。这样每一位客户不仅有一个真实的流失状态, 同时还对应由模型估计出来的流失概率。如果瞄准排在最靠前的客户, 捕获真实流失客户的可能性越大。本研究根据估计的流失概率确定不同的阈值, 利用这些阈值把现有的客户分类, 如利用分位数20%、40%、60%、80%, 20%意味着覆盖流失概率在前20%的客户, 那么在这些客户中真正流失的客户数就是捕获率。因此, 对于每一个阈值, 都可以计算相应的覆盖率和捕获率, 本研究的覆盖率与捕获率的曲线见图1。

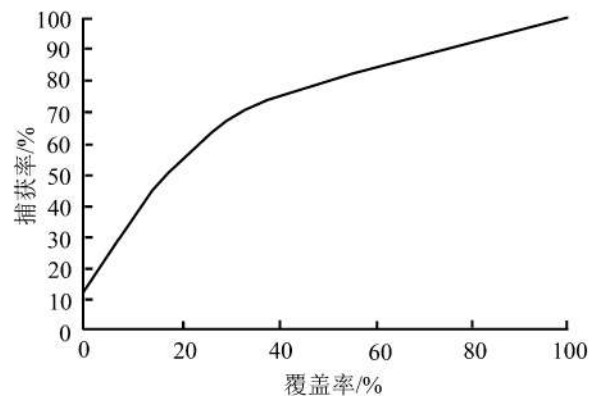


图1 覆盖率-捕获率曲线
Figure 1 Coverage-hit Rate Curve

由图1可知, 根据本研究模型, 只需覆盖20%的用户, 就可以达到大于50%的捕获率, 也就是说只对流失概率排名最靠前的20%客户进行营销, 在覆盖的这些客户中能识别出的真正流失的客户可以达到50%以上。模型不能做到百分之百的精度, 如果要抓住全部流失的客户, 也只能对所有客户进行营销, 而这时的成本也是最高的, 在实际中并不常用, 企业可以根据自身的情况(如营销预算成本)自主选择要覆盖多少用户。

4 管理实践

基于本研究构建的客户流失预警模型, 以2014年8月份的客户作为试验对象, 计算出8月份客户中流失风险高的客户, 然后企业的客服专员对这些识别出的高风险客户进行电话回访, 从而识别出真正流失的客户。具体做法如下: 首先根据模型的参数估计结果计算出8月份还在网的客户的流失概率, 按照流失概率从高到低对客户进行排序, 由表3的历史离网率可知, 平均每月离网率大概维持在1.200%左右, 所以本研究选择1%作为阈值, 计算出的流失概率大于1%的即为流失风险高的客户, 小于1%的为流失风险低的客户。经统计流失风险高的客户有16 128人, 由于电话回访还需要一定的人力成本, 所以公司在16 128人中随机选取4 997个客户进行电话回访, 回访的主要目的是询问客户对当前的服务是否满意, 是否有些抱怨。最终的统计结果表明, 电话接通率为62%, 在接通的人群里有348人表达了不同程度的抱怨情绪, 抱怨率为11.233%。从电话回访的结果看, 已经有部分客户对服务感到不满意了, 这可以说是流失之前的一个很重要的预警, 而且电话的接通率也相对较低, 这也是流失前的一个重要预警。为了验证模型的精度, 在9月份结束后, 本研究又从公司获得9月份客户流失情况, 统计结果见表6。

从之后的统计可知, 根据本研究模型选出的16 128个高风险客户中, 有472个流失的客户, 捕获率为72% ($\frac{472}{653} = 72.282\%$)左右。而此时的覆盖率只有

34% ($\frac{16\ 128}{47\ 539} = 33.926\%$)左右。所以这次业务实施给了企业很大的信心用该模型进行客户流失风险预警。

表6 2014年9月份客户流失情况
Table 6 Customer Attrition in September 2014

		样本量	流失客户数	流失率/%
真实情况		47 539	653	1.374
模型预测情况	高风险组	16 128	472	2.927
	低风险组	31 421	181	0.576

5 结论

随着电信行业的迅猛发展,市场饱和程度越来越高,新用户的增长十分缓慢,而中国三大运营商激烈的竞争局面也使老用户的保留变得异常困难,因此识别影响客户流失的因素并有效防止客户流失已经成为该行业管理者普遍关心的问题之一。本研究着眼于客户流失影响因素研究,运用社交网络分析方法,通过构建与网络结构相关的变量进行影响因素的探讨,运用逻辑回归方法构建客户流失预警模型。本研究构建与自我网络特征相关的变量个体的度、联系的强度、个体的信息熵,逻辑回归结果表明,这3个变量对预测客户是否流失具有显著影响。具体的,个体的度显著为负,说明通话的人数越多,客户越不容易流失,表现稳定。联系的强度也显著为负,说明平均通话时长越长,客户越不容易流失,表现稳定。个体的信息熵显著为负,说明平均通话时长的分布越分散,用户的流失概率越低,表现相对稳定。本研究的3个假设得到验证。

与传统模型(模型1)相比,本研究提出的基于自我网络特征的模型预测精度更高。本研究结果对企业实践具有非常重要的意义,合作企业应用本研究模型进行流失风险高客户的识别,预测精度可以达到72%左右,达到了企业的实践预期。该流失预警模型可以帮助企业提前识别流失风险高客户,极大地降低企业维系客户的成本。根据本研究结果,建议企业管理者在未来更加关注与客户社交网络有关的变量,从网络结构的视角理解消费者行为,从而更好地进行客户关系管理。

本研究仍然存在一些不足和需要改进的方向。①关于流失的定义,本研究将确定性的流失(如客户主动申报离网)和不确定性的流失(如公司根据以往经验认为累计欠费达3个月的也是流失客户)都归为流失进行分析,但其实只有第一种确定性的流失才是真正的流失,因此后续研究可以把流失的种类分开讨论和建模。②本研究的数据抓取并不全面,未考虑其他电信运营商的客户数据,可以尝试获取跨平台的数据,这会使研究更加全面。③受到

互联网的冲击影响,很多用户的手机使用行为会受到微信等即时通讯工具的影响,未来可以考虑第三方互联网平台的数据。④对于企业来说,流失更像一个长期行为,而本研究用当月的消费行为数据预测下一个月的流失情况,更像是一个短期行为,后续研究可以尝试把时间区间拉长。

参考文献:

- [1] 钱苏丽,何建敏,王纯麟. 基于改进支持向量机的电信客户流失预测模型. *管理科学*, 2007, 20(1): 54-58.
QIAN Suli, HE Jianmin, WANG Chunlin. Telecom customer churn prediction model based on improved SVM. *Journal of Management Science*, 2007, 20(1): 54-58. (in Chinese)
- [2] 郭红丽. 客户体验维度识别的实证研究:以电信行业为例. *管理科学*, 2006, 19(1): 59-65.
GUO Hongli. Empirical research on the identification of customer experience dimensions: taking telecommunication industry as an example. *Journal of Management Science*, 2006, 19(1): 59-65. (in Chinese)
- [3] ZEITHAML A, LEONARD B, ANANTHANARAYANAN P. The behavioral consequences of service quality. *Journal of Marketing*, 1996, 60(2): 31-46.
- [4] VERHOEF C. Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing*, 2003, 67(4): 30-45.
- [5] 唐小飞,周庭锐,贾建民. CRM赢回策略对消费者购买行为影响的实证研究. *南开管理评论*, 2009, 12(1): 57-63.
TANG Xiaofei, ZHOU Tingrui, JIA Jianmin. An empirical research: how CRM win-back strategy influences customer satisfaction and share of wallet. *Nankai Business Review*, 2009, 12(1): 57-63. (in Chinese)
- [6] 李海芹,张子刚. CSR对企业声誉及顾客忠诚影响的实证研究. *南开管理评论*, 2010, 13(1): 90-98.
LI Haiqin, ZHANG Zigang. An empirical study on the effects of corporate social responsibility on corporate reputation and customer loyalty. *Nankai Business Review*, 2010, 13(1): 90-98. (in Chinese)
- [7] 崔艳武,苏秦,李钊. 电子商务环境下顾客的关系利益实证研究. *南开管理评论*, 2006, 9(4): 96-103.
CUI Yanwu, SU Qin, LI Zhao. An empirical research on relational benefits in e-business environment. *Nankai Business Review*, 2006, 9(4): 96-103. (in Chinese)
- [8] GOEL S, GOLDSTEIN D G. Predicting individual behavior with social networks. *Marketing Science*, 2014, 33(1): 82-93.
- [9] TRUSOV M, BUCKLIN R E, PAUWELS K. Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *Journal of Marketing*, 2009, 73(5): 90-102.
- [10] WANG J, ARIBARG A, ATCHADÉ Y F. Modeling choice interdependence in a social network. *Marketing Science*, 2013, 32(6): 977-997.
- [11] IYER G, KATONA Z. Competing for attention in social communication markets. *Management Science*, 2016, 62(8):

- 2304–2320.
- [12] WEI Y, YILDIRIM P, VAN DEN BULTE C, et al. Credit scoring with social network data. *Marketing Science*, 2016, 35(2):234–258.
- [13] STEPHEN A T, TOUBIA O. Deriving value from social commerce networks. *Journal of Marketing Research*, 2010, 47(2):215–228.
- [14] 冯芷艳, 郭迅华, 曾大军, 等. 大数据背景下商务管理研究若干前沿课题. *管理科学学报*, 2013, 16(1):1–9. FENG Zhiyan, GUO Xunhua, ZENG Dajun, et al. On the research frontiers of business management in the context of big data. *Journal of Management Sciences in China*, 2013, 16(1):1–9. (in Chinese)
- [15] VERBEKE W, MARTENS D, BAESENS B. Social network analysis for customer churn prediction. *Applied Soft Computing*, 2014, 14:431–446.
- [16] HASAN S, BAGDE S. Peers and network growth: evidence from a natural experiment. *Management Science*, 2015, 61(10):2536–2547.
- [17] WASSERMAN S, FAUST K. *Social network analysis: methods and applications*. Cambridge, UK: Cambridge University Press, 1994:169–215.
- [18] EVERETT M, BORGATTI S P. Ego network betweenness. *Social Networks*, 2005, 27(1):31–38.
- [19] HANNEMAN R A, RIDDLE M. *Introduction to social network methods*. Riverside, CA: University of California Press, 2005:169–215.
- [20] SHRIVER S K, NAIR H S, HOFSTETTER R. Social ties and user-generated content: evidence from an online social network. *Management Science*, 2013, 59(6):1425–1443.
- [21] DUBOIS D, BONEZZI A, DE ANGELIS M. Sharing with friends versus strangers: how interpersonal closeness influences word-of-mouth valence. *Journal of Marketing Research*, 2016, 53(5):712–727.
- [22] KATONA Z, ZUBCSEK P P, SARVARY M. Network effects and personal influences: the diffusion of an online social network. *Journal of Marketing Research*, 2011, 48(3):425–443.
- [23] RISSELADA H, VERHOEF P C, BIJMOLT T H A. Dynamic effects of social influence and direct marketing on the adoption of high-technology products. *Journal of Marketing*, 2014, 78(2):52–68.
- [24] FANG E, LEE J, PALMATIER R, et al. If it takes a village to foster innovation, success depends on the neighbors: the effects of global and ego networks on new product launches. *Journal of Marketing Research*, 2016, 53(3):319–337.
- [25] WANG R, GUPTA A, GREWAL R. Mobility of top marketing and sales executives in business-to-business markets: a social network perspective. *Journal of Marketing Research*, 2017, 54(4):650–670.
- [26] ADLER P S, KWON S W. Social capital: prospects for a new concept. *Academy of Management Review*, 2002, 27(1):17–40.
- [27] IYENGAR R, VAN DEN BULTE C, VALENTE T W. Opinion leadership and social contagion in new product diffusion. *Marketing Science*, 2011, 30(2):195–212.
- [28] TRAINOR K J, ANDZULIS J M, RAPP A, et al. Social media technology usage and customer relationship performance: a capabilities-based examination of social CRM. *Journal of Business Research*, 2014, 67(6):1201–1208.
- [29] GIUDICATI G, RICCABONI M, ROMITI A. Experience, socialization and customer retention: lessons from the dance floor. *Marketing Letters*, 2013, 24(4):409–422.
- [30] 陈爱辉, 鲁耀斌. SNS用户活跃行为研究: 集成承诺、社会支持、沉没成本和社会影响理论的观点. *南开管理评论*, 2014, 17(3):30–39. CHEN Aihui, LU Yaobin. Users' active behavior in SNSs: integrating commitment, social support, sunk cost and social influence perspective. *Nankai Business Review*, 2014, 17(3):30–39. (in Chinese)
- [31] LIN M, PRABHALA N R, VISWANATHAN S. Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, 2013, 59(1):17–35.
- [32] OH H, CHUNG M H, LABIANCA G. Group social capital and group effectiveness: the role of informal socializing ties. *Academy of Management Journal*, 2004, 47(6):860–875.
- [33] LU Y, JERATH K, SINGH P V. The emergence of opinion leaders in a networked online community: a dyadic model with time dynamics and a heuristic for fast estimation. *Management Science*, 2013, 59(8):1783–1799.
- [34] VAN DEN BULTE C, WUYTS S. *Social networks and marketing*. Cambridge, Massachusetts: Marketing Science Institute, 2007:29–36.
- [35] 周涛, 鲁耀斌. 基于社会资本理论的移动社区用户参与行为研究. *管理科学*, 2008, 21(3):43–50. ZHOU Tao, LU Yaobin. Explaining mobile community user participation based on social capital theory. *Journal of Management Science*, 2008, 21(3):43–50. (in Chinese)
- [36] 黄敏学, 王琦缘, 肖邦明, 等. 消费咨询网络中意见领袖的演化机制研究: 预期线索与网络结构. *管理世界*, 2015(7):109–121. HUANG Minxue, WANG Qiyuan, XIAO Bangming, et al. A study on the evolutionary mechanism of the idea leader in the consumer advice network of the expectancy clue and the network structure. *Management World*, 2015(7):109–121. (in Chinese)
- [37] GRAY R M. *Entropy and information theory*. New York: Springer-Verlag, 2011:17.
- [38] 夏维力, 王青松. 基于客户价值的客户细分及保持策略研究. *管理科学*, 2006, 19(4):35–38. XIA Weili, WANG Qingsong. Customer segmentation and retention strategy based on customer value. *Journal of Management Science*, 2006, 19(4):35–38. (in Chinese)
- [39] 罗彬, 邵培基, 罗尽尧, 等. 基于竞争对手反击的电信客户流失挽留研究. *管理科学学报*, 2011, 14(8):17–33. LUO Bin, SHAO Peiji, LUO Jinyao, et al. Research on customer churn detainment in telecom based on rival counterattack and customer detainment value maximization. *Journal of Management Sciences in China*, 2011, 14(8):17–33. (in Chinese)

- [40] SHANNON C E. A mathematical theory of communication. *Bell Labs Technical Journal*, 1948, 27(3):379-423.
- [41] WANG H. A note on iterative marginal optimization: a simple algorithm for maximum rank correlation estimation. *Computational Statistics & Data Analysis*, 2007, 51(6):2803-2812.

On the Influence of Ego Network Concerning Customer Attrition of the Telecommunication Industry

ZHOU Jing¹, ZHOU Xiaoyu², WANG Hansheng³

1 School of Statistics, Renmin University of China, Beijing 100872, China

2 School of Entrepreneurship and Management, Shanghai Tech University, Shanghai 201210, China

3 Guanghua School of Management, Peking University, Beijing 100871, China

Abstract: In the recent years, it is observed that, given the prospering development of telecommunication industry, the companies experience costly investment in customer acquisition as the market becomes more saturated. Meanwhile in China, the fierce competition between three major operators has intensively promoted the homogeneity of products and services. This leads to the difficulty of retaining old customers in the sense that customer attrition is increasing annually. Therefore, the identification of and prevention for customer attrition are regarded as a key issue in the field of telecommunication management.

This paper focuses on the study of customer attrition. Inspired by the popularly used social network analysis method, we construct some network related variables to explore the influencing factors. A logistic regression is proposed to build customer attrition model. From the perspective of social network, we establish a customer communication network using their point-to-point communication data. Under the framework of ego network, this paper constructs three ego network featured variables, namely, degree, tie strength and ego entropy. The empirical data comes from one of the three major telecommunication companies in mainland China. These data includes communication bill and point-to-point communication data. A logistic regression is used to investigate customer attrition model based on these variables.

It is found that, the degree, tie strength, and ego entropy are all significant indicators in predicting customer attrition. Specifically, if a customer has a larger degree, a higher tie strength and a bigger ego entropy, then his attrition rate will be lower than others. The out of sample AUC value is about 0.75 on average, which reflects a relatively high prediction accuracy.

The results of this paper are of great importance to the practice of enterprises. The model has been adopted by the cooperative enterprise. They use the model to identify high risk customers who are going to leave, and the prediction accuracy can reach 70%, which meets the expectation of the enterprise. The proposed attrition model can help companies to identify their high-risk customers in advance. This greatly reduce the cost of maintaining existing customers. Through this study, we strongly recommend business managers should pay much attention to the social network-related variables of customers. It can help the company to better understand consumer behavior from the perspective of network structure and thus better customer relationship management.

Keywords: social network; degree; entropy; customer attrition; ego network

Received Date: November 13th, 2016 **Accepted Date:** July 20th, 2017

Funded Project: Supported by the Research Funds of Renmin University of China, the Major Planning Project Study on Internet Statistics

Biography: ZHOU Jing, doctor in management, is a lecturer in the School of Statistics at Renmin University of China. Her research interests cover complex network modeling, social network analysis and user-generate content analysis. Her representative paper titled "Estimating spatial autocorrelation with sampled network data" was published in the *Journal of Business & Economic Statistics* (Issue 1, 2017). E-mail: zhoujing_89@126.com

ZHOU Xiaoyu, doctor in management, is an assistant professor in the School of Entrepreneurship and Management at Shanghai Tech University. His research interests include marketing strategy and entrepreneurial corporate governance. His representative paper titled "An empirical investigation on firms' proactive and passive motivation for bribery in China" was published in the *Journal of Business Ethics* (Issue 3, 2013). E-mail: zhoxuy@shanghaitech.edu.cn

WANG Hansheng, doctor in statistics, is a professor in the Guanghua School of Management at Peking University. His research interests include search engine marketing and social network analysis. His representative paper titled "Factor profiled sure independence screening" was published in the *Biometrika* (Issue 99, 2012). E-mail: hansheng@pku.edu.cn

□