



移动用户的动态信用评估模型的构建与应用

边凌燕

(中国电信股份有限公司上海研究院 上海 200122)

摘要:结合移动用户的特征数据分析,构建了科学、动态的用户信用评估模型,进而应用于面向不同信用等级用户的差异化信控停机服务,有效提升用户信用服务体验,加强用户黏性,挖掘用户潜在消费力,增加企业运营收入。

关键词:信用控制;差异化服务;信用评估;特征分析

doi: 10.3969/j.issn.1000-0801.2014.10.023

Construction and Application of Dynamic Credit Evaluation Model for Mobile User

Bian Lingyan

(Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China)

Abstract: Combined with the data analysis of mobile users' characteristics, a scientific and dynamic estimation models for user credit were built. It can be applied to different solutions of suspended service for different mobile users based on the diverse credit category. It is effective to improve user experience, strengthen user loyalty, exploit potential spending power and ultimately increase the operating income.

Key words: credit control, differentiation service, credit evaluation, characteristic analysis

1 引言

全国移动用户数已突破 10 亿户大关,通信市场渐近饱和,三大运营商围绕用户市场展开的竞争日趋激烈。差异化的用户服务正在成为运营商提升竞争力的核心要素。对比于航空、金融等行业愈加成熟的用户差异化信用服务^[1],运营商的信用控制体系存在着评估规则固化、针对不同信用等级用户的差异化服务缺失等问题。中国电信股份有限公司(以下简称中国电信)某省公司月均停机用户数量占移动出账用户总数的 25%~30%,会在 5 天内复机的约有 68%,但拥有信控金额的用户比例仅为 1.5%。初步测算,对那些信用良好且通信需求稳定的用户执行简单固化的信控停机策略,其造成

的不当停机,除了直接导致企业百余万元的月收入损失,也极易影响用户的满意度和忠诚度。

本文结合该省公司数据实例,主要阐述了利用用户通信行为数据构建用户信用评估模型,并将信用评估结果应用于存量用户的差异化信控停机策略,提高用户黏性,满足其潜在的消费需求,增加企业量收。

2 动态信用评估模型

用户信用评估是一个综合评价的过程,其模型的构建需要综合考虑用户基本信息、业务订购信息、历史信用、消费能力、通信行为、企业风险、交往圈影响等多方面因素以及各影响因素之间的差异性。模型输出的用户信用评分需要能够全面、真实地反映用户信用情况^[2]。

从国内目前对电信客户信用度评估的研究来看,参考文献[3]和参考文献[4]都是利用线性公式对相关客户属性的权值(通过遗传算法获得每个属性的权值)进行简单的相加,计算出信用度分值。但是,通过大量应用来看,客户信用度与各相关属性之间并不是简单的线性关系,而是非线性关系,因此,线性模型不能很好地逼近实际情况,有必要采用更精确的模型。而国外对个人信用评估方法的研究已经发展到把信用度分析看成是模式识别中的一种分类问题,具体做法是根据历史记录中每个类别(如约交纳话费、违约等)的若干样本,从已知的数据中发现其规律,从而总结出分类的规则,建立判别模型,用于对当前样本的判别^[5]。在理论状况下,它可能是最有效的方法,也是国际学术界视为主流的方法。但由于属性及其组合的权重选取具有多重性,实际应用中很难得出比较理想的结果。

因此,本文结合应用实际提出的信用评估模型是在确立关键评估维度的基础上,先通过因子分析法得到每个用户 5 个关键维度分,然后利用熵值法计算各个维度的权重,最后加权计算得到用户信用综合评分。模型通过结合用户每一段时间内的行为特征属性数据,周期性给出用户信用评分,实现动态评估。目标用户确定结合具体业务需求及规则,移动用户信用度评估模型的初始目标需排除以下 4 类用户:

- 无限信用用户(党、政、军、高价值及其他特殊用户);
- 零信用用户(OCS 用户、无线上网卡用户、欠费停机 30 天以上未复机的信用不良用户等);
- 无信控业务需求用户(标准后付费用户、公免用户、银行托收用户等);
- 不满足评级条件用户(入网 3 个月以内的用户、批量开通且未激活用户等)。

2.1 评估特征选取

用户信用度主要用来表征用户遵守守信的程度。外部环境、用户历史表现、用户态度及消费能力等因素都会影响用户的信用度。外部环境一般包括宏观经济环境、电信竞争格局等重大事件,其影响效应一般是群体性的。用户历史表现一般包括缴费及时率、大额欠费、欠费销户等有关用户信用的记录。用户态度包括用户忠诚度和满意度,其中忠诚度衡量用户对运营商业务和服务的依赖程度,譬如重复购买、预存话费、合约捆绑等情况,忠诚度高的用户,信用损坏的代价会更高;满意度衡

量用户的期望与需求被满足的差距,譬如通信趋势、投诉及业务退订情况,满意度低的用户有潜在补偿心理。用户消费能力的考察因素包括身份背景、消费力和消费意愿等^[6,7]。

将影响用户信用度的概念细化,由此确定历史信用、通信需求、社会属性、违约成本、支付能力等五大关键用户信用评估维度^[8]。结合评估需求,根据用户各个特征数据与信用关系的关联性探索结果,最终选择影响显著的 25 个变量,并依次归入上述 5 个评估维度。具体变量筛选结果见表 1。

表 1 用户信用评估模型输入变量

变量分类	变量名称
历史信用	近 6 个月双停次数
	近 6 个月平均双停时长/h
	最近一次双停时长/h
通信需求	近 1 个月活跃天数
	近 3 个月平均通话时长
	近 3 个月通话时长趋势
	近 6 个月漫游月份数
	近 3 个月平均短信次数
	同用户下的产品设备数(CDMA 网、宽带和固话)
	近 3 个月漫游通话时长占比
社会属性	近 3 个月平均流量
	近 3 个月流量趋势
	近 3 个月交往圈用户数
	用户年龄
	是否校园用户
违约成本	主叫本地交往圈用户占比
	主叫交往圈用户占比
	网间交往圈用户占比
支付能力	在网时长
	近 6 个月平均每次缴费金额
	是否捆绑用户(1 机补、0 话补、-9 未知)
	近 3 个月平均 ARPU
	ARPU 趋势
	终端价格

2.2 利用因子分析计算维度得分

计算维度得分之前,需要确定各个指标对用户信用度的影响方向,即确定指标是正向指标、负向指标还是适度指标,如用户近 3 个月欠费次数指标,指标越大,用户发生信用风险的可能性越高,因此该指标与目标呈反向关系,为负向指标。对负向指标和适度指标,要转化为正向指标,



负向指标可以采用取倒数或极大化的方式： $\frac{1}{x}$ 或 $\max -x$ ，其中 x 为原始指标；适度指标应趋于适度值（假设为 k ），则可以将指标转化为： $\frac{1}{|x-k|}$ 。

当所有指标都转化为正向指标时，对各维度的指标进行因子分析。因子分析的主要步骤如下。

(1) 对原始评估指标体系进行标准化，标准化为原始指标减去平均值除以标准差，标准化的指标均值为 0，方差为 1，消除指标量纲和数量级的影响。

(2) 根据标准化的结果求相关系数矩阵。

(3) 求解矩阵的特征值、特征向量和方差贡献率，把累计方差贡献率作为公因子的权重。

(4) 根据累计方差贡献率确定公因子个数，一般选择累计方差贡献率超过 85% 的前 K 个公因子，或者选择特征值大于 1 的 K 个公因子。

(5) 根据每个公因子的权重对因子得分加权求和，得到每个维度得分。

由于因子得分相对数性质不明显，可能出现负值，需要对因子得分进行归一化处理，采用逻辑函数对因子得分线性化：

$$factor_i' = \frac{1}{1 + e^{-factor_i}} \quad (1)$$

根据各因子的权重，计算各维度得分：

$$Score = \sum w_i \times factor_i' \quad (2)$$

根据图 1 可得，社会属性维度得分为：

$$\frac{57.461\ 8}{78.863\ 9(1 + e^{-factor1})} + \frac{21.402\ 1}{78.863\ 9(1 + e^{-factor2})} \quad (3)$$

2.3 利用熵值法计算维度权重

维度权重的计算采用熵值法。熵是对不确定性的一种度量，一般而言信息量越大，不确定性就越小，熵也就越小；信息量越小，不确定性越大，熵也越大。根据熵的特性，本文可以通过计算熵值来判断某个指标的离散程度，指标的离散程度越大，说明该指标对综合评价的影响越大。

首先计算每个维度的熵值，计算式如下：

$$H_j = \frac{-\sum_{i=1}^m f_{ij} \ln f_{ij}}{\ln m} \quad (4)$$

其中， m 是样本数， $f_{ij} = \frac{S_{ij}}{\sum_{i=1}^m S_{ij}}$ ， S_{ij} 是第 i 个样本在第

j 个维度上的得分，则第 j 个维度的权重为：

$$w_j = \frac{1 - H_j}{\sum_{i=1}^n (1 - H_j)} \quad (5)$$

其中， n 是维度的个数。

经过计算，得到历史信用、社会属性、通信需求、违约成本和支付能力各维度的权重见表 2。

表 2 5 个评估维度的权重

维度名称	权重
历史信用	0.385 0
社会属性	0.239 2
通信需求	0.112 6
违约成本	0.154 3
支付能力	0.108 8

最后根据维度得分和维度权重，加权求和可以得到每个用户的信用得分：

$$\sum_{i=1}^n \omega_i Score_i$$

2.4 模型效果评估

由此，得到模型目标用户的信用综合得分，分布如图 2 所示。信用评分呈现合理的长尾的正态分布形态，即信用较好用户的信用评分集中在高分值，信用较差用户的信用评分分散在低分值的尾部，表明信用差的用户与信用好的用户在信用评分上差异明显。

用户信用得分大小反映了用户信用状况的好坏，信用得分越高的用户信用状况越好。用户信用状况的好坏用 30 天

社会属性	近3个月交往圈用户数	社会属性			
	用户年龄	公因子	特征值	方差贡献率	累计方差贡献率
	是否校园用户	factor1	3.448	57.461 8	57.461 8
	主叫本地交往圈用户占比	factor2	1.284	21.402 1	78.863 9
	主叫交往圈用户占比				
	网间交往圈用户占比				

图 1 计算社会属性维度得分的示例

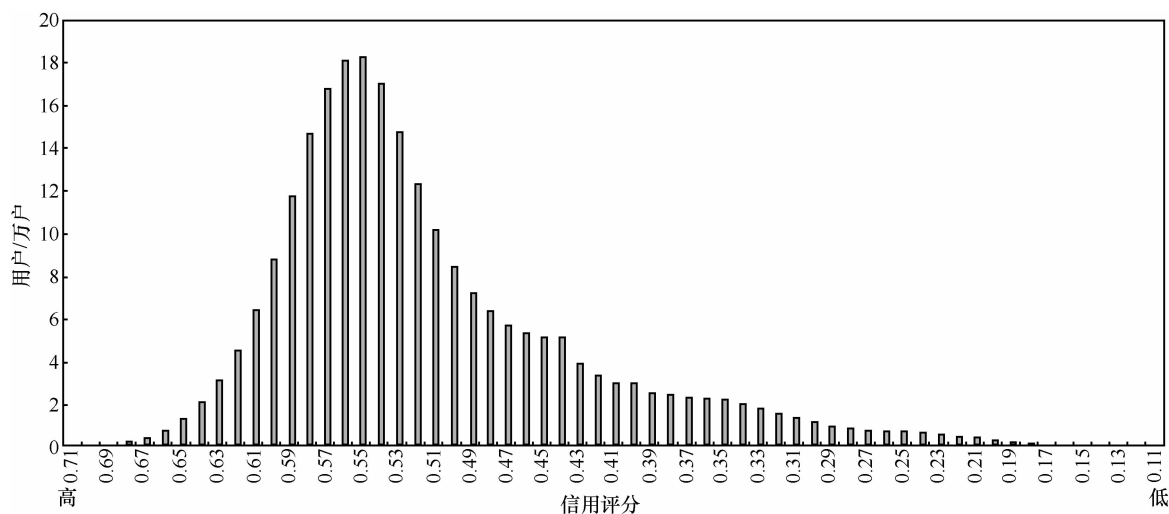


图2 目标用户的信用综合得分分布

复机率进行衡量,按照信用得分由高到低排序,比较不同信用得分下的30天复机率,结果如图3所示。

将用户信用评分由高到低排序,发现不同信用评分的用户的30天复机率也依次降低,说明信用评分模型能较好地反映用户信用的好坏。

3 基于评估模型的用户差异化信控停机服务应用

差异化信控停机服务的主要实现方式是依据用户信用评估模型的评估得分和具体通信需求,设计用户等级并给予差异化的信控额度,提供延迟停机服务,减少无谓或不当停机。实现流程上还可以优化增加信用额度告知、余额不足提醒及停机告知策略,提升用户在信控额度内的缴费率。

3.1 根据模型评分划分用户信用等级

按照信用评分将用户的信用划为4个等级,分别为很好(A)、良好(B)、好(C)、不好(D)。而每一个等级用户比

例的确定,主要参考移动停机用户 N 天复机率。该省公司把3天内复机的用户(占比50%)作为信用很好的用户,5天内复机的用户(占比10%)作为信用良好的用户,30天内复机的用户(占比10%)作为信用好的用户,30天后复机的用户(占比30%)作为信用不好的用户。考虑到是停机授信应用,除了控制风险以外,还需要针对真正具备通信需求的用户、最后信用等级划分,综合考虑信用评估模型得分、用户通信需求(活跃天数)、历史信用(最近6个月最长停机时长)3个方面。在该省公司应用时,最后得到的用户信用等级划分见表3。

3.2 各信用等级用户信控额度计算设定

通过历史数据分析得到,该省欠费金额在10~50元的用户停机后复机率最高,因此信控金额设置建议在10~50元。用户动态信控金额为:

$$\frac{Score_i \times A_i - \min(Score_i \times A_i)}{\max(Score_i \times A_i) - \min(Score_i \times A_i)} \times 50 \quad (6)$$

其中, $Score_i$ 为信用评估模型评分, A_i 为用户的最近3个

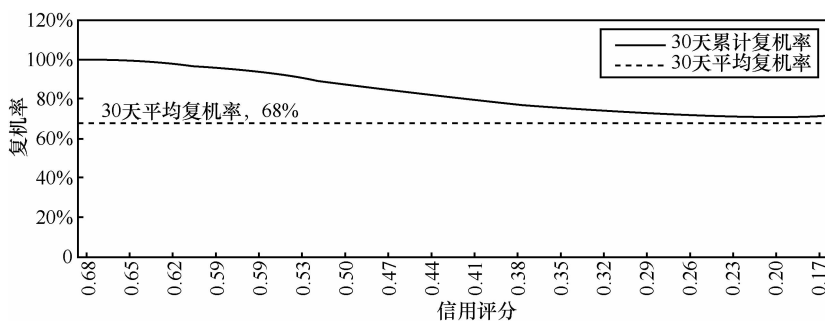


图3 DPI数据构建网页URL分类体系方案



表 3 某省公司评估模型目标用户信用等级划分结果

信用等级	用户规模/户	信用等级描述
A	228 972	模型评分前 50%的用户,且上月活跃天数 ≥ 18 天,最近 6 个月停机时长不超过 5 天
B	71 233	模型评分前 50%~60%的用户,且上月活跃天数 ≥ 8 天,最近 6 个月停机时长不超过 10 天
C	58 541	模型评分前 60%~70%的用户,且上月活跃天数 ≥ 1 天,最近 6 个月停机时长不超过 30 天
D	175 159	模型评分等级后 30%的用户以及月活跃天数及最近 6 个月停机时长不满足上述要求的用户

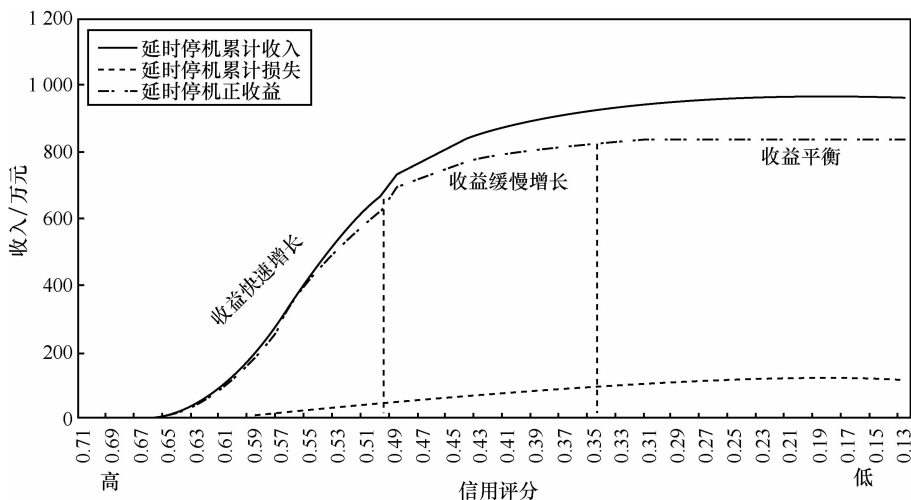


图 4 用户差异化信控停机服务后的收入增长示意

月的平均 ARPU 值。

实际应用中,将评级为 D 的用户授信直接置为 0 元, A/B/C 等级的用户按此计算式向上按 10 元取整数。

3.3 应用效果

图 4 是给不同信用得分用户授信实施差异化停机服务后,给企业带来的收入变化(收入与损失之差为延时停机给电信运营商带来的正收益)。可见,模型应用于用户差异化信控停机服务后,给企业带来的收益明显。

4 结束语

用户动态信用评估模型结合用户行为特征动态数据,较科学合理地计算得到用户信用评分。将其应用于存量用户的差异化信控停机服务,能适时满足用户通信需求,提升用户满意度与忠诚度,为企业增加可观的经济效益,具备一定的应用推广价值。

参考文献

- 1 翟清兰.基于 Logit 模型和人工神经网络的商业银行信用风险评估.东北财经大学硕士学位论文,2006
- 2 杨建刚.移动电话客户信用控制和管理.通信世界,2005(31):

30~31

- 3 张玉洁,孟祥武.基于遗传算法的电信客户初始信用度分配算法.北京邮电大学学报,2002,25(2):74~77
- 4 徐继生,五道恒.移动通信用户信用度的预测评估.电信科学,2002,18(8):17~20
- 5 吴国富,安万福,刘景海.实用数据分析方法.北京:中国统计出版社,1992
- 6 郭瑜敏,徐清宇.电信业客户信用控制过程的优化.电脑开发与应用,2009(8):70~71
- 7 朱晓明,刘治国.信用评分模型综述.统计与决策,2007(1):103~105
- 8 陈仁强.基于客户特征的信用控制模型.移动通信,2014(5):75~79

[作者简介]



边凌燕,女,中国电信股份有限公司上海研究院工程师,主要研究方向为数据分析、数据挖掘。

(收稿日期:2014-08-29)