

文章编号: 2095-2163(2019)03-0256-04

中图分类号: TP311.52

文献标志码: A

反欺诈中的用户手机号数据研究

石 强

(上海琰璟网络科技有限公司, 上海 201203)

摘 要: 本文对互联网金融反欺诈系统中的欺诈用户和正常用户的手机号数据进行研究, 针对手机号的号段、归属的电信运营商、省份城市等地域信息展开分析, 同时还详细分析了手机号对应的手机设备上一次插卡的电信运营商为海外运营商的特性, 将欺诈用户的画像更精细地呈现出来, 对互联网反欺诈系统和技术的优化升级具有较强的实际应用价值。

关键词: 反欺诈; 特征; 电信运营商; 地域

Research on the data of user's mobile phone number in anti-cheating

SHI Qiang

(Shanghai Yanjing Network Technology Co. Ltd., Shanghai 201203, China)

[Abstract] This paper studies the data of the mobile phone numbers of the cheating users and normal users in Internet financial anti-cheating system. The regional information of mobile phone numbers, telecom operators, provinces and cities are analyzed. At the same time, the characteristics of mobile phone numbers are also discussed whose last phone numbers in the same mobile devices correspond to overseas telecom operators. A more detailed portrait of the cheating user is presented. It has strong practical application value to the optimization and upgrading of Internet anti-fraud system and technology.

[Key words] anti-cheating; feature; telecom operators; geographical area

0 引言

近年来,随着互联网金融网络借贷的普及和发展,对于反欺诈作为风控的核心关键问题的研究也越来越深入。互联网金融中的网络借贷由于借款人来自线上,风险系数极高,而且目前各借贷平台的信息互不共享,借款人的违法骗贷成本极低,故借款人故意欺诈骗贷的概率很高^[1]。一旦通过了欺诈用户的借款,那么该笔借款的不良逾期风险极高,还款意愿也极低。目前,借款用户的手机号信息作为用户信息的主要属性^[2],是技术化防范金融欺诈的重要依据,对借款用户的手机号进行分析,可以发现欺诈的地域等特性,设置相关手机号黑名单或者风险概率,有助于业务的反欺诈。

本文依据大量已经确定为欺诈的用户和正常用户的手机号数据,对手机号的附属信息,比如地域、运营商等进行详细分析,将欺诈的画像予以更加精细的呈现,依据结果和结论,可以形成一些防范欺诈的手机号规则引擎^[3]或者反欺诈模型^[4-6],对行业的反欺诈应用具有重大的实际价值。

1 数据处理与特征分析

综合多种借贷产品的用户数据,通过历史的反

欺诈识别和风控验证,获得了数十万的欺诈用户和两百万左右的普通用户的手机号数据,对于每个手机号可以提炼出以下特征:欺诈用户或正常用户、手机号前2位、手机号前3位、手机号归属的电信运营商、手机号归属的省份、手机号归属的城市、手机号所属手机上一次插卡是否为海外电信运营商以及运营商名称等。对各特征、变量或指标进行单维度或组合维度的分析,可以发现一些欺诈用户具有的明显特性,同时可以了解目前用户数据情况并确定该特征是否明显不适合应用于反作弊模型。

1.1 手机号前2位或前3位的分析

通过对手机号前2位的分析,可得分析后结果如图1所示。由图1可以发现,17号段中的欺诈用户占比明显较高,欺诈用户占比为52.3%,正常用户占比48.7%,17号段明显较其它3个号段表现异常。18号段的欺诈用户占比很少,几乎可忽略。13号段和15号段均存在两、至三成的欺诈用户。同时,研究得出了正常用户和欺诈用户中不同手机号前2位的占比如图2所示。由图2中可以看到,从欺诈用户维度的号段分布看,大部分欺诈用户集中在13和17号段,对于正常用户的号段分布,除了欺诈用户占比很高的17号段,正常用户基本均匀分布在其他三个号段,正常用户分布在13、15、18号段上

作者简介: 石 强 (1984-),男,硕士,工程师,主要研究方向:数据分析与挖掘、大数据技术架构。

收稿日期: 2019-01-20

的占比均在 30%左右。

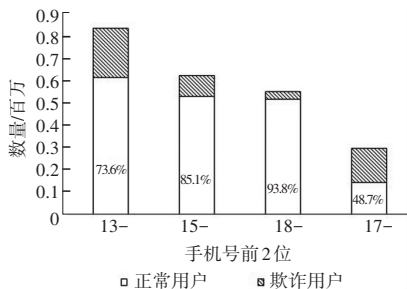


图 1 手机号前 2 位的正常用户和欺诈用户数量

Fig. 1 Number of the top two mobile phone numbers of the normal users and cheating users

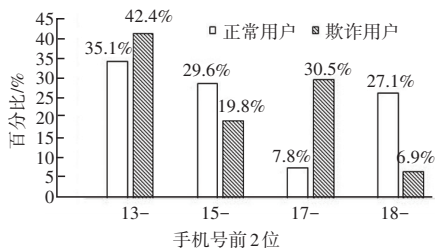


图 2 正常用户和欺诈用户中不同手机号前 2 位的占比

Fig. 2 The proportion of the top two mobile phone numbers in the normal users and cheating users

手机号前 3 位的正常用户和欺诈用户数量如图 3、图 4 所示。由图 3 和图 4 可以发现手机号前 3 位构成的号段的欺诈属性可以分为 4 种档次,比如第一档次为高风险档,主要是 170 和 171 号段,该号段的欺诈用户占比极高,170 号段不仅欺诈用户占比达 87.2%,同时 170 号段的用户绝对数量在所有号段里面也排第一,但 171 号段的用户绝对数量较小。第二档为中风险档,包括 130、131 和 132 三个号段,其中的欺诈用户占比近五成,而且用户绝对数量也很大,绝对数量位于 TOP4 内。第三档为低风险档,主要是 155 和 156 号段,欺诈用户占比近 3 成。第四档为无风险档,由其他所有号段构成。

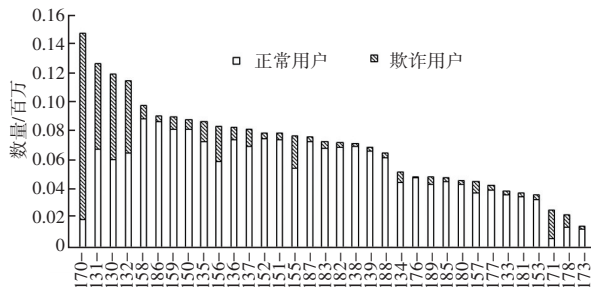


图 3 手机号前 3 位的正常用户和欺诈用户数量

Fig. 3 Number of the top three mobile phone numbers of the normal users and cheating users

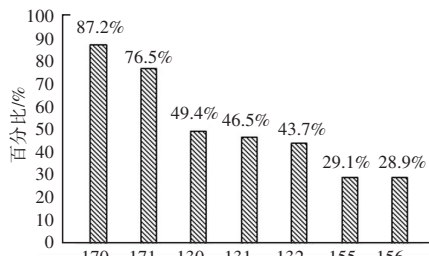


图 4 各手机号前 3 位中的欺诈用户占比

Fig. 4 The proportion of the cheating users in the top three mobile phone numbers

1.2 手机号归属的省份、城市、运营商

研究可知,欺诈用户的绝对数量较多的省份依次是:广东、浙江、广西、山东、四川。欺诈用户占比高,且绝对数量在一定的可信阈值之上的有:广西(49%)、浙江(38%)、广东(34%)、上海和四川(均在 25%左右)。欺诈用户的绝对数量较多的城市依次是:深圳、南宁、成都、广州等,这些城市的欺诈用户占比在 40%~60% 区间。

对于全体手机号集合,不同电信运营商中的欺诈用户占比如图 5 所示。由图 5 中可以发现中国电信和中国移动中的欺诈用户占比比较低,中国联通的欺诈用户占比最高,达 35.1%。从运营商角度向地域维度看,中国联通的浙江、广西、广东和上海的欺诈用户绝对数量最多,并且明显多于其它省份,广西南宁、浙江台州和广东深圳是中国联通中欺诈用户绝对数量排前三的城市。中国移动中,广东省的欺诈用户绝对数量最多,并且远远多于其它省份,广东省的深圳城市的欺诈用户数量最多。

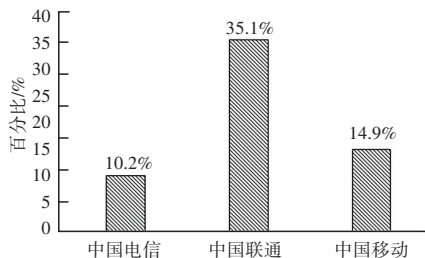


图 5 不同电信运营商中的欺诈用户占比

Fig. 5 The proportion of the cheating users in different telecom operators

从省份区域维度看运营商,广西省和浙江省的中国联通的欺诈用户占比均在 60% 以上,广东省的中国联通的欺诈用户占比也达 30%,比较高。在数量置信的条件下,广东、北京和四川的中国移动的欺诈用户占比在 30% 左右,是中国移动在全国省份范围内欺诈用户占比最高的省份。城市角度,北京在中国电信和中国联通中不含有欺诈用户,而中国移动的欺诈用户占比在 26% 左右。但是上海却在中

国电信和中国移动中不含有欺诈用户,而在中国联通中的欺诈用户占比一半。武汉的三大运营商中均不存在欺诈用户,说明按照城市维度,各城市的情况具有多样性,对于识别欺诈用户可以抽取细分的维度和规则去考虑。

1.3 手机设备上一次插卡为海外电信运营商的手机号数据分析

首先从欺诈的角度总体统计,对于目前手机号所属手机设备上一次插卡为海外电信运营商的,其中欺诈用户占比 51.5%,正常用户占比 48.5%,虽然在本集合中,欺诈用户占比与正常用户的占比相差不大,但是在产品全体手机号数据集合中,确定的欺诈用户占比不到 26%,也就是说目前手机号所属手机设备上一次插卡为海外电信运营商的风险明显偏高,可能是欺诈集团统一从海外收购了大批量手机来参与欺诈运作,本特征可以作为一个欺诈的高概率性特征。

目前,所属手机设备上一次插卡为海外电信运营商的手机号都会归属不同的电信运营商,对于本集合手机号,中国电信、中国联通、中国移动中欺诈用户占比分别为 56%、73%、27%;本集合手机号绝对数量方面,中国联通的数量也最多,与中国电信和中国移动的数量之和相当。可以发现本集合的手机在转入国内使用时,欺诈用户会优先使用中国联通。同时,从图 6 中可以看到本集合中的欺诈用户手机号分散在联通的各号段,并不仅是在其中的 170 或 171 虚拟运营商号段,且分布比例相对均匀,无明显的异常占比的特例。对于本集合中的手机号归属中国电信的,欺诈用户也是均匀分布在电信各号段,但是对于归属于中国移动的,欺诈用户集中在 170 号段(中国移动仅在 170 号段有虚拟运营商),所有欺诈用户的 40% 出现在 170 号段,同时 170 号段中正常用户仅占比 10.5%,其它中国移动号段中正常用户占比都在 80% 左右。

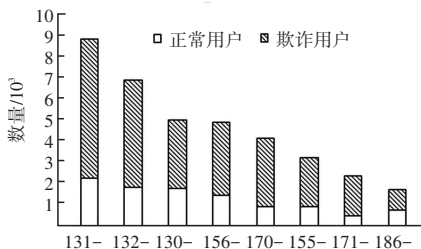


图6 所属手机设备上一次插卡为海外电信运营商的手机号前3位的正常用户和欺诈用户数量

Fig. 6 Number of the top three mobile phone numbers whose device was last plugged in by overseas telecom operators of the normal users and cheating users

正常用户和欺诈用户的目前手机号所属手机上一次插卡为海外电信运营商的数量比较结果如图7所示。从图7可以看到,目前手机号所属手机上一次插卡为海外电信运营商的前五大运营商中,只有KDDI中的欺诈用户比例较少,仅占 26.1%,其它的海外电信运营商中均是欺诈用户较多。关于所属手机上一次插卡为海外电信运营商的手机号目前的省份和城市分布,欺诈用户的绝对数量较多的省份依次是:广西、广东、浙江、山东、上海。欺诈用户占比较高,大于 60% (且绝对数量在一定的可信阈值之上)的有:广西(71%)、浙江(66%)、上海(80%)。欺诈用户的绝对数量较多的城市依次是:南宁、上海、东莞、北京,这些城市的欺诈用户占比均在 80% 左右。

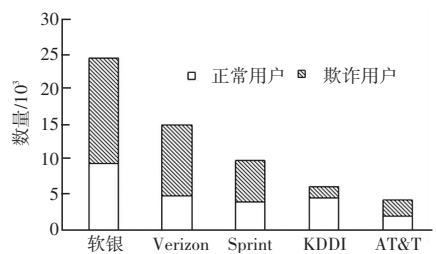


图7 正常用户和欺诈用户的目前手机号所属手机上一次插卡为海外电信运营商的数量比较

Fig. 7 The number comparison of current mobile phone numbers whose device was last plugged in by overseas telecom operator between the normal users and cheating users

如前文所述,通过对用户手机号数据的深入分析可以发现虚拟运营商等存在欺诈用户的可能性很高,特别是 170 和 171 号段,可以作为判断为欺诈用户的高概率特征。更深入地,比如中国联通的浙江、广西、广东和上海等较多区域欺诈风险较高,中国移动的少数区域存在欺诈风险,且概率不高。所属手机设备上一次插卡为海外电信运营商的欺诈特性更为明显,手机设备大多来自距离中国较远的日本(软银运营商),手机设备在转入国内使用时,欺诈用户会优先使用中国联通,特别是其虚拟运营商号段。此外,还可以发现一些基本不反映欺诈的维度或变量,比如 188 号段、中国电信运营商等,这些维度在识别欺诈用户时可以建议不予考虑,避免冗余影响。

2 结束语

本文研究了互联网金融反欺诈系统中的欺诈用户和正常用户的手机号数据,对手机号的附属信息, (下转第 262 页)