

基于XDR数据的双卡槽终端用户类型识别方法研究

Study on Identifying Terminal
User Types for Dual-SIM
Cards Based on XDR Data

陈玉锋¹,蔡子龙¹,曾 炜²,李广彬¹,司马超¹(1. 中讯邮电咨询设计院有限公司郑州分公司,河南 郑州 450007;2. 中国联通江西分公司,江西 南昌 330096)

Chen Yufeng¹,Cai Zilong¹,Zeng Wei²,Li Guangbin¹,Sima Chao¹(1. China Information Technology Designing & Consulting Institute Co.,Ltd. Zhengzhou Branch,Zhengzhou 450007,China;2. China Unicom Jiangxi Branch,Nanchang 330096,China)

摘要:

提出了一种用于识别双卡槽终端用户类型的分析方法,基于4G/5G网络XDR数据进行分析,通过对终端类型一致性、信令事件的时空同步性、业务量特征、行为特征、疑似Wi-Fi用户的分析,挖掘出属于同一终端的IMEI对、主副卡关系等,进而识别出双卡本网用户、单卡用户、异网主卡本网副卡用户。为存量用户的维系保有、副卡转化提升、二卡槽用户拓展等市场精细化运营提供参考。

关键词:

双卡槽终端;XDR;用户类型识别;信令;时空同步
doi:10.12045/j.issn.1007-3043.2024.12.009
文章编号:1007-3043(2024)12-0054-06
中图分类号:TN929.5
文献标识码:A
开放科学(资源服务)标识码(OSID):



Abstract:

It proposes an analysis method for identifying the types of dual-card slot terminal users. Based on 4G/5G network XDR data analysis, through the analysis of terminal type consistency, signaling event spatiotemporal synchronization, traffic characteristics, behavioral characteristics, and suspected Wi-Fi users, it can mine the IMEI pairs belonging to the same terminal, the relationship between the primary and secondary cards, and identify the dual-card local users, single-card users, and the users of the primary card in different networks and the secondary card in the local network, which provides reference for the maintenance and retention of existing users, improvement of secondary card conversion, and expansion of secondary card slot users in the market refinement operation.

Keywords:

Dual-SIM card slot terminal; XDR; User type identification; Signaling; Spatiotemporal synchronization

引用格式:陈玉锋,蔡子龙,曾炜,等. 基于XDR数据的双卡槽终端用户类型识别方法研究[J]. 邮电设计技术,2024(12):54-59.

1 概述

随着移动通信的发展,越来越多的人拥有2张以上的电话卡,其中有一些人拥有属于不同运营商的电话卡,双卡手机也已成为主流的配置方案。“BCI通信研究”的数据统计显示双卡终端市场份额由2016年初的80%上升到2018年7月的90%^[1]。法国权威市场调研公司Reportlinker发布了“全球双卡智能手机产业(Global Dual SIM Smartphones Industry)”报告,该报告

指出,到2030年全球双SIM卡的智能手机将达到46亿部,在2022—2030年期间的复合年增长率为14.3%;而中国将达到7.922亿部,在2022—2030年期间的复合年增长率为13.6%^[2]。

对于运营商来说,精确识别双卡槽终端中2张卡的业务状态,是进行市场拓展、深耕客户黏性、精准定位客户需求的基础。目前,运营商的网络数据中只包括注册到本运营商网络的SIM卡、IMEI信息,而不清楚在网的双卡槽终端是否插了双卡,无法直接得知哪2张卡是属于同一终端的2个卡槽、双卡的业务状态、主副卡关系,无法区分出异网主卡终端。

收稿日期:2024-10-09

本文基于XDR信令面和用户面数据开展大数据融合分析,通过对终端类型一致性、信令发生的时空同步性、业务量特征、行为特征、疑似Wi-Fi用户的分析,匹配出属于同一终端的IMEI对,进而对4G/5G双卡槽终端的双卡业务类型进行判断,对用户类型进行区分^[3-4]。以中国联通为例,双卡槽终端用户分为以下几种类型:单卡中国联通用户、双卡双中国联通用户、

中国联通主卡异网副卡用户、异网主卡中国联通副卡用户。主要目标是区分出异网主卡、中国联通副卡的用户,从而进行市场的营销活动^[5](见图1)。

2 算法介绍

2.1 数据源

数据源主要包括如下几项。

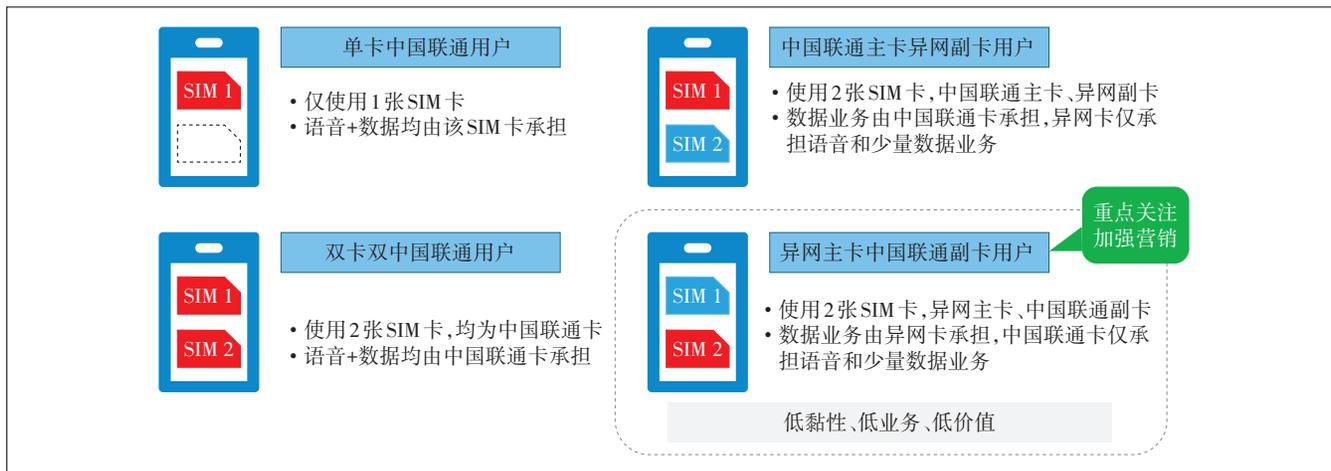


图1 双卡槽终端中国联通用户情况

a) XDR信令面和用户面部分接口的话单数据(见表1)。

表1 XDR话单数据

网络类型	接口	话单类型					
		Attach	Detach	TAU	Handover	-	-
4G	S1_MME	Attach	Detach	TAU	Handover	-	-
	S1_U	http	streaming	im	game	other	-
5G	N1N2	Registration	De_registration	-	Handover	-	-
	N3	http	streaming	im	game	other	-

b) 终端库。包含终端TAC、终端类型、是否双卡等信息。

c) 4G和5G工参。包含省份、地(市)、网络类型、小区编码(eci/gci)、经纬度等信息。

2.2 分析整体架构

本算法利用XDR数据、终端库、4G/5G工参为基础,提取出用户IMEI的多维特征,搭建起一套完整的算法模型,找到高可信的双卡终端IMEI对的匹配关系以及主副卡关系,从而将双卡终端用户按照不同类型进行区分。分析整体架构如图2所示。

2.3 算法思路和流程

算法实现思路和分析流程如图3所示。

第1步,对全网的4G/5G终端的XDR信令面IMEI进行汇总生成全量终端IMEI集。

第2步,全量IMEI集匹配终端库,筛选出双卡槽终端IMEI集作为分析对象。

第3步,进行双卡终端IMEI对匹配分析,IMEI对匹配分析是该算法的核心和难点,涉及信令同步特征分析、业务量特征分析、行为特征分析、IMEI号码分析等多个关键环节,通过构建算法模型,挖掘出属于同一终端的IMEI对,这部分IMEI属于双卡双中国联通用户,再结合业务量特征区分出主副卡。

第4步,对未成功配对的IMEI,结合用户面流量进行分析,如果用户面有数据流量,则识别为单卡中国联通用户。

第5步,对用户面未发生流量业务的IMEI,结合其活动范围分析识别出疑似Wi-Fi用户和异网主卡中国联通副卡用户。

2.4 双卡终端IMEI对匹配分析

2.4.1 IMEI特征

根据GSMA的规定,双卡手机的每个卡槽对应1个独立的IMEI,IMEI的前8位为该款终端的TAC(Type Allocation Code),用以区分手机品牌和型号的

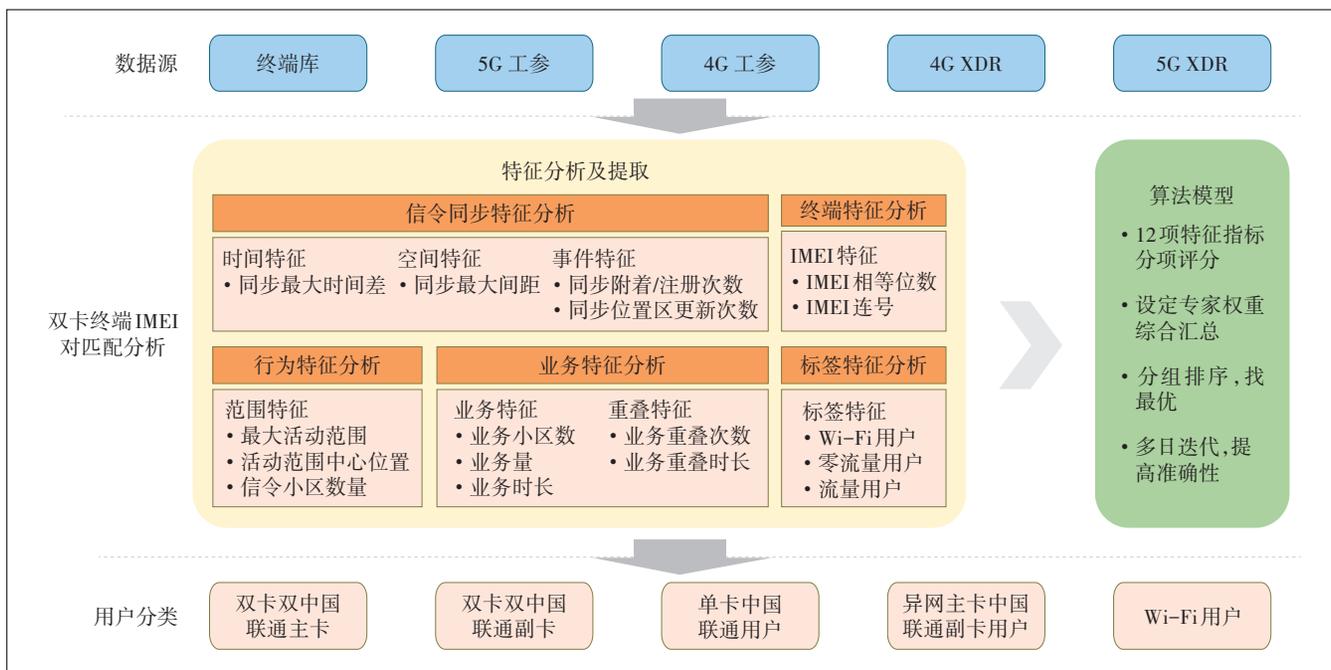


图2 分析整体架构

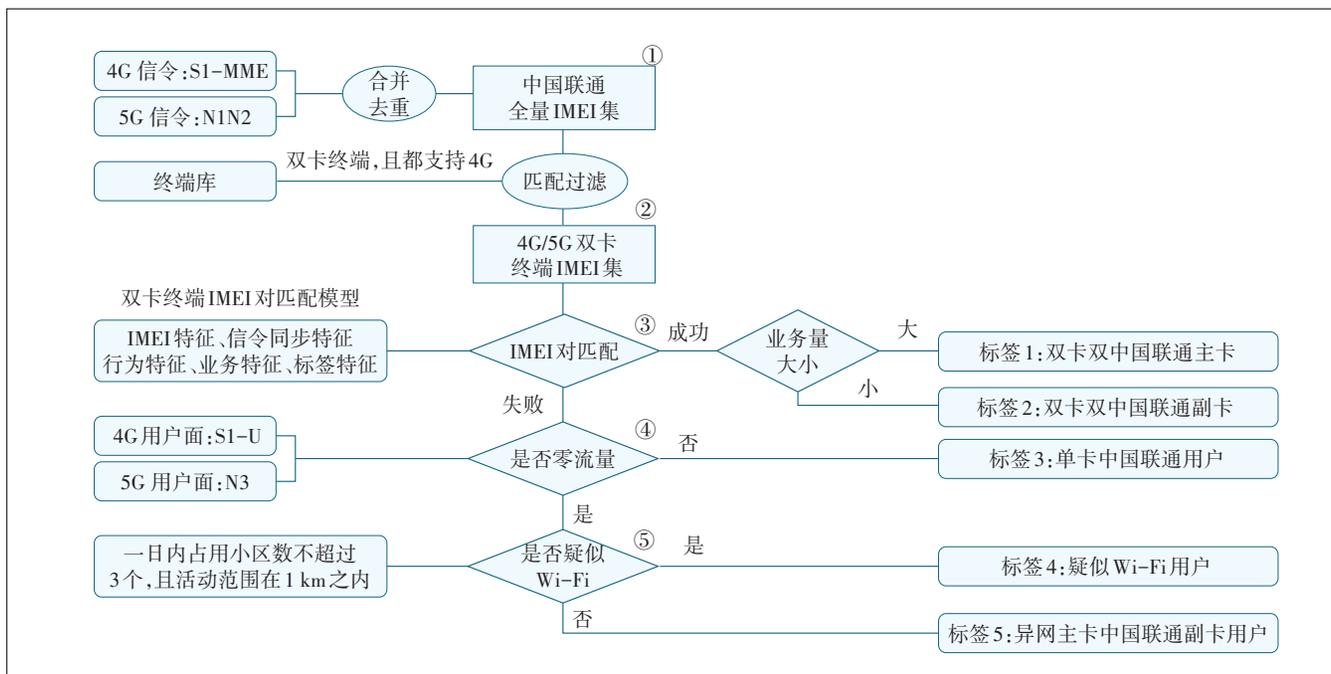


图3 分析流程

编码^[6]。2个IMEI属于同一部手机,则必须满足前8位(TAC)相同。另外,据统计发现双卡手机的2个IMEI不仅仅前8位(TAC)相同,大部分手机IMEI前10位都是一致的,有些双卡手机仅后2位不同,甚至有连号的情况,所以2个IMEI的相似程度对是否属于同一终端判断具有一定参考价值。

根据以上分析,通过对2个IMEI的简单对比分析可得出IMEI相等位数、IMEI是否连号2项IMEI特征。

2.4.2 信令同步特征

根据核心网信令流程来看,双卡手机的2个IMEI具有较高的信令同步特征,如同步开关机、附着、去附着、位置区更新等。所以一般情况下,在开关机、附

着、去附着、位置区更新等信令事件上发生时空同步的次数越多,则这2个IMEI属于同一手机的概率越大。

本算法对S1-MME和N1、N2接口的信令(Attach、Detach、TAU、Registration、De_registration)话单进行合并,以小区eci和gci为关联字段,与4G、5G工参进行关联,用以获取小区的经纬度信息。再按照设定的时间、空间阈值,对TAC相同(IMEI不等)的IMEI两两进行时空同步判断,最终得到A、B 2个IMEI的时空同步关系表。时空同步关系表包含A、B各自IMEI号、信令开始时间、小区编号、经纬度、时间差、小区间距离等信息,时空同步判断流程如图4所示。

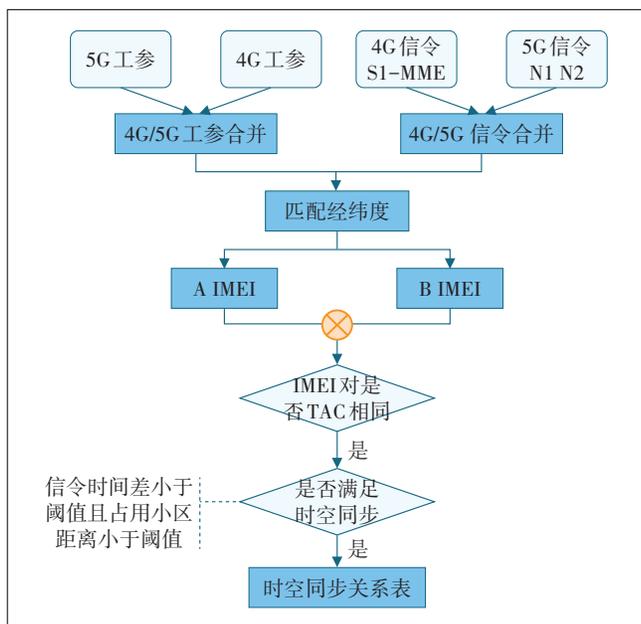


图4 信令同步判断流程

在图4中,时间同步是指信令时间差值在预设阈值范围内,建议为正负10s,可根据情况调整;空间同步是指发生信令事件时双方所占用的小区距离小于预设阈值,建议1km;4G、5G网络信令流程有所不同,4G的附着(attach)对应5G的Registration(type为1:初始化注册),去附着(detach)对应5G的De_registration;4G的位置区更新(TAU)对应5G的Registration(type为2:移动更新注册),详见表2。

根据以上时空同步关系判断,可以输出信令同步的4项特征指标,即同步最大时间差、同步最大间距、同步附着/注册次数、同步位置区更新次数。具体特征指标见第2.4.6条中的表3。

2.4.3 行为特征

表2 信令数据筛选规则

网络	信令名称	类型	value	过滤条件
4G	ATTACH	EPS attach type	①EPS attach	无
			②联合EPS/IMSI attach	
			③EPS紧急附着	
			⑦reserved	
	DETACH	Detach type	①EPS分离	无
			②IMSI分离	
③联合分离(UE发起的)				
TAU	SRVTYPE	①TA更新	去除③	
		①联合TA、LA更新		
		②带IMSI attach的联合TA、LA更新		
		③周期性更新		
		④、⑤同①		
5G	Registration	Registration type	①initial registration(初始注册)	①对应4G ATTACH ②对应4G TAU
			②mobility registration updating(移动更新注册)	
			③periodic registration updating(周期更新注册)	
			④emergency registration(紧急注册)	
De_registration	Deregistration type		①Normal deregistration(正常去注册)	无
			①Switch off(关闭5G)	

在判定A、B 2个IMEI的时空同步关系的可信度时,不仅要关注A与B相互间的同步特征,还要基于单IMEI的独立行为特征(如A、B各自的活动范围)进行侧面验证。如果两者独立行为特征差异较大,则说明之前判断的时空同步关系只是偶然发生的巧合,属于小概率事件,双卡时空同步可信度较低。

由于无线通信的复杂性,用户的精准位置很难获得。本算法采用4G、5G信令面话单(Attach、Detach、TAU、Registration、De_registration、Handover)中的服务小区位置代替用户位置用以计算用户行为特征,包括单日活动的最大距离、活动范围中心点位置、单日接入的小区数等。在此基础上,结合时空同步关系表计算A与B之间的活动范围差值和活动范围中心点距离指标,具体特征指标见第2.4.6条中的表3。

2.4.4 业务特征

在使用双卡手机上网的实际过程中,当A卡发生流量业务时,B卡必然处于空闲状态,因此,大多数双卡终端会出现2卡流量不均衡的情况,即1张卡的流量较大而另外一张卡的流量相对较小,比如1个IMEI流量小于100MB,另外一个IMEI流量大于1GB,则说明这对属于同一终端的双卡在业务特征是合理的,

反之则认为不合理。

本算法中用户的业务特征采用4G、5G用户面话单中业务量和业务小区数量进行评判。用户业务特征指标需要统计出该用户一天内的用户面话单(http、streaming、im、game、other)总的业务次数、业务时长、业务流量等指标。在进行同一终端IMEI对匹配时,可比较2个IMEI的业务量是否均衡来进行判定。此外,如果双卡手机A、B 2个卡的业务发生时间存在较长的重叠,则可认为这2个卡同属于一个终端的可能性较低。因此增加业务量是否均衡、业务重叠时长和业务重叠次数作为业务特征指标,具体特征指标见第2.4.6条中的表3。

2.4.5 标签特征

对于移动终端来说,上网的途径一般分为2种,一种是通过运营商提供的无线网络上网,另外一种就是利用家庭Wi-Fi上网。在实际分析中,Wi-Fi用户与副卡用户在业务特征上类似,产生的业务流量均较少或者为零,这会对流量均衡以及副卡的判断产生干扰。因此需要对Wi-Fi用户进行识别剔除,减少误判。

在现实生活中,Wi-Fi信号的覆盖范围一般在50m之内,这说明Wi-Fi用户的终端在移动范围上有明显的局限性。因此如果用户的活动范围较小而且流量也较低,基本可以判定为Wi-Fi用户。考虑到无线通信环境的复杂性,出现信号的波动和切换的情况也属正常。经常出现终端所占用的小区并不是距自己最近的小区的情况,而且有时候即使终端不移动也会存在与多个小区之间产生切换的情况。因此本算法中将单日业务流量为零的用户,且单日接入小区数不超过3个,活动范围小于1km的用户判定为Wi-Fi用户,从零流量用户中剔除。

据此,可将用户按流量用户、Wi-Fi用户和零流量用户区分,分别打上标签。标签判别流程如图5所示。

根据前文提到的流量不均衡原则,在进行同一终端IMEI对匹配时,可比较A、B 2个卡的用户流量标签,当其中一个卡为流量用户,另外一个卡为零流量用户,则属于同一终端的可能性较高,这类标签组合作为同一终端IMEI对匹配的标签特征,具体特征指标见第2.4.6条中的表3。

2.4.6 算法模型

综合以上特征指标分析,构建专家打分模型,计算每个IMEI对匹配综合得分。各分项特征建议权重见表3。

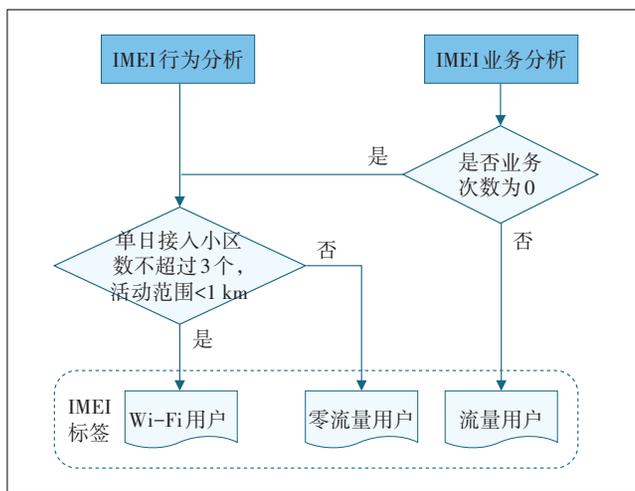


图5 标签判别流程

表3 配对特征指标表

配对特征分类	配对特征指标	指标单位	数据类型	专家权重/%
信令同步特征	同步最大时间差	s	浮点小数	10
	同步最大间距	m	浮点小数	10
	同步附着/注册次数	次	整数	10
	同步位置区更新次数	次	整数	10
IMEI特征	IMEI相等位数	位	整数	5
	IMEI是否连号	-	布尔	5
行为特征	活动范围差值	m	浮点小数	5
	活动范围中心距	m	浮点小数	5
业务特征	是否流量不均衡	-	布尔	10
	业务时间重叠次数	次	数字	10
	业务时间重叠时长	s	浮点小数	10
标签特征	是否零流量和流量用户	-	布尔	10

由于在进行同一终端IMEI对匹配过程中难免会出现1对多的情况,需根据综合得分进行分组排序,取组内得分最高的一对。以上分析以日为时间周期进行,考虑到单日数据存在偶然性的可能,本算法建议采用多日数据进行迭代,以提高结果的准确性。

此外,待掌握了足够的、真实有效的双卡终端IMEI对样本数据后,可借助机器学习构建新的算法模型,对算法进一步迭代,提高识别的准确性。

2.4.7 用户类型识别结果

根据算法模型输出结果,满足判断条件的IMEI对识别为双卡双中国联通用户,其中数据流量较大的IMEI为双中国联通主卡用户,数据流量较小的为双中国联通副卡用户。对于IMEI集中的其他IMEI,再结合用户流量标签进行分类判别,将标签为流量用户的识别为单中国联通用户,将标签为Wi-Fi用户的识别

为Wi-Fi用户,标签为零流量用户的识别为异网主卡中国联通副卡用户。

为验证方法的有效性,采集中国联通某地(市)3天的XDR数据进行验证。全网终端IMEI数量为200.7万。其中178.4万IMEI包含附着、去附着、注册、去注册和位置区更新、切换等完整信令数据,属于双卡终端的IMEI为141.9万。使用算法模型识别出双卡双中国联通终端IMEI数量为2.4万,单卡中国联通终端IMEI数量为117.65万,单卡中国联通用户数为117.65万,Wi-Fi用户数为9.79万,竞对主卡中国联通副卡用户数为12.06万,用户分类结构如图6所示。

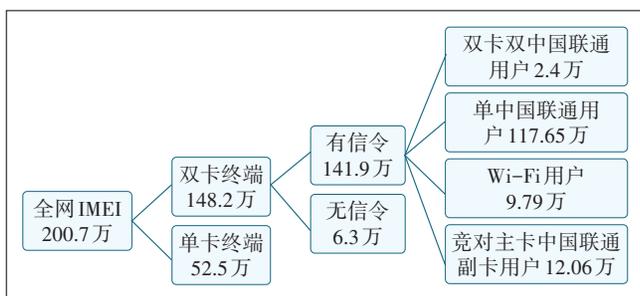


图6 中国联通某地(市)用户分类

经验证,该方法能够有效地识别出双卡槽终端不同的用户类型,准确率达到76%。

2.5 数据应用方向及建议

双卡槽终端用户识别的分析结果可为运营商市场精细化运营提供支撑。可以着重提升双卡用户的满意度,建立稳固的双卡用户基础,增加用户黏性;针对目标用户,开展精准营销,进行二卡槽拓展和副卡转化提升,不断扩大市场份额。本文给出如下建议。

(a) 建立持续激励机制,深化双卡用户忠诚度。

(a) 积分奖励与会员特权。为长期双卡用户提供积分累积和会员升级制度,让双卡用户享受更多优惠和独家特权,从而激励其持续使用。

(b) 深入双卡用户洞察。通过精细化的数据分析,深入了解双卡用户的行为、习惯和偏好,为后续提供个性化服务奠定基础。

(c) 定期用户反馈调研。定期主动与双卡用户互动,收集他们的需求和反馈,确保运营商的产品和服务始终贴近用户需求。

b) 精准营销,扩大双卡用户群体。

(a) 明确定位目标用户。利用大数据分析,明确双卡手机的主要用户群体,如商务旅行者、多卡需求者等,进行有针对性的营销活动。

(b) 定制化套餐与服务。根据目标用户的需求,

推出专属的通话、流量、国际漫游等优惠套餐以及一系列增值服务。

c) 加大优惠吸引力度,推动副卡转化升级。

(a) 独特特权吸引。提供一系列独家特权,如新功能的优先体验、专属客服支持、免费网络提速等,增加用户转化的动力。

(b) 优惠副卡套餐。为副卡槽用户提供极具竞争力的通话资费、流量优惠和国际漫游服务,鼓励其考虑将副卡切换升级。

3 结束语

本文提出了一种用于识别双卡槽终端用户类型的方法,该方法基于4G/5G网络XDR数据进行分析,通过对终端类型一致性、信令事件的时空同步性、业务量特征、行为特征、疑似Wi-Fi用户的分析,挖掘出属于同一终端的IMEI对、主副卡关系等,进而识别出双卡本网用户、单卡用户、异网主卡本网副卡用户。为存量用户的维系保有、副卡转化提升、二卡槽用户拓展等市场精细化运营提供条件。未来,笔者将继续优化算法性能,在掌握了足够准确的样本数据后,借助机器学习构建算法模型提高识别的准确性,为双卡槽终端的个性化服务和产品提供更可靠的依据。

参考文献:

- [1] 陈丰伟. 双卡手机简史(上)[EB/OL]. [2024-06-29]. <https://m.luxiu.com/article/264978.html>.
- [2] 周雅. 中国人的手机里,为什么要装两张SIM卡?[EB/OL]. [2024-06-16]. <https://baijiahao.baidu.com/s?id=1768837911677161784&wfr=spider&for=pc>.
- [3] 张一博. 基于数据挖掘的双卡用户识别模型研究[J]. 山西科技, 2018, 33(5): 67-69.
- [4] 赵建川, 崔忠伟. 双卡双待移动终端主卡挂起副卡过程研究[J]. 贵州师范学院学报, 2016, 32(9): 7-12.
- [5] 刘北阳, 万玉海, 双程, 等. 基于终端数据运营平台的市场分析方法及实践[J]. 信息通信技术, 2018, 12(3): 58-64.
- [6] GSMA. IMEI allocation and approval process: GSMA TS.06[S]. London: GSM Association, 2016.

作者简介:

陈玉锋,高级工程师,学士,主要从事移动通信无线网络规划设计、网络数据分析等工作;蔡子龙,高级工程师,硕士,主要从事移动通信无线网络规划设计、网络数据分析等工作;曾炜,工程师,学士,主要从事移动通信网络建设规划相关方面的工作;李广彬,高级工程师,硕士,主要从事移动通信无线网络规划设计、网络数据分析等工作;司马超,工程师,学士,主要从事移动通信无线网络规划设计、网络数据分析等工作。