

基于大数据分析的移动网城市

Research and Application of City Portrait
of Mobile Network Based on Big Data
Analysis

画像研究与应用

傅俊锋, 孙红 (中国联通网络技术研究院, 北京 100048)

Fu Junfeng, Sun Hong (China Unicom Network Technology Research Institute, Beijing 100048, China)

摘要:

目前通信行业市场竞争激烈、营收增长乏力,因此网络精细化运营被提到了很高的高度。要做到运营的精细化,首先需要针对不同城市进行差异化分析。通过大数据分析提出了一系列指标,对城市的用户、业务进行了精准刻画。同时,通过对用户、业务进行数学建模,预测下一周期城市画像的变化,为把握市场变化规律、提前对网络进行规划建设提供了一个有效的手段。本方法已在中国联通的“精准建设”中得到了广泛应用。

Abstract:

At present, the competition in the communications industry is fierce and the revenue growth is weak. Therefore, the meticulous operation of the network has been raised to a very high level. To achieve meticulous operation, the cities should be differentiated first. It puts forward a series of indexes through big data analysis, and accurately depicts the users and businesses in the cities. At the same time, by modeling the user and business, the change of the city portrait can be predicted, which provides an effective means for grasping the law of market change and planning and building the network in advance. It has been widely used in the precision construction of China Unicom.

Keywords:

Big data analysis; Decision tree; City portrait; Business forecast; Meticulous operation; Precise construction

引用格式:傅俊锋,孙红. 基于大数据分析的移动网城市画像研究与应用[J]. 邮电设计技术, 2019(6): 64-69.

0 引言

在移动互联网充分发展和提速降费持续推进的背景下,当前的个人移动通信市场,从用户数量上来看已经接近峰值;从用户结构来看4G用户占比也渐趋稳定。因此对于运营商而言,迫切需要利用大数据分析方法,对省市用户、业务有一个清晰的画像,从而进一步挖掘现有用户潜力、提升业务量和收入。

本文以单个城市为基本单元,提出了一套城市画像方法,用于对不同城市的用户、业务现状进行刻画,

通过对比,了解其在市场和网络方面的优势和短板。在此基础上以2G用户转网为例,研究用户、业务变更的规律,从而实现对不同城市在下一周期内发展趋势的预测。基于现状和预测,运营商在市场推广上可以早作准备,抓住市场发展先机;在网络建设上能提前部署,及时满足业务需求。

1 城市画像方法

用户画像是根据用户社会属性、生活习惯和消费行为等信息而抽象出来的一个标签化的用户模型。构建用户画像的核心工作是给用户贴“标签”,而标签是通过用户信息分析而来的高度精炼的特征标识。

收稿日期:2019-03-28

所谓城市画像,既包含用户画像的概念,又比用户画像更宏观,偏向于从整体上对城市的特点、规律进行把握。本文提出的城市画像方法从用户及业务总量、用户模型、用户结构这3个角度进行分析,通过提炼的关键指标对单个城市进行刻画,最终通过散点图和雷达图的方式将城市画像的最终结果直观地呈现出来。

下面以中国联通A省及A1市、B省及B1市为例,对其画像指标进行对比分析(原始数据已通过函数进行了变换,与现网真实情况无关)。

1.1 总量指标

总量指标反映了城市的用户和业务规模的大小,具体指标取定:用户数、收入。

2省的用户数、收入指标分布如图1所示(A省为橙色,B省为蓝色)。

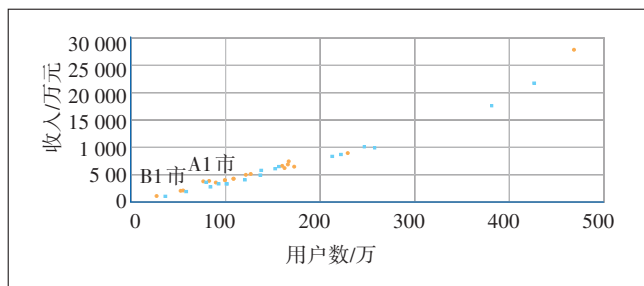


图1 2省用户数、收入指标分布

从图1中可以看出,收入指标同用户数指标呈正比例分布;2省大部分地(市)的分布区域比较接近,各有2个地(市)规模较为突出;A1市、B1市在所在省内规模都相对较小。

1.2 模型指标

模型指标反映了城市内用户的业务使用情况,具体指标取定:平均ARPU、平均DOU、平均MOU。

2省的平均ARPU、平均DOU指标分布见图2。

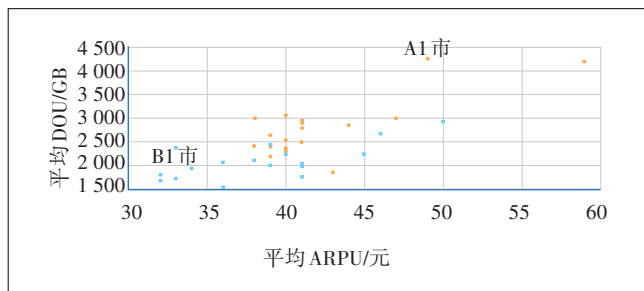


图2 2省平均ARPU、平均DOU指标分布

A省平均DOU普遍高于B省,说明A省用户使用

数据流量更为活跃;B省虽然平均DOU较低,但也有部分地(市)平均ARPU分布在头部;A1市在A省内的模型指标数值相对较大,平均DOU更是排在第1位;B1市在B省内的这2个指标都分布在尾部,用户较不活跃。

2省的平均DOU、平均MOU指标分布见图3。

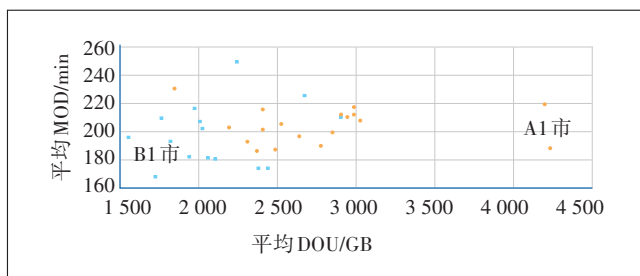


图3 2省平均DOU、平均MOU指标分布

2省平均MOU分布区域接近;B省有2个地(市)平均MOU显著高于2省其他地(市),而A省则有2个地(市)平均DOU显著高于2省其他地(市),说明B省少数地(市)侧重于语音业务;A1市的用户使用数据业务最多,而B1市语音、数据业务均不活跃。

1.3 结构指标

结构指标反映了城市内不同用户群的构成,具体指标取定:平均年龄、移动互联网用户占比、2G用户占比、2I2C用户(互联网套餐)占比。

2省的平均年龄、移动互联网用户占比指标分布如图4所示。

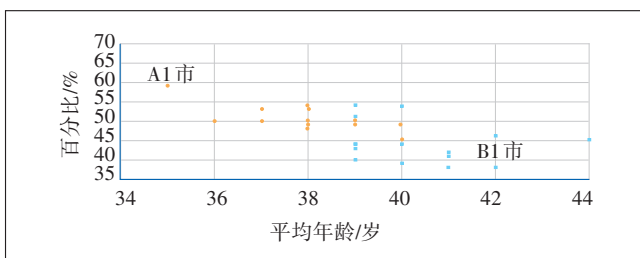


图4 2省平均年龄、移动互联网用户占比指标分布

A省用户普遍比B省年轻,其移动互联网用户占比更高;A1市平均年龄在2省各地(市)中分布在头部,相对更年轻,大部分用户都是移动互联网用户;B1市平均年龄分布在尾部,移动互联网用户也较少。

2省的2G用户占比、2I2C用户占比分布见图5。

A省2I2C用户占比普遍高于B省;B省部分地(市)2G用户占比偏高,也存在部分地(市)2G用户占比低;A1市2I2C用户占比最高,2G用户占比分布在

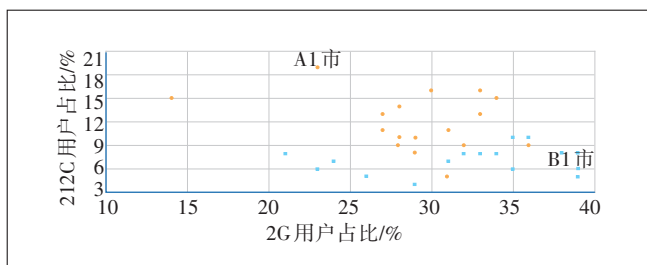


图5 2省2G用户占比、2I2C用户占比指标分布

头部,而B1市则正好相反。

1.4 城市画像指标汇总

最终的画像结果如图6所示(橙色为城市级,蓝色为省级,紫色为全国级;为方便图形显示及对比,以三者最大数值为单个指标轴的最高点)。从分布形状来看,A、B 2省的差异十分明显,A1、B1 2市的差异更加显著。

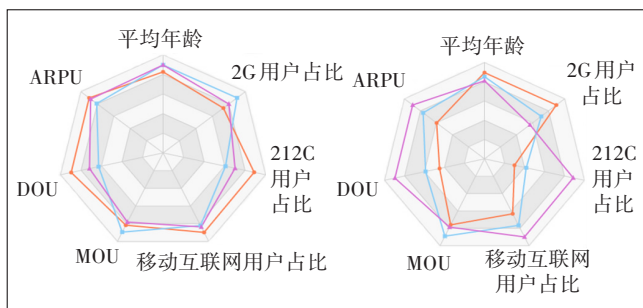


图6 A省及A1市、B省及B1市画像

2省用户及业务规模相当,但由于用户结构的较大差异,A省用户使用数据业务更多,B省用户则更偏向于语音业务。A省的收入主要由年轻的移动互联网用户产生,而B省的收入主要依赖于年龄偏大的语音型用户。

A1、B1 2市的用户及业务规模在省内都较小,但其用户构成完全不同:A1市的2I2C用户占比、平均DOU明显高出全国平均水平,其他指标同全国平均水平相近;而B1市的移动互联网用户占比、2I2C用户占比、平均ARPU、平均DOU都很低,而2G用户占比很高。

因此初步的结论是,B省及B1市需加大年轻用户的市场发展力度,同时利用市场手段将2G用户转为4G,在提升移动互联网用户占比的同时,也能提升其平均ARPU。

2 2G用户转网模型分析

在对不同城市进行精准画像的基础上,运用大数

据分析的方法,对用户行为的规律和原因进行深入分析,通过市场手段有针对性地对用户进行引导,从而能预测出下一个周期不同城市的画像变化情况。根据城市画像的变化,还可以进一步细化出用户和业务在地理上的分布,为网络的精准建设提供数据支撑。

中国联通运营4G网络已经多年,但目前仍有相当比例的2G网络用户,不少本地网2G网络用户占比达到50%。大量的2G用户导致2G网络无法退网,长期运营3张无线网络,运营成本居高不下。2G网络还占用了宝贵的低频资源,无法释放出来用于建设4GL900网络。因此,2G用户的转网对于提升ARPU值和企业收入、降低运维成本、提高4G建网效率具有重要的意义,是中国联通未来几年更好更快发展的一个关键因素。

本文以2G用户转网为例,通过SPSS Modeler对历史数据进行建模分析,得出影响2G用户流向的关键因素,预测哪些2G用户更有可能转为3G/4G用户,哪些2G用户会离网。对转网后的用户,通过统计其转网前后业务量的变化,从而得到实施2G转网对不同城市画像指标的影响。

2.1 模型指标

提取A省、B省最近2年的部分2G用户数据进行建模,根据2G用户流向的变化,打上4类标签:未转网、离网、转为3G、转网4G。用于建模的主要字段如表1所示。

表1 建模主要字段列表

序号	字段名称	注释	序号	字段名称	注释
1	DOU	每月数据流量	10	全部通话次数	按月统计
2	ARPU	每月通信费用	11	全部通话天数	按月统计
3	VIP级别	普通、一星、二星、三星、四星、五星	12	中联通通话次数占比	按月统计
4	付费模式	后付费、预付费	13	主叫通话次数占比	按月统计
5	入网年份	入网日期所在的年份	14	非漫游通话次数占比	按月统计
6	客户年龄	按身份证计算年龄	15	非漫游通话天数占比	按月统计
7	MOU	每月通话分钟数	16	非长途通话次数占比	按月统计
8	融合类型	智慧沃家、主副卡、沃家庭、无	17	非长途通话天数占比	按月统计
9	终端类型	2G、3G、4G			

2.2 建模分析

在SPSS Modeler软件中构建如图7所示数据流。

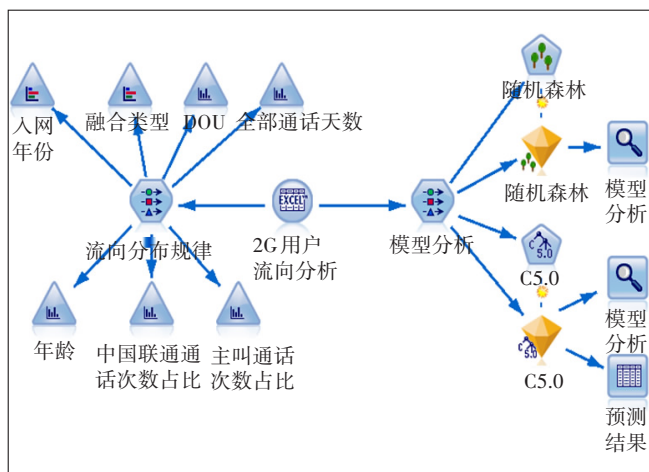


图7 SPSS Modeler数据流

在数据流中,通过“2G用户流向分析”导入整理好的源数据,经过多个模型的对比分析,最后选择2种模型进行用户流向的建模。

2.2.1 随机森林

随机森林指的是利用多棵树对样本进行训练并预测的一种分类器,主要应用于回归和分类,能够很好地处理高维度的数据。采用随机森林模型的预测结果,准确率达到82.9%。其预测的各个变量的重要性如图8所示。

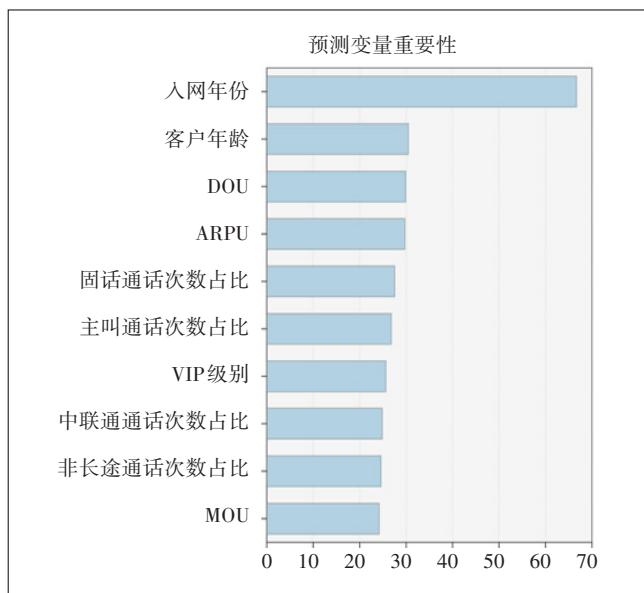


图8 随机森林预测变量重要性

可以看到,“入网年份”对预测结果的重要性最高,其次为“客户年龄”及“DOU”。

2.2.2 C5.0

C5.0是经典的决策树模型算法之一,可生成多分

支的决策树,目标变量为分类变量,使用C5.0算法可以生成决策树或者规则集。采用C5.0模型的预测结果,准确率达到93.2%。其预测的各个变量的重要性如图9所示。

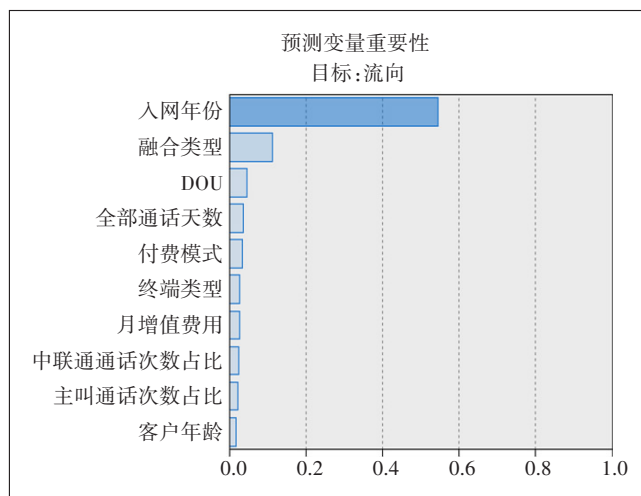


图9 C5.0预测变量重要性

可以看到,仍然是“入网年份”对预测结果的重要性最高,其次为“融合类型”及“DOU”。

综合这2种模型的预测结果,可知“入网年份”对2G用户流向的影响最大,另外这2个模型中都出现的重要变量还有“DOU”“主叫通话次数占比”“中国联通通话次数占比”“客户年龄”。

2.3 模型解释及验证

下面对2G用户流向在不同指标中的分布进行统计验证(见图10)。

从图10中可以了解到:

a) 入网年份。2008年以前入网的2G用户基本上都离网了,2008年以后入网的2G用户转为4G的比例更高。可能的原因是,2008年以前入网的用户到了其客户生命周期的末期,对号码的依存度较低;2008年后入网的用户,由于中国联通推出了3G/4G业务,增加了网络对用户的吸引力,因此转网用户更多。

b) 融合类型。图10中“1.0”为智慧沃家,“2.0”为主副卡,“3.0”为沃家庭,“0.0”为非融合套餐。融合套餐中转网3G/4G的用户比例高,非融合套餐离网比例高。说明融合套餐增加了用户黏性,对保有用户起到了较大的作用。

c) DOU、年龄。DOU低、年龄大的用户离网比例高,而DOU高且年轻的用户更倾向于转网3G/4G。说明中国联通3G/4G网络对年轻的移动互联网用户吸引

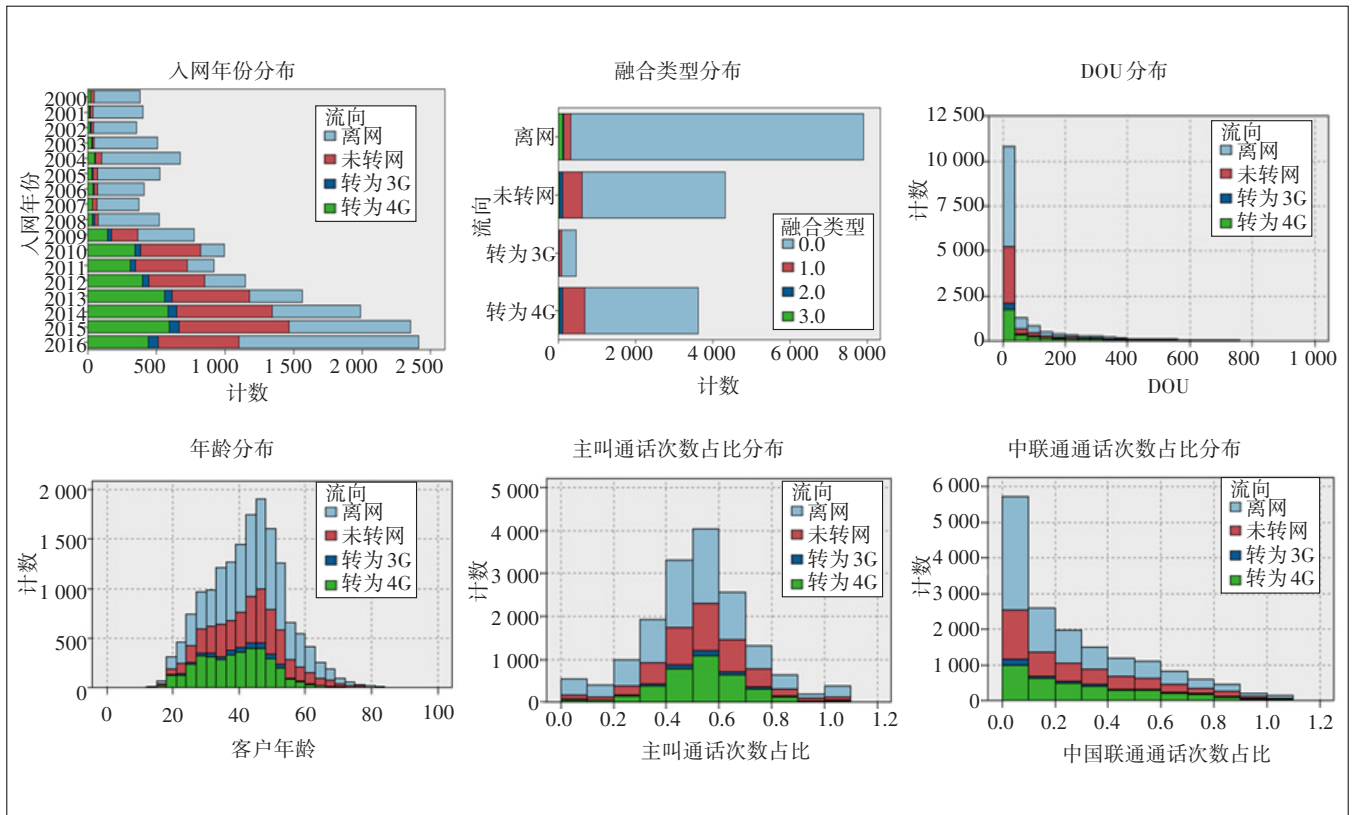


图10 2G用户流向分布

力更大。

d) 主叫通话次数占比、中国联通通话次数占比。这两个指标对2G用户转网的影响虽然不如前几个指标显著,但同转网3G/4G用户比例也存在着正相关的关系。说明对中国联通网络黏性大的用户转网3G/4G的概率更大。

最终选择C5.0模型对A省、B省的2G用户进行流向的预测。

3 城市画像预测

根据模型预测的结果,A省总的2G用户离网的比例为41%,转网3G/4G用户的比例为37%;B省总的2G用户离网的比例为47%,转网3G/4G用户的比例为26%。对2省的历史数据进行统计,发现2G用户转网后其平均ARPU增长为15元,平均DOU增长6倍,平均MOU基本持平,转网3G/4G的用户中10%转为2I2C用户。由此计算出2省各个地(市)的城市画像指标。

图11和图12分别给出了2G用户转网后2省用户数、收入指标、平均ARPU、平均DOU的分布情况,与图1和图2(2G用户转网前)相比可以看出,推动2G用户

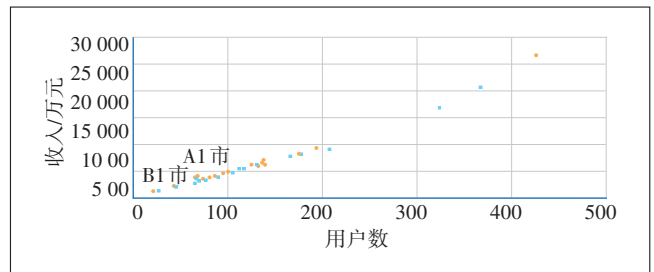


图11 2省用户数、收入指标分布

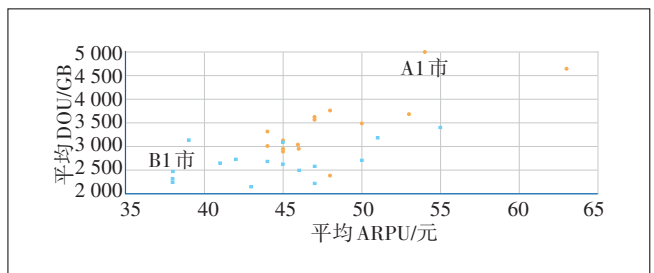


图12 2省平均ARPU、平均DOU指标分布

转网后,2省各地(市)的用户数、收入普遍略有下降,但用户模型和用户结构得到了改善。

从最终的城市画像图(见图13)来看:

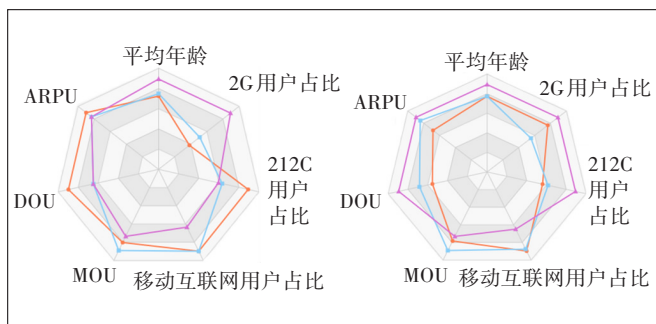


图 13 A省及A1市、B省及B1市画像

a) A1市变化较大的指标为平均年龄、移动互联网用户占比和2G用户占比,2G用户占比的变化较B1市更为显著,其原因是转网3G/4G的2G用户比例高,剩余2G用户更少。

b) B1市变化较大的指标为平均年龄、移动互联网用户占比和2G用户占比,平均年龄、移动互联网用户占比的变化较A1市更为显著,其原因是B1市3G/4G用户本身基数较小,在转网后年轻的3G/4G用户快速增加。

A1市2G用户占比从转网前的23.7%下降到10.0%,2G用户的减少在市区的高校、商场、写字楼等区域会表现得更显著。需关注这些区域的2G基站负荷,及时对负荷低的基站进行减频或者退网。

B1市用户平均DOU从转网前的1.7GB增加到2.3GB,增长了35.3%,而商场、景点等热点区域的流量增幅更大。因此从网络建设的角度,需了解转网后流量的地理分布情况,从而提前对网络进行扩容。

4 总结

本文通过对2G用户转网的建模分析,对比了转网前后2个省及2个典型地(市)的城市画像变化趋势。为了解用户和业务现状、了解业务发展趋势并在网络建设上及时应对提供了详细的数据支撑。本文提出的方法,还可以同其他热点专题相结合,如212C用户、视频业务等专题,运用大数据和模型分析,滚动刷新全国不同城市的画像指标,实现网络的精准运营。

参考文献:

[1] 李梅,杜翠凤. 基于大数据分析的移动通信网络规划方法[J]. 移动通信,2015(10):24-29.
[2] 黄亚洲,雷景智. 大数据在推动2G用户转网中的应用[J]. 邮电设计技术,2016(5):42-47.
[3] 吴昱. 大数据精准挖掘[M]. 北京:化学工业出版社,2014:131-

137.
[4] 徐振华. 浅议4G网络的建设和运营[J]. 信息通信,2014(10):244-245.
[5] VERMA D C. 大数据爆炸时代的移动通信技术与应用[M]. 北京:机械工业出版社,2016:17-21.
[6] 张文彤,钟云飞. IBM SPSS数据分析与挖掘实战案例精粹[M]. 北京:清华大学出版社,2014:390-392.
[7] 黄文彬,徐山川. 移动用户画像构建研究[J]. 现代情报,2016(10):54-61.
[8] 肖三保. 聚焦用户行为精准建设4G网络[J]. 江西通信科技,2015(2):4-6.
[9] 肖云,陈森. 移动通信业务预测方法研究[J]. 邮电设计技术,2011(2):53-57.
[10] 翟倩. 移动网络数据分析方法及应用综述[J]. 电信技术,2017(3):60-63.
[11] 鲁喆,陈水平. 数据挖掘技术及其在移动通信中的应用[J]. 科技广场,2010(11):27-30.
[12] 孟昉,郑雅婷. 基于大数据分析的4G电信运营业务精准运营平台的研究[J]. 纳税,2018(1):124-128.
[13] 李涛,曾春秋. 大数据时代的数据挖掘——从应用的角度看大数据挖掘[J]. 大数据,2015(4):57-80.
[14] 唐亘. 精通数据科学,从线性回归到深度学习[M]. 北京:人民邮电出版社,2018:238-240.
[15] 薛薇. 基于SPSS Modeler的数据挖掘[M]. 北京:中国人民大学出版社,2014:209-213.
[16] 秦洁,徐喜安,陈毓锋,等. 基于移动通信大数据的城市职住空间分析[J]. 移动通信,2017,41(17):73-77.
[17] 王峰,唐美华. 基于移动通信大数据的城市人口管理解决方案[J]. 移动通信,2014(13):38-41.
[18] 陈艳秋,吴礼华,江昊,等. 基于移动通信大数据的城市空间结构感知[J]. 科学技术与工程,2018,18(20).
[19] 赵卫丹. 基于大数据平台的城市地铁交通模式研究[D]. 长春:吉林大学,2016.
[20] 涂小华,张正. 基于移动通信大数据背景下的南昌城市空间结构研究[J]. 江西化工,2016(6):103-106.
[21] 秦萧,甄峰,熊丽芳,等. 大数据时代城市时空行为研究方法[J]. 地理科学进展,2013,32(9):1352-1361.
[22] 吉鸿雁,马新文. 大数据时代大型城市智慧化应用平台的实践与思考[J]. 中兴通讯技术,2014,20(4):21-24.
[23] 佚名. 数字空间视角下的城市数据画像理论思考[J]. 情报学报,2019,38(1):58-67.

作者简介:

傅俊锋,毕业于华中科技大学,高级工程师,硕士,主要从事移动网络的规划、研究及大数据分析工作;孙红,高级工程师,主要从事移动网络的规划、研究及大数据分析工作。

