

基于运营商大数据的

游戏用户画像构建研究

Research on Game User
Profile Construction Based
on Operators' Big Data

欧阳秀平, 廖娟, 冯焯, 刘卉芳 (中国联通广东省分公司, 广东 广州 510627)

Ouyang Xiuping, Liao Juan, Feng Ye, Liu Huifang (China Unicom Guangdong Branch, Guangzhou 510627, China)

摘要:

利用大数据分析技术对运营商海量数据进行挖掘, 提出构建运营商特色用户画像全流程方法。通过爬虫、NLP分词等技术实现APP智能归类, 以替代过去人工标注的方法; 通过挖掘APP上网日志数据, 可以构建用户的兴趣偏好画像, 改善E域数据未被充分利用的行业现状。以游戏行业为例, 构建主流游戏类别玩家多维度画像, 为精准营销提供数据参考。

关键词:

大数据; 运营商; 用户画像; 精准营销; 游戏行业
doi: 10.12045/j.issn.1007-3043.2019.09.009
中图分类号: TP274
文献标识码: A
文章编号: 1007-3043(2019)09-0040-05

Abstract:

Using big data analysis technology to mine operators' massive data, the whole process method of constructing operator's characteristic user profile is proposed. APP intelligent classification is realized by Web crawler and NLP participle technology to replace the manual labeling method in the past. The user interest preference profile is constructed by mining the APP log data, which can improve the current situation that the E-domain data is not fully utilized. The game industry is taken as an example, the multi-dimensional profiles of mainstream game players are built, which provides data reference for precision marketing.

Keywords:

Big data; Operator; User profile; Precision marketing; Game industry

引用格式: 欧阳秀平, 廖娟, 冯焯, 等. 基于运营商大数据的游戏用户画像构建研究[J]. 邮电设计技术, 2019(9): 40-44.

0 引言

随着5G时代和携号转网的到来, 传统电信业务带来的收入驱动力逐渐下降, 流量红利窗口也即将关闭。如何挖掘自身优势以提升竞争力、减少用户流失、发掘用户的非通信价值成为三大运营商普遍关注的问题。在此背景下, 深度挖掘用户信息成为大数据时代的研究热点。用户画像的定义是使用标签来表征用户的属性和特性, 以达到描述用户的目的, 即用户信息标签化, 可以分为用户属性研究和用户行为研

究两大内容。用户属性包含年龄、性别、归属地、职业、学历等基本信息, 用户行为类标签则是通过用户的行为数据挖掘出用户的行为习惯、兴趣偏好等隐藏信息。各大互联网公司都基于自身数据对用户进行建模分析, 构建用户画像系统^[1-7]。但是, 用户画像的构建很大程度上依赖于业务目的和数据源, 各个互联网公司都只拥有用户在自己平台的行为数据, 在数据不共享的情况下, 这些公司的用户画像都存在一定局限性。而运营商以号码为主键可整合位置、上网、通信、终端等各类数据, 用户数据的完整性是其他企业难以企及的, 因此, 基于运营商数据的用户画像可以实现更精准的个性化推荐服务, 这是运营商最大的优

收稿日期: 2019-07-16

势和竞争力,同时也能成为运营商流量变现的方式。

通过对广东联通部分用户的信息进行数据挖掘,着重分析其APP使用行为,提出构建APP画像和用户画像的方法,为运营商深入了解用户行为、实现异业合作、精准营销、增加后向收入提供参考。

1 用户画像构建框架

运营商可利用用户位置、上网、通信、终端等各类数据构建用户画像,对用户进行个性化推荐,这样不但可以为用户提供更高质量、更精准的服务,同时也为运营商提供流量变现的可靠方式,在运营商实现跨界的业务融合方面具有深远的意义和价值。目前,对运营商B域数据的研究相对成熟,许多学者基于用户账单数据构建用户个体画像和群体画像^[8-11]。但是,对于运营商E域数据如APP上网日志数据的研究由于解析困难等原因起步较晚,目前仍属于研究热点^[12-15]。

整体来说,目前关于运营用户画像研究存在以下问题:E域的APP上网日志数据没有被充分挖掘;APP的分类主要靠人工标注。针对上述现状,提出构建运营商特色用户画像全流程方法,其创新点主要有以下几方面。

a) 通过爬虫、NLP分词等技术实现APP智能归类打标,构建APP画像。通过APP画像可以快速筛选出满足任意给定特征属性词的所有APP,这不仅解决过去APP归类需要大量人工标注的问题,而且更快速、全面。

b) 通过挖掘APP上网日志数据,引入用户兴趣时间衰减函数,结合APP画像,构建用户的兴趣偏好画像。

用户画像具体构建流程框架如图1所示。首先,对E域通道数据进行DPI解析得到用户APP行为数据,通过网络爬虫、分词、正则匹配等技术挖掘互联网数据得到APP属性画像,然后,将用户的APP使用次数和流量等信息与APP画像关联,从而挖掘用户的兴趣偏好,再结合B域数据如年龄、性别、地域等基本信息得到个体用户画像,最后对用户的行为属性标签进行聚类整合,构建出某一类用户群体的整体特征,为行业洞察和差异化精准营销提供数据支撑。

2 用户画像构建实例——游戏行业

以游戏行业为例阐述构建运营商特色用户画像全流程方法。分析数据来自中国联通广东省分公司,

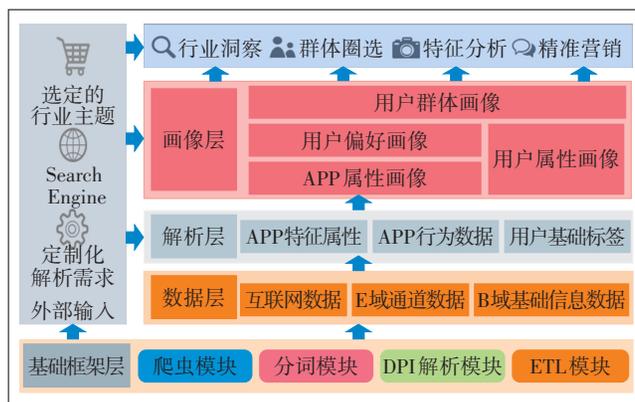


图1 用户画像构建框架

随机抽取1 000万个用户的基础信息和其在2018年7月至2019年1月期间的上网日志进行DPI解析,字段主要包括用户ID、性别、年龄、地(市)、终端、APP名称、APP使用起止时间、访问次数、使用流量等,该数据完全符合大数据的4V特征:数据体量巨大(Volume)、内容多样化(Variety)、价值密度低(Value)、增长速度快(Velocity)。

2.1 APP画像构建过程

APP画像可以反映该APP的类别、功能等信息,其构建方法如下:首先,通过搜索APP名称,对APP官网和百度百科进行网页爬虫得到关于该APP的信息,再通过NLP分词、知识图谱、正则匹配等方式沉淀一组该APP的属性特征词,最后通过统计词频、删除无意义词等方式得到APP画像。

通过上述方法构建的属性特征词和APP是多对多的映射关系,共囊括2 306个有DPI用户数据的APP和27 446个属性词。以游戏行业为例,共有330个APP被打上“游戏”标签,即被定义为“游戏APP”,将这部分游戏APP和其他APP的属性词做比较,可以得到392个游戏行业特有词,通过统计每个高频词关联APP的用户规模,选取用户量排名前9的属性词作为游戏特征类进行重点分析,如表1所示。若需要分析其他特征属性,方法流程类似。

表1 游戏行业主流特征类别

游戏类	MOBA	FPS	RPG	消除	模拟	仙侠	策略	棋牌	桌游
APP个数	28	19	29	10	11	13	72	46	7

MOBA:多人在线战术竞技游戏,FPS:第一人称射击游戏,RPG:角色扮演游戏

需要注意的是,一个APP有可能同时归属于多个类,例如“绝地求生:刺激战场”既属于MOBA类,又属于FPS类。

2.2 用户画像构建过程

由于各APP的访问次数、流量、时长等信息因性质不同不具有可比性,需要对数据进行标准化处理。采用“min-max标准化”方法,将变量的原始值映射成[0,1]之间的值,其公式为:

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

式中:

X^* ——标准化后的值

X ——原始值

X_{\min} ——该指标的最小值

X_{\max} ——该指标的最大值

通过构建用户偏好得分模型将用户游戏偏好划分为5档(0~4),分别代表非玩家、试探玩家、一般玩家、核心玩家、塔尖玩家。偏好画像的得分应满足以下3个条件。

a) 用户在此特征属性类中的APP上操作越多,得分越高。

b) 用户对各个特征类的喜好程度不同可以通过偏好得分区间体现。

c) 越久远的用户数据对得分的影响越小。

由于用户的兴趣爱好具有很强的时效性,根据艾宾浩斯遗忘曲线,用户产生行为时的时间戳 t 与用户兴趣的相关性函数定义如式(2)所示。

$$p(t) = 1 - a(T-t)^b \quad T \geq t \quad (2)$$

式中:

a, b ——兴趣衰减参数

T ——当前时间,显然,当 $t=T$ 时, $p(t)=1$ 达到最大值

若假设 $a=1/8, b=1/2$,则用户兴趣的时间衰减系数16天衰减一半,64天衰减完毕。具体参数的设置可以根据业务需求而定。

构建用户偏好得分公式如下

$$\text{Score} = \sum_{APP_i \in \text{特征属性类}} \sum_{t \in (0, T]} p(t) \times [w_1 \times PV(t, APP_i) + w_2 \times \text{time}(t, APP_i) + w_3 \times \text{GPRS}(t, APP_i)] \quad (3)$$

式中:

$PV(t, APP_i)$ ——第 t 天 APP_i 的访问次数,已通过式(1)进行标准化处理

$\text{time}(t, APP_i)$ ——第 t 天 APP_i 的使用时长,已通过式(1)进行标准化处理

$\text{GPRS}(t, APP_i)$ ——表示第 t 天 APP_i 的使用流量,已通过式(1)进行标准化处理

w_1, w_2, w_3 ——权重系数

$p(t)$ ——时间衰减系数,具体计算如式(2)所示

通过式(3)对每个用户在各类游戏中的表现进行打分,并划分为5个等级:top25%的为“塔尖玩家”,top25%~50%的为“核心玩家”,top50%~75%的为“一般玩家”,top75%以下的为“试探玩家”以及不玩这类游戏的“非玩家”。定义“塔尖玩家”和“核心玩家”为“爱好者”。

2.3 群体画像分析

群体用户画像是基于单个用户画像实例,将某一特征属性相似的用户群作为整体进行分析,从而得到该用户群的属性、行为特征。以表1中的9个特征属性为例,从基础信息和行为信息两大方面构建这9个特征爱好者用户群的群体画像,若需要分析其他用户群,方法流程与此类似。

2.3.1 基础信息画像

对9个特征群体的性别分别进行统计,结果如图2所示。所有有上网行为的用户中,男性占比为67%,与这个基准值相比,消除类游戏爱好者中女性玩家更多,其他类则是男性玩家占比更多。因此消除类APP应更注重女性体验,其他类游戏APP应更看重男性视角,从而增加用户黏性、吸引新用户。

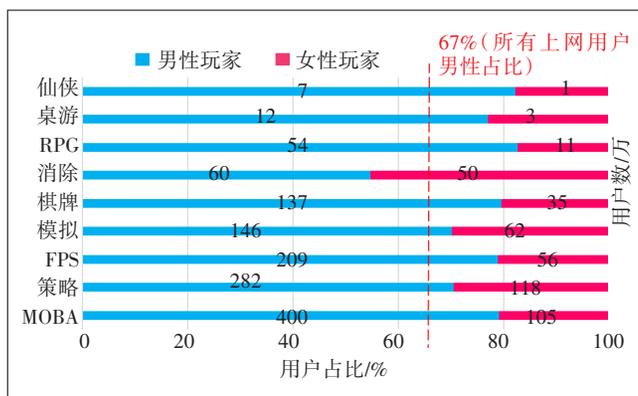


图2 不同类型手游玩家群体性别分布

年龄分布如图3所示。通过机器学习聚类算法,可以将9个特征群体分为2类,一类为策略、模拟、棋牌、仙侠、消除类游戏爱好者,年龄主要分布在18~33岁,同时在38~53岁的用户群中也有一定市场,说明这几类APP更偏向“老少皆宜”;另一类如FPS、MOBA、RPG、桌游类游戏爱好者,年龄主要集中在18~26岁,

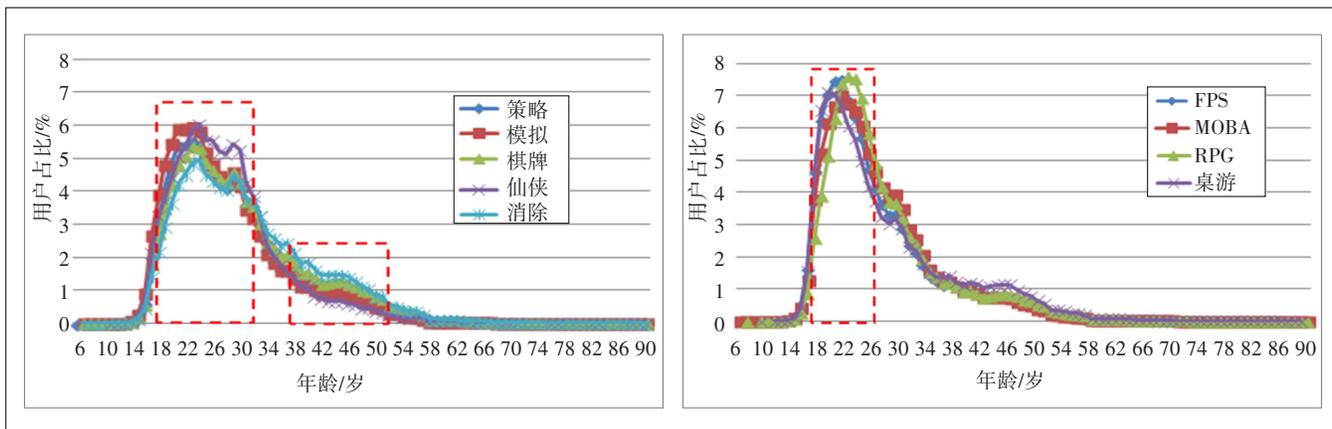


图3 不同类型手游玩家群体年龄分布

相比第1类更为“年轻化”,这几类APP可以着重向在校和职场新人推荐。

通过统计终端信息,发现99.9%的游戏爱好者使用4G终端,终端品牌分布如图4所示。以全部上网用户的终端分布为基准,可以看出各个用户群的终端偏好。其中,策略、棋牌类爱好者更偏好苹果手机,FPS、MOBA类爱好者更偏爱华为、OPPO、VIVO、小米四大品牌,同时可以发现OPPO用户大多喜欢玩游戏。运营商在推广对应权益产品时,可以参考这些信息进行终端合约营销。

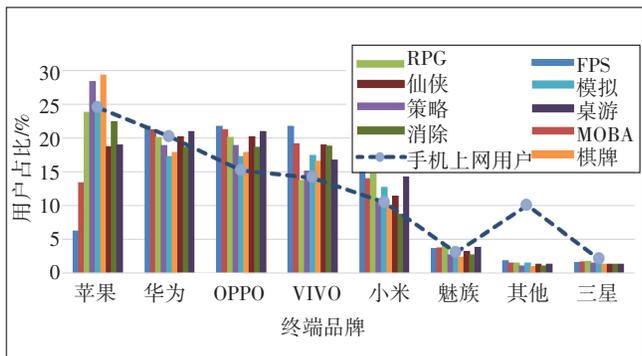


图4 不同类型手游玩家群体终端品牌分布

2.3.2 行为信息画像

选取2018年8月至2019年1月的数据,统计各类游戏玩家平均每月花在该类游戏上的时长,结果如图5所示。MOBA类玩家每月花在MOBA类游戏上的时间最长,达22h,其次为消除类和仙侠类。数据分析结果说明这几类游戏的玩家相对黏性较高,愿意花时间和精力投入在这些游戏中。

提取样本用户的上网日志的DPI解析数据,计算每个小时的活跃用户数占比,结果如图6所示。通过

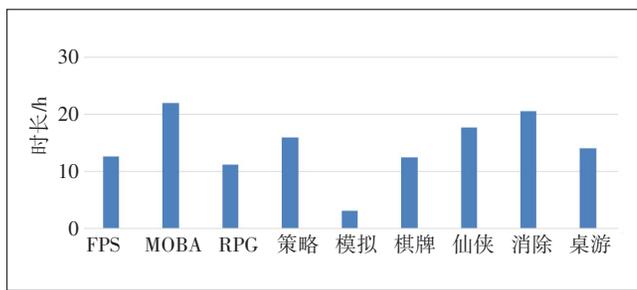


图5 不同类型手游玩家群体月均游戏时长分布

机器学习聚类算法,将9个特征群体分为2类,一类为FPS、MOBA、策略、模拟类爱好者,晚高峰的活跃用户数明显高于午高峰,说明这类用户群更偏爱在19点之后玩游戏,其次为11点—18点,在12点呈现小高峰。说明这类用户群喜欢在一天的工作或学习结束之后再玩游戏。另一类为RPG、棋牌、仙侠、消除、桌游类爱好者,在11点—23点之间的活跃用户数相对平稳,结合这类游戏性质,可以解释为这类用户群偏爱利用碎片化时间,即有空就玩一两局,或者直接挂机。

整体来说,MOBA类游戏爱好者偏向男性化、年轻化,更喜欢用安卓机(华为、OPPO、VIVO),偏爱晚上玩游戏,月均游戏时长22h。消除类游戏爱好者偏向女性化、中青年居多,更喜欢用OPPO、VIVO,晚高峰与午高峰持平,月均游戏时长21h。因此MOBA类APP可以重点向使用安卓机的年轻男性推荐,并且在晚上开放限时优惠等活动。而消除类APP可以重点向使用OPPO、VIVO的中青年女性推荐。其他游戏类别的分析与此类似。

3 总结

与互联网公司数据相比,E域上网日志数据是运

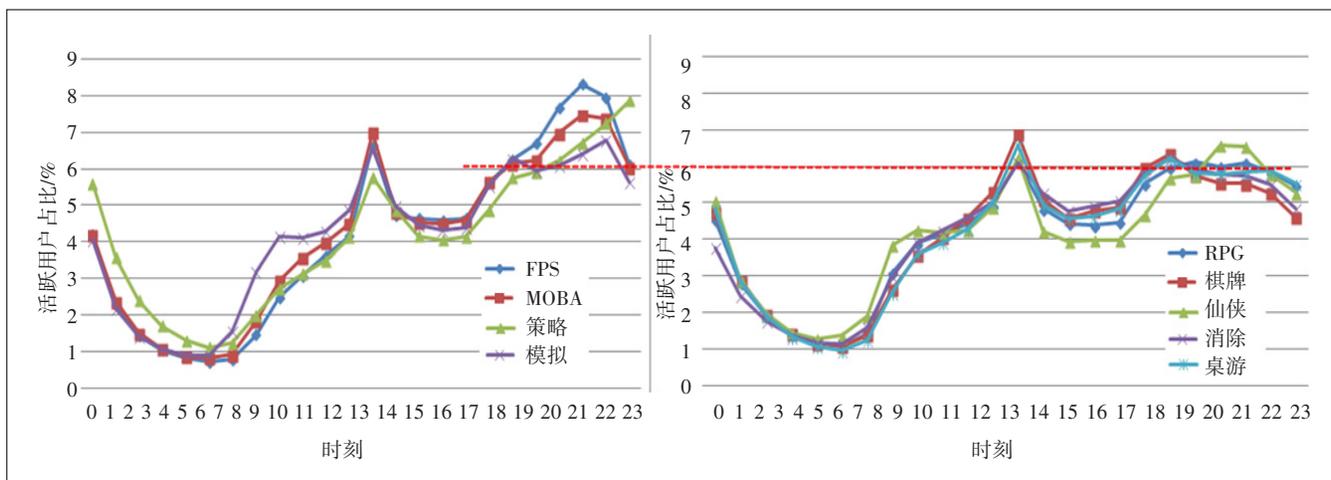


图6 不同类型手游玩家群体活跃时段分布

营商特有的数据源,基于此,本文提出一种构建运营商特色用户画像的全流程方法,并以游戏行业为例进行应用示范。一方面,通过APP智能归类打标解决过去APP归类需要大量人工标注的问题,另一方面,通过挖掘APP上网日志数据,结合APP画像,可以构建用户的兴趣偏好画像,将用户的兴趣偏好分为5个等级,改善E域数据未被充分利用的行业现状。基于个体用户画像和群体用户画像,输出游戏行业洞察分析报告,为精准营销提供数据参考。

参考文献:

[1] 杨捷. 基于用户画像和因子分解机的推荐算法研究[D]. 太原: 中北大学, 2018.

[2] 雷兵, 常知刚, 钟镇. 基于网络店铺订单数据的群体用户画像构建研究[J]. 河南工业大学学报(社会科学版), 2019, 15(1): 52-59.

[3] 单晓红, 张晓月, 刘晓燕. 基于在线评论的用户画像研究: 以携程酒店为例[J]. 情报理论与实践, 2018, 41(4): 99-104, 149.

[4] 郑建兴, 张博锋, 岳晓冬, 等. 基于友邻-用户模型的微博主题推荐研究[J]. 山东大学学报: 理学版, 2013, 48(11): 59-65.

[5] 林燕霞, 谢湘生. 基于社会认同理论的微博群体用户画像[J]. 情报理论与实践, 2018, 41(3): 142-148.

[6] 洪婷婷. 互联网广告人群标签库的设计及投放规则挖掘[D]. 北京: 中国科学院大学, 2015.

[7] 陈志明, 胡震云. UGC网站用户画像研究[J]. 计算机系统应用, 2017, 26(1): 24-30.

[8] 郑正广, 冯必成, 赵明月. 基于大数据技术的腾讯王卡用户画像构建研究[J]. 邮电设计技术, 2017(9): 52-56.

[9] 张海旭. 基于话单数据的移动通信用户画像研究[D]. 北京: 中国科学技术大学, 2018.

[10] 王晓霞, 刘静沙, 许丹丹. 运营商大数据用户画像实践[J]. 电信科学, 2018(5): 127-132.

[11] 王艺霖. 基于大数据技术的电信用户画像系统的实现与应用

[D]. 北京: 北京邮电大学, 2018.

[12] 程豪, 吕晓玲, 钟琰, 等. 大数据背景下智能手机APP组合推荐研究[J]. 统计与信息论坛, 2016, 31(6): 86-91.

[13] 程豪, 吕晓玲, 钟琰, 等. 基于智能手机大数据的APP使用时间分布规律研究[J]. 数学的实践和认识, 2018, 48(19): 158-164.

[14] 苗雨来. S省运营商公司大数据精准营销研究[D]. 北京: 北京邮电大学, 2018.

[15] 金石. 基于运营商管道大数据的智能电商推荐系统[D]. 南京: 南京邮电大学, 2018.

[16] 单晓红, 张晓月, 刘晓燕. 基于在线评论的用户画像研究——以携程酒店为例[J]. 情报理论与实践, 2018, 41(4).

[17] 刘海, 卢慧, 阮金花, 等. 基于“用户画像”挖掘的精准营销细分模型研究[J]. 丝绸, 2015, 52(12): 37-42.

[18] 刘海鸥, 孙晶晶, 苏妍娜, 等. 国内外用户画像研究综述[J]. 情报理论与实践, 2018, 41(11): 159-164.

[19] 隋国政. 浅谈用户画像构建及应用[J]. 电子世界, 2017(15): 144-144.

[20] 刘晓丽. 大数据背景下智能手机APP组合推荐研究[J]. 数码世界, 2017(10): 187-187.

[21] 黄勇军, 冯明, 丁圣勇, 等. 电信运营商大数据发展策略探讨[J]. 电信科学, 2013, 29(3).

[22] 陈涛, 鲁萌, 陈彦名. 运营商大数据技术应用研究[J]. 电信科学, 2017, 33(1): 135-139.

[23] 冉蹬. 基于用户画像的手机游戏用户个性化内容推荐研究[D]. 西安: 西安理工大学, 2018.

[24] 尤骏杰. 大数据营销理论及其在游戏运营中的应用[D]. 苏州: 苏州大学, 2016.

作者简介:

欧阳秀平, 中国联通广东分公司信息化部总经理, 中国联通广州软件研究院院长, 高级工程师, 主要研究方向为大数据、云计算、计算机软硬件; 廖娟, 硕士, 主要从事大数据分析、建模、算法研究等方面工作; 冯焯, 工程师, 主要从事大数据分析、建模、算法研究等方面工作; 刘卉芳, 工程师, 主要从事大数据分析、建模、算法研究等方面工作。