



运营商大数据用户画像实践

王晓霞, 刘静沙, 许丹丹

(中国联合网络通信有限公司研究院, 北京 100176)

摘要: 移动市场从增量发展模式进入存量发展模式, 在存量模式的情况下, 减少用户流失、延长用户在网的生命周期、提升用户价值、发掘用户的非通信价值已成为共识。以存量用户数量 and 价值的保有为目标, 利用大数据技术精准客户聚类画像, 从而达到获取客户需求、匹配维系产品资源、在重点目标客户群体上形成突破的营销目的。

关键词: 通信行业; 大数据; 用户画像; 存量经营; 精准营销; 标签

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018109

A practice of user profile in big data

WANG Xiaoxia, LIU Jingsha, XU Dandan

China Unicom Research Institute, Beijing 100176, China

Abstract: The development of telecom market focused on the development model from the incremental development model. In the case of incremental development model, it has become a consensus to prevent customer reduction, to extend customers' life cycle, to enhance revenues and earnings. User profile technology was used to defend telecom business and customer base in the face of pressure from OTT. Big data technology of user profile was used to describe customer demand, match customer and product, so as to make a breakthrough in marketing.

Key words: telecom industry, big data, user profile, stock management, precision marketing, label

1 引言

大数据时代下, 在企业的生产实践中, 利用大数据技术进行“用户画像”是各行各业的选择。用户画像、语音识别、图像识别和自然语言理解并列为当前人工智能领域最重要的四大应用场景^[1]。用户画像的本质是从业务角度出发对用户进行分析, 了解用户需求, 寻找目标客户。同一用户在不同时间、不同场景下的行为有差异, 而不同场

景下的不同用户画像更是千变万化。本文以流失预警这一业务场景为例, 介绍如何利用大数据用户画像来实现客户流失提前预警。

2 用户画像

用户画像(给用户打标签)是用户数据的形象化展示, 可以理解为用一连串的用户标签简短而形象地描述一个用户的特征^[2]。用户画像是勾画目标用户、联系用户诉求与设计方向的



有效工具，它是具体商业目标下标签的重新排列组合^[3]。

用户画像有两类，一是 user personal（用户角色），二是 user profile（用户属性）。用户角色是产品经理、策略设计师和运营人员从用户群体中抽象出来的典型用户，可以借助虚拟的角色，跳出离散的需求，聚焦用户需求。它的本质是用来描述用户需求的工具，适合产品早期目标用户人群的定义和研究，帮助开发者在产品设计、策略匹配阶段，站在用户的角度思考问题。用户属性是根据每个人在产品中的用户行为数据，生成描述用户的标签的集合。利用已有数据，通过以定量为主的研究方法描绘出用户的人口学、产品行为和人群数据属性，适合产品中后期实体用户的人群划分和聚类分析^[4]。用户角色适用于产品和策略设计，针对用户设计产品；用户属性适用于精准营销，针对产品找用户。

3 用户画像分级架构

用户画像的第一步就是进行画像架构的分级分类。画像的维度不同，画像结果千变万化，所以有效的画像必须是以商业结果为目标标签集合。例如存量经营的商业目的就是存量产品的精准营销，因此画像架构的分级依据为存量产品体

系。从存量产品体系的策略、需求、产品和渠道4个维度出发，建立了四级用户画像架构。

3.1 一级架构

一级架构为基本画像即用户的基本维度，是用户的生命周期画像，如图1所示。一级架构从产品策略维度出发，根据用户对业务的使用程度进行划分，将用户划分为4个群体，分别是体验用户、价值提升用户、稳定保有用户、离网挽留用户。体验用户为新入网用户，由于入网时间短，这类用户对产品的感知还处于探索阶段，消费需求尚不明确，其中一部分体验用户随着对产品的良好体验和品牌好感度的增加将转为价值提升用户；价值提升用户介于体验用户和稳定用户之间，入网时间大于半年，对产品的消费需求非常明确，消费特征显著，ARPU（每用户平均收入）连续3个月呈现提升的趋势，有些用户同时伴随着流量压抑、流量超套或者流量高饱和等行为，部分价值提升用户随着对产品持续性的良好感知将会转变为稳定用户；稳定用户的在网时长一般大于1年，在4种用户中品牌忠诚度最高，套餐变动频率低、对资费较为不敏感，由于用户感知良好，客服投诉量也较少；维挽用户为即将离网的用户，其消费特征为ARPU值连续3个月呈现持续降低的趋势，此阶段用户有可能投诉较多，这

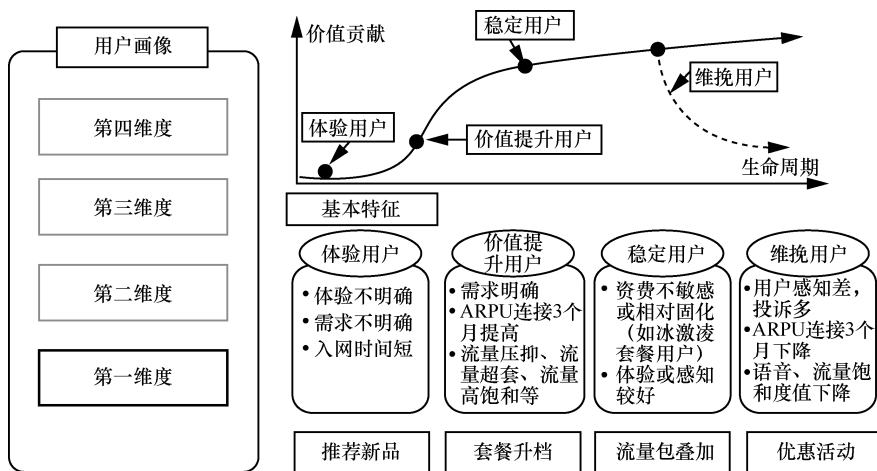


图1 生命周期画像

部分趋于离网的用户语音和流量的饱和度值呈下降状态。

3.2 二级架构

二级架构是行为画像。从用户的需求维度进行画像,行为画像根据不同的行为特征可细分为4类:通信行为、语音行为、流量行为和权益行为。通信行为反映用户的综合行为特征,如套餐是否经常超量,语音、短信和流量的使用度是否失衡等;语音行为用来描述用户语音套餐的使用程度,例如国内及国际语音是否有剩余或超额;流量行为用来描述用户流量套餐的使用程度,例如是否存在流量压抑、流量超出套餐以及流量高饱和的行为;权益行为用来描述用户对自身权益的敏感程度,例如是否资费敏感、ARPU值是否相对固定以及是否积极参加优惠活动等。

3.3 三级架构

三级架构是产品画像,如图2所示。产品画像针对具体的产品以及产品的具体内容进行目标

用户画像。将用户的标签属性与产品内容、产品特征相结合,进行相互匹配,筛选目标用户。例如根据现有产品冰激凌、畅越流量王、畅越视听宝等套餐的具体内容和特征进行目标用户画像寻找目标用户,实现精确的产品推荐。

3.4 四级架构

四级架构是触点画像,如图3所示。进行产品匹配后的用户,需要通过用户最常用和最容易接受的触点,进行产品营销。通过分析用户行为得到用户最常用的业务办理渠道,有针对性地在用户触点渠道进行产品营销。渠道类型分为电子渠道和实体渠道两类。其中电子渠道包括自助终端、网上营业厅、掌上营业厅、互联网和短信营业厅等;实体渠道包括大客户客户经理、营业厅和呼叫中心等。

4 用户画像步骤

首先对标签体系进行系统梳理,根据具体的

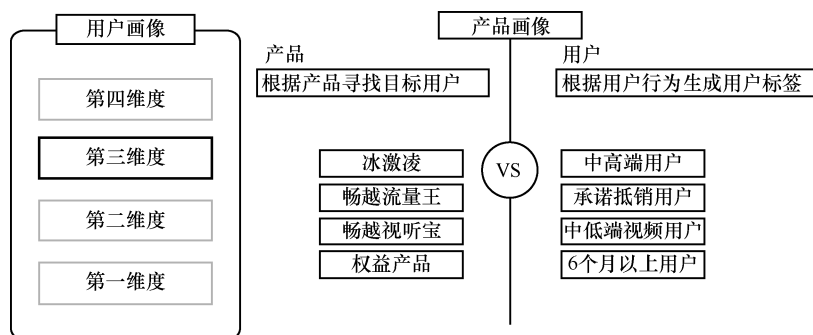


图2 产品画像

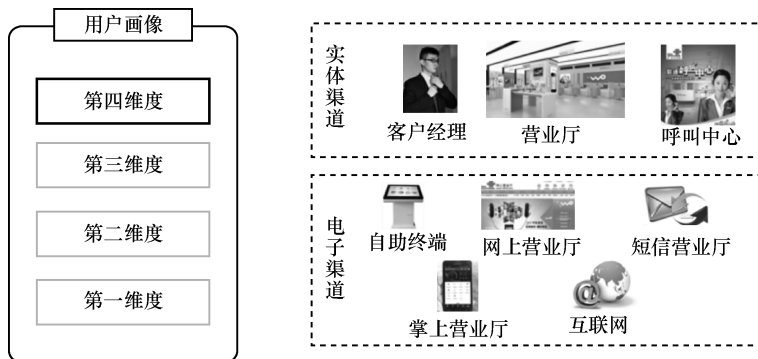


图3 触点画像



业务场景进行用户画像，再将用户画像形成的特征标签库应用于实际营销中精准的定位目标用户。画像流程分为 4 个步骤，分别是画像数据收集、标签梳理、用户画像和标签应用。

4.1 数据源

用户画像的第一步是列举相关的数据源。在对数据源进行列举分析之前应该收集所有可以提供用户数据的数据源，尽可能多地掌握用户的全量数据。用户信息分布在不同的系统中，一般情况下，客户的自然人信息主要集中在客户关系管理系统，消费特征主要集中在渠道和产品系统中。画像流程的第一步就是将这些分散在各个系统的用户信息收集起来集中处理。将五大数据源信息集中起来按照设定规则生成标签元数据体系。五大数据源包括互联网、通信网络、业务平台系统、IT 生产系统、合作伙伴。

其中来自通信网络、业务平台系统和 IT 生产系统的数据是企业内部数据，可信度高，采集方式为在系统中直接获取。IT 生产系统的数据源主要包括电子化销售服务管理系统（ESS）、业务支撑系统（BSS）、企业管理域（MSS）和运营支撑域（OSS）等。ESS 主要负责管理终端销售；BSS 主要负责客户关系管理、数据采集系统、计费账务、综合结算、营销支撑等功能，它包括总部 CRM、集中 PRM 和集中集团客户等系统；MSS 主要负责电信企业运营的流程管理，它包括 ERP、OA、财务和人力资源等系统；OSS 是一个综合的业务运营和管理平台，同时也是真正融合了传统 IP 数据业务与移动增值业务的综合管理平台。它主要由网络管理、系统管理、计费、营业、账务和客户服务等部分组成。

来自合作伙伴的数据源主要包括公共机构和数据/知识合作伙伴，属于外部数据，价值密度中等，包含了第三方的用户/企业的各方面信息，需要与内部高可信数据相互验证，获取方式可通过合作伙伴直接获取。

来自互联网的数据源主要包括互联网的访问行为、互联网网站内容，属于外部数据，价值密度低，包含互联网上的各种行为/言论，可信度低，需要与内部高可信数据验证。互联网信息库的数据源主要来自移动用户上网日志和互联网数据，其中移动用户上网日志数据是采用 FTP（文件传输协议）文件方式进行数据采集实现；互联网数据是建立多台服务器集群，通过实现分布式网络爬虫采集、插件采集、模拟采集等技术方式进行高效的互联网数据采集。

4.2 标签梳理

用户画像的第二步是根据用户数据的类型梳理用户标签。标签是指对客户行为数据与基本信息进行分析、归纳，提炼出的客户行为特征或属性特征（如性别标签），即描述男女两性区别的客户特征。基础标签库主要由电信业务基础标签库、互联网基础标签库两部分组成，分为五大类，即自然人标签、个人标签、通信标签、互联网标签以及应用标签。电信业务基础标签库以应用需求和用户的业务使用特征为基础，结合数据支撑能力，反映用户全生命周期的行为特征，包括基本属性、产品需求、业务特征、消费特征、渠道特征、终端偏好、客户服务评价及位置轨迹八大类内容。互联网基础标签库（互联网内容偏好标签库）的设计，参考主流互联网门户网站和应用商店的分类目录，借鉴电信行业经验，将用户上网行为分为新闻资讯、通信交流、娱乐休闲、生活服务、商务应用、自有业务、工具和使用偏好 8 类^[5]。

4.3 业务场景的理解

用户画像的第三步是对业务需求的理解，将业务语言转化为建模语言，为下一步的数据建模做准备。

- 根据具体的业务场景需求确定用户画像的目的，在此基础上制定用户画像标签，形成画像用户的特征标签库。
- 确定画像目的后，根据画像目的找到强相

关信息,剔除不重要的弱相关信息。强相关信息是指同业务场景需求直接相关的信息,可以是因果信息,也可以是相关程度很高的信息。强相关信息可以帮助定位目标客户,了解客户潜在需求,开发需求产品。只有强相关信息才能有效结合业务需求,创造商业价值。

- 集中了所有强相关数据标签之后,依据业务需求,对标签进行加工整理,需要对定性的标签进行定量以方便数据分类和筛选。这部分工作一般在数据仓库进行,定性信息进行定量分类是用户画像的一个重要工作环节,具有较高的业务场景要求,主要目的是将复杂数据简单化,形成用户特征标签库。

4.4 数据建模

通过对业务场景的理解可以确定数据建模的输入以及建模目的。对数据源进行分析,并对数据进行清洗后,就可以进行数据建模,数据建模主要包括数据核查、数据处理、变量相关性分析、变量筛选以及模型建立。

4.5 提供标签/数据可视化

在数据建模完成后,需要对数据模型进行可视化的展现。所以画像步骤的最后一步就是给用户打上标签,即以可视化的形式清晰简洁地将数据建模的结果传达出来。

5 用户画像基于用户流失预测的实践

5.1 画像概述

将业务场景与用户画像深度结合,筛选出具有价值的数据和客户,精准地定位目标客户,触达客户,对营销效果进行记录和反馈。以存量用户保有为例,具体步骤如下。

步骤 1 确定画像的目的,通过梳理各类流失模型应用场景得出流失定义,如不出账、极低三无、状态不正常行为等,然后根据流失定义确定用户画

像的目的,即对定义的预警用户进行数量保有。

步骤 2 根据画像目的来筛选预警用户强相关信息标签,如用户的生命周期、消费行为、通信行为、产品使用和行为偏好等标签。

步骤 3 根据步骤 1 筛选出来的强相关客户聚类信息,经过标签定性和定量形成用户特征标签库。最后通过流失预警用户特征标签库筛选出实际的目标用户,以提升客户使用感知为手段,通过匹配维系产品资源,为筛选出的目标客户推荐合适产品,实现客户维挽的最终目的。

5.2 流失预警的理解

将用户画像技术应用在流失预警上需要对业务场景有深入的理解。确定流失预警的定义为拆机离网、不出账、极低三无和状态不正常。其中,离网用户毋庸置疑是流失;不出账用户根据探索和市场部经验基本可以定义为流失;极低三无用户约占整个离网人数的 98%,且每个月离网速度在 14%左右,需将这部分用户定义为流失用户,作为模型目标变量,确保后期模型能够提前在极低三无前发现目标;状态不正常的用户后期有 67%转为不出账和极低三无,基本可以定义为流失用户。综上所述,得出流失定义为:如果用户在 T 月为在网出账、非极低三无、状态正常的用户,在 $(T+1)$ 月及以后发生离网、不出账、转为极低三无、转为状态不正常即视为流失用户。随着 3G 与 4G 网络与业务的不断普及,用户在行为上也出现了风格迥异的局面,对全网用户笼统地构建流失模型已经无法适应当前情况,需要针对性差异化构建流失模型。目前,市场维系产品以合约(终端合约、单卡类合约)产品、单产品(裸机、流量包和语音包)、固网融合业务产品为主。

5.3 模型建立

根据不同地区(城市人口结构、用户特点)以及不同业务体系(2G、3G、4G)用户行为的特点,分别搭建流失模型。如合约到期流失模型、可提前续约用户流失模型、逾期未续约用户流失



模型、合约成长期用户流失模型、单卡用户流失模型等。以提前续约用户流失模型为例，模型输入为：选取本网 3G 用户近 3 个月的用户信息（使用指标、消费行为指标、终端指标等），并根据经验创建出具有显著相关性的衍生变量，通过数据清洗、单变量分析、双变量分析和相关性分析等，最终确定建模所需变量。模型数据预处理过程如下：数据核查，根据业务标准及客观经验来判断提取的建模数据是否正确、可用；数据处理，对于值域过大的非错误数据进行数据去极值化处理；相关性分析，提取对目标变量强相关的变量，能够大大缩短模型的运行周期、提升模型的运行效率及精准度；变量筛选，根据业务经验和相关性分析选取重要的模型输入变量。然后再进行模型算法对比分析：逻辑回归对数据整体结构的分析优于决策树，而决策树对局部结构的分析优于逻辑回归。

离网预测准确率： S 为全体用户数， P 为预测流失用户数、 C 为真实流失用户数， P 与 C 的交集即正确预测用户数。离网预测准确率如图 4 所示^[6]。

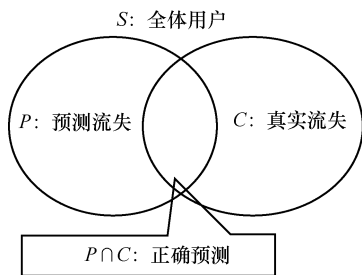


图 4 离网预测准确率

预测准确率计算式为：预测流失用户数 $P = \text{预测流失(实际非流失)} + \text{预测流失(实际流失)}$ ， $P \cap C = \text{预测流失(实际流失)}$ 。预测准确率（查准率）等于正确预测用户数除以预测流失用户数。

打标签用户数据组离网预测准确率计算为：预测流失用户数 $P = 72\,022$ ，预测流失（实际流失）

$$P \cap C = 61\,720, \text{查准率} \frac{P \cap C}{P} = \frac{61\,720}{72\,022} = 85.70\%。$$

提前续约用户流失模型训练结果见表 1。

表 1 提前续约用户流失模型训练结果

矩阵（训练）	非流失（实际）/人	流失（实际）/人	合计/人
非流失（预测）	69 959	18 544	88 503
流失（预测）	10 302	61 720	72 022
合计	80 261	80 264	160 525
查准率	85.70%		

5.4 模型运用及结果

将模型结果应用到合约到期流失预警名单、提前续约/逾期续约流失预警名单、单卡类用户流失预警名单中，每月输出移动互联网用户中流失概率为极高/高的用户并对其进行月度维系挽留，每天输出流失概率极高的特定人数用户并进行预警。模型应用对比结果如下。

（1）解决目标不明确

传统手段无法精准且全面地捕获流失用户；流失模型可以准确地捕获流失目标。

（2）解决方向不明确

传统的维系手段无法对不同流失原因的用户差异化营销；流失模型可以准确识别用户离网原因。

（3）解决节奏不明确

以往的维系节奏单一，所有用户统一时间段维系；流失模型能够明确紧急流失用户和缓期流失用户。

（4）解决资源浪费

传统手段无法区分流失风险大小，无法做到资源合理投放；流失模型针对不同流失风险的用户，精准化维系资源投放。

6 结束语

用户画像的本质是深度分析客户，掌握有价值数据，找到目标客户，按照客户需求来定制产品，利用数据实现价值变现^[7,8]。运营商拥有丰富的海量数据，具有天然的大数据优势，在目前存量发展模式下，利用画像技术实现企业存量用户

价值经营是必然的选择,更为有效的画像方法有待进一步思考和实践。

参考文献:

- [1] 李英坤. 大数据背景下用户画像的统计方法实践研究[D]. 北京: 首都经济贸易大学, 2016: 2-59.
LI Y K. Statistical methods of user portrait in big data background[D]. Beijing: Capital University of Economics and Business, 2016: 2-59.
- [2] 韩晶, 张智江, 王健全, 等. 面向统一运营的电信运营商大数据战略[J]. 电信科学, 2014, 30(11): 154-158.
HAN J, ZHANG Z J, WANG J Q, et al. The unified-operation-oriented big data strategy for telecom operators [J]. Telecommunications Science, 2014, 30(11): 154-158.
- [3] 高玉龙. 基于文本挖掘的用户画像研究[D]. 汕头: 汕头大学, 2014: 1-54.
GAO Y L. Users portrait research based on text mining[D]. Shantou: Shantou University, 2014: 1-54.
- [4] 陈如明. 大数据时代的挑战价值与应对策略[J]. 移动通信, 2012(17): 14-15.
CHEN R M. The challenges and strategies in big data[J]. Mobile Communications, 2012(17): 14-15.
- [5] 张敏, 宋杰, 刘晓峰. 电信运营商面对 OTT 的战略选择[J]. 电信科学, 2014, 30(2): 142-146, 151.
ZHANG M, SONG J, LIU X F, et al. Strategic selection of telecom operators to counter OTT [J]. Telecommunications Science, 2014, 30(2): 142-146, 151.
- [6] 尤海浪, 钱锋, 黄祥为, 等. 基于大数据挖掘构建游戏平台个性化推荐系统的研究与实践[J]. 电信科学, 2014, 30(10): 27-32.
YOU H L, QIAN F, HUANG X W, et al. Research and practice of building a personalized recommendation system for mobile game platform based on big data mining [J]. Telecommunications Science, 2014, 30(10): 27-32.

- [7] 郝胜宇, 陈静仁. 大数据时代用户画像助力企业实现精准化营销[J]. 中国集体经济, 2016(4): 61-62.

HAO S Y, CHEN J R. User portrait helps accurate marketing in big data era[J]. China Collective Economy, 2016(4): 61-62.

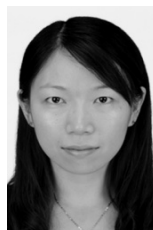
- [8] 袁海, 陈康, 陶彩霞, 等. 基于中文文本的可视化技术研究[J]. 电信科学, 2014, 30(4): 114-122.

YUAN H, CHEN K, TAO C X, et al. Research on visualization techniques based on Chinese texts [J]. Telecommunications Science, 2014, 30(4): 114-122.

[作者简介]



王晓霞(1987-), 女, 现就职于中国联合网络通信有限公司研究院大数据研究中心, 主要从事数据挖掘、数据建模、大数据处理等方面的研究工作。



刘静沙(1986-), 女, 现就职于中国联合网络通信有限公司研究院大数据研究中心, 主要从事电信领域存量经营、大数据挖掘和建模等方面的研究工作。



许丹丹(1989-), 女, 现就职于中国联合网络通信有限公司研究院大数据研究中心, 主要从事数据挖掘、数据可视化等方面的研究工作。