

# 基于大数据的用户画像构建及用户体验优化策略<sup>\*</sup>

□文 | 许志强 徐瑾钰

**[摘要]** 在审视大数据、用户画像与用户体验基本概念之间关系的基础上，首先探讨了基于数据驱动的用户画像生命周期管理思想，其次研究了基于大数据的用户画像构建逻辑，最后从价值、连接、生态、体验、创新五个维度揭示提升用户体验的优化策略，希望把大数据这个“最大变量”变成未来媒介融合发展的“最大增量”，助推传媒产业高质量发展。

**[关键词]** 大数据 用户画像 用户体验 媒介融合 标签

从工业视觉到计算机视觉、从人机交互到智能驾驶、从虚拟现实到物体自动识别、从智能解释到遥感解译，大数据正在像水电煤气一样持续为人类造福，并成为了信息基础设施中不可或缺的重要一环。大数据环境下的智能互联平台，一方面可将具有海量、异构、多维、多尺度等特性的用户数据从终端汇聚到云端并进行实时监测、分析和管理，另一方面无所不在、具备自我学习能力的终端又能不断地从云端汲取智慧，如此正向循环，成就着“大数据”向“智数据”的过渡。

## 一、大数据、用户画像与用户体验

在整个世界都由算法控制的互联网大数据时代，数据基于用户行为而生，大数据通过“量化一切”实现了世界的数字化。基于大数据的思维和工具，使得人类不仅可更快地通过智能设备完成数据采集、沉淀与归类等，还可得到用户的各种画像。这种变迁将对促进计算机与外部世界内容驱动和交互方式的演变，甚至对促进外部世界按照智能合约更加智慧的运行。

### 1. 大数据

数字时代，大数据实现了从“数据→信息→知识→决策”的一整套数据到应用变现，<sup>[1]</sup> 并从提供

数据支持的低级阶段进入到了拥有自身独立产业链的高级阶段。<sup>[2]</sup> 站在哲学的高度理解，大数据既可以描述客观物理世界，又可以刻画人类精神世界和人类社会，是融合“人类社会（Human society）- 信息空间（Cyberspace）- 物理空间（Physical space）”三元世界的纽带（见图1）。<sup>[3]</sup>

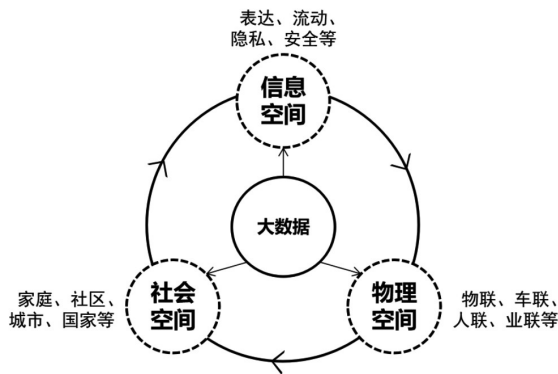


图1 大数据：三元世界的纽带

### 2. 用户画像

数字时代，用户数据（包括人口学特征、设备信息、商业属性等相对稳定的静态数据和访问频率与深度、停留时间、拖拽行为等实时变化的动态数据）直接或间接地反映出了用户行为习惯和态度偏好。若能将用户所有的数据快速归类与聚合分析并进行标签、权重与模型化处理，便能抽象出一个标

<sup>\*</sup> 本文系 2016 年度教育部人文社会科学规划基金项目“网络‘正能量谣言’的舆论生成与治理对策研究”（16YJA860006）、2017 年度国家社科基金重大项目“丝绸之路经济带沿线国家文化产业合作共赢模式及路径研究”（17ZDA044）、2016 年成都市科技惠民技术研发项目“全媒体视域下突发公共事件媒体应急平台研究”（2015-HM01-00632-SF）的成果之一

签化的用户模型，即用户画像。

### 3. 用户体验

数字时代，传统媒体“内容为王”的深度优势不再，“渠道为王”“终端为王”的格局也难撑全局，而“以人为本”的设计理念得到了包括传媒领域在内的各个领域的广泛认可，使得“用户为王”全面开启了新时代。由此可见，大数据与用户体验亦为正向循环，大数据为优化用户体验提供了有力工具，用户体验则为解锁大数据提供了密钥。

## 二、基于数据驱动的用户画像生命周期管理思想

研究发现，基于用户行为数据的用户模型呈现实时动态性。从访客到陌生人到最后成为高净值用户，在构建用户画像生命周期的不同阶段，传媒业界或可采取有针对性的拉新、转化、留存等运营策略。

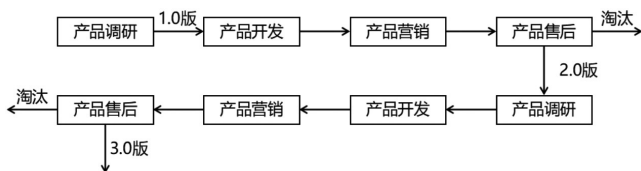


图2 基于数据驱动的用户画像生命周期

### 1. 产品调研

一般采用市场需求调查、产品 / 服务使用现状等方式来研究计划所属领域的整体用户属性特征，明确定位用户，并通过用户的人口统计、社会属性、用户消费、用户行为等基础数据初步构建起用户画像，从而完成对海量用户基本情况的调研。

### 2. 产品开发

一般采用用户体验、产品包装与产品定价等方式来测试用户与产品的匹配度。通过行为、用户、资源维度逐层钻取关联分析得到更精准细化的（用户或资源）群体或行为现象，对用户触点的理解是核心，解决标签与算法分发权重是关键，如此方能构建出包括用户标签、时间、行为类型、接触点（终端 + 内容）等信息在内的临时用户画像模型。

### 3. 产品营销

一般采用人群特征、广告创意、媒介选择与引流获客等受众与渠道分析的方式发现并获取潜在客

户、引导客户使用新产品并培养高净值客户。通过以上综合评估媒介质量及用户的使用习惯助力广告媒介的明智选择，实现精准投放。

### 4. 产品售后

一般采用购买行为、售后服务与神秘客等用户体验与留存的方式来重新激活或延长既得用户生命周期。通过对用户画像数字价值发现与应用不断地由粗到细、由大变小的筛选、过滤、剔除不符合条件的数据结果集，并向用户提供精确的服务指向，从而得到下一版本的用户画像的构建指标，如此循环、周而复始。

由此可见，消费不再是媒体流程的结束，一切数据汇聚到后台用户画像重构，每一个数据升级又要求算法做出更高反馈，最终复原出一个个数字化的“完整人像”。从某种意义上讲，人的思想、意识和行为方式，都可以通过数据的过滤及呈现，进行控制与干扰等。传媒机构可借助用户画像生命周期管理的理念打通并优化价值链，总结出适合自身目标定位的营销策划思路，选择一种或多种传播策略以多样化的形态竞争发展，最终在市场中立足扎根。

## 三、基于大数据的用户画像构建

构建用户画像，应从用户留在网络上的“数字足迹”的采集、整理与归类开始，在形成用户的个性化标签数据后，利用自然语言处理、机器学习、聚类分析等技术将业务广泛且复杂多变的多维度用户标签融合为一体，再对其重要特征进行分析、理解和可视化，从而更精准地推断出用户真实需求（见图3）。

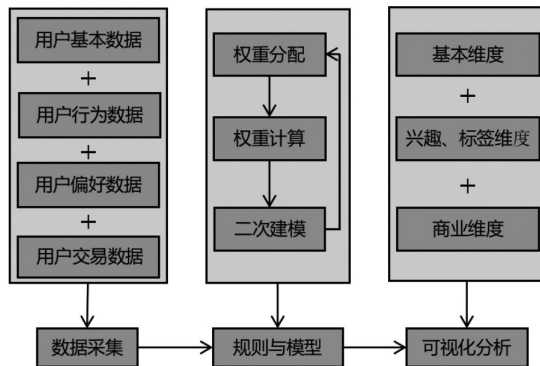


图3 用户画像构建流程

## 1. 数据采集与分析

伴随着信息技术从“计算”“连接”再到“智慧”的演进,传播正在经历从“简单的传播树”到“密集的传播森林”的进化。为更加精准地向用户提供服务,传媒机构可对用户的海量数据及潜在信息进行最大程度的采集、整理、挖掘与分析,并用高度精炼的特征标识来描述用户,从而构建出一整套完善的用户画像。抓取和分析用户“数字足迹”,对用户进行深浅度归类,判断其是否为活跃用户或者忠诚用户,并结合动态与静态标签从而构成完整的立体用户画像。以用智能设备播放视频作为采集用户多维属性数据为例,如果只看图4左侧的动作列表便无法预测用户将会做什么,因此还须知道如图4右侧所示的用户基本属性以及所处环境等并进行权重计算与分配。

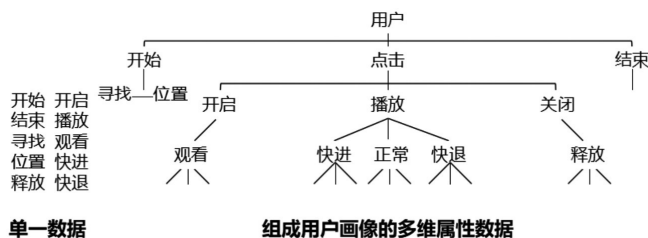


图4 组成用户画像的多维属性数据

## 2. 规则与模型

构建用户画像主要基于以下两个规则：一是以解决实际问题为目的,即传媒机构须结合自身性质,制定较明确的业务目标和推进能力；二是为进一步提升用户体验,即传媒机构在对用户基本属性有极高认知的前提下,将满足某种特定条件的用户群体提取出来,然后分析这一群体的行为特征从而对新增业务进行优化设计。

在行为建模时,可在界定数据源和清洗数据之后采用“4W”模式(when、where、who、what)给用户的行为标签赋予权重,提炼关键要素。when(时间),指收集用户时间因素(包含跨度和长度),即用户某种行为发生、结束的时刻和用户特定页面的浏览时间；where(地点),指收集用户接触点(内容和网址),即用户作用于的对象标签和用户行为发生的具体地点；who(对象),指收集用户标识因素(包括但不限于用户名称、电话、

微信、身份证号、邮箱等),且不同标识的可信度不同；what(行为),指收集用户事件因素(具象和抽象),即交互的商品/内容的标识和用户的网络行为。

由此可见,在内容展现与推荐过程中,当单个标签权重确定下来后,便可利用公式计算总的用户标签权重:标签权重=时间衰减因子×行为权重×接触点权重。<sup>[4]</sup>传媒业界可根据用户行为进行画像建模,在已有的用户数据中深挖并贴“标签”,刻画出用户的全网需求偏好模型,从而助推画像与场景精准匹配的媒介技术升级。

## 3. 数据可视化

传统的非可视化技术需要专业统计人员提前对数据进行提取汇总,但这种方式制定的图表属于静态性质,无法进行实时的动态调整。复杂广泛的用户数据,耗时耗力的汇集过程,终将导致传媒机构无法满足用户的多元化需求,也无法达到产品优化的目的。因此,如何将广泛复杂的用户数据转换成图形或图像并在屏幕上展现出来,则是数据可视化的关键所在。

目前传媒机构大多利用计算机图形图像技术,联系产品与用户交互的理论技术,达到数据可视化的目的。在实现数据可视化过程中,涉及计算机图形图像、视觉以及辅助设计等多个领域,成为研究数据表示、决策分析等一系列问题的综合技术。<sup>[5]</sup>传媒机构在信息触达环节中对用户进行数据可视化分析,一方面能形象地展示用户在某一接触点的历史行为数据(行为路径、行为特征等)合集,另一方面能形象地展示用户在各大网站上相互转换与兴趣偏好的数据合集。同时,还可根据用户价值来细分出核心用户、评估某一群体的潜在价值空间,从而提升信息和人的有效链接以及最大化的用户体验。

## 四、用户体验优化策略

数字时代,新闻传媒业面临着智能化3.0转型<sup>[6]</sup>与挖掘数据隐藏价值的双重挑战。媒体应从学理性思维和实践性思维的双重维度把握“媒介融合”,以内容、用户数据和服务为核心资源,以多元智能终端为载体,从价值、连接、生态、体验、创新等





多个维度提升用户体验,从而促使大数据从“最大变量”发展成为未来媒介融合发展的“最大增量”,助推传媒产业高质量发展。

### 1. 智能融合虚实两个世界,深挖用户价值

在信息爆炸和消费升级的双轮驱动下,可通过融合现实世界需求和虚拟分析空间的智能硬件自动采集客观反映人口属性、终端设备、用户行为等多维度特征数据,深层剖析量化偏好程度、挖掘隐藏在背后的模式、趋势和相关性,进行A(认知)、I(兴趣)、P(购买)、L(忠诚)的全链路运营,从而推进用户将“需求”转变为“消费”。与此同时,为实现对用户行为和事实的准确预测,还可运用新的数据分析技术和更完善的工具,将目标产品从面向海量广泛的用户市场转换为个别用户市场,从而提供精确、个性化的用户服务。

例如,北京歌华有线除收集收视率数据外,还密切关注点播次数、回看次数、暂停等“隐形”数据,其在完整记录用户实际喜好的同时,通过精确“推荐”的方式为用户提供个性化的智能电视收视服务,从而为节目制作、编排以及可能带来的收视效果提供科学参考。

### 2. 坚持数据导向,增强用户黏性

坚持以数据分析为导向,在充分了解用户基本状态属性和历史行为属性等数据的基础上,深入挖掘用户行为、兴趣和偏好等数据,并从多个触点和时间段的大数据分析中找到关键点推动客户最终完成购买,使用户连接从过去的“经验主义”的模糊方式迈向“数据驱动”的精确方式。在此基础之上,传媒机构可采取分流分级的方法实现对不同类型用户的管理。对潜在用户,通过对代表性常规属性以及行为属性数据的挖掘,提供针对性的服务将其转化为正式用户;对既得用户,通过可行性测试观察用户如何使用产品,尤其是使用过程中遇到的问题及原因进而不断迭代产品,从而提高用户的满意度、忠诚度,降低用户的使用成本。

### 3. 构建生态体系,引领用户连接

用户原始数据类型复杂、数量繁多,其价值更是不可估量。为确保可为用户提供精确、个性化的服务,各个细分行业须围绕数字化关键技术、能力和基础设施等不断进行突破。为此,传媒机构应尽快构建一种从数据存储、数据集成、数据处理到

数据分析以及数据可视化,相互依存、共同演化的“点一线一面一体”完整生态体系。该体系以大数据战略和架构为基础,除可随时采集用户实时数据与轻松浏览外,还可打开过去用户数据中的封闭结构,并通过降维攻击对产品功能进行更新升级从而高效地“连接”用户,推动快速实现价值(失败或成功)。

### 4. 融入互联网战略,构建用户画像

在产品生命周期的各个阶段,包括调研、研发、营销和售后等,数据导向战略均发挥着重要作用。通过对潜在或既得用户的大量行为数据的深度挖掘,可推动建立基于内容和用户画像的个性化推荐系统,从而为用户提供更准确的搜索结果、更优质的搜索体验。为优化用户选择最佳的收货地点,零售门店可提供任何地方购买、挑选或发货的能力,使用户感受到优质的服务体验;为更好地了解店内客流量以及更适当的人员配备要求,可利用传感器数据实时采集和分析入店率以及顾客消费情况等。

例如,零售巨头沃尔玛推出了一款针对购物的语义搜索引擎“Polaris”,其具备语义数据、文本分析、机器学习和同义词挖掘等的能力,可对Polaris上的数据进行挖掘和分析,从而构建画像并预测用户的行为。销售部人员通过用户在Polaris上的日常商品搜索数据统计了解当季最新最热产品,网站前端负责者则通过用户反馈不断完善网站类各板块属性,此举帮其在线购物的完成率提升了15%左右。

### 5. 借助新兴科技,提升用户体验

构建用户画像是为用户提供广泛的个性化推荐,从而提升用户体验的有效手段。在个性化推荐技术中,算法分发权重(编辑分发、社交分发、各种算法产出等)是个性化推荐质量的良好保证,而用户数据模型的质量好坏意味着能否正确表示用户的兴趣爱好。媒介融合环境下,基于内容的个性化推荐算法可提取用户的时空多维属性数据信息,融合线上与线下、融合现实与虚拟,实现“用户、场景、创意”在媒介融合环境下的精准传播,而这极有可能是一个涉及综合赋能iABCD(物联网、人工智能、区块链、云计算、大数据)等新兴科技的过程。<sup>[7]</sup>

在建构具体的算法模型时,可借鉴“M-P神

神经元模型”(McCulloch and Pitts, 1943),如图5所示。在传媒领域,用户行为便是神经网络中的最基本元素——神经元。该模型中,数据中心在接收到来自  $n$  个维度、具有自学习和自适应的能力的用户数据后,将接收到的所有数据互连、加权、累加,从而在不断传递的过程中形成一定数量的经验值,而所有经验值在整合后又可与数据中心的阈值进行比较,然后通过“激活函数”(Activation Function)向外输出优化建议<sup>[8]</sup>(其中  $\theta$  是所谓的“阈值”, $f$  就是激活函数)。用户数据经过此流程,层层筛选过滤,则可最终输出代表着该用户的完整画像  $V$  值,此举可为传媒机构精准化服务提供决策支撑。

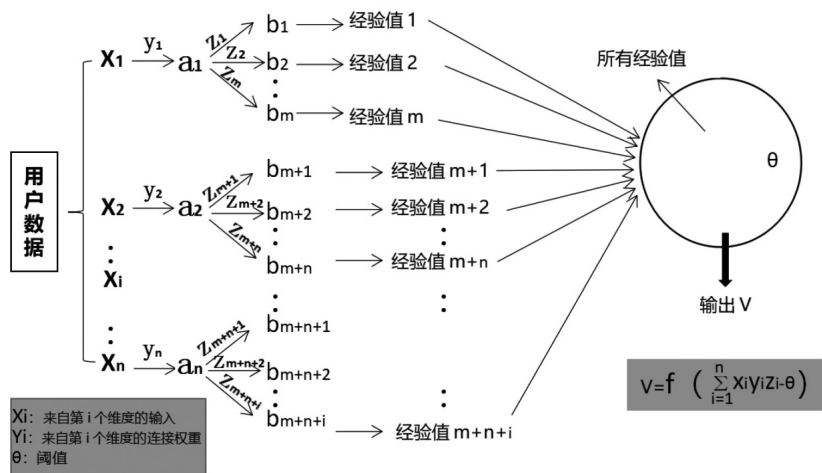


图5 “M-P神经元算法”模型图示

思想进行了探讨,然后根据现有环境以及技术,从深挖价值、增强黏性、引导连接、联网战略以及提升体验五个维度对用户体验的优化设计提出相应完善改进策略。

由于用户数据存在部分空缺,本文对用户的微观行为刻画比较有限,更多的是阐述了一种探析思路。今后,笔者将会以更多新算法技术与用户画像构建相结合的方式,获取更准确的用户行为和兴趣偏好,从而为媒介深度融合探索全路径,推进传媒产业高质量发展。

(作者单位:四川传媒学院“一带一路”文化传播中心 四川传媒学院数字媒体与创意设计学院)

## 五、结语

伴随着区块链、大数据智能分析等关键技术的交融发展,智能世界源于万物的感知被唤醒并助推千亿级别连接升级,链接将会增多变强并从不可信的“人-人”社交网络向高度可信、跨界融合的“人-机-物”复杂社交网络演进。本文首先从基于数据驱动的用户画像、用户体验及生命周期管理

### 参考文献:

- [1] 丁一,郭伏,胡名彩,等.用户体验国内外研究综述[J].工业工程与管理,2014,19(4):92-97
- [2] CSDN.问题来了,大数据的特性究竟有多少个V?[EB/OL].<https://blog.csdn.net/implok/article/details/79932798>[2018-4-13]
- [3] 用好大数据的六个诀窍,你知道多少?[EB/OL].[http://www.sohu.com/a/190131468\\_99922237](http://www.sohu.com/a/190131468_99922237)[2017-9-6]
- [4] 陈明.大数据可视化分析[J].计算机教育,2015(5):94-97
- [5] 王玉梅.数据导向的精准营销创新[J].中国广告,2015(07):128-129
- [6] 朱敏.智能媒体创新发展模式研究[J].中国出版,2016(12):17-21
- [7] 杨晶.用户兴趣模型及实时个性化推荐算法研究[D].南京:南京邮电大学,2013
- [8] E A Sobie, R C Susil, L Tung. A generalized activating function for predicting virtual electrodes in cardiac tissue[J]. Biophysical Journal, 1997, 73(3)