

## 12

## 产品研发中用户画像的数据模建

## ——从具象到抽象

Data Modeling of User Portrayal in the Product Development:  
From Concrete to Abstract

余孟杰 YU Mengjie

武汉理工大学, 武汉 430070 ( Wuhan University of Technology, 430070 Wuhan )

摘要：用户画像为进一步精准、快速地分析用户行为习惯、消费习惯等重要商业信息，提供了足够的数据库基础，完美地抽象出了用户的信息全貌，可以看作企业应用大数据的根基。

关键词：用户画像；数据建模；产品

Abstract: User portrayal can provide adequate data that perfectly grasp the overall information of users, which contains the important business information for enterprise to observe user's behavior and consumption habits more accurately and quickly. Therefore, it can be regarded as the foundation of data application for enterprises.

Key words: user portrayal; data modeling; product

中图分类号：J04 文献标识码：A doi: 10.3963/j.issn.2095-0705.2014.06.012 (0060-05)

从1991年Tim Berners-Lee发明了万维网开始，直到2011年，互联网真正走向了一个新的里程碑，进入了“大数据时代”。在经历了近两年热炒之后，人们逐渐冷静下来，更加聚焦于如何利用大数据挖掘潜在的商业价值，如何在企业中实实在在地应用大数据技术。在这个过程中，创新、个性化技术成为了一个重要落地点。相比传统的线下会员管理、问卷调查、购物篮分析，大数据第一次使得企业能够通过互联网便利地获取用户更为广泛的反馈信息，为进一步精准、快速地分析

用户行为习惯、消费习惯等重要商业信息，提供了足够的数据库基础。用户画像（user profile）的概念应运而生，它完美地抽象出一个用户的信息全貌，可以看作企业应用大数据的根基。

### 一、用户画像的创建

在用户体验设计中，最重要的原则就是了解用户。同样的概念其实早就存在，例如市场细分（market segment），使用者（actor），用户模型（user archetype）等。画像是近几年在个人人机

收稿日期：2014-11-17

作者简介：余孟杰，武汉理工大学艺术与设计学院博士生。

交互中时常提及的一个名词,用户画像通常是一些虚构出来的人物,用来代替最终的用户群体。明白光学知识的人都知道,一道彩虹里面存在上百种颜色,但是在光谱里,每种颜色都是不一样的,所以你得到的答案往往是7种,因为这七种颜色代表了上百万种颜色。同样,用户画像就像是彩虹里的7种颜色,用来代表上百万的不同的实际用户。举个简单的例子说明一下:

男,苏萌,31岁,已婚,月收入1万以上,博士生导师,最活跃的年轻创业家,新媒体营销数据分析专家,电子商务爱好者,喜欢红酒配香烟。这样一串描述即为用户画像的典型案列,如果用一句话来描述,即:用户信息标签化。如果用一幅图来展现(见图1):

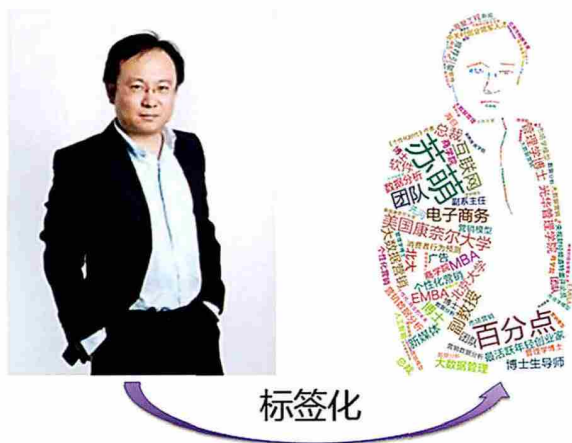


图1 用户图像的“标签化”

在这个例子中我们可以看到,画像是一个把数据形象化的方法。用户画像的核心工作是为用户打标签,打标签的重要目的之一是为了让人能够理解并且方便计算机处理。例如可以分类统计:喜欢红酒的用户有多少?喜欢红酒的人群中,男、女比例是多少?在产品设计的我们会遇到同样的问题,把抽象的用户数据转化到设计里面是不容易的。另外很多公司进行设计的和研究的不是同一个人,设计和研究之间的交流变得非

常重要。在许多的开发项目中,众多人员会在产品的不同阶段参与,这些人可能包括公司的高层管理、商业策划、项目经理等等。如果每个人心目中的用户都是有出入的,可以想象他们在产品的不同阶段所做的决定对产品有什么样的影响。在产品的开发中,我们需要一个工具,能够把抽象的数据具体化,也能够起到很好的交流作用。用户画像就是一个很好的工具,它把抽象的数据转化为虚拟的人物来代表个人的背景、需求、喜好等。设计师们可以通过考虑角色的需要,更好地推断一个真实人的需求。

用户画像在设计各个阶段起到交流的作用,有助于防止一些常见的缺陷。首先是所谓的“弹性用户”<sup>[1]</sup>,这是指在开发的过程中,相关的设计、开发人员和利益相关的决策者在描述用户需要什么、用户想干什么和用户希望什么的时候,因为用户未经定义,概念空洞广泛,所以这些相关的设计和开发人员几乎都能说任何他们想说的,而在实际操作中并没有真正的办法来反对这些观点。所以“用户想要什么,这类广泛而模糊的陈述应该能够被角色检验,以避免过去仅仅使用“用户”这个词来允许任何需求都可以被随便提出。用户的目、能力和使用情景不再空洞而广泛。用户画像也有助于防止“自我参考设计”,指的是设计师或开发人员可能会在不知不觉中,把自己的心智模式映射到产品设计中。可是设计师和开发人员的背景和理解与目标用户可能是截然不同的,角色在这里提供了实践中的检验,帮助排除通常不会发生的目标用户。

## 二、如何构建用户画像

一个标签通常是人为规定的高度精炼的特征标识,如年龄段标签:25~35岁,地域标签:北京。标签呈现出两个重要特征:语义化,人们能够很方便地理解每个标签的含义,这也使得用户画像模型具备实际意义;短文本,每个标签通常只是表示一种含义,标签本身无需再做过多文本分析等预处理工作,这为利用机器提取标准化信息提供了便利。人制定标签规则,并且能够通

过标签快速读出其中的信息，机器方便做标签提取、聚合分析。所以，用户画像即用户标签，向我们展示了一种朴素、简洁的方法用于描述用户信息。

### 1. 群体定量统计分析

每一个公司和项目都是不同的，建立角色的基础是不一样的。其中，就个体对用户画像的群体定量统计理解，可分为三个层次：第一层次是群体用户的调查分析；第二层次是数据分析的具象化个体描述；第三层次是抽象数据建模后产品的开发应用（见图2）。

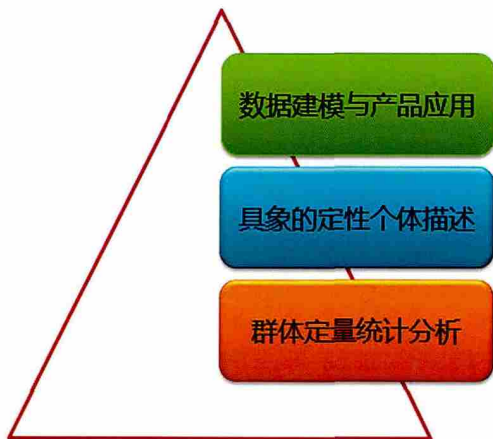


图2 群体定量统计

首先通过采用用户数据提取分析和市场问卷调查的方式设定用户画像的基础，并且根据产品的推进目标设定维度指标，从而对海量用户形成一个初步的了解。信息包括用户的注册资料、年龄、行为、需求、收入、家庭状况等，也可以从一些市场研究公司拿到用户的消费资料。分析的维度，可以按照结合定量和定性数据的方法。在建立角色之前，需要先想好参与团队，包括任何对最终用户有所了解的人，例如产品经理、公司的高层、商业策划、设计师、开发经理、销售等等。扩展信息包括产品类别、现场报告、活跃频率、个人喜好、使用习惯、消费习惯等。在用户细分的研究中，最理想的情况是把所有以上提到

的数据分为几个大类，并且指出哪些数据起主要作用，是用户分类的依据。

这里结合腾讯开放平台的数据进行定性定量分析（见图3）。从图3中我们可以看到，在网页游戏产品中，用户属性分析、年龄与性别的交叉分析、付费用户与年龄的交叉分析都需要研究的效度与信度论证。最理想的情况是把上面所有的数据放在一个大的矩阵中，然后进行聚类统计分析，每一个用户的大类，都可以作为一个角色。

案例一：网页游戏用户年龄与性别分布（见图3）

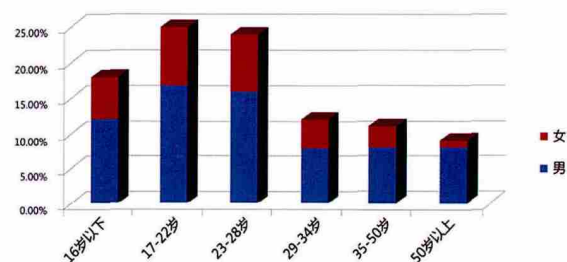


图3 网页游戏用户年龄、性别交叉分布

### 2. 具象的定性个体描述

定性个性描述也就是创造人物角色，Alan Cooper 最早提出了“Persona”这一概念，“Persona”是通过海量数据分析从而得到的一个虚拟在线用户。包括：（1）真实的访谈情境。（2）用户角色中包含姓名、照片和产品相关的描述。（3）与真实人物的配比度。（4）独特性和相似性。（5）该用户角色是否包含关键词和与产品相配合的高层次目标。（6）用户角色的记忆度。（7）设计决策是否容易执行。通过这七个方面去调研了解用户，根据他们的目标、观点和行为的差异，从而得到数据化的类比，然后从每种类型中抽取典型特征。对于网站来说，甚至可以在数据库里标注每一个用户的大类，可以作为一个角色，针对性地为不同的角色提供不同的功能和服务。2005年4月17日，美国《华盛顿邮报》上刊登了一篇为《零售业，为了利润而建的用户

模型——百思图商店迎合特定客户类型》的文章。文章中记载了百思图在改造的过程中,20%是针对商店里的产品,但是更多的80%是为了提高客户体验的。在基于一系列的数据分析,包括购买情况、当地人口普查数字、客户调查和有针对性的焦点访谈后,百思图开始改造67个美国加利福尼亚州的分店,以满足单个或者多个不同的客户模型。例如,为了让用户画像更加容易记忆,赋予个体一个名字、一张图片、一幅场景等描述,从而通过几条简单的关键特征进行描述,方便群体定位。

#### 案例二:蔡XX 单机游戏骨灰玩家

基本属性:男,23岁,本科毕业,已经使用YY3年,典型的单机任天堂3DS游戏玩家。目前的职业是外贸销售,在外贸公司工作,工作地方距离比较远,工作较忙,玩网游较少,曾经玩《梦幻西游》的时候上YY比较多。

上网习惯:常去的网站有:新浪微博、时光网、人人影视、目前多数时间在玩单机游戏。

收藏频道:DBK金牌大博客、泡泡恋爱官网、YY频道(ID:652521)。

频道进入:主要通过优酷、新浪微博组织的播客活动进入YY频道。习惯输入频道号直接跳转,以前很少去活动中心和频道排行。

用户角色的产生是一种具象化过程,是数据分析中假设和事实在一起的结合,它不但生动,而且可以大大提高团队的沟通效率。由于基于统一用户认知基础的沟通,“用户”可以让设计师找到“着力点”,从而在团队成员间对设计方案存在分歧时,依据“人物角色”就可以选择最佳的设计方案。接下来的大体过程是先把对用户理解的假设写出来,然后将有关用户的事实写出来和假设放在一起,根据大的类别做出角色的骨架,定出最后的角色,然后对角色进行比较详细的描述。这要把骨架里的列表变成一个虚拟的人,为了让角色比较好记忆,相片、短语和短句是很重要的方法,短句即是用很精练的一句话来概括这个角色。

### 3. 数据建模与产品应用

用户行为研究的价值,最终还是要体现在产品的应用上。一方面,使用统计分析的方法进行,设计师、产品经理在进行视觉设计、交互设计等方面要用团队的观察力和分析能力进行簇集分析和参考。另一方面,在使用数据做产品时,执行力的配合往往占据主要地位。

#### 案例三:英雄联盟对新用户的类别划分

新玩家进入游戏后,根据水平的不同会有不同的产品体验(见图4)。



图4 英雄联盟对新用户的类别划分

简单说来,这样一个过程是把假设和事实结合在一起,用亲和图的方式进行分类做出角色的过程。每个人把自己对这些用户的假设写出来,为了将来更加方便作亲和图分类。内容包括用户,以及该用户的目标、行为、活动,所遇见的问题等。接下来的工作要把角色骨架转化为有血有肉的人。当然,在实际工作中,也会遇到公司没有太多的人力和时间来系统地建立角色,在这种情况下,可以建立一些临时的角色。这是对用户需求 and 特点最好的粗略估计,不需要细节或者叙述,过多的描述反而给人虚假的精确。

此外,用户画像的核心在于对用户接触点的理解,接触点的内容直接决定了标签信息。接触点可以是网址,也可以是某个产品的特定功能界面。内容地址、行为类型、时间衰减,决定了权重模型是关键,权重值本身的二次建模则是水到渠成的进阶。此外,网址本身表征了用户的标签偏好权重,网址对应的内容体现了标签信息。用

户,对于电商有如下典型行为:浏览、添加购物车、搜索、评论、购买、点击赞、收藏等等。不同的行为类型,对于接触点的内容产生的标签信息,具有不同的权重。用户标签的权重可能随时间的增加而衰减,因此定义时间为衰减因子 $r$ ,行为类型、网址决定了权重,内容决定了标签,进一步转换为公式:标签权重 = 衰减因子  $\times$  行为权重  $\times$  网址子权重<sup>[2]</sup>。

如:用户A,昨天在品尚红酒网浏览一瓶价值238元的长城干红葡萄酒信息。

标签:红酒,长城

时间:因为是昨天的行为,假设衰减因子为:  
 $r=0.95$

行为类型:浏览行为记为权重1

地点:品尚红酒单品页的网址子权重记为0.9  
(相比京东红酒单品页的0.7)

则用户偏好标签是:红酒,权重是 $0.95 \times 0.7 \times 1 = 0.665$ ,即,用户A:红酒0.665、长城0.665。上述模型权重值的选取只是举例参考,具体的权重值需要根据业务需求二次建模,这里强调的是如何从整体思考,去构建用户画像模型,进而能够逐步细化模型。

比如游戏产品,典型接触点可能会是,关键任务,关键指数(分数)等等。百分点现已全面应用用户画像技术于推荐引擎中,在对某电商客户,针对活动页新访客的应用中,依靠用户画像产生的个性化效果,对比热销榜,推荐效果有显著提升。但是,接触点本身并不一定有内容,也可以泛化理解为某种阈值<sup>[3]</sup>,某个行为超过多少次,达到多长时间等。例如,同样一瓶矿泉水,超市卖1元,火车上卖3元,景区卖5元。商品的售卖价值,不在于成本,更在于售卖地点。标签均是矿泉水,但接触点的不同体现出了权重差异,即愿意支付的价值不同。

### 三、结论

以上更多的是阐述了一种分析思想,并未涉及具体算法,在计划构建用户画像时,能够给使用者提供一个系统性、框架性的思维指导。用户

画像也不是一成不变的,随着时间的推移,产品推介的重点可能会产生转移,画像角色也会发生更新。对于局部用户画像角色的描述,在比较大的系统设计中,有的时候在整个系统根据用户的作用建立角色以后,在设计某个部分时,会需要将其中的一个角色更加细分,这也是为了更好地适应市场需求。所以,目前在团队中不断强化和统一团队对于目标用户的认识,可以及时感知市场和用户的变化,保持用户画像的更新和生命力。

### 注释

图片来源于<http://blog.csdn.net/cuilanbo/article/details/39553189>。

图片来源:笔者收集、整理、绘制。

图片来源:2013年腾讯开放平台白皮书。  
<http://finance.qq.com/a/20140515/037286.htm>

图片来源于笔者搜集、整理。

Alan Cooper,“交互设计之父”,荣获视窗先锋奖(Microsoft Windows Pioneer)和软件梦幻奖(Software Visionary),他创建了Cooper交互设计公司,此公司一直致力于创建专为用户而设计的应用软件。

### 参考文献

- [1]黄建明,傅利民,饶培伦.人机交互:以用户为中心的设计和评估[M].北京:清华大学出版社,2013:64.
- [2]Jon Kolko.交互设计沉思录[M].方舟,译.北京:机械工业出版社,2012:67-68.
- [3]Hackos JT, Redish JC. User and Task Analysis for Interface Design[M]. New York: Wiley, 1998:35.

(责任编辑 曾璘)