

基于大数据的用户画像构建方法与运用

谭浩^{1,2}, 郭雅婷¹

(1.湖南大学, 长沙 431000; 2.汽车车身先进设计制造国家重点实验室, 长沙 431000)

摘要: **目的** 以大数据环境为研究基础, 为产品研发初期获取准确性较高的用户画像, 优化其构建方法。**方法** 基于定性与定量综合集成方法的三层理论框架, 首先通过定量研究挖掘线上社交平台的多模态用户数据并对其进行文本分析, 提取具有高影响的用户属性因子。其次, 使用优化模糊C—均值聚类算法(FCM 算法)对用户属性因子进行聚类分析, 获得几类用户属性原型。将其原型作为构建画像的维度框架, 通过线下定性研究方法对样本用户进行调研, 挖掘深层次需求动机。最终, 构建出几类用户画像原型。**结论** 以中国手机年轻用户相机拍照行为为研究对象, 通过线上社交平台和电商平台的公开数据获取用户拍照相关的评论信息, 对其进行文本挖掘和分析, 并对分析后的数据进行提取、聚类和定性研究, 从而获得3类不同拍照类型的用户画像模型。

关键词: 用户画像; 文本数据挖掘; 模糊聚类分析; 用户研究

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2019)22-0095-07

DOI: 10.19554/j.cnki.1001-3563.2019.22.017

Construction Method and Application of Personas Based on Big Data

TAN Hao^{1,2}, GUO Ya-ting¹

(1.Hunan University, Changsha 431000, China; 2.State Key Laboratory of Advanced Design and Manufacture for Vehicle Body, Changsha 431000, China)

ABSTRACT: The work aims to acquire accurate user portraits at the initial stage of product development with the big data environment as the research basis, and optimize its construction methods. Based on the three-level theoretical framework of the qualitative and quantitative comprehensive integration method, the multi-modal user data of online social platforms were firstly mined through quantitative research and text analysis to extract the user attributes with high impact. Secondly, fuzzy c-mean clustering algorithm (FCM algorithm) was used to cluster and analyze the user attributes, and several types of user attributes were obtained. The prototype was used as the dimension frame to construct the portrait, and the sample users were investigated through the offline qualitative research method to explore the deep-level demand motivation. Eventually, several types of user portrait prototypes were built. Taking the mobile phone camera photo behavior research project of young Chinese users as an example, the user's photo-related comment information is obtained through the public data on online data platform and e-commerce platform, and text mining and analysis is carried out. The analyzed data is extracted, clustered and qualitatively studied to obtain three kinds of user portrait models of different photo types.

KEY WORDS: persona; text mining; the fuzzy C-mean clustering algorithm; user research

近年众多企业在“互联网+”的带领下, 参与到了互联网产品的研发中。对于驱动用户体验的决策来说, 能够与用户行为绑定才是成功的标准。这些用户行为也一定是可以通过设计影响行为^[1]。网络社交文

化的迅猛发展连带出庞大的线上用户数据量, 用户在网络平台上的活跃使得用户数据呈现出多模态, 而这些数据经过分析可映射出用户潜在行为。因此, 从大数据的角度研究用户需求, 对于用户研究领域是有意义的。

收稿日期: 2019-08-21

作者简介: 谭浩(1977—), 男, 四川人, 博士, 湖南大学教授, 主要研究方向为汽车设计、人机交互和设计方法。

1 用户画像

在产品设计初期，为达到精准营销产品的目的，企业采取构建用户画像的方法，帮助设计和开发人员更好理解用户需求。阿兰·库珀^[2]1983 年提出以用户为中心的设计方法，其中就有过对用户画像相关概念的描述。通过识别用户特征，细分用户类型，以用户特征定义典型用户，可以发现用户使用产品的需求和倾向。Pruitt 和 Adlin 等^[3]研究了用户画像的概念及其深层含义，Goodwin^[4]认为用户画像将是行为、动机、目标相似的用户建立群组。

2 用户画像构建

根据数据来源的不同，目前对用户画像的构建方法主要分为两种：基于人文学科范畴的研究方向和基于计算机、数据统计等范畴的研究方向。其中，前者的构建数据来源于业务系统、事件系统和关系信息等，这个方向的缺点在于画像的构建无法做到更加准确、全方位。后者的数据来源主要依靠移动互联网、物联网、大数据等，可通过多渠道获取产品信息、用户社交信息、用户事件信息，从而构建一个全面多维度的用户画像^[5-7]。由于完全依靠数据，构建出来的画像相对而言缺少情感倾向，并且难以获得深层次的用户动机。

为了让用户画像能够更好促进产品与用户之间的关系，并获得用户的行为动机和期待趋势，管理决策领域中专家认为，如果只采用单一的模型或经验可能远远不够，需要将两者结合进行起来从而提升信度^[8]。向阳等^[9]基于人类解决复杂问题的思维模式提出了

定性与定量综合集成方法的三层次理论框架：定量分析层、定性处理层和综合集成求解层，其中定量分析层，主要是借助机器和定量的研究方法分析子问题，从而构建逻辑模型；定性处理层，可利用定性研究方法明确问题的概念属性和相互关系；综合集成求解层，通过组合子问题逻辑模型获得定量结论，运用相关领域知识进行定性归纳，形成解决问题关键，见图 1。

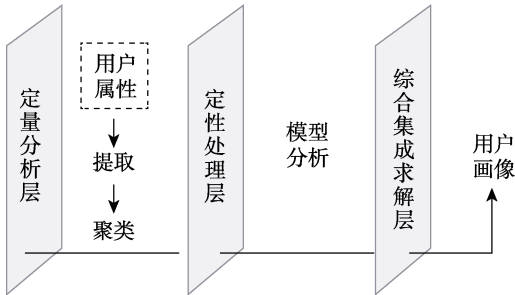


图 1 三层次理论获取用户画像
Fig.1 Persona obtained through three-level theory

2.1 用户属性

用户画像需要多维度的属性标签来构建，业界对于用户画像属性的相关课题有着大量研究，目前主要将用户画像的属性维度定义为 6 类：基本属性、兴趣属性、社交属性、行为习惯、心理属性、能力属性。刘蓓琳等^[10]在研究电子商务购买决策的人群分析中提出将这 6 类属性再分为一级属性和细化的二级属性标签，见图 2。其中二级属性标签主要是基础信息，一级属性标签为信息特征。

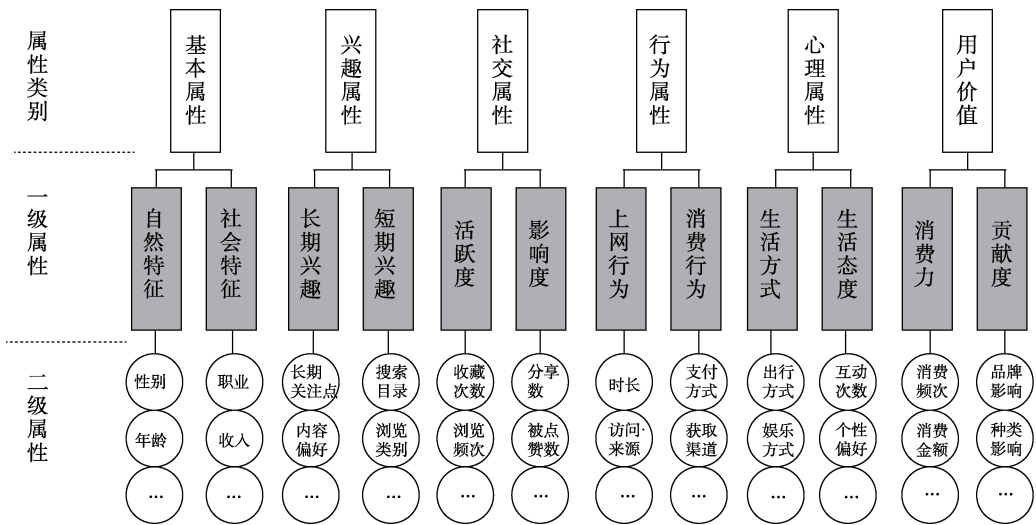


图 2 属性标签分类
Fig.2 Attribute label classification

2.2 文本数据挖掘

文本挖掘的概念最早由 Ronen Feldman 等提出，

可以在大量文本、语料数据中发现潜在具有价值的规律信息^[11]。文本挖掘的主要方法包括：对文本型数据

进行分类、聚类、观点挖掘、情绪分析、关键词自动取、话题的发现与聚类、语义分析、文档的自动摘要和实体关系模型等^[12]。卡耐基梅隆大学曾经研发的网页监视工具 WebWatcher, 可以通过监测用户的点击操作行为, 分析用户行为和兴趣。世界上最大的汽车销售站点 Auto Trader 通过数据挖掘获得用户对产品的喜爱偏好, 并且进行特定的服务。国内的专家学者认为文本挖掘是指从文本数据中获取可理解的、可用的知识, 同时运用这些知识更好地组织信息以便将来参考的过程。由于用户在网络平台上日趋活跃, 海量用户都成为了信息平台中一个具备传布和消费信息功能的节点, 大量的后台数据样本也为数据挖掘技术提供了富余的准备条件。

2.3 提取属性

Pazzani 等为得到用户属性, 收集已访问过的网

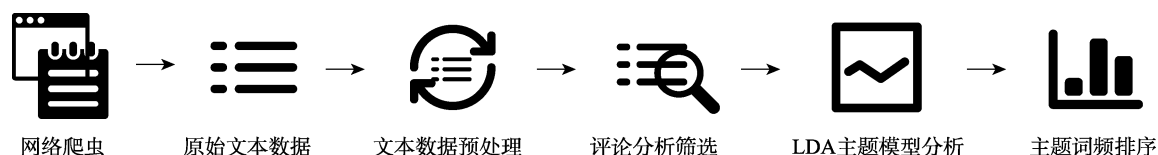


图 3 文本数据挖掘方法
Fig.3 Text data mining methods

2.4 聚类分析

为将用户建立成不同的组群, Hayman 等^[15]提出属性标签能够将拥有类似思想和共同关注点的用户聚类起来, 形成特定的社会群体。聚类分析是根据事物本身的特性对个体进行分类的方法, 将数据分成几个外部差异性大的群组, 但群组内部相似度高。现阶段的聚类分析研究领域主要有图像处理、模式识别、市场分析等。K-means 算法和模糊 c-均值聚类算法 (FCM 算法) 是目前应用最为广泛的两种划法。K-means 是一种硬划分算法, 它将对象排他地隶属于一个类簇, 如果有两个类簇有着较高重合度, 那么 K-means 将无法区分^[16]。Bellman 等在 20 世纪 60 年代提出了模糊集的概念, 用来处理聚类的问题。在模糊聚类分析中, 每个样本以一定的隶属度属于某个聚类, 这样的划分能更准确地描述出样本的真实分布, 可以获得更好的聚类效果。FCM 算法具备完善理论, 可以快捷处理大数据集, 但它的目标函数容易局限在极小值而无法达到聚类的最优解, 而基于这个问题崔适时等^[16]将优化的遗传算法与 FCM 算法结合, 从而提高了算法性能, 描述如下。

1) 创建初始群体和编码。选取 S 组对象, 每组 C 个簇, 然后按照类内距离最小, 类间距离最大的原则将 N-C 个对象划分到 C 个簇中, 这样就产生了 C 个聚类;

2) 计算适应度函数;

3) 执行选择、交叉及其变异操作, 保留父代以

页评分然后利用其进行机器学习用户画像的建立本身要立足于用户本身^[13]。由用户自发地在网络平台上写下的留言是企业商家获取客户实际需求的重要途径, 用户评论具备很多潜在的用户需求和情感信息。刘青, 郑献卫等专家学者基于主题抽取模型分析出产品特征或者基于情感分析评论文本的情感态度^[14]。获取文本型数据的方式有很多, 如 Web 页面数据, 它首先需要处理掉其 html 标签后, 再进行分词处理, 将其变为结构化的文本数据。在文本处理阶段, 常用的停用词以及出现频率很高信息价值却不高的常用词需要过滤与删除^[15], 之后把可用分词在文本中标, 通过统计得出高频词语片段, 并且邀请专家将这些分词进行评估赋予权重值, 作为用户属性, 见图 3。

由于用户画像需要以用户行为和目标需求作为属性来构建原型, 因此使用文本挖掘技术可以从大量文本中抽取潜在用户信息作为用户属性。

及下一代适应度高的个体, 合成新的下一代;

4) 如果达到了设定的繁衍代数, 算法结束; 否则继续第 3 步操作。

3 案例分析与运用

以中国手机年轻用户相机拍照行为研究项目为背景, 调研目前年龄范围在 18~25 岁间的中国年轻用户的手机拍照需求, 建立相关用户画像。定量分析线上数据样本来源于新浪微博等开放数据平台、4 个应用商城拍照 app 的用户评论; 线下样本选择主要针对中国一二三线城市展开, 并考虑南北、东西和中部相关用户的生活差异影响。采用一线城市 (北京, 上海, 广东, 深圳) 为支撑, 辐射到各类城市的方式共定 10 个城市, 30 个线下用户样本。流程见图 4。

3.1 评论数据收集与分析

GooSeeker 爬虫工具是一个集合网页抓取、数据抽取和页面信息提取的工具包, 操作便捷, 功能较为稳定; 它可以自动生成指令文件, 对指定网页 URL 和 HTML 内元素进行连续提取, 从而筛选出需要的信息, 输出打包文件内的文本数据带有语义结构。因此选用该软件包作为抓取评论数据的工具。为获得年轻用户的拍照相关的评论数据, 选取 2015 年至 2017 年中国年轻用户摄影社交活跃度高的平台的评论页面, 包括新浪微博摄影标签页和 Apple store、腾讯应用宝、百度手机助手和 360 手机助手中摄影类 APP 论页。

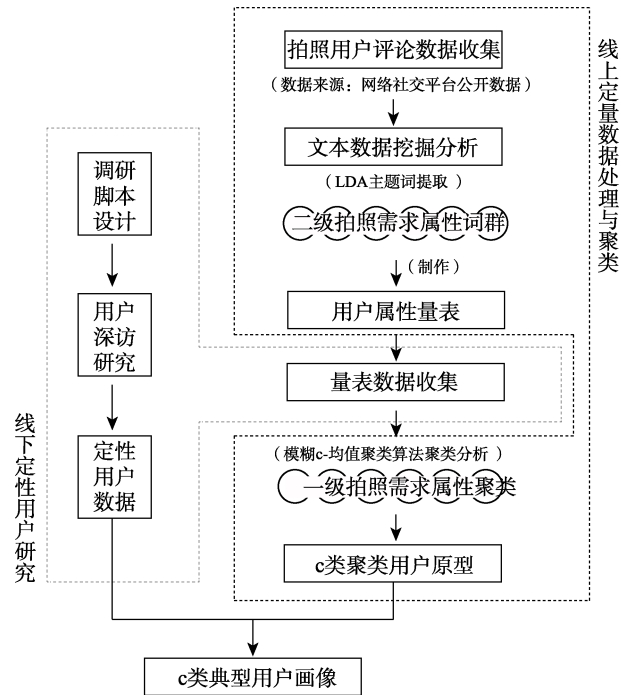


图 4 用户画像构建流程
Fig.4 Construction flow of the persona

通过工具的采集,得到结果大部分是以 Excel 格式储存,但由于一部分输出格式为 XML 文件无法直接使用和分析,需要将其利用 Swift 软件转化并把所有采集数据汇总成 Excel 文件。考虑到利用爬虫技术挖掘到的数据内会含有无效评论数据,若不除去将会对后面的模型分析造成影响,这里使用爬山虎采集器的高级筛选功能结合手工操作进行评论预处理,筛选条件为:(1)评论内容和拍照主题无关;(2)重复形

式的文案和三辨识度高的广告内容^[11],经过处理后获得 358743 条有效样本评论数据。

利用天据英眼工具通过机器学习的方式对导入的有效评论数据自动分词,该工具内带有 HowNet 情感词典可对自然语言进行处理,将常用的停用词表以及副词之类出现频率很高信息价值却不高的常用词过滤与删除,并且该工具具有分词打标功能,可统计出高频分词作为文本词语 N 库。LDA 主题模型是一种贝叶斯网络模型,具有清晰的层次结构“文档-主题-词语”,它也是用来挖掘文本主题的新方法^[14]。为了从数据集中获得 k 个主题词,即需求属性词,可在得知文本词语 N 的情况下通过概率方法反向得出主题词分布情况^[11]。

因此,将预处理过的有效评论作为文档,分词打标后获得文本词语 N 库作为条件,利用 R 语言编程分析得出近 158 个主题分词。由于机器处理获取的少数主题分词的词意可能无法成为有效的用户需求属性词,所以还需要对这近 158 个词进行再次筛选分析。我们邀请了 5 名相关专家,将这些主题分词结合相关评论文本句进行分词有效性评估,并算出对应的权重值。每个分词有效性评估的选项分为 5 个(等级):无效、有点效、有效、非常有效和极为有效,并分别赋值为 1, 2, 3, 4, 5。专家们分别进行评估后会得出每个分词 5 个等级的评价比例,选择“有效”以及以上数据统计,以 67%为界限,如果分词的“有效”、“非常有效”和“极为有效”的比例合计小于 67%,则删除该分词视为无效(删除 13 个分词)。通过计算每个分词的权重,选取权重值前 100 的分词作为需求属性词,见表 1。

表 1 需求属性词(主题分词)数据表(部分数据)
Tab.1 Data sheet (partial data) of requirement attribute words (subject segmentation)

主题词归纳(处理后)	权重值	有效性评级(%)					有效及以上(%)	相关评论条数
		无效	有点效	有效	非常有效	极为有效		
分享	0.013 9	0	20%	20%	40%	20%	80%	35 678 (条)
(变)好看	0.014 1	20%	0	40%	20%	20%	80%	30 243 (条)
记录	0.013 4	0	20%	0	40%	40%	80%	26 121 (条)
表情包	0.013 8	0	20%	20%	20%	40%	80%	18 746 (条)
...	...	...	...	...	...	...	...	...

3.2 用户属性量表

通过分析文本数据挖掘中获得的高频需求属性词,可以将这些用户需求属性词作为用户画像属性的一级属性和二级属性相关标签,作为构建用户画像的维度条件。5W2H 法是情境分析法(Scenario Analysis)的进一步深化和发展,通过概括基本影响因素可构建出产品基本框架^[17]。基于 5W2H 法的作用对用户属性量表进行设计,其中量表主要分为两个部分,第一部分将拍照过程细分为 10 个功能场景,客观记录用

户拍照中的操作轨迹(How-行为因素)和每个场景(功能)的使用频次(How much/many-程度因素)。第二部分定义出 5 个影响因素,获取用户深层次的拍照的需求和动机,包括,拍什么(Who-对象因素);拍的时候要什么(What-需求因素);为什么拍(Why-原因因素);什么时候会拍照(When-时间因素);会在哪里拍照(Where-空间因素)。基于表 1 主题词的类型区分,分别在每个影响因素下选择出 5 个词频排序高的需求属性词作为属性因子,并且赋值 0 和 1。部分数据见表 2。

表 2 用户属性量表收集结果（部分数据）
Tab.2 Results collected by user attributes scale (partial data)

序号	用户样本															
	因素	1	2	3	4	5	6	7	8	9	10	11	12	13	14	...
拍照行为轨迹	1 比例	7	3	5	2	6	7	9	4	2	5	2	1	3	2	...
	2 滤镜	3	1	1	2	3	3	4	1	0	2	1	0	1	1	...
	4 贴纸	2	3	3	5	5	2	0	5	6	2	1	6	0	2	...
拍什么	11 自拍	1	0	1	1	1	1	1	1	0	1	0	0	1	0	...
	12 （拍）宠物	1	0	0	1	0	1	1	1	0	1	1	0	1	1	...
	15 （拍）美食	1	1	1	1	1	1	1	1	1	0	1	1	1	1	...
拍照需求	16 贴纸 / 写字	1	0	1	1	1	1	0	1	1	0	0	1	0	0	...
	17 布局（美化）	1	1	1	1	1	0	1	0	1	1	0	1	1	0	...
为什么拍	21 分享	1	1	1	1	0	1	1	0	1	1	0	1	0	1	...
	24 记录（生活）	1	1	1	0	0	1	1	1	1	1	1	0	1	1	...
什么时候拍	26 旅游	1	0	1	1	0	1	1	0	1	0	0	1	1	0	...
	27 吃饭	1	1	1	0	0	1	1	1	1	0	0	0	1	1	...
	29 夜景	1	0	1	1	1	0	1	0	1	1	1	1	0	0	...
在哪拍	32 街拍	1	0	1	1	1	1	1	0	0	1	0	1	0	0	...
	35 主题性场所	1	0	1	1	1	1	1	1	1	0	1	1	1	0	...

3.3 聚类分析与结果

在用户画像的属性聚类分析阶段，模糊聚类的理论使用 matlab 数据处理工具进行属性因子聚类分析。Tatham & Anderson 等认为，聚类的数目为 3~6 个，

则研究发现将易于处理与沟通。在项目过程中通过和企业方沟通，设定聚类类别数 $c=3$ 。通过遗传优化 FCM 算法计算获得 3 类聚类原型的聚类中心矩阵表，见表 3（部分数据）。

表 3 聚类原型矩阵中心表（部分数据，显示小数点后一位）
Tab.3 Clustering prototype matrix center (partial data, to one decimal place)

因素序号	1	2	4	11	12	15	16	17	21	24	26	27	29	32	35
聚类一	3.1	1.2	0.3	0.5	0.6	0.7	0.6	0.7	0.5	0.7	0.6	0.4	0.6	0.5	0.6
聚类二	2.6	1.6	0.9	0.4	0.4	0.8	0.6	0.5	0.6	0.7	0.4	0.3	0.4	0.5	0.6
聚类三	3.3	1.8	0.6	0.6	0.5	0.8	0.5	0.4	0.7	0.6	0.5	0.4	0.5	0.6	0.4

运算适应度函数，执行选择、交叉、变异操作步骤，保留适应度高的个体形成下一代，如图 5 为适应度曲线结果，从图中看出平均迭代 30 次后开始适应度趋于最佳，100 次后终止运算。

将 30 个用户样本按照最大隶属度的对象分组，每个样本对 3 个聚类原型的隶属度的和为 1，最终得到图 6 为聚类树状图。从树状图中可以看到群组内部区相似度高，群组之间有明显差异。表 4 为 3 类聚类原型分布图以及其对应特征属性。

表 4 3 类聚类原型分布图以及其对应特征属性
Tab.4 Distribution diagram for three type of cluster prototypes and its corresponding characteristic attributes

聚类原型	用户样本
聚类原型一	18, 27, 29, 19, 30, 25, 8, 22, 2, 11, 14, 13
聚类原型二	4, 5, 9, 12, 15, 24, 16, 23, 17, 26, 28
聚类原型三	1, 3, 6, 10, 7, 20, 21

4.5 用户画像创建

为让用户画像更加生动形象，属性标签是不够的。产品设计和产品营销背后最打动用户的是情感需求，用户画像的创建应该更近一步挖掘用户需求背后

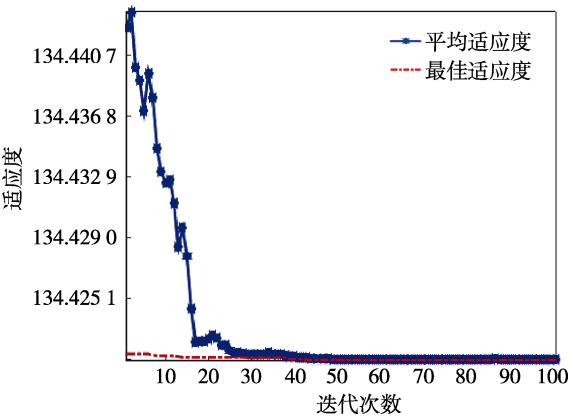


图 5 适应度曲线图
Fig.5 Fitness curve

的动机。为得到有价值、深层次的用户情感洞察。将通过定量研究分析出来的 3 类用户原型作为参照,对 30 名样本中和 3 类原型最近似的样本 1, 4, 18, 3, 5, 27 共 6 名用户(3 男 3 女, 每类用户样本 2 人)根本调研脚本进行深访。对其相关的生活形态、态度和拍照相关的因素进行深入调研,记录视频,录音数据等,见图 7。

最终获得三类用户画像,分别为文艺型、新潮型和俚人型,见图 8。在生活态度方面,文艺型用户享

受慢节奏生活方式,主动记录生活细节,喜好自己动手制作物品;新潮型用户追求与众不同,个性审美,有强烈的偏好取向;俚人型用户热爱网络社交,会主动自我展示,执着于自己的容貌,爱美。在拍照风格方面:文艺型用户追求氛围柔和、文静素雅,追求自然美;俚人型用户追求美颜画风,注重构图排版,画面需要突显自己;新潮型用户追求酷炫、夺人眼球的效果,运用特效、贴图等功能增强自己照片的艺术性。

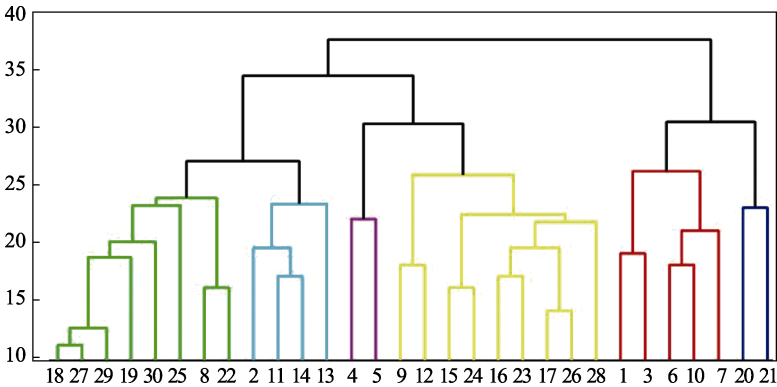


图 6 3 类聚类原型树状图
Fig.6 Tree diagram for three types of clustering prototype



图 7 调研现场
Fig.7 Field study research

画像类型	态度描述	拍照动机	用户拍照特征属性
画像一 文艺型	享受慢节奏; 自然形态生活方式; 喜好自己动手制作物品。	记录生活, 分享。	拍风景、拍美食、拍宠物; 滤镜渲染画面氛围; 旅游时、用餐时拍照; 海边风景照、充满氛围的室内照。
画像二 新潮型	追求独特性: 个性审美; 有强烈的偏好取向。	自我创作。	自拍、拍人像; 排版布局、特效功能处理照片; 贴纸、DIY 涂鸦装饰照片、多形式的照片效果, 如, 动态照片, gif; 恶搞他人, 制作表情包; 室外和公众场所拍照, 如街边或健身房等。
画像三 俚人型	热爱网络社交, 执着于自己的外在美。	展示自我, 获得他人认可。	自拍; 照片主要以人像为主; 修图功能, 人物颜值, 美颜功能(瘦脸瘦腿美白), 美妆功能、美妆功能试妆; 照片构图比例、画面效果氛围; 直播、推广自己或物品; 拍照空间不限制, 随时拍。

图 8 三类用户画像
Fig.8 Three types of personas

5 结语

本文基于大数据文本挖掘和模糊聚类的方法, 分析中国年轻用户群体拍照相关的行为属性特征。基于5W2H法的作用和文本挖掘分析提取的需求属性词, 制作出用户拍照相关属性的量表, 并从主客观等多方面收集用户量化值数据。利用优化FCM算法聚类出用户属性原型, 从而辅助创建出典型的用户画像和相关的产品机会点。相比传统的用户调研方法, 结合大数据分析的用户研究方法可以更加客观, 并且多渠道获得相关用户属性。由于数据基数大, 一方面增强了用户属性的信度, 同时也避免了过多的主观定义用户属性的因素, 在这基础上构建用户画像有利于促进产品研发的效率。本文主要研究的是从量化提取用户属性、属性维度聚类角度思考用户画像的构建方法, 但对于画像中属性维度如何转换成需求设计等问题并未综合考虑, 因此后续工作还需要进一步探讨研究。

参考文献:

- [1] GARRETT J J. The Elements of User Experience[M]. Beijing: China Machine Press, 2011.
- [2] COOPER A. The Origin of Personas[EB/OL]. [2003-08-12]. <http://www.cooper.com/journal/2003/08/>.
- [3] PRUITT J, ADLIN T. The Persona Lifecycle[M]. New York: Morgan Kaufmann, 2006.
- [4] Kim Goodwin. Perfecting Your Personas[M]. New York: Morgan Kaufmann, 2001.
- [5] IGLESIAS J A, ANGELOV P, LEDEZMA A, et al. Creating Evolving User Behavior Profiles Automatically[J]. IEEE Transactions on Knowledge & Data Engineering, 2012, 24(5): 854-867.
- [6] 席岩, 张乃光, 王磊. 基于大数据的用户画像方法研究综述[J]. 广播电视信息, 2017(10): 37-41.
XI Yan, ZHANG Nai-guang, WANG Lei. Overview of User Portrait Methods Based on Big Data[J]. Radio and Television Information, 2017(10): 37-41.
- [7] 丁伟, 王题, 刘新海. 基于大数据技术的手机用户画像与征信研究[J]. 邮电设计技术, 2016(3): 64-69.
DING Wei, WANG Ti, LIU Xin-hai. Research on Portrait and Credit Investigation of Mobile Phone Users Based on Big Data Technology[J]. Post and Telecommunications Design Technology, 2016(3): 64-69.
- [8] 祝世京. 定量模型与定性判断相结合的决策方法[J]. 控制与决策, 1993(4): 249-254.
ZHU Shi-jing. A Decision Making Method Combining Quantitative Model and Qualitative Judgment[J]. Control and Decision Making, 1993(4): 249-254.
- [9] 向阳, 于长锐. 复杂决策问题求解的定性与定量综合集成方法[J]. 管理科学学报, 2001, 4(2): 25-31.
XIANG Yang, YU Chang-rui. A Qualitative and Quantitative Integration Method for Solving Complex Decision Problems[J]. Journal of Management Science, 2001, 4(2): 25-31.
- [10] 刘蓓琳, 张琪. 基于购买决策过程的电子商务用户画像应用研究[J]. 商业经济研究, 2017(24): 49-51.
LIU Bei-lin, ZHANG Qi. Application Research of Ecommerce User Portrait Based on Purchase Decision Process[J]. Business Economics Research, 2017(24): 49-51.
- [11] 高玉龙. 基于文本挖掘的用户画像研究[D]. 汕头: 汕头大学, 2014.
GAO Yu-long. User Portrait Research Based on Text Mining[D]. Shantou: Shantou University, 2014.
- [12] FELDMAN R, DAGAN I. Knowledge Discovery in Textual Databases (KDT)[C]// International Conference on Knowledge Discovery and Data Mining AAAI, 1995.
- [13] PAZZANI M, MURAMATSU J, BILLSUS D. Syskill & Webert: Identifying Interesting Web Sites[C]// Thirteenth National Conference on Artificial Intelligence AAAI, 1996.
- [14] 刘青. 产品评论挖掘技术现状概述[J]. 电子制作, 2013(15): 249.
LIU Qing. Product Review Mining Technology Overview[J]. Electronic Manufacture, 2013(15): 249.
- [15] 李实. 中文网络客户评论的产品特征挖掘方法研究[J]. 管理科学学报, 2009, 12(2): 142-150.
LI Shi. Research on the Method of Product Feature Mining for Chinese Online Customer Reviews[J]. Journal of Management Science, 2009, 12(2): 142-150.
- [16] 崔适时, 刘震宇. 基于优化遗传算法的FCM[J]. 电脑编程技巧与维护, 2013(2): 17-18.
CUI Shi-shi, LIU Zhen-yu. FCM Based on Optimized Genetic Algorithm[J]. Computer Programming Skills and Maintenance, 2013(2): 17-18.
- [17] BEZDEK J C, EHRLICH R. Clustering Algorithm[J]. Computers & Geosciences, 1984, 10(2): 191-203.
- [18] HAYMAN S. Folksonomies and Tagging: New Developments in Social Bookmarking[C]// Ark Group Conference: Developing and Improving Classification Schemes, 2009.