

# 用户画像构建方法研究综述\*

高广尚

(桂林理工大学商学院 桂林 541004)

**摘要:**【目的】从设计与思维和数据类型两个角度分别探讨用户画像构建过程的机制。【文献范围】在 Google Scholar 和 CNKI 中分别以关键词“User Personas”、“User Profiles”和“用户画像”进行文献检索,再结合主题筛选,精读并使用追溯法获得用户画像研究的代表性文献共 90 篇。【方法】从设计思维角度研究画像的构建过程,具体结合目标导向、角色导向、参与导向、虚构导向这 4 个视角进行探讨分析;从数据类型角度研究画像的构建过程,具体结合本体或概念、主题或话题、兴趣或偏好、行为或日志、多维或融合这些概念进行探讨分析;对所述构建方法从逻辑思路、性能特点和局限性三个方面进行详细比较,最后对用户画像研究亟需解决的问题进行展望。【结果】用户画像技术在网络舆情治理、广告营销和个性化服务等诸多领域起着至关重要的作用。【局限】没有深入分析各用户画像算法的评价指标。【结论】尽管现有的用户画像构建方法能在一定程度上满足诸多应用的需求,但在大数据时代仍面临数据稀疏性、场景智能感知和用户兴趣迁移等挑战。

**关键词:** 用户画像 本体 主题 兴趣 行为日志

**分类号:** TP393

**DOI:** 10.11925/infotech.2096-3467.2018.0784

## 1 引言

以用户为中心的设计(User-Centered Design, UCD)是一种现代人机交互(Human-Computer Interaction, HCI)设计理念,其中,用户的需求、欲望和局限性等被探究和分析。作为用户研究的重要组成部分,用户画像(User Personas、User Profiles)是实现以用户为中心的交互设计的重要工具。通过用户画像,设计团队在产品、服务设计过程中能时刻关注用户及其需求,从而与用户达成共识。用户画像概念最初由交互设计之父 Cooper<sup>[1]</sup>提出,被用来作为一种交互式设计工具,以促进和巩固以用户为中心的设计思路<sup>[2]</sup>。用户画像又称为用户原型(User Archetypes)、用户模型(User Models)和生活方式快照(Life-style Snapshots)等<sup>[3]</sup>。

构建用户画像的过程本质上是以短文本(或加上图片)描述虚拟用户组的过程<sup>[4]</sup>,即把用户特征抽象成

短语标签,其中每个组内的虚拟用户具有相似的目标、需求和行为等<sup>[1,5]</sup>。这一描述过程中所涉及的短文本、图片称为画像描述(Personas Descriptions)<sup>[3]</sup>。现有研究中存在两类用户画像构建过程:一类是产品设计人员、运营人员根据用户需求从用户群体中抽象出典型用户(即 User Personas)<sup>[3]</sup>;另一类是根据每个用户在产品、服务中的行为、观点等数据,生成描述用户的标签集合(即 User Profiles)<sup>[6]</sup>。前者得到的画像本质上是一个描述用户需求的工具,用于帮助不同设计人员在产品、服务设计过程中站在用户的角度去思考问题。而后者得到的画像本质上是一个标签化的用户模型,用于刻画用户意图。很显然,前者重定性分析而轻定量计算,后者重定量计算而轻定性分析<sup>[5]</sup>。

用户画像已成为世界各国用户研究的热点,正日益引起业界、学术界的广泛关注<sup>[7-9]</sup>,更重要的是,它还是众多应用的关键技术之一,例如个性化推荐、广

通讯作者:高广尚, ORCID: 0000-0003-4140-1735, E-mail: 25969393@qq.com。

\*本文系国家自然科学基金项目“面向数据演化的增量实体解析方法研究”(项目编号: 71761008)和广西高校人文社会科学重点研究基地基金项目“面向企业数据治理的数据质量改善研究”(项目编号: 16YB010)的研究成果之一。

告系统等。鉴于此,笔者在 Google Scholar 和 CNKI 中分别以关键词“User Personas”“User Profiles”和“用户画像”,并不限定时间范围进行文献检索,通过阅读标题和摘要获取文献研究主题后,筛选得到中文文献 39 篇,英文文献 94 篇。经过精读并利用追溯法,最终对 90 篇有代表性的重点文献进行系统梳理,其中英文文献 72 篇,中文文献 18 篇。通过文献调研,本文从上述两类构建角度出发,对国内外现有研究中有代表性的用户画像构建方法进行梳理和总结,分析其中的研究成果及存在的问题,并在此基础上提出该领域中需要进一步研究的科学问题。

## 2 基于设计与思维的用户画像构建方法

通过调查问卷、用户访谈等方式,了解用户的共性与差异,然后据此分析和设计以形成不同的用户画像。鉴于这种构建方法固有的特殊性和复杂性,现有研究从 4 种视角分别提出具体的构建过程:目标导向视角(Goal-Directed)、角色导向视角(Role-Based)、参与导向视角(Engagement-Based)和虚构导向视角(Fiction-Based)<sup>[3]</sup>。其中,前三者的共同之处在于构建过程都以数据为基础,而最后一种却不以数据为基础,仅凭设计人员的直觉和假设。

### 2.1 目标导向的构建法

目标导向的构建方法主要关注用户通过使用设计团队所开发的产品(或服务)来达到什么目的,即围绕用户使用产品的目的构建用户画像。Cooper 等<sup>[10]</sup>认为这种方法构建出的画像能使设计团队更好地理解用户,并提出用于帮助完成构建过程的 7 个步骤。与此类似,Nielsen<sup>[3]</sup>从更细粒度的角度提出用于构建过程的 10 个步骤,尽管有些项目不需要遵循所有 10 个步骤。事实上,这种构建方法基于这样一种潜在假设:设计团队已开展了足够多的用户调研,并能明确其产品会给用户带来价值。然而,这种假设存在一定问题,它假定用户不是一般普通人,而是关注特定细节的独特人物。尽管这种构建方法能提供集中的设计思路并作为沟通工具完成讨论,然而它却因为低估用户参与的价值,而遭到业界批评<sup>[11]</sup>。

### 2.2 角色导向的构建法

角色导向的构建方法除了关注上述目标导向外,重点关注用户在其所属组织以及更多生活场景中所扮

演的角色<sup>[12-13]</sup>,即用户行为。该方法的出发点是传统的 IT 系统开发方法和目标导向的构建法所受到的批评,其中,前者在用户描述中缺乏明确性和一致性,后者低估了用户参与的价值。鉴于此,角色导向的构建方法做了重要补充:利用定性和定量材料对画像描述进行补充说明;明确数据和画像描述之间的关系<sup>[11]</sup>。事实上,考虑用户真实生活场景中的角色,能帮助设计团队在构建用户画像时做出更有利的设计决策。该方法被认为是一种不可单独使用的可用性方法,应与其他方法一起配合使用。

### 2.3 参与导向的构建法

参与导向的构建方法根植于用故事创造参与和洞察力的能力<sup>[12]</sup>。通过对故事及其人物的理解,设计团队有可能据此创造出更生动、更接近现实情况的用户画像。该方法的目的是让设计团队认识到用户具有他人无法识别的刻板印象和无法想象的生活,从而让设计团队积极参与故事中人物的生活,即虚拟用户的生活。这种方法的优点是不给画像描述带来风险,例如歧视性描述,因为它考虑用户的诸多方面,例如情感因素、心理因素和社会背景等,而不仅专注于用户行为<sup>[14]</sup>。事实上,这种方法产生的画像描述能在数据和有关真实应用程序与虚构信息(旨在创造同理心)的知识之间取得平衡。此外,这种方法反对自动思维,并且满足上述两种构建方法的需求。

### 2.4 虚构导向的构建法

虚构导向的构建方法强调通过设计团队的假设推导出可能的典型用户,其产生的用户画像经常被用于探索如何进行设计,以及在现场产生讨论和洞察力<sup>[15]</sup>。在该方法构建的用户画像中,即席用户画像(Ad Hoc Personas)和极端人物(Extreme Characters)是比较典型的例子。其中,Norman<sup>[16]</sup>认为即席用户画像用于在设计过程产生移情焦点(Empathetic Focus),其构建过程基于设计人员的直觉和经验。类似地,Djajadiningrat 等<sup>[17]</sup>认为极端人物有助于产生设计见解并探索设计空间的边缘。Blythe 等<sup>[18]</sup>认为 Pastiche 场景(模仿场景)可以构建出像 Bridget Jones 或 Ebenezer Scrooge 这样的虚构用户,从而帮助设计人员在创造场景的时候具有反射性反应(Reflexive),即本能反应。这种方法的优点是在设计初期可以作为初步推测用户需求的工具,但不适合用来进一步指引产品和服务设计。

### 3 基于本体或概念的用户画像构建方法

利用本体或概念中定义的结构化信息和关系信息来刻画用户。事实上,本体是一种强大的知识表示手段和合理的推理机制<sup>[19]</sup>。现实系统通常包含用户本体(定义不同用户特征及其关系)、领域本体(定义电子商务系统中产品目录及其关系)和交互本体(定义用户与系统交互的语义)等。

#### 3.1 基于本体的构建法

Middleton 等<sup>[20]</sup>研究了学术论文推荐系统中用户画像方法,其中用户被表示为一个个基于主题的本体,而文章转化为对应的词向量去与之进行匹配。值得说明的是,表示用户的主题实体来自于用户的反馈和浏览历史,而且在实际的 QuickStart 和 Foxtrot 系统中,基于本体的推断可以方便地提升用户画像准确率,并能应对冷启动问题。Edwards<sup>[21]</sup>设计并实现一个可重用、可扩展和语义 Web 兼容的客户特征本体 OntoProfi,它不仅具有表示电子商务系统的实体(例如产品、订单履行、通信),以及这些实体之间关系的类层次结构,还具有各种规则和限制,因而使推理者能将每个顾客划分为特定类别。Golemati 等<sup>[22]</sup>首先创建一个用户概要本体,该本体包含用于构建用户画像的概念和属性,然后在此基础上构建通用的、全面的和可扩展的用户画像。该方法的优点是考虑与用户上下文和概要信息领域相关的现有文献、应用程序和本体。Jayanthi 等<sup>[23]</sup>通过监视和存储用户的浏览习惯来动态地构建和更新配置文件,从而实现隐式构建基于本体的用户画像。Calegari 等<sup>[24]</sup>研究从现有 YAGO 本体中提取知识以构建用户画像的方法,该方法不仅结合用户个性检索信息,还对其中词语进行相应赋权。Skillen 等<sup>[25]</sup>研究从用户概要本体(User Profile Ontology)中构建可扩展的用户画像方法,该方法的一个显著特征是它关注用户的动态和静态信息。

在国内方面,单晓红等<sup>[26]</sup>提出以在线评论数据为基础,从用户信息属性、酒店信息属性和用户评价信息属性三个维度来构建用户画像,并采用 Protégé 工具建立本体实现用户画像属性之间的关联,最终完成对酒店用户特征的完整刻画。这种方法的缺点是对数据源的准确性和酒店用户评论的数量有较大的依赖性。郑建兴等<sup>[27]</sup>针对微博信息文本短、信息少、主题兴趣更新快、主题热度变化快等特点,提出基于百度百科

建立本体分类关系,利用本体的部分结构表示用户模型。牛温佳等<sup>[28]</sup>提出基于知识的用户画像,其核心思想是从知识工程的角度出发,利用本体对用户画像中的标签进行表示、验证、推理和解释等。

#### 3.2 基于概念的构建法

Leung 等<sup>[29]</sup>基于搜索引擎日志和点击记录,并使用排序学习方法,构建基于概念的用户画像,其中的概念向量(Concept Vector)既具有正值也具有负值。类似地,Sasikala 等<sup>[30]</sup>基于用户浏览文档的内容和搜索历史抽取概念并建立概念关系图,然后利用排序支持向量机(Ranking Support Vector Machine, RSVM)算法学习用户偏好,其中用户偏好使用加权概念向量表示,最后构建出基于概念的用户画像。此外,有学者尝试建立一个具有层级的概念结构来表示、组织用户的画像描述<sup>[31-32]</sup>。

### 4 基于主题或话题的用户画像构建方法

通过主题模型或话题模型发现文本信息中隐含的主题或话题,进而据此刻画出用户。用户产生的文本信息通常包含多个主题或话题(Topics),它们表现为一系列相关的词语,且具有不同的概率。

#### 4.1 基于主题的构建法

Pazzani 等<sup>[33]</sup>尝试利用朴素贝叶斯(Naive Bayes)算法<sup>[34]</sup>分类用户的话题表示用户。在 Syskill & Webert 系统中,用户被表示为热点页面(频繁访问的页面)上具有概率值的词语表。事实上,由于这种基于话题词语的表示方法:一方面可以方便地使用先验知识(如过去的用户概要、词汇关系数据库 WordNET<sup>[35]</sup>等);另一方面可以利用朴素贝叶斯算法分类少量数据时所具有的最优性<sup>[36]</sup>,因此它能保证准确率和运算时间,进而改进用户画像质量。此后,Billsus 等<sup>[37]</sup>尝试从新闻媒体的隐式、显式反馈建模用户的概要,同时包含基于 KNN 的短期概要和基于朴素贝叶斯的长期概要,以自适应地获得用户信息需求。

LDA 以较好的建模能力以及相对较低的计算复杂度被广泛研究和应用<sup>[38]</sup>。Tang 等<sup>[39]</sup>基于主题模型方法构建包含语义的学者网络概要信息,其中包含概要信息提取、名字消歧、用户兴趣发现(用户-会议-主题概率模型)三个步骤。Kim 等<sup>[40]</sup>从用户的阅读水平、主题分布这两方面研究如何刻画用户,所提出的概率主

题模型包含阅读内容和阅读任务的多样性等信息。Venington 等<sup>[41]</sup>结合用户历史浏览行为建立用户兴趣主题和文档主题,计算每个文档个性化评分,进而重新进行排序,将结果用于个性化信息检索。林燕霞等<sup>[42]</sup>从社会认同理论出发,使用主题模型文本挖掘出用户感兴趣的微博主题,并利用得到的用户偏好主题的概率分布空间向量,计算相应的用户相似度,实现用户所属群体的分类。由于缺乏用户浏览行为、关注分组等信息,该方法无法构建更加全面、精准的用户画像。Chen<sup>[43]</sup>通过对微博用户信息特征的研究和分析,定义微博用户模型,提出将主题表示与向量空间模型表示相结合的微博用户模型构建方法;此外,还设计一种基于动态更新用户模型的学习算法,并通过实验验证该学习算法的有效性。

## 4.2 基于话题的构建法

曾鸿等<sup>[44]</sup>将对同一话题感兴趣的一类人聚合到一起,以获取这类人群的信息,然后借助算法提取该类人群标签。尽管该方法应用在新浪微博上对明星粉丝人群进行画像,但并未从对比角度给出有效性验证。李文峰<sup>[45]</sup>通过借助用户感兴趣的主题特征信息,提出结合主题特征的半监督模型,在模型中通过用户相关文本的主题特征信息建立主题特征文本,然后将主题特征文本引入到主题模型的建模过程中,让主题特征文本影响模型中主题分布。由此,通过利用用户相关文本和主题特征信息,间接地构建了用户画像。该模型未通过引入社交网络信息来加强对用户特性的探索,并找到用户之间差别的根源。郭光明<sup>[46]</sup>提出基于社交数据用户隐行为模式的用户信用画像方法,综合考虑行为数据和文本数据,同时在更细粒度上捕获用户的隐含行为模式、个性和习惯。该方法既具有 LDA 建模文本数据的优势,也具有普通特征提取方法建模行为数据的优势。值得一提的是,这种方法在很大程度上也适应用户产生的其他类型社交大数据,例如 Facebook 数据、微信数据等。

## 5 基于兴趣或偏好的用户画像构建方法

利用用户经常浏览或关注的信息刻画用户。事实上,这些用户信息可看作是用户兴趣偏好的外在表现。

### 5.1 基于兴趣的构建法

Pazzani 等<sup>[47]</sup>总结多种针对用户兴趣的用户画像

构建方法,并在此基础上提出基于用户所喜欢商品的描述信息、用户对于特定商品的交互历史这两类信息,构建用户画像的方法。有学者从用户产生的标签入手发现用户的兴趣概要。Carmagnola 等<sup>[48]</sup>依据词法学和 WordNet 分类标签构建和丰富用户的兴趣概要模型(用户画像)。Li 等<sup>[49]</sup>则直接使用高效率的关联规则挖掘算法<sup>[50]</sup>计算用户兴趣模型。张小可等<sup>[51]</sup>通过对移动互联网用户的历史数据进行多元回归分析,得到贝叶斯网络的初始参数,再利用 Netica 软件构造一个贝叶斯网络模型,并根据用户实时发生的业务行为更新网络的参数,以此刻画每个用户的画像。Wu 等<sup>[52]</sup>提出基于用户兴趣与用户社交关系的共同演化模型,通过该模型能在社交元素动态演化条件下更好地构建用户画像。Wang 等<sup>[53]</sup>通过挖掘自适应新闻系统的 Web 日志数据,分析用户点击行为与新闻报道类别之间的关系,以建立用户兴趣模型。Li 等<sup>[54]</sup>提出一个系统的用户兴趣分布挖掘框架(User Interest Distribution Mining, UIDM),从用户行为数据中提取可解释的用户兴趣。在解决方案中,作者首先使用概率潜在语义分析(Probabilistic Latent Semantic Analysis, PLSA)发现用户与其行为之间的关系,可以将其描述为隐藏的主题;然后通过手工标注特征实体矩阵,构建隐藏主题与用户兴趣之间的映射矩阵。Hoang<sup>[55]</sup>提出社区和个人兴趣(Community and Personal Interest, CPI)模型,利用内容和行为,将微博用户与社区用户的兴趣进行联合建模。CPI 模型还提供一个通用框架适应多种类型的用户行为。模型的优点是它并没有假设个人兴趣与社区兴趣之间存在等级关系,即一个完全基于另一个。

### 5.2 基于偏好的构建法

Sugiyama 等<sup>[56]</sup>利用用户浏览行为、结果评分等信息构建用户偏好概要,这类似 Billsus 等<sup>[37]</sup>的研究,根据不同的搜索会话(Search Session)将用户偏好分为长期和短期两大类。Mianowska 等<sup>[6]</sup>根据用户的搜索行为动态地调整用户偏好,同时使用加权的词汇表示用户的偏好概要,并在大学图书馆的搜索系统中验证了这种动态性的有效性。Agichtein 等<sup>[57]</sup>发现独立于查询(Query)的浏览和交互特征,对于用户点击数据理解解非常重要,这些信息也是用户偏好的一部分。Filipova 等<sup>[58]</sup>研究使用贝叶斯网络、决策树和神经网络等数据挖掘技术确定影响客户消费模式和购买偏好的因素。



## 6 基于行为或日志的用户画像构建方法

利用丰富的行为、日志和点击历史记录数据刻画用户。事实上,这些数据中的大部分能反映用户当时的真实心理需求。因此,分析这部分数据可对画像描述起到一个重要的补充作用。

### 6.1 基于行为的构建法

Fawcett 等<sup>[59]</sup>在欺诈检测应用中通过规则发现的方法获取电话记录中的欺诈指数,利用这些指数构建用户行为画像。Rafter 等<sup>[60]</sup>将基于行为的用户画像技术归类为被动技术(Passive Techniques),并且根据用户的阅读时间和重访问行为构建用户概要信息。Adomavicius 等<sup>[61]</sup>开展了面向个性化的用户画像工作,以发现隐藏在用户商品交易记录中的行为概要信息,使用包括分类规则、关联规则在内的数据挖掘方法<sup>[62]</sup>。另外,他们又研究了如何评估挖掘出的行为习惯规则的质量、如何使用专家知识推动规则发现的用户画像系统<sup>[63]</sup>。Moudani 等<sup>[64]</sup>提出一种利用知识发现和数据挖掘技术对海量异构和噪声数据进行精确业务预测的模型,并在此基础上开发一种评估和分析客户行为的模糊方法,以更好地理解客户群体。

Nawaz 等<sup>[65]</sup>提出一种多用户个性化电子邮件社区检测方法,基于邮件用户的结构和语义亲密度发现他们的分组。张慷<sup>[66]</sup>通过挖掘用户的上网行为特征,同时结合套餐信息、终端信息、业务订购等信息,进行深度数据融合及交叉分析,构建通信用户画像。汪强兵等<sup>[67]</sup>以移动平台下的 Web 阅读系统为工具,通过收集用户在移动设备上浏览文献时产生的用户手势行为(单击、双击、滑动、拖动、放大/缩小等)数据,以及用户手势行为相对应的文本内容,并结合对应文本内容的浏览时间,构建用户模型。该方法的缺点是构建过程没有考虑点赞、转发和评论文章等影响因素。

此外,为从点击行为角度构建用户画像,Wang 等<sup>[68]</sup>从点击流数据中捕获用户行为,并以直观的方式对检测到的行为进行可视化,在此基础上,通过划分相似图识别相似用户的“集群”(节点是用户,边是通过点击流相似度加权)。随后,Wang 等<sup>[69]</sup>进一步研究描述大型在线服务中用户行为的点击流模型,希望实现两个目标:捕获不同的用户组,用于检测恶意账户;从用户组中提取语义信息,以理解捕获的行为。Zhang 等<sup>[70]</sup>提出一种自底向上的量化数据驱动方法构建用户画

像,通过点击和点击流数据直接合并用户行为识别 workflow,然后使用混合模型聚合常见的工作流,以创建数据驱动的用户画像。以往依赖调查或用户访谈数据识别 workflow 的方法与此截然不同,因为那些数据与用户行为没有直接关系,弱反映了用户在产品中的实际工作流程,收集成本高,只局限于几百个响应,并且随着用户 workflow 的发展而过时。

### 6.2 基于日志的构建法

Nasraoui 等<sup>[71]</sup>研究如何根据动态网站上的网络使用日志构建动态可演化的用户行为画像,他们提出的网络使用日志挖掘框架可以挖掘、跟踪、验证动态的多种用户画像信息。Iglesias 等<sup>[72]</sup>则从用户在 UnixShell 上的命令日志研究动态用户画像问题,以此获得计算机用户的行为画像。为利用挖掘出的结果发现相似用户, Ma 等<sup>[73]</sup>提出一种新的方法以克服以往行为模式空间的稀疏性,其中涉及将用户交互记录转换为更一般的表示,以规范化每个用户的原始上下文日志。Chen 等<sup>[74]</sup>通过改进和应用频繁的顺序模式挖掘技术,提取用户经常访问的位置序列日志,并使用它们建模用户的移动概要。Zhu 等<sup>[75]</sup>通过分析用户情景日志,提出一种情境感知的移动用户画像构建方法。Wang 等<sup>[76]</sup>研究搜索日志数据中的用户画像问题,并在非参数贝叶斯框架内提出生成模型 dpRank。通过对用户搜索行为的生成假设,dpRank 以联合的方式识别每个用户的潜在搜索兴趣和不同的搜索结果偏好。段建勇等<sup>[77]</sup>通过对网络查询日志进行聚类分析,将相似度大的查询词聚类,并构建用户兴趣模型,该方法的缺点是聚类过程开销比较大。黄文彬等<sup>[78]</sup>在分析和挖掘移动用户记录基站数据的基础上,采用频繁模式挖掘、构建概率矩阵、计算熵等方法,从用户基站日志所包含的地理位置信息中构建移动用户行为画像。该方法因为用户数据存在部分空缺而无法对用户的微观行为进行有效刻画。

## 7 基于多维或融合的用户画像构建方法

通过多种特征类型的数据从多个维度对用户进行刻画。很显然,多维数据是构建立体用户画像的可靠来源,能显著改善用户画像的质量。

### 7.1 基于多维的构建法

Rosenthal 等<sup>[79]</sup>同时利用文本特征和社交特征构建博主用户画像并进行年龄分类研究。Mueller 等<sup>[80]</sup>

针对 Twitter 用户的用户性别推断问题,提出一种新的名称分类器,而不是仅仅依赖于一个名称数据库。该方法从用户名中提取特征,并使用这些特征将名称分配给一个性别。类似地,Marquardt 等<sup>[81]</sup>建立一个特征集以让性别及年龄预测的准确性最大化。An 等<sup>[82]</sup>利用 YouTube 社交媒体数据,包括人口数据、客户互动和话题兴趣等多维数据,实时自动构建用户画像,不需要人工开发过程;此外,为生成更丰富的画像描述,还从其他社交媒体渠道收集用户数据,并将其与相应的用户数据进行匹配。Jung 等<sup>[83]</sup>通过从主要社交媒体平台挖掘汇总的大规模用户在线数据构建用户画像,该方法的优点是能构建基于反映真实用户的真实数据的用户画像,并具有灵活性和适应性,适用于作者正在探索的各种环境。王凌霄等<sup>[84]</sup>从用户资历、用户参与度、用户回答质量以及用户发展趋势 4 个方面构建用户画像,并结合具体用户数据验证结果。尽管作者还提出标记用户特征的一些方法,并讨论处理社会化回答社区用户特征时可能遇到的问题及应对措施,但没有从文本内容(例如话题、问题)等更为细致的角度分析用户画像。

7.2 基于融合的构建法

费鹏等<sup>[85]</sup>提出一种多视角融合框架构建用户画

像,利用双通道对不同特性的用户分别建模预测;针对不同类型的数据提出多种特征萃取方法,以构建用户多源特征体系(包括类别特征、聚类特征和数值特征等);为充分利用多源特征解决高维特征的问题,进一步提出基于双层 Xgboost 的多视角融合模型,其中 Xgboost 是一种经过优化的提升树模型<sup>[86]</sup>。尽管该框架具有一定的合理性和有效性,但是计算量较大。李恒超等<sup>[87]</sup>认为现有的构建方法在短文本相关性预测、各标签属性之间的关联预测方面表现不佳,鉴于此,基于用户的查询词历史记录,提出一种用于预测用户多维标签的二级融合算法框架。尽管进一步提升了预测的准确性,但该框架仍存在一些可以改进的方面,例如选择合适的外部语料训练词向量、使用更复杂的网络结构等。

8 构建方法之间的详细比较

总体而言,上述这些方法的构建过程从流程上可分为三个主要步骤:收集用户数据;研究用户信息、细分标签;建立和丰富用户画像描述。目前,数据的多样性、异质性,以及庞大的数据规模等给构建过程带来了极大挑战。本文中所述构建方法在逻辑思路、性能特点和局限性这三个方面的详细比较如表 1 所示。

表 1 用户画像构建方法详细比较

构建方法	逻辑思路	性能特点	局限性
基于设计与思维	通过调查问卷、用户访谈等方式,了解用户的共性与差异,据此分析和设计以形成不同的用户画像。	1.通过开放性的问题获得用户真实的心理需求,具象用户特征; 2.特征或者维度都相对容易解释; 3.数据有限时需要结合定性与定量方法。	1.无法确定用户群体是不是真的; 2.画像结果在很大程度上取决于设计人员对目标用户群的直觉判断。
基于本体或概念	利用本体或概念中定义的结构化信息和关系信息刻画用户。	1.有较强的语义表达和逻辑推理能力; 2.本体的结构化和语义性可以弥补社会化标签的不足。	1.需要对不同的领域分别构建相应的本体结构; 2.需要专家参与; 3.费时费力且成本高。
基于主题或话题	通过主题模型或话题模型发现文本信息中隐含的主题或话题,进而刻画出用户。	1.不考虑词语的顺序只考虑它们出现的次数; 2.可以捕捉隐含的语义信息; 3.将用户和资源文本一起考虑。	1.无法真正地将用户本身的特征引入到模型中; 2.分析正面情感、负面情感时可能存在偏差。
基于兴趣或偏好	利用用户经常浏览或关注的信息刻画用户。	1.充分利用不同用户相似的兴趣偏好; 2.一定程度上克服过多的无关词语带来的负面影响。	1.不能很好地处理用户兴趣、偏好中的动态性、实时性; 2.只能从有限的历史行为中估计用户的兴趣和偏好。
基于行为或日志	利用丰富的行为、日志和点击历史记录数据刻画用户。	1.利用丰富的用户个性化信息; 2.充分利用不同用户相似的行为模式。	1.无法很好地处理用户行为日志中的非理性因素; 2.未能很好地解决用户行为日志中的数据稀疏性、数据异质性问题。
基于多维或融合	通过多种特征类型的数据从多个维度对用户进行刻画。	1.综合考虑用户多方面的特征; 2.采用多级模型或从多角度分析属性之间的关联信息。	1.多级模型本身需要较大计算量且复杂; 2.涉及特征维数可达上万级别。

## 9 结 语

用户画像技术在网络舆情治理、广告营销和个性化服务等诸多领域起着至关重要的作用,主要体现在让信息形成关联、目标用户明确和重难点开始清晰等方面。诚然,用户画像技术发展的深入,正在改变着行业的产品开发方向、营销方式和应用场景,也会加速消费者日常生活方式的数字化转变。本文对以往研究中一些有代表性的用户画像构建方法及相关思路进行分析探讨,为下一步的深入研究打下良好基础。展望未来,用户画像技术在数据稀疏性、场景智能感知和用户兴趣迁移等方面仍存在开放性的研究方向。

### (1) 数据稀疏性

在现实生活中,用户买过什么商品、看过什么电影、看过什么书、甚至在什么商品面前停留了多长时间等行为所产生的数据非常稀疏,具体表现为特征稀疏和语义稀疏两个方面<sup>[88]</sup>。这些数据相对整体而言具有较低的覆盖度,而且不同用户之间重合度也很低,传统方法难以从中找到相似性,或获得有用的信息以生成有效的属性标签。尽管这些数据很难用,但却很有用,因此如何在构建用户画像时对这些数据进行合适的平滑处理,或从不同角度采用多种方法进行处理是很有意义的研究方向。

### (2) 场景智能感知

在现实生活中,用户可能处于多种生活场景之中,例如社交场景、消息场景、资讯场景、搜索场景和网购场景等<sup>[89]</sup>。由于这些场景本身之间存在较大的差异,因此用户在不同场景下会有不同的行为、需求等。为将特定场景中的用户精准细分出来,需要构建方法能充分挖掘具体的应用场景数据,甚至衍生组合数据,这样才能让画像更符合实际应用场景。因此如何在构建用户画像时考虑加入时间段、地理位置等信息,从更细粒度准确感知、识别场景,进而确定哪些用户何时处于什么样的特定场景将是最前沿的研究方向之一。

### (3) 用户兴趣迁移

在现实场景中,用户兴趣的迁移、变化可能会受多个因素的影响,例如用户自身兴趣发生改变、用户兴趣受外部因素的影响等,具体表现为容易受到人群、环境、热点事件、行业等方面的影响<sup>[90]</sup>。从本质上来讲,用户兴趣的迁移不仅呈现出多样性、差异性

和碎片化特征,而且还呈现出特有的长期兴趣(兴趣随时间缓慢变化)、短期兴趣(兴趣随某些事件的发生剧烈变化)等现象。因此如何在构建用户画像时平衡用户长期和短期的兴趣,并加强针对兴趣特征的分析,以避免画像失真是重要的研究方向。

## 参考文献:

- [1] Cooper A. The Inmates are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity[M]. Sams Publishing, 2004.
- [2] Pruitt J, Adlin T. The Persona Lifecycle: Keeping People in Mind Throughout Product Design[M]. Morgan Kaufmann Publishers Inc., 2010.
- [3] Nielsen L. Personas - User Focused Design[M]. London: Springer, 2013.
- [4] Lerouge C, Ma J, Sneha S, et al. User Profiles and Personas in the Design and Development of Consumer Health Technologies[J]. International Journal of Medical Informatics, 2013, 82(11): e251-e268.
- [5] Brickley J, Walczak S, Burgess T. Comparing Semi-Automated Clustering Methods for Persona Development[J]. IEEE Transactions on Software Engineering, 2012, 38(3): 537-546.
- [6] Mianowska B, Nguyen N T. Tuning User Profiles Based on Analyzing Dynamic Preference in Document Retrieval Systems[J]. Multimedia Tools and Applications, 2013, 65(1): 93-118.
- [7] Nielsen L. Personas in a More User-Focused World[A]// Nielsen L. Personas-User Focused Design[M]. Springer, 2013: 129-154.
- [8] Salminen J, Jung S G, An J, et al. Findings of a User Study of Automatically Generated Personas[C]// Proceeding of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 2018: 1-6.
- [9] Chen R, Liu J. Personas: Powerful Tool for Designers[A]// Luchs M G, Swan K S, Griffin A. Design Thinking: New Product Development Essentials from the PDMA[M]. Wiley, 2016: 27-40.
- [10] Cooper A, Reimann R, Cronin D, et al. About Face: The Essentials of Interaction Design[M]. The 4th Edition. Wiley, 2014.
- [11] Grudin J, Pruitt J. Personas, Participatory Design and Product Development: An Infrastructure for Engagement[C]// Proceedings of the 2002 Participatory Design Conference.

- 2002: 144-161.
- [12] Sønderstrup-Andersen E. Personas: En Domæneanalytisk Tilgang[J]. Dansk Biblioteksforskning, 2007, 3(2): 61-75.
  - [13] Nielsen L. Introduction: Stories About Users[A]// Personas-User Focused Design[M]. Springer, 2013.
  - [14] Nielsen L. Engaging Personas and Narrative Scenarios[OL]. [2018-03-22]. <https://openarchive.cbs.dk/bitstream/handle/10398/6448/working%20paper%202004%2016.pdf?sequence=1>.
  - [15] Floyd I R, Jones M C, Twidale M B. Resolving Incommensurable Debates: A Preliminary Identification of Persona Kinds, Attributes, and Characteristics[J]. Artifact, 2008, 2(1): 12-26.
  - [16] Norman D. Ad-hoc Personas & Empathetic Focus[A]// The Persona Lifecycle: Keeping People in Mind During Product Design[M]. Morgan Kaufmann Publishers Inc., 2006: 154-157.
  - [17] Djajadiningrat J P, Gaver W W, Fres J. Interaction Relabelling and Extreme Characters: Methods for Exploring Aesthetic Interactions[C]// Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques. ACM, 2000: 66-71.
  - [18] Blythe M A, Wright P C. Pastiche Scenarios: Fiction as a Resource for User Centred Design[J]. Interacting with Computers, 2006, 18(5): 1139-1164.
  - [19] Liu W, Jin F, Zhang X. Ontology-Based User Modeling for E-Commerce System[C]// Proceedings of the 3rd International Conference on Pervasive Computing and Applications. IEEE, 2008: 260-263.
  - [20] Middleton S E, Shadbolt N R, De Roure D C. Ontological User Profiling in Recommender Systems[J]. ACM Transactions on Information Systems(TOIS), 2004, 22(1): 54-88.
  - [21] Edwards N. E-commerce Website Personalisation Based on Ontological Profiling[D]. Wales: Cardiff University, 2015.
  - [22] Golemati M, Katifori A, Vassilakis C, et al. Creating an Ontology for the User Profile: Method and Applications[C]// Proceedings of the 1st International Conference on Research Challenges in Information Science. 2007: 407-412.
  - [23] Jayanthi J, Jayakumar K S, Surendran S. Generation of Ontology Based User Profiles for Personalized Web Search[C]// Proceedings of the 3rd International Conference on Electronics Computer Technology. 2011: 240-244.
  - [24] Calegari S, Pasi G. Personal Ontologies: Generation of User Profiles Based on the YAGO Ontology[J]. Information Processing & Management, 2013, 49(3): 640-658.
  - [25] Skillen K L, Chen L, Nugent C D, et al. Ontological User Profile Modeling for Context-Aware Application Personalization[C]// Proceedings of the 2012 International Conference on Ubiquitous Computing and Ambient Intelligence. 2012: 261-268.
  - [26] 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): 1-10. (Shan Xiaohong, Zhang Xiaoyue, Liu Xiaoyan. Research on User Portrait Based on Online Review: Taking Ctrip Hotel as an Example[J]. Information Studies: Theory & Application, 2018, 41(4): 1-10.)
  - [27] 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. (Zheng Jianxing, Zhang Bofeng, Yue Xiaodong, et al. Research on Themes Recommendation in Micro-Blogging Scenario Based on Neighbor-User Profile[J]. Journal of Shangdong University: Natural Science, 2013, 48(11): 59-65.)
  - [28] 牛温佳, 刘吉强, 石川. 用户网络行为画像: 大数据中的用户网络行为画像分析与内容推荐应用[M]. 北京: 电子工业出版社, 2016. (Niu Wenjia, Liu Jiqiang, Shi Chuan. User Network Behavior Portrait: The Analysis of Users' Network Behaviors and the Application of Content Recommendation in Big Data[M]. Beijing: Electronic Industry Press, 2016.)
  - [29] Leung K W T, Lee D L. Deriving Concept-Based User Profiles from Search Engine Logs[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(7): 969-982.
  - [30] Sasikala P, Vidhya V. An Efficient Concept-based Mining Model for Deriving User Profiles[J]. International Journal of Applied Information Systems, 2014, 1(6): 26-34.
  - [31] Bloedorn E, Mani I, Macmillan T R. Machine Learning of User Profiles: Representational Issues[OL]. arXiv Preprint, arXiv: 9712002.
  - [32] Trajkova J, Gauch S. Improving Ontology-Based User Profiles[C]// Proceedings of the RIAO'04 Coupling Approaches, Coupling Media and Coupling Languages for Information Retrieval. 2004: 380-390.
  - [33] Pazzani M, Billsus D. Learning and Revising User Profiles: The Identification of Interesting Web Sites[J]. Machine Learning, 1997, 27(3): 313-331.
  - [34] Rish I. An Empirical Study of the Naive Bayes Classifier[C]// Proceedings of the 2001 IJCAI Workshop on Empirical Methods in Artificial Intelligence. 2001: 41-46.
  - [35] Miller G A. WordNet: A Lexical Database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
  - [36] Domingos P, Pazzani M. On the Optimality of the Simple



- Bayesian Classifier Under Zero-One Loss [J]. Machine Learning, 1997, 29(2-3): 103-130.
- [37] Billsus D, Pazzani M J. User Modeling for Adaptive News Access[J]. User Modeling and User-Adapted Interaction, 2000, 10(2-3): 147-180.
- [38] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [39] Tang J, Yao L, Zhang D, et al. A Combination Approach to Web User Profiling[J]. ACM Transactions on Knowledge Discovery from Data(TKDD), 2010, 5(1): Article No.2.
- [40] Kim J Y, Collins-Thompson K, Bennett P N, et al. Characterizing Web Content, User Interests, and Search Behavior by Reading Level and Topic[C]// Proceedings of the 5th ACM International Conference on Web Search and Data Mining. ACM, 2012: 213-222.
- [41] Veningston K, Shanmugalakshmi R. Combining User Interested Topic and Document Topic for Personalized Information Retrieval[C]// Proceedings of the 2014 International Conference on Big Data Analytics. Springer, 2014: 60-79.
- [42] 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. (Lin Yanxia, Xie Xiangsheng. User Portrait of Diversified Groups in Micro-blog Based on Social Identity Theory[J]. Information Studies: Theory & Application, 2018, 41(3): 142-148.)
- [43] Chen Z. Modeling Research on Micro-blog Users[C]// Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering. 2013.
- [44] 曾鸿, 吴苏倪. 基于微博的大数据用户画像与精准营销[J]. 现代经济信息, 2016(16): 306-308. (Zeng Hong, Wu Suni. User Image and Precision Marketing on Account of Big Data in Weibo[J]. Modern Economic Information, 2016(16): 306-308.)
- [45] 李文峰. 基于主题模型的用户建模研究[D]. 北京: 北京邮电大学, 2013. (Li Wenfeng. Topic Model Based User Modeling[D]. Beijing: Beijing University of Posts and Telecommunications, 2013.)
- [46] 郭光明. 基于社交大数据的用户信用画像方法研究[D]. 合肥: 中国科学技术大学, 2017. (Guo Guangming. User Credit Profiling Techniques for Online Users with Big Social Data[D]. Hefei: University of Science and Technology of China, 2017.)
- [47] Pazzani M J, Billsus D. Content-Based Recommendation Systems[A]// Brusilovsky P, Kobsa A, Nejdl W. The Adaptive Web[M]. Springer, 2007: 325-341.
- [48] Carmagnola F, Cena F, Gena C. User Modeling in the Social Web[C]// Proceedings of the 2007 International Conference on Knowledge-Based and Intelligent Information and Engineering Systems. Springer, 2007: 745-752.
- [49] Li X, Guo L, Zhao Y E. Tag-Based Social Interest Discovery[C]// Proceedings of the 17th International Conference on World Wide Web. ACM, 2008: 675-684.
- [50] Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules[C]// Proceedings of the 20th International Conference on Very Large Data Bases. 1994: 487-499.
- [51] 张小可, 沈文明, 杜翠凤. 贝叶斯网络在用户画像构建中的研究[J]. 移动通信, 2016, 40(22): 22-26. (Zhang Xiaoke, Shen Wenming, Du Cuifeng. Research on Bayesian Network in User Portrait Construction[J]. Mobile Communications, 2016, 40(22): 22-26.)
- [52] Wu L, Ge Y, Liu Q, et al. Modeling Users' Preferences and Social Links in Social Networking Services: A Joint-Evolving Perspective[C]// Proceedings of the 30th AAAI Conference on Artificial Intelligence. 2016: 279-286.
- [53] Wang W, Zhao D, Luo H, et al. Mining User Interests in Web Logs of an Online News Service Based on Memory Model[C]// Proceedings of the IEEE 8th International Conference on Networking, Architecture and Storage. 2013: 151-155.
- [54] Li J, Zuo X Q, Zhou M Q, et al. Mining Explainable User Interests from Scalable User Behavior Data[J]. Procedia Computer Science, 2013, 17: 789-796.
- [55] Hoang T A. Modeling User Interest and Community Interest in Microbloggings: An Integrated Approach[C]// Proceedings of the 2015 Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, 2015: 708-721.
- [56] Sugiyama K, Hatano K, Yoshikawa M. Adaptive Web Search Based on User Profile Constructed Without Any Effort from Users[C]// Proceedings of the 13th International Conference on World Wide Web. ACM, 2004: 675-684.
- [57] Agichtein E, Brill E, Dumais S, et al. Learning User Interaction Models for Predicting Web Search Result Preferences[C]// Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2006: 3-10.
- [58] Filipova B T, Martinovska C. Analysing Customer Profiles Using Data Mining Techniques[C]// Proceedings of the 34th International Conference on Information Technology Interfaces. 2012: 73-78.
- [59] Fawcett T, Provost F. Combining Data Mining and Machine Learning for Effective Fraud Detection[C]// Proceedings of the 2nd International Conference on Knowledge Discovery

- and Data Mining. 1996: 14-19.
- [60] Rafter R, Smyth B. Passive Profiling from Server Logs in an Online Recruitment Environment[C]// Proceedings of the 17th International Joint Conference on Artificial Intelligence. 2001.
- [61] Adomavicius G, Tuzhilin A. User Profiling in Personalization Applications Through Rule Discovery and Validation[C]// Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 1999: 377-381.
- [62] Witten I H, Frank E, Hall M A. Data Mining: Practical Machine Learning Tools and Techniques[M]. United State: Morgan Kauffman, 2011.
- [63] Adomavicius G, Tuzhilin A. Using Data Mining Methods to Build Customer Profiles[J]. Computer, 2001, 34(2): 74-82.
- [64] Moudani W, Zaarour G, Mora-Camino F. Fuzzy Classification of Customer Insolvency in Mobile Telecommunication[J]. International Journal of Decision Support System Technology, 2014, 6(3): 1-29.
- [65] Nawaz W, Khan K U, Lee Y K. A Multi-User Perspective for Personalized Email Communities[J]. Expert Systems with Applications, 2016, 54: 265-283.
- [66] 张慷. 手机用户画像在大数据平台的实现方案[J]. 信息通信, 2014(2): 266-267. (Zhang Kang. Implementation Scheme of Mobile User Portrait in Big Data Platform[J]. Information & Communications, 2014(2): 266-267.)
- [67] 汪强兵, 章成志. 融合内容与用户手势行为的用户画像构建系统设计与实现[J]. 数据分析与知识发现, 2017, 1(2): 80-86. (Wang Qiangbing, Zhang Chengzhi. Constructing Users Profiles with Content and Gesture Behaviors[J]. Data Analysis and Knowledge Discovery, 2017, 1(2): 80-86.)
- [68] Wang G, Zhang X, Tang S, et al. Unsupervised Clickstream Clustering for User Behavior Analysis[C]// Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 2016: 225-236.
- [69] Wang G, Zhang X, Tang S, et al. Clickstream User Behavior Models[J]. ACM Transactions on the Web(TWEB), 2017, 11(4): 1-37.
- [70] Zhang X, Brown H F, Shankar A. Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry[C]// Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 2016: 5350-5359.
- [71] Nasraoui O, Soliman M, Saka E, et al. A Web Usage Mining Framework for Mining Evolving User Profiles in Dynamic Web Sites[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(2): 202-215.
- [72] Iglesias J A, Angelov P, Ledezma A, et al. Creating Evolving User Behavior Profiles Automatically[J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(5): 854-867.
- [73] Ma H, Cao H, Yang Q, et al. A Habit Mining Approach for Discovering Similar Mobile Users[C]// Proceedings of the 21st International Conference on World Wide Web. 2012: 231-240.
- [74] Chen X, Pang J, Xue R. Constructing and Comparing User Mobility Profiles[J]. ACM Transactions on the Web (TWEB), 2014, 8(4): Article No.21.
- [75] Zhu H, Chen E, Xiong H, et al. Mining Mobile User Preferences for Personalized Context-Aware Recommendation [J]. ACM Transactions on Intelligent Systems & Technology, 2014, 5(4): 1-27.
- [76] Wang H, Zhai C, Liang F, et al. User Modeling in Search Logs via a Nonparametric Bayesian Approach[C]// Proceedings of the 7th ACM International Conference on Web Search and Data Mining. ACM, 2014: 203-212.
- [77] 段建勇, 魏晓亮, 张梅, 等. 基于网络日志的用户兴趣模型构建[J]. 情报科学, 2013, 31(9): 78-82. (Duan Jianyong, Wei Xiaoliang, Zhang Mei, et al. Web Query Log Based User Interest Model[J]. Information Science, 2013, 31(9): 78-82.)
- [78] 黄文彬, 徐山川, 吴家辉, 等. 移动用户画像构建研究[J]. 现代情报, 2016, 36(10): 54-61. (Huang Wenbin, Xu Shanchuan, Wu Jiahui, et al. The Profile Construction of the Mobile User[J]. Journal of Modern Information, 2016, 36(10): 54-61.)
- [79] Rosenthal S, McKeown K. Age Prediction in Blogs: A Study of Style, Content, and Online Behavior in Pre- and Post-Social Media Generations[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. 2011: 763-772.
- [80] Mueller J, Stumme G. Gender Inference Using Statistical Name Characteristics in Twitter[C]// Proceedings of the the 3rd Multidisciplinary International Social Networks Conference on Social Informatics. ACM, 2016: 1-8.
- [81] Marquardt J, Farnadi G, Vasudevan G, et al. Age and Gender Identification in Social Media[C]// Proceedings of CLEF 2014 Evaluation Labs. 2014: 1129-1136.
- [82] An J, Kwak H, Jansen B J. Automatic Generation of Personas Using YouTube Social Media Data[C]// Proceedings of the 50th International Conference on System Sciences. 2017.
- [83] Jung S G, An J, Kwak H, et al. Persona Generation from Aggregated Social Media Data[C]// Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in

- Computing Systems. ACM, 2017: 1748-1755.
- [84] 王凌霄, 沈卓, 李艳. 社会化问答社区用户画像构建[J]. 情报理论与实践, 2018, 41(1): 129-134. (Wang Lingxiao, Shen Zhuo, Li Yan. User Profiling of Socail Q&A Community[J]. Information Studies: Theory & Application, 2018, 41(1): 129-134.)
- [85] 费鹏, 林鸿飞, 杨亮, 等. 一种用于构建用户画像的多视角融合框架[J]. 计算机科学, 2018, 45(1): 179-182, 204. (Fei Peng, Lin Hongfei, Yang Liang, et al. Multi-view Ensemble Framework for Constructing User Profile[J]. Computer Science, 2018, 45(1): 179-182, 204.)
- [86] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 785-794.
- [87] 李恒超, 林鸿飞, 杨亮, 等. 一种用于构建用户画像的二级融合算法框架[J]. 计算机科学, 2018, 45(1): 157-161. (Li Hengchao, Lin Hongfei, Yang Liang, et al. Two-level Stacking Algorithm Framework for Building User Portrait[J]. Computer Science, 2018, 45(1): 157-161.)
- [88] 李琳, 刘锦行, 孟祥福, 等. 融合评分矩阵与评论文本的商品推荐模型[J]. 计算机学报, 2018, 41(7): 1559-1573. (Li Lin, Liu Jinhang, Meng Xiangfu, et al. Recommendation Models by Exploiting Rating Matrix and Review Text[J]. Chinese Journal of Computers, 2018, 41(7): 1559-1573.)
- [89] 叶舒雁, 张未展, 齐天亮, 等. 一种基于传感器与用户行为数据分析的移动学习场景感知分类方法[J]. 计算机研究与发展, 2016, 53(12): 2721-2728. (Ye Shuyan, Zhang Weizhan, Qi Tianliang, et al. A Sensor and User Behavior Data Analysis Based Method of Mobile Learning Situation Perception[J]. Journal of Computer Research and Development, 2016, 53(12): 2721-2728.)
- [90] 尚燕敏, 曹亚男, 韩毅, 等. 基于主题和大众影响的用户动态行为倾向预测[J]. 计算机学报, 2018, 41(7): 1434-1447. (Shang Yanmin, Cao Ya'nan, Han Yi, et al. Recommending the Right Items for User Temporal Interest with Matrix Factorization Through Topic Model[J]. Chinese Journal of Computers, 2018, 41(7): 1434-1447.)

### 利益冲突声明:

作者声明不存在利益冲突关系。

收稿日期: 2018-07-18  
收修改稿日期: 2018-09-23

## A Survey of User Profiles Methods

Gao Guangshang

(Business School, Guilin University of Technology, Guilin 541004, China)

**Abstract:** [Objective] This paper discusses the mechanism of User Profiles construction process from the perspectives of design thinking and data types. [Coverage] We used Google Scholar and CNKI to search literatures with the keywords “User Personas” and “User Profiles”. Then we selected 90 representative literatures on User Personas in conjunction with topic screening, intensive reading and retrospective method. [Methods] Firstly, this paper studies the construction process of User Profiles from the perspective of design thinking, specifically combining the four perspectives of Goal-Directed, Role-Based, Engagement-Based and Fiction-Based. Second, it analyzes construction process of User Profiles from the perspective of data types, specifically combining Ontology or Concept, Subject or Topic, Interest or Preference, Behavior or Log, Multidimension or Fusion. Next, the construction methods are compared in detail from three aspects: logical ideas, performance characteristics and limitations. Finally, the next step for research on User Profiles is prospected. [Results] User Profiles technology plays a vital role in many areas such as online public opinion governance, advertising marketing and personalized services. [Limitations] There is no in-depth analysis of the evaluation indicators of User Profiles algorithms. [Conclusions] Although the existing methods of User Profiles can meet the needs of many applications to a certain extent, in the era of big data, it still faces the challenges of data sparsity, scene intelligence perception and user interest migration.

**Keywords:** User Profiles Ontology Topic Interest Behavior Log