

● 单晓红, 张晓月, 刘晓燕(北京工业大学经济与管理学院, 北京 100124)

基于在线评论的用户画像研究^{*} ——以携程酒店为例

摘要: [目的/意义] 用户画像描述产品用户群体行为特征, 为商家提供多方位的用户信息, 使商家了解、认知用户, 改进产品或服务, 为精准营销带来可能。[方法/过程] 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构建用户画像模型的概念模型, 并采用 Protégé 工具建立本体来实现用户画像属性之间的关联, 完成对酒店用户特征的完整刻画。[结果/结论] 以携程网北京地区酒店的在线评论为例, 从用户信息、酒店、评论 3 个维度进行可视化, 构建用户画像, 为酒店了解用户需求, 开展精准营销提供决策依据。

关键词: 用户画像; 本体; 在线评论; 案例分析

Research on User Portrait Based on Online Review: Taking Ctrip Hotel as an Example

Abstract: [Purpose/significance] User portrait describes the behavioral characteristics of the user group of a specific product, and provides merchants with multi-faceted user information, so that merchants can understand their customers, improve products or services, and bring possibilities for precision marketing. [Method/process] Based on the online review, the conceptual model of user portrait is constructed from three dimensions: user information attribute, hotel information attribute and user evaluation information attribute. Then the tool of Protégé is used to establish the ontology to realize the association between user portrait attributes, and accomplish the complete description of the hotel users' characteristics. [Result/conclusion] Taking the online reviews of Ctrip hotels in Beijing as an example, the dimensions of user, hotel, and comment information are visualized to build user portrait. The result provides basis for merchants to understand customers' needs and implement the strategy of precision marketing.

Keywords: user portrait; ontology; online review; case analysis

用户画像, 是建立在一系列数据之上的目标用户模型。通过用户画像可以从多个角度精细化地刻画用户特征, 帮助商家发现用户关注点, 改进产品或服务, 实现精准营销。已有研究往往使用用户基本信息、网页的浏览、收藏、购买、话题讨论以及手机移动端的上网时长、通信记录、套餐信息等数据构建用户画像。但是这些数据无法反映出用户对产品的评价和情感倾向, 不能完整地描述用户的特征, 为商家提供准确的决策依据。随着电子商务和网络的发展, 越来越多的用户愿意在网上购买产品或服务并发表自己的观点, 这些在线评论信息使得商家能够通过互联网便利地获取用户的反馈信息。因此结合在线评论进行用户画像的构建可以弥补传统用户画像信息不全面的缺陷。

本文基于在线评论, 从用户信息、酒店信息和用户评价信息 3 个维度构建用户画像模型, 以携程酒店为例, 采用本体技术实现用户画像的构建, 从多个角度刻画携程酒店用户特征, 不断提升市场营销、用户维系和服务工作的精准性, 为酒店改进及发展提供方向。

1 相关研究综述

用户画像是基于用户属性和用户行为提取用户的兴趣, 了解用户的过程。用户画像多用于内容推送^[1]、应用推荐^[2]、产品研发^[3]、广告投放^[4]、手机个人征信^[5]、移动个性化服务^[6]领域。

用户画像的数据来源包括静态数据和动态数据两个部分。静态数据是用户相对稳定的数据, 如用户的基础信息^[1,7], 如年龄、地域等。动态数据是随着用户行为不断变化的信息, 不同的应用领域动态数据也有所差别。在内容推送领域, Amazon 通过用户在网站上的浏览、购买和评分等行为进行用户画像。曾鸿等^[4]使用用户参与话题讨论的数据构建微博用户画像。在应用推荐领域, 张小可

* 本文为北京市自然科学基金项目“基于技术交易数据的京津冀科技协同创新潜力挖掘”(项目编号: 9172002) 和国家自然科学基金项目“基于多 Agent 制造业供应链产销协同冲突协商研究”(项目编号: 71371018) 的成果。

等^[2]通过移动互联网用户对手机应用的登录次数、逗留时间计算兴趣度，得到不同的应用兴趣集。吴明礼等^[1]通过移动端收集用户使用手机的浏览信息、点击 URL 信息、使用手机时间点、用户位置信息、用户经常使用的 APP 信息等动态信息，对用户精细化画像，能够针对移动终端用户情景的变化自动触发相应的推送任务。在移动个性化服务领域，张慷^[8]基于用户基础信息、终端信息、业务订购等信息，形成手机用户画像。丁伟等^[5]和黄文彬等^[6]用手机用户的主被叫通信记录作为社交网络数据、网站浏览行为作为行为偏好数据描述用户特征。

构建用户画像多采用统计、贝叶斯网络、神经网络、主题模型、聚类分析等方法。G. Xu 等^[9]提出通过数据统计的方法对用户的浏览记录和点击情况进行分析。王庆福^[10]和张小可等^[2]采用贝叶斯网络构建用户兴趣模型，从而刻画用户画像。辛菊琴等^[11]提出了基于产品特征建立用户画像，然后采用神经网络训练得到具体的用户画像模型。基于主题的用户画像，大大减少了用户画像的存储空间，描述了用户特征的多样性，并陆续出现了 ATM^[12]，ARTM^[13]，AITM^[14]，LITM^[15]，Topic-link LDA^[16]等多种改进方法。张慷^[8]通过提取用户的上网行为特征，结合相关数据进行数据融合，构建通信用户的画像模型。黄文彬等^[6,17]在分析移动用户的基站轨迹基础上，采用频繁模式挖掘、构建概率矩阵等方法，构建包含地理位置信息的用户画像模型。李冰等^[18]提出基于 K-means 算法的卷烟零售用户特征画像。

综上所述，用户画像的研究为刻画用户特征，为企业提供精准营销决策提供了基础。但是现有的用户画像缺乏层次性和完整性，没有反映出用户对产品的评价。在线评论作为文本形式为主的用户产品评价，是用户购买产品或服务后发表的对产品或服务的体验，体现了用户对产品或服务的情感和认同度，并且在线评论获取不涉及用户隐私，因此基于在线评论构建的用户画像能够全面刻画用户特征。本体是一种形式化的，对共享概念体系明确而又详细的说明，通过本体实现用户画像能够表现用户特征的概念层次结构。因此，本文提出以网络在线评论为数据源构建用户画像的方法，通过对评论进行处理，抽取用户关注的产品特征并计算情感极性，借助本体的方法实现用户画像并进行分析。

2 用户画像模型构建流程

基于在线评论的用户画像模型的构建主要包括用户画像概念模型、基于本体的用户画像模型实现、用户画像多维可视化分析 3 个步骤，如图 1 所示。

1) 用户画像概念模型。从用户信息属性、酒店信息

属性和用户评价信息属性 3 个维度构建用户画像，确定各个维度包含的信息。

2) 基于本体的用户画像模型实现。主要包括：数据源获取、酒店用户本体构建和导入 Protégé 实现本体模型。

3) 用户画像多维可视化分析。利用 OntoGraf 插件实现可视化并进行分析。

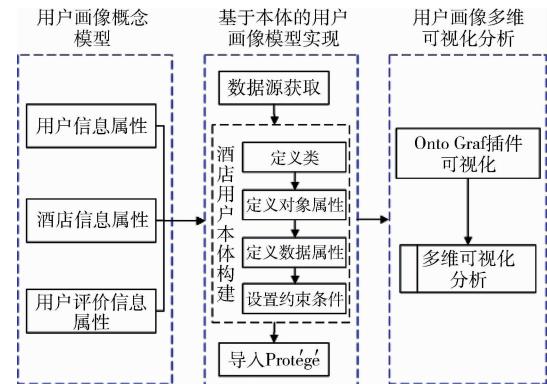


图 1 基于在线评论的用户画像模型构建流程

3 用户画像概念模型

3.1 用户信息属性

用户信息属性描述用户基本特征，反映了酒店用户入住的原因，包括用户名称和用户出行方式。用户名称是区别于其他用户的最基本的属性；用户出行方式包括商务出差、朋友出游、家庭亲子、情侣出游和其他，是刻画用户群体行为的特征属性。

3.2 酒店信息属性

酒店信息属性是酒店的基本特征，反映了用户对酒店的偏好，包括酒店名称、区域、详细位置、房间类型和价格。酒店区域描述了酒店所处区，详细位置是酒店的具体位置，房间类型包括大床、标准间等，价格是用户入住房间的价格。这些信息可以帮助商家分析用户选择酒店的动机，有针对性地开展市场营销活动。

3.3 用户评价信息属性

用户评价信息表达了用户对酒店的观点，包括评论内容、评论日期、点赞数和总评分。评论内容蕴含了用户关注的酒店特征及相应特征的态度；总评分是用户对酒店的整体评价；其他酒店用户通过点赞来表示对评论的认同度；评论日期可以反映评论发生的时效性。用户评价信息通过情感分析的方法计算用户关注的不同特征的情感极性，以正向评价 1 分、负向评价 -1 分，乘以程度副词和否定词的权重来表示。

4 基于本体的用户画像模型实现

本体作为一种知识重用、共享和建模的技术，在智能

信息处理、知识工程领域已经得到广泛的应用。如郑建兴等^[19]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；由于本体的结构化和语义性可以弥补社会化标签的不足，唐晓波等^[20]提出基于本体和标签的个性化推荐模型；文献[21-23]以在线评论为基础构建了情感本体，对产品整体以及部件的情感倾向进行了分析。虽然本体在用户画像领域的研究不多，但是由于用户画像的属性概念具有多样性、层次性和领域性的特点，适合本体实现，因此本文选择本体作为用户画像实现的技术。

以用户画像概念模型为基础，通过定义类、对象属性、数据属性及设置约束条件构建酒店用户画像本体。选取 Protégé 作为本体构建工具，OWL 作为描述语言，实现基于本体的用户画像实例化。基于本体的用户画像实现主要包括：数据源获取、酒店用户画像本体的构建、Protégé 中本体模型的实现。

4.1 数据源

本文数据来源于携程网站的酒店基本信息和在线评论，借助八爪鱼采集携程网站的酒店基本信息（包括酒店的名称、区、详细位置、房间类型、价格）、用户基本信息（包括用户名称和出行方式）、在线评论（包括评论内容、评论日期、点赞数、总评分）。

4.2 酒店用户本体的构建

本体构建和实现需确定类、对象属性、数据属性及约束条件的设置。本文设置评论类、“特征观点对”类、日期类、用户信息类、酒店信息类、酒店位置类、价格类、房间类型类 8 个类，具体设置如下：

1) 评论类 (comment)。评论类实现用户评价信息属性，包括评论内容 (content)、点赞数 (point_number) 和总评分 (total_number)。

2) “特征观点对”类 (character_sentiwords)。特征观点对是从评论中抽取的、用户关注的产品特征及相应情感极性的精炼信息，是刻画用户画像的关键。“特征观点对”类是评论类的子类，实现用户评价信息属性。

3) 日期类 (date)。用户发表评论的日期，日期类是评论类的子类，实现用户评价信息属性。

4) 用户信息类 (userinfo)。用户信息类包括用户名称 (user_ID) 和用户出行方式，实现用户信息属性。

5) 酒店信息类 (hotelinfo)。酒店信息类包括酒店名称和酒店详细位置 (hotel_location)，实现酒店信息属性。

6) 酒店位置类 (hotel_location)。酒店位置类表示酒店位于哪个区，是酒店信息类的子类，实现酒店信息属性。

7) 价格类 (price)。价格类是入住酒店相应房间的

价格，是酒店信息类的子类，实现酒店信息属性。

8) 房间类型类 (room_type)。房间类型类是用户入住的房间类型，是酒店信息类的子类，实现酒店信息属性。

设置评论内容 (content)、点赞数 (point_number)、总评分 (total_number)、用户名 (user_ID)、详细位置 (location) 为数据属性。在构建的 8 个类中，评论类作为主体类存在，其他类通过对象属性与评论类进行联系构成用户画像模型。类的约束条件主要是指类之间的互斥关系，具体设置如：comment 类下的“disjoint with”项添加其他 7 个类，表示 comment 类与其他类之间是没有交集的。对象属性的约束条件设置是通过对对象属性的定义域和值域的限定完成，如对象属性 come 的定义域为 comment 类，值域为 userinfo 类。数据属性的约束也是对其定义域及值域的限制，如：数据属性 content 的定义域为 comment 类、值域为 string 类型的限制。

4.3 “特征观点对”类的构建

“特征观点对”是从用户发表的评论中提取出来，可以直观地了解用户对每个产品特征的情感倾向。流程如图 2 所示。

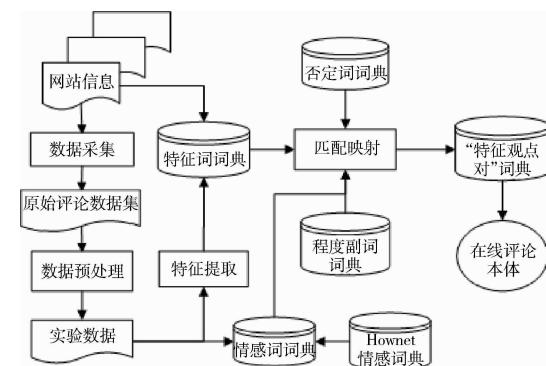


图 2 “特征观点对”类的构建流程

1) 特征词词典的构建。首先，利用 python 中的结巴分词对原始评论数据进行预处理。从酒店简介信息中人工抽取特征形成第一部分候选特征词，借鉴文献 [24]，以名词 (/n) 和动词 (/v) 作为候选特征，利用 textrank 算法得到最重要词语作为第二部分候选特征。其次，根据余弦相似度计算候选特征同义词，并在实验数据中进行替换，构成特征词词典候选特征。最后，结合文献 [5] 中的方法确定特征词的下位词，实现候选特征词的上下位划分，得到最终的特征词词典。

2) 情感词典构建。以 Hownet 情感词典为基础，结合产品评论中的形容词构建情感词典。情感词典包括正向通用情感词、负向通用情感词、正向专用情感词、负向专用情感词。通用情感词可以表达多个特征的情感，例如

“好、很好”既可以描述“房间”，也可以描述“wifi”。专用情感词专门用来描述产品的某个特征，例如“干净”用来描述“卫生”这一特征。利用评论的语料提取和标注观点词，计算基准观点词的情感极性，通过语义相似度的方法计算其他观点词的情感极性，构成情感词典。

3) 否定词词典和程度副词词典构建。用户在对产品评论过程中常使用程度副词来强烈表现自己的情感倾向，考虑用户这一特性，将程度副词分成了4个等级：低、中、高和极高，分别赋予0.5, 1, 1.5和2的权值，构成程度副词词典。汉语中还有一些否定词，这些否定词和情感词搭配使用时，词的情感会发生逆转，本文将包括“没有、不、无”等13个常用否定词构成否定词词典。

4) 特征观点对匹配。特征观点对匹配是指从大量的评论数据中提取出描述产品特征的特征观点对，从而建立产品特征与用户情感倾向之间的联系。

本文将评论按标点符号进行切分，Python编程提取每一个句子的特征观点对。抽取结果表示为($C, N, D, S, p(C, N, D)$)，其中 C 表示特征词， N 表示否定词， D 表示程度副词， S 表示情感词， p 表示情感极性得分。借鉴文献[21]，配对模式为：①特征词与观点词为一对一关系。②特征词与观点词为多对一关系。③特征词与观点词为一对多关系。④隐式产品特征识别。中文在线评论存在主语省略的现象，但可根据上下文推断出隐含的特征词语，这些特征被称为隐式特征，根据专用情感词自动匹配隐式特征。

4.4 用户画像模型的实现

在本体构建的基础上，采用Protégé和OWL的人工与半自动化结合的方式来实现本体实例化。通过定义类、属性、关系构建本体，从用户信息属性、酒店信息属性、用户评价信息属性3个维度实现酒店用户画像。

5 携程酒店用户

画像实证分析

基于前文提出的用户画像模型，根据北京八大城区酒店数量比例，从携程采集北京市八大城区朝阳区、东城区、西城区、丰台区、海淀区、石景山区、大兴区、房山区的酒店基本信息、用户信息及2014年7月到2017年5月11日的16319条在线评论。经去重等操作去除无用评论后得到有效数据10579条，以此

为数据源构建酒店用户画像，分别从酒店整体、用户评价、酒店价格、酒店品牌对酒店用户进行画像分析。

5.1 酒店用户整体画像呈现

通过用户画像可以从整体上刻画携程酒店用户特征。以朝阳区、丰台区、西城区3个区的74个酒店为例对北京酒店进行用户画像。朝阳区有41家酒店、441条评论，丰台区有21家酒店、115条评论，西城区有12家酒店、65条评论。

整体用户画像为：用户出行方式以商务出差为主。用户入住的房间类型以大床房居多，大床房细分为商务、豪华、自主、标准、榻榻米等。朝阳区酒店的价格大多位于200~300元之间，丰台区酒店的最低价格为162元，西城区价格相对偏高，基本在300元以上，其中纽斯特酒店达到833元/晚，说明价格与地理位置、酒店类型有关。用户对酒店的评价涉及整体、房间、环境、服务和交通几大类，包括餐饮、价格、卫生等细节方面。其中丰台区多条评论提及房间，如房间_精致_1.0等。朝阳区提及景点和机场，西城区多条评论涉及交通、位置、wifi，说明用户选择酒店与交通和出行目的相关。因此，酒店可以通过用户画像制定精准营销方案和酒店发展方案。如针对西城区酒店和丰台区酒店，丰台区在营销时可以以价格作为优势，吸引如以出行方式为家庭亲子、朋友出行等特定人群。

为了更加清晰详细地通过酒店在线评论分析用户画像，对用户评价特征、同一品牌酒店特征、同等价位酒店分别进行画像分析。

5.2 用户评价特征画像

在呈现出的用户画像中，通过精确匹配检索方式检索某一酒店某一用户的精细画像，检索结果为图3。

如图3所示，北京7天连锁酒店（北京燕莎美国大使

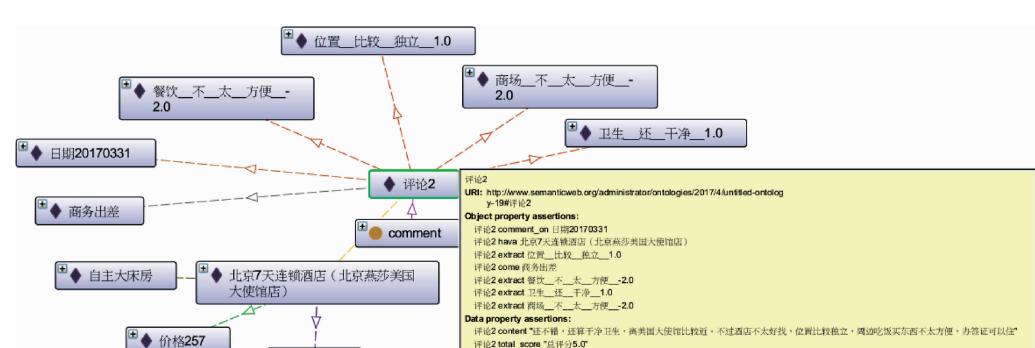


图3 用户评价特征画像

馆店）通过“评论2”得到的酒店用户画像为：酒店位于朝阳区，房间类型为“自主大床房”，价格为257元，用户的出行方式为“商务出差”，评论的日期为2017年3月

31 日，酒店的评价可以看出，该用户对酒店的位置及卫生还算满意，但餐饮及商场因为不方便的原因给出了消极评价。另外，可以通过数据属性看到评论的内容，总评分为 5.0，点赞数为 0。因此该酒店可以在营销中宣传其位置及卫生，并对餐饮方面进行改善。

通过 extract 属性展开得到涉及“位置”评价的具体画像，如图 4 所示。

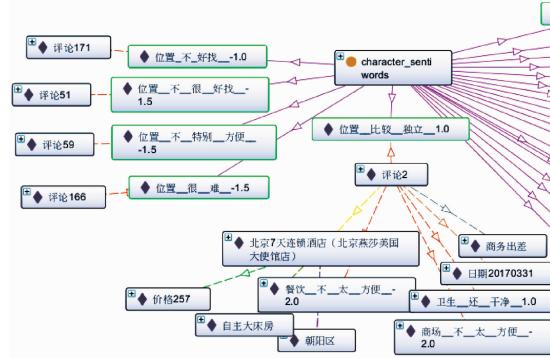


图 4 酒店用户位置画像

用户在发表的评论中有 40 条评论提及位置，表现出用户对位置的在乎程度，其中 4 条评论对位置持消极态度，原因是“不好找”或“不方便”，36 条评论中 7 条表示对位置很满意。同理，可将评论分别按照 have 属性进行展开，可以准确地看到哪个酒店的哪条评论涉及位置，在营销中有位置优势的酒店要强调地理位置，位置处于劣势的酒店可尽可能少提及位置。

5.3 酒店特征画像

通过正则表达式检索方式检索显示朝阳区有 3 家桔子酒店，将 3 家酒店分别按 sale、room_type、have、come 属性展开，得到朝阳区的桔子酒店的用户画像，结果如图 5 所示。

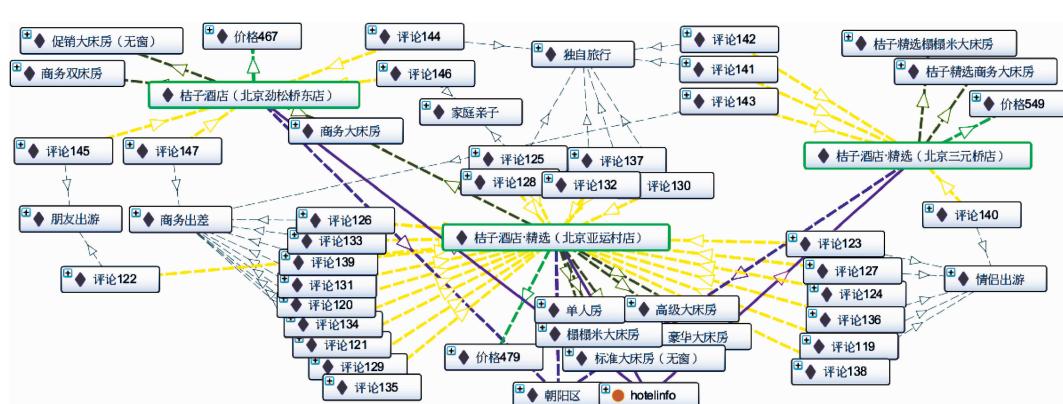


图 5 酒店特征画像

从图 5 可以看出，朝阳区 3 家桔子酒店（北京劲松桥东店、北京三元桥店和北京亚运村店）的用户画像各具

特色，房间价格、房间类型、用户出行方式、评论数量都不相同，桔子酒店·精选（北京亚运村店）拥有的评论最多，并且均以商务出差、情侣出游和家庭亲子类型的用户居多，说明北京亚运村店相对来说能更好地满足用户需求。因此本文对这家酒店的评论进行进一步分析。

图 6 是朝阳区桔子酒店·精选（北京亚运村店）关于用户评价信息属性中特征观点对的精细画像结果。通过

整理分析可以发现，用户对酒店的评价信息涉及服务、交通、餐饮、房间、价格、设施等方面，涉及最多的是服务、交通和设施，说明这三个是用户最关注的需求。用户对

该酒店的各方面评价基本呈现积极的态度，个别消极评价如“餐饮_没_-1.0”“房间_很_味_-1.5”“房间_不大_-1.0”“价格_很_高_-1.5”。表明酒店未能完全满足用户需求，在房间以及价格方面仍有需要改进的地方。

5.4 同等价位酒店比较

价格是用户选择酒店最关注的特征之一。根据正则表达式检索价格在 200 ~ 300 的酒店，共有 18 家酒店的价格在 200 ~ 300 之间，其中价格是 219 元的有 3 家酒店，分别是北京 7 天连锁酒店（北京宋家庄地铁站店）、北京 7 天连锁酒店（北京青年路地铁站店大悦城店）、北京 7 天连锁酒店（北京方庄桥店），对这 3 家酒店进行进一步的分析。

3 家酒店的用户画像为：北京 7 天连锁酒店（北京宋家庄地铁站店）位于丰台区，用户类型为“家庭亲子”“商务出差”和“其他”，2/3 的评论提到离地铁远。北京 7 天连锁酒店（北京青年路地铁站店大悦城店）位于朝阳区，用户类型为“朋友出游”和“商务出差”，用户对酒店的床、房间、设施

特色，房间价格、房间类型、用户出行方式、评论数量都不相同，桔子酒店·精选（北京亚运村店）拥有的评论最多，并且均以商务出差、情侣出游和家庭亲子类型的用户居多，说明北京亚运村店相对来说能更好地满足用户需求。因此本文对这家酒店的评论进行进一步分析。

具有负面情感，酒店应当重视用户的这种情绪，并及时进行改进。北京 7 天连锁酒店（北京方庄桥店）位于丰台区，用户类型为“家庭亲子”和“商务出差”，3/8 的用户表示房间小或破旧问题，提及隔音的

1/4 的用户评论都表明隔音效果不好，以及提及的床单及毛巾的问题可以看出酒店的设施有待改善；从对餐饮、景点、位置等评论可以看出该酒店的地理位置不错。综上，北京方庄桥店的用户多的原因可能是地理位置的原因，表明用户对地址位置需求比较重视，所以在价格成本相同情况下，选址应是酒店考虑的最重要因素之一。

综上所述，通过对用户评价特征、同一品牌酒店特征、同等价位酒店分别进行画像分析，更准确地描述了酒店用户特征，从而帮助酒店实现精准营销。还可以从用户出行方式、同一地区不同品牌酒店、不同地区同一品牌酒店进行画像分析，由于篇幅原因，在此不做进一步解释。

6 结束语

本文以携程酒店在线评论为基础，从用户信息属性、酒店信息属性和用户评价信息属性 3 个维度完整全面地刻画酒店用户的特征，并借助本体实现了用户画像的层次化、概念化和可视化，充分展示了多元主体间的联系。其中用户评价属性通过特征观点对表达出的用户感兴趣的产品特征以及相应的情感极性，实现用户对酒店特征和情感倾向的分析。

本文提出的基于在线评论的携程酒店用户画像在理论上和实际上都具有一定的价值和意义。通过用户画像可以从在线评论中找到用户关注的产品特征以及用户对特征的态度，为商家提供全方位的用户信息，剖析核心诉求，使产品的服务对象更加聚焦。同时拓展了本体在精准营销、个性化推荐、用户分析等领域的应用。本文方法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，可以通过提高评论的有用性提升用户画像质量。□

参考文献

- [1] 吴明礼, 等. 用户画像在内容推送中的研究与应用 [J]. 电脑知识与技术: 学术交流, 2016, 12 (32): 255-259.

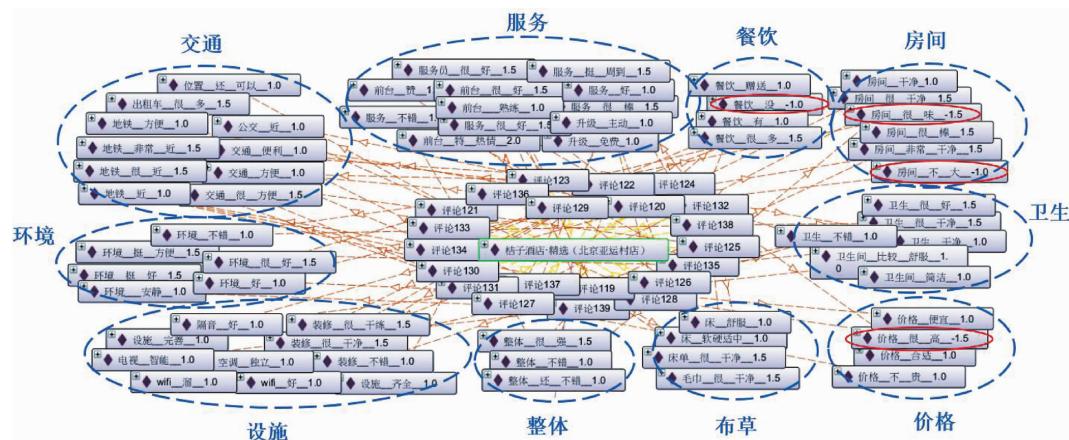


图 6 桔子酒店 · 精选 (北京亚运村店) 用户画像

- [2] 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.
- [3] 余孟杰. 产品研发中用户画像的数据模建——从具象到抽象 [J]. 设计艺术研究, 2014 (6): 60-64.
- [4] 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.
- [5] 丁伟, 王题, 刘新海, 等. 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.
- [6] 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.
- [7] BAKSHY E, ECKLES D, YAN R, ROSENN I. Social influence insocial advertising: evidence from field experiments [C] // Proc. of the 13th ACM Conference on Electronic Commerce. ACM, 2012: 146-161.
- [8] 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.
- [9] XU G, ZHANG Y, ZHOU X. Towards user profiling for Web recommendation [J]. Lecture Notes in Computer Science, 2005, lncs 3809: 415-424.
- [10] 王庆福. 贝叶斯网络在用户兴趣模型构建中的研究 [J]. 无线互联科技, 2016 (12): 101-102.
- [11] 辛菊琴, 蒋艳, 舒少龙. 综合用户偏好模型和 BP 神经网络的个性化推荐 [J]. 计算机工程与应用, 2013, 49 (2): 57-60.
- [12] MCCALLUM A, CORRADA-EMMANUEL A, WANG X. The Author-Recipient-Topic Model for topic and role discovery in social networks: experiments with enron and academic email [J]. Emmanuel, 2004 (6).
- [13] MIMNO D, MCCALLUM A. Expertise modeling for matching papers with reviewers [C] // ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, California, USA, August. DBLP, 2007: 500-509.
- [14] KAWAMAE N. Author interest topic model [C] // Proceeding of the, International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, Geneva, Switzerland, July. DBLP, 2010: 887-888.

具有相同的评价效力；年度 h 指数、累积 h 指数与 5 年影响因子的相关系数的引证高峰分别在 6 年和 5 年。需要指出的是，本研究结果仅是针对信息科学与图书馆学期刊，对其他学科是否也表现出相同的特性有待进一步研究；另外，由于入选研究对象的限制，本研究仅计算到了 10 年的引证时间窗口，更长时间引证时间窗口的 h 指数对期刊的评价效力如何仍有待研究。

考虑 h 指数的引证时间窗口问题，对于 h 指数在学者、机构等评价中的应用也具有一定意义，如可以对同一时段内被评价者的学术水平进行分析，以便于纵向或横向对比等。总之，对 h 指数的多方位研究，有利于挖掘其使用价值，为完善学术评价体系提供新思路。□

参考文献

- [1] HIRSCH J E. An index to quantify an individual's scientific research output [J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102 (46): 16569-16572.
- [2] BRAUN T, GLÄNZEL W, ACHUBERT A. A hirsch-type index for journals [J]. Scientometrics, 2006 (1): 169-173.
- [3] XU F, LIU WB, MINGERS J. New journal classification methods based on the global h-index [J]. Information Processing & Management, 2015, 51 (2): 50-61.
- [4] CHRISOVALANTIS M. Some variations on the standard theoretical models for the h-index: a comparative analysis [J]. Journal of the Association for Information Science and Technology, 2015, 66 (11): 2384-2388.
- [5] BERTOLI-BARSOTTI L, LANDO T. A theoretical model of the relationship between the h-index and other simple citation indicators [J]. Scientometrics, 2017: 1-34.
- [6] ROUSSEAU R. A case study: evolution of JASIS' Hirsch index [J]. Science Focus, 2006, 1 (1): 16-17.
- [7] 俞立平, 方丹丹. 一个新的文献计量学存量指标——历史 h 指数 [J]. 情报科学, 2016, 34 (4): 108-111.
- [8] 美捷登. 谷歌学术 H5 指数及其与汤森路透影响因子对比分析 [EB/OL]. (2015-08-05) [2017-07-15]. <http://blog.sciencenet.cn/blog-475824-910701.html>.
- [9] HIRSCH J E, BUELA-CASAL G. The meaning of the h-index [J]. International Journal of Clinical and Health Psychology, 2014, 14 (2): 161-164.
- [10] SCHREIBER M. Restricting the h-index to a publication and citation time window: a case study of a timed Hirsch index [J]. Journal of Informetrics, 2015, 9 (1): 150-155.
- [11] 刘雪立, 周晶, 盖双双, 等. 基于被引频次分布特征的影响因子缺陷矫正——位置指标应用于期刊评价的实证研究 [J]. 图书情报工作, 2016 (9): 99-105, 115.
- [12] 盛丽娜, 顾欢. “影响因子百分位”与 h 指数、累积 h 指数对期刊的评价效力分析 [J]. 中国科技期刊研究, 2017, 28 (2): 166-170.
- [13] 刘雪立. 期刊累积影响因子与年度影响因子的比较研究 [J]. 中国科技期刊研究, 2016, 27 (7): 774-778.
- [14] 王梅英, 周志新, 刘雪立. 2009 年 SCI、SSCI 收录图书情报学期刊及其文献计量学指标综合分析 [J]. 情报科学, 2011 (5): 727-734.
- [15] 盛丽娜, 顾欢. 基于文献类型矫正影响因子在信息科学与图书馆学期刊中的实证分析 [J]. 中国科技期刊研究, 2016, 27 (11): 1202-1207.
- [16] 刘雪立. 基于 Web of Science 数据库预测 SCI 期刊影响因子的方法 [J]. 科技与出版, 2014 (2): 87-91.
- [17] 俞立平. 学术期刊 h 指数的时间演变规律研究 [J]. 情报杂志, 2015, 34 (1): 96-99.
- [18] ROUSSEAU R, STIMULATE G. On the relation between the WoS impact factor, the Eigenfactor, the SCImago Journal Rank, the Article Influence Score and the journal h-index [EB/OL]. (2009-01-01) [2016-11-01]. https://www.researchgate.net/publication/28810385_On_the_relation_between_the_WoS_impact_factor_the_Eigenfactor_the_SCImago_Journal_Rank_the_Article_Influence_Score_and_the_journal_h-index.

作者简介: 顾欢 (ORCID: 0000-0001-5484-7694), 男, 硕士。盛丽娜 (ORCID: 0000-0003-1359-4074, 通讯作者), 女, 副编审。

录用日期: 2017-09-25

(上接第 104 页)

- [15] KAWAMAE N. Latent interest-topic model: finding the causal relationships behind dyadic data [C] //ACM International Conference on Information and Knowledge Management. ACM, 2010: 649-658.
 - [16] LIU Y, NICULESCU-MIZIL A, GRYC W. Topic-link LDA: joint models of topic and author community [C] //International Conference on Machine Learning. ACM, 2009: 665-672.
 - [17] 黄文彬, 吴家辉, 等. 数据驱动的移动用户行为研究框架与方法分析 [J]. 情报科学, 2016, 34 (7): 14-20.
 - [18] 李冰, 王悦, 刘永祥. 大数据环境下基于 K-means 的用户画像与智能推荐的应用 [J]. 现代计算机: 专业版, 2016 (24): 11-15.
 - [19] 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.
 - [20] 唐晓波, 钟林霞, 王中勤. 基于本体和标签的个性化推荐 [J]. 情报理论与实践, 2016 (12): 114-119.
 - [21] 郑丽娟, 等. 基于情感本体的在线评论情感极性及强度分析: 以手机为例 [J]. 管理工程学报, 2017, 31 (2).
 - [22] 杜嘉忠, 徐健, 刘颖. 网络商品评论的特征——情感词本体构建与情感分析方法研究 [J]. 现代图书情报技术, 2014, 30 (5): 74-82.
 - [23] 唐晓波, 等. 基于情感本体和 KNN 算法的在线评论情感分类研究 [J]. 情报理论与实践, 2016 (6): 110-114.
 - [24] 尹裴, 王洪伟. 面向产品特征的中文在线评论情感分类: 以本体建模为方法 [J]. 系统管理学报, 2016, 25 (1): 103-114.
- 作者简介:** 单晓红 (ORCID: 0000-0002-7535-1943, 通讯作者), 女, 1976 年生, 博士, 副教授。张晓月 (ORCID: 0000-0003-0617-3925), 女, 1993 年生, 硕士生。刘晓燕 (ORCID: 0000-0002-5835-6021), 女, 1974 年生, 博士, 副教授。
- 录用日期:** 2017-09-26