

引用格式:徐菲菲,刺利青, Ye Feng. 基于网络数据文本分析的目的地形象维度分异研究——以南京为例[J]. 资源科学, 2018, 40(7): 1483-1493. [Xu F F, La L Q, Ye F. A research on destination image and perceived dimension difference based on big data of tourists' comments: a case of Nanjing[J]. *Resources Science*, 2018, 40(7): 1483-1493.] DOI :10.18402/resci.2018.07.16

基于网络数据文本分析的目的地形象 维度分异研究 ——以南京为例

徐菲菲¹, 刺利青¹, Ye Feng²

(1. 东南大学人文学院旅游管理系, 南京 210096;

2. 美国普渡大学酒店与旅游管理系, 西拉法叶 47906)

摘要:随着 Web2.0 技术的发展, 网络文本已成为目的地形象研究的重要数据来源。本文选取南京作为案例地, 以马蜂窝游客对南京的 10 077 条在线评论文本为数据, 运用网络大数据文本挖掘法、词频统计法和共现网络法, 分析得出目的地形象维度重要性存在差异, 旅游吸引物是第一层次, 公共基础设施是第二层次, 休闲和娱乐、旅游环境和地方氛围是第三层次, 旅游基础设施是第四层次。其次, 南京感知形象中突出的认知成分源于旅游吸引物和公共基础设施维度的感知, 而情感形象为正面, 并产生了一定的意动形象, 形成了网络口碑效应。此外, 南京整体形象主要源于旅游吸引物和公共基础设施维度的感知, 进一步验证了感知维度的重要性存在差异。本文验证了旅游形象的相关理论, 丰富了旅游大数据研究方法, 并可为南京旅游形象建设和发展提供科学依据。

关键词:大数据; 目的地形象; 文本分析; 南京

DOI :10.18402/resci.2018.07.16

1 引言

Web2.0 技术的出现推动了以社会互动和旅游者之间信息交换为特征的旅游 2.0 现象的出现^[1]。现代旅游者在网络上对旅行体验进行评价和记述, 甚至出现了专门的旅游社交平台, 比如马蜂窝旅游网站为游客分享自己的出行游记及对某旅游产品的评价提供了一个专门性的平台。旅游者在网络上发表和更新的与旅游相关的内容被学者称为“游客生成内容”(Tourist-Generated Content), 以下简称为 TGC。这已经吸引了国内外学者们的广泛关注^[2, 3]。

TGC 由旅游者自发生成, 与目的地营销机构有目的提供的信息相比有更高的可信度, 同时能较为全面地反映游客对旅游目的地形象的感知。TGC 作为自发的用户反馈, 具有低成本、容易获取、广泛

使用的特征, 可以帮助研究者正确理解消费者的喜好、感知和需求信息, 可以弥补传统研究方法的不足, 帮助解释一些重要问题^[4]。另外, 研究表明 TGC 具有强烈的口碑效应, 并且能在不同时间不同地点的游客之间传播^[5]。TGC 的可信度和可获得性使其成为塑造目的地形象和影响游客决策的重要媒介^[6, 7]。

尽管 TGC 在影响游客行为和决策方面有显著的重要性, 但依然缺乏对这种新型媒介的了解。第一, TGC 的本质是对旅游体验的反映, 涵盖了食住行游购娱等旅游经历的方方面面^[7], 为研究者分析目的地形象提供了丰富的原真信息。目的地形象维度是进行形象测量和评价的主要途径。因此, 有必要从目的地形象维度出发分析 TGC 反映的目的地形象。第二, 和其他信息(如目的地营销组织和

收稿日期: 2017-11-15 修订日期: 2018-02-08

基金项目: 国家自然科学基金面上项目(41571133); 国家自然科学基金国际合作项目(41711530650); 中央高校基本科研业务经费(2242017S10005)。

作者简介: 徐菲菲, 女, 江苏南京人, 博士, 教授, 主要从事大数据与智慧旅游等方面研究。E-mail: 101011780@seu.edu.cn

通讯作者: 刺利青, E-mail: 2320510455@qq.com

问卷信息)不同,TGC为研究目的地的感知形象提供了可信度高的未经处理的丰富的信息^[8]。很多学者采纳 Gartner^[9]的观点,将目的地感知形象分为认知、情感、意动形象3个构成成分,并证明它们对目的地的选择有直接或间接的积极影响^[10-12]。因此,有必要研究TGC反映的目的地形象。

本文选取南京作为案例地,采用大数据文本分析方法,借助语义分析软件,对马蜂窝上游客关于南京的在线评论文本进行词频统计和语义网络分析,研究目的包括:①探索感知目的地维度的层次性;②分析游客评论反映的南京旅游目的地感知形象。希望以此促进对游客评论等网络大数据的研究,并对目的地形象管理和营销提供科学依据。

2 研究方法

2.1 目的地形象维度及成分

旅游目的地形象维度是形象测量和评价的主要途径,有许多学者从国家或城市等不同尺度上,用结构法或非结构法测量目的地形象^[7,13-15],提出了不同的形象维度及其属性;其中 Beerli 等的形象维度划分最具影响,即划分为自然资源、公共基础设施、旅游基础设施、旅游休闲和娱乐、文化/历史和艺术、政治和经济因素、自然环境、社会环境、地方氛围等9类^[16]。国内学者吴晋峰在此基础上将目的地感知形象维度划分为旅游吸引物、基础设施、休闲和娱乐、旅游环境和地方氛围等5类,其维度划分较为系统和完善^[17]。但因基础设施包括一般的基础设施和

主要为旅游服务的基础设施两类,其内部具有差异性,因此为区分这两类设施,本文在吴晋峰的研究基础上,将基础设施分为公共基础设施和旅游基础设施两个维度,形成旅游吸引物、公共基础设施、旅游基础设施、旅游休闲和娱乐、旅游环境和地方氛围6类维度。结合 Beerli 等的研究结果,对各维度的属性进行界定,见表1。目的地形象感知维度的重要性存在差异^[18],例如游客在参观自然景点时则会对自然吸引物维度给予更多的关注。在研究中学者在测量形象感知中会对测量维度赋予不同的权重,比如史坤博等对目的地形象感知维度赋予了不同的权重达到更好地测量大陆学生对台湾形象感知的目的^[19]。

形象反映了游客对他们记忆的目的地看法^[20]。基于属性的目的地形象的概念最初是由 Gartner 提出的,他认为目的地形象包含三个组成部分,即认知形象、情感形象和意动形象^[9]。目的地的认知评价包括旅游者对旅游目的地物质属性的信念和知识^[21],反映游客对感知的目的地属性的评价^[22],如某地气候宜人,某旅游地有独特的生活方式和地方习俗。情感形象成分表示游客对目的地的情绪反应或评价^[23],如高兴、愉快、令人兴奋等,而意动形象成分指的是游客积极考虑一个地方作为潜在的旅行目的地,描述游客想要进行的期望的未来的状态^[24,25],如想去、推荐等。然而,以往的文献经常忽略对意动形象的研究^[17,26-28],因为许多学者认为意动

表1 旅游目的地形象维度及属性

Table 1 Dimensions and attributes of tourist destination image

维度	属性
旅游吸引物	自然类: 地文景观、水域景观、生物景观、天象与气候景观 人文类: 博物馆、历史建筑物、节事活动、手工艺品、美食、宗教、民俗……
公共基础设施	道路、机场和港口、私人 and 公共交通、邮电和通信设施
旅游基础设施	宾馆等住宿设施、床位数量、餐饮设施、酒吧、俱乐部、游客服务中心、旅游厕所、旅游停车场、旅游标识系统……
旅游休闲和娱乐	主题公园、休闲活动、体育活动、冒险活动……
旅游环境	自然环境: 景观的吸引力、干净、空气污染、噪音污染、交通阻塞…… 社会环境: 当地居民的热情和友好程度、生活质量、语言阻碍…… 政治经济环境: 政治稳定、经济发展、安全、犯罪率、价格……
地方氛围	奢华、时尚、高知名度、充满异国情调、神秘、令人放松、令人紧张、有趣、令人高兴、令人厌倦、有吸引力……

2018年7月

形象与行为意图相同或类似^[29]。但是,最近也有一些学者表示意动形象和行为意图有不同之处^[30],在研究目的地感知形象方面有不可或缺的作用。Echtner等^[14]提出了“整体形象”的概念,认为整体形象是认知、情感、意动三个组成部分的总和,而一些学者的认识存在差异,认为它大于其各部分的总和^[22]。但正如Echtner所说的“整体形象是对某个特定的目的地界定的重要概念”。因此,将三个成分与整体形象相结合可以更加准确地界定目的地的感知形象。

2.2 样本网站和案例地选取

本文以马蜂窝(m.mafengwo.cn)作为游记样本选取网站,主要原因如下:

(1)马蜂窝是中国最大的游客分享网站,拥有众多用户,提供了海量的数据。马蜂窝自2006年上线运营以来,注册用户量持续攀高,截止2018年,用户量已超1亿¹⁾。依靠庞大的用户群体,马蜂窝拥有超2000万条真实点评,与南京直接相关的游记高达45 836条。

(2)与携程、同程等在线旅行社(OTA)相比,马蜂窝是一家专注于用户生成内容(UGC)的旅游社交分享网站。它更强调用户原创和分享内容,其评论具有真实性、主动性的特征,因而具有较高的研究价值,可以从了解到用户真实的感受。

南京是首批中国优秀旅游城市、国家历史文化名城,至2016年末,南京共有A级景区56家,其中4A级以上景区22家,接待海内外旅游者共11 206万人次,实现旅游总收入1909.26亿元,国际旅游创汇收入6.76亿美元^[31]。南京向来有“六朝古都”、“十朝都会”之称,在国内外享有很高声誉。由于南京是热门旅游目的地,互联网上存在大量与南京旅游相关的网络日志及网友点评,便于从中分析游客对南京旅游形象的感知,并进一步提升南京旅游形象,因此,本文选择南京作为研究对象。

2.3 网络数据研究方法

2.3.1 网络数据研究方法的优点

研究旅游目的地形象最为常用的数据获取方式是问卷调查。问卷法调查要求问卷的设计、发放都必须具有一定的科学性,问卷调查法虽然能对主题进行深入的研究,但对统计上概率较小的信息却

很少涉及。目前,文本内容分析的研究方法也引起了学者关注。随着互联网技术的迅猛发展和用户规模的爆发性增长,互联网已经步入“大数据”时代,网络数据提供了文本分析所需要的数据源。例如新浪微博^[32]、博客^[33]、在线评论^[34]等用户生成内容反映了游客对旅游目的地形象的感知,具有真实性、反馈及时性、内容丰富性的特点^[35],成为分析旅游目的地形象的重要研究资料来源。但之前的文本研究对于网络大数据样本的选取很大程度上受到了研究者主观因素的影响,选取的样本量较小,分析方法主要采用人工的语义分析;而文本挖掘技术实现了从海量的Web数据中抽取感兴趣的潜在的有用的和隐藏的信息,实现直接从网络上抓取信息,避免了主观因素的影响,从而发现一些问卷调查可能忽略的信息^[36]。此外,传统的问卷调查方式要求被调查者对李克特量表上的一组事先设定好的目的地属性评分,在了解游客喜好和旅行经历方面是一个较为死板的工具。相反,由于TGC本身是游客对旅行经历的未经处理的自发的丰富的表达和评价,从而它是收集旅游者反馈和进行质量控制的一种更为经济和有效的方式^[7]。

2.3.2 数据收集

本文收集了2013年1月—2015年12月马蜂窝用户所有关于南京的游记和点评,形成本地大数据库。首先利用网络爬虫工具抓取了在南京留下足迹的2332个评论用户ID,之后对这些用户关于南京的游记及评论进行收集。共抓取了10 248条旅游在线评论,经过合并重复ID以及剔除171条无效数据,剩余有效数据10 077条,总字数375 230字。数据收集时间从2016年3月开始截止至2016年6月底,所获得的用户评论发表时间为2013年1月至2015年12月。

2.3.3 数据处理及分析

本文使用KH coder内容挖掘软件进行数据处理。KH coder是较为普及的用于非结构化文本分析的开源软件,利用自然语言处理技术成功实现了对特定单词的精确检索与定位、词频统计、词间关系以及可视化等文本挖掘过程^[37],并广泛应用于国内外的文本分析研究中^[38, 39]。

1) 百度百科.词条马蜂窝. <https://baike.baidu.com/item/马蜂窝/22375486?fromtitle=马蜂窝&fromid=778717>.

(1) 采用 Stanford POS Tagger 进行数据预处理。Stanford POS Tagger 主要是对词性标记方法的 Java 实现^[40]。预处理过程包括句子分割、标记(即将一段文本分割为标记)、去除停用词、词性标注(例如,对名词、动词、形容词等具有相似语法性质的词的分类)和词干提取(即将不同形态变化的一个词组合在一起)。由于社会语言的变化发展,KH coder 分词工具可能存在无法准确提取词干的问题,本文对 KH coder 语料库进行了测试和更新,测试中发现南京的景点类词汇在初次测试中缺乏识别和提取,从而对语料库增加了“鸡鸣寺”、“玄武湖”“中山陵”等景点词汇。预处理后的游客评论数据由 324 123 个标记和 29 309 个字类型组成。

(2) 为了确定代表游客感知的主要词汇和主题,预处理过后的数据采用词汇分析、关联统计和数据可视化进行分析。

(3) 为了识别评论使用的重要的词汇,基于每个词汇出现的频率即词频(term frequency/TF),对高频词进行提取。经过分词工具处理之后的评论,剔除了情感倾向判断,比如将“不推荐”提取为“推荐”,本文通过人工检查筛除这种情况从而对这一误差进行了校正。按词频从高到低对各个词性下的词频进行统计,并将其可视化生成词频分布图。

(4) 为了检验高频词在文献中的分布情况(即它们如何在一个评论中相互联系),进行 R 统计程序下 igraph 包的词频共现网络的分析。网络的节点是高频词。网络的边是由这个词对的 Jaccard 系数决定^[41]。Jaccard 系数是一个比较有限样本集(即词)之间的相似性的统计指标,它被定义为样本集的交集除以合集。一个词对 A 和 B 的 Jaccard 系数是:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

网络的布局由 Fruchterman 等^[42]的算法实现,它使用基于力的图形绘制技术呈现网络。

3 研究内容

3.1 文本信息

本文通过文本挖掘方法收集并分析了马蜂窝网站(m.mafengwo.cn)上的游客评论。网站用户上传了个人资料,包括性别、居住地、职业、年龄、肖像等。其中,性别和居住地是用户必须上传的信息。

通过网络爬虫工具,本文抓取了 2013—2015 年共 2332 名评论用户的信息(一名用户可能发表了多篇游记或点评,因此抓取的评论数为 10 248 条)。评论用户包括 1167 名男性和 1165 名女性,男女比例基本相同,约为 50%。用户居住地涵盖中国 34 个省级行政区,北京上海等一线城市、江苏本省和浙江、广东等东南沿海省份的用户占比在 30% 以上。港澳台以及西北、西南等内陆省份的用户占比小于 2%。职业、年龄等信息只有极少数人进行了公布。大多数游记文本描述旅行者的一次旅行经历,也有一些例外。比如:一个用户上传了 3 篇游记描述了他南京的一次旅行经历。此外,游记还包括旅行的出发时间、出行天数、出行人物、人均费用以及上传时间等详细信息。

将用户评论整理合并为一个 TXT 文本,剔除无意义的词,如“这儿”“那儿”“第二”等词语,并对意思相同表达不同的词语进行合并,如“秦淮和秦淮河”,“景区和风景区”,“中山陵和中山陵园”,对文本进行词频统计。图 1 表示了评论文本的词频(TF-Term Frequency)分布。游客评论中词频的平均值是 7.08(每个词汇平均出现 7 次)。词频的分布符合长尾理论,出现次数少于 20 次的词汇约占总词汇的 95%,而高频关键词只占有所有词汇的 5%,包含有限的词汇类型。这表明有限的高频词能够有效地反映游客评论中的重要信息,对高频词的分析是本文进行大数据文本分析的重要途径。

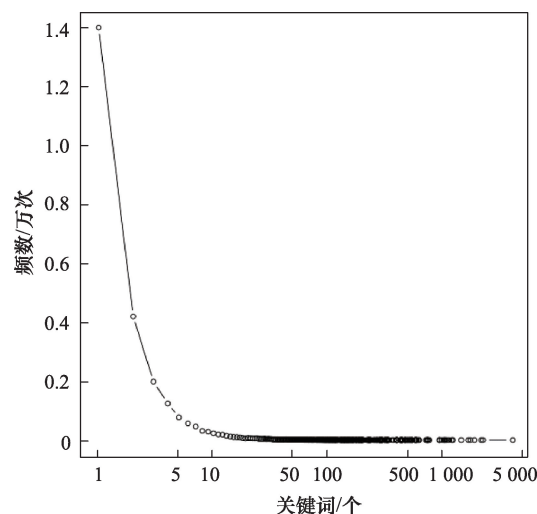


图1 游客评论

Figure 1 Tourists comments

2018年7月

3.2 目的地形象感知维度分析

游记描述了游客的旅游经历,涉及到食住行游购娱等旅行经历的方方面面^[7],它所提供的原真的、自发的和丰富的信息也反映了目的地感知形象。感知维度是学者进行目的地形象测量和评价的主要途径^[36],而各个维度在游客感知的重要性中存在差异。通过分析游客评论,看出游客讨论吸引物、设施、环境等不同维度的词汇及篇幅存在较大差异。于是本文从旅游吸引物、公共基础设施、旅游基础设施、休闲和娱乐、旅游环境和地方氛围6个感知维度对高频词进行分类,见表2。通过分析对各维度涉及的高频词的数量和频次,探究各个感知维度的重要性。

研究结果显示,“旅游吸引物”的高频词数量最多,频率最高,是目的地形象维度的第一层次。旅游吸引物包括自然类和人文吸引物。自然吸引物包括6个词汇,占9.5%,人文吸引物包括29个词汇,占46.0%。首先,以旅游景点为代表的核心旅游吸引物,是吸引游客前往旅游目的地的最主要因素^[6],也是南京旅游形象的主要载体。研究结果揭示夫子庙、中山陵作为南京的代表性景点,成为游客感知目的地的重要场所。其次,高频词也包括“新街口”“先锋书店”等词汇。“新街口”作为南京商业集聚区,吸引了游客前往购物。“先锋书店”则以其独特的艺术设计和文艺气息吸引了大量游客来此游览。因此,商业圈、商业街、纪念品商店等场所也是

游客感知目的地的重要场所。最后,一些非传统吸引物,如“大学”(385次)成为游览景点的高频词汇,南京大学、东南大学、南京师范大学成为一些年轻游客的必去之地。研究结果显示南京的旅游吸引物既包括地文景观、水域风光等自然类吸引物,也包括建筑、文化、历史、美食等人文类吸引物,其中人文类吸引物最为突出,所占的高频词最多,成为游客感知南京形象的重要维度。

公共基础设施相关的高频词有7个,占11.1%,仅次于“旅游吸引物”,是目的地形象维度的第二层次。公共基础设施包括公共交通(地铁、火车、公交)、交通节点(火车站、南站、机场)。联结客源地到目的地、各主要景区、景点之间的交通方式是游客旅游体验中不可或缺的部分^[17]。在不同交通方式中地铁的关注度最高(371次),或许因为南京地铁线路发达,联结了大多数旅游景点,成为游客行程中重要的交通工具。此外,各类交通节点成为游客感受目的地形象的重要场所,一些特色的交通节点如浦口火车站甚至成为了重要的旅游景点。

旅游基础设施相关的高频词有4个,占6.3%,高频词比例最低。旅游基础设施包括住宿餐饮设施(酒店、房间、饭店、宾馆),并未覆盖酒吧、俱乐部等设施。这表明游客感知到的南京旅游基础设施种类单一,缺少民宿等特色住宿设施和酒吧、俱乐部等娱乐设施。

休闲和娱乐有关的高频词有6个,占9.5%,游

表2 高频词分类

Table 2 The classification of high frequency keywords

维度	高频词(频次)	词汇比例/%
旅游吸引物	自然类 秦淮河(645)玄武湖(393)夜景(92)莫愁湖(64)紫金山(63)钟山(109)	9.5
	人文类 夫子庙(942)中山陵(460)大学(385)总统府(382)建筑(339)风景(334)历史(330)小吃(269)明孝陵(262)粉丝(244)鸡鸣寺(224)美食(195)城墙(191)博物馆(159)新街口(158)雨花台(158)南京大屠杀遇难同胞纪念馆(139)孙中山(136)汤包(126)狮子桥(125)长江大桥(116)大牌档(101)江南贡院(98)文化(94)先锋书店(83)中华门(81)灵谷寺(66)音乐台(66)锅贴(62)	46.0
公共基础设施	地铁(371)火车(270)公交(150)交通(147)火车站(132)南站(103)机场(85)	11.1
旅游基础设施	酒店(459)房间(149)饭店(62)宾馆(56)	6.3
休闲和娱乐	步行(158)旅游/旅行(109)拍照(101)游览(94)参观(88)游船(65)	9.5
旅游环境	自然环境 方便(180)干净(68)	3.2
	社会环境	0.0
	政治经济环境 贵(172)价格(146)免费(129)	4.8
地方氛围	民国(162)六朝(84)著名(81)漂亮(67)热闹(51)自然(42)	9.5

2018年7月

等基础设施的感知和评价更为细腻。旅游环境维度下“价格”出现在男性评论高频词中,这与惯常认为的女性比男性更为关心商品价格相反。综上,男性更关注旅游吸引物这一维度并对地方环境中的价格有特殊关注,而女性对旅游基础设施、公共基础设施维度的感知更为细腻。感知维度呈现的性别差异可能与男性女性的出游动机和生理心理特征不同有关。男性的出游动机更多与增长见闻、锻炼、获得知识等有关,因而会更加关注目的地吸引物。女性对服务质量和舒适度的感知比男性游客更为挑剔,于是更注重旅游过程中的住宿、交通并有更多情感性的评价,验证了其他学者的研究结果^[29]。

3.3 目的地形象分析

游客对目的地形象的感知来源于不同的感知维度。形象反映游客对目的地的具体看法和评价,感知维度则反映的是此种看法(或评价)的形成途径。感知维度和感知形象相辅相成,紧密联系,将二者相结合进行分析,可以帮助深入了解目的地形象及其形成过程。为进一步分析感知维度对感知形象的影响,下文通过形象组成成分对南京感知形象进行分析。

目的地形象包括认知形象、情感形象、意动形象、整体形象四个组成成分。游客评论为学者探究目的地形象提供了丰富的原真的信息。通过文本挖掘方法,从认知形象、情感形象、意动形象三个方面对游客评论高频词进行分类,见表3,并通过语义网络图(图4)对整体形象进行分析。

认知形象相关词汇包括“中山陵”、“夫子庙”、“总统府”等旅游景点和“环境”、“酒店”、“建筑”、“风景”、“历史”等目的地属性词汇,在高频词中占72.7%。游客评论反映出游客感知南京的认知形象

较为丰富,既有中山陵、火车站等景点和交通节点,也有丰富多样的目的地属性特征,如环境等自然属性,建筑、历史、文化等人文属性。可以看出游客对南京的认知主要源于旅游吸引物维度,也有少数认知源于公共基础设施(火车站、机场)和旅游基础设施(酒店)的感知。

情感形象的词汇占20.0%。“漂亮”“热闹”“美丽”“雄伟”“壮观”是对游览景观的情感态度,“方便”“干净”“好吃”分别是对交通、住宿、饮食的正向评价,“喜欢”则直接的表达出游客积极的情感态度。由此,TGC反映出情感形象是正面的,来源于旅游吸引物、公共基础设施、旅游基础设施和地方氛围维度的感知。总体说来,南京作为热门旅游目的地,其旅游吸引物、设施和地方氛围都使游客满意。

意动形象相关词汇本身相对匮乏,只有一个,即“推荐”。意动形象指游客积极考虑目的地作为潜在旅行地。“推荐”表明游客积极同意其他人将南京作为旅行目的地。这也反映了积极的网络口碑效应将会促进南京旅游进一步发展。

由上可知南京目的地形象成分中认知成分主要来源于旅游吸引物,情感形象总体上为正面,并产生了网络口碑效应。根据游客评论及上述各形象成分的分析也可以进一步得知游客感知南京目的地的过程。Baloglu^[13]认为认知形象是情感形象的基础,认知形象、情感形象、意动形象结合形成整体形象。南京有中山陵、夫子庙等历史文化资源和秦淮河、玄武湖等自然旅游资源,且传统观光旅游产品占大多数,因而游客感知到的认知形象主要是热门旅游景点和建筑、历史、文化、饮食等旅游资源。在认知形象基础上,游客对南京的饮食、交通、住宿、景点、旅游氛围等做出了好吃、方便、干净、漂

表3 高频词旅游形象成分分类

Table 3 Tourism image components classification of high frequency words

类别	高频词	词汇比例/%
认知形象	景点 中山陵、夫子庙、总统府、玄武湖、明孝陵、南京大屠杀遇难同胞纪念馆、雨花台、鸡鸣寺、南京博物院、长江大桥、大学、新街口、狮子桥、南京大牌档、火车站、南站、机场、先锋书店、中华门、灵谷寺、音乐台	72.7
	内容 环境、酒店、建筑、风景、历史、中国、火车、小吃、味道、江南、门票、美食、城墙、金陵、民国、日军、文化、夜景、六朝	
情感形象	喜欢、好吃、方便、干净、漂亮、热闹、美丽、雄伟、壮观	20.0
意动形象	推荐	7.3

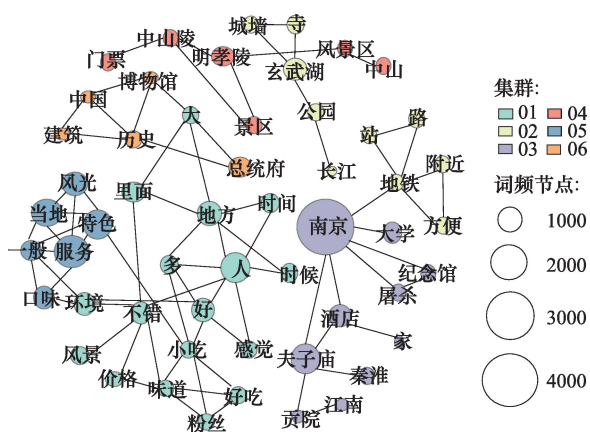


图4 南京语义网络

Figure 4 Co-occurrence network of Nanjing

亮、热闹等正向的情感评价,并有推荐他人游玩的意动形象,从而结合形成整体形象。

游客评论可以真实地反映南京的整体形象,词频共现网络图能够对数量庞大的评论数据进行关键信息的提取、关联分析和可视化处理,因此本文通过绘制游客评论的共现网络图(图4)对南京的整体形象进行分析。

图4的共现网络中“南京”是游客评论的核心,其余词汇直接或间接与南京相关联。共现网络呈现6个节点群。集群1主要代表对地方氛围和旅游环境的感知(人、地方、多、好、感觉、粉丝、好吃、价格、风景、环境等),其中饮食产生正面影响,而较大的客流量产生负面影响。集群2包括对玄武湖、长江等景点和地铁代表的交通设施的感知,“方便”表示游客对南京以地铁为代表的交通设施评价较高。集群3是以南京为核心关联的夫子庙周边景点、大学和南京大屠杀遇难同胞纪念馆等吸引物和以酒店为代表的旅游基础设施。集群4是相隔较近的中山陵和明孝陵,“门票”作为高频词与中山陵相关联,这是因为中山陵实行了免费开放,在游客心中留下深刻的印象。集群5包括当地、风光、特色、服务、口味、一般6个节点,代表游客对南京的总体评价为一般。集群6(历史、建筑、中国、博物馆和总统府)表明历史和建筑是南京旅游形象的重要组成部分。因而,游客感知的南京整体形象由景点、美食、交通、历史、建筑构成,主要源于旅游吸引物和公共基础设施维度,这也进一步验证了目的地形象感知维度的重要性存在差异。景点词汇最多,印证

了旅游景区是形成旅游目的地游客感知形象的基础^[43],其中夫子庙是南京的代表性景点,在游客感知南京形象过程中有关键作用。交通和美食呈明显的正向评价。

4 结论及建议

本研究以南京为例,采用KH coder内容挖掘软件和文本分析法,对游客评论数据进行分析,通过实证研究从理论上完善了旅游形象的文本研究方法,探索了TGC反映的目的地感知形象,相比以往对旅游目的地形象的研究,本次研究采用的方法更具有实时性,能够更为高效地获取游客对某一旅游地的评价状况,并更为真实和全面地了解游客的真实意愿,从而得到了一些新的发现。

(1)本文从旅游吸引物、公共基础设施、旅游基础设施、休闲娱乐、旅游环境和地方氛围6个维度对高频词分类分析得出,目的地形象维度的重要程度不同,按重要性排序,旅游吸引物(主要指人文类)是第一层次,公共基础设施是第二层次,休闲娱乐、旅游环境和地方氛围是第三层次,旅游基础设施是第四层次。这在以往学者^[7, 36]研究的基础上对维度的层次性进行了界定;此外,男性和女性的感知维度也存在一定差异。

(2)对目的地形象成分的研究结果显示,南京的认知形象主要源于旅游吸引物和公共基础设施维度的感知,说明经过长期的形象构建,南京的认知形象得到游客的肯定;情感形象为正面,源于对旅游吸引物、旅游基础设施和地方氛围的感知,并产生了推荐游览的意动形象,形成了网络口碑效应;南京整体形象由景点、美食、交通、历史、建筑构成,主要源于对旅游吸引物和公共基础设施维度的感知,这也验证了感知维度的重要性存在差异。其中游客对交通和美食呈明显的正向评价,而负面印象主要源于较大的客流量。

(3)研究结果显示南京“人文绿都”的目的地形象基本得到游客认可,民国、六朝、历史、自然等成为南京的特色形象。

针对上述结论,本文对南京旅游提出以下发展建议:

(1)吸引物和公共基础设施维度是形象感知的重要维度,因此南京旅游形象塑造和宣传应突出这

2018年7月

一层次,以丰富、免费和好玩的旅游吸引物,以便利的交通设施和服务为宣传重点。

(2)南京认知形象的重点为人文景点,但也出现了高校等新的热门景点,南京可以通过影视剧的宣传高校校区等新的吸引物,使南京的认知形象不断更新,以吸引年轻的游客群体。

(3)南京的情感形象偏向正面,可通过提高游客对吸引物、旅游基础设施和地方氛围维度感知的满意度,从而利用网络口碑效应推动南京旅游的发展。

(4)南京整体形象的负面认知主要是游人多,拥挤、排队等不良的游客体验。可以通过实行游客分流、智慧旅游等手段控制客流量,提高满意度。

TGC为研究游客对旅游目的地感知形象提供了丰富真实的有效信息。以往对TGC的研究多是对有限的样本数量进行人工的定性分析,不可避免存在人工处理上的误差,本文采用大数据的收集和定量与定性结合的分析方法,将TGC和大数据研究方法相结合增强了结论的可靠性和科学性。但是,TGC的研究也存在一定的局限性,TGC覆盖的人群主要是习惯在网站上对目的地进行评论的年轻人,因而对TGC的研究可能会忽略那些不在网站上进行评论的群体。此外,本研究只覆盖了马蜂窝一家旅游评论网站,评论数据也只有1万多条,尚未实现真正意义上的网络数据全覆盖。今后的研究可以选取不同的网站和样本群体,进行更大范围的数据覆盖和分析。

参考文献(References):

- [1] Leung D, Lee H A, Law R. Adopting Web 2.0 technologies on chain and independent hotel websites: a case study of hotels in Hong Kong[J]. *Springer Vienna*, 2011, doi: 10.1007/978-3-7091-0503-0_19.
- [2] Sun M H. Using Chinese travel blogs to examine perceived destination image: the case of New Zealand[J]. *Journal of Travel Research*, 2015, 54(4): 192-205.
- [3] Xiang Z. Role of social media in online travel information search[J]. *Tourism Management*, 2010, 31(2): 179-188.
- [4] Guo Y, Barnes S J, Jia Q. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation[J]. *Tourism Management*, 2017, 59: 467-483.
- [5] Litvin S W, Goldsmith R E, Pan B. Electronic word-of-mouth in hospitality and tourism management[J]. *Tourism Management*, 2008, 29(3): 458-468.
- [6] Banyai M, Glover T D. Evaluating research methods on travel blogs[J]. *Journal of Travel Research*, 2012, 51(3): 267-277.
- [7] Pan B, MacLaurin T, Crofts J C. Travel blogs and the implications for destination marketing[J]. *Journal of Travel Research*, 2007, 46(1): 35-45.
- [8] Mak A H N. Online destination image: comparing national tourism organisation's and tourists' perspectives[J]. *Tourism Management*, 2017, 60: 280-297.
- [9] Gartner W C. Image Formation Process[A]. Uysal M, Fesenmaier D. *Communication and Channel Systems in Tourism Marketing* [M]. New York: The Haworth Press, 1993.
- [10] Baloglu S, Love C. Association meeting planners' perceptions and intentions for five major US convention cities: the structured and unstructured images[J]. *Tourism Management*, 2005, 26(5): 743-752.
- [11] Chew Y C Y T, Jahari S A. Destination image as a mediator between perceived risks and revisit intention: a case of post-disaster Japan[J]. *Tourism Management*, 2014, 40(2): 382-393.
- [12] Qu H, Kim L H, Im H H. A model of destination branding: Integrating the concepts of the branding and destination image[J]. *Tourism Management*, 2011, 32(3): 465-476.
- [13] Baloglu S, McCleary K W. A model of destination image formation[J]. *Annals of Tourism Research*, 1999, 26(4): 868-897.
- [14] Echtner C M, Ritchie J R B. The measurement of destination image: an empirical assessment[J]. *Journal of Travel Research*, 1993, 31(4): 3-13.
- [15] Stepchenkova S, Zhan F. Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography[J]. *Tourism Management*, 2013, 36(3): 590-601.
- [16] Beerli A, Martín J D. Factors influencing destination image[J]. *Annals of Tourism Research*, 2004, 31(3): 657-681.
- [17] 吴晋峰. 旅游目的地形象“拼图”及测评方法[J]. 陕西师范大学学报(自然科学版), 2014, (6): 85-93. [Wu J F. Jigsaw puzzle of tourism destination image and its evaluation index[J]. *Journal of Shaanxi Normal University (Natural Science Edition)*, 2014, (6): 85-93.]
- [18] 贾衍菊, 林德荣. 旅游者服务感知、地方依恋与忠诚度—以厦门为例[J]. 地理研究, 2016, 35(2): 390-400. [Jia Y J, Lin D R. Tourists' perception of urban service, place attachment and loyal behaviors: a case study of Xiamen[J]. *Geographical Research*, 2016, 35(2): 390-400.]
- [19] 史坤博, 杨永春. 大陆大学生在台湾的旅游动机及旅游形象感知评价[J]. 资源科学, 2015, 37(3): 593-604. [Shi K B, Yang Y C. Chinese mainland university students' tourism motives and evaluation of the tourism destination image of Taiwan[J]. *Resources Science*, 2015, 37(3): 593-604.]
- [20] Cai L A. Cooperative branding for rural destinations[J]. *Annals of*

- Tourism Research*, 2002, 29(3): 720–742.
- [21] 陈楠, 乔光辉. 基于感知-认知因素的奥运会后北京旅游形象变化研究-以入境游客为例[J]. *资源科学*, 2009, 31(6): 1000–1006. [Chen N, Qiao G H. Change of tourism images of Beijing after 2008 Beijing Olympics based on cognitive and perceptive elements: a case study of inbound tourists[J]. *Resources Science*, 2009, 31(6): 1000–1006.]
 - [22] Bignalcaiz E, Snchez Garca I, Sanz Blas S. The functional-psychological continuum in the cognitive image of a destination: a confirmatory analysis[J]. *Tourism Management*, 2009, 30(5): 715–723.
 - [23] Hallmann K. Perceived destination image: an image model for a winter sports destination and its effect on intention to revisit[J]. *Journal of Travel Research*, 2015, 54(1): 94–106.
 - [24] Dann G M S. Tourists' images of a destination—an alternative analysis[J]. *Journal of Travel & Tourism Marketing*, 1996, 5(1–2): 41–55.
 - [25] White C J. Ideal standards and attitude formation: a tourism destination perspective[J]. *International Journal of Tourism Research*, 2014, 16(5): 441–449.
 - [26] Tasci A D A. Social distance: the missing link in the loop of movies, destination image, and tourist behavior [J]. *Journal of Travel Research*, 2008, 47(4): 494–507.
 - [27] 吴小根, 杜莹莹. 旅游目的地游客感知形象形成机理与实证-以江苏省南通市为例[J]. *地理研究*, 2011, 30(9): 1554–1565. [Wu X G, Du Y Y. Empirical study on perceptive image formation of tourists on tourism destination: a case study of Nantong, Jiangsu Province[J]. *Geographical Research*, 2011, 30(9): 1554–1565.]
 - [28] 郑鹏. 旅游地认知形象中功能性-心理属性对旅游者游后行为的影响-以西安欧美旅游者为例[J]. *资源科学*, 2012, 34(5): 948–955. [Zheng Peng. The effects of functional-psychological continuum in the cognitive image of a destination on tourists' behavior after the trip: a case for European and American tourists in Xi'an [J]. *Resources Science*, 2012, 34(5): 948–955.]
 - [29] Chen C F, Phou S. A closer look at destination: Image, personality, relationship and loyalty[J]. *Tourism Management*, 2013, 36(3): 269–278.
 - [30] Stylos N, Bellou V, Andronikidis A, et al. Linking the dots among destination images, place attachment, and revisit intentions: a study among British and Russian tourists[J]. *Tourism Management*, 2017, 60: 15–29.
 - [31] 南京市统计局. 南京市2016年国民经济和社会发展统计公报[EB/OL]. (2017-03-01)[2018-01-04]. <http://www.nanjing.gov.cn/xxgk/szf/201703/t20170330-4418767.html>. [Nanjing Bureau of Statistics. Statistics Bulletin on National Economic and Social Development of Nanjing in 2016[EB/OL]. (2017-03-01)[2018-01-04]. <http://www.nanjing.gov.cn/xxgk/szf/201703/t20170330-4418767.html>.]
 - [32] 于鹏, 张宏梅, 黄薇薇. 韩国旅游形象研究: 基于韩国旅游发展局新浪微博的内容分析[J]. *安徽师范大学学报(自然科学版)*, 2015, 38(4): 377–383. [Yu P, Zhang H M. Study on tourist image of South Korea: based on a content analysis of Korea tourism organization's Sina microblog[J]. *Journal of Anhui Normal University (Natural Science Edition)*, 2015, 38(4): 377–383.]
 - [33] 冯捷蕴. 北京旅游目的地形象的感知-中西方旅游者博客的多维话语分析[J]. *旅游学刊*, 2011, 26(9): 19–28. [Feng J Y. Perception of the image of Beijing's tourist destinations: an Analysis of the multi-dimensional discourses on the blogs from Chinese and Western tourists[J]. *Tourism Tribune*, 2011, 26(9): 19–28.]
 - [34] 谢彦君, 马天, 卫银栋. 宣传片、在线评论和游记对目的地形象改变的实证研究-以大学生对台湾旅游形象认知变化为例[J]. *北京第二外国语学院学报*, 2014, 36(1): 77–84. [Xie Y J, Ma T, Wei Y D. An empirical research on destination image change of promotional videos, online-reviews and travel notes— a case study for the change of Taiwan's image[J]. *Journal of Beijing International Studies University*, 2014, 36(1): 77–84.]
 - [35] 陈静, 章锦河, 刘泽华, 等. 旅游博客研究进展与启示[J]. *地理科学进展*, 2013, 32(10): 1511–1519. [Chen J, Zhang J H, Liu Z H, et al. Progress in travel blog research and its implications [J]. *Progress in Geography*, 2013, 32(10): 1511–1519.]
 - [36] 王媛, 许鑫, 冯学钢, 等. 基于文本挖掘的古镇旅游形象感知研究-以朱家角为例[J]. *旅游科学*, 2013, 27(5): 86–95. [Wang Y, Xu X, Feng X G, et al. Study on tourist image of South Korea: based on a content analysis of Korea tourism organization's Sina microblog[J]. *Tourism Science*, 2013, 27(5): 86–95.]
 - [37] 程慧荣, 黄国彬, 郑琳. 非结构化文本分析软件比较研究——以KH Coder和Wordstat为例[J]. *图书与情报*, 2015(4): 110–117+122. [Cheng H Y, Huang G B, Zheng L. Comparison of Software for Unstructured Text Analysis: KH Coder vs. Wordstat[J]. *Library & Information*, 2015(4): 110–117+122.]
 - [38] 刘文字, 毛伟伟. “教授”集体身份在报纸媒体与微博中的话语建构差异[J]. *现代教育管理*, 2017(9): 70–74. [Liu W Y, Mao W W. A Comparative Study on Discursive Construction of Professors' Collective Identity on Printing Media and Weibo[J]. *Modern Education Management*, 2017(9): 70–74.]
 - [39] Tussyadiah I P, Zach F. Identifying salient attributes of peer-to-peer accommodation experience[J]. *Journal of Travel & Tourism Marketing*, 2017, 34(5): 636–652.
 - [40] Toutanova K, Klein D, Manning C, et al. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network[C]. Los Angeles: Proceedings of HLT-NAACL, 2003.
 - [41] Romesburg H C. Cluster Analysis for Researchers. [M]. Belmont: Lifetime Learning Publications, 1984.
 - [42] Fruchterman T M J, Reingold E M. Graph drawing by force-directed placement[J]. *Software - Practice and Experience*, 1991, 21

(11): 1129–1164.

[43] 张春晖, 白凯, 马耀峰, 等. 入境游客视角下中国旅游形象的景区代言[J]. 地理研究, 2013, 32(5): 924–941. [Zhang C H, Bai K,

Ma Y F, *et al.* Scenic spot endorsement of Chinese tourism image from the perspective of inbound tourists[J]. *Geographical Research*, 2013, 32(5): 924–941.]

A research on destination image and perceived dimension difference based on big data of tourists comments: a case of Nanjing

XU Feifei¹, LA Liqing¹, YE Feng²

(1. *Tourism Department, School of Humanities, Southeast University, Nanjing 210096, China;*

2. *Hospitality and Tourism Management, Purdue University, West Lafayette 47906, U. S)*

Abstract: With the development of web 2.0 technology, internet data has become an important data source for the study of tourism destination image. Taking Nanjing as an example, this paper collected 10077 online reviews about Nanjing from a specialized travel blog site Ma Feng Wo by using the web crawler tool. Through data mining techniques such as word frequency, network semantic analysis, the paper discussed the dimensions of destination perceived image by analyzing TGC (Tourist Generated Comments). First, the results of data analysis show that the importance of dimensions of destination perceived image is different. It is suggested that for the six dimensions of destination perceived image, the tourism attraction is the first level, general infrastructure is the second level, tourist leisure and recreation, tourism environment and atmosphere of the place are the third level, tourist infrastructure is the fourth level. Second, the salient cognitive components in Nanjing's perceived image stem from the dimensions of tourist attractions and public infrastructure. The emotional image of Nanjing is positive. The conative image of Nanjing has been formed and generated the effect of “word of mouth”. In addition, we found that the overall image of Nanjing mainly originated from the perception dimensions of tourism attractions and public infrastructure, which also verified the hierarchy of perception dimensions of destination image. This paper verifies the relevant theories of tourism destination image, enriches the research methods of big data in the field of tourism research, and provides a scientific basis for the construction and development of tourism image in Nanjing.

Key words: big data; destination image; text mining; Nanjing