

# 基于网络评论的旅游目的地评价研究 ——以我国 31 个省市自治区为例

龚 箭<sup>1,2\*</sup>, 杨舒悦<sup>1,2</sup>

(1. 华中师范大学城市与环境科学学院, 武汉 430079; 2. 中国旅游研究院武汉分院, 武汉 430079)

**摘要:** 以创新旅游目的地评价方法为目的, 基于旅游网络评论数据, 从旅游目的地发展规模和质量两个维度构建旅游目的地关注度—满意度分析(ASA)模型。整合国内 8 大在线旅游平台评论数据, 采集全国 23 496 个景点的 1 368 052 条有效评论, 建立我国大陆 31 个省市自治区基础数据库。通过模型化和可视化的方法对省域旅游目的地发展水平进行分析, 结果显示: 1) 网络旅游关注度呈现长尾和极化特征, 约 80% 的在线旅游评论集中在 10% 的核心旅游景点; 核心旅游景点空间聚集度呈现出由高到低的“东 中 西”格局; 2) 不同类型景点在网络关注度和满意度上均有明显差异, 且游客对不同类型景点具有不同的正负面情感特征。基于上述特征分析, 结合 31 个省市自治区 ASA 定位分析结果, 提出了我国大陆 31 个省市自治区旅游发展优化路径。

**关键词:** 网络文本分析; 关注度满意度分析(ASA)模型; 旅游目的地评价

中图分类号: F59

文献标识码: A

随着旅游目的地竞争的日益激烈, 旅游目的地评价研究成为旅游研究者、政府决策者与企业管理者关注的热点<sup>[1]</sup>。Kozak 与 Rimmington 提出旅游目的地评价首先需要确定目的地的直接竞争对手(竞争组)<sup>[2]</sup>。基于此, 国内外学者构建了众多旅游目的地竞争力评价模型, 从不同空间尺度对旅游目的地进行了评价和比较。研究所采用的数据主要为国际权威组织或国家公开发布的数据和游客问卷调查数据。公开数据相对容易获取, 但统计口径、统计质量的差异使数据往往缺乏可比性, 而问卷调查数据又存在样本量、信度、效度等问题。信息化和分享经济背景下, 旅游网络信息分享渐成风潮, 海量的旅游评论内容不仅为旅游者旅游决策提供了丰富信息, 也为广大学者研究旅游消费者行为和旅游目的地发展提供了重要资源。与传统的公开数据和调研数据相比, 网络文本信息是由消费者根据其自身体验而提供的关于旅游产品或服务的正面或负面陈述信息<sup>[3]</sup>, 更具时效性和真实性。因而, 近年来也有许多学者开始尝试基于网络文本数据探究和揭示旅游目的地发展状况及旅游市场特征。

现有的基于旅游网络信息的旅游目的地研究,

主要从 2 个维度对旅游目的地发展进行评价。一是从旅游目的地市场规模和影响力方面, 研究旅游目的地网络关注度。多数研究使用了 Google 搜索<sup>[4]</sup>、百度搜索及其提供的百度指数模型<sup>[5-8]</sup>, 也有一些学者开始尝试挖掘如 TripAdvisor<sup>[9]</sup>、新浪博客<sup>[10]</sup>等旅游社区平台网络信息数据, 以评论或游记数量作为旅游目的地网络关注度评价指标, 研究旅游目的地关注度时空格局、网络关注度与现实旅游流的耦合关系等。二是从旅游目的地发展质量方面, 研究游客对旅游目的地的感知或情感。主要利用内容分析、情感分析等方法, 基于网络文本研究旅游目的地感知形象或对游客满意度进行定量测评。如 Steve Pan 等通过对旅游照片的内容分析探究了旅游动机、地方情感和感知形象间的关系<sup>[11]</sup>; 王媛等分析了新浪博客中游客对朱家角古镇的旅游形象感知<sup>[12]</sup>; 刘欢、朱峰对婺源的网络感知形象进行了历时演化分析<sup>[13]</sup>; 何琼峰<sup>[14]</sup>和孙小培<sup>[15]</sup>分别基于网络评论对目的地游客满意度进行定性和定量研究。

旅游目的地的发展和评价既要注重规模也要注重质量, 但是已有的研究大多限于关注度或满意度的单一考量, 研究的应用对象主要是旅游企业,

收稿日期: 2017-06-20.

基金项目: 国家社会科学基金重点项目(14AJY024); 教育部哲学社会科学发展报告项目(11JGP041).

\* E-mail: 343901552@qq.com.

无法适用于目的地的整体评价,对目的地管理的指导作用有限。鉴于此,本研究尝试将上述两个维度的指标和数据进行整合,基于旅游网络文本构建新的旅游目的地评价模型,并以我国大陆 31 个省市自治区数据为例,通过模型化、可视化的方式对目的地旅游发展作出评估,以方便目的地管理者利用大数据对目的地旅游发展进行有效诊断,也为旅游目的地评价研究提供新的视角和方法。

## 1 研究设计

### 1.1 数据来源

本文研究数据主要为《2016 中国旅游业发展报告》专题二中的基础数据<sup>[16]</sup>。在武汉酷犬数据科技有限公司的技术支持下,利用大数据挖掘技术采集携程网、同程网、百度旅游、去哪儿网、驴妈妈、蚂蜂窝、大众点评和途牛网这 8 个国内主要在线旅游平台的用户评论数据(图 1)。数据包含了自 2015 年 7 月 1 日至 2016 年 6 月 30 日期间以上 8 大平台上的用户评论数共 5 267 988 条。在去除重复、无意义内容后,对数据进行进一步删选:评论必须含有具体信息和情感内容。经过数据清洗后,最终得到全国 23 496 个景点的有效评论数 1 368 052 条。

表 1 我国大陆 31 个省市自治区旅游评论数据库  
Tab. 1 Tourism comments database of 31 provinces in China mainland

编号	省市自治区	景点数/个	评论数/条	编号	省市自治区	景点数/个	评论数/条
1	江苏省	1 902	226 937	17	重庆市	386	25 347
2	广东省	1 841	153 751	18	湖南省	706	24 748
3	浙江省	2 263	128 632	19	辽宁省	691	24 248
4	上海市	765	120 087	20	黑龙江省	444	20 478
5	北京市	1 155	81 293	21	河北省	779	19 374
6	四川省	1 190	65 573	22	江西省	680	16 904
7	山东省	1 270	60 688	23	山西省	604	16 337
8	福建省	1 143	49 415	24	贵州省	497	12 011
9	云南省	863	47 524	25	甘肃省	342	7 366
10	海南省	284	47 207	26	西藏自治区	302	6 419
11	湖北省	681	38 920	27	内蒙古自治区	460	6 344
12	陕西省	675	35 686	28	吉林省	238	5 307
13	河南省	848	30 953	29	新疆维吾尔自治区	415	4 965
14	安徽省	761	29 827	30	青海省	245	4 616
15	广西壮族自治区	654	27 370	31	宁夏回族自治区	134	4 234
16	天津市	278	25 491				

### 1.2 研究方法

IPA(Importance-Performance Analysis)分析

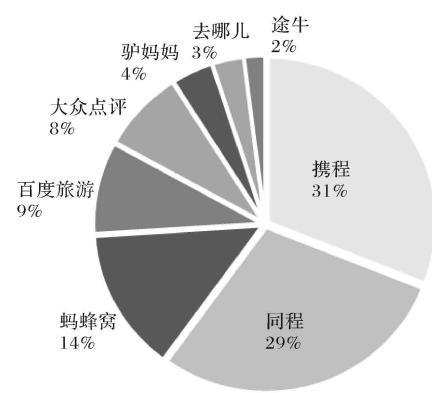


图 1 数据来源  
Fig. 1 Data sources

旅游目的地的研究可以从景区、城市、省域、国家等不同空间尺度展开,当前我国目的地研究中主要以城市、景点等小尺度空间为主,省域旅游目的地评价研究相对较少。但在全球旅游竞争加剧的背景下,地方和区域层面的目的地的作用日益显现<sup>[17]</sup>。因此本研究选择以省域旅游目的地为研究对象,基于景点的地理信息数据将用户评论按省域进行归档整理,建立了我国大陆 31 个省市自治区网络关注度、满意度基础数据库(表 1)。

法即重要性及表现分析法,以其形象直观、方便诊断和决策而被广泛运用于各类研究。本研究在参照

Lai等<sup>[18]</sup>及戴光全和肖璐<sup>[19]</sup>所采用的IPA分析法的基础上,构建了旅游目的地评价的“关注度—满意度”分析模型(Attention-Satisfaction Analysis,简称ASA模型)。模型的具体构建方法为:1)确定关注度和满意度的分值范围;2)分别求出关注度及满意度的平均值,并且以这两个平均值为交叉点,基于该交叉点划分“关注度—满意度”图的四象限:第I象限是高关注度、高满意度的旅游发展优势区;第II象限是低关注度、高满意的旅游发展机会区;第III象限是高关注度、低满意度的旅游发展修补区;第IV象限是低关注度、低满意度的旅游发展重点改进区;3)根据测算的目的地旅游网络关注度和满意度值,确定各旅游目的地在四象限的位置;4)用气泡标注目的地,气泡大小表示该旅游目的地旅游资源丰度。

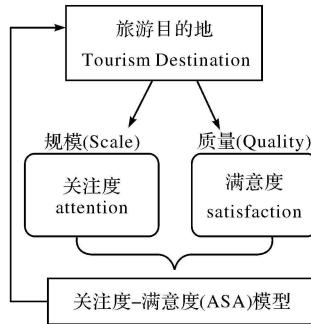


图2 旅游目的地评价模型研究框架

Fig. 2 Research framework of TD evaluation model

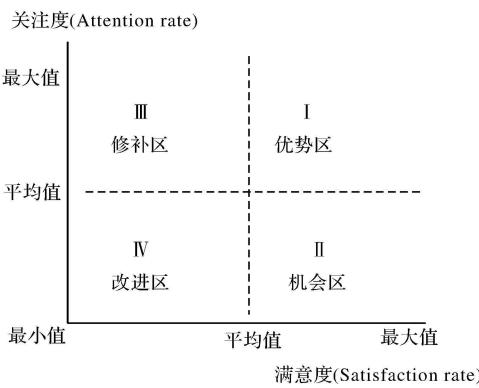


图3 关注度-满意度分析(ASA)模型

Fig. 3 Attention-Satisfaction analysis model

省域旅游目的地网络关注度、满意度及旅游资源丰度的具体测算方法如下。

1) 网络关注度。目前网络关注度的测算方法主要有两种,一种是直接使用百度或Google提供的用户关注度数据;另一种是对网络游记或评论进行统计,以游记数、评论数或目的地词频数表征网

络关注度。本研究所采取的是第二种测算方法,即与目的地相关的有效评论数越多,则该目的地的网络旅游关注度越高。网络旅游关注度  $A_i = \frac{X_i}{\sum X_i} \times 100$ ,  $X_i$  为  $i$  目的地有效评论数。

2) 网络满意度。通过自然语言处理对评论信息进行语义分析、情感分析,判断评论的情感极性,对有效评论中的正面印象数和负面印象数进行计量统计分析。满意度  $S_i = \frac{P_i}{P_i + N_i} \times 100$ ,  $P_i$  为  $i$  目的地有效评论中的正面印象数,  $N_i$  为  $i$  目的地有效评论中的负面印象数。

3) 旅游资源丰度。共享经济及全域旅游背景下,除A级景区外,高校、城市公园等也成了目的地重要旅游资源,而以往的研究往往忽视了对这类资源的统计。不同于以往文献中用A级景点个数测定旅游资源丰度<sup>[20]</sup>,本文将通过目的地拥有的具有有效评论的旅游景点数对旅游资源丰度进行测定,并通过归一化处理,使旅游资源丰度的气泡值  $D_i$  在  $[0,1]$  区间变化,归一化处理公式为  $D_i = \frac{V_i - V_{\min}}{V_{\max} - V_{\min}}$ ,  $V_i$  为  $i$  目的地旅游景点个数,  $V_{\min}$  为所有目的地中所含景点数最小值,  $V_{\max}$  为所有目的地中所含景点数最大值。

## 2 分析与发现

### 2.1 关注度分布特征

有效评论数是目的地关注度的直接体现,为更好地揭示省域旅游目的地网络旅游关注度差异及其影响因素,本文首先以景点评论数据为基础分析了网络旅游关注度的分布特征。对我国31个省市自治区23 496个景点有效评论数进行统计分析可以发现,景点关注度离散度大,极化特征明显,呈现“长尾现象”。有效评论数超过1 000的景点仅有245个,但有效评论数总量却达600 611条;有效评论数在100条以下的景点多达21 514个,但有效评论数仅为270 936条,即头部1.0%的景点,其有效评论数占总有效评论数达43.9%;中部7.4%的景点,其有效评论数占总有效评论数36.3%;而尾部91.6%的景点,其有效评论数却仅占总有效评论数的19.8%,说明游客主要集中于少数“热门”景点。

用ArcGIS软件对头部245个高关注度景点进行可视化处理,从空间分布看,北京、上海、江苏、浙江和广东为高关注景点的集聚中心。高关注度景点主要分布在东部沿海地区,而西部地区高关注度

景点较少,空间聚集度总体上呈现出由高到低的“东—中—西”格局,与区域经济梯度相耦合。这与王琨等基于 Tripadvisor 的中国旅游地国际关注度及空间分布格局研究结果也相一致<sup>[9]</sup>。

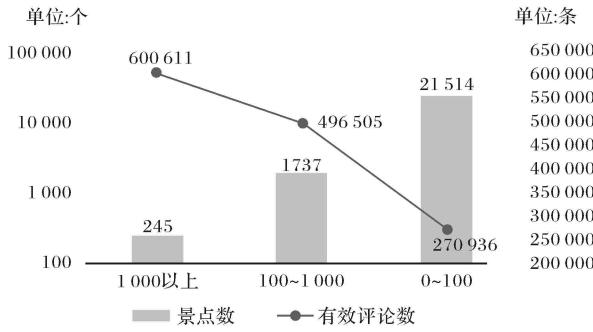


图 4 有效评论数与景点数的层级对应关系

Fig. 4 Levels corresponding relation between number of comments & tourist attractions

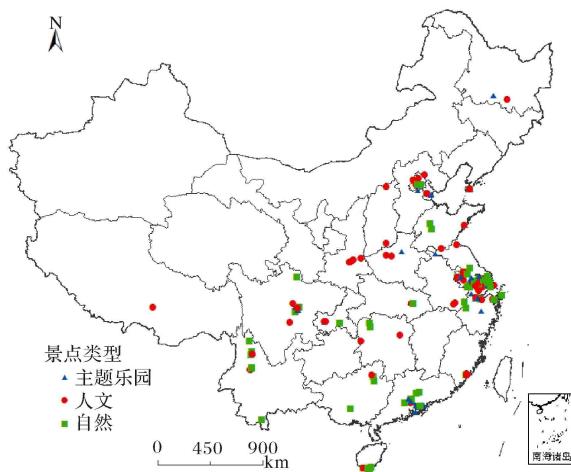


图 5 245 个高关注度旅游景点的空间分布图

Fig. 5 Distribution of 245 tourist attractions with high attention rate

从景点类型看,主题乐园类景点有效评论数均值最大,为 237.29 条,关注度远高于自然类景点(52.70 条)和人文类旅游景点(44.96 条);自然类

景点关注度略高于人文类景点。这可能主要与用户的群体特征相关,在线旅游社区用户主要以青年学生和年轻妈妈为主体,且青年学生和年轻妈妈群体的外出休闲时间和上网时间都相对较多,而这两类主体也正是主题乐园类景点的主要市场群体,因而在网络游客评论数据中主题乐园类景点关注度相对传统人文或自然类景区要高许多。

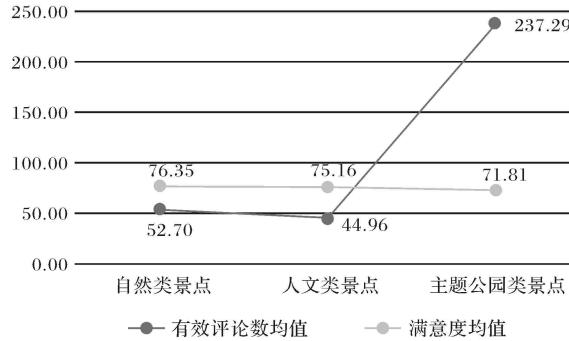


图 6 不同类型景点有效评论数与满意度均值比较

Fig. 6 Comparison of mean comments numbers & satisfaction rate between different types of tourist attractions

## 2.2 满意度情感特征

通过对景点评论中正面及负面词汇进行识别和统计,共得到正面印象数 1 980 814 个,负面印象数 710 513 个,总体而言正面印象数远高于负面印象数,旅游景点的总体网络满意度均值为 73.95。人们在社交网络上的表达都是有选择性的表达,而表达积极正面情绪更容易获得社会认同,因此人们在社交网络上会不自觉地表现得积极向上<sup>[21]</sup>。从景点类型上看,自然类景点满意度均值 76.35,人文类景点满意度均值为 75.16,主题乐园类景点满意度均值为 71.81,自然类景点网络满意度相对较高。为更好地揭示旅游目的地网络旅游满意度差异及其影响因素,本文以景点评论数据为基础,结合高频词表和社会语义网络结构进行了网络旅游满意度情感特征分析。

表 2 不同类型景点正负面高频词及情感特征

Tab. 2 High frequency words and emotional characteristics of different types of tourist attractions

景点类型	正面高频词(前 10)	特征提取	负面高频词(前 10)	特征提取
自然类	风景、美、值得、漂亮、好、自然、赞、喜欢、推荐、神奇	景观质量 游览体验	远、人多、可惜、遗憾、辛苦、一般、累、贵、失望、挤	旅游交通 景区拥挤度
人文类	特别、历史、喜欢、值得、美、特色、免费、文化、推荐、古朴	历史文化 景区服务	一般、差、破坏、商业化、坑、贵、没意思、乱、失望、随便	景区氛围 旅游服务价格
主题乐园类	好玩、刺激、欢乐、开心、值得、棒、喜欢、赞、推荐、特色	景区氛围 娱乐项目	贵、不值、累、人多、排队、一般、没意思、不敢、无聊、失望	门票价格 景区服务

通过高频词及语义网络结构对游客评论进行情感特征提取,自然类景点正面评论主要集中在两方面:一是对自然类景点风景资源的高度认同,体现在“风景、美、漂亮、神奇”等高频词上;二是反映游客对游览体验的满意,主要体现在“值得、好、赞喜欢、推荐”等高频词上。而自然类景点负面高频词“远、辛苦、累、贵”和“人多、可惜、挤”等则具体反映出游客对旅游交通和景区拥挤度的抱怨。人文类景点正面评论高频词“特别、历史、美、文化、古朴”等主要反映了游客对景点历史文化的赞美,而“喜欢、值得、免费、特色、推荐”等词则是游客对景区服务的肯定。人文类景点负面评论高频词如“破坏、商业化、乱”等主要集中于对景区氛围的批评,“坑、贵、随便”等词则主要反映游客对旅游服务价格过高的

不满。主题乐园类景区正面评论高频词通常与景区整体氛围和娱乐项目设施有关,高频词主要有“好玩、刺激、欢乐、开心、值得”等,而负面评论高频词则主要反映了景区门票价格高和景区服务质量差这两方面,高频词主要有“贵、不值、累、人多、排队”等。

### 2.3 我国大陆31个省市自治区ASA定位分析

根据关注度和满意度数据,计算得到我国大陆31个省市自治区网络旅游关注度均值为3.32,网络旅游满意度的均值为73.95,在ASA图中找到点(73.95,3.32)。基于该点,得到ASA图四象限。根据网络旅游关注度、满意度值逐一定位31个省市区在四象限中的位置,并根据旅游资源丰度画出气泡图(图7)。

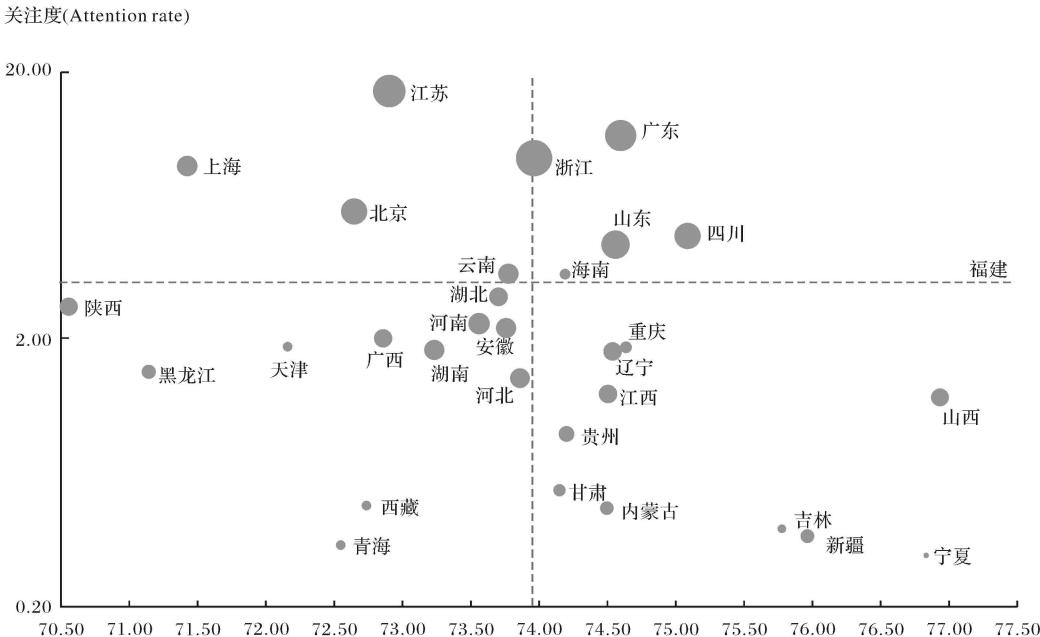


图7 我国大陆31个省市自治区ASA分析图

Fig. 7 ASA chart of 31 provinces in China mainland

位于第Ⅰ象限竞争优势区的省市区共6个,分别为:广东、浙江、山东、四川、福建和海南,这6个省市区在网络旅游关注度和满意度方面得分均比较高,在省域目的地竞争群中处于竞争优势地位;位于第Ⅱ象限竞争机会区的省市区共10个,分别为:重庆、辽宁、江西、贵州、山西、甘肃、内蒙古、吉林、新疆和宁夏,虽然这10个省市区网络关注度并不高但网络满意度相对较高,具有一定发展潜力;位于第Ⅲ象限修补区的省市区4个,分别为:江苏、上海、北京和云南,这4个省市区网络旅游关注度较高但满意度表现欠佳;位于第Ⅳ象限重点改进区的省市区共11个,分别为:湖北、河南、安徽、陕

西、广西、湖南、河北、天津、黑龙江、西藏和青海,这些省市区在网络旅游关注度和满意度方面还有待提高,目前在省域在线旅游竞争中仍处于相对劣势地位。

### 2.4 我国大陆31个省市自治区的ASA优化路径分析

基于网络文本的ASA分析法,以形象直观的气泡图形式为各省市区提供了基于比较的旅游目的地发展定位及发展标杆,能够使各省市区明确本区域网络关注度及满意度状况。为了更好地帮助省域旅游目的地在网络背景下提升旅游发展规模和质量,本文结合网络关注度分布特征、满意度情感

特征和 ASA 定位分析结果,提出了我国 31 个省市区旅游关注度及满意度优化路径。

1) 机会区:拉关注度长板。整体上看,ASA 气泡图中气泡呈现“上大下小”,即一般来说省域目的地所含旅游资源越丰富,网络旅游关注度就越高。但也存在个别省域在线旅游资源丰富度不高,但网络旅游关注度表现较好,如海南省虽然景点个数不多,但景点质量均较高,关注度分布相对均衡,因而目的地整体关注度较高。事实上,约 80% 的在线旅游评论集中在 10% 的核心旅游景点。旅游资源,特别是高级别的旅游资源是吸引旅游者、发展旅游业的物质基础<sup>[19]</sup>。因此对于省域旅游目的地来说,网络旅游关注度的提高主要还在于扩大核心旅游吸引物的影响力,而非增加旅游景点数量。对于位于第Ⅱ象限机会区的省域来说,在资源有限的情况下,与其全盘发力导致关注度分散,不如集中资源打造核心优势。

2) 修补区:补满意度短板。网络负面口碑对目的地品牌资产存在稀释效应<sup>[22]</sup>。对于省域旅游目的地而言,任何一个景区的负面评论都会对省域旅游整体满意度产生重要影响。因此,在满意度方面,更应重视控制负面评论,加强对低满意度景区的管理。对于位于第Ⅲ象限修补区的省域来说,应重点监测低满意度景点,且对不同类型的景区应有不同的管理侧重点:自然类旅游景点应注重改善旅游交通和控制旅游流;人文类景点需重视景区氛围营造和旅游服务价格管理;主题乐园类景点应注重控制景区门票价格和提高景区服务质量。

3) 改进区:先补满意度短板,后拉关注度长板。旅游网络口碑对旅游目的地的形象建构和旅游者购买选择有着重大的影响<sup>[21]</sup>。共享经济背景下,口碑效应能够迅速为目的地聚集人气,因此对于位于重点改进区的省市区来说,首先应致力于提高区域旅游满意度,继而通过网络口碑效应不断提高网络关注度,逐步提高区域旅游竞争力。因此,对于位于第Ⅳ象限重点改进区的省域来说,改进路径应为:“改进区”→“机会区”→“优势区”。

### 3 结论与讨论

#### 3.1 结论

基于网络文本数据,通过对关注度和满意度的综合分析,得到网络旅游关注度的分布特征及网络旅游满意度的情感特征:1) 网络旅游关注度呈现长尾和极化特征,约 80% 的在线旅游评论集中在 10% 的热点旅游景点,且头部高关注度景点的空间

聚集度呈现“东一中一西”依次递减的分布特征,与区域经济梯度耦合;2) 主题乐园类景点网络旅游关注度高于传统自然和人文类景点;3) 网络旅游满意度均值为 73.95,自然类旅游景点满意度高于人文类旅游景点,主题乐园类景点满意度最低;4) 不同类型景点满意度情感特征不同,影响自然类景点游客情感的正面因素主要是景观质量和天气状况,负面影响主要是旅游交通和景区拥挤度;影响人文类景点游客情感的正面因素主要是历史文化和景区服务,负面影响主要是景区氛围和旅游服务价格;影响主题乐园类景点游客情感的正面因素主要是景区氛围和娱乐服务,负面影响主要是门票价格和景区服务。基于上述特征分析,根据 ASA 定位分析结果,提出省域旅游目的地网络旅游满意度及关注度优化提升路径:位于第Ⅱ象限机会区的省域应重点提高核心旅游吸引物吸引力,拉关注度长板;位于第Ⅲ象限修补区的省域应重点监测低满意度景区,补满意度短板;位于第Ⅳ象限改进区的省域应先补满意度短板再拉关注度长板。

#### 3.2 讨论

本研究数据来源于 2015 年 7 月 1 日~2016 年 6 月 30 我国 8 大在线旅游平台,相对较长的时间尺度和较全的空间范围有利于更好更全面地揭示网络关注度和满意度时空分布的客观规律。与以往单一的网络关注度或满意度研究相比,本研究创新性地将满意度和关注度相结合构建了 ASA 分析模型,并以气泡图的形式呈现目的地评价结果,能够更直观更全面地为旅游目的地管理者提供基于比较的旅游目的地发展定位,也为基于大数据的旅游目的地评价研究提供新的方法和视角。

但正如前文所提到的,在线旅游社交平台的用户群体主要是中青年,因而本研究基于网络文本数据所得到的分析和排名结果与传统调研结果可能存在较大差异。大数据分析和传统线下调研如何相互补充、相互结合,也是我们今后研究的重要方向。此外,目前本研究对于关注度在时间上没有进行如月、周等更小尺度的细分,在之后研究中可进一步分析关注度月内、周内变化规律;在满意度研究方面,还可对目的地旅游交通、旅游住宿、旅游餐饮等满意度进行分维度细化测量;在研究的空间尺度上,还可从城市、景点等尺度出发,对模型进行进一步完善。

#### 参考文献:

- [1] 史春云,张捷,沈正平,等.旅游目的地竞争力定量评价

- 综述[J].人文地理,2006,21(3):72-77.
- SHI C Y, ZHANG J, SHEN Z P, et al. An overview of quantitative evaluation of the tourist destination competitiveness[J]. Human Geography, 2006, 21(3):72-77. (Ch).
- [2] KOZAK M, RIMMINGTON M. Measuring Tourist Destination Competitiveness: Conceptual Considerations and Empirical Findings [J]. Hospitality Management, 1999, 18: 273-283.
- [3] AKEHURST G. User generated content: the use of blogs for tourism organizations and tourism consumers [J]. Service Business, 2009, 3(1): 51-61.
- [4] 王章郡,方忠权,杜坤,等.中国自驾车旅游网络空间关注度的时空演变——基于Google搜索解析的分析[J].地域研究与开发,2011,30(5):112-117.
- WANG Z, FANG Z, DU K. Temporal-spatial envelopment characteristics of cyberspace attention index of self-driving tours in China: based on Google Insight[J]. Areal Research & Development, 2011, 30(5):112-117. (Ch).
- [5] 李山,邱荣旭,陈玲,等.基于百度指数的旅游景区络空间关注度、时间分布及其前兆效应[J].地理与地理信息科学,2008,24(6):102-107.
- LI S, QIU R X, CHEN L. Cyberspace attention of tourist attractions based on Baidu Index: temporal distribution and precursor effect[J]. Geography and Geo-Information Science, 2008, 24(6):102-107. (Ch).
- [6] 龙茂兴,孙根年,马丽君,等.区域旅游网络关注度与客流量时空动态比较分析——以四川为例[J].地域研究与开发,2011,3:93-97.
- LONG M X, SUN G N, Ma L J, et al. An analysis on the variation between the degree of consumer attention of travel network and tourist flow in regional tourism: a case of Sichuan province[J]. Areal Research & Development, 2011, 3:93-97. (Ch).
- [7] 张晓梅,程绍文,刘晓蕾,等.古城旅游地网络关注度时空特征及其影响因素——以平遥古城为例[J].经济地理,2016,36(7):196-202,207.
- ZHANG X M, CHENG S W, LIU X L, et al. Spatial-Temporal characteristics and influencing factors of network attention to ancient city destination: a case of Pingyao[J]. Economic Geography, 2016, 36(7):196-202,207. (Ch).
- [8] 宋增文.区域旅游络空间关注度与客源市场相关性研究[J].中国人口·资源与环境,2016,26(S2):270-273.
- SONG Z W. Study on correlation of regional tourism cyberspace attention distribution and tourist origins[J]. China Population, Resources and Environment, 2016, 26 (S2): 270-273. (Ch).
- [9] 王琨,郭风华,李仁杰,等.基于Tripadvisor的中国旅游地国际关注度及空间格局[J].地理科学进展,2014,33(11):1462-1473.
- WANG K, GUO F H, LI R J, et al. Tourism attention degree about China from overseas and its spatial patterns based on Tripadvisor[J]. Progress in Geography, 2014, 33(11): 1462-1473. (Ch).
- [10] 静恩明,郭风华,李仁杰,等.基于新浪旅游博客的河北省A级景区网络关注度研究[J].地理与地理信息科学,2015,31(3):118-122.
- JING E M, GUO F H, LI R J, et al. Analysis of scenic network attention-degree of A-grade tourist attractions in Hebei province: base on Sina travel blogs[J]. Geography and Geo-Information Science, 2015, 31(3):118-122. (Ch).
- [11] PAN S, LEE J, TSAI H, et al. Travel photos, Motivations, image dimensions, and affective qualities of places [J]. Tourism management, 2014, 40(2):59-69.
- [12] 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. Research on tourists' perceived image of ancient town using web text mining methods:a case study of Zhujiajiao[J]. Tourism Science, 2013, 27(5):86-95. (Ch).
- [13] 刘欢,朱峰.旅游目的地网络感知形象历时演化分析[J].资源开发与市场,2013,29(4):440-443.
- LIU H, ZHU F. Analysis of diachronic online image of tourism destinations[J]. Resource Development & Market, 2013, 29 (4):440-443. (Ch).
- [14] 何琼峰.基于扎根理论的文化遗产景区游客满意度影响因素研究——以大众点评网北京5A景区的游客评论为例[J].经济地理,2014,1:168-173,139.
- HE Q F. Grounded theory approach to influence factors of tourist satisfaction in scenic area of cultural heritage taking tourist comments on Beijing 5A scenic area from the website of www.dianping.com[J]. Economic Geography, 2014, 1:168-173,139. (Ch).
- [15] 孙小培.基于网络评论的目的地游客满意度研究[D].华东师范大学,2011.
- SUN X P. A study on the tourists' satisfaction based on the online reviews[D]. East China Normal University, 2011. (Ch).
- [16] 胡静,谢双玉,冯娟,等.2016年中国旅游业发展报告[M].北京:中国旅游出版社,2016: 246.
- HU J, XIE S Y, FENG J, et al. China's Tourism Development Report 2016[M]. Beijing: China Travel & Tourism Press,2016: 246. (Ch).
- [17] 许峰,李静,弗朗索瓦·贝达德,等.全球视野下优秀旅游目的地评价系统的发展与检验[J].旅游科学,2013,27(06):1-12.
- XU F, LI J, BDARD F, et al. Development and experimentation of excellent tourist destinations assessment system from global perspective[J]. Tourism Science, 2013, 8(17):457-467. (Ch).
- [18] IVAN K, MICHAEL H. Importance-performance analysis in tourism: a framework for researchers[J]. Tourism management, 2015, 48(1):242-267.
- [19] 戴光全,肖璐.基于区域联系和IPA的节事游客地方认同空间特征——以2011西安世界园艺博览会为例[J].人文地理,2012,4:115-124.

- DAI G Q, XIAO L. Spatial characteristics of event-goers' place identity based on regional relation&IPA model—a case study of international horticultural exposition 2011 in Xi'an[J]. Human Geography, 2012, 4:115-124. (Ch).
- [20] 马丽君, 孙根年, 黄芸玛, 等. 城市国内客流量与游客网络关注度时空相关分析[J]. 经济地理, 2011, 4:680-685.
- MA L J, SUN G N, HUANG Y M, et al. A correlative analysis on the relationship between domestic tourists and network attention[J]. Economic Geography, 2011, 4:680-685. (Ch).
- [21] QIU L, LIN H, LEUNG A K, et al. Putting their best foot forward: emotional disclosure on Facebook. [J]. Cyberpsychology Behavior and Social Networking, 2012, 15(10):569.
- [22] 胡兴安, 王立磊, 杨懿. 网络负面口碑对旅游目的地品牌资产稀释效应——熟悉度和易感性的调节作用[J]. 人文地理, 2015, 5:126-133.
- XU X A, WANG L L, YANG Y. Dilution effects of online negative word-of-mouth on brand equity for tourism destination: moderating effects of familiarity and susceptibility[J]. Human Geography, 2015, 5:126-133. (Ch).

## Study on tourism destination evaluation based on web reviews —— taking 31 provinces in China for example

GONG Jian<sup>1,2</sup>, YANG Shuyue<sup>1,2</sup>

(1. College of Urban and Environmental Science, Central China Normal University, Wuhan 430079, China;  
2. Wuhan Branch of China Tourism Academy, Wuhan 430079, China)

**Abstract:** In order to innovate the research method of tourism destination evaluation, this paper evaluates tourism destination from two dimensions, i. e. scale and quality, and proposes an Attention-Satisfaction Analysis (ASA) model based on tourism web text. 1,368,052 effective comments of 23,496 tourist attractions are collected and a foundation database of 31 provinces in mainland China is founded by integrating the comments from 8 main online travel platforms. The development level of provincial tourism destination is analyzed by modeling and visualization. The analysis result shows that: 1) network tourism attention rates present long-tail distribution and polarization characteristics, approximately 10% hot tourist attractions accounting for 80% online travel comments; spatial concentration of popular tourist attractions decreases from eastern regions to central regions and then to western regions; 2) there are significant differences between different kinds of tourist attractions in terms of online tourism attention and satisfaction rate; tourists hold different positive and negative emotional characteristics to different types of tourist attractions. Based on the analysis of the characteristics and the ASA positioning analysis result of 31 provinces, the tourism development optimization path is put forward for 31 provinces in China mainland.

**Key words:** web text analysis; Attention-Satisfaction Analysis model; tourism destination evaluation