

基于Tripadvisor的中国旅游地 国际关注度及空间格局

王 琨¹, 郭风华², 李仁杰^{1,3}, 傅学庆^{1,3}

(1. 河北师范大学资源与环境科学学院, 石家庄 050024; 2. 河北省科学院地理科学研究所, 石家庄 050011;
3. 河北省环境演变与生态建设实验室, 石家庄 050024)

摘 要: 用户贡献内容(UGC)已逐渐成为旅游行为与感知研究的重要数据源。区别于通常利用搜索引擎关键词数量描述网络关注度的方法, 本文引入电子社区层次结构为权重因子, 建立了基于社区UGC的旅游关注度模型, 能够灵活调节模型表达的重点, 优化计算结果。针对著名旅游电子社区Tripadvisor的研究发现, 国外社区用户对中国旅游的关注呈现3个典型特征: ①旅游关注集中在“长城、泰山、黄山、九寨沟、张家界”等少数旅游吸引物, 和“北京、香港、上海、桂林”少量目的地城市; 大量吸引物和目的地关注度较低, 呈现“长尾现象”与极化特征。②吸引物与目的地城市的关注空间具有明显耦合性, 关注度较高的吸引物多邻近或隶属于关注度较高的城市, 如桂林阳朔、北京长城、成都都江堰和九寨沟、杭州西湖等。③旅游关注空间整体呈现出由高到低的“东—中—西”格局, 与中国区域经济的“东—中—西”梯度格局基本耦合; 北京、香港、广州、深圳、上海、成都等关注中心也与区域经济中心一致。旅游资源禀赋、电子口碑传播模式、地理区位、经济水平和关注者国家的文化背景、经济发展状况、地理区位等是影响旅游者关注度及其空间格局变化的主要因素。旅游关注度模型旨在解决互联网用户对区域旅游关注的定量计算问题, 为基于互联网UGC的旅游地理学研究提供新思路。

关 键 词: 旅游关注度; 空间格局; 电子社区; 用户创建内容(UGC); Tripadvisor

doi: 10.11820/dlkxjz.2014.11.004

中图分类号: P28

文献标识码: A

1 引言

采用智能化手段处理互联网上的海量信息是智慧信息时代的重要特征。旅游业是一个信息敏感性产业, 旅游者从旅行前的检索、规划和决策, 到旅行中的行为与感知, 再到旅行结束后的评价和反馈, 都在生产与旅游相关的信息, 这些信息中包含大量用户贡献内容(UGC, User Generated Content)。众多的网络论坛、博客等电子社区和照片等多媒体分享网站都是UGC的生产平台。这类互联网应用也受到在线旅游者的欢迎(Xiang et al, 2010), 他们愿意与他人分享旅行经历(Akehurst, 2009), 其中, 有关旅游地和产品的评论可以直接影响旅游者决策(Lu et al, 2012)。UGC越来越被认为是可信的口碑传播形式(Stepchenkova et al, 2013),

这也促使旅游产业运营方式和旅游者行为模式发生根本性改变。因此, 通过智能化手段对信息进行处理, 了解UGC中包含的旅游关注度信息就显得非常必要。

旅游关注度与旅游满意度、旅游感知研究一样, 都用于描述旅游者对旅游活动相关区域、设施和服务等的态度与情感, 其研究结果可以服务于旅游政策制定、规划管理与决策。满意度与感知研究侧重表达已经体验过的旅游活动, 而关注度研究则不区分参与者是否真正经历了旅游活动, 因此还包括潜在旅游者的贡献。

中国学者在基于网络信息的旅游关注度研究中, 多数使用了百度搜索及其提供的百度指数模型(李山等, 2008; 龙茂兴等, 2011; 马丽君等, 2011; 林志慧等, 2012; 龙茂兴等, 2013); 另外也有学者使用

收稿日期: 2014-04; 修订日期: 2014-07。

基金项目: 国家自然科学基金项目(41101105, 41171105); 河北省高校重点学科建设项目; 河北省软科学研究计划项目(13406002D)。

作者简介: 王琨(1988-), 女, 河北元氏人, 硕士研究生, 研究方向为旅游地理信息挖掘, E-mail: vickyohara1988@163.com。

通讯作者: 李仁杰(1975-), 男, 河北鹿泉人, 教授, 博士, 研究方向为地理信息可视化与空间建模, E-mail: lrjgis@163.com。

Google搜索进行案例研究(王章郡等, 2011)。基于搜索引擎及其提供的固定模型研究具有方便、快捷的特点,但也存在模型参数较难调整,数据源多样化带来的分类和深度筛选困难等问题。

旅游电子社交网络为旅游休闲业的发展提供了机遇(Sparks et al, 2013),也为开展旅游关注度研究提供了海量数据,与基于网络信息的旅游关注度研究相比,数据更具旅游方向性。其生产的UGC相对于封闭式问卷能提供更丰富的数据内容,可用于分析旅游者满意度、印象感知,发现旅游引起的文化、生态变化特征(Ricker et al, 2013)。例如,将网络游记作为部分数据来源,获取旅游客流数据,研究跨界旅游空间格局及优化路径(彭红松等, 2014);利用旅游者上传的照片分析其对旅游地的关注度(王守成等, 2014),或基于公众上传照片分析城市环境感知的空间模式(Feick et al, 2013)等。

Tripadvisor^①是全球第一的旅游评论网站,拥有超过1000万的注册会员和每月2亿的活跃用户数,已成为一个大型在线“数据库”,积累了大量关于旅游的UGC。本文旨在建立一个基于电子社区内容的旅游关注度计算模型,以Tripadvisor中关于中国旅游的文本信息为基础,开展区域旅游关注度的实证研究,并结合GIS数据库挖掘关注度的空间格局,为旅游地理学利用UGC进行研究提供参考。

2 电子社区的旅游关注度模型

电子社区中有关某一事物或现象的论坛、话题和评论数量的多少是社区成员关注该事象热度高低的最直接表现。因此,电子社区中的区域旅游关注度可以通过相关社区中的论坛、话题和评论的数量多少进行初步表征。

2.1 关注度模型的主要因子

2.1.1 旅游地名的使用频率

城市名称、旅游地名称等是旅游电子社区用户创建相关讨论区、发表话题和评论的关键词,本文将相关旅游文本中出现城市 and 旅游地名称的频次作为构建区域旅游关注度模型的核心参数。由于社区用户关注的旅游地名频次差异较大,为了便于关注度模型的表达和计算结果的区域间对比,需要对地名统计频次进行标准化处理。在保证原始数据数值变化特征不变的前提下,模型中以论坛、话

题、评论作为不同的统计版块,基于各版块同类地名的总频次,对该版块文本中的旅游地名、城市地名出现频次进行标准化处理,使结果落到[0~1]区间,得到各旅游地或城市名称的出现频率(x^*),具体公式为:

$$x^* = \frac{x}{\sum x_i} \quad (1)$$

式中: x 为该版块出现的某一地名的频次; $\sum x_i$ 为某版块出现的同类地名的总频次。旅游地或城市地名的出现频率将作为电子社区对其关注度的基本指数,为后续基于社区层次结构的加权计算和不同量级数据的关注度集成奠定基础。

2.1.2 电子社区的层次结构

旅游电子社区一般拥有清晰的层次结构,并在各层次上由管理员或用户划分出不同的主题版块,以吸引相关用户针对同一主题或内容开展交流探讨。层次结构与版块的设计有利于信息分类整合和搜集,例如Tripadvisor旅游社区的层次结构是“Forum(论坛)—Topic(话题)—Review(旅评)”(图1),论坛是Tripadvisor的基本单元,注册用户可以创建或选择已存在的论坛,在相关论坛下发布话题,或在话题下发表评论。通过系统分析发现,由于旅游社区的特殊性,Tripadvisor用户大量使用旅游目的地名称或相关标志性地物名称作为论坛、话题和评论的关键词,说明地名可以反映用户对旅游目的地的关注情况。

鉴于社区层次结构划分对旅游描述信息组织的重要作用,用户在社区不同层次上使用地名对相

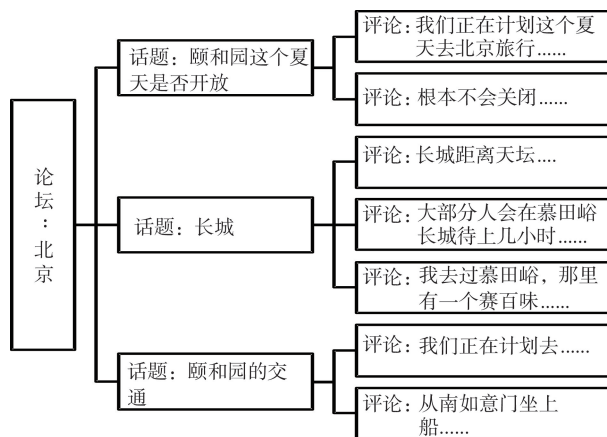


图1 Tripadvisor旅游社区的层次结构图(以论坛Beijing为例)

Fig.1 Hierarchical structure of the tourism community of Tripadvisor

① 网址:www.tripadvisor.com

关区域旅游关注度的贡献也不一样。在综合电子社区层次结构特征和Tripadvisor的旅游信息特点的基础上,分别赋予论坛0.6、话题0.25、评论0.15的权重分值,用于旅游关注度模型中的地名频率加权。

2.2 旅游吸引物和目的地关注度模型

在旅游关注度模型中涉及对旅游吸引物的关注和旅游目的地的关注。狭义的旅游吸引物是指有形的旅游资源,广义的旅游吸引物还包括旅游服务、社会制度等无形旅游资源。本文主要关注的是狭义的旅游吸引物,即有形的自然或人文旅游资源。旅游目的地是旅游吸引物和服务设施的所在地。旅游关注度模型可进一步区分为旅游吸引物关注度和旅游目的地关注度。在较大空间尺度上,旅游者的认知中往往难以明确区分旅游吸引物与旅游目的地,旅游者对旅游吸引物的关注也是对旅游目的地关注的一种间接表现,因此,旅游目的地关注度模型中应该考虑吸引物关注度的贡献。

2.2.1 旅游吸引物关注度模型

旅游吸引物是本文旅游关注度研究的最小空间单元,所以其关注度不受下级单元的影响,根据旅游吸引物相关地名使用频率和旅游电子社区层次结构设计的吸引物关注度模型为:

$$F_1 = W_A \times P_A + W_B \times P_B + W_C \times P_C \quad (2)$$

式中: F_1 为旅游电子社区用户对特定区域的旅游吸引物的关注度; P_A 、 P_B 、 P_C 分别为旅游电子社区中“论坛”、“话题”和“评论”文本中出现的旅游吸引物名称的频率; W_A 、 W_B 、 W_C 分别为“论坛”、“话题”和“评论”不同层次对旅游关注度地名贡献权重。

2.2.2 旅游目的地关注度模型

旅游目的地关注度模型的主体与吸引物关注度模型一致,主要考虑旅游目的地相关地名使用频率和旅游电子社区层次结构对旅游者关注度的影响。另外,目的地关注度模型应考虑旅游吸引物对区位隶属目的地的贡献部分。在多种空间尺度下(城市、省区、国家等)计算目的地关注度时,还应考虑小空间尺度单元对其区位隶属的大空间尺度单元的贡献。例如,基于城市单元的旅游目的地关注度计算公式如下:

$$F_2 = \ln\{(1 - \alpha) \times (W_A \times Q_A + W_B \times Q_B + W_C \times Q_C) + \alpha \sum F_1\} \quad (3)$$

式中: F_2 为旅游电子社区用户对特定旅游目的地城市的关注度。 Q_A 、 Q_B 、 Q_C 分别为论坛、话题、评论文本中的目的地城市名称的出现频率; α 为旅游吸引

物对其区位隶属目的地的关注度的贡献权重。

旅游吸引物对其区位隶属目的地的关注度的贡献权重通过Tripadvisor网站本身的吸引物与目的地共同出现的概率进行计算。贡献权重因子等于旅游吸引物和目的地名称同时出现的旅游文本数量之和,除以目的地名称出现的文本数。随机选取自然和人文旅游吸引物各10个,对旅游文本中出现的吸引物和目的地名称进行统计计算,求平均值,得出Tripadvisor中旅游吸引物对目的地关注度的贡献因子 α 为5.41%。

3 关注度模型计算中的数据库构建

3.1 旅游文本数据库

本文选取全球最大的旅游电子社区Tripadvisor作为对象,采用网络爬虫程序自动采集社区用户发布的旅游地UGC作为数据来源,构建与研究区域相关的旅游文本数据库。首先,通过Tripadvisor网站嵌入的内部搜索引擎检索与“China”相关的旅游文本信息,通过爬虫采集规则筛选查询结果中的旅游评论和话题。旅游评论信息主要采集“标题、回复标题、时间、评论文本、详细信息(含所属话题)”5项内容,共采集中国旅游相关评论文本371939条;话题信息主要采集话题名称及其隶属的论坛名称等,共75214条。由于话题信息中已经包括了论坛名称,因此不再单独采集论坛名称信息。通过对话题信息进行解析与整理,共获取7767个与“China”相关的旅游论坛名称。对旅游评论文本进一步去重等整理,得到有效评论371092条,建立了Tripadvisor社区的中国旅游文本数据库。

3.2 中英文旅游目的地与吸引物地名数据库

中国旅游资源丰富,分布区域广泛,各类旅游景区和吸引物已达到2万多个,但众多的旅游景点主要接待的是国内旅游者。在不影响研究成果质量的基础上,为尽量减少数据样本,本文采用国家旅游局英文官方网站推荐的69个旅游吸引物为研究样本,采用地市级以上670个城市作为旅游目的地研究样本,构建计算中国旅游关注度的中英文地名数据库。

数据库中的英文地名主要用于英文旅游文本信息挖掘的关键词匹配,对应的中文地名则主要用于和基础地理信息数据库建立空间链接,实现旅游关注度的空间可视化与格局分析。目的地城市名

称中的专名在英文社区中的使用相对比较规范,主要采用国际通用拼音表示法。旅游吸引物名称的使用则比较灵活,可能导致关键词统计中的不确定性。由于旅游资源的自然、文化属性丰富,一个景区可能有多个英文表达法;而且,电子社区用户发布的旅游UGC中还经常使用简称、别名等表达方式。研究中考虑了英文地名大小写的转换、词语倒装、国际惯例与国内用法的差异、标准名称与别名等因素,尽量减小多样化吸引物地名表达对地名统计精度的影响。

数据库中的空间信息主要来自两个方面。目的地城市空间信息来自国家测绘科学数据共享网下载的中国比例尺1:500万行政区图。旅游吸引物空间信息则通过综合百度百科和维基百科中的经纬度描述信息建立地址文本文件,并利用百度、谷歌地图进行位置验证。将确认准确空间信息之后的吸引物地址信息导入地名数据库,通过中文标准名称与吸引物其他属性建立关键词链接,即可完成旅游吸引物信息的空间化。

4 Tripadvisor的中国旅游关注度格局及影响机理

4.1 Tripadvisor对中国旅游关注度统计特征

旅游文本中的地名出现频次是信息发布者对区域旅游关注的直接反应。将7767个论坛名称文本、75214条话题文本和371092条评论文本分别合并为3个文本集合。以旅游目的地和吸引物地名数据库中的英文地名为基础构建关键词库,统计Tripadvisor不同层次文本中关于中国旅游吸引物和目的地名称出现的频次,如图2所示。

图2a-2c显示了官方推荐的旅游吸引物地名在Tripadvisor的论坛、话题和评论文本中的频次分布情况。整体上出现频次较高的旅游吸引物地名主要是长城、九寨沟、黄山、张家界、泰山等。其中,长城在论坛名称和话题文本中均是出现频次最高的旅游地名;泰山则是评论文本中出现频次最高的旅游地名。经过关注度模型计算,Tripadvisor对中国旅游吸引物的关注度分布曲线如图3所示,其中,

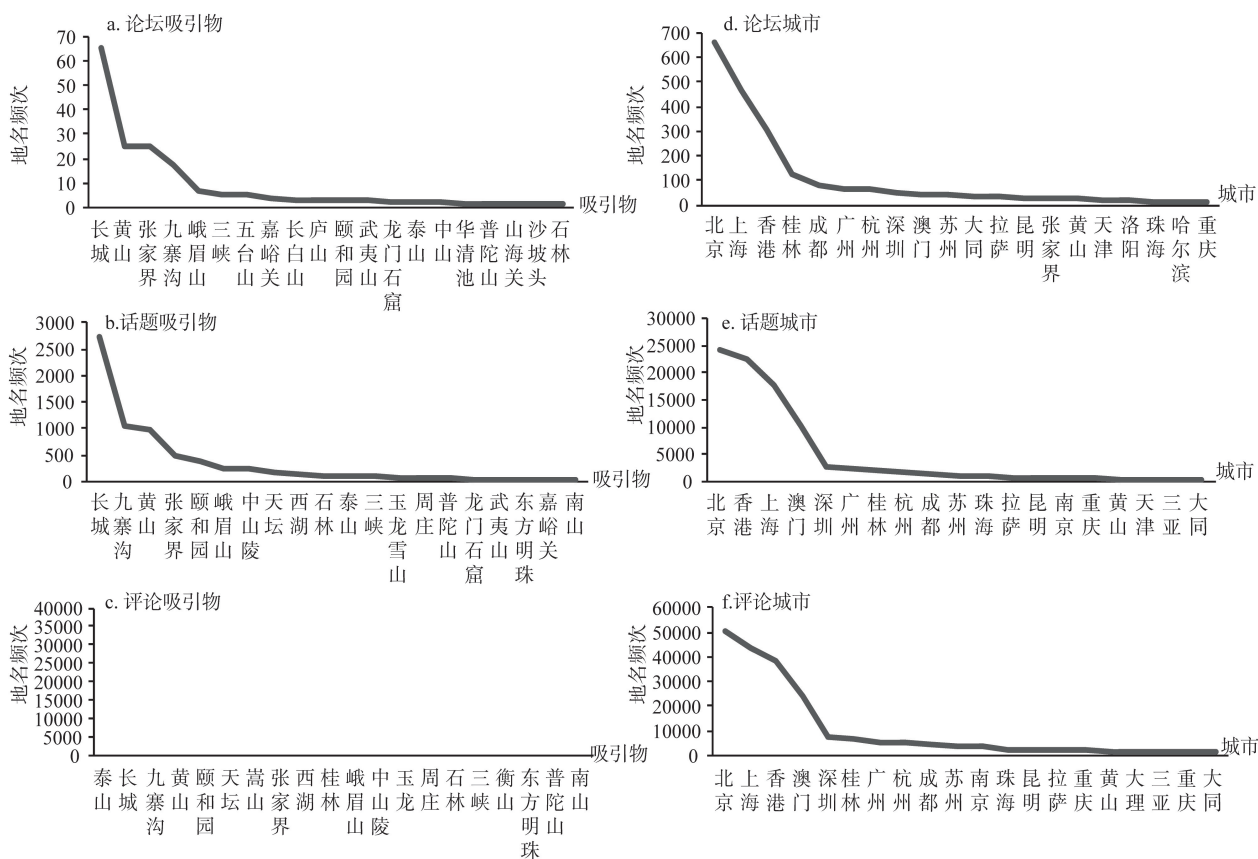


图2 Tripadvisor论坛、话题和评论中的地名频次情况分布曲线

Fig.2 Frequency of the place names that appeared in the forums, topics, and comments of Tripadvisor: (a) by attraction names (b) by destination city names

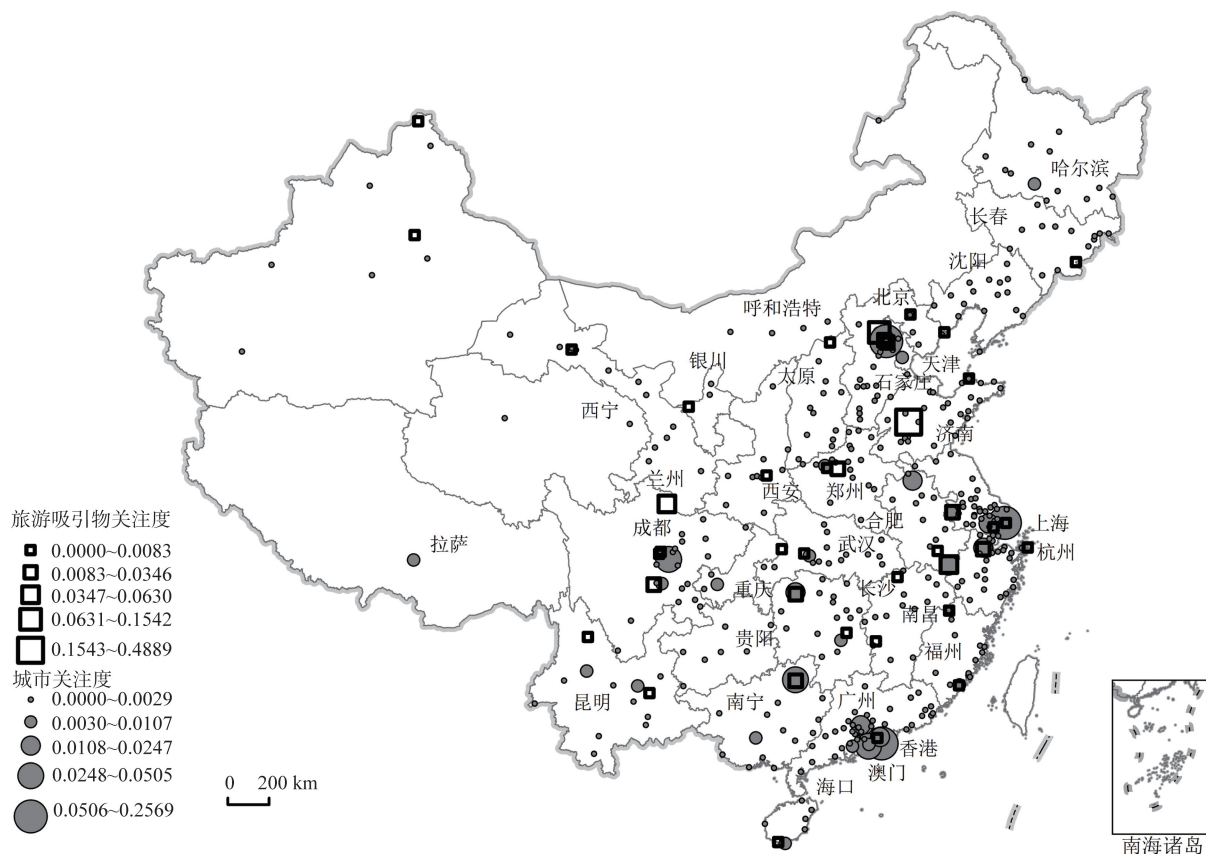


图5 Tripadvisor网站对中国旅游吸引物和目的地城市的关注度空间分布格局

Fig.5 Spatial pattern of the attention degree for the Chinese tourism attractions and destination cities

东北地区旅游关注度较低。总体看来,电子社区中旅游吸引物关注度空间格局差异较大,集中分布于少量著名景点。关注度较高的旅游吸引物主要分布在华北、华南和西南地区,其中华北地区的长城的关注度最高,这与Google发布的2013中国旅游国际网络关注度研究报告中指出的“长城是各国网民心中最具有代表性的中国旅游元素”表述相一致,西南地区的九寨沟关注度最高,华东地区的黄山关注度最高。

从旅游目的地城市层面上看,电子社区对中国城市的关注范围很大,覆盖中国大部分地区,但明显东多西少。华北地区关注点覆盖范围广,但多数关注集中于北京。华东地区、华南地区与华北地区相比分布的聚集性更为明显,而且华南地区除广州以外还有香港、深圳等关注热点。西部关注度较高的城市主要分布在西南地区,成都、西昌、丽江、大理、昆明、自贡等城市也表现出很强的聚集特征,这与旅游电子社区中的路线推荐有很大关系。拉萨与西部其他地区相比关注度较为突出,这与西藏目前旅游业快速发展的现状相符。

Tripadvisor用户的中国旅游关注度总体呈现“东—中—西”由高到低的空间分布格局。关注度较高区域主要分布在华北、华南、华东和西南,西北、华中和东北地区大部分区域关注度较低。北京是华北地区的关注度中心;香港、广州、深圳等为华南地区的关注度中心;上海为华东地区的关注度中心,成都则是西南地区的关注度中心。研究结果与马耀峰等(2001)根据中国入境旅游流统计资料得出的分区研究结论具有很强的一致性。

Tripadvisor对中国旅游的关注呈现高度空间聚集现象。东部关注度较高的城市和旅游吸引物分布集中,主要围绕北京、上海、广州等地集中分布,环京津、长三角、珠三角成为社区用户的高关注度聚集区域。另外,成都周边也形成了较为明显的关注度聚集现象。旅游吸引物与目的地城市的关注度具有明显的空间耦合性,关注度较高的旅游吸引物在空间上多邻近或隶属于关注度较高的城市,如桂林阳朔、北京长城、成都周边都江堰和九寨沟、杭州西湖、苏州周庄等。关注度较高的旅游地也表现出空间上的聚集性,例如都江堰—峨眉山—九寨沟。

沟形成的旅游景区链,周庄—西湖等形成的旅游关注热点等,这说明网络空间中的外国旅游者对中国旅游目的地城市的关注度受其所辖旅游吸引物关注度的影响。

通过观察发现,关注度较高地区可大致分为两类,一类为经济发达地区,如上海、香港、广东等地;另外一类是其旅游业体现强烈的民族特色或地方自然风光的区域,如四川、云南。关注度较低的地区也可以分为两类,一类为具有旅游业发展的基础自然人文条件,但受观念、认识、经济和基础设施建设以及开发不足等因素制约(赵欣等,2007),旅游辐射能力急需加强的地区,如新疆和甘肃;另一类是不受经济发展水平等因素的制约,但因缺少丰富的特色旅游资源而对外国人的吸引力较弱的地区,如黑龙江、吉林。

4.3 Tripadvisor 中国旅游关注者的空间分布情况及不同国家对中国旅游关注度的差异分析

为了便于深入了解游客的行为模式,了解客源国的市场状况,本文对电子社区中各个国家用户的发文数量进行统计分析,了解 Tripadvisor 电子社区中国旅游的关注者的空间分布状况,如图6所示。从图6可以看出,发文数最多的是美国、澳大利亚、加拿大和英国,其次是新加坡、马来西亚、菲律宾、印度、泰国等国。总体来说,对中国旅游关注较多的国家大部分分布于北半球,且主要集中于北美、东南亚和西北欧地区,欧洲大部分地区及中东地区

对中国旅游关注也相对较多,其在空间分布上表现出邻近性;非洲大部分地区 and 南美洲对中国旅游的关注较少。

通过对关注者空间分布情况的分析确定了对中国关注较多的国家,不同国家的 Tripadvisor 用户对中国旅游地的关注度也存在差异,如图7所示。分别对美国、澳大利亚、加拿大、英国、新加坡和马来西亚六国对中国旅游的关注度进行计算,选取了各个国家关注度最高的前八位吸引物地名和城市地名,并对其进行统计特征分析。美国、澳大利亚、加拿大、英国4个国家中,对中国旅游吸引物地名“长城”的关注度最高,黄山、天坛、颐和园次之,九寨沟、西湖、玉龙雪山等地关注度排在上述4个景区之后。新加坡和马来西亚关注度最高的旅游吸引物地名为九寨沟,其次是长城,而对天坛、西湖等地名关注度相对偏低,且不一致。在目的地城市中,美国、澳大利亚、加拿大、英国关注度最高的目的地城市为北京,上海次之,对桂林、深圳、广州、成都、杭州等地的关注也占据了很大比重。新加坡和马来西亚关注度最高的城市为上海,北京次之,且对广州、深圳、珠海、杭州的关注度也普遍较高。在对其他城市的关注中,美国、澳大利亚、加拿大、英国对桂林和深圳关注度远高于其他城市,新加坡和马来西亚则对包括珠海、广州、深圳在内的珠三角城市群兴趣更大。从关注度变化特征来看,美国、澳大利亚、加拿大、英国4个国家的关注度变化较

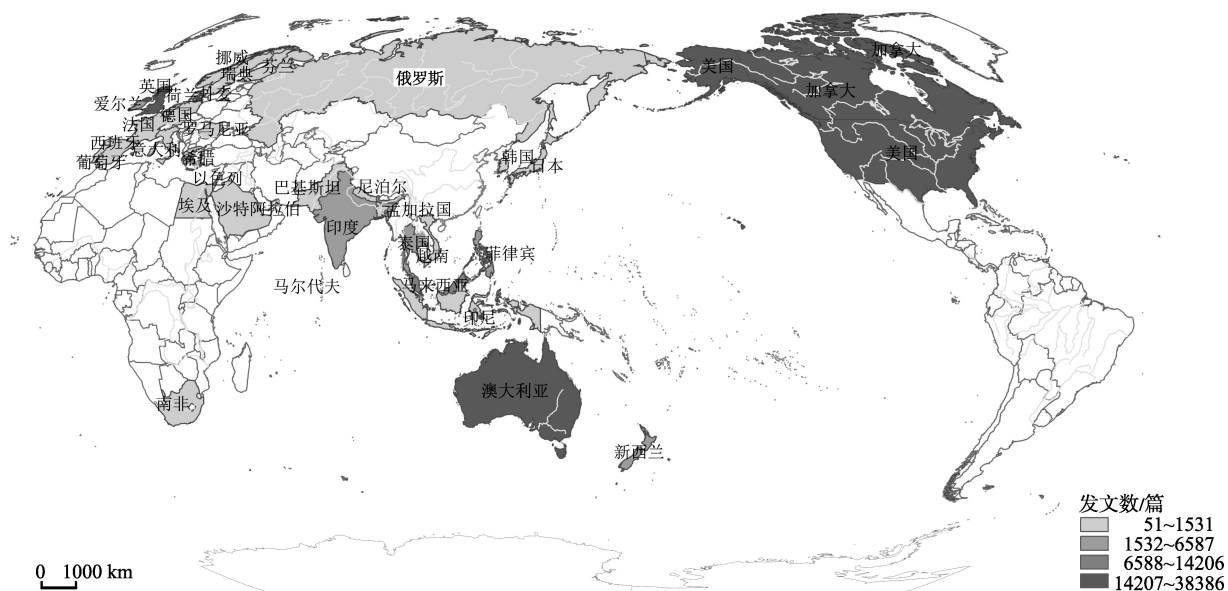


图6 Tripadvisor 网站中国旅游关注者的空间分布格局

Fig.6 Spatial distribution of the followers of the Chinese tourism in Tripadvisor

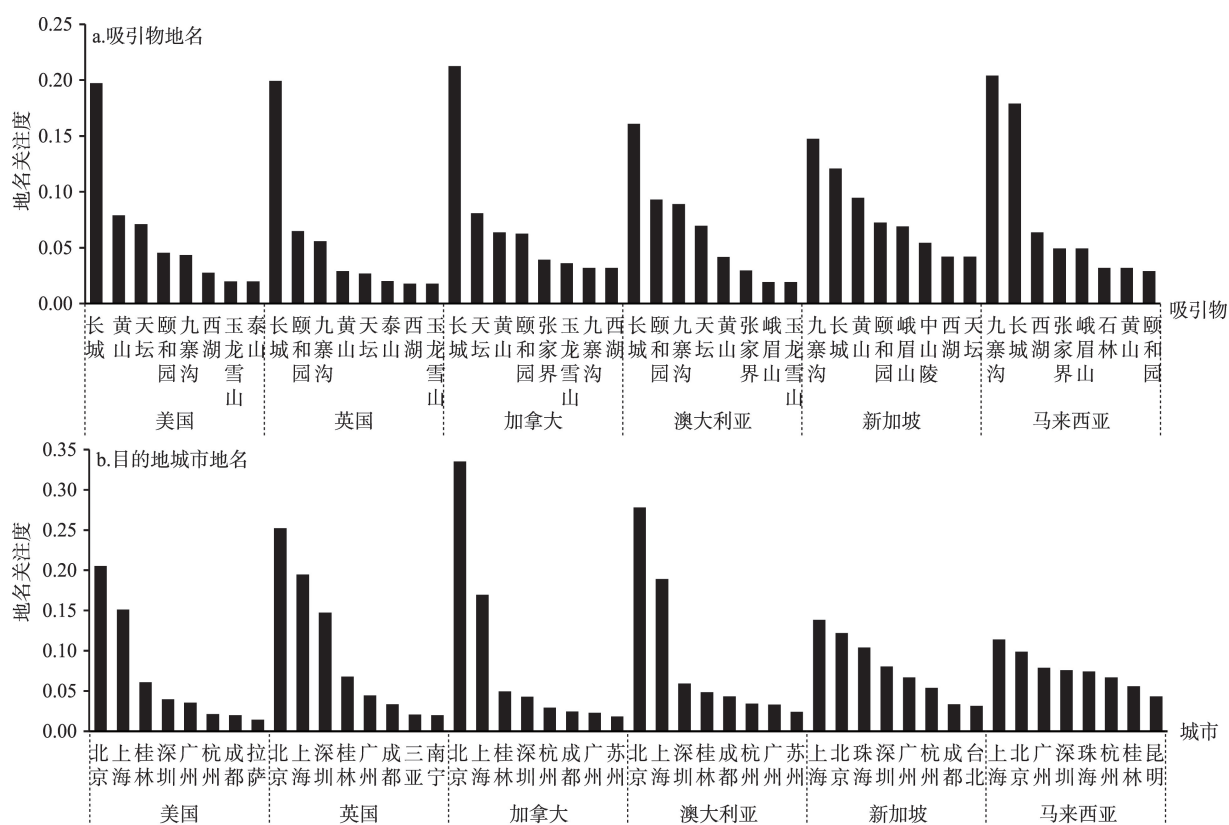


图7 Tripadvisor网站中不同国家对中国旅游关注度的统计特征

Fig.7 Statistical characteristics of the Chinese tourism attention degree of different countries in Tripadvisor:

(a) by attraction names (b) by destination city names

为激烈,新加坡和马来西亚的变化相对平缓,这说明美国、澳大利亚、加拿大、英国对中国旅游的关注度更为集中。

总体来说,北半球发达国家对中国旅游地的关注度普遍较高,且表现出邻近特征。其中,美国、澳大利亚、加拿大、英国4个国家关注的旅游地表现出很强的耦合性,关注度最高的旅游地均为北京和长城。除此之外,在4个国家其他关注度较高的吸引物地名和城市地名中,长城、黄山、天坛、颐和园、西湖、泰山为世界文化遗产,北京、上海、深圳、广州、杭州等城市又都以人文景观著称。因此,人文旅游吸引物和城市的数量占绝对优势,这也说明美国、澳大利亚、加拿大、英国的关注者对人文景观的关注度要远高于自然风光。这与他们的欧美文化背景有很大关系,如北京城内的天坛、颐和园,杭州的西湖等拥有悠久历史文化和丰富文物古迹的旅游吸引物充分满足了欧美关注者对中华文化的猎奇心理。

新加坡和马来西亚对中国旅游地的关注同样表现出耦合性,均对九寨沟和上海的关注度最高。

因为新、马两国同中国有特殊的人种、血缘、文化关系(周鹏等, 2004),对历史文化景观的猎奇需求较弱,所以在旅游吸引物层面上,对欧美国家关注度最高的历史文化吸引物的关注度不如九寨沟等自然风光吸引物突出。从城市层面上看,新加坡和马来西亚对西南地区旅游城市和珠三角城市群更为关注,这说明西南和华南地区的东南亚旅游市场较为成熟,研究结果与张娟娟等(2011)证实的马来西亚、新加坡为昆明、成都主要客源市场的研究结果相似,也与广州“辐射泛珠三角、影响东南亚”的旅游发展目标契合。

4.4 关注度分布曲线与空间格局的驱动机制分析

从社会宏观视角探究,旅游流受属性耦合驱动的主要有“旅游资源”、“地区经济”、“交通区位”和“接待能力”(郑鹏, 2013)。旅游资源丰度、经济发展水平和地理区位均是影响游客网络关注度空间分布的主要因素(马丽君等, 2011)。在深入研究 Tripadvisor 中国旅游关注度分布曲线与空间格局基础上,结合中国旅游资源特点和互联网空间信息传播模式,对比参考其他研究者的同类研究成果,分析

总结了影响国外电子社区用户对中国旅游关注度的以下4个重要因素:

(1) 旅游资源禀赋与特色是 Tripadvisor 用户对中国旅游关注的基础动力

旅游资源禀赋是影响中国区域旅游业发展差异的主要因素之一(敖荣军等, 2006), 也是吸引旅游者关注的主导因素。世界遗产、旅游风景名胜、森林公园、自然保护区、地质公园、历史文化名城等类型的旅游资源是旅游者关注的重点。Tripadvisor 用户对中国旅游关注度最高的旅游吸引物中, 长城、泰山、黄山、九寨沟、张家界都是世界自然或文化遗产; 关注度较高的旅游目的地城市中, 北京、成都、桂林、杭州和苏州均为世界著名的风景游览城市 and 历史文化名城, 拥有有丰富的自然、文化遗产和国家级风景名胜区、森林公园、地质公园等。因此旅游资源禀赋与特色成为电子社区用户对中国旅游关注的基础动力。

(2) 中国旅游业的国际化是旅游关注度总体特征的外源动力

国外电子社区对中国旅游关注度的不断攀升与中国旅游业的国际化发展密切相关。北京、上海、深圳和广州是中国旅游国际化水平最高的四大城市, 桂林、成都、杭州和苏州等城市的旅游国际化水平也比较高, 而且东部沿海地区是旅游国际化高水平集聚区(闻飞等, 2012)。城市的旅游国际化水平有助于为外国游客了解中国旅游、适应旅游环境提供良好的软硬件条件, 成为中国旅游关注度总体特征的外源动力。

(3) 电子口碑传播模式是关注度数据分布曲线的内在动力

电子社区使传统口碑相传变成电子口碑, 而且社区用户更相信以电子口碑方式传播的旅游信息(Arsal, 2010), 认为由其他消费者发表的具有明确指向的旅游信息值得信任。旅游者开始从传统的旅游项目选择方式转变为依据网络上获取的电子口碑做出决策。基于“六度空间理论”^②, 旅游信息在好友间的分享、转发可以实现高效率口碑相传, 口碑好的目的地和旅游吸引物关注度会快速上升, 这也是关注度曲线均呈现“长尾”特征的内在动力。

(4) 经济发展水平与地理区位是影响关注度空间格局的核心动力

Tripadvisor 对中国旅游吸引物和目的地城市的关注空间覆盖了全国大部分区域, 但关注度较高且出现空间聚集的区域主要是长三角、珠三角、京津冀, 与中国经济的区域结构有较好的耦合度。这一特征与马耀峰等(2003)研究证实的京—沪—穗、京—沪—深、京—桂—沪等入境旅游流金三角框架也有较高的一致性。经济发达的区域中心城市多为关注度较高的旅游目的地城市, 如上海、香港、澳门、广州、深圳等。由于地理位置的邻近性也是影响外国游客选择旅游目的地的重要因素, 高关注的区域中心城市周边也聚集了大量中低关注度的目的地城市。例如, 成都及其周边城市和旅游地的网络关注度空间格局, 与史春云等(2007)研究证实的四川省“旅游区域核心—边缘空间结构”非常相似。总体来看, 经济发展水平与地理区位是影响关注度空间格局的核心动力。

(5) 关注者国家的文化背景、经济发展状况、地理区位是影响关注度空间格局变化的深层动力

电子社区中的关注者覆盖全球大部分地区, 但关注较多的地区主要集中于北半球发达国家, 说明对中国旅游的关注度与各国经济发展水平有很大关系。不同国家与不同地区、不同种类旅游地的亲疏程度不同, 因为不同文化背景的关注者有不同的旅游需求。以美国、澳大利亚、英国和加拿大为代表的欧美国家对中国人文景观的关注度要高于自然风光, 对历史文化名城北京的关注度高于作为金融中心的上海, 而新加坡和马来西亚对自然风光更感兴趣。另外, 新加坡和马来西亚对城市的关注度主要集中于华东和华南地区, 这也是受地理区位上的邻近性的影响。因此, 关注者国家的文化背景、经济发展状况、地理区位是影响关注度空间格局变化的深层动力。

5 结论与讨论

5.1 结论

区别于基于搜索引擎关键词搜索统计的网络关注度计算方法, 本文提出了一种针对电子社区用户的区域旅游关注度计算模型。该模型在引入加权参数方面具有更好的灵活性, 可以通过社区内容的层次结构有效调节、修正关注度模型计算结果;

^②“六度空间理论”也称“小世界现象”, 理论假设世界上所有互不相识的人只需要很少中间人就能建立起联系, 是社交网络的理论基础。

另外,由于统计结果与数据源文本、发文者信息等直接相关,相关信息有助于更好的分析模型结果背后的驱动机制。模型为量化表达特定用户对区域旅游的关注情况提供了方法,为发现旅游关注的特征、空间格局、电子社区中的客源市场分布和不同国家的电子社区用户关注情况差异分析奠定了基础。

基于Tripadvisor社区旅游文本的中国旅游关注度计算结果发现,国外社区用户对中国的旅游关注呈现以下3个典型特征:

(1) 旅游关注度的“长尾现象”与极化特征。Tripadvisor社区对中国旅游吸引物与目的地城市的关注度曲线呈现典型的“长尾”特征;高关注度区域集中在长城、泰山、黄山、九寨沟和张家界等少数旅游地,以及北京、香港、上海、桂林等少量城市,呈现关注极化特征。该研究结果与Pan等(2011)美国公民对跨境中国旅游地形象认知研究中发现的形象词汇长尾曲线特征基本一致。

(2) 旅游吸引物与目的地城市的关注度具有明显的空间耦合性。关注度较高的旅游吸引物在空间上多邻近或隶属于关注度较高的城市,如桂林阳朔、北京长城、成都周边都江堰和九寨沟、杭州西湖、苏州周庄等等。关注度较高的旅游地也表现出空间上的聚集性,例如都江堰—峨眉山—九寨沟形成的旅游景区链,周庄—西湖等形成的旅游关注热点等。

(3) 旅游关注度的空间聚集及其与区域经济梯度的耦合特征。Tripadvisor社区对中国旅游的关注空间中,环京津、长三角、珠三角城市群成为高关注度聚集区,该研究结果与汪德根等(2011)得出的“广东、北京和上海是中国东部沿海地区的旅游产业综合竞争力最强的省份”研究结果相一致。成都周边等也形成了明显的较高关注度聚集现象。关注度整体上呈现“东—中—西”由高到低的空间分布格局,与中国区域经济呈现的“东—中—西”经济梯度格局高度相关;北京、香港、广州、深圳、上海、成都等旅游关注度中心也与区域经济中心相一致。关注度空间格局说明了区域经济基础与旅游发展水平的耦合关系。

5.2 讨论

受限于数据量和计算工作量,本文仅对国家旅游局官方网站对外推荐的旅游吸引物和地级以上城市目的地进行了研究,但该方法可用于对任意旅游吸引物的关注度分析,也可扩展至下级城镇、上

级省域,或各类自然和经济区以及国家等多种空间尺度。

本文仅关注了吸引物和目的地城市相关地名的使用情况,未来研究将从以下3个方面开展:一是通过文本挖掘进一步分析旅游者对特定目的地和吸引物的态度和情感,以在更多维度上开展旅游地理认知研究;二是通过建立更为详细的旅游景观和目的地内的主要标志地物数据库,实现更小空间尺度上的关注度分析,突破当前关注度模型仅从整体上描述吸引物和目的地关注量度的限制;三是充分挖掘文本数据中的时间信息,开展关注度的时空格局与驱动机制研究,为区域旅游规划和空间决策服务。

参考文献(References)

- 敖荣军, 韦燕生. 2006. 中国区域旅游发展差异影响因素研究: 来自1990-2003年的经验数据检验. 财经研究, 32(3): 32-43. [Ao R J, Wei Y S. 2006. A study on the regional tourism resources and the unbalanced development of the tourism industry in China. Journal of Finance and Economics, 32(3): 32-43.]
- 李山, 邱荣旭, 陈玲. 2008. 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. 地理与地理信息科学, 24(6): 102-107. [Li S, Qiu R X, Chen L. 2008. Cyberspace attention of tourist attractions based on Baidu index: temporal distribution and precursor effect. Geography and Geo-Information Science, 24(6): 102-107.]
- 林志慧, 马耀峰, 刘宪锋, 等. 2012. 旅游景区网络关注度时空分布特征分析. 资源科学, 34(12): 2427-2433. [Lin Z H, Ma Y F, Liu X F, et al. 2012. Spatial and temporal features of network attention of scenic areas. Resources Science, 34(12): 2427-2433.]
- 龙茂兴, 孙根年, 龙珍付. 2013. 遵义红色旅游网络关注度的客流响应研究. 地理与地理信息科学, 29(5): 98-101, 118. [Long M X, Sun G N, Long Z F. 2013. Tourist flow's response to degree of consumer network attention to Zunyi tourism. Geography and Geo-Information Science, 29(5): 98-101, 118.]
- 龙茂兴, 孙根年, 马丽君, 等. 2011. 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. 地域研究与开发, 30(3): 93-97. [Long M X, Sun G N, Ma L J, et al. 2011. 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. Areal Research and Development, 30(3): 93-97.]
- 马丽君, 孙根年, 黄芸玛, 等. 2011. 城市国内客流量与游客网络关注度时空相关分析. 经济地理, 31(4): 680-685. [Ma L J, Sun G N, Huang Y M, et al. 2011. A correlative

- analysis on the relationship between domestic tourists and network attention. *Economic Geography*, 31(4): 680-685.]
- 马耀峰, 李旭. 2003. 中国入境游客旅游选择模式研究. 西北大学学报: 自然科学版, 33(5): 575-580. [Ma Y F, Li X. 2003. A study of China inbound tourists' selection patterns. *Journal of Northwest University: Natural Science Edition*, 33(5): 575-580.]
- 马耀峰, 李永军. 2001. 中国入境后旅游流的空间分布研究. 人文地理, 16(6): 44-46, 35. [Ma Y F, Li Y J. 2001. Research on spatial region-division of inbound tourist flow in China. *Human Geography*, 16(6): 44-46, 35.]
- 彭红松, 陆林, 路幸福, 等. 2014. 基于旅游客流的跨界旅游区空间网络结构优化: 以泸沽湖为例. 地理科学进展, 33(3): 422-431. [Peng H S, Lu L, Lu X F, et al. 2014. Spatial network structure and optimization of cross-border tourism area based on tourist flow: a case study in Lugu Lake. *Progress in Geography*, 33(3): 422-431.]
- 史春云, 张捷, 尤海梅, 等. 2007. 四川省旅游区域核心: 边缘空间格局演变. 地理学报, 62(6): 631-639. [Shi C Y, Zhang J, You H M, et al. 2007. Core-periphery spatial structure and its evolution of tourism region in Sichuan Province. *Acta Geographica Sinica*, 62(6): 631-639.]
- 汪德根, 陈田. 2011. 基于竞争力评价的区域旅游产业发展差异: 以中国东部沿海三大旅游圈为例. 地理科学进展, 30(2): 249-256. [Wang D G, Chen T. 2011. Differences of regional tourism development based on competitiveness evaluation: a case study of the three circles in eastern coastal China. *Progress in Geography*, 30(2): 249-256.]
- 王守成, 郭风华, 傅学庆, 等. 2014. 基于自发地理信息的旅游地景观关注度研究: 以九寨沟为例. 旅游学刊, 29(2): 84-92. [Wang S C, Guo F H, Fu X Q, et al. 2014. A study of the spatial patterns of tourist sightseeing based on volunteered geographic information: the case of the Jiuzhai Valley. *Tourism Tribune*, 29(2): 84-92.]
- 王章郡, 方忠权, 杜坤. 2011. 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. 地域研究与开发, 30(5): 112-117. [Wang Z J, Fang Z Q, Du K. 2011. Temporal-spatial evolvement characteristics of cyberspace attention index of self-driving tours in China: based on Google insight. *Areal Research and Development*, 30(5): 112-117.]
- 闻飞, 王娟. 2012. 中国热点旅游城市旅游国际化水平定量评价研究. 人文地理, 27(2): 145-150. [Wen F, Wang J. 2012. Quantitative evaluation research on tourism internationalization level of hot tourism cities in China. *Human Geography*, 27(2): 145-150.]
- 张娟娟, 马耀峰. 2011. 西安、成都、昆明入境旅游市场动态演化对比分析. 资源开发与市场, 27(8): 736-739. [Zhang J J, Ma Y F. 2011. Contrast analysis on inbound tourism market evolution and development in Xi'an, Chengdu and Kunming. *Resource Development & Market*, 27(8): 736-739.]
- 赵欣, 陈建波, 刘永泉. 2007. 西北地区旅游业发展现状及对策浅议. 新疆师范大学学报: 自然科学版, 26(3): 303-305. [Zhao X, Chen J C, Liu Y Q. 2007. The development and plan of tourism in Western region. *Journal of Xinjiang Normal University: Natural Science Edition*, 26(3): 303-305.]
- 郑鹏. 2013. 基于社会宏观与个体微观视角的入境旅游流驱动模式. 地理科学进展, 32(10): 1501-1510. [Zheng P. 2013. Driving force model for inbound tourist flow from macroscopic and microcosmic perspectives. *Progress in Geography*, 32(10): 1501-1510.]
- 周鹏, 任建兰. 2004. 中国入境旅游客源市场格局及发展预测. 地域研究与开发, 23(5): 68-72. [Zhou P, Ren J L. 2004. Analysis and anticipation on overseas inbound guest-market of China. *Areal Research and Development*, 23(5): 68-72.]
- Akehurst G. 2009. User generated content: the use of blogs for tourism organizations and tourism consumers. *Service Business*, 3(1): 51-61.
- Arsal I, Woosnam K M, Baldwin E D, et al. 2010. Residents as travel destination information providers: an online community perspective. *Journal of Travel Research*, 49(4): 400-413.
- Feick R, Robertson C. 2013. A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems*. doi: 10.1016/j.compenvurbsys.2013.11.006.
- Lu W, Stepchenkova S. 2012. Ecotourism experiences reported online: classification of satisfaction attributes. *Tourism Management*, 33(3): 702-712.
- Pan B, Li X R. 2011. The long tail of destination image and online marketing. *Annals of Tourism Research*, 38(1): 132-152.
- Ricker B A, Johnson P A, Sieber R E. 2013. Tourism and environmental change in Barbados: gathering citizen perspectives with volunteered geographic information (VGI). *Journal of Sustainable Tourism*, 21(2): 212-228.
- Sparks B A, Perkins H E, Buckley R. 2013. Online travel reviews as persuasive communication: the effects of content type, source, and certification logos on consumer behavior. *Tourism Management*, 39(6): 1-9.
- Stepchenkova S, Zhan F. 2013. Visual destination images of Peru: comparative content analysis of DMO and user-generated photography. *Tourism Management*, 36(3): 590-601.
- Xiang Z, Gretzel U. 2010. Role of social media in online travel information search. *Tourism Management*, 31(2): 179-188.

Tourism attention degree about China from overseas and its spatial patterns based on Tripadvisor

WANG Kun¹, GUO Fenghua², LI Renjie^{1,3}, Fu Xueqing^{1,3}

(1. College of Resource and Environment Science, Hebei Normal University, Shijiazhuang 050024, China;

2. Institute of Geographical Sciences, Hebei Academy of Sciences, Shijiazhuang 050011, China;

3. Hebei Key Laboratory of Environmental Change and Ecological Construction, Shijiazhuang 050024, China)

Abstract: Inbound tourism is an important part and indicator of tourism development in a country. International tourism is influenced by a series of factors. With the development of Internet, tourism electronic community is becoming increasingly more influential in the selection of travel destinations. User-generated content (UGC) that tourism electronic communities produce has gradually become the important source of tourism behavior and perception research. In contrast to the method of describing online attention degree only by using the number of search engine keywords, this article introduces the electronic community hierarchy as the weighting factor and establishes the model of tourist attention degree based on community UGC to adjust the focus of the model expression flexibly and optimize calculation results. Based on the web information collection technology, this study first designed an information collection process for retrieving information about Chinese tourism from an electronic community, Tripadvisor, and reconstructed this information to build a tourism information text database. The authors then built a place name database in order for using the text mining methods to obtain information about the place name frequency. The place names include two types: tourism attractions and destination cities. Based on the study of the tourist electronic community Tripadvisor, we found that attention of international users of Tripadvisor to tourism in China has three typical characteristics: (1) the focus of tourism is on the Great Wall, Mount Taishan, Mount Huangshan, Jiuzhaigou, Zhangjiajie and a small number of other tourist attractions, and a small number of destination cities such as Beijing, Hong Kong, Shanghai, and Guilin; attention degree for a large number of attractions and destination cities is low, with a long tail distribution and polarization. (2) Attention for the attractions and destination cities clearly match spatially. The attractions with high attention degree are near to or affiliated with the cities with high attention degree, such as Yangshuo in Guilin, the Great Wall in Beijing, Dujiangyan and Jiuzhaigou in Chengdu, and the West Lake in Hangzhou. (3) Tourism attention decreases from east to central and west and this pattern is consistent with the level of regional economic development in China; popular destination areas such as Beijing, Hong Kong, Guangzhou, Shenzhen, Shanghai, and Chengdu are also regional centers of economic activities. Tourism resources endowment, dissemination pattern of electronic word-of-mouth, geographic location, and level of economic development and the location, economic situation, and culture of the followers' country are the main factors influencing tourists' attention degree and its spatial pattern. This study designed a new data collection process to realize the effective management of a toponym database and tourism text database and set up a e-community tourism attention degree model that aims to solve the problem of quantitative calculation of Internet users' attention to the regional tourism and to provide a new thought for tourism geography study based on UGC from the Internet. In the visualization study, the authors analyzed the spatial distribution characteristics of attention and temporal variations. This method provides a new way for inbound tourism research.

Key words: tourism attention-degree; spatial pattern; electronic community; user-generated content (UGC); Tripadvisor