

综合LDA与特征维度的丽江古城意象感知分析

梁晨晨¹, 李仁杰^{1,2*}

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

摘要: 论文通过建立基于LDA(Latent Dirichlet Allocation)模型和包含两级特征维度的旅游地意象感知研究框架, 将LDA主题模型用于旅游微博文本分析, 以特征维度半定量刻画旅游地意象感知特征, 减少LDA主题凝练的主观性, 帮助研究者在特定维度框架约束下准确、客观地提取旅游地意象特征。丽江古城案例证明, 一级特征维度可以完整勾勒出丽江古城意象感知的基本框架, 包括以聚落形态、音乐意境、标志人物、休闲空间和纳西美食为核心的5组空间与景观元素, 深度旅游者、城市居民、年轻人、女孩子4类人群的特殊感知体验, 及旅游者与环境要素的不同互动特征; 二级特征维度进一步精细解读丽江古城的意象感知特征, 表现为丽江古城慢活性、夜生活和浪漫之都的文化意象、旅游者对地方文化与现代风情融合的凝视与体验等。结合特征维度的LDA模型, 准确构建了意象基本框架, 成功刻画了丽江古城的形象及精细特征, 并能进一步解析意象的形成机制, 为旅游地意象感知研究提供了新视角, 有助于深度解读意象形成的地方意义, 厘清认知、情感和行为意象间的关系。

关键词: LDA模型; 特征维度; 共现关系; 意象特征; 新浪微博; 丽江古城

旅游地意象感知是游客对旅游地的心理图像, 是认知、行为、情感的综合体现^[1-2]。当前, 许多旅游者通过社交媒体分享旅游经历、感知和评价^[3-4], 这类用户贡献内容(User Generated Contents, UGC)对其他访问者来说具有更高可信度^[5], 游客发布的UGC对旅游形象塑造具有越来越重要的作用^[6], 因此, 旅游UGC数据成为旅游地意象、旅游满意度、情感分析等研究的重要数据源^[7]。

虽然基于UGC的旅游意象感知研究持续增长, 但许多研究仍采用人工处理分析数据的方法^[8], 使方法扩展性和结果可靠性受到质疑。引入自然语言处理和数据挖掘方法是解决问题的重要途径^[9]。例如, 采用词频分析、网络分析等方法对在线评论、博客、游记和官方媒体文本进行挖掘, 识别旅游地形象特征^[10]和旅游者意象感知^[11]。词频分析通过计算文本高频词挖掘游客对旅游地的印象和情感, 获得旅游地形象概要特征, 但结果语义模糊且容易产生歧义^[12]; 语义网络分析通过文本特征词的中心度、

共现关系等挖掘游客意象结构^[13], 但缺少文档集合的主题聚类。机器学习中的主题分类技术能够把非规则化UGC数据按照特定主题进行分类, 解决上述文本分析面临的问题^[14], 使基于UGC的旅游意象特征提取结果可靠性提高。

LDA(Latent Dirichlet Allocation)是机器学习领域的潜在主题模型之一, 无需人工构造训练数据, 在基于在线评论的满意度等研究中逐渐得到应用^[15], 但对于噪声较多的短微博文本应用受到较多质疑^[16-18]。已有研究尝试添加级别标签^[16], 或基于空间聚合^[19-20]等方式解决噪声多的短文本问题。级别标签能较好标识微博内容结构, 但并未解决内容短小和表述随意引起的主题语义模糊问题; 空间聚合则适于较大空间尺度的主题分类, 对较小空间范围的应用效果并不明显。更关键的问题是, LDA分类结果虽能明确一组文本具有相似的主题, 但在主题命名或主题语义描述和凝练方面, 则存在明显的主观性。

收稿日期: 2019-04-01; 修订日期: 2019-05-24。

基金项目: 国家自然科学基金项目(41471127)。[Foundation: National Natural Science Foundation of China, No. 41471127.]

第一作者简介: 梁晨晨(1994—), 女, 河北保定人, 硕士生, 主要从事社会文化地理信息系统研究。E-mail: lccgis@163.com

*通信作者简介: 李仁杰(1975—), 男, 河北石家庄人, 教授, 博士生导师, 主要从事GIS与旅游地理研究。

E-mail: lrjgis@hebtu.edu.cn

引用格式: 梁晨晨, 李仁杰. 综合LDA与特征维度的丽江古城意象感知分析[J]. 地理科学进展, 2020, 39(4): 614-626. [Liang Chenchen, Li Renjie. Tourism destination image perception analysis based on the Latent Dirichlet Allocation model and dominant semantic dimensions: A case of the Old Town of Lijiang. Progress in Geography, 2020, 39(4): 614-626.] DOI: 10.18306/dlkxjz.2020.04.008

本文尝试引入特征维度分类方法改进LDA模型分析结果,对丽江古城新浪微博进行主题提取分析,以解决LDA在短文本应用中面临的问题,保证不同研究者能从大量UGC中提炼出基本一致的主题语义,最终获得旅游地意象感知的精细结构与特征。综合已有研究提出的级别标签和空间聚合2种方法,设计本方法的基本思路:选择区域较小的古城旅游地为案例;在基本噪声清洗基础上开展LDA主题分类;对结果特征词按照旅游地意象的特征维度进行后期标注;基于标注后的主题特征词进行主题语义提炼。本文尝试回答2个问题:①对LDA分类结果进行后期标注,是否可以使LDA模型适于旅游微博短文本的主题分析;②基于LDA提取的丽江古城意象感知特征是否与其他研究一致,并具有更好的结构性和精细度。

1 理论与方法

1.1 LDA主题模型

LDA是一种非监督机器学习技术,利用3层贝叶斯概率模型识别大规模文档中潜藏的主题信息。它的主要思想是:一篇文档是以一定概率选择了某个主题,并从该主题中以一定概率选择某个词语而构成,即一篇文档代表若干主题构成的一个概率分布,而每一主题又代表若干词语构成的一个概率分布^[21-22]。LDA模型的计算结果可以得到文档-主题和主题-词语2个概率分布。主题-词语概率分布由一系列特征词及其在该主题中出现的概率值表示,特征词概率值越大对该主题贡献率越高,与该主题关联度越大,由此反映每个主题的内部结构;文档-主题概率分布获得每个主题下的文档支持权重,权重越大表示该文档与主题的关联度越大^[23]。

假设 d 代表文档, w 代表文档中的词语, t 代表主题,则词语在文档中出现的概率 P 可表示为^[24]:

$$P(w|d) = P(w|t) \times P(t|d) \tag{1}$$

式中: $P(w|d)$ 表示词语 w 在文档 d 中出现的概率,是已知的;而 $P(w|t)$ 表示词语 w 在主题 t 中出现的概率, $P(t|d)$ 表示文档 d 对应主题 t 的概率,二者是未知的。LDA模型则利用统计学抽样方法,通过一个已知的参数推算出2个未知的参数,从而实现文档的主题分析。本研究使用Python的gensim工具包调用LDA模型,实现微博文本主题分析,为凝练旅游地意象感知特征提供结构化处理的专题数据。

1.2 旅游地意象感知的特征维度

虽然LDA模型可以基于一组大数据文本特征词的相关性进行主题分类,但每个分类结果的主题到底如何凝练和表达,并未形成标准或统一方法。已有研究一般由研究者基于研究目标自行判断主题语义^[19,25],因此,即使采用相同文本集合得到同样的主题分类结果,不同研究者可能提炼出不同的主题语义。针对这一问题,由不同参与者分别提炼主题后,判断结果的相关性,减少主题凝练结果的主观性^[26]。本文提出对每个主题的高概率特征词进行基于意象感知特征维度的后期标注方法,为LDA分类结果的主题语义提炼提供标准和约束框架,以控制语义描述的一致性。

对特征词进行意象感知维度分类的基本思路为:首先,分析旅游地意象感知的研究成果,明确旅游地意象是由认知意象、情感意象与行为意象支撑的基本结构维度^[1];第二,将旅游地意象的结构维度与影响因素转换为具体的表征维度;第三,考虑旅游空间与景观语义的时空属性,引入时间与行程维度,刻画旅游行程对意象感知形成的影响;第四,根据旅游地意象感知基本结构与表征维度,并考虑实际旅游文本描述特点,设计了包含4个一级维度、10个二级维度的旅游地意象感知特征维度结构体系(图1、表1)。

“空间与景观”维度主要是认知意象的具体表征和呈现载体,反映旅游者对目的地区位、旅游环境和景观要素的认知与印象。“空间”强调旅游行程

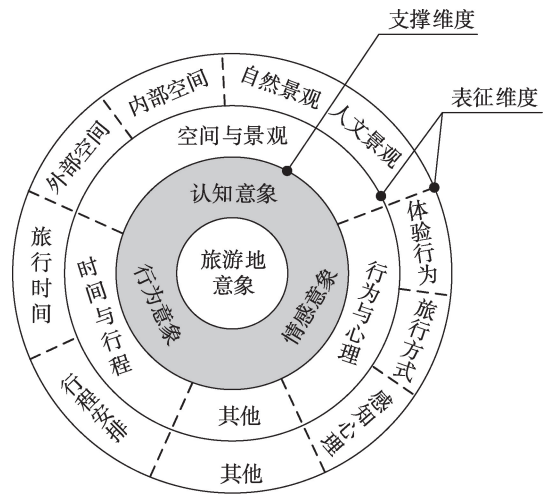


图1 意象感知基本结构与特征维度的关系概念图
Fig.1 A conceptual model of the relationship between the basic structure of image perception and characteristic dimensions

表1 旅游地意象感知特征维度结构体系与解释

Tab.1 Structure of characteristic dimensions of tourism destination image perception and its interpretation

一级维度	二级维度	维度释义	特征词样例(以丽江古城为例)
空间与景观 (SL)	外部空间(SL ₁)	目的地周边旅游景区、客源地等空间单元	昆明、大理、拉市海、虎跳峡、北京
	内部空间(SL ₂)	目的地内部的客栈、餐馆、酒吧等空间单元	客栈、酒店、房间、院子、阳台、酒吧、电影院
	自然景观(SL ₃)	描述地文、水域、生物、气候等自然景观要素	阳光、蓝天、微风、小雨、白云
	人文景观(SL ₄)	包括建筑、民俗、文化、历史等人文景观要素	四方街、情人节、鲜花饼、小倩、乐队
行为与心理 (BP)	体验行为(BP ₁)	旅行中融入目的地空间,与景观交互的行为活动	骑马、喝咖啡、散步、晒太阳、唱歌、泡吧、跳舞
	旅行方式(BP ₂)	游览观光的出行方式,如交通工具、步行等	搭车、行走、火车、飞机
	感知心理(BP ₃)	旅游者对环境的感知与旅游体验心理活动	开心、愉快、激动、热闹、崩溃
时间与行程 (TR)	旅行时间(TR ₁)	描述旅游行为的特定时间节点	旺季、暑假、周末、白天、早晨
	行程安排(TR ₂)	描述旅游行程的节奏与旅行过程安排	两天、第一天、一站、早餐、起床、出门
其他类(OT)	其他类(OT)	不属于以上任一维度的特征词语	日子、听说、商业、距离

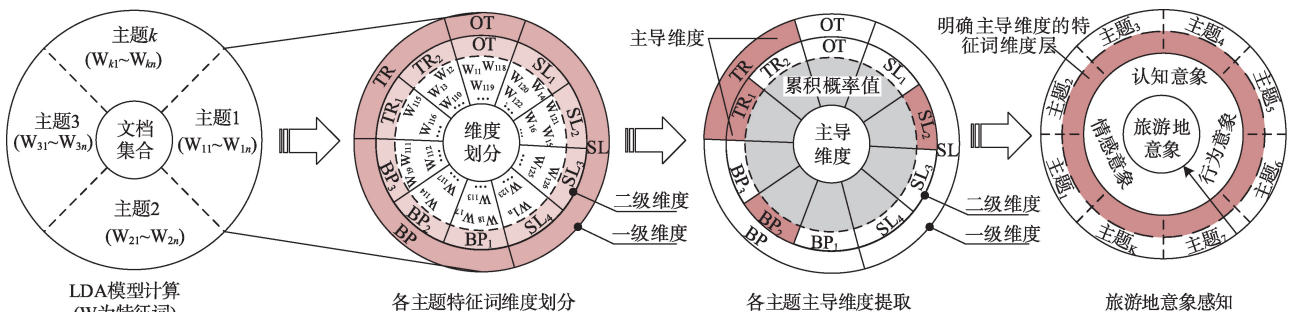
中的区位与环境表述,而“景观”强调景观要素或空间单元对旅游者感知、体验的意义,例如丽江古城“四方街”对旅游者来说更突出人文景观意义,而非古城环境空间。“行为与心理”维度主要表征行为意象和情感意象,是旅行方式与景观体验的具体表征,实质是“空间与景观”维度的互动呈现。“时间与行程”从景观与旅游者双重视角对意象感知的时性进行刻画,既表征景观意义的时间特征,又体现旅游者在特定时间的景观融入(例如古城夜晚泡吧)。“其他类”维度是对旅游地意象凝练的自由度表征,为识别旅游者特殊视角下建构的旅游地意象提供扩展接口。

特征维度划分方法具有以下特点:①充分考虑旅游地意象的概念与内涵,实现基于主题分类计算结果的结构化提炼框架,控制旅游地意象感知提取结果的结构一致性;②特征词维度划分使旅游地意象感知的提炼结果具有系统性和层次性,并有助于显性表达维度间的关系;③研究者可根据特征维度结构框架,提炼认知意象、情感意象与行为意象,并

最终形成旅游地整体意象。因此,特征维度划分框架从旅游地理视角,为主题分类结果的深度分析与挖掘提供了标准。

1.3 研究框架

本文综合利用LDA模型和结果主题词语的特征维度,设计了旅游地意象感知的基本研究框架,用于分析提炼游客对旅游地的意象感知特征(图2)。首先,通过LDA模型对预处理后的旅游位置微博文本进行主题提取,将大量文本聚类为若干主题。第二,将构成每个主题的高概率特征词划分不同层次特征维度。第三,根据不同维度的特征词累积概率特征,确定每个主题的一、二级主导特征维度。第四,以每个主题的主导维度确定该主题表征的核心意象特征:一级维度决定意象基本框架,二级维度刻画意象特征细节。第五,根据各个主题的两级主导维度结构特征确定主题间的结构关系,根据每个主题的高概率特征词之间的结构关系(例如文本共现关系等),从认知意象、行为意象和情感意象3个方面提炼旅游地整体意象特征。



注:维度编码含义见表1。

图2 旅游地意象感知研究框架

Fig.2 Research framework of tourism destination image perception

2 实证研究

2.1 案例区及数据基本情况

2.1.1 案例区基本特征分析

丽江古城位于云南省丽江市古城区,又名大研镇,坐落在丽江坝中部。选择丽江古城为案例区主要基于3个方面考虑。第一,古城旅游地的典型性。丽江古城有良好的人居环境和独特的纳西东巴文化,集世界文化遗产、世界自然遗产和世界记忆遗产于一身。丽江古城从整体布局到民居形式,均完好保存了古代风貌;包括藏族、白族、彝族、回族、傈僳族、普米族等居民在此定居,纳西古乐、占卜文化以及纳西族火把节等多彩的地方文化与民族习俗别具一格^[27],诸多民族和文化交汇并存、碰撞,构成丽江独特的文化背景^[28],一度被称为“东方威尼斯”“高原姑苏”,成为中国历史文化名城和国际著名旅游城市。第二,自丽江古城成为世界文化遗产以来,吸引了世界各地的大量游客,古城文化经历着不断变迁与创造的过程,引起学界极大关注,涌现出丰富研究成果。第三,丽江古城面积较小,仅有7.279 km²,符合研究方法对较小旅游地开展研究的设计要求。

2.1.2 数据基本情况

本文以旅游者在丽江古城发布的带定位信息的微博文本为数据源。课题组在取得新浪微博官方授权后,利用新浪微博开放平台和应用程序接口,接入基于Python的网络爬虫程序,获取古城范围内的全部位置微博数据。数据覆盖时间范围为2013年1月1日至2017年12月31日。整个古城共获取位置微博302921条,微博数据内容包含微博文本、发布时间、地理坐标、用户信息、微博ID等信息。

数据预处理和数据清洗的基本流程:去除文本中的表情符号、英文字符、标点符号、特殊字符以及网址链接等;利用正则表达式定义规则对微博文本进行中文提取;采用结巴分词系统中的精确模式对中文提取后的微博文本进行分词;去除“的”“是”等无意义停用词;由于所有数据都位于古城内部,而微博数据中出现“丽江古城”一词的数量极高,为避免该词对其他词语的概率产生影响,本文将该词收入停用词表中。最后参与LDA模型计算的微博数据共263431条。

2.2 LDA计算结果

2.2.1 主题分布情况与选择

将清洗后的微博数据作为一个文本集合,输入

LDA模型计算程序,以概率分布中概率值最大的主题作为每条微博的所属主题^[29],反复测试不同数量的主题分类结果的区分度,最终确定包含30个主题的计算结果,获得每条微博的主题概率分布。含微博数量较少的主题对丽江古城意象感知的影响较小,如主题28的微博共4046条,仅占全部微博量的1.54%。因此,统计各主题所含微博数量后,剔除微博数比例小于2%的3个主题,最后获得27个有效主题。

2.2.2 主题特征词概率分布

经试验,LDA模型输出的各主题中,概率较大的前50个词累积贡献概率多数超过60%,只有2个主题累积概率为52.8%和59.3%,说明前50个特征词能够代表该主题的语义表征。主题提取结果的部分样例如表2所示。由主题提取结果可见,LDA模型抽取的主题效果较好,各主题由不同概率的词语分布构成,主题特征词语的概率分布具有“长尾”分布特征,在词语概率分布的尾部,词语概率值趋于平缓,概率值约小于0.005,概率值较小的词语对主题的贡献率较小。

3 结果分析

按照以上两级特征维度的划分标准,对27个主题包含的特征词进行分类,可以发现各主题特征词在不同维度的分布有明显差异。计算每个主题中各特征维度包含的特征词累积概率,以确定各主题的主导特征维度及其语义内涵,对各主题进行分类与对比分析,完成旅游地意象感知特征的凝练与提取。

3.1 一级特征维度的分布特征及其意象表征

按照一级特征维度进行特征词累积概率值计算,并确定每个主题一级维度中的主导维度。由于一级维度主要解决意象感知的框架描述,因此主导维度的确定标准为:包含贡献率较高的一个或几个相同维度特征词,且所有该维度特征词的累积贡献率明显高于其他特征维度。根据27个主题分类结果,可以划分为2种主导类型:单维度主导型和双维度主导型(表3)。

3.1.1 单维度主导型主题呈现的意象特征

单维度主导型主题是提取旅游地意象框架的主要依据。丽江古城的单维度主导型以“行为与心理”维度主导型为主,“空间与景观”维度主导型其次,“时间与行程”维度没有主导型主题。

表2 主题-特征词概率分布示例

Tab.2 Examples of topic-characteristic word probability distribution

主题6			主题12			主题17			主题22			主题24		
概率	词语	维度	概率	词语	维度	概率	词语	维度	概率	词语	维度	概率	词语	维度
0.222	客栈	SL ₂	0.168	旅行	BP ₁	0.145	偷笑	BP ₃	0.078	朋友	SL ₄	0.069	玉龙雪山	SL ₁
0.047	老板	SL ₄	0.068	心情	BP ₃	0.076	浮云	SL ₃	0.028	唱歌	BP ₁	0.048	睡觉	BP ₁
0.042	月亮	SL ₃	0.059	风景	SL ₃	0.071	音乐	SL ₄	0.027	爱情	SL ₄	0.044	纳西	SL ₄
0.033	离开	BP ₁	0.047	一米阳光	SL ₄	0.065	天气	SL ₃	0.026	日子	SL ₄	0.038	火锅	SL ₄
0.032	舒服	BP ₃	0.034	小伙伴	SL ₄	0.044	夜晚	TR ₁	0.023	忘记	BP ₃	0.033	真心	BP ₃
0.026	惬意	BP ₃	0.024	江湖	OT	0.042	来到	OT	0.021	啤酒	SL ₄	0.031	酸奶	SL ₄
0.022	院子	SL ₂	0.022	早晨	TR ₁	0.036	夏天播放	SL ₄	0.021	一场	OT	0.022	好吃	BP ₃
0.022	房间	SL ₂	0.021	哈气	OT	0.035	酒吧	SL ₂	0.017	不用	OT	0.021	美食	SL ₄
0.019	一家	OT	0.016	最美	BP ₃	0.023	播放	BP ₁	0.017	浪漫	BP ₃	0.020	小吃	SL ₄
0.016	晚餐	TR ₁	0.014	小桥流水	SL ₃	0.018	听歌	BP ₁	0.016	孩子	SL ₄	0.019	自然	SL ₃
0.015	印象	BP ₃	0.013	一座	OT	0.018	还好	BP ₃	0.016	一群	OT	0.018	到达	BP ₁
0.015	外面	OT	0.012	景色	SL ₃	0.016	见到	OT	0.014	听说	OT	0.016	烤鱼	SL ₄
0.013	入住	BP ₁	0.011	木有	OT	0.013	妹子	SL ₄	0.014	旅程	TR ₂	0.016	海拔	SL ₃
0.012	笑哈哈	BP ₁	0.01	倾城	BP ₃	0.011	辛苦	BP ₃	0.013	静静的	BP ₃	0.013	拜拜	BP ₁
0.011	环境	OT	0.01	向往	BP ₃	0.010	懂得	BP ₃	0.013	收获	OT	0.012	雪山	SL ₁

注:维度编码含义见表1。

表3 主题特征词一级维度累积概率计算结果及分类

Tab.3 Results and classification of cumulative probability calculation in the first-level dimensions

(%)

类别	主题	空间与景观	行为与心理	时间与行程	其他类	类别	主题	空间与景观	行为与心理	时间与行程	其他类
单维度主导	主题1	20.2	29.9	5.7	9.6	单维度主导	主题19	13.9	26.1	9.8	16.8
单维度主导	主题3	12.0	26.8	6.4	17.3	单维度主导	主题20	17.6	29.8	7.6	13.8
单维度主导	主题4	15.1	28.5	1.9	22.5	单维度主导	主题21	9.7	37.6	2.6	21.4
单维度主导	主题5	21.7	31.7	4.4	16.3	单维度主导	主题23	8.8	45.6	1.3	21.1
单维度主导	主题6	39.0	17.1	3.0	14.5	单维度主导	主题24	38.4	18.8	0.0	7.4
单维度主导	主题7	9.1	26.2	1.5	16.0	双维度主导	主题0	20.6	13.9	14.9	13.6
单维度主导	主题8	25.3	9.5	15.8	19.9	双维度主导	主题2	24.2	29.1	0.4	15.6
单维度主导	主题10	24.6	33.7	6.2	12.1	双维度主导	主题9	27.6	30.9	1.9	19.5
单维度主导	主题11	6.3	15.6	3.5	36.6	双维度主导	主题16	24.6	29.7	0.0	15.4
单维度主导	主题12	22.0	33.8	3.5	14.7	双维度主导	主题17	35.0	27.2	4.4	13.0
单维度主导	主题13	40.5	17.4	4.6	15.2	双维度主导	主题22	23.0	16.5	1.4	17.6
单维度主导	主题14	9.9	29.8	0.0	19.6	双维度主导	主题25	22.1	17.6	3.7	26.7
单维度主导	主题15	31.2	22.8	10.5	13.0	双维度主导	主题26	26.0	31.8	1.4	12.8
单维度主导	主题18	19.2	32.7	1.2	11.7						

注:加粗数字为主导维度累积概率值。

“行为与心理”维度主导型主题共计13个,以主题5、10、12、18、21、23最具代表性,6个代表性主题中的主导维度特征词累积贡献均超过30%,从4个视角表现了不同类型旅游者的古城情感意象框架:

(1) (深度旅游者)“发呆、晒太阳”是丽江古城最悠闲的生活方式;

(2) (城市居民)“期待”逃离繁华城市带来的“烦恼、讨厌、痛苦”;

(3) (年轻人)“快乐、享受、喜欢、向往、馋嘴”是古城慢生活深度体验的直接感受;

(4) (女孩子)在浪漫之城偶遇时自然流露的“害羞”体验。

主题6、8、13、15、24是“空间与景观”维度主导型主题,该维度特征词在5个主题中的累积贡献均超过25%,以5组核心空间与景观元素刻画丽江古城的认知意象框架:

- (1) 以大研古城、束河古镇为主要内部空间载体的丽江古城;
- (2) 以班布、醉叶居、必胜客为代表的客栈、餐厅、酒吧、酒吧街聚落风格;
- (3) 以妹子、妹妹、女神为形象标志的人物特征;
- (4) 以房间、院子、门口为典型单元的休闲空间;
- (5) 以火锅、酸奶、烤鱼、鸡豆、凉粉、鲜花饼、过桥米线为代表的美食文化。

3.1.2 双维度主导型主题呈现的意象特征

双维度主导型主题能较好地体现不同维度间的相互联系,提高形象感知提炼结果的可解释性。双维度主导型包括主题0、2、9、16、17、22、25、26共8个主题。其中5个主题是“空间与景观-行为与心理”类型,表现了旅游者与目的地空间和景观交互的过程,有助于发现旅游者如何融入目的地空间,进而感知与体验景观的意义并形成地方感。此类主题表征了人与环境要素的4类互动特征:

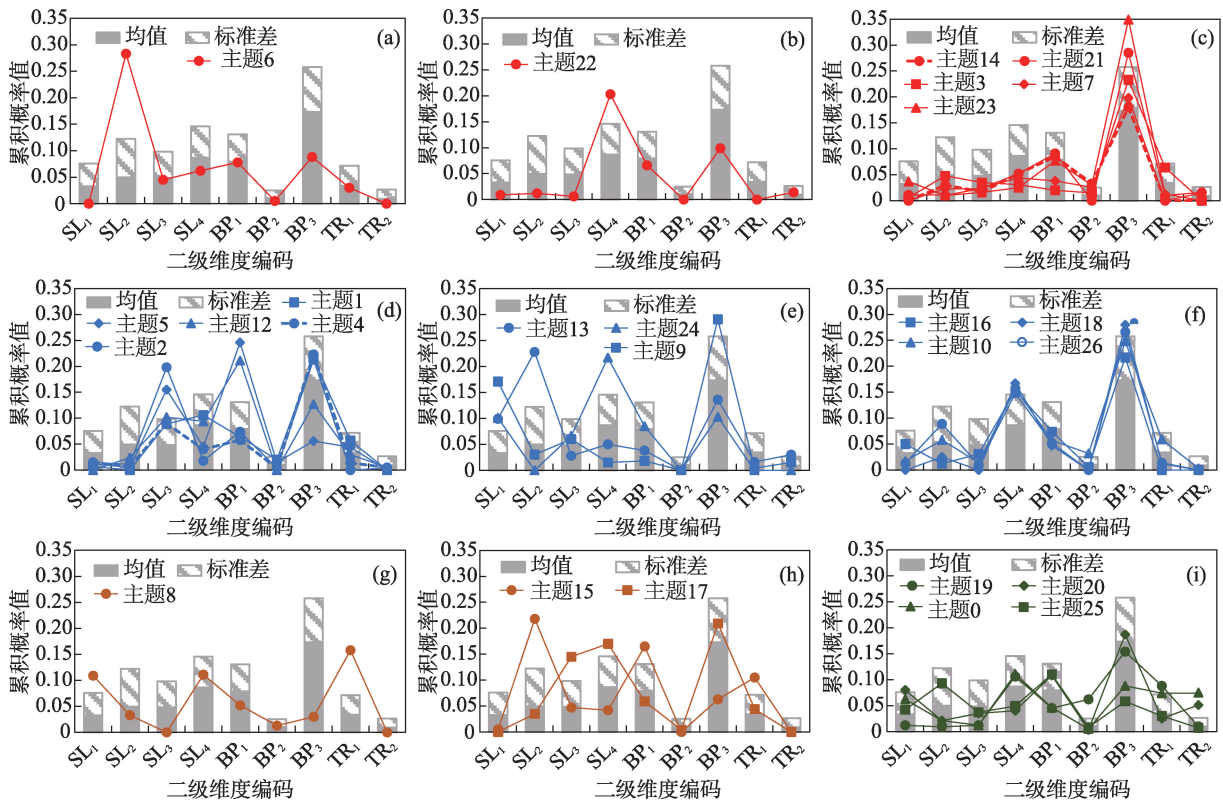
- (1) 从“阳光、太阳”等自然景观获得“温暖、休闲、美美”的感知体验;
- (2) 从“米线、烧烤”等美食获得“好吃、美味”的

- 感知体验;
- (3) “一条街、风花雪月、大冰、小屋、排骨、美女、水车、故事、七夕”等景观要素,使旅游者的体验“浪漫之都”的行为与心理有了自然与文化意境;
- (4) “北京、昆明”等客源地旅游者获得“幸福、开心”的感知体验,他们希望“再见、遇见”包括“大理、洱海、双廊、雪山”等在内的大丽江。

3.2 二级特征维度的分布特征及其意象表征

各主题的第二级特征维度可以更细致、精确地描述旅游者的古城意象感知。为准确提炼各主题表征的意象感知,二级维度的主导标准定义为:不仅是该主题中出现概率较高的维度,而且高于(或接近)所有主题中该维度特征词的概率均值加1个标准差。根据上述标准,按照各主题二级特征词概率分布曲线,将27个主题分为单维度主导型、双维度主导型、多维度主导型和主导维度不明显型4种类型,并根据主题意象感知提炼视角进行分组展示(图3)。

由于二级特征维度需要精细刻画旅游地意象特征,为了进一步减小基于主导维度提取意象特征时的主观性,本文又引入了特征词共现关系网络方



注:图a、b、c为单维度主导型,d、e、f为双维度主导型,g、h为多维度主导型,i为主导维度不明显型。

图3 根据二级维度特征词累积概率分布特征的主题分类结果

Fig.3 Classification of topics based on the cumulative probability distribution in the secondary-level dimensions

法,定量描述每个主题包含的高概率特征词在博文文本中共同出现的次数,以发现共现网络的中心和特征词共现规律。特征词共现关系网络是主题内部特征词语义关系结构的细化表达。9个代表性主题高概率特征词的共现关系网络如图4所示。

3.2.1 单维度主导型主题呈现的意象特征

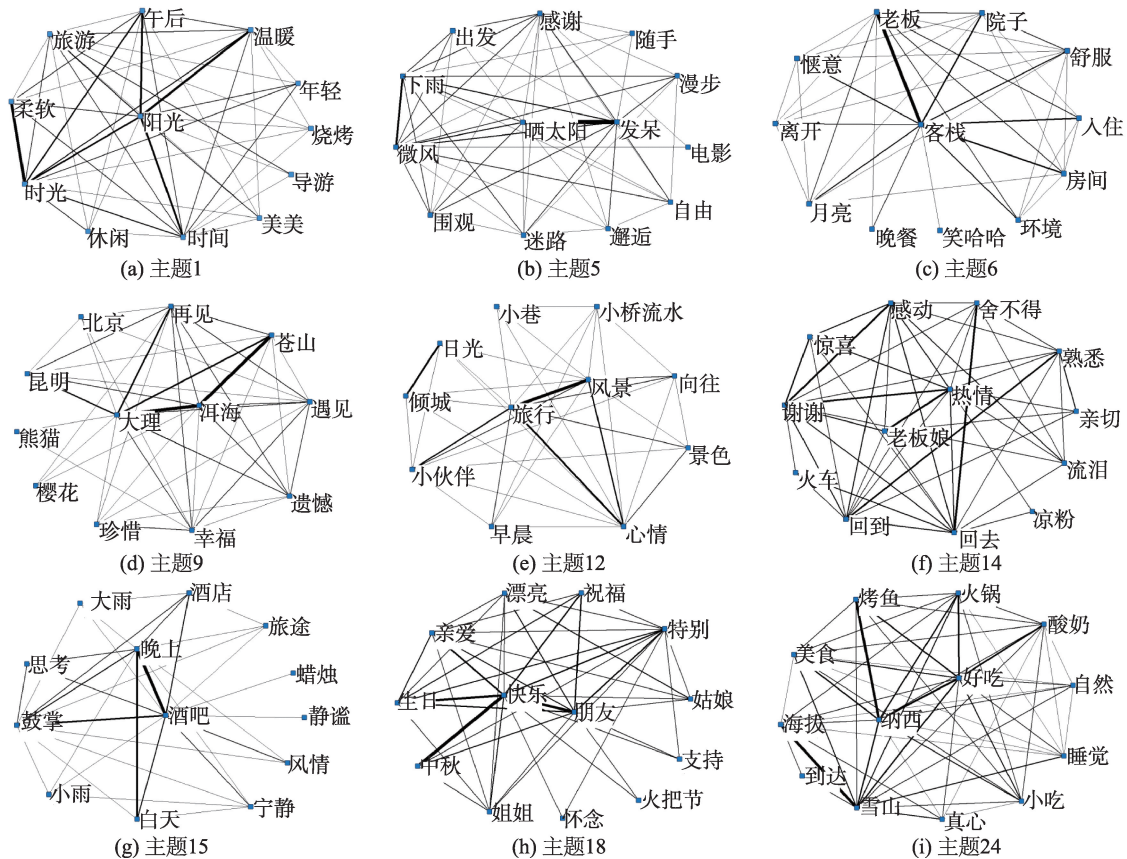
二级特征维度下的单维度主导型主题是丽江古城意象感知特征的精细刻画,共有7个主题,其中1个内部空间维度主导型、1个人文景观维度主导型、5个感知心理维度主导型。

主题6是内部空间维度主导主题。图3a和图4c表明,客栈是旅游者体验古城慢生活的重要空间载体之一;“老板”这一人文景观要素对旅游者形成“客栈”认知意象具有关键作用;高概率的“房间、院子、门口”是客栈内关键景观元素;客栈及其景观元素与旅游者感知体验特征词呈现的较高共现度,使旅游者在古城客栈中形成“舒服、惬意”的情感意象体验更加具象化。

主题22为人文景观维度主导的主题(图3b),通

过“朋友、爱情、啤酒、女朋友、情人节”等人物、节日等景观要素描写,与旅游者“唱歌、浪漫、陪伴”等行为与感知词相呼应,高概率特征词“朋友”与其他特征词均有共现,且与“爱情”“陪伴”共现度最高,隐含了旅游者对友情和爱情的期许。

主题3、7、14、21和23是感知心理维度突出的5个主题(图3c),既有大众旅行常规的景观偏好和感受,也有逃离常住地进入非惯常旅游地的强烈感知对比,还体现了旅游者与地方经营者的情感交互。主题3和7更像是传统大众旅游感知体验,包含“不想、美的、无聊、威武、可惜、给力、爱上”等直接感知体验描述词,以及“两天、离开、徒步、逛逛”等较多旅行方式与行程描述,“忠义市场”作为具体表征的空间意象载体,说明丽江古城游客还包括传统观光式旅游群体。主题14中的“谢谢、舍不得、感动、惊喜、亲切”等感知特征词与“老板娘、热情”等词汇的高共现关系组合(图4f),巧妙折射出旅游者与古城经营者的融洽关系,与主题6对老板人物的呈现形成极好的耦合。主题21和主题23的特征词共现网



注:连接线越粗表示词语共现次数越多。

图4 代表性主题的高概率特征词共现网络

Fig.4 Co-occurrence network of high frequency characteristic words of representative topics

络核心均为“生活”,与之共现的特征词包括“喜欢、享受、热闹、安逸”及“痛苦、抱怨”等,呈现明显相反的感知心理;结合主题23的“城市、回家”等高概率词,说明丽江古城在疏解旅游者城市生活压力方面具有特殊意义。

3.2.2 双维度主导型主题呈现的意象特征

主题9、13和24均包含外部空间主导维度(图3e)。3个主题借助“昆明、大理、苍山、洱海、玉龙雪山、虎跳峡、西藏、拉市海、束河”等特征词,集中体现了与丽江古城形成资源整合、互补的不同距离范围的外部空间目的地。主题9包括的“大理、苍山、洱海”高共现网络核心,是丽江古城外部中等距离的旅游空间,游客的直接感知体验包括“幸福、遗憾、遇见、再见”等(图4d);主题24是外部空间与人文景观双维度主导,其中高概率的“火锅、酸奶、烤鱼”等美食词,和“雪山、纳西”两个核心特征词具有较高共现度(图4i),使纳西文化中的自然崇拜和特色美食形象跃然纸上;主题13则通过“古镇、庭院”等内部空间,结合“家人、老公”等人物描写,使温馨的家庭、厚重的文化和生态的自然在丽江古城融为一体。

主题1、2、4和主题5、12分别是自然景观与感知心理和体验行为组合的双维度主导型主题(图3d),突出表达人与自然的交互,细化一级特征维度下旅游者对丽江阳光的美好体验。首先,5个主题均直接和隐含表达了旅游者邂逅“下雨”的优美意境,古城阳光因此多了一份清新和温润,使旅游者遗忘日常生活的“喧闹”,以“发呆、晒太阳、漫步”和品味美食的方式享受悠闲时光。第二,主题1包含的高概率“午后”特征词,为古城“阳光”意境定义了时间坐标(图4b)。第三,“微风”吹过来和“小桥流水”为古城旅行体验增加了触觉和视觉上的动感元素(图4b、图4e)。第四,充分体现了“蓝月谷”作为丽江古城外部空间上新的旅游资源组合,与较为成熟的拉市海一起,将古城的历史文脉与自然水景观巧妙融为一体。

主题10、16、18、26是人文景观与感知心理双维度主导型主题(图3f),相对于人与自然交互中追求宁静和清新的感知,这组主题更突出旅游者与地方文化交互的愉悦感受,特别是与地方人物的互动感增强。主题16和18(图4h)2个特征词共现网络的中心词分别是“开心”和“快乐”,分别与“雪山、四方街、青年、东巴、客栈、酒吧”和“朋友、生日、中秋

(节)、姑娘、姐姐”等呈现较高共现关系,加上较多的“愉快、喜欢、幸福”等积极感知特征词,共同描述旅游者借助古城及外围核心景观凝视、体验传统东巴文化与现代文艺风情的美好情景。主题10以“美女、男人、女人、烤肉、排骨”等人文景观和“花心、晚安、特色”等感知体验为主导,是旅游者对古城浪漫风情的直接内心体验。主题26则是从内部空间景观元素对浪漫之都的刻画,在“小院、小屋”等内部空间中融入“故事、蛋糕、七夕、大冰、风花雪月、情侣”等丰富的人文元素,刻画出“宁静”的浪漫意境。旅游者心中的“艳遇之都”并非真正寻求艳遇,而实为“浪漫之都”的意象表征:厚重的人文底蕴,承载了旅游者逃离喧嚣城市生活的愉悦,对浪漫、自由不羁的向往;宁静的浪漫意境,消除了相互间的文化隔离和认知束缚,衬托古城的地方意义和解读艳遇的真正意象内涵。

3.2.3 多维度主导型主题呈现的意象特征

主题8由外部空间、人文景观和旅行时间3个主导维度构成(图3g)。“云南”主导了外部空间特征维度,是对丽江古城旅游的区位刻画;“鲜花饼、过桥米线、丁丁(酸奶)、餐馆”等特征词是云南饮食文化景观的表达;“明天、傍晚、半夜、一早”等特征词是对旅游时间与行程的描述。主题8与主题24相呼应,但主题8并不包含具体的外部空间要素,而侧重于描述丽江古城在云南饮食文化中的独特地位。

主题15与17都是基于3个主导维度直接或间接描述古城酒吧夜生活精细画面的典型主题(图3h)。主题15由内部空间、体验行为和旅行时间3个维度主导,特征词共现网络的中心词是“酒吧”,并与“鼓掌、晚上、蜡烛、思考、酒店、大雨、小雨”等景观要素和行为特征词有较高共现关系(图4g),另外出现了“班布、醉叶居、必胜客”等高概率酒吧与客栈名称词,形成鲜明的酒吧生活主题。主题17是自然景观、人文景观和感知心理三维度主导的主题,“天气、夏天播放、珍贵、音乐、酒吧、姑娘、乐队、手鼓、民谣、鼓掌、唱歌、喜欢”等特征词组成的共现网络,形成“夏天播放、酒吧、民谣”3个网络中心,精细刻画了音乐、酒吧和人物等元素构成的酒吧夜生活画面。主题15和17的主导维度和特征词共现网络,充分证明了“音乐始终是丽江古城非常鲜明的地方符号,融入了旅游者对新空间的地方感知”^[30];品牌酒吧、著名乐队、民族音乐元素、美丽姑娘,是古城音乐酒吧夜生活的标志元素;或大或小的雨又

为强烈的动感画面嵌入一份静谧的思考空间,与主题26主导的宁静浪漫意境形成动静呼应之势。

3.2.4 主导维度不明显主题呈现的意象特征

主题0、19、20和25是4个主导维度不明显的主题(图3i),但从4个主题的维度分布及特征词共现情况,也能反映出旅游者对古城内部景观及其外部空间的旅游意象感知特征。

主题0的旅行时间与行程维度显著高于所有主题均值,但在本主题中并不突出;该主题的“时间”加上“早餐、起床、出门、午饭、等待、离开”等行程特征词,结合“歌手、酒吧、歌声、世界、成都”等人物与空间特征词的共现特征,说明丽江是区域旅游线路的关键节点。主题19的感知体验和人文景观维度累积概率相对较高,拥有“味道、悲伤、梦想、精彩、有缘”等较多感知体验特征词,另外,“骑马、划船”等旅游行为,“昨天、早上、下午”等时间行程描述,以及高概率特征词“明信片”及其与“喜欢、旅行、未来”等词的较高共现度,刻画了特殊的古城意象感知,使特色明信片成为古城旅游者现实与梦想、旅行中的此地方与彼地方的联系纽带。主题20突出了“西安”等客源地及与“泸沽湖”的稳定线路组合,拥有共现度较高的行程安排和体验行为特征词。主题25则以高概率特征词“生活”为中心,从“古道、公园、咖啡馆、观景台”等内部空间维度,呈现旅游者在丽江古城“悠闲”惬意的感知体验。

4 结论与讨论

4.1 结论

针对LDA主题模型是否适于微博等社交媒体短文本,以及分类结果语义提炼时存在主观性的问题,本文建立了一个适用于旅游地意象感知描述的特征维度分类框架,包括“空间与景观、行为与心理、时间与行程、其他”4个一级特征维度和10个二级特征维度。基于特征维度分类,LDA分类结果可以从语义描述框架和主导维度2个角度,保证不同研究者基于LDA结果提炼同一旅游地意象时保持较好一致性。首先,每个主题包含的特征词按设定的特征维度进行分类标注,保证主题分析具有一致的维度框架和基本结构。然后,在两级特征维度下,基于每个主题不同维度特征词的累积概率分布,确定每个主题的主导维度,作为各主题提炼时的核心依据。在利用文本型社交媒体内容进行旅

游地意象感知提取时,结合特征维度的LDA模型分析,能够实现意象感知描述框架和感知精细特征提取的有效统一。选择丽江古城为研究案例,利用位置微博数据进行LDA主题分析,通过两级特征维度划分及主导维度分析,获得了古城多维度意象感知特征,初步证明结合特征维度的LDA模型可适用于微博等社交媒体短文本分析;在LDA分析基础上,构建各主题特征词的共现关系网络,可以进一步解释意象特征形成的原因和机制。主要结论如下:

(1) 各主题一级特征维度勾勒出丽江古城意象感知的基本框架。① 空间与景观方面提取出5组核心空间与景观元素:大研古城、束河古镇为代表的古城及其客栈、酒吧为主的聚落形态;乐队、手鼓、民谣、灯光元素烘托的音乐意境;以妹子、妹妹、女神为标志的人物;以房间、院子、门口为代表的休闲空间;以火锅、酸奶、烤鱼、鸡豆、凉粉为代表的纳西美食。② 行为与心理方面提取4类人群的特殊感知:深度旅游者享受“发呆、晒太阳”的悠闲惬意,城市居民期待逃离城市的烦恼与痛苦,年轻人的喜爱、向往、享受和馋嘴,女孩子在浪漫之城的羞涩。③ 空间与景观、行为与心理共同表征的核心意象是“再见、遇见”,体现了旅游者与环境的4类互动模式:阳光中感受温暖与休闲;纳西美食的柔软与美味;大冰小屋、排骨、美女、水车等景观元素映衬下的浪漫行为与心理;由丽江古城、玉龙雪山和大理等共同承载的幸福、开心的旅行感知体验。

(2) 各主题二级特征维度及特征词间的共现关系,全面呈现旅游者与景观的交互特征,精细刻画了丽江古城的意象特征。① 从6个视角精细解读丽江古城慢活性:“客栈”及其“房间、院子和门口”等具象化单元是慢活性呈现的主要空间;“发呆、晒太阳、漫步、品味美食”是慢活性的主导体验方式;邂逅“下雨”、“微风”拂面和“小桥流水”,是慢活性的动感元素和清新点缀;“午后”是慢活性的时间坐标;逃离惯常环境的喧哗、压力与烦恼是追求慢活的情感驱动;现代经营者,特别是热情的老板娘,有意无意承担了休闲慢活性意象的直接传播媒介。② 从5个方面烘托古城夜生活和浪漫之都的文化意境:“酒吧、酒店、酒吧街”是夜生活的核心场所;音乐主题酒吧则是夜生活的鲜明地方符号;美女和美食是浪漫之都的直接感知;音乐(民谣)、酒吧、灯光和美女构成的画面成为浪漫意境的主要支撑;在中秋、情人节等节日中与朋友唱歌、品茶,隐含了旅

游者对友情和爱情的期许。③用4种载体支撑旅游者对地方文化、自然与现代风情融合的凝视与体验:古城地标、纳西美食和雪山崇拜是凝视、体验民族地方文化与现代文艺风情的主导载体;拉市海、玉龙雪山和蓝月谷等古城近邻旅游地,将古城历史文脉与自然山水景观巧妙融合;昆明、大理、泸沽湖、茶马古道等更远距离和更大空间尺度的资源组合,承载旅游者的自然与地方文化体验;特色明信片则成为古城旅游者现实与梦想、此地方与彼地方的联系纽带。

4.2 讨论

尽管已有研究已经指出慢活性的历史和文化根基,但从旅游者直接的感受来看,现代古城的经营者已经有意无意地承担了休闲、慢活性古城意象传播的重要媒介,客栈老板特别是老板娘的热情和客栈空间景观元素的设计,无疑成为旅游者体验古城生活的直接感受。

引入特征维度的LDA模型,以标准化和半定量方式从大量旅游者UGC中快速、精确刻画了丽江古城的意象感知特征,部分揭示了意象特征形成的原因和机制,其价值主要体现在3个方面:一级特征维度有助于准确把握旅游地意象特征的基本框架,二级特征维度有助于定量、精细刻画旅游地意象特征的核心维度和具体形象特征;双维度或多维度主题,有助于合理解读意象感知特征的地方意义,进而厘清认知、情感和行为意象特征之间的关系;通过主题包含的特征词进一步提炼主题内涵与语义,可一定程度上保证不同研究者按照基本一致的方式总结主题语义。

从丽江古城案例分析结果看,基于微博的LDA主题分析取得了与问卷或访谈等社会调查方式接近的研究结果。例如,白凯等^[11]通过扎根理论对不同类型用户的深度访谈文本进行分析,总结了丽江慢活性呈现的多种维度,与本文的分析结果高度一致。崔庆明等^[32]的研究中被游客提及最多的是丽江古城古建筑、历史文化、特色酒吧和客栈、自然景观和气候等属性,以及“寻求艳遇、逃离和归属感”等古城的旅游价值,这些观点在本研究中通过一级特征维度全部刻画出来。古城的音乐元素,特别是音乐与酒吧文化组合对浪漫之都的地方构建,也能与李大为等^[33]关于丽江古城地方音乐对地方构建研究的观点基本一致。

与小数据量文本分析和访谈调查研究不同,基

于大量微博的LDA分析所获得的意象特征呈现了更多的维度、视角和载体,使形象特征更加丰富、立体、细腻且有画面感,能捕捉人与环境互动界面的特征。例如,孙九霞等^[34]基于访谈和观察法的研究发现,酒吧扩张使人们刻意追求丽江艳遇,使小资文化走向庸俗化,丽江古城正在演绎文化解构与重构的过程,但并未给出文化重构是好还是坏的结论。本文的LDA模型分析发现,酒吧与地方音乐的组合呈现是酒吧发展的特质,正如旅游者对古城的凝视一样,旅游者并非期待真实的艳遇,艳遇的本质是对浪漫之都的特殊呈现。再如,已有文献还没有充分讨论过丽江古城明信片的特色价值,除了联系旅游者的现实与梦想、此地方与彼地方,是否也隐约中传递对真正爱情的期许呢?

然而,要想全面解读丽江古城意象的形成原因和机制,还需在LDA分析基础上进一步综合运用更多方法和更深入的文献分析进行解释。例如,综合本文与白凯等^[11]的研究成果,丽江古城慢活形象的形成源于3个方面:由自然环境、历史文化积淀造就的原住民慢活理念;外来经营者对客栈和酒吧的营造和推广,特别是老板娘的形象,使其强化和鲜活;逃离城市、享受休闲时光的需求,使古城慢活更为显化。②丽江古城浪漫之都的形象也有其形成原因和机制:舒适的高原气候,独特纳西美食、民俗与建筑文化显然成为浪漫意境的主要环境基础;夜色中热闹的音乐主题酒吧、宁谧温馨的风情客栈,动静画面的自然交互中,更是把浪漫意境呈现得淋漓尽致;无数文人、艺术家定居古城,精美的明信片,以及充满浪漫气息的客栈与酒吧名称,也为古城添加了丰富的浪漫元素。

总之,将LDA模型与特征维度方法应用到旅游地意象特征的半定量提取,为旅游地意象感知研究提供了新的思路和方法。选择丽江古城意象感知的案例分析,其意义已并不局限于旅游地内部空间,而是对整个丽江地方文化的凝视与理解;有关丽江古城形象特征的结论,可以为旅游地政策制定、游客管理和可持续发展策略实施提供参考依据。

当然,本文仍有需要改进的地方。例如LDA模型并未考虑文本词语顺序,且微博文本相对较短,语言表述的随意性强,引入特征维度后的词语分类仍存在一定主观性等。因此,基于LDA模型进行的微博分析,获取的旅游者古城意象特征可能并不完整,或存在一些偏差。首先,主题提取结果是以正

面形象感知为主的,对部分旅游学者忧虑的文化变迁、过度商业化^[35]等发展特征的体现并不明显。这一结果符合微博用户以年轻人为主的属性特征,他们更能快速接受地方文化与现代精神的融合及其对地方意义的建构,甚至会成为参与建构的主体。第二,丽江古城包括大研、束河和白沙3个古镇,但LDA模型分析结果显示的丽江古城意象空间与完整的丽江古城并不耦合,特别是白沙古镇在主题分类结果中涉及较少。总之,实际应用LDA模型时,仍有必要参考地方资料和实地考察认知,以避免对特征词的过度解读或曲解。

参考文献(References)

- [1] 白凯, 赵安周. 城市意象与旅游目的地意象研究中的趋同与分野 [J]. 地理科学进展, 2011, 30(10): 1312-1320. [Bai Kai, Zhao Anzhou. Studies on convergence and divergence of city image and destination image. Progress in Geography, 2011, 30(10): 1312-1320.]
- [2] Çakmak E, Isaac R K. What destination marketers can learn from their visitors' blogs: An image analysis of Bethlehem, Palestine [J]. Journal of Destination Marketing & Management, 2012, 1(1-2): 124-133.
- [3] 乌铁红, 张捷, 李文杰. 地域文化差异对旅游者的旅游地意象感知影响: 以内蒙古自治区草原旅游地为例 [J]. 旅游学刊, 2010, 25(6): 66-72. [Wu Tiehong, Zhang Jie, Li Wenjie. The impact of regional cultural differences on tourists' perception of tourism destination image: A case study of tourists' perception of prairie image in Inner Mongolia Autonomous Region. Tourism Tribune, 2010, 25(6): 66-72.]
- [4] 滕茜, 杨勇, 布倩楠, 等. 基于网络文本的景区感知及互动研究: 以上海为例 [J]. 旅游学刊, 2015, 30(2): 33-41. [Teng Qian, Yang Yong, Bu Qiannan, et al. Research on the perception and interaction of tourist attractions in Shanghai based on web texts. Tourism Tribune, 2015, 30(2): 33-41.]
- [5] Aرسال I, Woosnam K M, Baldwin E D, et al. Residents as travel destination information providers: An online community perspective [J]. Journal of Travel Research, 2010, 49(4): 400-413.
- [6] 郭风华, 王琨, 张建立, 等. 成都“五朵金花”乡村旅游地形象认知: 基于博客游记文本的分析 [J]. 旅游学刊, 2015, 30(4): 84-94. [Guo Fenghua, Wang Kun, Zhang Jianli, et al. Rural tourism destination image of the case in "Five Golden Flowers", Chengdu: Based on the text mining of blogs. Tourism Tribune, 2015, 30(4): 84-94.]
- [7] 谢永俊, 彭霞, 黄舟, 等. 基于微博数据的北京市热点区域意象感知 [J]. 地理科学进展, 2017, 36(9): 1099-1110. [Xie Yongjun, Peng Xia, Huang Zhou, et al. Image perception of Beijing's regional hotspots based on microblog data. Progress in Geography, 2017, 36(9): 1099-1110.]
- [8] Lu Weilin, Stepchenkova S. User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software [J]. Journal of Hospitality Marketing & Management, 2015, 24(2): 119-154.
- [9] Alazzawi A N, Abdelmoty A I, Jones C B. What can I do there? Towards the automatic discovery of place-related services and activities [J]. International Journal of Geographical Information Science, 2012, 26(2): 345-364.
- [10] 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 Yuan, Xu Xin, Feng Xuegang, et al. Research on tourists' perceived image of ancient town using web text mining methods: A case study of Zhujiajiao. Tourism Science, 2013, 27(5): 86-95.]
- [11] 孟奕爽, 邓森文, 徐佳. 旅游目的地投射意象与游客感知意象差异分析: 以平江石牛寨景区为例 [J]. 中南林业科技大学学报(社会科学版), 2018, 12(3): 67-73. [Meng Yishuang, Deng Senwen, Xu Jia. Study on the differences between projection image and tourist perception image of tourist destination: Case of Shiniuzhai scenic spot, Pingjiang. Journal of Central South University of Forestry & Technology (Social Sciences), 2018, 12(3): 67-73.]
- [12] Büschken J, Allenby G M. Sentence-based text analysis for customer reviews [J]. Marketing Science, 2016, 35(6): 1-23.
- [14] 王永明, 王美霞, 李瑞, 等. 基于网络文本内容分析的凤凰古城旅游地意象感知研究 [J]. 地理与地理信息科学, 2015, 31(1): 64-67, 79. [Wang Yongming, Wang Meixia, Li Rui, et al. Destination image perception of Fenghuang Ancient Town based on content analysis of travelers' web text. Geography and Geo-information Science, 2015, 31(1): 64-67, 79.]
- [15] Weerdenburg D, Scheider S, Adams B, et al. Where to go and what to do: Extracting leisure activity potentials from web data on urban space [J]. Computers, Environment and Urban Systems, 2019, 73: 143-156.
- [16] Gao Baojun, Hu Nan, Bose I. Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews [J]. Decision Support Systems, 2016, 95: 1-11.
- [17] Andrienko G, Andrienko N, Bosch H, et al. Thematic patterns in georeferenced tweets through space-time visual analytics [J]. Computing in Science & Engineering,

- 2013, 15(3): 72-82.
- [18] Ramage D, Dumais S T, Liebling D J. Characterizing microblogs with topic models [C]// Association for the Advancement of Artificial Intelligence. Proceedings of the fourth international conference on weblogs and social media. Washington D C, USA: AAAI Press, 2010: 130-137.
- [19] Cheng Xueqi, Yan Xiaohui, Lan Yanyan, et al. BTM: Topic modeling over short texts [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(12): 2928-2941.
- [20] Lansley G, Longley P A. The geography of twitter topics in London [J]. Computers, Environment and Urban Systems, 2016, 58: 85-96.
- [21] Brandt T, Bendler J, Neumann D. Social media analytics and value creation in urban smart tourism ecosystems [J]. Information & Management, 2017, 54(6): 703-713.
- [22] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3 (4-5): 993-1022.
- [23] Ghosh D, Guha R. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and geographic information system [J]. Cartography and Geographic Information Science, 2013, 40(2): 90-102.
- [24] 关鹏, 王曰芬. 基于LDA主题模型和生命周期理论的科学文献主题挖掘 [J]. 情报学报, 2015, 34(3): 286-299. [Guan Peng, Wang Yuefen. Topic mining in scientific literature based on LDA topic model and life cycle theory. Journal of the China Society for Scientific and Technical Information, 2015, 34(3): 286-299.]
- [25] 谢昊, 江红. 一种面向微博主题挖掘的改进LDA模型 [J]. 华东师范大学学报(自然科版), 2013(6): 93-101. [Xie Hao, Jiang Hong. Improved LDA model for microblog topic mining. Journal of East China Normal University (Natural Sciences), 2013(6): 93-101.]
- [26] Hu Nan, Zhang Ting, Gao Baojun, et al. What do hotel customers complain about? Text analysis using structural topic model [J]. Tourism Management, 2019, 72: 417-426.
- [27] Guo Yue, Barnes S J, Jia Qiong. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation [J]. Tourism Management, 2017, 59: 467-483.
- [28] 南亭. 丽江古城的前世今生 [J]. 商业文化, 2017(34): 88-91. [Nan Ting. Past and present of Lijiang Ancient City. Business Culture, 2017(34): 88-91.]
- [29] 王鲁民, 吕诗佳. 建构丽江: 秩序·形态·方法 [M]. 北京: 生活·读书·新知三联书店, 2013. [Wang Lumin, Lv Shijia. Constructing Lijiang: Order, form, method. Beijing, China: SDX Joint Publishing Company, 2013.]
- [30] Huang Yuqian, Li Yue, Shan Jie. Spatial-temporal event detection from geo-tagged tweets [J]. ISPRS International Journal of Geo-Information, 2018, 7(4): 150. doi: 10.3390/ijgi7040150.
- [31] 白凯, 胡宪洋, 吕洋洋, 等. 丽江古城慢活地方性的呈现与形成 [J]. 地理学报, 2017, 72(6): 1104-1117. [Bai Kai, Hu Xianyang, Lv Yangyang, et al. Study on the identity with placeness of slow living in Lijiang. Acta Geographica Sinica, 2017, 72(6): 1104-1117.]
- [32] 崔庆明, 和琳珊, 徐红罡. 遗产旅游动机的核心-边缘结构研究: 以丽江为例 [J]. 旅游学刊, 2016, 31(10): 84-93. [Cui Qingming, He Linshan, Xu Honggang. The core-periphery structure of heritage tourism motivation in Lijiang. Tourism Tribune, 2016, 31(10): 84-93.]
- [33] 李大伟, 李诚固, 马佐澎, 等. 旅游影响下的地方音乐及其对地方的建构: 以丽江古城为例 [J]. 地理科学进展, 2018, 37(6): 833-843. [Li Dawei, Li Chenggu, Ma Zuo-peng, et al. Influence of tourism on local music and its construction to the place: A case study on the Lijiang Ancient Town. Progress in Geography, 2018, 37(6): 833-843.]
- [34] 孙九霞, 王心蕊. 丽江大研古城文化变迁中的“虚无”与“实在”: 以酒吧发展为例 [J]. 旅游学刊, 2012, 27(9): 73-83. [Sun Jiuxia, Wang Xinrui. The "nothing" and "something" during the cultural changes in Lijiang Ancient Town: A case study of Lijiang Ancient Town's bars. Tourism Tribune, 2012, 27(9): 73-83.]
- [35] 保继刚, 苏晓波. 历史城镇的旅游商业化研究 [J]. 地理学报, 2004, 59(3): 427-436. [Bao Jigang, Su Xiaobo. Study on control of tourism commercialization in historic town and village. Acta Geographica Sinica, 2004, 59 (3): 427-436.]

Tourism destination image perception analysis based on the Latent Dirichlet Allocation model and dominant semantic dimensions: A case of the Old Town of Lijiang

LIANG Chenchen¹, LI Renjie^{1,2*}

(1. School of Resources and Environmental Sciences, Hebei Normal University, Shijiazhuang 050024, China;

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

Abstract: To explore whether the Latent Dirichlet Allocation (LDA) topic model is suitable for analyzing the short text of tourism microblogs and whether the result can be consistent with the research results based on interviews and other data, this study established a destination image perception framework including four first-level dimensions and 10 second-level dimensions. Then the meaning of the topics is defined based on the dominant dimensions of each topic, which can reduce the subjectivity of researchers and help them to use the LDA model to extract the destination image perception quantitatively and objectively. The case study of the Old Town of Lijiang shows that in the first-level dimensions, the basic framework of image perception can be fully outlined through the five groups of core spatial and landscape elements, including the human settlements, music culture, character, leisure space and Naxi cuisine, and the special perception of the deep tourists, urban residents, young people and girls, and the characteristics of interaction of human and environmental elements. In the second-level dimensions, more detailed perception of destination image can be vividly presented from three aspects: the slow living in the Old Town of Lijiang, the culture of nightlife and romance, and tourists' perception of the fusion of local culture and modernity. This study proves the feasibility and advantage of this method—it shows that LDA is suitable for short text analysis of social media such as Weibo. Topic analysis based on dominant semantic dimensions successfully portrays the image perception of the Old Town of Lijiang and further analyzes the mechanism of image formation, and provides a new perspective for destination image perception, which has three values. It helps to accurately establish the basic framework of destination image perception; quantitatively extract the core dimensions of image perception; and deeply interpret the local meaning of destination image and clarify the relationship between cognitive, affective, and behavioral images.

Keywords: LDA model; dominant semantic dimensions; co-occurrence relationship; image characteristic; Sina Weibo; the Old Town of Lijiang