

基于多维度特征与 LDA 模型的城市旅游画像演化分析^{*}

叶光辉 徐 彤 毕崇武 李心悦

(华中师范大学信息管理学院 武汉 430079)

摘要:【目的】以具有时间属性的公众城市认知数据为研究样本,探寻城市画像主题变化特征与规律。【方法】选择城市旅游业为研究分面,利用 LDA 主题模型与多维度城市画像主题描述框架,从主题发展历程、一级与二级特征维度下的主题演化趋势这三个角度共同刻画城市画像在时间轴线上的变迁。【结果】就中国香港而言,城市旅游画像在主题演化过程中并未出现周期性波动规律,但在主题感知维度的表现却有着明显的主次之分;旅游游览、旅游交通以及旅游娱乐是构成公众对于中国香港城市形象认知体系的主要属性;其中,旅游游览往往总领主题演化的发展方向,旅游娱乐则主要分布于演化轴线的前期与末期,而旅游交通处于演化中期;除此以外,演化路径中各主题在各主导维度下又分别有着稳定的标志性意象载体。【局限】研究结论有待综合多样化的城市数据以及分析方式实现进一步深化与拓展。【结论】本文提出的基于旅游特征维度的主题演化研究思路,实现了城市旅游画像的演化轨迹描绘,同时研究结论能够为政府及行业相关举措的规划与实施提供参考。

关键词: 城市画像 演化分析 特征维度 LDA 模型

分类号: TP391

DOI: 10.11925/infotech.2096-3467.2020.0606

引用本文: 叶光辉, 徐彤, 毕崇武等. 基于多维度特征与 LDA 模型的城市旅游画像演化分析[J]. 数据分析与知识发现, 2020, 4(11): 121-130. (Ye Guanghui, Xu Tong, Bi Chongwu, et al. Analyzing Evolution of City Tourism Portraits with Multi-Dimensional Features and LDA Model[J]. Data Analysis and Knowledge Discovery, 2020, 4(11): 121-130.)

1 引言

城市画像技术旨在利用数据识别、理解公众视域下更为全面与清晰的城市特征,从而在公共服务、城市治理等方面嵌入有利要素,以此提高人民生活水平,推动城市稳健发展。

随着旅游经济在城市整体经济中比重不断增大,旅游行业备受关注,逐渐成为城市发展的基本单元与关键节点。但事实上,不同于传统旅游服务行业,科技的飞速发展使得打造智慧旅游行业成为城

市创新变革的重要发力点。新业态旅游产业的成长需要不断倒逼学界重视。然而,就目前来看,有关于城市旅游画像的分析思路与模型不够成熟与深入,现有研究大多局限为借助地理位置数据以揭示群体人流、物流等移动规律,从而从宏观角度刻画城市旅游产业结构,但却无法通过主题化或标签化的数据更加准确、快速地服务于旅游行业的需求升级。

为此,本文以社交媒体语料为研究样本,在城市旅游画像描述框架下,引入时间维度,从旅游分面揭示公众视域下城市画像的核心演化路径,在更长的

通讯作者: 叶光辉, ORCID: 0000-0001-8111-5034, E-mail: 3879-4081@163.com。

^{*}本文系国家自然科学基金项目“基于标签语义挖掘的城市画像计算与应用模型研究”(项目编号: 71804055)和中央高校基本科研业务费资助(创新资助项目)“基于时序分析的城市画像动态演化研究”(项目编号: 2019CXZZ130)的研究成果之一。

时间周期内探索城市核心主题演化的特征与规律。本研究充分借助网络信息技术,整合并分析旅游方面上城市画像相关资源,一方面有利于旅游业态的个性化与智慧化发展,另一方面也将推进政府及有关部门关于精细化城市治理方案的拟定,为城市可持续发展注入新的活力。

2 研究现状

目前,围绕城市画像的研究可细分为基于多源多面城市数据的融合机制、实现模式研究以及垂直分面下的城市画像研究两大类。前者侧重分析城市画像与城市管理之间的逻辑关联,典型研究如:杜智涛等着眼于城市系统固有的复杂多样性特征,提出“资源-领域-实现”三维度城市画像应用体系^[1];张杨焱等深度剖析城市数据内容以及特征,构建自底向上的面向应用服务的数据融合体系与框架^[2];毕崇武等以用户生成内容数据源,利用标签相似度和凝聚式层次聚类算法,获得多个维度下公众对各城市整体状况的共同感知^[3]。后者是一系列的实证研究,更多地探索城市不同业务的特征体系,如胡燕祝以城市安全生产监管为研究分面,构建辅助当地安全生产监管力量部署的城市安全画像^[4];任思源等对城市中不同社会群体与城市区域的快递面单数据进行分析,构建寄递大数据城市画像,以发现城市居民寄递行为规律与异常^[5]。与此同时,移动网络及通讯技术的进步和广泛应用使得互联网中累积了海量涵盖城市文化、功能与特性的城市数据,其可获得性与可用性的日益提升也在不断推进高质量城市画像科研成果的产出以及研究领域的横纵双向拓展。

此外,针对社交媒体中文本信息的主题演化相关理论、技术、方法和模型也在不断发展与完善。其中,LDA模型因能够提供科学可靠的量化研究主题方法,受到学界的广泛关注。典型性研究如:吴查科等以11种核心期刊文献为语料库,通过LDA模型刻画图书馆学领域的主题演化路径^[6];蒋翠清等针对产品论坛构建主题演化模型,并从主题标签、热度以及主题词热度三个方面探讨在线产品评论的主题发展过程^[7];秦晓慧等利用主题关联过滤规则优化基于LDA模型构建主题演化路径,更为准确地解释主

题新生、消亡、继承、分裂和合并的演化轨迹^[8]。

这些研究一定程度上肯定了LDA模型在主题挖掘中的重要价值,但是在具体研究中主题命名的确定或者主题语义的描述和凝练等方面,仍存在较强的主观性以及不确定性。因此,区别于本团队先前关于城市画像描述框架及网络结构的城市画像领域相关研究^[3,9],本文从游记文本信息着手,开展城市旅游画像主题特征的探索,利用多维度的城市画像主题描述框架,对LDA模型提取结果进行类目划分,并引入主题维度相似性的判断规则以复刻公众感知视域下城市画像发展轨迹,由此探究城市画像的演化规律以及演进过程中城市旅游主题在一级与二级维度下的变迁特征。

3 研究方法设计及实现机制

3.1 研究框架

基于已有研究成果,本研究设计了公众视域下城市画像主题研究框架,如图1所示。具体来说,本文将从以下5个方面展开研究:

(1)借助LDA模型的城市旅游画像主题生成:以预处理后各个时间窗口内的游记文本为研究样本,借助LDA模型进行主题提取工作,从大量文本中聚类若干旅游分面上的城市画像相关主题。

(2)城市旅游画像主题描述维度框架的拟定:参照文献调研结果,综合本研究目的等因素,明确旅游分面上城市画像主题描述维度框架。

(3)旅游主题-特征向量的计算:对照(1)中主题-词矩阵以及(2)的城市旅游画像主题描述维度框架,将每个主题下的高概率特征词划分到不同层次的维度内,形成各时间段内的主题-特征向量。

(4)核心主题演化路径挖掘:根据各月份主题-特征向量,利用主题余弦相似度的关联规则,确定核心演化路径,探寻城市旅游画像主题演化规律。

(5)核心演化路径下的主题内容变迁分析:从多维度城市旅游画像主题描述框架涉及的各个方面如旅游娱乐、旅游游览等着手,明晰城市画像主题的变迁特征,其中一级维度决定城市核心主题基本框架,二级维度刻画游客对于该主题下的关注侧重点。综合二者分析结果可提高城市画像主题演化历程描述的形象性与全面性。

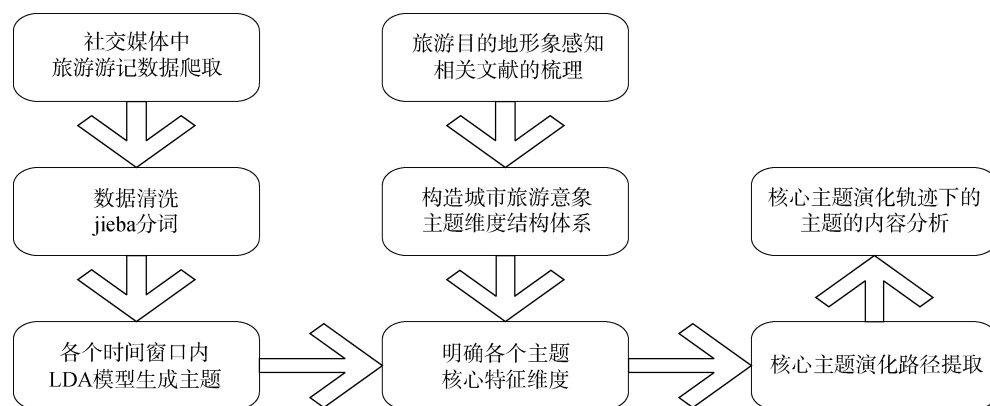


图1 公众视域下城市画像主题研究框架

Fig.1 The Research Framework of Urban Portrait Subjects in the View of the Public

3.2 旅游分面上多维度城市画像主题描述体系

通过文献资料收集与分析,梳理围绕旅游目的地感知、城市旅游主题等的相关研究成果,如表1所示。

表1 围绕旅游相关主题的研究成果

Table 1 Relevant Research Around the Perception of Tourist Destinations, Urban Tourism Themes, etc.

作者	特征维度选取
凯文·林奇 (Kevin Lynch) ^[10]	道路;边缘;区域;节点;标志物
李茜等 ^[11]	旅游氛围(气候、价格、市民素质……);旅游传播(旅游口号、历史文化……);旅游基础设施(交通方式、城市环境……);旅游资源(景点内涵、资源丰富……);旅游服务(景点讲解、纪念品)
陆易 ^[12]	游客基本特征(兴趣偏好、地点偏好……);认知形象(吃、住、行……);情感感知(资源环境、景观特色……);总体感知(满意度、重游意愿)
马天等 ^[13]	地域条件(天气、空气质量……);文化(人文素质、民风……);社会环境(市容市貌、治安条件……);旅游吸引物(自然风光、历史遗迹);旅游配套设施与服务(住宿、食物……)
张文等 ^[14]	游客基本特征(旅游次数、停留天数……);认知形象(自然与人文景点、旅游设施与服务……);情感形象(愉快、激动);重游意向
Gallarza等 ^[15]	景观环境;自然状况;文化吸引;夜间娱乐;购物设施;信息指引;体育设施……
陈建明 ^[16]	规划布局;特色建筑;生态环境;经济发展服务接待;气候条件;历史文化;通讯设施;旅游咨询;景色优美;节庆活动;旅游商品……
韦晨 ^[17]	旅游吸引物(自然景观、风味美食……);旅游设施与服务(市内交通、购物……);旅游环境(价格水平、卫生状况……)

出于不同研究目的需要,学者们对于旅游目的地形象感知维度拟定存在一定的差异。不过这一点又是可以理解的,因为完全列举出形象感知要素及其组合是非常困难,也是没有意义的。因此,本文借鉴已有的类目体系,综合本研究的实际情况进行调整,以此构造基于中国香港旅游的多维度城市画像主题描述体系,如表2所示。

特征维度一共7个一级维度以及18个二级维

表2 多维度城市画像主题描述体系

Table 2 Description System of City Portrait

一级指标	二级指标	特征示例词
旅游娱乐 A	特色美食 A ₁	甜品、茶餐厅、生蚝、牛腩
	购物商场 A ₂	面膜、海港城、青衣城、护肤品
	节事活动 A ₃	庆典、万圣节、舞火龙、新春
旅游游览 B	自然风光 B ₁	大屿山、海滨、麦理浩径、日落
	人文景观 B ₂	渔村、教堂、博物馆、香港大学
	现代城市景观 B ₃	中环半山扶梯、旺角、尖沙咀
	主题乐园 B ₄	迪士尼乐园、海洋公园
旅游环境 C	气候环境 C ₁	小雨、天气、热、太阳
	生态环境 C ₂	牛、清新、松叶、小溪、野猪
	景区秩序 C ₃	排队、工作人员、服务态度
旅游服务 D	酒店 D ₁	住宿、半岛酒店、旅舍、民宿
	旅行社 D ₂	旅行团、导游
旅游交通 E	本地交通 E ₁	天星小龙、缆车、地铁
	城市对外交通 E ₂	飞机、高铁、航空公司、机场
城市氛围 F	城市治安 F ₁	社区服务、政府、交通警察
	市容市貌 F ₂	干净、著名、繁华
	市民素质 F ₃	居民、素质、本地人
其他 G	通行证、行程安排等	通行证、港币、流量

度,特征示例词是指用来表述对应主题维度的词语。

3.3 旅游分面上城市画像主题挖掘的实现机制

(1) 基于LDA模型的主题演化分析方式

LDA (Latent Dirichlet Allocation)是文档主题生成模型,是一种典型的词袋模型,视一篇文档为由一组词构成的一个集合。LDA通过数据训练得到,实现文档-词矩阵到文档-主题与主题-词矩阵的转化,达到主题聚类或文本分类的目的。LDA模型生成过程如图2所示。

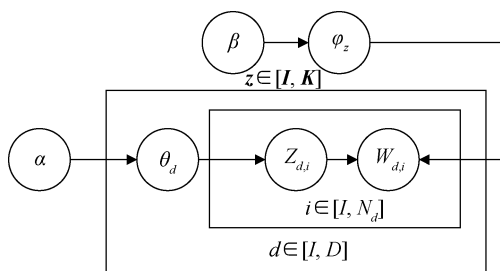
图2 LDA模型生成过程^[6]

Fig.2 The Process of LDA Model Generation^[6]

D 表示原始数据中的文档, Z 为数据集中的主题, α 是反映潜在主题强弱的先验概率, β 为描述所有潜在主题词概率分布的先验概率; θ_d 与 φ_z 分别代表文档-主题分布与主题-词分布; $W_{d,i}$ 为文档的词向量。

本文采用先离散分析(Pre-discretized Analysis)方法开展城市画像主题动态演化研究。先依据文本的时间属性将信息离散到时间序列中的时间窗口内,然后针对各个时间窗口内的文本集合进行处理,以追踪主题的演化路径。

(2) 各时间窗口内的主题发现

在 Python 环境下利用 jieba 将游记转换为文档-词矩阵,运用 Gensim 工具包进行主题建模。在主题模型训练过程中,依据困惑值确定需要预先设定的主题数 K ,困惑值越小,模型性能越好。在明确最优主题数的基础之上生成最终主题模型,并输出主题-词矩阵,为后续开展的基于主题特征维度的主题分布以及主题演化研究奠定基础。

(3) 时间轴线上的主题关联规则建立

研究主题之间演化关系的常用方法包括直接关联法、相似度阈值法、距离阈值法等,本文依照以下

步骤以明确主题之间的关联关系:

① 相似度计算

计算邻近时间窗口之间的余弦相似度衡量主题之间的关联强度,如公式(1)所示,其中涉及的变量含义如表3所示。

$$Sim(T_i^t, T_i^{t+1}) = \frac{\sum_{k=1}^n P_k(T_i^t) \times P_k(T_i^{t+1})}{\sqrt{\sum_{k=1}^n P_k^2(T_i^t)} \times \sqrt{\sum_{k=1}^n P_k^2(T_i^{t+1})}} \quad (1)$$

表3 主题相似度计算公式变量解释

Table 3 The Meaning of Variables in the Topic Similarity Calculation Formula

字符	含义
T_i^t	时间窗口 t 下的主题
T_i^{t+1}	时间窗口 $t+1$ 下的主题
$P_k(T_i^t)$	T_i^t 对应的主题-特征矩阵中的概率值
$P_k(T_i^{t+1})$	T_i^{t+1} 对应的主题-特征矩阵中的概率值

②主题关联构建

将相邻时间窗口内的关联主题按照相似度大小进行排序,相似度最大的 T_{i+1} 视为该主题 T_i 的后向主题, T_i 为 T_{i+1} 的前向主题。一个主题至少拥有一个后向主题;一个主题同时为多个主题的后向主题。

通过各个时间段内主题相似度的排序,对主题关联强弱展开过滤,最终排序结果服务于核心主题演化路径的确定。

4 城市旅游画像演化分析

4.1 数据获取与处理

(1) 数据采集

选择马蜂窝作为研究样本来源网站,以“中国香港”为检索关键词在该平台社区版块搜索2019年1月-2019年12月的相关游记,利用网络爬虫技术采集游记数据共5 367条。每条记录包括用户ID、出行时间、游玩时间、同行成员以及游记文本等7个字段。为保证数据质量,确保后期数据分析工作的科学性与严谨性,对于收集得到的数据集进行必要的筛选、查重、格式规范等操作,得到有效记录2 943条。其中80%以上游记的文本长度大于500字,相较于针对某一景点或旅游产品的评价短文本而言,其属于长文本,满足LDA主题模型对于数据样本的

基础要求。

(2) 时间划分

本文探寻不同时间段内城市画像主题演化路径,因此以月份作为时间划分单位,对游记总集进行切割,共计生成 12 个时间窗口。各个时间窗口内的游记文本数量统计如表 4 所示。

4.2 城市旅游画像主题特征分布

对照 3.2 中多维度城市画像主题描述体系,对主题下的特征词进行类目划分。以 2019 年 1 月数据为例,当主题数 $K=5$ 时,模型困惑值最小,由此得到该 LDA 模型下生成的 5 个主题。各主题特征词所属维

表 4 不同时间窗口城市旅游分面数据量

Table 4 The Amount of Urban Tourism Facet Data in Different Time Windows

时间窗口	游记数量	时间窗口	游记数量
2019.1	163	2019.7	233
2019.2	444	2019.8	93
2019.3	469	2019.9	77
2019.4	440	2019.10	89
2019.5	444	2019.11	40
2019.6	402	2019.12	49

度划分如表 5 所示,各主题在不同维度下的累积概率如表 6 所示。

表 5 各主题特征词所属维度划分

Table 5 The Dimension Division of Each Topic Feature Word

Topic 1	概率	所属维度	Topic 2	概率	所属维度	Topic 3	概率	所属维度	Topic 4	概率	所属维度	Topic 5	概率	所属维度
香港海洋公园	0.013	B_4	沙茶面	0.003	A_1	油麻地	0.002	B_3	大澳	0.019	B_2	酒店	0.019	D_1
酒店	0.005	D_1	酒店	0.009	D_1	旅舍	0.002	B_1	挪亚方舟	0.010	B_4	迪斯尼	0.010	B_4
味道	0.004	A_1	机场	0.008	E_2	麦理浩	0.008	B_1	餐厅	0.008	A_1	山顶	0.008	B_3
维多利亚港	0.003	B_3	码头	0.008	E_1	码头	0.006	E_1	当代艺术中心	0.008	B_4	地铁	0.008	E_1
海洋公园	0.003	B_4	维多利亚港	0.006	B_3	装备	0.006	G	城市规划	0.007	F_2	港币	0.007	G
甜品	0.003	A_1	长洲岛	0.006	B_1	西贡	0.004	B_1	粥	0.007	A_1	中环	0.007	B_3
拉面	0.003	A_1	旺角	0.005	B_2	水库	0.004	B_2	小朋友	0.007	G	机场	0.007	E_2

表 6 各主题在不同维度下的累积概率

Table 6 The Cumulative Probability of Each Topic in Different Dimensions

主题	旅游娱乐	旅游游览	旅游环境	旅游服务	旅游交通	城市氛围	其他
Topic 1	0.056	0.046	0.003	0.005	0.010	0.001	0.010
Topic 2	0.031	0.061	0.013	0.010	0.031	0.000	0.012
Topic 3	0.007	0.053	0.019	0.002	0.018	0.001	0.005
Topic 4	0.048	0.101	0.015	0.000	0.042	0.013	0.022
Topic 5	0.035	0.111	0.004	0.028	0.064	0.002	0.023

综合文献[18]以及本次研究的实验结果,发现各主题下的特征词出现概率呈现出明显的长尾效应,过半主题下第 100 个词后概率值变化趋于平缓,由于尾部概率值较低的特征词对于主题的语义表征贡献较少,因此选择各个主题下概率较大的前 100 个特征词进行特征维度划分,以分别计算各个主题中的特征词在不同维度下的累积概率。由于部分特征词无法划分到更为精细化的二级指标类目,因此

划分时便不再向二级维度细化展开,例如特征词“交通”所属维度为 E,即旅游交通;特征词“风景”所属维度为 B,即旅游游览。

4.3 城市旅游画像演化路径获取

运用 4.2 节对于某一时间窗口内的主题特征分析方式处理每个月份的游记数据,并将游客对于城市旅游主题的感知差异抽象为主题-特征向量的表现形式,如 $T_i = \{(V_1, P_1), (V_2, P_2), \dots, (V_7, P_7)\}$ 。其中, V_i 为 T_i 中的特征词所属维度, P_i 为该维度下特征词的累积概率值,由此生成 12 个相互独立的主题-特征向量。

结合主题关联关系挖掘方法,从中抽取强关联关系,而后依据主题相似度均值从中过滤部分主题关联关系,剩下的为城市旅游画像主题核心演化路径的待选集合。从整体主题演化关系中截取出来的部分主题相似计算结果,如表 7 所示(Pre-Theme 表示前向主题, Next-Theme 表示后向主题;波浪线标识的为经过上述方式过滤后保留的主题)。

表7 主题关联关系计算(部分)

Table 7 The Calculation of Subject Relationship (Partial)

Pre-Theme	Theme	Sim	Theme	Next-Theme	Sim
1-topic1	2-topic6	0.959 5	2-topic1	3-topic10	0.967 6
1-topic2	2-topic3	0.954 6	2-topic2	3-topic10	0.986 7
1-topic3	2-topic3	0.986 1	2-topic3	3-topic23	0.981 5
1-topic4	2-topic3	0.951 7	2-topic4	3-topic26	0.987 3
1-topic5	2-topic4	0.981 0	2-topic5	3-topic11	0.981 8
			2-topic6	3-topic4	0.997 0
			2-topic7	3-topic1	0.986 8
Average		0.966 5	Average		0.984 1

综合上表7与其他月份的分析数据得知,各个主题可能具有新生、成长以及消亡三种状态:

(1)新生:当前主题仅拥有后向关联主题,如2-topic2;

(2)成长:当前主题拥有前向与后向主题,如2-topic4;

(3)消亡:当前主题仅拥有前向主题。

在整个研究时间段内,主题演化关系可能存在继承、合并两种情况,如图3所示。

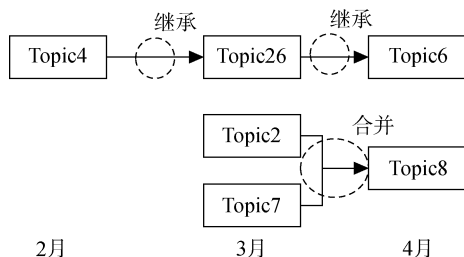


图3 主题之间的关联类型

Fig.3 The Type of Association Between Topics

(1)继承:当前主题为上一时段内的某单一主题的延续;

(2)合并:当前主题与上一时段内的两个或两个以上的主题具有关联关系。

与其他类似研究不同,由于本文对于主题相似度的精确值为小数点后4位数,因此导致主题之间仅存在合并关系,尚未呈现出明显的分裂关系。

综合上述对于三种主题状态以及两种可能的关联类型,参考文献[19]对于核心主题演化路径的确定方式,本研究将具有新生特征且连续演化两个月以上(即为继承、合并)的主题演化关系定义为城市

旅游画像演化核心路线。

由此,在2019年1月-2019年12月共挖掘得到5条核心演化路径。其中一条演化路线如图4所示,该路径涉及13个相关主题,2月份的topic2在4个月的时间推移中,与其他的11个主题相互作用,最终融合为2019年6月的topic5,该路径中共包含继承以及合并两种主题演化关系类型,演化结构并不复杂。

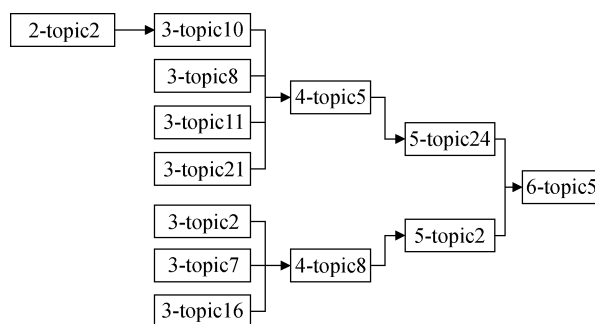


图4 城市旅游画像核心演化路线(部分)

Fig.4 The Core Evolution Route of Urban Tourism Portrait (Partial)

4.4 核心演化路径中的主题内容分析

为探求主题演化特征与规律,综合核心演化轨迹中各主题的特征词分类结果,从特征词所属一级与二级维度出发,更为精细地描绘游客感知的城市旅游画像演化脉络。

(1) 一级维度的主题特征演化分析

以图4所示的核心主题演化路径为例,该路径下的13个主题所属一级维度的概率变化如表8所示(其中波浪线标识的为各主题所属一级主导维度下的累积概率值)。该主题演化路径中,游客主要关注“旅游游览”、“旅游娱乐”以及“旅游环境”这三个方面。在主题演化历程的第一阶段,主题2-topic2是以“旅游娱乐”为主导维度的主题;“旅游环境”仅在第二阶段出现一次;而“旅游游览”则是占据核心地位的感知维度,分布于近乎3/4的演化轴线上。该维度共涉及9个关联主题,以3-topic8、3-topic11最具有代表性。这也表明城市赋有的景点形象为外来游客认识与记忆城市的关键分面。

通过对5条演化路径分析的综合可以发现以下规律:

①相较于具有周期性特征的旅游淡季,当细

表 8 第一条核心主题演化路径下的主题所属一级维度概率

Table 8 The Probability Statistics of the First Dimension of the Theme Under the First Core Theme Evolution Path

所属阶段	主题	旅游娱乐	旅游游览	旅游环境	旅游服务	旅游交通	城市氛围	其他
1	2-topic2	<u>0.030</u>	0.100	0.018	0.001	0.010	0.002	0.006
	3-topic10	0.045	0.207	<u>0.053</u>	0.000	0.040	0.004	0.033
	3-topic8	0.008	<u>0.196</u>	0.077	0.000	0.055	0.000	0.028
	3-topic11	0.014	<u>0.175</u>	0.004	0.000	0.006	0.014	0.011
2	3-topic21	0.027	<u>0.136</u>	0.008	0.004	0.038	0.004	0.001
	3-topic2	0.037	<u>0.064</u>	0.021	0.012	0.031	0.005	0.030
	3-topic7	0.038	<u>0.083</u>	0.017	0.022	0.025	0.002	0.018
	3-topic16	0.041	<u>0.084</u>	0.011	0.035	0.01	0.000	0.026
3	4-topic5	0.017	<u>0.127</u>	0.032	0.005	0.028	0.001	0.012
	4-topic8	0.035	<u>0.137</u>	0.017	0.033	0.044	0.002	0.053
4	6-topic5	0.015	<u>0.130</u>	0.005	0.019	0.02	0.000	0.029

化到外来人员对于一个城市的旅游形象感知时,可以发现城市旅游画像演化过程中并未呈现出明显的周期性规律。但是,公众通过旅游的方式得到的关于城市印象主题感知维度却有显著的主次之分。

②公众对于城市旅游画像感知以“旅游游览”、“旅游娱乐”以及“旅游交通”为重点。其中,“旅游游览”往往总领主题演化的整体发展方向,“旅游娱乐”主要分布于演化轴线的前期与末期,而“旅游交通”则更多地分布于演化中期。此外,上述三者之间的演化关系在所有演化路径中也表现突出。一方面体现出游客对这三个方面的交叉关注;另一方面也表明,在旅游分面上,游客对于在城市内发生的行为活动以及城市自身所具有的空间与景观特征更为在意。

③“城市氛围”以及“旅游服务”是公众所感知到的城市旅游画像主题的次要维度,在多个演化阶段曾出现为 0 的特征词累积概率。不过,关注度较低并不代表可以忽略其在城市旅游业发展中发挥的重要作用。对于城市旅游业规划与发展而言,其社会环境治理以及基础服务设施的建设仍然有助于提升游客旅游体验,这已成为各界普遍认同的观点。同时,在 KANO 模型中,用户的需求可以细分为必备型、魅力型、无差异型和反向型^[20]。对于中国香港而言,“城市氛围”以及“旅游服务”可能属于必备型的公众旅游需求,即这是一个旅游城市应该具有的因素,只有在城市建设中感受到这两方面的完全缺失

才会改变城市在公众心中的印象。但是需要指出的是,这一结论仍需进一步的社会调研工作加以佐证。

(2) 二级维度的主题特征演化分析

基于对核心主题演化路径中主题所属一级主次维度的挖掘结果,针对这三个主要维度展开更为细致和精确的二级特征维度分析。

①旅游游览

旅游游览包含自然风光、人文景观、现代城市景观以及主题乐园 4 个二级指标,其在 5 条演化路径中一直属于核心感知主题。具体而言,主要从“主题乐园”、“自然风光”以及“现代城市景观”这三个视角表现了二级维度下游客对于中国香港旅游画像的感知框架:

1) 主题乐园。3-topic10、4-topic8 以及 6-topic5 是贡献率最高的相关主题,三个主题借助特征词“迪士尼乐园”、“香港海洋公园”、“挪亚方舟”等直接的场馆名称,并结合“米奇”、“旋转木马”、“企鹅乐园”等项目描述以及“小朋友”、“宝贝”等人物描写,使得温馨、欢乐的游乐氛围与主题乐园融为一体。

2) 自然风光。4-topic10、6-topic1、5-topic17 是自然景观感知的主要主题,重点刻画游客对中国香港自然风景的美好体验。三个主题都直接或隐含表达山与水是游客来港拥抱自然的落脚点。并且充分体现了“麦理浩径线”是中国香港重要旅游资源组合线路,它将“万宜水库”、“浪茄湾”、“西贡”等地,结合“徒步”这一新式旅游释压方式呈现给更多的潜在游客。

3) 现代城市景观。该维度涉及 40 个主题,其中以 5-topic18、6-topic3、7-topic8、10-topic1 最具代表性。直接或间接描述中国香港中西区以及油尖旺区的城市景点,特征词包括“乐华南邨”、“彩虹邨”等描绘具有特色风格的城市建筑;“维多利亚港”、“太平山顶”等领略中国香港最美夜景的打卡点;“油麻地”、“庙街”等蕴含地道中国香港风情与风味的街道,精细刻画了充满“文艺气息”、“江湖气”等氛围要素的多元繁华现代城市画面。

此外,在这 5 条核心演化路径中,第一条与第二条的主题二级聚集维度分别为“主题乐园”、“自然风光”,其他三条均以“现代城市景观”为主导维度,该分析结果也与文献[21]通过实证研究得出的结论

“游客对于中国香港景点的感知大多集中体现在现代都市景观”在一定程度上保持一致。

② 旅游交通

旅游交通一级维度下包括城市本地交通与城市对外交通这两个二级指标。交通是顺利出行、愉快旅行的充分必要条件,因此其核心程度仅次于旅游游览。从各条演化路径包含的主题中可以得知,相较于如何抵达目的地的交通方式与渠道的选择,游客更加注重当地交通的实际便利性。

在城市对外交通中,飞机更为普遍,而随着广深港高速铁路中国香港段的开通,动车出行也逐渐成为更多游客抵港的选择。

至于城市本地交通,通过主题包含的特征词可以看出“地铁”、“天星小轮”、“巴士”是具有游客共性的当地出行选择。由于中国香港地铁其站点富有复古感,且体现不同文化路线的装修风格,因此地铁对于外来游客而言,成为兼具出行与游览二重功效的交通方式。而“天星小轮”与“巴士”对于游客而言,都具有相似的旅游体验。“天星小轮”是旅客游览维多利亚港最质优价廉的渡海交通工具,而中国香港独有的巴士——“叮叮车”也是游客体验中国香港文化的方式。从上述分析可以看出,中国香港本地交通不同于其他城市,其具有个性化的城市风格,已经成为游客对于中国香港旅游画像感知的一个重要节点。

③ 旅游娱乐

旅游娱乐是除去景点游览之外的,游客在港所能感知与体验的其他旅游事物与活动,而且由上文可知,旅游娱乐为主导维度的主题通常出现在主题演化轴线的前期与末期,一定程度上表明旅游娱乐在中国香港旅游吸引力方面发挥着不容小觑的作用。从该维度所包含特色美食、购物商场、节事活动这三个二级指标出发,以明晰游客对于中国香港旅游娱乐的感知内容:

1) 从特色美食容纳的特征词可以看出,游客对中国香港美食的印象以“翠华餐厅”、“华嫂冰室”、“澳洲牛奶公司”、“添好运点心专门店”、“池记”等店铺为代表,以“车仔面”、“鱼蛋”、“菠萝包”、“叉烧包”等菜品为特色,从侧面体现出茶餐厅在中国香港饮食文化中的独特地位。

2) 至于购物商场,相关主题中的高频特征词包

括“化妆品”、“面膜”、“护肤品”、“香水”等详细的购物产品;“青衣城”、“海港城”、“东荟城”等大型商场为中心内部空间描述词;以及“折扣”、“便宜”、“价格”、“性价比”等消费心理描述词,折射出游客来港购物热情,与中国香港一直以来的“购物天堂”的美誉呈现出极好的耦合。

3) 节事活动是节庆、事件等精心策划的各种活动的简称,根据资料显示中国香港包括年宵花市、舞龙舞狮、音乐节等多种多样的节庆样式。但在5条演化路径中,可以清晰地发现,其热度与特色美食、购物商场之间有明显的差异,仅在为数不多的主题内有所体现,如4-topic6、7-topic12等,且包含的特征词更多地聚焦于有着浓厚节日氛围的“新春”、“庆典”、“万圣节”、“舞火龙”等直接感知体验描述词。

根据以上分析,外来人员对于中国香港旅游画像感知在时间上并没有波动性的变化规律,但是在各个维度上却分别有较为稳定的标志性意象载体。由此推测,游客对于未来各主题特征维度的感知趋势在不出意外的情况下仍会保留一直以来的中心与热点。

5 结 语

本文以城市旅游为研究分面,提出基于旅游特征维度的主题演化研究思路,借助LDA模型识别社交媒体中游记文本中潜在主题,而后利用主题相似性获取主题之间关联关系,探测时间轴线上城市旅游主题演进路径,并从各主题所属一级与二级维度视角挖掘演化规律。

本研究是在2019年1月-2019年12月这一时间周期下,引入特征维度LDA模型的城市画像分析的一种尝试,但尚存一些内容有待完善:

(1) 现在网络平台上发布的游记大多数以自由行旅行记录为主,因此仅依赖网络文本数据研究,造成缺乏一定量的跟团游、半自由行等以其他形式旅行的游客旅游感受数据的收集,这可能导致对城市旅游画像在各维度上的意象载体梳理的不完整以及解释的不够充分。因此,在今后的研究中可结合社会调查等研究方法,从多维度以及更长时间跨度进一步展开城市旅游画像研究。

(2) 仅探讨了公众对于旅游分面上的城市形象感知,缺乏探寻公众给予城市意象的情感分析,在后

续的研究中将予以完善,使得城市画像研究结果更加丰富饱满。

参考文献:

- [1] 杜智涛,李纲.面向精细化治理的城市画像:构成要素与应用体系[J].图书情报知识,2019(4):43-51.(Du Zhitao, Li Gang. City Portrait Oriented to Refined Governance: Constitutive Elements and Application System[J]. Documentation, Information & Knowledge, 2019(4): 43-51.)
- [2] 张杨斌,谢辉,毛进,等.面向城市数据画像构建的多源数据需求与融合方法研究[J].情报理论与实践,2020,43(6):88-96.(Zhang Yangyi, Xie Hui, Mao Jin, et al. Multi-source Information Need and Data Fusion Methods Oriented to the Construction of City Profile[J]. Information Studies: Theory & Application, 2020, 43(6): 88-96.)
- [3] 毕崇武,叶光辉,李明倩,等.基于标签语义挖掘的城市画像感知研究[J].数据分析与知识发现,2019,3(12):41-51.(Bi Chongwu, Ye Guanghui, Li Mingqian, et al. Discovering City Profile Based on Tag Semantic Mining[J]. Data Analysis and Knowledge Discovery, 2019, 3(12): 41-51.)
- [4] 胡燕祝.城市安全生产画像构建与应用探讨[J].安全,2019,40(10):1-7.(Hu Yanzhu. Discussion on the Construction and Application of Urban Safety Production Portrait[J]. Safety & Security. 2019, 40(10): 1-7.)
- [5] 任思源,郭斌,张曼,等.寄递大数据城市画像[J].浙江大学学报(工学版),2019,53(9):1779-1787.(Ren Siyuan, Guo Bin, Zhang Man, et al. Urban Profiling Using Express Big Data[J]. Journal of Zhejiang University(Engineering Science), 2019, 53(9): 1779-1787.)
- [6] 吴查科,王树义.基于LDA的国内图书馆学研究主题发现及演化研究[J].新世纪图书馆,2019(7):90-96.(Wu Chake, Wang Shuyi. Research on the Theme Discovery and Evolution of Domestic Library Science Research Based on LDA[J]. New Century Library, 2019(7): 90-96.)
- [7] 蒋翠清,吕孝忠,段锐.基于主题模型的产品在线论坛主题演化分析[J].系统工程学报,2019,34(5):598-609.(Jiang Cuiqing, Lü Xiaozhong, Duan Rui. Analyzing Topic Evolution of Online Product Forum Based on Topic Model[J]. Journal of Systems Engineering, 2019, 34(5): 598-609.)
- [8] 秦晓慧,乐小虬.基于LDA主题关联过滤的领域主题演化研究[J].现代图书情报技术,2015(3):18-25.(Qin Xiaohui, Le Xiaoqiu. Topic Evolution Research on a Certain Field Based on LDA Topic Association Filter[J]. New Technology of Library and Information Service, 2015(3): 18-25.)
- [9] 叶光辉,徐彤.基于演化分析的动态城市画像研究[J].数据分析与知识发现,2020,4(9):100-110.(Ye Guanghui, Xu Tong. Dynamic City Profile Based on Evolutionary Analysis [J]. Data Analysis and Knowledge Discovery, 2020, 4(9): 100-110.)
- [10] 林奇·凯文.城市的印象[M].项秉仁译.北京:中国建筑工业出版社,1990.(Lynch K. The Image of the City[M]. Translated by Xiang Bingren. Beijing: China Architecture & Building Press, 1990.)
- [11] 李茜,王东红,李士娟.西安市旅游目的地形象感知测量研究[J].西北大学学报(自然科学版),2013,43(5):815-820.(Li Qian, Wang Donghong, Li Shijuan. The Study on Tourism Destination Image Perception Measurement in Xi'an City[J]. Journal of Northwest University (Natural Science Edition), 2013, 43(5): 815-820.)
- [12] 陆易.基于内容分析法的杭州经济旅游感知要素评估的研究[J].现代营销(经营版),2018(11):64-65.(Lu Yi. Research on Evaluation of Hangzhou Economic Tourism Perception Elements Based on Content Analysis[J]. Modern Marketing (Business Edition), 2018(11): 64-65.)
- [13] 马天,屈册.大陆游客台湾旅游目的地形象感知与情感分析[J].商业研究,2014(9):169-174.(Ma Tian, Qu Ce. Perceptions and Emotional Analysis of Tourism Destination Image of Taiwan from Mainland China Tourists[J]. Commercial Research, 2014 (9): 169-174.)
- [14] 张文,顿雪霏.探讨大陆游客对台湾旅游目的地形象的感知——基于网上游记的内容分析[J].北京第二外国语学院学报,2010,32(11):75-83.(Zhang Wen, Dun Xuefei. Study on Image Perception of Mainland Tourists to Taiwan——Based on Content Analysis of Online Travels[J]. Journal of Beijing International Studies University, 2010, 32(11): 75-83.)
- [15] Gallarza M G, Saura I G, García H C. Destination Image: Towards a Conceptual Framework[J]. Annals of Tourism Research, 2002, 29(1): 56-78.
- [16] 陈建明.旅游目的地形象感知分析——以湘西凤凰古城为例[D].长沙:中南林业科技大学,2007.(Chen Jianming. The Analysis of the Tourist Destination Image Perception——A Case Study of Xiangxi Fenghuang[D]. Changsha: Central South University of Forestry and Technology, 2007.)
- [17] 韦晨.基于游客情感分析的旅游目的地形象感知测量——以南京夫子庙秦淮风光带为例[D].南京:南京师范大学,2017.(Wei Chen. Tourism Destination Image Perception Measurement Based on Tourist Sentiment Analysis——Taking the Confucius Temple in Nanjing for Example[D]. Nanjing: Nanjing Normal University, 2017.)
- [18] 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[J]. Progress in Geography, 2020, 39(4): 614-626.)
- [19] 颜端武,苏琼,张馨月.基于时序主题关联演化的科学领域前

沿探测研究[J]. 情报理论与实践, 2019, 42(7): 144-150. (Yan Duanwu, Su Qiong, Zhang Xinyue. Research on Frontier Detection in Scientific Field Based on Sequential Topic Association Evolution[J]. Information Studies: Theory & Application, 2019, 42(7): 144-150.)

[20] Kano N, Seraku N, Takahashi F, et al. Attractive Quality and Must-Be Quality[J]. The Journal of the Japanese Society for Quality Control, 1984, 14(2): 39-48.

[21] 张丹. 基于网络文本分析的旅游目的地形象感知研究——以香港为例[J]. 旅游纵览(下半月), 2019(1): 31-32. (Zhang Dan. Research on Image Perception of Tourist Destination Based on Web Text Analysis: A Case Study of Hongkong[J]. Tourism Overview, 2019(1): 31-32.)

作者贡献声明:

叶光辉:论文选题与论文审定;

徐彤:论文撰写,模型实现,实验数据处理和分析;
毕崇武:提出研究思路,采集实验数据,修改论文;
李心悦:修改论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: 56947634@qq.com。

[1] 徐彤. 中国香港旅游相关数据集.rar. 中国香港游记文本与用户信息、LDA主题生成代码等实验数据。

收稿日期: 2020-06-24

收修改稿日期: 2020-08-27

Analyzing Evolution of City Tourism Portraits with Multi-Dimensional Features and LDA Model

Ye Guanghui Xu Tong Bi Chongwu Li Xinyue

(School of Information Management, Central China Normal University, Wuhan 430079, China)

Abstract: [Objective] This paper explores the characteristics of a city portrait's evolution based on visitor's cognitive data with time attributes. [Methods] First, we chose the urban tourism industry as our research subject. Then, we developed a method using the LDA model and multi-dimensional theme description framework for the city. Finally, we revealed the changing trends of the city portraits from three perspectives: the theme development process, as well as the theme evolution trends in the first and second feature dimensions. [Results] We examined our new model with Hong Kong and found its urban tourism portraits showed no significant periodic changes. However, the tourists' perceptions on Hong Kong always had the primary and secondary dimensions. Sightseeing, transportation and entertainment were the main factors of tourists' perceptions on Hong Kong. Specifically, sightseeing was the most important one during the entire process, entertainment was mainly in the early and late stages, and transportation was more likely at the middle stage. We also found each topic node in the evolutionary path had a stable iconic. [Limitations] We need to evaluate our method with other cities. [Conclusions] Our research will benefit urban planning and policy implementation.

Keywords: City Profile Evolution Analysis Feature Dimension LDA Model