

基于运营商客流大数据的乡村旅游点类型划分研究

——以南京市江宁区为例

丁 鼎¹, 葛军莲¹, 龙 毅^{1,2,3}, 周贵鹏¹

(1. 南京师范大学地理科学学院, 江苏 南京 210023)

(2. 南京师范大学虚拟地理环境教育部重点实验室, 江苏 南京 210023)

(3. 江苏省地理信息资源开发与利用协同创新中心, 江苏 南京 210023)

[摘要] 基于运营商客流大数据, 以南京市江宁区 22 个星级乡村旅游点为研究对象, 从客流时间序列特征与客流空间结构特征视角出发, 利用层次聚类分析方法对其进行类型划分。结果显示: 按照客流时间序列特征可分为客流单峰型、客流双峰型和客流多峰型 3 类乡村旅游点, 按照客流空间结构特征分为强市场吸引型、中市场吸引型和弱市场吸引型 3 类乡村旅游点, 且这两种划分结果有高度的一致性, 并探讨了产生一致性的原因。

[关键词] 客流大数据, 乡村旅游点, 类型划分, 南京市江宁区

[中图分类号] K901.2 [文献标志码] A [文章编号] 1001-4616(2018)03-0116-06

The Type Classification of Rural Tourist Sites Based on the Passenger Flow Big Data of Operators

—A Case Study of Jiangning District of Nanjing

Ding Ding¹, Ge Junlian¹, Long Yi^{1,2,3}, Zhou Guipeng¹

(1. School of Geography Science, Nanjing Normal University, Nanjing 210023, China)

(2. Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, China)

(3. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China)

Abstract: This paper takes 22 star-rating rural tourist sites in Jiangning District of Nanjing as research area, and uses hierarchical cluster analysis to divide them into different types, based on the passenger flow big data of operators. The study has two perspectives, one based on the time series characteristics of passenger flow, while the other on the spatial structure characteristics. The result shows that, the rural tourist sites can be divided into unimodal, bimodal or multimodal passenger flow patterns from the former perspective, or strong, medium and weak market attraction patterns from the latter. The classifications of both have high resemblance, the reasons of the resemblance are also discussed in the study.

Key words: passenger flow big data, rural tourist sites, type classification, Jiangning District of Nanjing

近年来乡村旅游正在以前所未有的发展速度和全新的面貌展现在公众面前, 引起了国内各界的广泛关注, 乡村旅游成为扶贫和富民新渠道。

乡村旅游点分类 20 世纪 60 年代就出现在国外的研究学界, 在西方发达国家出现了集生产、游乐为一体的多功能观光农场, 这些农场通过在周边的城市投放广告吸引城市居民前往度假休闲, 围绕特色农业种植园、特色养殖场等兴建旅游设施, 形成了不同类型和独特卖点的乡村旅游点。国外学者多角度探讨了乡村旅游点的分类, Bulter 从旅游地生命周期的角度进行了类型划分^[1], Lucyna 从乡村农业活动的角度探讨了乡村旅游点的类型^[2]。

国内乡村旅游点分类研究出现在 20 世纪 90 年代, 对乡村旅游概念内涵不同的理解导致乡村旅游点

收稿日期: 2018-04-30.

基金项目: 国家自然科学基金(青年基金)(41301144)、江苏省普通高校研究生科研创新计划项目(CXZZ13_0404)、南京师范大学英才培养计划项目。

通讯联系人: 葛军莲, 博士研究生, 讲师, 研究方向: 旅游信息化. E-mail: gejunlian@njnu.edu.cn

分类的不一致性. 国内对乡村旅游类型的划分主要有以下 4 种维度:一是从旅游景观维度划分,可分为乡村生态环境类、乡村园林类、现代新农村类等^[3];二是从乡村旅游点功能的维度划分,分为务农型、品尝型、购物型、疗养型和观赏型等^[4];三是从乡村旅游者的维度来划分,按乡村旅游者旅游的目的一般划分为观光旅游型、休闲度假型、参与体验型、文化娱乐型、学习教育型、回归自然型^[5-6],按乡村旅游者年龄特征将乡村旅游点划分为青少年型、青年型、中老年型等^[7];四是从观光农业旅游开发的地域模式的维度划分,将其分为依托自然型和依托城市型^[8].

综上所述,国内外学者划分乡村旅游点类型主要基于乡村旅游的内涵、概念、理论模型,重点从农业旅游资源、乡村旅游点的功能、旅游者特征等方面进行了有益的探索,为本文研究提供了翔实的基础,但仍存在一些局限性:目前划分方法主要以定性研究方法为主,依据市场以及游客行为与偏好的研究不足. 随着我国乡村旅游发展迅猛,乡村旅游点数量繁多且形式多样,现行乡村旅游点分类方法已难以适应乡村旅游的发展.

1 研究设计

1.1 研究区域

江宁区位于南京市中心的南部,总面积 1 561 km²,东面与镇江市相邻,东南面与本市溧水区接壤,西南与西面是安徽省. 江宁区经济发达,2017 年全年实现地区生产总值 1 936 亿元.

江宁区旅游资源丰富,拥有 1 个国家级旅游度假区,15 个国家等级景区. 据江宁区旅游局公布,2017 年全区接待游客达到 3 306 万人次,旅游收入超过 350 亿元,全区乡村旅游接待游客超过 1 000 万人次,实现旅游收入超 50 亿元. 江宁区有 24 个星级乡村旅游点,是全省拥有星级乡村旅游点最多的区县,根据国家旅游局大数据显示,江宁区已迈入全国乡村旅游热点目的地前 3 位.

1.2 数据获取

本研究所用的客流数据由中国电信南京分公司提供,数据源接入主要包括位置更新、网优无线话单、2G/3G/4G 基站位置、深度报文检测(DPI)、4G 上网、兴趣点(POI)等数据. 根据手机定位数据对进入江宁区各乡村旅游点的手机用户进行游客判别,判别规则为:在乡村旅游点开园时间段内,抓取在划定区域内停留大于 60 min 小于 300 min 的访客,将 1 个月内累计出现 20 d 以上的访客作为黑名单(工作人员)予以剔除,将 1 个月内在夜间 00:00 至 06:00 时段中出现 20 次以上的访客作为黑名单(常住居民)予以剔除. 选择 2018 年元旦假期前一周(12 月 21 至 28 日)和元旦假期周(12 月 29 日至 1 月 4 日)共计 15 d 作为研究时段. 研究共计获取 1 008 910 条游客信息数据,获取的客流属性数据包括各个景点的客流总量、每日客流量和客源地比例等.

1.3 研究方法

本研究采用了层次聚类方法. 层次聚类方法首先明确类间距离的定义和距离的计算方法,按照距离计算的结果,将相近的数据归入一类,依次类推,直到数据完全归为一类为止^[9-10]. 根据客流大数据提取出的多项特征值,划分为客流时间序列特征和客流空间结构特征.

1.3.1 客流时间序列特征

全球定位系统(GPS)的广泛使用,使得通讯运营商产生了大量的观测数据,在旅游业中被尝试应用,这些游客位置数据有别于传统的静态空间数据,具有向时间维扩展属性,构成时间序列特征. 时间序列特征分析是对动态数据进行深度处理的重要方法,能够揭示一直处在变动中的旅游客流的内在规律. 本文根据每个乡村旅游点客流时间序列的偏度、峰度、均值和标准差对客流时间序列模型进行描述. 这些统计特征的时间序列聚类在一定程度上克服了噪声问题,提高了计算效率^[11-13].

1.3.2 客流空间结构特征

运营商客流大数据是实名制的身份数据,具有准确、实时与连续性特征,能够揭示客流的空間分布与行为特点. 本研究根据客流的身份数据与空间分布数据,利用客源地地理集中指数和市场平均距离对乡村旅游点的客流空间特征进行描述. 客源地地理集中指数常用的定量计算式是:

$$G = 100 \times \sqrt{\sum_{i=1}^n \left(\frac{x_i}{T} \right)^2}, \quad (1)$$

式中, G 为乡村旅游客源地地理集中指数, x_i 为第 i 个地级市来乡村旅游点的游客数量, T 为乡村旅游点接待游客总量, n 为客流来源的地级市总数. G 值越小,代表客流来源的地级市分布越分散,江宁区游客来源地级市的数目越多,反之亦然^[14].

客流来源地与乡村旅游点之间的总体远近关系,用市场平均距离来表示,其计算公式为:

$$AD_{od} = \sum_{i=1}^n \left(\frac{x_i}{T} \right) d_i,$$

(2)

式中, AD_{od} 是客流来源的地级市到江宁区乡村旅游点的公路距离, T 为乡村旅游点接待游客总量, x_i 为第 i 个地级市来乡村旅游点的游客数量, n 为乡村客流来源的地级市总数, d_i 为第 i 个客流来源的地级市到江宁区乡村旅游点的公路里程. AD_{od} 值越大,表明客流来源的地级市到江宁区乡村旅游点的距离越大,乡村旅游点的吸引力越强,反之亦然^[15].

2 乡村旅游点类型划分

2.1 客流时间序列特征分类

将每个乡村旅游点客流数据视为一段时间序列,每个时间序列对应 4 项特征值,分别提取 22 个乡村旅游点特征值进行层次聚类. 为了消除量纲的影响,对数据进行 Z 分数标准化,通过聚类系数和树形图可以判断出最合适的分类结果和种类数. 选择聚类数为 3 种的聚类结果,并对该分类结果进行单因素方差分析,得到表 1,所有特征值都呈现显著性,说明组间差异显著,分类效果较好. 分类结果图中,所有数据在绘制折线图前均经过 Z 分数标准化处理,依据数据标准化后各乡村旅游点在时间上呈现的变化特征,分别命名为:客流单峰型乡村旅游点(见图 1)、客流双峰型乡村旅游点(见图 2)和客流多峰型乡村旅游点(见图 3).

表 1 时间序列聚类 ANOVA 表
Table 1 ANOVA table of time series clustering

特征值		平方和	自由度	均值平方	F 值	显著性
平均数	组间	49 806 558.300	3	16 602 186.100	26.186	0.000
	组内	12 680 334.940	20	634 016.747		
	合计	62 486 893.240	23			
标准偏差	组间	2 414 515.030	3	804 838.343	40.451	0.000
	组内	397 937.178	20	19 896.859		
	合计	2 812 452.208	23			
偏度	组间	11.338	3	3.779	18.142	0.000
	组内	4.166	20	0.208		
	合计	15.504	23			
峰度	组间	33.880	3	11.293	9.028	0.001
	组内	25.019	20	1.251		
	合计	58.899	23			

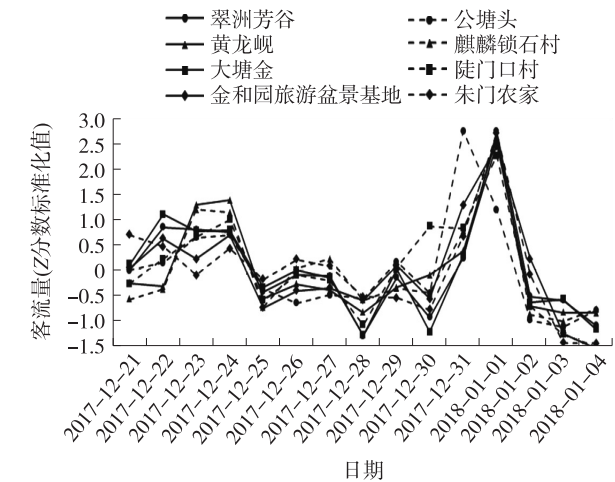


图 1 客流单峰型乡村旅游点客流量
Fig. 1 Passenger flow of unimodal rural tourist sites

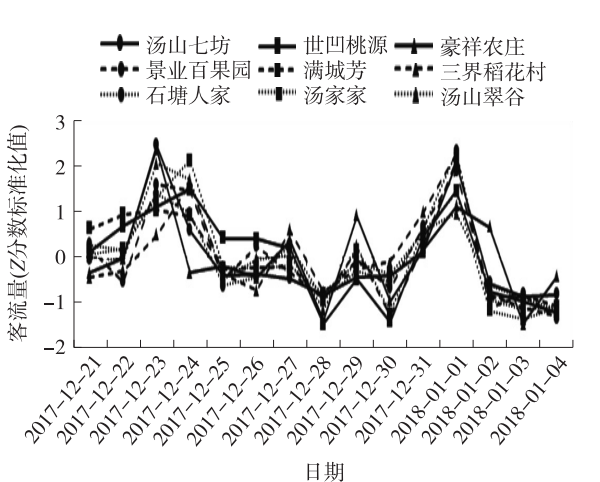


图 2 客流双峰型乡村旅游点客流量
Fig. 2 Passenger flow of bimodal rural tourist sites

2.1.1 客流单峰型乡村旅游点

客流单峰型乡村旅游点整体呈现节假日客流远高于双休日、双休日客流高于平日的特征,日客流量变化曲线整体呈单峰状. 此类乡村旅游点双休日客流迎来一个峰值,在元旦假期客流仍然能够迎来双休日两倍或以上的峰值,即节假日会迎来爆发性的客流量增长,整体呈节假日客流峰值远远高于双休日特征,故将此类乡村旅游点定义为客流单峰型乡村旅游点. 南京市江宁区的客流单峰型乡村旅游点共计 8 个,分别是:翠洲芳谷、公塘头、黄龙岘、麒麟锁石村、大塘金、陡门口村、金和园旅游盆景基地和朱门农家. 此类乡村旅游点客流量较大,日均客流为 2 443 人,在接待量排名前 5 的乡村旅游点中,有 4 个为单峰型乡村旅游点.

2.1.2 客流双峰型乡村旅游点

客流双峰型乡村旅游点整体呈现节假日客流量与双休日接近、平日客流量较少的特征,客流量变化曲线整体呈双峰状. 此类乡村旅游点的主要客流量来自于双休日和节假日,平日客流量所占比重较少. 南京市江宁区的客流双峰型乡村旅游点共计 9 个,分别是:汤山七坊、世凹桃源、豪祥农庄、景业百果园、满城芳、三界稻花村、石塘人家、汤家家和汤山翠谷. 客流量为中等水平,日均客流在 865 人.

2.1.3 客流多峰型乡村旅游点

客流多峰型乡村旅游点整体呈现客流出现峰值的次数较多,以及双休日、节假日和平日客流峰值区别不明显的特征,日客流量变化曲线整体呈多峰状. 此类乡村旅游点在平日的客流峰值与双休日客流峰值差值较小,元旦假期的客流量相较于双休日的客流峰值较为接近. 客流多峰型的乡村旅游点共计 5 个,分别是:湖山村、马场山、七仙大福村、台湾农民创业园和杨柳村. 这些乡村旅游点日均客流大约 442 人,客流量较少.

2.2 客流空间结构特征分类

根据式(1)和(2)的计算结果进行层次聚类,同样在分类前对数据进行 Z 分数标准化. 选择聚类数为 3 种的聚类结果,并对聚类结果同样进行单因素方差分析,得到表 2,特征值均为显著性,组间差异较大,分类效果较好. 根据特征值分类的统计结果,分出 3 类,结合分类特征值属性,分别命名为强市场吸引型乡村旅游点、中市场吸引型乡村旅游点和弱市场吸引型乡村旅游点(见表 3).

表 2 客流空间结构聚类 ANOVA 表

Table 2 ANOVA table of passenger flow spatial structure clustering

特征值		平方和	自由度	均值平方	F 值	显著性
地理集中指数	组间	1 063.415	3	354.472	18.272	0.000
	组内	387.996	20	19.400		
	合计	1 451.411	23			
市场平均距离	组间	25 752.138	3	8 584.046	20.696	0.000
	组内	8 295.363	20	414.768		
	合计	34 047.501	23			

表 3 客流空间结构分类汇总表

Table 3 Summary table of passenger flow spatial structure classification

乡村旅游点类型	市场集中度	市场平均距离/km	平均客流量/人	类型占比/%
强市场吸引型	72.389	122.785	41 455	31.819
中市场吸引型	74.228	114.588	10 040	45.454
弱市场吸引型	86.054	85.303	5 243	22.727

2.2.1 强市场吸引型乡村旅游点

强市场吸引型乡村旅游点呈现市场集中度低、市场平均距离远、客流量大的特征. 此类乡村旅游点客

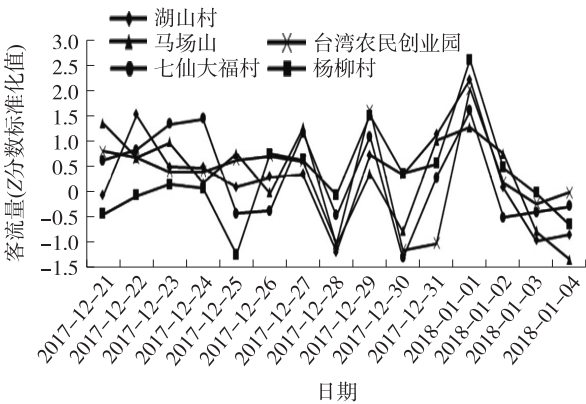


图 3 客流多峰型乡村旅游点客流量

Fig. 3 Passenger flow of multimodal rural tourist sites

源地数量多,客源地分布广泛,能够吸引较远地区游客,市场知名度较高. 强市场吸引型乡村旅游点共计 7 个,分别是石塘人家、麒麟锁石村、黄龙岷、公塘头、大塘金、陡门口村和翠洲芳谷.

2.2.2 中市场吸引型乡村旅游点

中市场吸引型乡村旅游点呈现市场集中度较低、市场平均距离较远、客流量较大的特征. 此类乡村旅游点客源地分布较广泛,能够吸引部分周边地级市游客,具有一定市场知名度. 中市场吸引型乡村旅游点共计 10 个,分别是景业百果园、汤山七坊、豪祥农庄、满城芳、金和园旅游盆景基地、七仙大福村、朱门农家、汤山翠谷、汤家家和世凹桃源.

2.2.3 弱市场吸引型乡村旅游点

弱市场吸引型乡村旅游点呈现市场集中度高、市场平均距离小、客流量较小的特征. 此类乡村旅游点游客以本地游客为主体,市场知名度不高. 弱市场吸引型乡村旅游点共计 5 个,分别是马场山、湖山村、三界稻花村、杨柳村和台湾农民创业园.

3 结论与讨论

3.1 结论

本研究以南京市江宁区的 22 个乡村旅游点为研究对象,基于运营商旅游客流大数据,利用层次聚类分析方法,从客流时间序列特征和客流空间结构特征出发,对南京市江宁区的乡村旅游点进行类型划分(见图 4),发现这两种分类方法的分类结果之间存在较强的一致性:

(1)客流单峰型与强市场吸引型乡村旅游点主体是重合的. 翠洲芳谷、公塘头、大塘金、陡门口村、黄龙岷和麒麟锁石村 6 个乡村旅游点是一致的. 这些乡村旅游点有着客流量大、市场吸引力强等共性特征.

(2)客流双峰型与中市场吸引型乡村旅游点大部分也是重合的. 汤山七坊、世凹桃源、豪祥农庄、景业百果园、满城芳、汤家家和汤山翠谷 7 个乡村旅游点是一致的. 这些乡村旅游点有着客流量较大、市场吸引力较强的共性特征.

(3)客流多峰型与弱市场吸引型乡村旅游点分类结果也有较高的一致性,分别是湖山村、马场山、台湾农民创业园和杨柳村. 这些乡村旅游点有着客流量小、市场吸引力弱、以本地游客为主等共性特征.



图 4 乡村旅游点类型划分结果

Fig. 4 Results of type classification of rural tourist sites

3.2 讨论

乡村旅游点客流时间序列的偏度、峰度、均值和方差的特征分类结果和客流空间结构的地理集中指数、市场平均距离的特征分类结果存在一致性的原因主要有两个方面。一是旅游客流时间序列特征是乡村旅游分类的有效依据。江宁区乡村旅游点数量较多,时间序列的结构特征规律难以分辨,通过对连续客流原始数据进行分析与转换,展现客流时间序列整体特征,表达客流内在变化机理,可以揭示不同乡村旅游点旅游客流时间序列上的变化规律,从而实现乡村旅游点的科学分类。二是乡村旅游点旅游客流空间结构特征的分类方法,验证了时间序列的特征分类方法具有一定的科学性。地理集中指数是旅游业中常用的景点类型划分方法,市场平均距离则是市场特征挖掘与吸引力描述常用的方法,这些方法都会受到统计数据准确性的限制,而本文采用准确性高、实时性强的通讯运营商数据来揭示旅游客流的时空特征,不仅弥补传统方法上的数据不足,同时也验证了时间序列特征分类方法的科学性。

本文尝试应用不同方法对乡村旅游点进行科学分类,揭示不同类型乡村旅游点市场吸引力与客流时空分布规律,据此针对不同类型的乡村旅游点制定差异化的产品、价格、渠道和促销策略,以期对乡村旅游研究进行有益尝试,为乡村旅游点的可持续发展及乡村振兴提供借鉴。

乡村旅游点的分类对乡村旅游营销、农民致富和乡村振兴都具有长远的理论和实践意义,值得进行深入广泛的研究。由于客观条件和作者水平限制,仅仅进行了有限尝试,还存在许多不完善的地方,今后还须对乡村旅游开展更为深入的定量研究与分析。

[参考文献]

- [1] BUTLER R W, MOHRI K, ISHII A. The concept of a tourist area cycle of evolution[J]. Rikkyo university bulletin of studies in tourism, 2002(4): 5-12.
- [2] PRZEZBORSKA L. Classification of agri-tourism/rural tourism SMEs in Poland(on the example of the Wielkopolska Region) [C]//Xlth International Congress of EAAE, Copenhagen, Denmark, August 23-27, 2005. Netherlands: European Association of Agricultural Economists, 2005: 1-15.
- [3] 卢云亭. 两类乡村旅游地的分类模式及发展趋势[J]. 旅游学刊, 2006, 21(4): 6-8.
- [4] 卢云亭, 刘军萍. 观光农业[M]. 北京: 北京出版社, 1995: 15-16.
- [5] 王云才, 许春霞, 郭焕成. 论中国乡村旅游发展的新趋势[J]. 干旱区地理, 2005(6): 862-868.
- [6] 王云才, 郭焕成, 徐辉林. 乡村旅游规划原理与方法[M]. 北京: 科学出版社, 2006: 142-143.
- [7] 夏林根. 乡村旅游概论[M]. 上海: 东方出版中心, 2007: 50-51.
- [8] 舒伯阳. 中国观光农业旅游的现状分析与前景展望[J]. 旅游学刊, 1997(5): 40-42.
- [9] 张文彤, 董伟. SPSS 统计分析高级教程[M]. 2 版. 北京: 高等教育出版社, 2013: 298-299.
- [10] 周涛, 陆惠玲. 数据挖掘中聚类算法研究进展[J]. 计算机工程与应用, 2012, 48(12): 100-111.
- [11] NANOPOULOS A, ALCOCK R, MANOLOPOULOS Y. Feature-based classification of time-series data[M]. New York: Nova Science Publishers, 2001: 49-61.
- [12] 宋辞, 裴韬. 基于特征的时间序列聚类方法研究进展[J]. 地理科学进展, 2012, 31(10): 1307-1317.
- [13] 尹芹, 孟斌, 张丽英. 基于客流特征的北京地铁站点类型识别[J]. 地理科学进展, 2016, 35(1): 126-134.
- [14] 保继刚, 楚义芳. 旅游地理学[M]. 修订版. 北京: 高等教育出版社, 1999: 54-55.
- [15] 丁正山. 南京国内旅游流时空演变研究[J]. 旅游学刊, 2004, 19(2): 37-40.

[责任编辑: 丁 蓉]