

# 基于Apriori算法的大学生选择旅游景点的研究分析

曹青青, 胡贵育, 霍奥莉, 刘思雨, 汤佳婧

河南科技大学商学院, 河南 洛阳

收稿日期: 2024年1月4日; 录用日期: 2024年1月26日; 发布日期: 2024年2月28日

## 摘要

大学生作为旅游行业消费群体的主力军, 其对旅游景点的选择极大地影响着旅游业的发展。本文基于Apriori算法, 对近几年大学生对旅游景点特征组合选择的关键因素进行了研究。文章首先使用因子分析法提取了五个主要因素, 其次将处理后的数据通过Apriori算法提取关联特征, 然后使用层次分析法寻找权重占比较高的因素。根据实验得出的重要因素, 最终找到影响大学生选择旅游景点的三个主要因素为民俗、物价和曝光度。

## 关键词

Apriori算法, 旅游景点, 因子分析法, 层次分析法

## Research and Analysis of College Students' Selection of Tourist Attractions Based on Apriori Algorithm

Qingqing Cao, Guiyu Hu, Aoli Huo, Siyu Liu, Jiajing Tang

School of Business, Henan University of Science and Technology, Luoyang Henan

Received: Jan. 4<sup>th</sup>, 2024; accepted: Jan. 26<sup>th</sup>, 2024; published: Feb. 28<sup>th</sup>, 2024

## Abstract

As the main force of consumer groups in the tourism industry, college students' choice of tourist attractions greatly affects the development of tourism. Based on Apriori algorithm, this paper re-

searches the key factors of college students' choice of feature combinations for tourist attractions in recent years. The article firstly extracted five main factors using factor analysis, secondly extracted the associated features from the processed data by Apriori algorithm, and then used hierarchical analysis to find the factors with higher weight share. According to the important factors derived from the experiment, the three main factors affecting college students' choice of tourist attractions were finally found to be folklore, price and exposure.

## Keywords

Apriori Algorithm, Tourist Attractions, Factor Analysis, Hierarchical Analysis

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

据统计报告显示, 2023 年旅游市场消费持续增长明显, 国内旅客数量呈现大幅度增长, 国内 128 万家 A 级旅游景区逐渐对游客开放, 占国内 A 级景区总数 80% 以上。据《2023 年上半年国内旅游数据情况》显示, 2023 年上半年, 旅游总人数高达 23.84 亿, 国内旅游总人次同比增幅分别为 63.9%, 其中第一季度, 国内旅游人数共 12.16 亿, 第二季度国内旅游人数共 11.68 亿。在五一旅游高峰期, 国内旅游人数合计 274 亿人, 同比增长 70.83%, 是 2019 年同期的 119.09% [1]。

大学生正处于培养个人爱好、选择释放压力、提升自我阅历与价值的关键时期, 在经过疫情三年封控后, 高校大学生的社交与娱乐需求呈现出显著增长[2]。据文献调查显示, 河南省 72.25% 的大学生表示需要进行外出旅游, 对旅游抱有一定的期待[3]。在旅游业消费群体中, 相比社会人员, 大学生具有基数大、空闲时间长等特点, 因此大学生消费群体已逐渐成为旅游业的重点关注对象, 其对旅游景点的选择极大地影响着旅游行业的发展。在疫情期间, 大学生旅游受到限制, 某些学生甚至没有探索过学校所在城市周边的旅游景区, 对外界有着强大好奇心的大学生极易受此影响, 当旅游的好奇心被抑制时, 其旅游需求得到不断的强化, 从而驱使大学生参与到旅游发展中[4]。

## 2. 方法介绍

### 2.1. 层次分析法

层次分析法是由来自美国的运筹学家萨蒂(20 世纪 70 年代初)提出的一种层次权重决策分析方法。该方法主要用于解决一个复杂的多因素系统决策问题, 其首先将问题及解决方案分为目标层、准则层和方案层, 其次通过定性的方法将问题分解为多指标的多个层次, 形成一个多层次分析结构模型, 最后通过定量的方法计算各方案的权重, 利用权重排序, 得出各方案的优劣次序[5]。

### 2.2. 因子分析法

因子分析法是由查尔斯 - 斯皮曼(1904 年)提出的一种在多变量中提取综合因子的统计方法, 在日常生活中, 变量之间是不可能做到毫无关系的, 当变量之间有一定的相关关系时, 它将具有相关关系的变量利用正交变换综合为少数且独立的几个因子。以表现原有变量与因子之间的相互关系, 讨论多个可直接获取数据的变量是受几个内在的独立因子影响, 根据研究需要从中可以取出几个较少的综合因子尽可

能地减少原始变量信息的丢失[6]。

### 2.3. Apriori 算法

Apriori 算法是由阿格拉沃尔等学者(1993 年)提出的挖掘出布尔关联规则频繁项集的算法。Apriori 算法利用频繁项集的先验知识,通过逐层搜索的迭代方法扫描数据集中的关系,最终形成大于最小支持度阈值的关联规则,其过程由连接和剪枝组成[7]。

## 3. 方案实施

### 3.1. 关键特征提取

通过在河南科技大学校内发放调查问卷,一定时间后我们收回了 257 份有效问卷,部分数据展示如表 1 所示,然后利用 SPSS 软件运用因子分析法,从影响大学生对旅游地点选择的因素:文化底蕴、曝光度、自然景观、口碑、交通出行、物价、服务与设施完善程度、政策支持、地域标识、安全保障、民风民俗和休闲娱乐 12 个特征值中提取出了 5 个关键特征,通过将问卷选项与价值尺度评价表相对应,消除各数据之间的量纲,并通过正交旋转最大方差,使得这 12 个特征值有高负载的变量数目达到最小。结果显示, KMO 值为 0.885,可知特征值之间的相关性较强,适合做因子分析。

Table 1. Feature selection value table

表 1. 特征选择价值表

城市	序号	文化底蕴	曝光度	自然景观	口碑	交通出行	物价	服务与设施完善程度	政策支持	地域标识度	安全保障	民风民俗	休闲娱乐
北京	1	4	5	4	4	5	4	4	3	2	4	4	5
洛阳	2	3	5	5	4	4	3	4	5	5	4	4	4
洛阳	2	4	5	4	5	5	4	4	5	5	5	5	5
洛阳	2	3	5	4	5	5	5	3	2	1	2	5	5
重庆	3	5	5	5	5	5	3	2	3	4	4	3	4
洛阳	2	3	4	5	5	5	4	4	4	5	5	5	5

Table 2. Component matrix after rotation

表 2. 旋转后的成份矩阵

指标	政策支持度	服务与设施完善度	安全保障	地域标识	曝光度	自然景观	口碑	交通出行	休闲娱乐	物价	民风民俗	文化底蕴
1	0.834	0.663	0.616	0.578	0.509							
2						0.780	0.681	0.567				
成份 3										0.868	0.681	
4								0.636	0.768			
5												0.889

由表 2 可知,根据相关系数矩阵计算累计方差贡献率和方差贡献率,第一个成分因子占有所有因子方差的 20.416%,前五个因子累计的方差贡献率为 70.684%,能够反映出原始问卷中的大部分信息,故

问卷信息失真较少，因子分析效果较为理想，具有实际研究意义，因此可以将提取出的 5 个关键特征作为本次研究的关键影响因素。其中，政策支持度、服务与设施完善度、安全保障、地域标识、曝光度位于第一个因子成分上，其主要表现为政策支持度、服务与设施完善度、安全保障，这些都是城市建设方面的内容，故取名为城市建设因子；自然景观、口碑、交通出行位于第二个因子成分上，主要为自然景观与口碑，因此取名为景观评价因子；物价、民风民俗位于第三个因子成分上，解释为物价与民风因子；交通出行、休闲娱乐位于第四个因子成分上，解释为出行与娱乐因子；文化底蕴位于第五个因子成分上，故解释为文化底蕴因子。综上所述，通过因子分析能够将这 12 个特征指标归类为以上 5 种关键因子。

### 3.2. 提取关键特征组合

围绕城市建设、景观评价、物价与民风、出行与娱乐、文化底蕴 5 个关键因子设计问卷，并在河南科技大学校内以及校外发放问卷，共收回了 804 份有效问卷，以此作为数据来源，对采集的数据进行预处理、筛选，将问卷选择结果以二分类数据进行转换，使用 SPSS 软件对数据进行 Apriori 算法，挖掘关联组合，部分数据如表 3 所示。

Table 3. Secondary classification data table

表 3. 二分类数据表

序号	景观评价	物价与民风	城市建设	出行与娱乐	文化底蕴
1	T	T	F	F	F
2	T	F	F	T	F
3	T	F	F	T	F
4	F	T	T	T	T
5	F	T	F	T	F
6	F	T	F	F	F
7	F	T	T	T	T
8	F	T	T	f	F

注：若该项被选择则标记为 T，反之标记为 F。

Table 4. Correlation combination extraction results

表 4. 关联组合提取结果

序号	前项	后项	支持度	置信度
1	文化底蕴，城市建设	物价与民风	20.373%	78.659%
2	文化底蕴，物价与民风	城市建设	20.497%	78.182%
3	城市建设	物价与民风	55.652%	76.562%

根据关联组合提取结果，设置最小支持度为 20%，最小置信度为 70%。见表 4，以第 1 行为列，前项为文化底蕴、城市建设，后项为物价与民风的支持度为 20.373%，置信度为 78.659%，说明在 804 份数据中，有 164 份数据表明该规则同时发生，在文化底蕴、城市建设为前项的出现条件下，物价与民风出现的概率为 78.659%。其中，第一列的支持度和置信度最高，说明文化底蕴、城市建设、物价与民风关键特征是大学生在选择旅游景点时较为关注的具有相互关联关系的关键特征组合。

### 3.3. 主要因素的确定

#### 3.3.1. 建立 AHP 层次模型

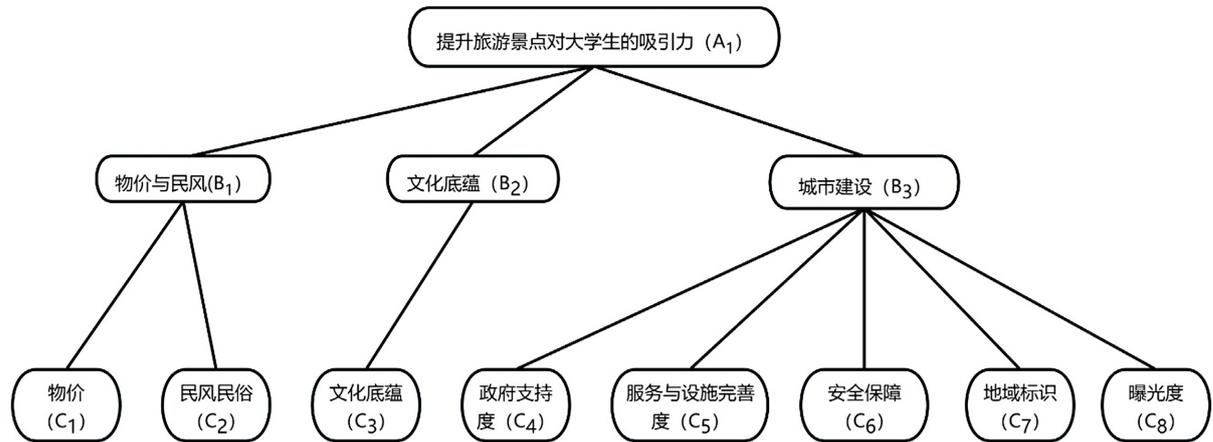


Figure 1. Hierarchical structure model

图 1. 层次结构模型

根据上文中的数据过程，将最初提出的影响大学生对旅游地点选择的 12 个因素进行关联组合，通过关键特征值的提取与因子分析结果，建立如图 1 的层次模型，下一步运用层次分析法得出旅游业应该着重从哪几个方面去提高自身，从而提升对大学生的吸引力。

#### 3.3.2. 构建判断矩阵及计算

根据整理的第二次问卷数据，将影响因素进行两两比较并进行赋值，从而构造判断矩阵，对判断矩阵的数据使用归一化处理并进行一致性检验。

Table 5. Weights of indicators

表 5. 各指标权重

目标层	一级指标	含义	$W_i$	二级指标	含义	$W_i$	综合权重	排序
A <sub>1</sub>	B <sub>1</sub>	物价与民风	0.570	C <sub>1</sub>	物价	0.333	0.190	2
				C <sub>2</sub>	民风民俗	0.667	0.380	1
	B <sub>2</sub>	文化底蕴	0.100	C <sub>3</sub>	文化底蕴	1.000	0.097	4
				C <sub>4</sub>	政府支持度	0.072	0.024	8
	B <sub>3</sub>	城市建设	0.330	C <sub>5</sub>	服务与设施完善度	0.118	0.039	7
				C <sub>6</sub>	安全保障	0.170	0.057	6
				C <sub>7</sub>	地域标识	0.242	0.081	5
				C <sub>8</sub>	曝光度	0.398	0.132	3

注: C.I. < 0.1、C.R. < 0.1 均通过一致性检验。

最终，得到各影响因素的指标权重，见表 5，如表所示，通过计算各指标的权重值，其中民风民俗、物价、曝光度的权重分别为 0.380、0.190、0.132，排名为前三，且权重占比均大于 0.1，则表格结果可以说明在影响大学生对旅游景点选择的 12 个主要因素中民风民俗、物价和曝光度占据重要方面，因此可以说明在大学生选择旅游景点时的主要影响因素有民风民俗、物价和曝光度，故文旅行业可以从以上三个

角度着重下手提出相应的改革政策，以提高旅游景点对大学生的吸引力。

## 4. 结论与建议

为找出影响大学生在选择旅游景点时的关键因素，本研究通过问卷调查得到研究数据，并通过关键特征组合的提取，成功构建层次结构模型，最终得出影响大学生选择旅游景点的三大因素分别为民风民俗、物价和曝光度。可知，大学生在旅游时会主要考虑景点的这三个方面是否达到他们的预期。

综上，通过对本项目的研究，能为一些正处于发展瓶颈的旅游相关行业带来一定的指导。旅游行业若想得到长远的发展，可以考虑从基数大的大学生群体入手，参考本项目研究得出的三大影响因素，从而对旅游景点的宣传和发展进行提高与完善。对此，旅游景点应正确处理好文化创新与民风民俗之间的关系，同时确保景区物价合理规范，加强景区运行的数字化营销，借文化兴起之势，在提高本地文化底蕴的同时，吸引更多高校学生走进景区。

### 4.1. 民俗方面的改进

互联网宣传能够从多方面提升旅游景点的市场竞争力，利用互联网优势提高旅游景点的曝光度。通过知名画家、博主等群体在社交媒体、旅游网站等平台打造独特的城市形象。以多方位、多方式的角度展示和吸引大学生眼球，明确城市定位，挖掘城市独特的文化内涵和个性。运用新媒体与传统媒体的同时，发动本地居民的力量，激发他们的认同感和自豪感，引导他们宣传城市代表性标签，在内容创作上进行创新，避免同质化宣传内容，如旅游景点推广策略、宣传用语等过于相似[8]。

### 4.2. 物价方面的改进

大学生外出旅游必然会考虑其经济情况以及当地物价，定价过高的景点周边及景区对于学生的吸引力不高。执行各种价格优惠政策是提高当地景区对大学生吸引力的有效措施之一，有利于直接提高景区周边的收益，促进景区发展[9]。同时，景区需要控制当地商品的品质，按照旅游商品市场发展制定合理的商品价格，景区管理者需要全面了解商品的进产销，明确商品的成本，制定适宜的价格战略[10]。

### 4.3. 曝光度方面的改进

针对曝光度的提升，景点首先要找到自身特殊性，比如特殊的地理位置、风俗习惯、景色和美食等等，然后挖掘自身文化底蕴，明确城市定位，丰富城市独特的文化内涵，打造一个独特的城市形象。其次要进行自我营销、合理利用互联网，要想通过网络提升曝光度，可以将传统内容年轻化、将严肃内容趣味化、将纪实内容戏剧化；媒体和文旅相互融合[11]。一个城市或景点是否能够“长红”，还要看能否持续曝光，首要任务就是发展经济。若想取得长期发展，我们必须首先审视自身。经济实力、人才储备、创新精神和丰富的文化底蕴，这些都是取得成功的基石[12]。

## 5. 结语

本研究通过问卷调查对大学生选择旅游景点的关键因素进行了分析，综合考虑了物价、文化、城市等多方面的影响因素，深入探讨了这些因素对大学生选择旅游景点的影响程度。综合分析的方法可以帮助我们全面地看待问题，不遗漏潜在的影响因素，从而有利于为行业提出具有针对性的建议。同时，通过因子分析和 Apriori 算法提取出这些因素中的主要因素，最后通过层次分析法抓取到关键因素，能够提出更有针对性的对策建议。这些研究使得本项目不仅为旅游行业提供了新的见解和证据，同时也为实践者提供了有价值的参考。

## 基金项目

河南省大学生创新创业训练计划创新重点项目《基于 Apriori 算法的疫情开放后大学生对旅游景点选择的研究分析》(项目编号: 202310464072)。

## 参考文献

- [1] 魏彪. 上半年国内旅游数据情况出炉旅游人次和收入实现大幅增长[N]. 中国旅游报, 2023-07-14(001).
- [2] 索冰洁, 宋贞迪, 石晓雪, 等. “深度旅游 + 社交”模式影响因素以及发展路径解析[J]. 中国市场, 2023(9): 1-4.
- [3] 梁留科, 钱宏胜, 陈德广. 河南高校大学生旅游市场初探[J]. 西北农林科技大学学报(社会科学版), 2007, 7(2): 77-80.
- [4] 程思, 罗希. 大学生旅游: 既爱“走马打卡”也爱“沉浸体验” [N]. 中国青年报, 2023-04-28(006).
- [5] 汪应洛, 主编. 系统工程[M]. 北京: 机械工业出版社, 2008: 122-131.
- [6] 张文彤, 主编, 董伟, 副主编. SPSS 统计分析高级教程[M]. 北京: 高等教育出版社, 2004: 245.
- [7] 孟海东, 宋宇辰, 著. 大数据挖掘技术与应用[M]. 北京: 冶金工业出版社, 2014: 64.
- [8] 高华. 互联网宣传对区域旅游市场竞争力的影响及对策研究[J]. 西部旅游, 2023(8): 94-96.
- [9] 孙萃. 旅游饭店执行淡季优惠的探讨[J]. 科技视界, 2012(13): 59-60.
- [10] 付英娟, 宋丹璞. 旅游景区商品高物价对旅游业的发展影响分析[J]. 中国市场, 2020(11): 34-35+42.
- [11] 李浩然. 如何利用媒体融合助力城市文旅宣传——以重庆广电为例[J]. 西部广播电视, 2020(12): 51-53.
- [12] 张蕊, 周添舒. 多媒体视阈下的“网红城市”——以重庆为例[J]. 今传媒(学术版), 2021, 29(3): 51-53.