

文章编号: 2095-2163(2021)04-0188-05

中图分类号: TP39

文献标志码: A

Python 熵值法对旅游目的地安全度评价

王晨光

(山西工程技术学院 管理工程系, 山西 阳泉 045000)

摘要: 中国旅游业近年来实现了跨越性发展,但与此同时,旅游安全事故也时有发生,不同程度地影响着旅游业的发展。为此,笔者利用 python 语言对熵值法进行了算法实现,并选取山西省作为案例,对山西省 11 地市旅游年鉴等数据进行了整理。通过文献汇总、专家询问构建出旅游目的地安全评价指标体系,再利用熵值法对指标层的权重系数进行了测算,形成一套完整的评价指标体系,来对山西省旅游目的地综合评价,做出等级划分。

关键词: python;熵值法;旅游目的地;旅游安全度

Python implements entropy method to evaluate the safety of tourist destinations

WANG Chenguang

(Department of Management Engineering, Shanxi Institute of Engineering and Technology, Yangquan Shanxi 045000, China)

[Abstract] China's tourism industry has achieved leapfrog development in recent years. At the same time, tourism safety accidents occur frequently, with many influencing factors, which affect the development of the tourism industry to varying degrees. For this reason, the author uses python language to implement the entropy method. And selected Shanxi Province as a case to organize the tourism yearbooks and other data of 11 cities in Shanxi Province. Through literature summary, experts asked to construct a tourism destination safety evaluation index system, and then use the entropy method to calculate the weight coefficient of the index layer. Based on the calculations, a complete evaluation index system was formed to comprehensively evaluate the tourist destinations in Shanxi Province and make a grade division.

[Key words] Python; Entropy method; Tourism destination; Tourism safety

0 引言

近年来,旅游业蓬勃发展,旅游目的地的服务水平也在不断提高。国家不断推出新政新规,以及新的路线。如,全域旅游、旅游+互联网、智慧旅游等等。作为旅游行业,如何利用大数据,利用计算机来更好的发展,无疑成为行业比较关注的热点问题。众所周知,旅游业受到了诸多因素的影响,且影响程度各有不同,如何将旅游服务质量数据化,从数据中找出发展方向,是本文要解决的问题。

本文查阅 2002-2020 年中国城市统计年鉴,挖掘到山西省 11 个地市中与旅游业相关的 35 个指标,初步假设旅游安全度评价体系,利用 python 语言,设计熵值法算法,对山西省旅游目的地安全度指标进行权重分析,并对各地安全度进行综合评价,引用中国旅游发展规划“十二五”的公共服务的 5 个体系作为变量层,来形成目的地旅游安全度的评价模型,进而创建旅游目的地安全的评价体系。

1 指标体系设计

旅游目的地安全评价模型是一个复杂而庞大的

系统。笔者通过联系现在全域旅游背景,在选取安全度评价指标的过程中,遵循层次性、客观性、可比性、动态性、合理性原则。根据旅游目的地安全的特点,结合旅游目的地的实际情况,通过 CNKI 数据库,对国内外学者文献的梳理,初步筛选出 45 个评价指标。结合中国旅游发展“十二五”规划中提到的旅游公共服务体系,归纳为 5 个变量层,即信息咨询、安全保障、便民惠民、行政服务和交通便捷。采用德尔菲法,对部分指标进行了调整,删除了对旅游目的地安全影响不大的指标,最终构建出旅游目的地安全评价体系。

在该体系中,旅游目的地安全的评价指标主要包括 5 个变量层,信息咨询包括租赁和商业服务业从业人员(万人)、信息传输、计算机、服务和软件业(万人)、互联网宽带接入用户数(万户)、移动电话年末用户数(万人)4 个评价指标^[1-5];安全保障包括保险深度(%)、保险密度(元/人)、保费收入(万元)、卫生、社会保障和社会福利业(万人)、医院、卫生院(个)、医院、卫生院床位数(个)、医生数(人)、人均地区生产总值地区(万元)、生产总值增长率(%)9 个评价指标^[6-10];便民惠民包括星级酒店

作者简介: 王晨光(1987-),男,硕士,实验师,主要研究方向:计算机应用、旅游管理。

收稿日期: 2020-12-12

(个)、金融机构网点数(个)、住宿与餐饮业(万人)、旅游从业人员(万人)、金融业(万人)、居民服务、修理和其他服务业(万人)、水利、环境和设施管理业(万人)、旅行社数(个)、建成区绿化覆盖率(%)、人均城市道路面积(平方米) 10 个评价指标^[11-15]; 行政服务包括公共财政支出(万元)、城市维护建设资金支出(万元)、污水处理集中处理率(%)、生活垃圾无害化处理率(%)、一般工业固体废物利用率(%) 5 个评价指标; 交通便捷包括交通运输、仓储和邮政业(万人)、铁路客运量(万人次)、公路客运量(万人次)、年末实有公共汽车营运车辆数(辆)、全年公共汽车客运总量(万人次)、年末实有出租车数(辆) 6 个评价指标; 共计 34 个评价指标。

旅游目的地评价指标中, 所有指标均采用客观统计数据, 利用 2000 年-2018 年的《中国城市统计年鉴》、《山西省统计年鉴》、《国民经济与社会发展统计公报》等官方数据。系统中有些数据缺失, 有些数据单位不统一, 对于这些数据, 采用了平均赋值法、趋势递推法等来赋值, 保证整个系统评价指标数据的完整性。

2 各指标权重计算

计算权重的方法主要有主观赋值法、客观赋值法。主观赋值法主要有德尔菲法、层次分析法等; 客观分析法主要有离差最大化法、熵值法、变异系数法等方法。本文采用熵值法, 计算在计算旅游目的地的客观评价指标。

设: m 个研究区域, n 个评价指标, x_{ij} 为第 i 研究区域的第 j 个指标的数值 ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$), 可构造初始指标判断矩阵 $\mathbf{X} = \{x_{ij}\}_{m \times n}$ 。若指标值 x_{ij} 的差距越明显, 该指标在综合评价函数中所起作用越大, 否则无作用。在信息论的概念中, 信息熵是对系统无序程度的度量, 公式为:

$$H(x) = - \sum_{i=1}^n p(x_i) \ln p(x_i).$$

式中, 等号左边为信息熵, 右边为信息, 二者的绝对值相等, 但是符号相反。当系统内某项指标值变异程度越大, 则该指标提供的信息量就越大, 相应的权重系数也越大, 信息熵也就越小, 反之亦然。由于年鉴中整理的评价指标的计量单位存在不统一的情况, 为保证其计算结果, 需先进行归一化处理。即先对其进行标准化处理, 把指标的绝对值转化为相对值。因此, 在熵值法中进行了以下改进, 对各指标的数值进行无量纲化和非负化处理。

(1) 指标无量纲化处理

为解决指标量纲或者测度量级的不同对计算结果造成影响, 根据功效函数法, 对于正向指标(正向指标数值越高越好)、负向指标(负向指标越低越好)的计算, 分别进行无量纲化处理, 可得到指标值 x'_{ij} 。

正向指标处理为:

$$x'_{ij} = \frac{x_{ij} - \min\{x_{ij}, \dots, x_{nj}\}}{\max\{x_{ij}, \dots, x_{nj}\} - \min\{x_{ij}, \dots, x_{nj}\}}.$$

负向指标处理为:

$$x'_{ij} = \frac{\max\{x_{ij}, \dots, x_{nj}\} - x_{ij}}{\max\{x_{ij}, \dots, x_{nj}\} - \min\{x_{ij}, \dots, x_{nj}\}}.$$

(2) 指标非负化处理

指标经过第一步无量纲化处理之后, 有些会出现 0 值。因对数的底数不能为 0, 只能是非负数, 则需对该值进行非负化处理。具体为:

$$x'_{ij} = (1 - \alpha) + \alpha \times x'_{ij}.$$

其中, $0 < \alpha < 1$ (本文设 $\alpha = 0.99$)。

(3) 指标同度量化处理

$$P_{ij} = \frac{y'_{ij}}{\sum_{j=1}^n y'_{ij}}.$$

(4) 计算第 j 项指标的熵值

$$H_j = -k \sum_{i=1}^m P_{ij} \ln P_{ij}.$$

其中, $K > 0$, $\ln P_{ij}$ 为自然对数, $H_j \geq 0$ 。如果 x_{ij} 对于给定的 j 全部相等, 则:

$$P_{ij} = \frac{y'_{ij}}{\sum_{j=1}^n y'_{ij}} = \frac{1}{m}.$$

此时, H_j 取极大值, 即

$$H_j = -k \sum_{i=1}^m \left(\frac{1}{m}\right) \ln\left(\frac{1}{m}\right).$$

若假设 $k = 1/\ln m$, 则有 $0 \leq H_j \leq 1H_j$ 。

(5) 计算第 j 项指标的差异性系数

对于第 j 项指标, 当指标值 x'_{ij} 的差异性越小, H_j 越大; 当 x'_{ij} 无差异, $H_j = H_{max} = 1H_j$, 此时对于区域间的比较, 指标 x'_{ij} 无作用; 当指标值 x'_{ij} 差异性越大, H_j 就越小, 该项指标对于区域间比较所起到的作用就越大。定义差异性系数 F_j 公式如下:

$$F_j = 1 - H_j H_j.$$

式中, 当 F_j 越大时, 指标重要性越强。

(6) 确定第 j 项指标的信息权重

$$\lambda_j = \frac{F_j}{\sum_{j=1}^n F_j}.$$

由此可得出各指标较合理的权重 ($\lambda_1, \lambda_2, \dots, \lambda_n$)。

将上述算法使用 python 语言进行编程,数据路径设为原始数据的存放地点,运行后得出结果。(受篇幅所限,在此略去程序代码)

3 计算结果

按照上述步骤进行运算,计算出 2002 年-2020

年 19 年各个评价指标的权重。权重系数是测量每个指标对于变量层的影响系数,由于每个指标在不同年份对变量层影响有所变化,为后续的计算方便,利用均值法,得出各指标的均值,作为此指标对于变量层的综合影响系数,以此作为该指标的权重系数,每个变量层的权重系数为对应的指标层权重系数求和。由此,可以推算出各指标层系数,为后续计算提供依据。旅游目的地安全度指标权重数据见表 1。

表 1 旅游目的地安全度指标权重

Tab. 1 Tourism destination safety index weight

系统层	变量层	权重	指标层	单位	权重	编码	类型
旅游目的地安全度	信息咨询	0.125	租赁和商业服务业从业人员	万人	0.034	xx ₁	+
			信息传输、计算机、服务和软件业	万人	0.036	xx ₂	+
			互联网宽带接入用户数	万户	0.032	xx ₃	+
			移动电话年末用户数	万人	0.023	xx ₃	+
			保险深度	%	0.012	aq ₁	+
			保险密度	元/人	0.032	aq ₂	+
	安全保障	0.19	保费收入	万元	0.025	aq ₃	+
			卫生、社会保障和社会福利业	万人	0.019	aq ₄	+
			医院、卫生院	个	0.021	aq ₅	+
			医院、卫生院床位数	张	0.023	aq ₆	+
			医生数	人	0.023	aq ₇	+
			人均地区生产总值	万元	0.022	aq ₈	+
			地区生产总值增长率	%	0.013	aq ₉	+
			星级酒店	个	0.012	bm ₁	+
			金融机构网点数	个	0.021	bm ₂	+
			住宿与餐饮业	万人	0.054	bm ₃	+
	便民惠民	0.328	旅游从业人员	万元	0.052	bm ₄	+
			金融业	万人	0.022	bm ₅	+
			居民服务、修理和其他服务业	万人	0.067	bm ₆	+
			水利、环境和设施管理业	万人	0.035	bm ₇	+
			旅行社数	个	0.029	bm ₈	+
			建成区绿化覆盖率	%	0.014	bm ₉	+
			人均城市道路面积	平方米	0.022	bm ₁₀	+
			公共财政支出	万元	0.032	xz ₁	+
			城市维护建设资金支出	万元	0.046	xz ₂	+
			污水处理集中处理率	%	0.018	xz ₃	+
	行政服务	0.131	生活垃圾无害化处理率	%	0.023	xz ₄	+
			一般工业固体废物利用率	%	0.012	xz ₅	+
			交通运输、仓储和邮政业	万人	0.048	jt ₁	+
			铁路客运量	万人次	0.036	jt ₂	+
			公路客运量	万人次	0.024	jt ₃	+
			年末实有公共汽车营运车辆数	辆	0.039	Jt ₄	+
	交通便捷	0.226	全年公共汽车客运总量	万人次	0.044	Jt ₅	+
			年末实有出租车数	辆	0.035	Jt ₆	+

4 综合评分及等级划分

按照上述步骤进行运算,得出了各指标层、以及各变量层的权重系数,根据上述权重系数,将山西省 11 地市 19 年的数据,导入评价体系模型,结合对应的权重系数,算出每个地市每一年的旅游安全度综合评分,最为后续等级划分的依据。各城市旅游安

全度综合评分情况见表 2。

对旅游目的地安全等级的划分,是对旅游目的地安全的一种刻画,也是旅游目的地安全度的衡量标准。目前,中国官方还没有形成规范、统一的旅游目的地安全等级划分标准。美国国土安全部在“9·11”事件后,对恐怖袭击进行了等级划分,将这些突发事件按严重性从重到轻分为绿、蓝、黄、橙、红 5

个等级。中国社会治安事件按事件的严重性分为特大警情、重大警情、较大警情、一般警情、没有警情5个等级。《国家突发公共事件总体应急预案

(2006)》按事件的危害程度、紧急程度等,分为特别严重、严重、较重和一般4个等级。

表2 各城市旅游安全度综合评分

Tab. 2 Comprehensive score of tourism safety of each city

年度	太原	大同	阳泉	长治	晋城	朔州	晋中	运城	忻州	临汾	吕梁
2002年	0.743	0.34	0.194	0.186	0.18	0.073	0.188	0.234	0.155	0.213	0.118
2003年	0.788	0.288	0.11	0.194	0.18	0.073	0.197	0.219	0.148	0.231	0.122
2004年	0.778	0.288	0.113	0.186	0.18	0.082	0.188	0.219	0.155	0.213	0.131
2005年	0.793	0.297	0.127	0.195	0.186	0.088	0.222	0.228	0.176	0.289	0.162
2006年	0.808	0.304	0.144	0.195	0.187	0.09	0.222	0.23	0.149	0.309	0.122
2007年	0.829	0.461	0.167	0.208	0.188	0.09	0.223	0.234	0.155	0.295	0.134
2008年	0.863	0.321	0.226	0.209	0.244	0.073	0.237	0.249	0.155	0.325	0.171
2009年	0.864	0.34	0.221	0.254	0.264	0.101	0.245	0.249	0.16	0.32	0.198
2010年	0.869	0.34	0.218	0.25	0.268	0.14	0.253	0.253	0.161	0.293	0.205
2011年	0.876	0.422	0.208	0.237	0.195	0.115	0.254	0.243	0.162	0.279	0.19
2012年	0.878	0.347	0.177	0.246	0.205	0.141	0.272	0.249	0.164	0.259	0.168
2013年	0.864	0.347	0.182	0.21	0.206	0.128	0.253	0.249	0.184	0.25	0.171
2014年	0.897	0.39	0.184	0.225	0.218	0.147	0.223	0.249	0.145	0.235	0.18
2015年	0.876	0.297	0.194	0.237	0.218	0.13	0.265	0.253	0.161	0.245	0.19
2016年	0.913	0.273	0.208	0.246	0.242	0.128	0.272	0.257	0.184	0.293	0.168
2017年	0.878	0.288	0.218	0.332	0.244	0.13	0.276	0.265	0.188	0.295	0.205
2018年	0.929	0.428	0.22	0.209	0.264	0.14	0.276	0.269	0.211	0.309	0.267
2019年	0.933	0.441	0.221	0.257	0.268	0.141	0.281	0.277	0.215	0.32	0.285
2020年	0.939	0.461	0.226	0.332	0.285	0.147	0.283	0.305	0.218	0.325	0.285

国内学者们对社会安全及社会相关系统的研究也很多,根据评价结果及研究内容的特点出现了许多安全等级的划分标准。本文经过查阅相关文献,

充分借鉴前人的研究成果,将旅游目的地的安全度划分为5个等级:(0-0.25]、(0.2-0.5]、(0.5-0.75]、(0.75-1]、(1,+∞]。详情见表3。

表3 旅游安全度等级划分

Tab.3 Classification of tourism safety degree

等级	指数范围	安全状态	安全度指数意义	预警信号灯
☆	(0-0.25]	恶劣	旅游目的地的安全系统非常糟糕,旅游安全系数很低。	红色
☆☆	(0.2-0.5]	较差	旅游目的地的安全系统的安全度需增强,减少旅游安全风险。	橙色
☆☆☆	(0.5-0.75]	一般	旅游目的地的安全系统一般,但是需不断加强,风险系数一般。	黄色
☆☆☆☆	(0.75-1]	较安全	旅游目的地的安全系统较好,风险系数较低。	蓝色
☆☆☆☆☆	(1,+∞]	安全	旅游目的地的安全系统非常理想,风险极低。	绿色

5 结束语

通过上述计算,可以得出太原市安全度最高,处于较安全状态;大同市、临汾市、运城市处于一般状态;长治市、晋城市、晋中市等7个地市处于恶劣状态,其中吕梁市安全度最低。在当前中国全域旅游战略的背景下,旅游安全的研究是非常必要的。通

过研究发现,中国在旅游安全的理论和实践研究上还存在很多可提升之处。例如对于旅游安全度指标体系不够健全的问题、旅游年鉴中数据统计口径不统一的问题、如何让政府与市场在旅游目的地安全中发挥作用、如何建立科学的旅游安全体系标准等等,都成为今后需要深度研究的问题。

(下转封三)