

大数据背景下云南省户外旅游关注度研究

陈丹¹, 胡俊^{1,2}

¹云南大学数学与统计学院, 云南 昆明

²云南农业大学理学院, 云南 昆明

Email: danchen@ynu.edu.cn

收稿日期: 2020年10月29日; 录用日期: 2020年11月13日; 发布日期: 2020年11月20日

摘要

作为大数据应用前景最广阔的行业之一, 旅游大数据从庞大的旅游数据中对有意义的信息进行专业化处理, 为整个旅游行业发展及旅游决策奠定坚实的基础, 实现旅游产业的增值和盈利, 体现了现代旅游信息的综合。本研究通过查阅相关资料, 设计调查问卷的方式进行数据收集, 对云南省户外旅游大数据的关注度及信息化平台应用进行研究。结果表明云南省户外旅游市场的经营模式主要是“线下俱乐部 + 线上俱乐部平台”和爱好者自发组织的方式。云南省户外旅游爱好者对大数据的关注度较低, 户外旅游信息资源分散、旅游信息化程度不高、各信息模块间不能互相连通, 数据无法交换、效率低下。最后, 针对云南省信息化平台的应用情况, 作者提出了在大数据背景下云南省户外旅游模式的重构策略, 以促进云南省户外旅游产业的全面信息化和转型升级。

关键词

户外旅游, 大数据, 因子分析, 云南省

Research on Yunnan Outdoor Tourism Attention under the Background of Big Data

Dan Chen¹, Jun Hu^{1,2}

¹School of Mathematics and Statistics, Yunnan University, Kunming Yunnan

²College of Science, Yunnan Agricultural University, Kunming Yunnan

Email: danchen@ynu.edu.cn

Received: Oct. 29th, 2020; accepted: Nov. 13th, 2020; published: Nov. 20th, 2020

Abstract

As one of the widest application prospects of big data, tourism big data protect the professional information from huge tourism data. Tourism big data support a solid foundation for the devel-

文章引用: 陈丹, 胡俊. 大数据背景下云南省户外旅游关注度研究[J]. 可持续发展, 2020, 10(5): 770-781.

DOI: 10.12677/sd.2020.105094

opment of the whole tourism industry and tourism decision-making, realize the value-added and profit-making of tourism industry and reflect the integration of modern tourism information. In this paper, we study the attention of Yunnan outdoor tourism big data and the application of information platform by consulting relevant information and designing a questionnaire to collect data. The results show that the business model of Yunnan outdoor tourism market is mainly "offline Club + online club platform" and the spontaneous organization of the enthusiasts. Outdoor tourism enthusiasts pay less attention to big data, the information resources of outdoor tourism are not clustered together, the degree of tourism informatization is lower, the information modules cannot interconnection each other, tourism data cannot exchange and the efficiency is low. Finally, according to the application of the information platform in Yunnan Province, the author puts forward the reconstruction strategy of Yunnan outdoor tourism mode under the background of big data in order to promote the comprehensive informatization and transformation and upgrading of Yunnan outdoor tourism industry.

Keywords

Outdoor Tourism, Big Data, Factor Analysis, Yunnan Province

Copyright © 2020 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

当前,大数据已经不同程度地渗透到了各行各业,并发挥着巨大的应用价值,越来越多的行业把收集、分析和利用“大数据”视为未来行业发展的重要途径。旅游大数据体现了现代旅游信息的综合,对旅游数据进行整合,建立科学的、合理的数学模型,研究数据的计算方法并找出有价值的信息,是扎实有序地推进旅游大数据必要的理论支撑和技术保障,同时也为整个旅游行业发展及旅游决策奠定坚实的基础。我国各地旅游业都在蓬勃发展,但大部分地区仍处于低消费、大众化、低水平的初期阶段[1] [2]。随着生活质量的提升,越来越多的游客不再满足于当下的旅游项目,与传统的跟团旅游不同,具有强烈个性化特征的户外旅游是一种正在发展的、新兴的旅游形式,是一种较高水平、更体现游客需求的旅游[3] [4] [5]。

目前,旅游大数据的应用主要集中在旅游市场分析领域。开展旅游目的地大数据分析的企业通过对手机位置数据、搜索数据、交通数据、交易数据、社会数据等进行系统集成、数据挖掘、降维和统计分析[6] [7] [8] [9] [10],建立了旅游目的地综合客户市场分析。携程,去哪儿,飞猪等互联网平台利用平台运营过程中产生的交易数据,评价数据和日志数据分析用户全方位画像和消费行为,在产品信息浏览和产品知名度等多方面形成精准统计分析,赋予游客多维属性标签[11] [12]。平台企业根据旅游产品类型和旅游标签组合的匹配程度推荐产品和服务。

英国航空将散布在各个系统的所有与旅客有关的碎片数据和历史数据整合起来,发现很多以前被忽视或无法看到的有价值信息,并将数据分析推向最前线,大大提升了企业内部的运营效率[13];著名在线旅游网站 Travelocity 将大数据分析用于定价、库存和广告,采用拟归模型,最佳交易分析和推荐引擎向目标客户推送最恰当的产品,并提供和支持实时决策[14];旅游广告公司 Sojern 收集并聚合来自各航空公司、酒店、汽车租赁商和信用卡公司的数据,通过机器学习和高级分析丰富用户个人资料,其数据分析结果已被航空公司、连锁酒店和汽车租赁商采用,适时调整价格和服务名目,制定合理的交叉销售策略并调整库存[13] [14] [15] [16] [17]。

在大数据背景下,对户外旅游数据进行分析处理研究,对不同类型的非结构化数据采取有针对性、灵活的统计建模方式[18][19][20],加快实现线上平台服务的户外旅游模式,使得户外旅游的产品更加标准化、结构化、多维度化,线路产品难度的分级、各种衡量维度也将更加精准。本文选择以云南省户外旅游大数据作为切入点,通过对云南省户外旅游俱乐部及其会员的问卷调查,以户外运动爱好者和户外行业从业人员对旅游大数据的关注度入手,对云南省户外旅游信息平台的使用进行调查研究,探讨云南省户外旅游大数据的现状和信息平台应用需求,旨在提供准确实用的户外旅游信息,从规划、开发和安全保障等方面提出户外旅游的发展策略,实现户外旅游游客、户外旅游区、相关旅游企业和旅游管理者三者之间的信息对称。

2. 研究设计

2.1. 问卷设计

研究采用问卷调查的形式进行调研,问卷包括2个部分:1)云南省户外旅游爱好者或者从业人员的基本情况;2)云南省户外旅游大数据的关注及应用情况。第一部分的问题包括被调查者的性别、年龄、学历、职业、月收入、户外旅游年消费、参加户外旅游的方式、获得户外旅游有关信息的途径、对户外旅游信息感兴趣的模块、对户外旅游大数据是否有了解并关注。变量类型主要是分类变量。第二部分的问题包括对户外旅游大数据的关注度,在此基础上,针对“了解并关注户外旅游大数据”的人群进行信息化应用平台的重要程度进行分析。

2.2. 样本采集

本研究分为两个阶段进行数据采集,考虑到户外旅游受众较小、针对性较强的特点,研究者于2019年1月对特定的户外旅游爱好者采用问卷星的方式,在手机上填答,问卷星给与数据反馈的方式获得78份有效问卷。对所获数据进行可信度、有效度分析,删除四个容易混淆、与主题关系不大的选项,并对调查问卷的提问方式和表述进行了调整。2019年4月,研究者到昆明市32家户外旅游俱乐部,通过面对面、随机偶遇的方式对俱乐部管理人员及其部分会员,采用现场填答现场回收的方式,共发放220份问卷,加上预调研的78份问卷,共发放298份问卷,其中有效问卷为278份,总体有效率为93.3%。

2.3. 样本描述性统计分析

1) 在被调查的参与人群中,参与人数最多的主要集中在31~55岁之间,占调查总人数的75.6%(表1)。可以看出参与人群主要以中年为主,这部分人事业、家庭稳定,在参加多年的工作以后,职位、收入、家庭、假期等各方面稳定,同时在这个阶段他们开始注重锻炼。

2) 户外旅游参与者通常有较高的教育背景。如图,被调查人群中,大学本科学历所占比例最高为50.0%(表1),大专学历次之为28.2%。户外运动爱好者通常受过良好的教育,有一定教育背景的这类人群对于新事物的接受能力较强,学习新知识、新技能的能力较强。

3) 户外旅游爱好者的月收入以5000~15,000元所占比例最高,达到71.5%(表1),对比昆明市的消费水平和工资水平,这部分人处于昆明总收入人群水平的中等偏上。因为户外活动的特性(即装备和到交通较远的地方开展),收入具有一定的稳定性,更具备参加户外活动的经济基础。

4) 参与户外运动人群的职业类型多样化(图1),其中专业技术人员(教师、医生、艺术、体育工作者)所占比例最多为34.6%,自由职业者占比第二为24.4%,企事业单位一般工作人员占17.9%,学生占比为7.7%,公务员占6.4%,党政机关企事业单位管理人员占比为5.1%,离退休人士占比为3.8%。专业技术人员、自由职业者和企事业单位人员占了户外旅游爱好者群体的76.9%,说明户外运动参与人群大多属于中产阶级职业人群,他们有稳定的工作环境或闲暇时间,有充足的经济基础和精力去参加户外活动。

Table 1. Basic situation of the participants in the questionnaire
表 1. 调查问卷参与者基本情况

年龄	人数	百分比(%)	学历	人数	百分比	月收入	人数	百分比(%)
19~30 岁	43	15.4	高中及以下	18	6.4	5 千元以下	53	19.1
31~45 岁	128	46.2	大专学历	78	28.2	5 千元~1 万元	142	51.1
46~55 岁	82	29.4	大学本科	139	50	1 万元~1.5 万元	57	20.4
56 岁以上	25	9.0	研究生以上	43	15.4	1.5 万元以上	26	9.4

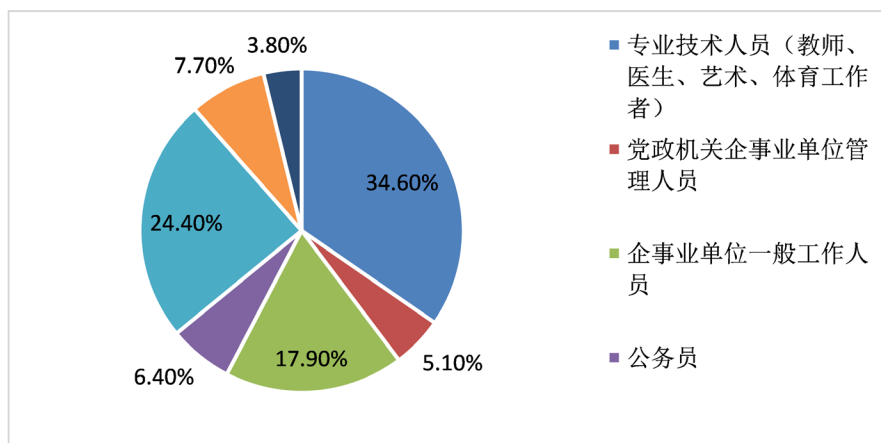


Figure 1. Occupational distribution of outdoor travel enthusiasts
图 1. 户外旅游爱好者职业分布情况

5) 户外旅游爱好者获取旅游相关资讯的途径主要是通过他人告知或邀请、报名参加旅行社、俱乐部或其它社团的方式, 占比约为 72.0% (表 2); 对于参加户外运动的方式, 则以个人完全资助或参加俱乐部为主, 占比为 72.8%。说明昆明户外运动仍处在起步阶段, 大部分参与者信息来源单一, 发达的网络对户外资讯的传播起到了很大的促进作用, 但影响力不够, 需要更多的俱乐部或旅行社通过定期发布信息吸引更多的人群参与到户外活动中。

Table 2. The situation of outdoor travel enthusiasts obtaining information and participating in outdoor sports
表 2. 户外旅游爱好者获取信息和参与户外运动的情况

获取途径	人数	百分比(%)	参加方式	人数	百分比(%)
他人告知或邀请	128	38.7	个人自助	130	41.5
报纸、书刊、杂志等纸质读物	11	3.3	网络召集	35	11.2
电视广播、广告宣传	32	9.7	报名参加俱乐部	98	31.3
旅游社、俱乐部或其他社团活动	106	32.0	旅行社组团	30	9.6
小程序、电子期刊、社交媒体	45	13.6	学校单位组织	12	3.8
其他途径	9	2.7	其他方式	8	2.6

6) 图 2 是云南省户外旅游的项目的参与情况, 最受欢迎的旅游方式是进行自助游、自驾游。在进行户外旅游的时候, 自驾出行比较方便自由, 不受到外界因素的影响, 在进行游览活动时只需要关注自身需求即可, 所以受到人们的欢迎。在进行户外旅游时, 更倾向于选择平时常见的徒步、野营、登山等户

外运动项目, 其次是参与探险、水上运动、野外生存项目, 选择飞行、攀岩、滑雪、轮滑、小轮车、极限运动等项目的较少, 倾向性较低。

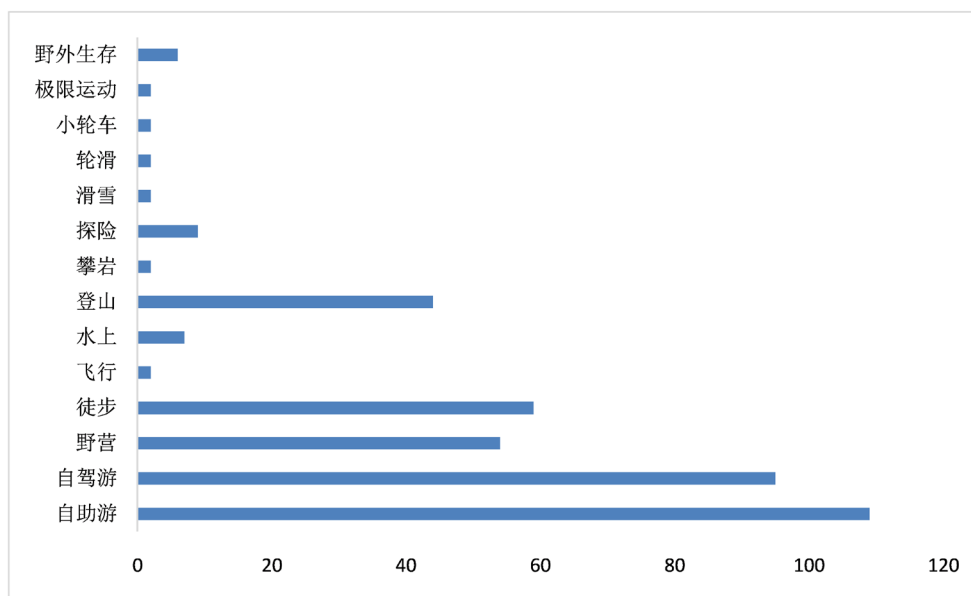


Figure 2. Outdoor travel enthusiasts participating in sports

图 2. 户外旅游爱好者参加运动项目的情况

3. 大数据背景下云南省户外旅游的研究

3.1. 对户外旅游大数据的关注度研究

通过“对户外旅游大数据是否了解并关注”的问卷调查发现, 仅仅听说过户外旅游大数据但尚未关注的, 占比最大约为 47.4%; 不了解也没有关注的约为 29.5%, 有一定了解并且较为关注的约为 23.1%。这与两个原因有关: ① 户外运动是一个较为宽泛的概念, 虽然户外运动的发展呈逐年上涨的趋势, 参与人数也越来越多。但是参与者未能完全了解户外旅游的真正含义, 或者对大数据户外旅游这一概念关注度还不是太高; ② 大数据体现了现代旅游信息的综合, 是旅游、工业、金融、交通等众多产业与互联网、信息技术深度融合, 其本质是从庞大的旅游数据中对有意义的信息进行专业化处理, 实现旅游产业的增值和盈利。人们对大数据的了解也仅限于听说过, 关注度并不高。

基于户外旅游和大数据两个概念, 我们选择从学历(高中及以下、大专、大学本科及研究生以上)和职业(专业技术人员(教师、医生、艺术、体育工作者)、党政机关企事业单位管理人员、企事业单位一般工作人员、公务员、自由职业者(个体)、学生、离退休人士)两个角度, 采用列联表统计分析方法对户外旅游大数据的关注度进行研究。

由表 3 可以看出, 各个学历层次中, 对户外旅游大数据有一定了解并关注的被调查者中, 学历是高中及以下的为 0.7%, 学历为大专的为 4.3%, 学历为大学本科的为 12.2%, 研究生以上为 5.8%; 对户外旅游大数据仅听说过但尚未关注的被调查者中, 学历是高中及以下的为 2.6%, 学历为大专的为 11.5%, 学历为大学本科的为 24.4%, 研究生以上为 9.0%; 没听说过户外旅游大数据的被调查者中, 学历是高中及以下的为 6.1%, 学历为大专的为 9.0%, 学历为大学本科的为 12.9%, 研究生以上为 1.4%。结果表面: 对户外旅游大数据有了解和关注的人群大部分为受教育程度较高的人群, 但是关注度也并不是很高, 说明需要进一步推广户外旅游大数据的相关概念和应用。

Table 3. The attention degree of outdoor tourism big data * cross table of academic qualifications
表 3. 户外旅游大数据的关注度 * 学历交叉表

		学历				总计	
		高中及以下	大专	大学本科	研究生以上		
是否对户外旅游大数据有了了解	没听说过	计数	17	25	36	4	82
		占是否对户外旅游大数据有了解的百分比	20.7%	30.5%	43.9%	4.9%	100.0%
		占学历的百分比	65.4%	36.2%	25.6%	8.8%	29.5%
		占总计的百分比	6.1%	9.0%	12.9%	1.4%	29.5%
	仅听说过, 尚未关注	计数	7	32	68	25	132
		占是否对户外旅游大数据有了解的百分比	5.4%	24.3%	51.4%	18.9%	100.0%
		占学历的百分比	26.9%	46.4%	49.3%	55.6%	47.4%
		占总计的百分比	2.6%	11.5%	24.4%	9.0%	47.4%
	有一定了解并关注	计数	2	12	34	16	64
		占是否对户外旅游大数据有了解的百分比	3.1%	18.8%	53.1%	25.0%	100.0%
		占学历的百分比	7.8%	17.4%	24.6%	35.6%	23.1%
		占总计的百分比	0.7%	4.3%	12.2%	5.8%	23.1%
总计	计数	26	69	138	45	278	
	占是否对户外旅游大数据有了解的百分比	7.2%	23.4%	49.6%	19.8%	100.0%	
	占学历的百分比	100.0%	100.0%	100.0%	100.0%	100.0%	
	占总计的百分比	9.4%	24.8%	49.6%	16.2%	100.0%	

为了检验列联表各变量之间的关联度是否显著相关, 我们用皮尔逊卡方检验对表 3 进行显著性水平为 0.05 的假设检验。由表 4 可知, 皮尔逊卡方检验的值是 0.047, 小于 0.05 的显著性水平, 说明不同学历人群的对户外旅游大数据的关注度是显著相关的。

Table 4. Chi-square test of the cross-contingency table of educational background of outdoor tourism big data
表 4. 户外旅游大数据学历交叉列联表卡方检验

	值	自由度	渐进显著性(双侧)
皮尔逊卡方	5.565 ^a	6	0.045
似然比(L)	6.640	6	0.036
线性关联	0.020	1	0.876
有效个案数	278		

a. 5 个单元格(41.7%)的期望计数小于 5。最小期望计数为 1.15。

由表 5 可知, 不同职业的被调查者中, 对户外旅游大数据有一定了解并关注的被调查者中, 专业技术人员占 6.5%, 党政机关企事业单位管理人员为 4.7%, 企事业单位一般工作人员为 4.0%, 公务员为 3.2%, 自由职业者为 2.2%, 学生为 2.5%, 离退休人士为 0%; 对户外旅游大数据仅听说过但尚未关注的被调查者中, 专业技术人员占 13.7%, 党政机关企事业单位管理人员为 9.7%, 企事业单位一般工作人员为 7.6%, 公务员为 6.8%, 自由职业者为 5.8%, 学生为 3.2%, 离退休人士为 0.7%; 没听说过户外旅游大数据的被

调查者中, 专业技术人员占 7.6%, 党政机关企事业单位管理人员为 3.6%, 企事业单位一般工作人员为 3.2%, 公务员为 4.0%, 自由职业者为 9.4%, 学生为 1.1%, 离退休人士为 0.7%。

Table 5. Attention of outdoor tourism big data * Occupation cross table
表 5. 户外旅游大数据的关注度 * 职业交叉表

		职业							总计	
		专业技术人员(教师、医生、艺术、体育工作者)	党政机关企事业单位管理人员	企事业单位一般工作人员	公务员	自由职业者(个体)	学生	离退休人士		
是否对户外旅游大数据有了解	没听说过	计数	21	10	9	11	26	3	2	82
		占是否对户外旅游大数据有了解的百分比	25.6%	12.2%	11.0%	13.4%	31.7%	3.7%	2.4%	100.0%
		占职业的百分比	27.3%	20.0%	22.0%	28.2%	54.2%	15.8%	50.0%	29.5%
		占总计的百分比	7.6%	3.6%	3.2%	4.0%	9.4%	1.1%	0.7%	29.5%
	仅听说过, 尚未关注	计数	38	27	21	19	16	9	2	132
		占是否对户外旅游大数据有了解的百分比	28.8%	20.5%	15.9%	14.4%	12.1%	6.8%	1.5%	100.0%
		占职业的百分比	49.4%	54.0%	51.2%	48.7%	33.3%	47.4%	50.0%	47.4%
		占总计的百分比	13.7%	9.7%	7.6%	6.8%	5.8%	3.2%	0.7%	47.4%
	有一定了解并关注	计数	18	13	11	9	6	7	0	64
		占是否对户外旅游大数据有了解的百分比	28.1%	20.3%	17.2%	14.1%	9.4%	10.9%	0.0%	100.0%
		占职业的百分比	23.4%	26.0%	26.8%	23.1%	12.5%	36.8%	0.0%	23.1%
		占总计的百分比	6.5%	4.7%	4.0%	3.2%	2.2%	2.5%	0.0%	23.1%
总计	计数	77	50	41	39	48	19	4	278	
	占是否对户外旅游大数据有了解的百分比	34.6%	5.1%	17.9%	6.4%	24.4%	7.7%	3.8%	100.0%	
	占职业的百分比	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	
	占总计的百分比	34.6%	5.1%	17.9%	6.4%	24.4%	7.7%	3.8%	100.0%	

Table 6. Chi-square test of the occupational cross-contingency table of outdoor tourism big data
表 6. 户外旅游大数据职业交叉列联表卡方检验

	值	自由度	渐进显著性(双侧)
皮尔逊卡方	20.940 ^a	12	0.043
似然比(L)	22.780	12	0.032
线性关联	3.439	1	0.065
有效个案数	278		

a. 15 个单元格(71.4%)的期望计数小于 5。最小期望计数为 0.69。

结果表面: 云南省大部分户外旅游爱好者对户外旅游大数据处于“听说过但尚未关注”的状态, 其职业类型以专业技术人员(教师、医生、艺术、体育工作者)、企事业单位一般工作人员和自由职业者(个

体)为主, 这类职业的特性属于有稳定的收入和可支配的假期, 是长期活跃在户外旅游的固定人群, 在后续的大数据技术及其应用推广过程中, 可以提高他们对户外旅游大数据的关注度。同理, 我们用皮尔逊卡方检验对表 5 进行显著性水平为 0.05 的假设检验。由表 6 可知, 皮尔逊卡方检验的值是 0.041, 小于 0.05 的显著性水平, 提示不同职业人群的对户外旅游大数据的关注度是显著相关的。

3.2. 大数据背景下旅游信息化平台的应用研究

我们对户外旅游大数据“有一定了解并关注”的人群展开旅游信息化平台的应用研究。户外旅游项目含有一定的风险, 爱好者对户外旅游关注度最高的是旅游线路、住宿、交通等情况, 其次是户外旅游的设施及用品、户外运动专业人才培养、户外应急与救援。在户外旅游大数据的时代背景下, 我们对云南省户外旅游信息化平台的重要程度进行了分析。

因子分析模型通过因子载荷以及旋转矩阵综合评价影响因素的重要程度, 其模型如下所示:

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \cdots + a_{2m}F_m + \varepsilon_2 \\ \vdots \\ X_p = a_{p1}F_1 + a_{p2}F_2 + \cdots + a_{pm}F_m + \varepsilon_p \end{cases}$$

矩阵表示式为:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}, A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{pmatrix}, F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$

构造的模型表达为 $X = AF + \varepsilon$ 。

首先, 通过计算数据的 KMO 值和巴特利特检验, 户外旅游大数据应用平台调查数据的 KMO 值为 0.616, 显著性为 0.000, 远小于 0.05, 应该拒绝零假设, 认为相关系数不可能是单位阵, 即原始变量之间存在相关性, 适合于作因子分析。

因子分析中变量共同度表示的是各变量中所含原始信息能被提取的公因子所解释的程度, 它描述全部公共因子对于变量的总方差所作的贡献, 说明了全部公共因子反映出原变量信息的百分比。以云南省户外旅游个信息应用平台作为变量, 借助主成分分析方法提取各变量的共同度, 结果显示提取的各个变量共同度值都比较大, 说明变量空间转换为因子空间时保留了比较多的信息, 因子分析的效果较为显著。同时, 对各变量进行总方差解释分析, 由表 7 可知, “初始特征值” 一行显示有前三个特征值大于 1, 所以我们选择前三个主成分: “提取载荷平方和” 一行显示第一主成分的方差贡献率是 69.443%, 前两个主成分的方差占有所有主成分方差的 77.706%, 前三个主成分的方差占有所有主成分方差的 83.473%, 由此可见, 选前三个主成分已足够替代原来的变量, 几乎涵盖了原变量的全部信息, “旋转载荷平方和” 一行显示的是旋转以后的因子, 提取结果与未旋转之前差别较大。

由表 8 可知, 户外旅游大数据的信息应用方面, 具有较大载荷的是: 户外目的地营销系统(0.896)、景区信息化管理系统(0.776)、爱好者信息化管理系统(0.756)、户外旅游 APP 客户端(0.744)、从业技术人员信息平台(0.740)、户外电子商务系统(0.737)、户外住宿信息共享平台(0.692)和户外交通信息化平台(0.665), 这些大数据平台主要反应了户外旅游信息状况, 对于户外旅游爱好者和从事户外旅商业的人群都是了解相关户外旅游信息的核心。搭建的户外旅游大数据平台将开放接口给各类旅行服务机构、旅游管理部门、户外俱乐部、领队及个人组织者, 鼓励各类经营者利用本平台组织各类旅行活动, 使得该平台成为链接政府、企业、游客、景区的重要数据中枢;

Table 7. Interpretation of total variance
表 7. 总方差解释

成分	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%	总计	方差百分比	累积%
1	12.500	69.443	69.443	12.500	69.443	69.443	6.330	35.166	35.166
2	1.487	8.263	77.706	1.487	8.263	77.706	5.437	30.205	65.371
3	1.038	5.767	83.473	1.038	5.767	83.473	3.258	18.102	83.473
4	0.718	3.990	87.463						
5	0.645	3.584	91.047						
6	0.411	2.283	93.330						
7	0.337	1.873	95.203						
8	0.291	1.615	96.818						
9	0.206	1.147	97.965						
10	0.112	0.620	98.585						
11	0.078	0.434	99.018						
12	0.071	0.395	99.413						
13	0.056	0.314	99.727						
14	0.030	0.168	99.895						
15	0.017	0.093	99.988						
16	0.002	0.009	99.998						
17	0.000	0.002	100.000						
18	-3.467E-16	-1.926E-15	100.000						

提取方法：主成分分析法。

Table 8. The component matrix after rotation
表 8. 旋转后的成分矩阵^a

	信息应用	安全保障	展示与收集
从业技术人员信息平台	0.740	0.311	0.311
呼叫中心系统	0.071	0.766	0.224
户外定位服务系统	0.300	0.875	0.119
户外应急救援系统	0.518	0.786	0.167
户外旅游APP客户端	0.744	0.254	0.499
户外目的地营销系统	0.896	0.121	0.207
户外电子商务系统	0.737	0.519	0.188
爱好者信息化管理系统	0.756	0.464	0.260
景区信息化管理系统	0.776	0.447	0.342
户外住宿信息共享平台	0.692	0.544	0.266
户外交通信息化平台	0.665	0.459	0.364
信息化政策法规	0.401	0.847	0.245
拓展培训	0.257	0.465	0.756
户外旅游人才培养	0.483	0.498	0.448

Continued

监管系统	0.498	0.649	0.467
微信公众号	0.582	0.185	0.706
信息安全保障系统	0.573	0.672	0.137
数据收集、存储及分析系统	0.223	0.103	0.887

提取方法：主成分分析法。旋转方法：凯撒正态化最大方差法。a. 旋转在 7 次迭代后已收敛。

在户外旅游大数据的安全保障方面，具有较大载荷的是：户外定位服务系统(0.875)、信息化政策法规(0.847)、户外应急救援系统(0.786)、呼叫中心系统(0.766)和信息安全保障系统(0.672)，这些大数据平台主要反应了户外旅游安全保障内容，为户外旅游从政策制定、信息安全及救援提高保障。建立户外旅游装备、后勤、服务收费以及客户评价的标准化体系，系统分析和全面审视云南省户外旅游领域研究的重点：如户外运动法律法规体系的健全、户外运动安全救援体系的构建、户外运动俱乐部的健康发展、户外运动市场的壮大等问题，为促进云南省户外旅游业的可持续发展、旅游文化多样性提供理论参考；

在户外旅游大数据的展示与收集方面，具有较大载荷的是：数据收集存储及分析系统(0.887)、拓展培训(0.756)和微信公众号(0.706)。这个部分主要为后续政策的制定和户外旅游大数据的进一步发展提供数据支持和理论基础。在此基础上，开发新的旅游路线实现个性化旅游定制服务；利用统计预测技术进行预警提示可实现错峰、错流旅游等，从而达到提升整个户外旅游的信息化水平。

4. 结论与建议

4.1. 结论

1) 市场关注度较低，市场规模很小

目前，云南省对于户外旅游大数据的关注度、技术开发及应用都处于概念阶段和发展初期，户外旅游开发主体大多以俱乐部、社会组织、个人等方式开展，实力较强的旅行社还未涉及户外旅游部分，市场规模的主体普遍较小。

2) 户外旅游资源尚未得到充分开发，智慧旅游和信息化水平不足

云南省拥有类型多样、特色鲜明、分布广泛的户外旅游资源，但目前的产品主要集中在徒步、登山、野营等方面，户外运动俱乐部开展的活动形式相对较单一。整个云南省户外旅游行业的数据资源差异较大，大部分重要数据尚未收集，户外旅游数据挖掘深度严重不足。同时，户外旅游数据尚缺乏准确性和完整性，平台的智能化和监管水平处于较低水平，仍处于数据孤岛和未开放状态。

3) 从业技术人员缺乏，运营监管力度不够

民间组织的无序化，云南省户外旅游发展与专业人才缺乏的矛盾突出，大多数专业教练员由领队或爱好者担当，俱乐部对员工培训不到位，稳定、优质的服务难以形成。户外运动的安全隐患问题，安全救援保障程度低，且受到经济、人员、装备的影响，救援能力不足。

4.2. 建议

云南省户外旅游业的高速发展实现了大数据应用给该行业带来的机遇，户外旅游大数据的应用从微观和宏观两个角度入手，调整特定户外旅游资源的资源和旅游计划，评估户外旅游市场的资源划分、营销方法和准确服务，还能对相关旅游部门，户外旅游龙头企业、平台和户外旅游从业人员共同促进该行业发展提供技术支持和数据保障，提高整个云南省户外旅游的整体发展水平和消费者的户外旅游理念，从而更好地促进我国户外旅游管理的运营。

1) 利用户外旅游大数据分析和预测旅游各环节的需求, 实现精准营销

针对云南省户外旅游的发展要把握户外旅游运动的特殊性, 加大安全保障投入, 强化安全意识; 充分发挥云南省丰富多样、特色鲜明的户外运动旅游资源优势, 因地制宜、分类指导, 推动不同地区、不同类型的户外旅游运动特色化、差别化发展, 提升云南省户外旅游的影响力; 建立对民间自由活动组织进行监管审批制度, 规范民间组织发展, 相关机构人员专业培训不可忽视, 加大培训力度, 推动稳定、优质服务的形成。

2) 利用大数据及时调整和控制户外旅游管理

在产业扶贫方面, 以住宿床位、协作向导、驮马和现有构筑物 and 建筑物的升级利用为基础, 带动当地居民增加收入。同时, 鼓励沿线居民根据自身条件, 开发提供民族文化、工艺品和各类生活体验项目。无论是扶贫攻坚, 还是环境保护, 只有当地居民的收入可持续增长, 且与环境质量正相关的前提下, 才可能具备最为可靠的群众基础, 旅游行业得以健康、可持续的发展。

3) 通过大数据实现户外旅游信息的共享和整合

旅游产业综合性强、关联度大、产业链长, 其广泛涉及并交叉渗透到许多相关行业和产业中, 影响、带动和促进与之相关联的行业飞速发展。户外旅游大数据的建立和发展将提升云南省旅游产业质量, 增加就业机会、促进传统旅游服务业向现代服务产业链转化, 加快实现旅游业及其相关产业向现代服务业升级。

基金项目

国家社会科学基金一般项目“大数据背景下户外旅游数据分析方法研究”(17BTJ038)。

参考文献

- [1] 蒙睿, 熊剑峰, 胡文琨. 云南户外运动旅游发展规划研究[M]. 北京: 中国旅游出版社, 2018.
- [2] 常亮, 曹玉婷, 孙文平, 张伟涛, 陈君同. 旅游推荐系统研究综述[J]. 计算机科学, 2017, 44(10): 1-6.
- [3] 钟薇, 沈洁楠, 郭志清. 我国户外运动救援体系的现状及发展研究[J]. 体育科技文献通报, 2019, 27(3): 151-154.
- [4] 王利. 户外运动的社会学思考[J]. 中国校外教育, 2009(12): 321-324.
- [5] 张鹏飞, 陆晶晶. 我国户外运动研究现状[J]. 体育成人教育学报, 2009, 25(5): 47-49.
- [6] Shen, D., Shen, H., Zhu, H., et al. (2016) The Statistics and Mathematics of High Dimension Low Sample Size Asymptotics. *Statistica Sinica*, **26**, 1747. <https://doi.org/10.5705/ss.202015.0088>
- [7] Allasonnière, S., et al. (2015) Convergent Stochastic Expectation Maximization Algorithm with Efficient Sampling in High Dimension. Application to Deformable Template Model Estimation. *Computational Statistics & Data Analysis*, **91**, 4-19. <https://doi.org/10.1016/j.csda.2015.04.011>
- [8] Zhao, J. and Shao, J. (2015) Semiparametric Pseudo Likelihoods in Generalized Linear Models with Nonignorable Missing Data. *Journal of the American Statistical Association*, **110**, 1577-1590. <https://doi.org/10.1080/01621459.2014.983234>
- [9] Kim, J.K. and Yu, C.L. (2011) A Semiparametric Estimation of Mean Functionals with Nonignorable Missing Data. *Journal of the American Statistical Association*, **106**, 157-165. <https://doi.org/10.1198/jasa.2011.tm10104>
- [10] Zhang, J., Zhang, R. and Lu, Z. (2015) Quantile-Adaptive Variable Screening in Ultra-High Dimensional Varying Coefficient Models. *Journal of Applied Statistics*, **43**, 1-12. <https://doi.org/10.1080/02664763.2015.1072141>
- [11] 李云鹏. 旅游场景驱动的大数据应用[J]. 旅游学刊, 2017(9): 4-6.
- [12] 王维晴. 基于网络文本分析的旅游目的地形象感知研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2019.
- [13] 大数据分析引发旅游业第二次技术革命[EB/OL]. <http://www.duozhishidai.com/article-67199-1.html>, 2019-05-07.
- [14] 全球十大在线旅游网站模式分析[EB/OL]. <https://wenku.baidu.com/view/9dd76fe1162ded630b1c59eef8c75fbfc77d944a.html>

-
- [15] 李红艳. 户外运动的理论与实践研究[D]: [博士学位论文]. 北京: 北京体育大学, 2006.
- [16] Buckley, R. (2016) Outdoor Tourism in China: A Foreigner's 30-Year Retrospective. *Progress in Geography*, **35**, 665-678. <https://doi.org/10.18306/dlkxjz.2016.06.001>
- [17] Buckley, R.C. (2014) Adventure Tourism as a Research Tool in Non-Tourism Disciplines. *Tourism Recreation Research*, **39**, 39-49. <https://doi.org/10.1080/02508281.2014.11081325>
- [18] Jing, L., Ng, M.K. and Huang, J.Z. (2007) An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data. *IEEE Transactions on Knowledge & Data Engineering*, **19**, 1026-1041. <https://doi.org/10.1109/TKDE.2007.1048>
- [19] 刘长生, 简玉峰. 我国旅游业发展与经济增长的关系研究——基于不同省份的个体数据和面板数据分析[J]. 旅游科学, 2008, 22(5): 23-32.
- [20] 杨桂华. 旅游资源与开发[M]. 昆明: 云南大学出版社, 2010.