

关键,大数据时代,对AI,也就是我们,也有新要求。

我们期待融合,首先不能回避的是自身的融合。

五、跟风热潮与战略定力

“只有当潮水退去的时候,才知道是谁在裸泳。”这是巴菲特的一句名言。当今时代,经济、社会、文化发展变幻令人目不暇接。光怪陆离中诱惑多多,产业发展确实需要战略定力,旅游大数据发展同样也需要战略定力。“战略”一词,是指从全面和长远两个维度考虑谋划,是理念层面,需要我们认准方向,整体考量。“定力”一词,是宗教用语,指正念坚固、不为境转,在实践层面要求我们持之以恒,百折不挠。

旅游大数据建设不能一蹴而就,大数据的作用展现也非一朝一夕,成功不仅需要选对道路,找对方向,重要的是比拼沿着既定方向前行的忍耐力。

回顾互联网经济发展的十年,有两句话值得回味,一是“后发优势”,二是“风口理论”,这两句话的形成和因果,包含了太多的色彩。在这个大数据的十年,特别值得回味。

“战略定力”是理性的代名词,是练达的标志,也是一个成熟产业的基本内涵。旅游大数据的冷与热,折射出产业转型期的躁动特征,也正在考验着我们的战略定力。

(第一作者系该中心主任,第二作者系该中心运营处处长;收稿日期:2017-06-10)

大数据旅游研究和应用中的几个问题

杨 旻^{1,2},刘法建³

(1. 天普大学旅游与酒店管理系,美国 费城 19122;
2. 天普大学美亚旅游与酒店研究中心,美国 费城 19122;
3. 安徽大学商学院,安徽 合肥 230601)

Doi: 10.3969/j.issn.1002-5006.2017.09.002

科学研究的发展往往始于技术的进步。近年来,在信息科学、网络技术的进步和融合的推动下,获取便捷、类型多样的大数据开始为众多研究领域所关注。大数据也成为国内外旅游学界和业界的热门话题,尤其在旅游流、旅游体验等对数据精度、规模、结构要求日益严苛的领域,大数据使得一些原本数据搜集复杂或者可操作性低的研究课题可以得到实施。然而,随着研究领域、数据来源的不断拓展,一些大数据应用的问题也随之而来。

一、虚假评论数据

作为最具代表性的旅游大数据来源,在线顾

客评论在已有的旅游大数据研究中占有重要的地位。此类研究,往往基于在线评论能客观反映旅游服务设施接待水平和游客体验的前提下展开分析。但事实上,一些点评网站并不能识别评论者是否购买过该项服务(如TripAdvisor),这导致了大量虚假评论的滋生,势必会影响到分析结果的有效性和科学性。目前,在国际顶级的商学期刊上,已有使用机器学习和人工智能方法来筛选和过滤虚假在线评论的研究成果发表。因此,相关旅游研究在使用在线评论数据时,也要特别注意可能存在的虚假评论,并且基于研究情境考虑使用合适的方法筛选出可能的虚假评论,提高数据本身的可信度。

二、不同大数据源之间的交互验证

对于同一个研究目的/研究问题,不同的数据源可能提供来自不同侧面的重要信息。这些数据本身或多或少会存在系统性及非系统性误差,使用时就需要对数据源进行交互验证(triangulation),以确定这些误差可能的来源,并为后续数据汇总过程中的误差控制提供思路。例如,在使用酒店在线评论数据时,很多研究者会从不同的酒店评论平台采集数据。但由于不同的平台可能是面向不同层次、不同类型的住客群体,这些评论数据存在系统差异。研究过程中首先就需要对这些平台数据进行交互验证,这不仅有助于降低误差,提高评论数据综合后的结果可信度,而且可以了解不同样本群体的特性差异,增强对结果的理解和解释。

三、数据样本的代表性

在传统的数 据采 集中,样 本的代表性(representativeness)是统计推断是否稳健和有效的重要前提。在社会统计和商业调查中,面对不具代表性的数据样本,再强大的统计模型也很难得出准确的、可以推广的分析结论。因此,科学、有效的抽样调查方法在传统数据获取中被普遍重视。在大数据时代,尤其是在大数据的旅游业应用中,如何确保样本和数据的有效性、代表性也是业界和学界重点关注的问题。实际上,很多大数据源产生的数据,并不能有效的代表完整的研究目标群体。最明显的例子就是从互联网获得的大数据中,中老年群体的样本比例严重不足。传统统计学中可使用的权重调整方法这时也较难使用,因为此类大数据样本中个体属性,尤其是社会人口属性特征往往缺失,使得权重的估计也变得难以操作。此外,计量经济学中常见的样本选择偏差(sample selection bias)也会导

致样本的代表性问题。一个常见的例子就是不满意的酒店住客可能倾向于在评论网站上留下详细的评论。如果这种样本选择偏差不得以修正,就容易过度放大负面信息,得出与事实出入很大的结果。

四、游客样本的甄别

除一些旅游业专用大数据源外,旅游研究中还会使用到一些综合性的、非定制的大数据源,比如微博数据、汽车GPS轨迹数据和手机信令数据。这时,首先需要解决如何甄别有效游客样本的问题。一般常用的方法是建立统计模型对数据中的时空信息进行数据挖掘,识别出游客群体。由于建立统计模型需要识别诸多参数,这可以通过小规模实地调查来获得,从而有效提升游客样本的甄别准确率。当然,由于不同地方的景点规模、目的地空间尺度、空间结构、交通设施水平的差异,不同地区的实地调查结果不能盲目推广。

五、研究方法“旧瓶装新酒”

许多标称“大数据应用”的旅游研究文章,使用的仍是传统和基础的统计和数据发掘方法。严格意义上说,难以纳入“大数据”研究的范畴。因为它们并没有涉及到“大数据革命”的核心之一:对于海量实时数据处理方法的革新。很多情况下,这些大数据旅游实证文章,仅仅是从大数据源中获得常规数据,增加的只是数据量,并没有带来创新性的、突破性的研究方法,未能实现研究思路和研究框架上的质变。

综上所述,在旅游的大数据研究和应用中,不应该盲目迷信和崇拜大数据源获得的数据。大数据与生俱来的诸多弊端,需要使用者在数据的获取、处理和分析过程中时刻注意和警惕。首先,研究人员和从业人士不能简单地把大数据分析看成一个“黑箱”,不假思索、教条式地使用统计或者人工智能的方法分析数据。而应把一部分注意力放在保证数据的质量上,增加针对性的数据清理和调试,以确保数据的有效性和科学性。其次,在大数据的应用领域,不同大数据源信息的融合越来越被重视。这种信息的汇总、融合可以提高数据的精度,提升结果的可靠性。最后,对于学界而言,大数据的兴起与流行并未给传统的旅游研究范式带来根本性的改变。作为社会科学的分支,旅游研究的“问题导向”(research question)和“理论建构”并重的属性依然关键,其学术研究的价值在很大程度上仍取决于研究问题的水平和质量。

(第一作者系该系助理教授,第二作者系该中心访问学者;收稿日期:2017-06-10)

旅游场景驱动的大数据应用

李云鹏

(首都经济贸易大学工商管理学院,北京100070)

Doi: 10.3969/j.issn.1002-5006.2017.09.003

百度指数(即搜索热度)显示,“大数据”一词从2012年5月27日开始有搜索记录,随后每年都会出现一个周搜索峰值,5年来最高搜索量达到10435次/周。近半年以来,“大数据”的搜索热度超过了“人工智能”,对旅游业来讲大数据应用更有价值和意义。

一、大数据在旅游业的应用现状:基于游客数据的统计分析

在旅游业这样一个边界模糊的复杂开放系统中,伴随着游客空间位置实时变化,动态数据随时产生、随时发生变化。旅游业拥有庞大、类型多样的数据,且涉及到很多关联行业的数据,总之,大数据在旅游业的应用研究是非常复杂的,目前旅游大数据应用主要聚焦于客源市场分析领域。

1. 旅游目的地大数据分析。目前,承担旅游目的地大数据分析业务的企业,通过对通讯运营商手机位置数据、搜索数据、交通流量数据、交易数据、社交数据等不同来源数据的系统整合分析,为旅游目的地构建了全面的客源市场分析。旅游目的地基于大数据的统计分析日渐成型,实现了使用大样本量(几乎接近全体样本)数据取代传统抽样调查进行客源市场分析,且在分析维度和颗粒度上,基于大数据的统计分析比抽样调查更加丰富,帮助目的地旅游管理部门更加全面地认知客源群体,实时了解客源群体在目的地空间分布状况,从而对客流预警和调控形成一定的帮助。

2. 互联网平台交易大数据分析。平台型旅游企业,如携程、去哪儿、飞猪、金棕榈等利用其平台经营过程产生的交易数据、评价数据和日志数据,对用户进行全方位画像和消费行为分析,形成对于其产品信息浏览情况、产品受欢迎程度等诸多方面准确的统计分析,赋予游客多个维度的属性标签。平台企业根据旅游产品类型与游客标签组合的匹配程度进行产品和服务推荐,企业利用互联网平台产生的大数据可以更全面地了解用户及企业自身业务开展情况。

基于游客跨时空行为模式及用户在线使用情况、形成的游客驱动的旅游大数据分析,主要价值在于对全样本数据特征的深度了解,其优势是游客数据相对容易获得,然而对于这种游客(用户)

驱动的大数据分析,其问题主要是数据分散在不同的网络载体和实体空间中,不同网络载体和实体空间之间难以实现同源数据之间关联,数据的准确性和真实性时常难以判断。此外,数据清洗能力也影响了大数据分析结果的实际应用价值。因此,要探究大数据分析在旅游业应用的更多价值,需要改变一种视角来考察大数据在旅游业的应用。

二、基于旅游场景的大数据分析:大数据在旅游业应用的新方向

2017年5月28日在贵阳举办的“数博会”上,马化腾提出,发展大数据“更重要的一个要素是场景”。智能时代的用户,每时每刻都被智能场景包围着,场景中不同角色之间的互动,使得场景具备了特定的细分标签,这样的细分标签进一步抽象,成为个性化推荐的基础。

旅游场景是鲜活、具象化的实体和虚拟空间的总和,不同来源的大数据在旅游场景中关联,使数据维度更加丰富和有效。因此,大数据在旅游业应用从基于全体用户/游客数据的孤立视角,转变为在特定场景下用户/游客作为场景参与者之一,通过不同大数据来源建立起每个用户/游客与场景之间相互关联的数据化关系谱系,带动场景中相关主体进行大数据关联挖掘和共享,形成一个可控、可操作且可以延展更多服务内容的大数据应用。

场景对消费者购买行为具有直接影响,成为购买产品决策的一部分。场景减少了产品和用户之间的距离,改善了信息不对称,使商品与消费者彼此之间可以更便捷的发现,产生情景交融的互动方式。在网络化时代,实体场景与虚拟网络空间多层次互动,消费者更加容易被触动从而产生购买欲望。伴随着移动互联网的发展以及LBS(基于位置的服务)服务内容更加多样化,这种场景驱动型的搜索与发现模式,为旅游业带来了巨大的增长空间。

场景化是旅游业的一个重要特征,场景对游客游览行为和消费行为都有重要影响。场景差异是生活和旅游的本质区分,场景对游客在不同活动空间下的行为影响相呼应,即场景链接带动消费行为发生。旅游产品在本质上就是在一种特定的场景化体验,每一个旅游场景都具备融合能力,通过场景融合产生不同的产业形态。

旅游场景是旅游业供需关系之间的一种反映方式,不同场景串联构成了旅游业的供需空间,旅游场景也是实体空间与虚拟空间互动关系的一种反映方式,浓缩不同类型的旅游供需空间,集聚各种大

数据,通过大数据分析反映每一种具体旅游场景的运行规律,可以为旅游业创造更多样化的发展模式。

三、典型旅游场景的大数据应用

为分析方便,我们把若干种旅游场景归类为4种典型场景,包括旅游消费场景、旅游管理场景、旅游营销场景和旅游服务场景。

1. 旅游消费场景:用户在旅游决策、旅游活动过程中的各类场景中都可能涉及到某种形式的消费,消费内容会受到偏好和场景激发的共同影响。利用用户在不同网络渠道留下的痕迹,对用户偏好进行画像,场景中的企业对消费内容进行个性化设计和推荐。场景激发是要找到与用户画像相匹配的场景设置,如用户在居住地生活偏好与所在旅游场景的匹配,根据用户标签和所在场景进行用户化产品设计,提供景点周边特色餐馆、景点的营业时间、停车场位置等等指南性信息。需要大数据分析技术从海量、实时的信息中提取客户真正需要的数据,满足游客的这些个性化需求。任何场景都可以与周边的各种商品和服务资源产生交互和连接,产生交互和连接之后叠加相应的分类信息、比价信息、点评信息等,可以极大的丰富游客在旅游场景中的消费体验。

2. 旅游管理场景:景区大门口或者景点周边区域,是典型的舆情分析和客流预警分流场景。景区管理者通过通信运营商位置数据了解游客来源情况,预测进入区域人员动态图,监控人员流向、流速并及时调整车辆运营班次;通过闸机与景区电子商务平台数据实时对接,可以实时控制买票数量和门票可用时间;景区管理部门可以通过在微信、微博上的发帖的实时数据了解舆情动态。

3. 旅游营销场景:在旅游目的地营销场景,需要通过大数据了解游客选择某个旅游目的地的动因,这些动因中哪些是可以通过营销活动进行引导的,哪些是偏刚性需求,哪些是与家庭结构有关,哪些与工作性质相关等,这些问题都需要通过分析在不同平台和渠道之间的数据来解决。

4. 旅游服务场景:游客咨询服务中心和旅行社门店是服务场景中最常见的服务场景,当用户进入场景时,服务人员根据游客咨询信息进一步推荐产品和服务给游客;酒店(住宿场所)是常见的线下服务场景,服务人员对客人身份识别及快捷服务、酒店消费及本地服务产品集成和推荐,离不开对游客数据分析和本地资源数据集成。酒店可以帮助客人建立在线临时社区,通过酒店wifi链接形成游客自组织,从实时数据挖掘中发现有价值的服务内容

并提供相关服务。

四、旅游场景驱动的大数据应用促进旅游创新发展

实际旅游业务场景是上述场景的组合,例如,消费场景中也包含了管理业务、营销业务、服务业务。每一个旅游业务场景都是一种综合的大数据分析和处理情境,如何增加游客便利性、可信度、综合价值,是旅游场景应用大数据的一个核心诉求。例如,在一个旅游购物场景中,每一件实物产品所承载的各种数据,如产地、生产者、原材料、品质等,构成了游客购买的可信度基础。游客通过社交媒体或者电商平台了解到其他游客购买评价、通过第三方支付平台完成在线支付,从而完成整个购买决策,大数据都能带给游客便利性价值。

场景驱动的创新对于旅游发展变得越来越重要。在特定场景之下,游客个性化需求才能被有效定义,在大数据分析基础上的场景服务或供给才更加有效。随着人工智能、大数据等技术进一步成熟,旅游场景的大数据应用将带来旅游业供需空间的优化,带来旅游产品和服务与个性化需求精准匹配,带给游客便捷、高效率和高水平的体验过程。

旅游场景的大数据应用,能够大大提高旅游资源的利用效率,为各行各业的跨界者和跨界产品提供数据支撑,实现跨界创新和跨产业融合。在每个特定场景下,发挥大数据的统筹能力,将用户与场景串联起来,形成若干智能化场景的链接,就可以有效促进游客活动全过程的智慧化,带来的企业组织方式变革和业务流程重组,带动全域范围内旅游休闲活动的优化,提高旅游业运行效率、带动旅游业效益提升。

(作者系该院副教授;收稿日期:2017-06-10)

大数据和旅游统计

文玲¹,李瑛²

(1.海南波罗密信息科技有限公司,海南海口 570216;

2.西北大学经管学院,陕西西安 710127)

Doi: 10.3969/j.issn.1002-5006.2017.09.004

一、现状评估:市场需要专业的旅游数据分析公司

伴随着智慧旅游建设,越来越多的省份把旅游大数据建设提到了议事日程,甚至很多省份已经完成了阶段性平台建设工作。一个共同现象,是由IT公司做建设方案,用政府采购方式进行开发。几年

下来,建设和应用的实际融合,仍然是面临的主要问题。以笔者多年业内实际工作的经历,从第三方的角度看,核心是必须改变以IT公司为主导建设旅游大数据的局面。旅游产业远不像作为旅游者所理解的那样简单,实践已经发展到了供给细分的阶段,市场需要专业的旅游数据分析公司。

比如,现在公认的旅游大数据在旅游行业的应用,主要是旅游市场细分、旅游营销诊断、景区动态监测、旅游舆情监测等。目的是提升协同管理和公共服务能力,推动旅游服务、旅游营销、旅游管理、旅游创新等变革。但几乎所有的大数据建设方案都不包含旅游统计。

大数据、旅游数据、旅游统计数据,是旅游大数据涉及到的3个数据层次。“大数据”提供了可供测算的基础和技术方法,“旅游数据”是将各种单一数据汇集、挖掘形成各类需要的数据,以便为旅游管理工作提供数据支撑。而“旅游统计”则是按照国家旅游局和国家统计局用法规形式规定的数据产生格式形成并发布的权威数据,具有法律效力。旅游统计是旅游数据中最重要的一个部分,是旅游数据的顶层数据。旅游数据的应用是建设大数据平台的核心数据需求,远离核心数据需求的旅游大数据平台,就像无水之源、无本之木,对旅游管理者来说,无异于隔靴搔痒。

让IT公司做旅游统计,是其技术背景所无法承担的,也偏离了IT公司本身的业务方向。即使IT公司搭建出了符合旅游统计需求的大数据平台,后续持续的数据分析工作也是IT公司所无法承担的。如果能够承担,只能说明IT公司转型成功,已经成为一家专业的旅游数据分析公司。

二、专业评估:对旅游统计的大数据认识

(1)旅游统计

长期以来,我国的旅游统计依据国家旅游局和国家统计局颁布的《旅游统计调查制度》,统计项为“人数”和“收入”。2000年,国家旅游局引入了和国际接轨的旅游卫星账户核算方式,计算旅游产业的规模和贡献。但数据采集方式仍然是基于传统统计学的填报和抽样调查。

比如游客数量,既不是到达目的地的总人数,也不是过夜人数,而是规模以上酒店入住游客的前台登记人数,叫“人次”。游客的定义是离开惯常地10千米超过6小时,不以赚钱为目的的活动。定义比较宽泛,界定困难,特别是一日游游客的界定。另外,旅游收入的计算依据是按照万分之三的样本