

基于互联网大数据的旅游景点推荐研究

贺治皓 张桂花 王科

四川大学锦城学院 四川 成都 611731

【摘要】本文针对在互联网大数据时代下对各地的旅游景点以及层出不穷的网红景点作出相应的推荐,借助大数据的优势,助力推动发展疫情过后的旅游行业 and 解决当前旅游行业所存在的问题,使其恢复生机。方法:本次调查在收集统计整合所有班级的满意度调查表数据之后,共有 626 名教师得分有效。

【关键词】大数据; 互联网; 旅游信息; 数据挖掘

在这个科技高度发达的时代,互联网给人们带来了诸多效益。特别是在近几年内,大数据技术突飞猛进,让这个数据量庞大的时代得到了改变,给人们的生活带来了诸多的便利,给各大行业带来了更多的发展机会。面对突如其来的疫情,对当今竞争激烈的旅游市场带来了不小的动荡,阻碍了其正常的发展。对于刚被开放旅游限制的旅游行业来说,首要的目的是吸引更多的游客游览,并不断开发创新旅游产品周边等,以此来弥补疫情所带来的冲击。

1 大数据

一个简单的“大”字,足以体现这是个信息化高度发达的时代,如今的数据量早已不同于以往,即使是在极小的时间段内,都会产生数 TB 的数据量,传统的处理方式早已无法容纳处理如此庞大的数据,也意味着大数据时代的到来是发展的必然结果。IBM 曾经提出大数据的 5V 特点: Volume (大量), Velocity (高速), Variety (多样), Value (价值), Veracity (真实性)。五大特点形象地描述了大数据:极短的时间内会产生大量的数据,且数据的传递具有高速的性质,大数据的数据有多种来源一般可分为五大类: a. 交易数据,如 POS 机数据; b. 移动通信数据,如手机漫游数据; c. 社交媒体数据,如通过微信、博客等产生的数据流; d. 机器和传感器数据,如来自感应器、GPS 系统的数据; e. 互联网上的开放数据来源,如政府机构免费提供的数据^[1]。通过对所拥有的数据进行分析,将其转换为有意义的信息,也就实现了数据的价值,数据并不只是一味的真实,它可能存在一定的虚伪性以及会获取到一些脏数据和“噪音”数据等,都会影响到其真实性的存在。即对人们而言,可以通过这五大特性来衡量大数据。

2 大数据技术在旅游行业的运用

2.1 当今旅游行业存在的问题

现如今的旅游行业,充斥着各种鱼龙混杂的旅游机构,当然也不乏有一些口碑优秀的机构。如今的人们在选择出行旅游时,大部分会选择从一些生活 APP 上(如美团,大众点评等)去做一个选择,传统的旅游机构已不再具有优势,人们在选择旅游时拥有的信息量会更多,也意味着人们对于旅游行业的要求会越来越高。并且人们的旅游偏好比起以往有了较大的改变,在以前旅游行业没那么发达的时候,只有从旅行机构单一地获取信息来源,所以人们往往只会知道当地最出名的历史文化遗址地点,各个的地方的旅游景点也趋于单一,能够选择的地方往往也只有零星几个,当旅游行业想要收集旅游客户反馈数据时,会显得十分麻烦以致于不能及时地改善服务,达不到实时性的效果。现在这些情况都得到了一定的解决,但伴随着疫情的影响,整个旅游行业充斥着动荡,相较于往年少了几个数量级的营业额,无法体现出旅游为国家或地区推动经济发展的效果。随着限制解除,人们才慢慢相继出去旅游,这就引出一系列问题:一些以往著名的旅游景点兴许能恢复当初的热闹,但一些曾经就门可罗雀的景点,在这一波影响之后,更是跌入了低谷,阻碍了其当地经济的发展。旅游的限制是解除了,但并不代表着就是百分百的安全,一旦某些景点的人数超过理应收纳的最大量,也许会导致不堪设想的后果。不仅如此,游客们总是会去讨论批评一些地方景点强买强卖的这种行为,让本应该享受旅途带来的乐趣大打折扣。这些都是目前旅游行业存在的一些致命性问题。

2.2 大数据技术如何改变旅游行业

通过大数据技术,我们能及时地接受从四面八方传来的数据,能实时对数据进行处理转化为有意义的消息,并实时对行业服务以及问题做出改善。从对 app 进行改

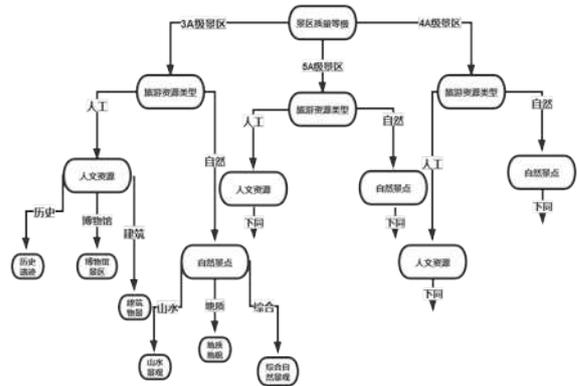
动开始，当人们在搜索栏中输入相关的景点名称时，后台可以通过相应的大数据技术——flume 能实现后台对 app 产生数据的实时监听，它能将 app 所产生的数据储存在存储器中（如 HDFS），到手的数据可以经由 Hadoop 技术的处理，将其转化为有价值的信息。通过这些数据的收集，能将人们的旅游偏好进行汇总分析，了解到哪些景点是人们旅游的热点，人们的旅游偏好和一些消费行为，从而能更快地将人们期望的旅游景点路线推送到 app 上，实现实时的内容推送，并且可以对市场的预期做出一个评估，有效推动旅游行业的经济建设。而对于那种被埋没的景点，应由 app 更为主动地去进行一个内容推荐，在推送一些著名景点的同时，在旁边用一些不太出名的景点进行点缀，实现内容推荐上的全面化。

在目前的大部分 app 中，尽管将一些景点分成了几个大类（如清凉玩水，自然风光，动植物园等），但真正当人们去选择时，往往又会出现选择疑难症，这时可以通过数据挖掘的算法——决策树来进行更为详细的分类。在数据挖掘和数据分析中，决策树起着非常重要的作用，其可以构建一个决策系统，对数据进行预测。决策树的结构类似树型结构，利用树的结构将记录进行分类，树的一个叶结点就代表某个条件下的一个记录集，以此重复，便可生成决策树^[2]，下面为测试集导入 python 中的部分核心代码：通过根节点的景区质量等级分类，划分为主要 3 个等级的景区，在各个等级的景区中再引入两项不同的旅游资源类型，通常将其分为两类，一类为人文资源，一类为自然景点，两类不同景点再经过细化，人文资源又囊括了历史遗迹，博物馆景区等，自然景点也包括山水景观，地质地貌以及综合自然景观。

```
def createDataSet():
    dataSet = [['3A级景区', '人工', '历史遗迹', ''],
              ['3A级景区', '自然', '', '山水景观'],
              ['4A级景区', '人工', '博物馆景区', ''],
              ['4A级景区', '自然', '', '地貌地质'],
              ['5A级景区', '人工', '建筑物景', ''],
              ['5A级景区', '自然', '', '综合自然景观']]
    labels = ['景区质量等级', '旅游资源类型', '人文资源', '自然景点'] #四个特征
    #change to discrete values
    return dataSet, labels
##### 按给定的特征划分数据 #####
def splitDataSet(dataSet, axis, value): #axis是dataSet数据集下要进行特征划分的列号
    #例如景区质量等级是0列，value是该列下某个特征值
    retDataSet = []
    for featVec in dataSet: #遍历数据集，并抽取按axis的当前value特征进行划分的数据集
        if featVec[axis] == value: #
            reducedFeatVec = featVec[:axis] #chop out axis used for splitting
            reducedFeatVec.extend(featVec[axis+1:])
            retDataSet.append(reducedFeatVec)
            # print axis, value, reducedFeatVec
    # print retDataSet
    return retDataSet

##### 生成决策树主方法
def createTree(dataSet, labels):
    classList = [example[-1] for example in dataSet] # 返回当前数据集下标签列所有值
    if classList.count(classList[0]) == len(classList):
        return classList[0] #当类别完全相同时则停止继续划分，直接返回该类的标签
    if len(dataSet[0]) == 1: #遍历完所有的特征时，仍然不能将数据集划分成仅包含
        #唯一类别的分组 dataSet
        return majorityCnt(classList) #由于无法简单的返回唯一的一类标签
        #这里就返回出现次数最多的类别作为返回值
    bestFeat = chooseBestFeatureToSplit(dataSet) # 获取最好的分类特征索引
    bestFeatLabel = labels[bestFeat] #获取该特征的名字
```

在经过决策树的运算后，从而得到一个比较纯净的数据集。例如下面导图的分类：



通过以上更为直观地划分景点，在保证利用简单的条件将数据集变纯的同时，也控制了决策树的深度问题，实现了很好的物以类聚的效果，人们在选择上又有了更为精确的方向，从而能更快更精准地定位到自己想去的地方。现在的人们对于旅游地点的选择呈现出多元化，对于以往的一些传统景点，他们可能并没有那么强烈的欲望想要去选择，这也就诞生了现在俗称的“网红景点”，它们称不上是真正意义上的景点，但随着来往的人们越来越多，也逐渐成为了一条旅游道路，吸引着众多年轻人，也一定程度上推动了经济效益。于是，在 app 上不再只有传统景点的选择，通过大数据技术收集分析各个地方的网红景点，将许多网红景点推送出现在推荐中，这样可供选择的地点将会不计其数，之前也许只会去传统景点游玩 1-2 天，在有了网红景点后，人们更会倾向于再待个几天去走遍一座城市的网红景点，客观上地延长了游客们的旅游时间，能随着网红景点带来更多可能性的发展。

在景点进行愉快游玩的同时，也不能忽视卫生安全的问题，虽然景区会要求人们佩戴口罩，但在人数众多的情况下，也并非那么安全，这时可以通过景区所安装的摄像头来对当前景区实时人数进行一个统计，再将实时人数数据反馈到 app 页面，这样人们在 app 上对景点进行选择时，能看到实施人数有多少人考虑去否，亦或是选择另外的景点，实现交叉式游览，让一个景点不会应人数过于而提高风险，同时也能为一些人数较少的景点带去可观的客流量，实现经济共赢。在大数据时代，游客们可以通过在 app 上对各个景点进行评价，这也使得相关旅游行业必须更加重视游客们的旅游体验，景点拥有一个良好的旅游服务，是城市的一个重要加分项，也能多多少少折射出当地人的一些品行。强买强卖这个现象在众多景区都是司空见惯的东西，严重影响游客的体验，导致美好的心情被整得郁郁寡欢。要改变这一现

象, 政府以及相关部门要从后台抓取数据, 从游客们的评价中制定规范准则, 用于督促旅游行业的服务标准, 对于那种不按规范准则的相关机构一定做到严惩不贷, 共同维护良好的旅游环境, 实现旅游服务的高标准。

3 结语

互联网大数据时代是一个崭新的充满机遇的时代, 疫情过后的旅游行业需要迅速恢复生机, 就要充分利用好大数据技术以及大数据分析, 结合目前旅游行业的结构组成, 利用数据挖掘获取足够多的游客信息, 对其偏好和需求进行分析。并借着互联网的优势, 有效地推广本地景点的特色以及特色周边, 目前人们最大也是最迅速的信息来源之一就是互联网, 相关旅游行业部门可以加大技术支持, 从而得到更具成效的结果。传统旅游行业转型升级是必然趋势, 明确旅游行业定位, 优化旅游

服务质量是旅游行业发展的必由之路, 只有这样才能够促进旅游行业由“景点旅游”到“全域旅游”的转变, 实现旅游产业的优势地位^[3]。疫情后的中国经济需要借助互联网大数据之手来更好发展, 不只是旅游行业, 各行各业要利用大数据技术从充满“噪音”杂乱的数据中去发掘有价值的信息, 真正意义上地实现科技带动经济不断发展, 成就更为辉煌的祖国。

【参考文献】

- [1] 曾忠禄. 大数据分析: 方向、方法与工具 [J]. 情报理论与实践, 2017, 40(1): 1-5.
- [2] 薛亚楠, 杨晓东. 基于决策树算法在学生成绩上的应用 [J]. 科技资讯, 2019, 17(36): 83-85.
- [3] 李传娟. 大数据视域下旅游服务质量优化路径研究 [J]. 西部皮革, 2019, 41(24): 85.