

大數據在旅遊研究中的運用

——基於國內文獻研究

王 興

(澳門理工學院,澳門)

摘 要:通過對國內旅遊大數據研究文獻的收集和整理,分析了國內旅遊大數據的運用領域、主要的數據來源和研究方法等,希望為國內旅遊大數據利用提供參考。研究文獻來自中國期刊全文數據庫,文獻分析主要採用內容分析法。通過對中國期刊全文數據庫收錄的 123 篇旅遊大數據實證研究文獻的分析,發現國內旅遊大數據的運用主要集中在以下五方面:遊客偏好、遊客預測、遊客流分析、遊客滿意度和旅遊營銷等。數據來源主要來自攜程、去哪兒和馬蜂窩等旅遊類網站和新浪微博、豆瓣等社交平台的 UGC(用戶產生的內容)數據、Flickr 網站的圖片數據,以及百度指數和手機定位數據等。本研究認為旅遊大數據能彌補傳統旅遊研究的部分缺陷,但旅遊大數據自身也有較大的缺陷,研究者需要根據具體問題和研究目的,正確地選擇數據來源和分析方法,減少數據的“噪音”污染。

關鍵詞:大數據;旅遊大數據;數據運用;內容分析法

中圖分類號:TU981

The Application of Big Data in Tourism Research: Review of Literature Based in China

Wang Xing

(Macao Polytechnic Institute, Macao)

作者簡介:王 興,澳門理工學院博士研究生。

Abstract: Through collecting and reviewing China's tourism research on big data, this paper analyzed the application field, the major data sources and the research methods, in an attempt to provide references for the application of big data in China's tourism. The research literature came from China Academic Journals full-text database. The literature analysis primarily adopted the content analysis method. Through analyzing 123 empirical research articles on tourism big data included in the China Academic Journals full-text database, it was found that the application of tourism big data in China mainly focused on the following five areas: tourist preference, tourist prediction, tourist flow analysis, tourist satisfaction and tourism marketing. The data sources were mainly from UGC (user-generated content) data from such tourism websites as Ctrip, Qunar and Mafengwo and social platforms like Sina Weibo and Douban, image data from Flickr website, as well as Baidu index and mobile roaming data. This study argued that tourism big data might make up for some of the shortcomings of the traditional tourism research, but tourism big data also had its own major shortcomings. Researchers need to choose the right data sources and research methods according to their specific problems and research purposes to reduce the "noise" pollution of data.

Key words: big data; tourism big data; data application; content analysis

引 言

大數據日益進入人們的生活,在很多行業和領域發揮重要作用,比如網上零售行業、通訊行業以及公共衛生領域、國家安全領域等。由於通訊技術的發展,人們創造信息和獲取信息的能力不斷增強,通過互聯網和通訊設備,可以較容易進行信息交換,信息的大量增加使得人類進入了大數據時代(曾忠祿,2015)。科技發展也使得人類保存數據的能力越來越強,隨時隨地產生的數據都可能被記錄和保存下來。比如,個人在互聯網上的搜索和瀏覽記錄,以及每天行為軌迹的地理位置信息等。個人產生的數據不斷被記錄和保存,通過大數據分析的方法,可以進行人物畫像、人口流動趨勢預測,以及危機預警等。所以,大數據分析已經成

為各行各業不可缺少的分析方法,也是促進各行業發展的一個主要助推器。

旅遊活動涉及吃、住、行、遊、購、娛等方面,旅遊業是產生各種信息數據的沃土。在旅遊前,人們可能需要比較不同的旅遊目的地,搜索目的地景點、酒店、天氣、交通和風土人情等信息;在旅遊中可能產生大量的行為軌迹信息,包括照片分享、撰寫旅遊日誌、發表感想等;在旅遊後可能產生對旅遊的評價和反思信息。這些旅遊信息以文字、圖片、視頻和地理位置信息等各種各樣的數據形式存在,數量龐大、結構複雜多樣、數據產生速度快、數據密度價值低,非常符合大數據的特徵。如何利用這些龐大的數據更好地瞭解遊客,瞭解市場趨勢,改善旅遊服務,對旅遊界、學術界和政府決策者都具有十分重要的意義。然而,國內旅遊大數據目前的運用趨勢、具體的運用領域、數據的主要來

源,以及優勢及缺陷等,目前還未發現有系統的研究。因此,本文希望通過對我國旅遊大數據研究文獻的分析,補充這方面的缺陷。

1 文獻來源與基本情況

1.1 資料來源

本研究的資料來源全部來自中國期刊全文數據庫。該庫是世界上最大的中文期刊全文數據庫,內容覆蓋自然科學、工程技術、醫學、人文社會科學等各個領域。目前,收錄國內期刊 11,453 種,全文文獻總量 69,203,690 篇。

資料收集時間為 2021 年 7 月 31 日,在中國期刊全文數據庫中搜索篇名中有“大數據”與“旅遊”二字的核心期刊文章,檢索到 74 篇文章(核心期刊指入選北大核心或者 CSSCI 來源的期刊)。擴大搜索範圍,增加了“互聯網”+“旅遊”作為關鍵詞檢索篇名,最終獲取超過 500 多篇文章。通過閱讀文章摘要,刪去研究綜述、理論探討、筆談、書評、短論等不涉及實證研究的文章,最終獲得 123 篇樣本。本研究採用內容分析法,通過人工閱讀,對 123 篇文章進行分析,總結大數據在旅遊研究中的運用情況。

1.2 基本情況

通過對論文的篩選和分析,發現國內利用 UGC(用戶產生的內容, User-generated Content)數據進行旅遊研究的文章最早可以追溯到 2006 年,當時有作者根據網上的自駕車旅遊網絡論壇上的帖子討論旅遊網絡論壇。但收錄的帖子僅 499 個,還算不上大數據。此外,該研究也不是討論旅遊,而是討論網絡論壇。刪除該文章,最早的文章是 2009 年的一篇文章。文章對大陸和台灣共 41 個網站中關於台灣旅遊的文本進行了收集,並採用 ROST WordParser 軟件對文本進行內容分析來研究台灣旅遊形象(肖亮和趙黎明,2009)。該文可以算是國內最早關於旅遊大數據的實證研究論文。但國內連續的涉及大數據的旅遊研究真正開始的時間是 2011 年,當年共有 4 篇研究論文。

根據對 123 篇旅遊大數據實證研究文章發表時間的統計,可以發現,國內旅遊大數據的實證研究大致可以分為三個階段:2009-2013 年為萌芽期,2014-2016 年為初步發展期,2017-2019 年為高速發展期。從 2020 年開始,旅遊大數據的研究開始減少,這可能與全球疫情爆發,旅遊活動的大量減少有關。具體發文量見圖 1。



圖 1 國內旅遊大數據研究發文量趨勢

從圖 1 可以看出,在 2013 年之前,每年關於旅遊大數據的文章還不足 5 篇;而 2014-2016 年間,旅遊大數據研究開始有明顯的增長,每年的發文量已經接近 10 篇;2017 年之後,旅遊大數據研究突然猛增,每年利用大數據進行旅遊實證研究的文章超過 20 多篇。因此,我們得出結論,我國旅遊研究對大數據的重視是在 2013 年以後,2018 年達到高峰,我國大數據在旅遊的運用趨勢與世界趨勢基本一致。Song(2017)研究了大數據在旅遊領域運用的世界趨勢,認為在 2006 年之前,大數據並沒有吸引旅遊研究者的興趣,直到 2015 年,該領域的研究才出現爆炸性的增長。

2 國內旅遊大數據主題、信息源與研究方法

利用文本分析中的主題分析方法對 123 篇旅遊大數據文章按主題進行分類,發現國內的旅遊大數據運用,主要集中在五個領域:遊客偏好、遊客預測、遊客流分析、遊客滿意度和旅遊營銷等。在國內旅遊大數據研究中,遊客偏好的研究論文高達 37 篇,佔比 30%;遊客預測 21 篇,佔比 17%;遊客流分析 20 篇,佔比 16%;遊客滿意度 11 篇,佔比 9%;旅遊營銷 11 篇,佔比 9%;其它 23 篇,佔比 19%。

可見,遊客偏好、遊客預測和遊客流分析是國內旅遊大數據研究的重點領域,這三個主題的發文量總體佔比高達 63%。

2.1 遊客偏好

由於遊客在網絡上對旅遊目的地、旅遊產品和服務等進行評價和反饋,很容易捕捉

到遊客的偏好內容。所以,採用網絡文本分析遊客偏好是一種有效途徑。吳寶清等(2015)對搜狗博客中 213 篇關於巴丹吉林沙漠的遊記進行了分析,發現沙漠旅遊者偏好自助旅遊的方式在沙漠中進行徒步、越野和探險等活動;高大的沙山、神秘的沙鳴、美麗的胡楊林和神奇的湖泊泉水等都是旅遊者最為偏好的自然景觀,藏傳佛教、蒙古族生產和生活文化所組成的沙漠景觀是旅遊者偏好的人文景觀,且旅遊者對自然景觀的偏好高於人文景觀。通過研究沙漠旅遊者偏好,可以為保護沙漠自然景觀,營造人文景觀提供參考意見。王新亮(2011)在百度中搜索了 102 篇關於周莊的遊記,分析發現周莊古鎮的遊客追求偏好呈現多樣化,不再局限於傳統的觀光旅遊,更加偏好在旅遊過程中的參與感;旅遊時間方面,遊客偏好“雙日遊”,遊客飲食偏好主要以萬三蹄、阿婆菜和阿婆茶等周莊特色菜為主,沈廳和張廳等景點是遊客偏好度最高的景點。研究遊客偏好,對景區旅遊發展具有重大意義,可以幫助景區提高服務質量,在未來的競爭中佔據優勢地位。

旅遊目的地選擇是遊客偏好的一種典型體現。蜜月遊可以體現出遊客的選擇偏好,吳恒和陳燕翎(2017)利用爬蟲工具收集了攜程旅行網上 2,650 篇蜜月遊的遊記,發現遊客選擇蜜月遊目的地時偏好優質的自然風光、富有特色的美食、豐富的娛樂活動、完善的服務設施和便捷的交通等因素。王菲等(2015)通過對一起遊、新浪博客和天涯社區等 10 個網站上 1,000 篇遊記的分析,發現遊客選擇蜜月遊目的地的偏好因素包含:人文景觀、自然風光、交通可達性、配

套設施、情感映象、飲食和住宿等。對蜜月遊目的地的遊客偏好分析,可以為目的地開發與管理提供非常有價值的建議。

社會文化背景差異可能引發遊客偏好差異,劉逸等(2017)在攜程旅行網、馬蜂窩和窮遊網抓取了 36,148 條中國遊客評論,並從 Tripadvisor 抓取了 2,350 條國際遊客的評論,對赴澳大利亞的中國遊客和國際遊客進行了分析,發現中國遊客更加偏好建築、動植物、海濱,以及觀光休閒類的活動要素,而國際遊客偏好藝術類和當地文化生活類活動要素。採用網上圖片進行旅遊大數據分析是一個新的趨勢。鄧寧等(2019)收集了中國港澳台、美國和英國等地遊客在 Flickr 網站上發布的 27,054 張與北京相關的圖片元數據,包括圖片的拍攝日期、經緯度、圖片的描述等 20 條信息指標,分析發現不同地區的遊客在文化藝術、人物和食物拍攝內容方面差異較大。在文化藝術方面,中國港澳台遊客偏好磁器、玉器等古代藝術珍品,英、美遊客則偏好神獸雕刻、彩繪和表演等文化藝術;在人物方面,中國港澳台遊客偏好自拍,英國遊客偏好拍攝天安門站崗士兵和服務人員,美國遊客則偏好拍攝居民生活場景和中國旅遊者;在食物方面,中國港澳台遊客偏好中國菜肴,英國遊客只展示了少量西餐圖片,美國遊客偏好燒烤、小吃和啤酒等。以上關於出境遊和入境遊的遊客偏好研究,一方面可以幫助旅行社合理規劃旅遊綫路,另一方面,可以幫助旅遊目的地根據遊客特色開發旅遊產品和服務。

遊客偏好的分析對旅遊產業發展至關重要,通過瞭解遊客偏好,能够幫助旅行社合理規劃旅遊綫路,有利於旅遊目的地調整

旅遊項目和服務,迎合遊客需求,使得旅遊目的地或者旅遊吸引物持續得到遊客的青睞。遊客偏好主要採用遊記和網絡評論等文本類和圖片類的 UGC 數據進行分析。分析方法主要是內容分析法和社會網絡分析法;37 篇中有 34 篇採用了內容分析法,3 篇採用了社會網絡分析法。

2.2 遊客流

旅遊流是連接旅遊客源地和目的地的紐帶,廣義的旅遊流包括旅遊客流、物質流、資金流、文化流和能量流等,狹義的旅遊流僅指旅遊客流。旅遊客流是指旅遊者因為相似的旅遊需求,發生的集體性空間移動現象(閆閃閃等,2017)。關於旅遊流的大數據研究主要是指旅遊客流這一狹義概念,而且研究重點主要集中在旅遊客流的時空分布特徵和網絡結構特徵兩方面。

2.2.1 遊客流時空分布特徵

由於旅遊客流在流動過程中,不僅是空間上發生位移,還包括時間上的變化,所以旅遊客流的時空變化對旅遊容量控制和交通安全等具有重要意義。一方面,旅遊流的時空分布特徵表現為節假日和一些特殊旅遊節日的旅遊流時空特徵。戢曉峰等(2019)採用雲南省交通流量數據檢測平台的大數據對節假日旅遊流時空分布特徵進行了研究,發現旅遊流在節假日具有明顯的聚集性特徵,並且長綫旅遊流的時間集中度高於短綫,自駕遊的旅遊流時間集中度高於團隊遊。丁鼎等(2018)利用中國電信南京分公司的 1,008,910 條遊客手機定位數據,對江寧區鄉村旅遊進行了分析,發現單峰型鄉村旅遊點遊客量呈現出節假日、雙休日、

工作日依次遞減；雙峰型鄉村旅遊點節假日和雙休日的遊客量差別不大，工作日客流較少；多峰型鄉村旅遊點節假日、雙休日和工作日客流量峰值無明顯區別。手機定位數據屬企業內部數據，在當前幾乎人人都有手機的時代，能夠較為全面的反應一個地區遊客量的真實情況。

旅遊客流時空特徵的另一方面表現為城市旅遊的時空特徵。閆閃閃等(2017)採用177,995條遊客在新浪微博簽到數據對洛陽市旅遊流進行了研究，發現洛陽市年內旅遊客流體現為“雙峰山嶺”特徵，雙峰分別對應牡丹文化節和國慶黃金周，春、夏、秋、冬四季客流量依次遞減，周末客流量大於周內，日內客流量主要集中於9-24時。王新越和曹嬋嬋(2019)採用攜程旅行網和馬蜂窩網站的1,891篇遊記數據，對青島旅遊流進行了研究，發現青島市旅遊流年內淡旺季明顯，4-10月旅遊客流較大，在我國法定節假日期間旅遊客流呈增長高峰，雙休日旅遊客流大於工作日。城市旅遊流的時空特徵對城市規劃和公共服務設施建設具有重要影響，掌握好城市旅遊流的時空特徵，有助於做好遊客接待服務，提升城市品牌形象。

2.2.2 遊客流空間網絡結構

空間網絡指的是空間單元及其關係的集合，包含節點、關係邊和流動物三個要素，節點指的是空間單元，關係邊指的是各個空間單元之間的連線，流動物可以是知識或信息類的虛體，也可以是人口或者資金之類的實體形式(周慧玲和許春曉,2016)。旅遊流空間網絡結構能夠揭示旅遊系統的空間規律，可以分析旅遊目的地之間的競爭關係或者互補關係(蔚海燕等,2018)。研究旅

遊流空間網絡結構，有利於政府對旅遊區域進行合理規劃，也有利於旅遊景區之間的相互合作等。周惠玲和許春曉(2016)從攜程旅行網、去哪兒、百度旅遊和途牛網收集了1,197篇遊記對湖南省旅遊流空間網絡結構進行了分析，發現旅遊流空間集聚性較高，鳳凰古城、張家界、天門山、岳麓山和橘子洲等旅遊流量大，聚集性突出；節點功能具有差異性，優勢子結構具有區域指向並形成跨越之勢。蔚海燕等(2018)從驢媽媽旅遊網篩選出迪士尼在上海開業前2,037條遊客信息和開業後2,610條遊客信息，對旅遊網絡結構進行了對比，發現存在明顯差異。迪士尼開業前，上海旅遊網絡以東方明珠和上海野生動物園為中心節點，迪士尼開業後，上海旅遊網絡中心變化為迪士尼、東方明珠和上海野生動物園，且迪士尼的開業使得上海旅遊網絡集中度增強，熱門景點客流量更多，其它景點則受到衝擊。遊客量空間網絡結構的研究能夠呈現出旅遊景點之間的關係，可以為景區聯合推出優惠套票或者其它服務提供有價值的參考。

旅遊客流網絡結構研究的數據主要是遊客的位置信息，比如新浪微博的簽到信息，攜程旅行網、去哪兒、馬蜂窩、途牛網和驢媽媽等旅遊網站上的遊記，這些信息都包含了時間、地點等有關遊客的流動情況。另外，地方的交通流量數據監測平台和電信公司根據收集信號的定位信息等，也是反映遊客流動的重要數據來源。

旅遊客流網絡結構主要是研究各個節點之間的遊客流動情況，所以研究者較多使用社會網絡分析的方法進行研究。常見的社會網絡分析方法有中心度分析、子群分

析、關聯性分析和 QAP 假設檢驗(蔚海燕等,2018)。除此之外,時間層次分析法也被用來進行旅遊流網絡結構分析。20 篇研究遊客流的論文中,12 篇主要採用了社會網絡分析方法,3 篇主要採用了時間層次分析方法。

2.3 遊客預測

精準的旅遊預測對於政府和企業及時制定相關政策具有重要意義,傳統的旅遊預測主要基於歷史數據,存在一定的滯後性(李曉炫等,2017)。網絡搜索信息在一定程度上代表了用戶的意願,可以作為提前預判用戶行為的參考依據。信息搜索是消費者在旅遊之前的一個重要環節,通常會通過網絡對目的地、交通、天氣和住宿等信息進行瞭解,而這些搜索信息都被記錄下來,形成了搜索詞的搜索指數(李曉炫等,2017)。利用搜索指數進行旅遊預測成為提高預測準確性的一個新途徑。

利用網絡搜索數據可以對景區、城市或者省級等行政區域進行遊客量的預測。李曉炫等(2017)利用百度指數對九寨溝景區遊客量進行了預測分析,認為基於網絡搜索的 CLSI-EMD-BP 預測效果優於時間序列、BP 神經網絡和網絡搜索的基準模型。孫燁等(2017)分別採用移動端百度指數和 PC 端百度指數對三清山景區日客流量進行了預測,發現移動端百度指數的預測效果更具解釋力。任樂和崔東佳(2014)利用百度指數和歷史數據構建了北京國內遊客量預測模型,發現對客流量預測的擬合優度高達 97.23%,月平均誤差僅為 5.08%。可見,通過網絡搜索數據進行旅遊客流研究,能夠解

決傳統預測方法的滯後性。但是用戶的搜索行為具有一定的複雜性,利用網絡搜索對遊客流量的預測需要注意“噪音”污染,所以在預測之前對數據進行降噪預處理是必要的(李曉炫等,2017)。遊客量的準確預測能夠幫助旅遊企業提前做好遊客接待工作,為遊客提供優質的旅遊產品與服務。

受網絡搜索數據可以預測旅遊客流量的啓發,研究者也開始探索使用網絡搜索數據來預測旅遊收入。張斌儒等(2015)採用百度指數對海南旅遊收入進行了預測,發現利用合成指數構建的模型比基準模型的預測精確度提高了 32.42%。旅遊收入的準確預測可以讓政府和企業提前預判收益情況,能夠推動政府積極制定相應的旅遊政策,促進旅遊發展。

基於網絡搜索的旅遊預測,其數據來源主要是百度指數,通常會根據研究問題確定關鍵詞,然後將所有關鍵詞的百度指數進行合成,最終形成合成指數。21 篇研究旅遊預測的論文中,有 14 篇採用了百度指數進行分析。旅遊預測的研究方法主要是根據研究者建立的模型進行回歸分析或者其它量化分析,最終得出預測結果。

2.4 旅遊營銷

採用網絡大數據進行旅遊營銷的研究,主要是以旅遊目的地形象的傳播為主。旅遊目的地形象自 20 世紀 70 年代以來,就成為旅遊研究中最熱門的主題之一,隨著研究的深入,其概念也不斷發展,經過從認知到情感的過渡,最終形成了複雜的系統(吳林芝等,2018)。互聯網的發展使得信息傳播更加方便,傳播速度也越來越快。網絡中旅

遊目的地的信息一方面向人們傳遞了目的地的形象,另一方面,遊客在旅遊之後將自己對目的地的旅遊印象記錄在網絡上,也參與了旅遊目的地形象的塑造。研究者對旅遊目的地形象傳播的研究主要側重於群體差異。肖亮和趙黎明(2009)選取台灣旅遊官方網站、台灣綜合旅遊網站、台灣旅行社網站、大陸綜合旅遊網站和大陸旅行社網站等5類共41個網站中關於台灣旅遊的數據進行了分析,發現台灣網站在營銷宣傳中側重於塑造台灣的自然生態、度假和鄉村形象,而大陸網站側重於塑造台灣的文化和城市形象。不同的營銷宣傳信息,會影響旅遊者的選擇,通過對網站宣傳信息的分析,可以幫助旅行社調整營銷宣傳策略,吸引更多的遊客前往旅遊目的地。吳林芝等(2018)採用官方微博、官網網站和遊客遊記等數據對南京旅遊形象進行了分析,共收集官方微博1,306條、官方網站文本124,832字、遊記120篇,發現三者所要傳遞的南京旅遊行為存在差異,官方微博重點宣傳鄉村風光形象,官方網站重點宣傳休閒度假,而網絡遊記中的遊客感知主要是城市景觀形象。黃沙等(2012)對武夷山和九鯉湖景區管理者進行了採訪,並採用驢媽媽網站上遊客評論,分析了武夷山和九鯉湖的景區形象,發現與景區管理者所要傳達的景區形象和遊客網上評論的景區形象不吻合,對其原因進行分析後,構建了網絡信息嵌入後的旅遊目的地形象雙向傳導模式。遊記代表了遊客的真實旅遊體驗,官方宣傳代表了政府和景區希望塑造和宣傳的旅遊形象,當兩者不相符合的時候,政府和旅遊企業需要考慮調整宣傳的內容,以便更好地吸引遊客。

旅遊形象傳播有利於提高旅遊目的地的知名度,是旅遊營銷的重要方式,可以促進旅遊發展,所以是研究者熱衷的研究主題之一。主要數據來源一方面是以攜程旅行網和驢媽媽旅遊網等專業的旅遊類網站,另一方面,也包含了政府旅遊部門和旅遊企業的旅遊宣傳內容。11篇文章中有5篇採用了以攜程旅行網為代表的專業旅遊網站的資料,4篇採用了政府部門發布的旅遊相關資料,2篇文章的資料來源包括以上兩種途徑。研究方法主要也是採取內容分析法,也有研究會增加訪談之類的傳統方法進行多源數據的研究。11篇文章中有7篇採用了內容分析法,2篇採用了案例分析法。

2.5 遊客滿意度

遊客滿意度雖然在旅遊學界是一個比較陳舊的研究話題,但是對於旅遊業界來說,確是非常現實的一個問題,也是旅遊業界必須關注的重點。王雨文等(2018)從攜程旅行網、百度旅遊、途牛網和Tripadvisor收集了湄洲島33,233字的在綫評論,通過網絡點評數據對湄洲島遊客滿意度進行了分析,發現遊客主要對風景因素和文化因素滿意度較高,對交通和價格滿意度較低,而環境和聲譽因素對提升遊客滿意度起著關鍵作用。研究旅遊目的地或者景區的遊客滿意度,可以幫助政府改善公共設施,幫助旅遊企業合理調整旅遊產品和服務的價格等。曹小曙和劉丹(2018)通過收集百度旅遊網關於中國大陸31個省(市、自治區)中334個地級市的16,458條旅遊交通評論,對我國城市旅遊交通滿意度進行了分析,發現在區域尺度上,遊客的旅遊交通滿意度沒有

顯著差異,但是在省域和市域尺度上則存在顯著差異;交通發展水平是影響市域旅遊交通滿意度的主要因素。旅遊必然涉及交通,交通滿意度的研究可以為政府進行交通規劃提供重要依據。劉向前等(2018)採用大眾點評網和攜程旅行網網上遊客對開封的美食評價數據,結合傳統問卷調研分析,研究開封美食夜市遊憩者的滿意度,發現價格、供應速度、供應準確性、文化氣息和洗手間等公共設施對外地遊客和本地居民的滿意度均有正向影響。美食是遊客在旅遊過程中追求的一個重要因素,美食滿意度的研究,對於旅遊企業改善服務質量、完善設施設備具有重要參考價值。

遊客滿意度的大數據信息來源主要是以大眾點評網和攜程旅行網為主的生活服務類和旅遊類網站,採用的研究方法也主要是內容分析法。11 篇文章中,有 9 篇採用了內容分析法,1 篇採用了網絡文本和問卷相結合的方法,1 篇採用了電信部門開發的輿情監測系統數據進行了分析。由於滿意度的研究在學界已經探討得非常多,在理論上,以後較難有新的突破,未來的研究可能主要還是以應用研究為主。

2.6 其 它

旅遊是一個交叉學科,既有人文社會科學的內容,也有自然科學的因素,旅遊大數據的研究範圍遠不止本文總結的幾個方面。旅遊過程中的社會認知(宋振春等,2018)、旅遊網絡社區的人際互動(張驍鳴和常璐,2019)、旅遊輿情危機(付業勤等,2014)、住宿業的空間格局演變(王娟等,2018)、節慶活動的評價(林仁狀和周永博,2019),以及

旅遊資源的分布情況(張坤等,2021)等都是旅遊大數據研究的範疇,而且在未來隨著大數據分析技術的發展,旅遊大數據研究的內涵將會不斷深化,外延也將不斷擴展。

3 結論與未來展望

3.1 結 論

本文主要分析了大數據在我國旅遊領域中的運用。研究發現如下:

國內採用大數據進行旅遊研究的時間雖然較短,但是研究文獻的增長量較快,說明大數據在旅遊研究中日益受到重視。以文本和圖片等 UGC 數據成為旅遊大數據的主要來源,百度指數的搜索數據、手機定位信息和政府交通監測平台的信息也是國內旅遊大數據研究的重要來源。在分析方法方面,研究者們也一定程度上達成了共識,比如文本數據的處理,幾乎都採用 ROST-CM 軟件進行內容分析。

目前旅遊大數據的運用主要集中在瞭解遊客偏好、遊客流、旅遊營銷、遊客滿意度,以及遊客預測等方面。其中遊客偏好、旅遊營銷和遊客滿意度等三個主題的旅遊大數據來源主要是攜程旅行網、去哪兒和馬蜂窩等旅遊類專業網站的評論或者遊記等文本信息,分析方法主要以內容分析法為主。遊客流和遊客預測主要是以百度指數和旅遊專業類網站 UGC 數據中的時間和地點等信息為主,採用的研究方法以社會網絡分析法和建模預測為主。

旅遊大數據分析能彌補傳統旅遊研究的部分缺陷,對旅遊發展具有重要價值,但旅遊大數據自身也有較大的缺陷。首先,由

於網絡 UGC 數據的易獲取性和低成本性 (Li, et al., 2018), 使得 UGC 數據成爲目前旅遊大數據研究的最主要來源, 但是以網絡數據爲主的旅遊大數據, 雖然能夠較好反映使用網絡的人群旅遊特徵, 但是對於較少使用網絡的人群缺乏代表性。其次, 數據越大並不一定是優勢, 也可能給研究帶來負面影響(曾忠祿, 2017), 大數據的低密度價值使得數據中可能包含了很多“噪音”污染, 干擾了分析的準確性。所以, 大數據分析需要根據具體問題和研究目的, 正確地選擇數據來源和分析範圍(曾忠祿, 2017)。

3.2 未來展望

大數據在旅遊中的運用雖然已經有很多經典的案例, 但是還有巨大的潛力有待挖掘。

從大數據在旅遊中運用的來源來看, UGC 數據比例最大。雖然網絡 UGC 數據方便獲取, 但是數據質量問題一直存在。因爲遊客可能會因爲某種原因給出虛假評論, 在綫文本數據是否可靠的問題無法避免 (Schuckert, et al., 2015)。未來可能需要進一步開發手機定位數據、WiFi 和藍牙等電子設備產生的數據, 以及遊客交易過程中產生的數據。這些數據有非常好的真實性和可靠性, 如果能夠開發利用, 將可能產生重大的價值。但需要注意的是, 在採用這些數據的時候, 需要保護好個人隱私, 要切記合法使用, 這是大數據開發和利用的前提。

從研究方面來看, 如何提高非結構化的數據處理能力, 是未來旅遊大數據研究的一個方向。比如文本處理中大量的使用高頻詞統計、詞雲圖等, 但是文字表達的邏輯關

係目前處理能力還有待提高。圖片分析的研究中, 主要可以識別圖片的內容、位置、時間等信息, 但是圖片所要表達的含義還沒有較好的大數據處理方法。非結構化的數據可能需要推動機器學習, 逐漸積累, 從而提升對非結構化數據的處理能力。

參 考 文 獻

- [1] Li, J., Xu, L., Tang, L., et al. (2018). Big data in tourism research: A literature review. *Tourism Management*, 68(10):301-323.
- [2] Song, J. A. (2017). Systematic literature review: Existing hospitality & tourism research on big data. UNLV Theses, Dissertations, Professional Papers, and Capstones.
- [3] Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and tourism online reviews: Recent trends and future directions. *Journal of Travel & Tourism Marketing*, 32(5), 608-621.
- [4] 曹小曙, 劉丹. 大數據視角下中國城市旅遊交通滿意度的空間分異特徵及影響因素. *熱帶地理*, 2018, 38(6):35-44.
- [5] 鄧甯, 劉耀芳, 牛宇, 等. 不同來源地旅遊者對北京目的地形象感知差異: 基於深度學習的 Flickr 圖片分析. *資源科學*, 2019, 41(3): 416-429.
- [6] 丁鼎, 葛軍蓮, 龍毅, 等. 基於運營商客流大數據的鄉村旅遊點類型劃分研究——以南京市江甯區為例. *南京師大學報(自然科學版)*, 2018, 41(3):116-121.
- [7] 付業勤, 鄭向敏, 張俊. 旅遊網絡輿情危機事件的時空分佈規律研究. *財經問題研究*, 2014(9):124-128, 封3.
- [8] 黃莎, 陳金華, 陳秋萍. 基於網絡資訊嵌入性的旅遊目的地形象傳導研究. *人文地理*, 2012(3):115-120.
- [9] 戢曉峰, 戈藝澄, 陳方. 基於公路交通流大數據的節假日旅遊流時空分異特徵——以雲南

- 省 2017 年 7 個節假日為例. 旅遊學刊, 2019, 34(6):37-47.
- [10] 李曉炫, 呂本富, 曾鵬誌, 等. 基於網絡搜索和 CLSI-EMD-BP 的旅遊客流量預測研究. 系統工程理論與實踐, 2017, 37(1):106-118.
- [11] 林仁狀, 周永博. 美食文化節事大數據監測與評價初探——以浙江省“詩畫浙江·百縣千碗”旅遊美食推廣節事為例. 美食研究, 2019, 36(4):23-28.
- [12] 劉向前, 梁留科, 元媛, 等. 大數據時代美食夜市遊憩者滿意度雙視角研究. 美食研究, 2018, 35(2):27-34.
- [13] 劉逸, 保繼剛, 陳凱琪. 中國赴澳大利亞遊客的情感特徵研究——基於大數據的文本分析. 旅遊學刊, 2017, 32(5):46-58.
- [14] 任樂, 崔東佳. 基於網絡搜索數據的國內旅遊客流量預測研究——以北京市國內旅遊客流量為例. 經濟問題探索, 2014(4):67-73.
- [15] 宋振春, 趙彩虹, 李旭東. 中國出境旅遊的社會認知研究——跨文化交流視角的網絡文本分析. 旅遊學刊, 2018, 33(3):75-88.
- [16] 孫燁, 張宏磊, 劉培學, 等. 基於旅遊者網絡關注度的旅遊景區日遊客量預測研究——以不同用戶端百度指數為例. 人文地理, 2017(3):158-166.
- [17] 王菲, 宋萌, 信心, 等. 基於網絡內容分析的中國市場蜜月旅遊目的地選擇研究. 北京大學學報(自然科學版), 2015, 51(3):503-510.
- [18] 王娟, 李亞娟, 呂麗, 等. 基於互聯網數據的住宿業競爭力及其空間格局——以武漢主城區高端酒店為例. 地理科學進展, 2018, 37(10):1405-1415.
- [19] 王新亮. 基於網絡遊記視角下古鎮周莊旅遊者消費偏好研究. 安徽農業科學, 2011, 39(12):7188-7190, 7192.
- [20] 王新越, 曹嬋嬋. 基於網絡遊記的青島市國內旅遊客源市場結構與旅遊流時空特徵分析. 地理科學, 2019, 39(12):1919-1928.
- [21] 王雨文, 駱培聰, 劉瑩楠, 等. 基於網絡點評的湄洲島遊客滿意度研究. 福建師範大學學報(自然科學版), 2018, 34(5):83-92.
- [22] 蔚海燕, 戴澤飢, 許鑫, 等. 上海迪士尼對上海旅遊流網絡的影響研究——基於驢媽媽遊客數位足跡的視角. 旅遊學刊, 2018, 33(4):33-45.
- [23] 吳寶清, 吳晉峰, 劉佳, 等. 基於網絡文本的巴丹吉林沙漠旅遊者偏好研究. 中國沙漠, 2015(4):201-207.
- [24] 吳恒, 陳燕翎. 基於 UGC 文本挖掘的遊客目的地選擇資訊研究——以攜程蜜月遊記為例. 情報科學, 2017, 35(1):101-105.
- [25] 吳林芝, 周春林, 黃子睿, 等. 網絡語境下官方目的地宣傳形象與遊客感知形象的差異——以南京市為例. 地域研究與開發, 2018, 37(3):90-94, 100.
- [26] 肖亮, 趙黎明. 互聯網傳播的台灣旅遊目的地形象——基於兩岸相關網站的內容分析. 旅遊學刊, 2009, 24(3):76-82.
- [27] 閔閃閃, 梁留科, 索誌輝, 等. 基於大數據的洛陽市旅遊流時空分佈特徵. 經濟地理, 2017, 37(8):216-224.
- [28] 曾忠祿. 大數據時代的企業資訊情報戰略. 情報雜誌, 2015, 34(3):5-8.
- [29] 曾忠祿. 大數據分析中的誤區——基於紐約市的成功實踐的案例研究. 情報雜誌, 2017, 36(4):93-97.
- [30] 曾忠祿. 大數據分析: 方向、方法與工具. 情報理論與實踐, 2017, 40(1):1-5.
- [31] 張斌儒, 黃先開, 劉樹林. 基於網絡搜索數據的旅遊收入預測——以海南省為例. 經濟問題探索, 2015(8):154-160.
- [32] 張坤, 蘇欣蕾, 蘇凱紅, 等. 基於 POI 大數據的京津冀旅遊資源空間分異研究. 地域研究與開發, 2021, 41(1):103-108, 114.
- [33] 張曉鳴, 常璐. 擬劇理論視角下的旅遊網絡社區人際互動研究——以豆瓣網“窮遊”社區為例. 旅遊學刊, 2019, 34(7):98-109.
- [34] 周慧玲, 許春曉. 基於遊記行程的湖南旅遊流空間網絡結構特徵. 經濟地理, 2016, 36(10):203-208.