Big Data Analytics in Tourism Industry

Weixuan Wang
Indiana University Bloomington
Bloomington, Indiana 47405
wangweix@indiana.edu

ABSTRACT

This study focused on how the tourism industry has been impacted by the development of the Internet and improvements in information and communication technologies. This study explored how big data are generated related the tourism industry and how big data analytic has influenced and can further affect tourism research.

1 INTRODUCTION

Information Communication Technologies (ICTs) have been transforming tourism business globally and revolutionizing the world of Tourism. It transforms tourism from a labor-intensive to an information-intensive industry[11]. Developments in search engines, capacity, and speed of networks have influenced the number of tourists around the world that use technologies for planning and experiencing their travels. In addition, ICTs enable travelers to access reliable and accurate information and make reservations faster, cheaper and more convenient than the traditional way [3].

The development of ICTs also enables Internet users to both create and distribute content (multimedia information), which is called user-generated content (UGC) or consumer-generated content (CGC)[3].Platforms for UGC or CGC such as blogs, virtual communities, wikis, social networks, collaborative tagging, and media sharing sites play an increasingly important role as information sources for tourists. Social networks like TripAdvisor, Instagram, Facebook, Yelp, and booking.com are essential for tourists for multiple reasons: for their preparation of trips (booking hotels), during a trip (choose restaurant) and sharing their experience (writing hotel or restaurant reviews). Millions of data records are produced daily in regards to tourism by tourists, businesses an public services [9]. These data can be distinguished from big data by its volume, velocity (the speed it is produced), variety (different formats), variability (diversity of sources) and volatility (different level of production) [8].

Big data has been attracting more and more attention from tourism business and tourism researchers alike [4]. Big data analytics which is the activities of the specification, capture, storage, access, and analysis of such data sets to make sense of its content, provide new opportunities and challenges for tourism practitioners and researchers to understand tourists' behavior [8]. This study explores how big data are generated in the tourism industry and used in tourism research, further explore the implication and influence of big data and big data analytics for future research.

2 BIG DATA IN THE TOURISM INDUSTRY

Most activists in tourism industry had been generating a huge amount of data for several years.booking a plane ticket, reserving a hotel room and renting a car all leaves a data trail [9]. These data could add up to more than hundred of terabytes or petabytes structured data in the conventional databases. Discussion of travel arrangement on travel community online, status and post on social media like facebook and twitter, compliments and compliant on review websites like TripAdvisor constructs as more challenging and live unstructured data that arrives at a much faster pace than a conventional database [1]. Tourism practitioners are trying to understand tourists' behavior by accepting and analyzing these big data [9].

Airline and hotel chains have been using their big data which is the large volume of structured information that has been produced internally. Airlines and hotels have employed this tool to analyze of hotels' prices. Moreover, airlines have optimized the details of planning for the crew and routing [2, 9]. The online sector of the tourism industry has also quickly adopted big data to improve internal decisions and understand customers [1]. The online sector of the industry include online travel agencies (like Kayak, Expedia), meta-search engines (like Google) and some information companies that distribute tourism information (TripAdvisor).For example, Amadeus has developed a program for special results and the ability to search for its customers and Kayak has developed a program for predicting the prices [9].

2.1 Use of Social Media data in the Tourism Industry

Tourists in the digital age often use a variety of tools to access information that the tourism industry or other users have provided [12]. A tourist produces high volume of data when they are searching for travel websites, reporting data on mobile applications, sharing traffic information in the cities, searching and posting on social media, taking and sharing photos, reporting experience on travel websites and social media [1, 9]. All these data that are produced constantly can demonstrate tourists' motivation, interests, and their planning patterns and so on.

Previous studies have demonstrated several different usage and formats of big data in the travel and tourism industry [13]. Social media is one of them that has a huge effect on the tourism industry. Social media includes blogs, review sites, media sharing, social networks, and wikis. The remarkable growth of these data sources has inspired new strategies to understand the socio-economic phenomenon in various fields [9]. Discussions on social media are considered as electronic word-of-mouth (eWOM) that has in some degree substituted tradition face-to-face word-of-mouth for information exchange of tourist experience [3]. According to a study

on travelers' counseling with social media for travel planning in the US in 2014, 44 percent in the age group 18-34 years old use information in social media before planning for travel [10].

Photo post on photographic sharing website also can also provide extensive information on the tourists. Previous studies have connected photos posted on Panoramio, Flickr, and Instagram [2, 8]. Because when a tourist post pictures on these websites, their photo is tagged with geographic locations and ordered chronologically. Therefore analyzing photos posted by tourists can provide a photo density map to better understand tourists' behaviors, and potentially provide opportunities to detect atypical tourists behavior and characterize communities behaviors. However, the study also has its own limitation because of the limitation of technology to better exploit the data [2]. Another study focused on the sequence of locations in shared geotagged photos by tourist to identify and recommend travel routes which helped the travel recommender system to generate personalized recommendation according to interests and time available [6].

2.2 Other Big Data in the Tourism Industry

Beside the use of social media content to analyze tourists behavior, previous studies by Statistics Netherlands has also proposed using other innovative ways to understand tourists behavior by using mobile phone [5]. First method is using log data collected by an app installed on mobile devices, which allowed researchers to tract accurate movements of a person or family. This app also can pop up different questions that be triggered by location or change of time, such as purpose of the journey, satisfaction and activities. This innovative design combined the traditional survey with log data from smartphone measurements produced a rich and valuable sets of data [5]. However, this kind of method may be hard to get willing participants, because of privacy concerns and also technical issues such as people may not know how to download and use such application.

Another project from Statistic Netherlands uses aggregated mobile phone meta-data based on call detailed records from 2012 to 2014. This study collaborated with two telecom providers. Call detailed records contained information of the date and time and location where a communication through mobile network is used. The study uses these information and roaming data to identify unique foreign tourists, was able to detect different groups of foreign tourists and what are their favorite touristic sites within Netherlands [5].The limitation of this research is also restricted because it requires collaboration with telecom providers and its privacy concerns. With the technology development and widespread of WI-FI, when tourists go to another country they may not need to have roaming service in their destination [5].

3 BIG DATA IN TOURISM RESEARCH

Although tourism scholar has recognized the importance of UGC data such as travel blogs, online reviews and social media post as a form of eWOM has a huge influence in creating destination image. Tourism scholar has also done content analysis on online reviews and travel blogs, but recognizing big data and using big data in tourism research is still limited [3, 11].

Most tourism research utilizing big data are still focusing on CGC or UGC, especially online reviews for a hotel. A recent study conducted by Guo, Barnes and Jia used data mining approach and linguistic analysis to extract meaning from 266,544 online reviews for 25,670 hotels. They mined their customer review data from TripAdvisor using a web crawler. Through their linguistic analysis of their data and cross-comparing with perceptual mapping of the hotels, they find 19 controllable dimensions that are important for hotels to manage their interactions with visitors (such as the price for value, check in and check out) [4].

Another study also focused on UGC and trying to find out determinants of hotel customer satisfaction by discriminating among customers by language group. This study collected 412,784 reviews on TripAdvisor for 10,149 hotels in China. They have found out that tourists speaking different languages (such as Chinese, English, German, French, Russian etc.) differs substantially in terms of their emphasis on various attributes of hotels, and forming different satisfaction rating for hotels [7].

Both of the two studies mentioned above were from tourism or hospitality journals, were conducted by tourism researchers. Another study from outside of tourism research cohort provided a different study using big data to understand tourist behavior. This study designed and evaluated a big data analytics method using geotagged photos shared by tourists on Flickr to support destination management organization in analyzing and predict tourist behavior patterns at destinations (for this study itis Melbourne, Australia). The study designed a geotagged photo analytic artifact with textual meta-data processing geographical data clustering, representative photo identification and time series data modeling. This study demonstrated how to analyze unstructured big data to enhance strategic decision making in tourism destinations, provided insight on how city tour can be designed to better reflect tourists' interests and enrich their travel experience [8].

4 CONCLUSION

This study has explored the literature of big data and its implication in the tourism industry. Both tourism practitioners and tourism researcher has recognized the influence of big data and big data sources for tourism development. Big data in the tourism industry are generated by tourists directly, compared to traditional data sets that are gathered from surveys. Therefore, big data presented us opportunities to better understand tourist behavior, their motivations, and interests. However, big data also poses challenges for tourism practitioner and tourism researchers.

Like these two studies from tourism and hospitality journals, they share similarities in terms of data collection methods. Tourism researchers have recognized the importance of user-generated data which was able to provide them the volume of data they need for better generalization. One limitation of this kind of tourism research is that they only focus on hotel reviews, but their method could extend to other tourism sectors such as attraction and event to evaluate or review dimensions of tourist satisfaction. Another limitation they have is that they are only focusing on the text-based data from review website. How to integrating and getting useful information from other unstructured data such as image, video, post on Facebook and Twitter is still challenging for tourism researchers.

However, studies outside of tourism domains can be helpful in helping tourism researchers to utilize other formats of big data to understand tourist behavior. Therefore, collaboration with other fields and utilizing unstructured big data, and big data analytics in relation to tourism are much needed for tourism research.

ACKNOWLEDGMENTS

REFERENCES

[1] Rajendra Akerkar. 2012. *Big Data & Tourism.* Technical Report. Technomathmatics Research Foundation.

[2] G. Chareyron, J. Da-Rugna, and T. Raimbault. 2014. Big data: A new challenge for tourism. In *2014 IEEE International Conference on Big Data (Big Data).* 5–7. https://doi.org/10.1109/BigData.2014.7004475

[3] Jin Chung and Dimitrios Buhalis. 2009. *Virtual travel community: bridging travellers and locals.* IGI Global, Chapter Chapter 8, 130–144.

[4] Yue Guo, Stuart J. Barnes, and Qiong Jia. 2017. Mining meaning from online ratings and reviews: Tourist satisfactionanalysis using latent dirichletallocation. *Tourism Management* 59, Supplement C (2017), 467 – 483. https://doi.org/10.1016/j.tourman.2016.09.009

[5] Nico Heerschap, Shirley Ortega, Alex Priem, and May Offermans. 2014. *Innovation of tourism statistics through the use of new big data sources.* Technical Report.

[6] Takeshi Kurashima, Tomoharu Iwata, Go Irie, and Ko Fujimura. 2013. Travel route recommendation using geotagged photos. *Knowledge and information systems* 37, 1 (2013), 37–60.

[7] Yong Liu, Thorsten Teichert, Matti Rossi, Hongxiu Li, and Feng Hu. 2017. Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management* 59, Supplement C (2017), 554 – 563. https://doi.org/10.1016/j.tourman.2016.08.012

[8] Shah Jahan Miah, Huy Quan Vu, John Gammack, and Michael McGrath. 2017. A Big Data Analytics Method for Tourist Behaviour Analysis. *Information & Management* 54, 6 (2017), 771 – 785. https://doi.org/10.1016/j.im.2016.11.011 Smart Tourism: Traveler, Business, and Organizational Perspectives.

[9] S. Shafiee and A. R. Ghatari. 2016. Big data in tourism industry. In *2016 10th International Conference on e-Commerce in Developing Countries: with focus on e-Tourism (ECDC).* 1–7. https://doi.org/10.1109/ECDC.2016.7492979

[10] Statistica. 2014. Travelers who consult social media when travel planning in the United States as of April 2014, by age group. (2014). https://www.statista.com/statistics/305150/travelers-using-social-media-for-travel-planning-by-age-us/ accessed 2017.

[11] N.L. Williams, A. Inversini, N. Ferdinand, and D. Buhalis. 2017. Destination eWOM: A macro and meso network approach? *Annals of Tourism Research* 64 (2017), 87–101. https://doi.org/10.1016/j.annals.2017.02.007 cited By 0.

[12] Zheng Xiang, Zvi Schwartz, John H. Gerdes, and Muzaffer Uysal. 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management* 44, Supplement C (2015), 120 – 130. https://doi.org/10.1016/j.ijhm.2014.10.013

[13] Karen L. Xie, Kevin Kam Fung So, and Wei Wang. 2017. Joint effects of management responses and online reviews on hotel financial performance: A data-analytics approach. *International Journal of Hospitality Management* 62, Supplement C (2017), 101 – 110. https://doi.org/10.1016/j.ijhm.2016.12.004

5 BIBTEX ISSUES

Warning–empty publisher in GJT14

Warning–empty address in GJT14

Warning–empty address in chung2009

Warning–empty institution in heerschap2014innovation

Warning–empty publisher in Shafiee16

Warning–empty address in Shafiee16

(There were 6 warnings)

6 ISSUES

6.1 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords