THINK INDIA JOURNAL

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

Big Data impacts in Tourism Research

Nitin Sharma

P.hD Scholar

Computer Science & Engineering

Sunrise University, Jaipur

nitin.sharma448@gmail.com

Dr. K. P. Yadav

Research Supervisor

Professor

Sunrise University, Alwar

kpyadavsangam@gmail.com

Abstract

Big data is one of the most in demand and generally use term to illustrate mounting increase and availability of data in modern age which is believed to be maintained or either it is also expected to grow in the nearby future. Talking about tourism research even at an early stage big data has shown very wondrous improvement in tourism research. This might be first attempt to present a capacious literature review on different types of big data tourism research. By data informant the big data related to the tourism comes into web search data, online booking data, web page visiting data, etc. Which carries different information, different data types address and different tourism related issues? For mentioned above each type a structural analysis is conducted from the perspective of research focuses, the characteristics of data and the techniques which are utilized in the analysis, major challenges and further directions. This survey simplifies a meticulous understanding of this sunrise research and offers a valuable vision into its future possibilities.

Keywords : Big Data, Tourism Management, Tourist Behavior, Literature Review

1.Introduction

Social media sites have become a very important part of our lives nowadays and somewhere it affects our ability to decide. Big data generated on the social media sites has fabricated countless opportunities for bringing more sagacity to decision makers. Few studies on big data analytics have demonstrated its support for tactical decision making. Beside, a

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

formal method for analyzing the big data generated on social media for decision support is yet to be developed, especially in tourism sector. The aim of this study is to design and valuate a 'big data analytics' method to support strategic decision making tourism destination management by using design science research approach. Everyday many types of data are constantly growing within social media (Such as Twitter, Facebook, Instagram, Line & weChat etc.) because of their large user base, and frequent uploads of digital photos and videos.

Big data has allowed companies to track study shopping patterns, recommendations, and the purchasing behaviour that are known to influence sales and it has provided an exceptional insight into customers decision making process. Big data has given more power to the agencies and the merchants involved in tourism that they can find ingenious ways to use a variety of data resources to connect with a particular visitor or the potential visitor at every stage of a trip and they can use these big data sources to enhance an timely understand the fastest growing visitors demographics.

We are confident that both consumers and tourism product providers will be benefited by using big data. It is clear that big data can provide better, targeted, and profitable products to consumers (Pries & Dunnigan,2015). Personalized marketing and targeted product designs are very powerful opportunities for both consumers and tourism product providers. For instance, Big data gets the information about the interest of the consumer from photos posted on social networks by the consumer or what he searches for. This provide a very useful information to the product providers about the type of product the user is looking for. It is up to the researchers how efficiently they are going to examine the digital footprint left by activities on internet and make the use of these data or whether these data will create a new research epitome that tangles new methodologies and will develop our hypothetical comprehension of the travel industry. Till date the online data source is mainly used in applied research, and by means of that advantage was taken of the large and often free-of-charge volumes of data that brings the understanding of the activities of the tourism travel industry and its patrons.

2. Theoretical Background

Business Intelligence and Big data are just about to unfold their full potential for the tourism domain. Big data can be a very useful

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

source of information for the tourism sector seems to be still ignored by the researchers. Same situation seems to happen with Business intelligence also. Big data is to be considered an incredible opportunity it can provide answers about people's behaviours and views (Rodolfo Baggio 2016). In the last years, we need to recognised customer-focused approach in order to increase the tourism. Big data have also started to be a source for Business Intelligence Activities. The tourism scholars have well understood the capability of BD to provide insights that are useful or should be used for enriching the BI practices for destinations and operators. This note presents the results of an analysis of the recent literature on BD and BI and the application of the related technique to the fields of travel and tourism.

Smart tourism is become more and more important for the local government to raise the economy. Nowadays, many researchers draw much attention to the smart tourism. By the use of new technologies we can extract useful information from social media contents and then it help people to make more efficient and economic travel decisions. We use urban tourism check algorithm in which can find out the visitors who come from other country using social media check-in-data. We choose the Shanghai as a research target to analyze the data and the tourism place are Renmin Square, Yu Garden, The Bund and Chenhuang Temple in 2016 from Sino Weibo.We divide the tourists in two parts, one is local tourists and the other is outside tourists (Kai Yang, Wanggen Wan 2016).In order to make sure the accuracy of algorithm, we use the PLSA algorithm to analyze the Web text content. Through this analysis of outsider tourists we can provide a reference to the government which can use the tourism to rise in urban economy.

Social websites, smart phones and other equipments of customers including PCs and laptops have allowed billions of people across the world to participate in data production and nowadays, data has been increased rapidly. Analysis of large sets of data that is referred to as big data. Big data helps in analyzing the traveler trends by collecting information from different consumer centers and developing a specific marketing strategy for the target audience or customers. Definitely companies such as eBay, Google, Facebook, and Linkedln have been established on big data since the beginning. Because users produce high amount of data in different formats and they have to adopt new methodologies and technologies (Sanaz Shafiee 2016). Big data can be regarded

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

as an innovation source for tourism organizations and tourism industry.

Aditya patel, Lalit Nagla 2014, purposes a model for doing customer analytics on social media using big data for improving target advertising and improved business decision making. Using social media networking, they influence others and reveal their buying interests in products and services. Various research papers, case studies and industry white papers recently published, demonstrate the use of big data related technologies and models to harness the large amount of data for strategic business advantage. Organizations capture as much as information as they can about customers and analyze it effectively to discover patterns, trends and other vital clues. Social media marketing is very powerful medium for customer engagement, collaboration and active communication.

A paper presents a knowledge infrastructure which has recently been implemented as a genuine novelty at the leading Swedish mountain tourism destination. They proposed a Destination Management Information System are derives knowledge creation and application as a precondition for organizational learning at tourism destination (Matthias Fuchs 2016).They addressed both the generation of customer-based knowledge within a tourism destination as well as the BI based supplier- oriented knowledge application to support supplier's decision making. The knowledge generation layer, though methods of information gathering, extraction and storage, makes knowledge accessible to stakeholders. The knowledge application layer offers e-services that inform about supply elements and tourist's activities. The knowledge level of tourism stakeholders can be significantly increased. The agenda considers, for instance, the application of real- time Business Intelligence to gain real-time knowledge on tourists' on-site behaviour at tourism destination.

Jingjing Li present a comprehensive literature review on the application of big data to tourism research. In such a big data era, a variety of data is used in various fields. Big data have been generated from three primary sources-users, devices and operations. Among them, UGC data were dominant with 47% after that the device data have 36% and in contrast the transaction (operations) data was 17%. Web search data have improved tourism demand prediction and online marketing. Based on this big data sources, tourist behaviour and tourism market can be better explored and understood by both academia and industries. This paper

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

attempts to fill such a literature gap to present a comprehensive literature review on different types of big data in tourism research and provide a systematically analysis.

Xiaolu Zhou, effectively detected and ranked popular tourism destinations in multiple cities and also leveraged cloud computing to expedite Flickr tag processing and similarity graph preparation. This study provides a method for extracting tourist spots using Flickr images and a big data approach. The number of geo-tagged digital photos has grown exponentially. Increasing numbers on digital photos with geo-tags are available on many photo-sharing websites such as Flickr and Instagram. Flickr, the most popular, hosts over 5 billion photos. We have great opportunities to study people's travel experiences. The temporal and geographic features of the photos are very useful in separating the tourist-related photos from event-related photos. This study also demonstrates the application of using RHadoop cloud computing infrastructure to crunch big geospatial data. Our result identifies the most famous tourists spots.

Slava Kisilevich, Daniel Keim , present a GIS based decision support system that can both, estimate objective hotel room rates using hotel locational characteristics and predict temporal room rates. This paper makes three major contributions. First, we present a GIS based decision support system like for hotel brokers. Second the DSS can be applied virtually any part which makes it attractive business tool. Third, it integrates data mining framework and it offers the possibility of adding new algorithms. The system has been designed and evaluated in close cooperation with a company that develops travel technology solutions. Company has also provided real large datasets to evaluate the system.

3. Big data sources:

3.1. Online textual data:

In today's scenario, the thrive in technology and social media has fiercely changed the way the travelling industry was by providing a vast platform to share user generated data. And these data has been used to promote tourism research, including two types: one is the online textual data like the reviews given by the users of a particular product and by means of blogs on social media sites and another is through the online shared photos on photo sharing sites.

Due to the unexpected growth of the internet in the last few decades' social media offers a copious platform for tourists to spread a variety of information related to tourism, such as travelling reviews and experiences.

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

The user can generate a big mine of online reviews data by expressing their likes and dislikes toward tourism products. They can also share their few point and experiences in blogs or websites such as Facebook, Twitter and Instagram. Which provide valuable information to the required user. And this online reviews data, blogs data in the textual style compose special type of big data in tourism research online textual data transmitting emotions and opinion of tourists.

The very first process is to collect the online textual data (including reviews related to tourism, blogs and online shared photos) from the social media sites by using web crawling technology. For instance, Xiang pt al.(2017) used the web crawlers, by the programming languages like python and java, to acquire hotel related reviews. Guo et al.(2017) developed a web crawler to assemble review data from trip advisor periodically.

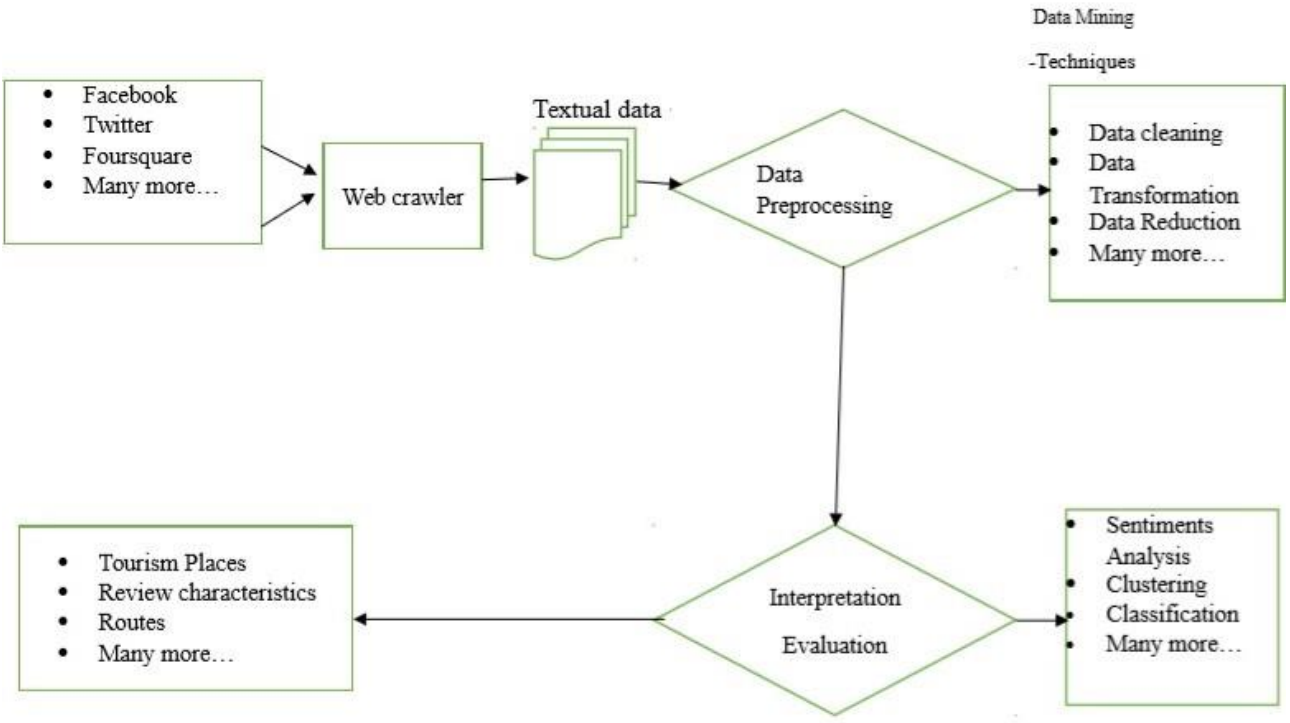

Figure: Model for Processing Textual Data

In the second step the online textual data that we have collected is evaluated to excerpt the useful knowledge for tourism research, by means of two sub-stages: first is the data pre-processing and another is interpretation evaluation.

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

Data Cleaning

a) Fill the missing values (attribute or column value)
b) Ignore the tuple: normally done when class label is missing.
c) To fill the missing values use the attribute mean.
d) For filling all the samples that belong to the same class use the attribute mean (or majority nominal value).
e) Remain missing values can be predicted by using a learning algorithm: consider the attribute with the missing value as a dependent (class) variable and run a learning algorithm (Bayes or decision tree) to predict the missing values.
f) Identify Outliers and smooth out noisy data:
    - Binning: Sort the attribute value and split them into bins then smooth by bin median, bin means, bin boundaries.
    - Clustering: Group the values in cluster and then identify and eliminate outliers.
    - Regression: smooth the data by fitting the data into the regression function.
g) Correct inconsistent data: use domain knowledge or expert decision.

Data Transformation:

a) **Normalization:** Scaling attribute values to fall within a specified range.
b) **Aggregation:** Moving up in the concept hierarchy on numeric attributes.
c) **Generalization:** Moving up in the concept hierarchy on nominal attributes.
d) **Attribute construction:** Replacing or adding new attributes inferred by existing attributes.

Data Reduction:

a) Reducing the number of attributes
b) Data cube aggregation: applying roll-up, slice or dice operation.
c) Removing irrelevant attributes: attributes selection (filtering and wrapper methods), searching the attribute space.
d) Principle component analysis (numeric attributes only): searching for a lower dimensional space that can be best represent the data.
e) Reducing the number of attribute values & tuples :
    - Binning: Reducing the number of attributes by grouping them into intervals.
    - Clustering: grouping values in clusters.
    - Aggregation or Generalization

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

- Sampling

3.2. Online Photo Data

Various data are posted and spread on social media sites that come in the category of user generated data. Like textual data and especially photos uploaded by the tourists. The user generated data contains very much information in relation to the users, location and time, providing a new aspect to study the behaviour of the tourist, tourism recommendation, and tourism marketing. Therefore this online data mainly the photos uploaded by the users (tourists) has aroused an increasingly large attention in tourism research. Online photo data have also been used in promoting tourism marketing. Online photo data have become a very efficient medium to form tourism destination for potential visitor. For example intended a machine learning based model to select the photo elements from the viewers perspective and assist destination for photo selection.

The Big data related to the photo in the tourism research mainly comes from various social media sites and photo sharing websites or platform: Flickr, Instagram, pinterest etc. Among these platforms, Flickr is the most dominant source of photo data used in tourism research. Various useful APIs are provided by this platform, for the easy access of photo data and the embedded metadata.

An online photo contains a very useful message in terms of metadata. Because of the heterogeneous information embedded in photos. There are three primary categories for metadata in tourism research.

- **User Related Information**
  - Photo ID: ID of the photo downloaded from websites.
  - User ID: ID of the tourist uploading the photo.

- **Temporal Information:**
  - Taken Date: The date and time to take the photo.
  - Uploaded Date: The date and time to upload the photo.
- **Textual Information:**
  - Textual metadata: Tourist defined textual information like title, description and tags of the photo.

4. Major Challenges & Solutions:

In tourism research varieties of big data generated by the users have been applied, making ominous employments and innovations. These big data originated from

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

user end. Which carries different information, different data type addresses, different tourism issues? The user generated data remain dominant in tourism research which gives the information about tourist sentiment analysis, behaviour of the tourist analysis, marketing and recommendation related to tourism.

Online textual data enforced in tourism research is of two types, i.e., Reviews by the user which can be called as reviews data and the another one is blogs data. Reviews data and blogs data that carries peculiar information have their own respective research focuses. Tourists attitude towards the tourism products can be known by reviews data and it can also be applied for measuring tourist satisfaction. This topic includes exploring the attributes of tourist satisfaction in the connection between tourist satisfaction and other related points such as guest experience and its position in the competition.

Tourism research adopting online photo data can be enhanced by flourishing research areas and perfecting analytic approach. As far as research areas are concerned, discovering tourism spots (i.e., where to travel) from the endemic dimension of online photos has been studied very abundantly. On the subject of analytic techniques to investigate about tourism related data in the present tourism research clustering and sequencing analysis have been used popularly. For estimation, classification, prediction, homogeneity grouping and association rule new big data mining technique can also be introduced into a fresh research. Apart from metadata the photo data themselves contain a plenty of information, and powerful mining techniques that will work directly on the images are needed.

The increase in the users of the social media in the last few years has created a positive effect on tourism research, as it has made the researchers to pay more attention on applying online textual big data reviews and blogs to tourism research. But there are still many things left in this field to work on and develop new research. Therefore, for future researches big sample size is despicable. Lot of work has done on the user generated data i.e., in the field of knowing tourists sentiments (or opinions) towards tourism product or destination. But, still how to use this to tourism product design and tourism marketing in practice is somewhat lacking.

But some issues are still there like data reliability. For example, sometimes the user

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

give fake reviews about the destination or product for avoiding unnecessary troubles or for gaining bonus (Schuckert et at.,2015 B). This type of reviews needs to be studied very deeply in the nearby future. Due to the amorphous composite feature of tourism related online reviews and blog data, the processing analytic techniques were fooled with some tricky challenges.

Lots of thing can be done to collect the information about the interest of the tourist at a particular place and what attracts the most and the real time number of the tourist can be known through the checking particular place by getting the data from different sources like Facebook, Foursquare, Flicker etc. So that the tourist can get the idea about the gather of that particular place. And it will help in finding the foot-falls of the tourism places.

For helping the tourist to know about the following tourism plan or getting plan information about them. A Decision Support System (DSS) can be built. And the data can be collected through API provided by different sources like Facebook, Foursquare etc. (On the other hand the researches on the different types of big data is very much data characteristics dependent. For example, for analyzing tourist satisfaction, about tourism product or destination. Online textual data (conveying tourist sentiments) is very much helpful. On the other side web search data shows the public attention towards tourism markets, parallel to inherent demands have improved tourism demands prediction and online marketing. It is important to see that even some type of big data has been used in the same issue; each of them has its own peculiar analysis prospective.

5. Conclusion:

Instead of prodigious improvement and innovation, there are still abundant places left to develop the tourism research using big data, mainly from the viewpoint of exploring research area and developing analytic techniques. To embellish the tourism research using big data, some more valuable big data and captivating tourism issues should be considered. Big data relatively makes very small contribution to tourism research, due to some tricky issues like high cost and privacy concerns. Therefore, the application of these big data to tourism research can be predominantly improved by intensifying the corporation and tourism industries. However, the data contains very useful information about the things that can be utilized to provide a new context of understanding tourist behaviour,

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

tourism management and tourism market. For instance, with the progress in smart tourism, most of the tourism places are equipped with monitoring system, which generates a large volume of video data. And this data is analyzed for understanding the behaviour of the tourist that will support the tourism management. Beside this tangential data or cross-domain data such as health, insurance and education data are also very important for tourism research, because they can influence the tourist preference greatly. Till now, most of the studies are focused on single type of big data; it may be insufficient to capture characteristics of complex tourism system. Therefore, study on multi-data type is vigorously recommended for development of tourism research.

There are some common issues like tourism demand prediction, tourist sentiments analysis, tourist behavior analysis and tourism recommendation that are using big data. But there are some issues left to work on like tourism precaution, tourism online marketing, scene spots programming, tourism product design and tourism capacity estimation. Moreover, various analytic methods such as machine learning and deep learning are really promising for tourism research.

References:

[1] M. M. S Silva, & R. Schegg , " The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis" Tourism Management, 2015, PP. 130-141

[2] K.H.Pries & R. Dunnigan, "Big Data Analytics: A practical guide for managers" . CRC Press, 2015.

[3] N. Deng & Li "Feeling a destination through the **'right'** photos: A machine learning model for DMOs' photo selection" Tourism Management , 2017, PP. 267-278.

[4] M. Schuckert, X. Liu & R. Law "Hospitality and tourism online reviews: Recent trends and future directions" . Journal of Travel & Tourism Marketing, 2015, PP. 608-621.

[5] C. Artola, F. Pinto & D.P. García "Can internet searches forecast tourism inflows?" International Journal of Manpower, 2015, 103-116.

[6] K.Berezina ,A. Bilgihan, C. Cobanoglu, & F. Okumus "Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews" Journal of Hospitality Marketing & Management, 2016, 1-24.

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

[7] A. Gandomi & M. Haider "Beyond the hype: Big data concepts, methods, and analytics"International Journal of Information Management, 2017 , PP. 137-144.

[8] B.Fang , D. Kucukusta & R. Law "Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics" Tourism Management, 2016, PP. 498-506.

[7] Hu, Y. H., Chen, Y. L., & Chou, H. L "Opinion mining from online hotel reviewse A text summarization approach" Information Processing & Management, 2017, PP. 436-449.

[8] S. Park & J.L.Nicolau "A symmetric effects of online consumer review" Annals of Tourism Research, 2015, PP. 67-83.

[9] E.C.L.Yang, C. Khoo-Lattimore & C. Arcodia "A systematic literature review of risk and gender research in tourism" Tourism Management, 2015, PP. 89-100.

[10] K.L.Xie, Z. Zhang "The business value of online consumer reviews and management response to hotel performance" International Journal of Hospitality Management, 2014, PP. 1-12.

[11] Z. Xiang, Ma, Y., & W. Fan "A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism" Tourism Management, 2017, PP. 51-65.

[12] K.H.Lim "Recommending tours and places-of-interest based on user interests from geo-tagged photos" ACM SIGMOD on PhD Symposium 2015.

[13] M. Fuchs, W. Höpken & M. Lexhagen "Applying business intelligence for knowledge generation in tourism destinations – A case from Sweden" Tourism and leisure – Current issues and perspectives of development in research and business (pp. 161–174). Wiesbaden: Springer.

[14] W. Höpken, M. Fuchs & M, Lexhagen "The knowledge destination: Applying methods of business intelligence to tourism. In: J. Wang (Ed.) Encyclopedia of business analytics and optimization (pp. 307–321). Pennsylvania: IGI Global Publisher.

[15] Slava Kisilevich, Daniel Keim , Lior Rokach, A GIS-based decision support system for hotel room rate estimation and temporal price prediction: The hotel brokers' context , Journal of Decision Support System , Elsevier January 2013.

ISSN: 0971-1260

Vol-21-Issue-1-January-March-2018

[16] Matthias Fuchs , Wolfram Höpken, Maria Lexhagen, "Big data analytics for knowledge generation in tourism destinations – A case from Sweden" Journal of Destination Marketing & Management , Elsevier August 2014.

[17] R. Baggio & G. Del Chiappa "Real and virtual relationships in tourism digital ecosystems" Journal of Information Technology & Tourism, 2014 , PP. 3–19.

[18] T. Chekalina, M. Fuchs & M. Lexhagen "A value creation perspective on the customer-based brand equity model for tourism destinations" Journal of Tourism Research, 2014, PP. 7–23.

[19] D. Jannach, M. Zanker & M. Fuchs "Leveraging multi-criteria customer feedback for satisfaction analysis and improved recommendations" Journal of Information Technology and Tourism, 2014, PP. 119–149.

[20] P. F. Bangwayo-Skeete & R. W. Skeete "Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach" Tourism Management, 2015, 454-464.

[21] G. Bordogna, L. Frigerio, A. Cuzzocrea & G. Psaila "Clustering geo-tagged tweets for advanced big data analytics" IEEE international congress on big data, San Francisco, USA 2016.

[22] B. Fang, Ye, Q., D. Kucukusta, & Law, R."Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics" Tourism Management, 2016, PP. 498-506.

[23] A. Gandomi & M. Haider "Beyond the hype: Big data concepts, methods, and analytics" International Journal of Information Management, 2016, PP. 137-144.

[24] Y. Guo, S.J.Barnes & Q. Jia "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichl et al location" Tourism Management, 2017, PP. 467-483.

[25] S.S.Huang & G. Chen "Current state of tourism research in China" Tourism Management Perspectives, 2016, PP. 10-18.

[26]Y.H. Hu, Y.L. Chen & H.L.Chou "Opinion mining from online hotel reviewseA text summarization approach" Information Processing & Management, 2017, PP. 436-449.

[27] M. Duggan, N.B.Ellison, C. Lampe A. Lenhart and M. Madden "Social Media Update 2016," Pew Research Center, Jan. 2017.