Book Chapter for: Management Science in Hospitality and Tourism: Theory, Practice and Applications

Monitoring and Forecasting Tourist Activities with Big Data

Bing Pan, Ph.D.

Associate Professor

Department of Hospitality and Tourism Management

School of Business

College of Charleston, Charleston, SC 29424-001, USA

Telephone: 1-843-953-2025

Fax: 1-843-953-5697

E-mail: bingpan@gmail.com

Yang Yang, Ph.D.

Assistant Professor

School of Tourism and Hospitality Management

Temple University

Telephone: 1-215-204-5030

Fax: 1-215-204-8705

E-mail: yangy@temple.edu

January 26, 2015

Bing Pan, PhD, is associate professor in the Department of Hospitality and Tourism Management at the College of Charleston in Charleston, South Carolina. His research interests include information technologies in tourism, destination marketing, search engine marketing, forecasting, and big data analytics.

Yang Yang, PhD, is assistant professor in the School of Tourism and Hospitality Management at Temple University. His areas of research interest include tourism demand analysis, tourist behavior, and location analysis in the hospitality and tourism industry.

Abstract

Driven by Moore's Law of technology advancement, various electronic devices that are used by tourists can provide large amounts of "big data", which can be captured and processed to monitor and predict tourist activities. This chapter provides a conceptual framework that connects the types of big data with stages of travel. This study reviews literature on the use of big data sources in the tourism industry, including data gathered from search queries, Web analytics, customer reviews, location tracking data, and social media. Most existing studies have focused on building behavioral models and validating correlational relationships between travel behavior and big data. Research on personalization, optimization, and resource allocation is lacking, and studies involving forecasting with big data for specific properties or businesses are also rare. Nonetheless, the combination of multiple data sources possesses a huge potential to dramatically improve the accuracy of forecasting and monitoring. Privacy concerns and business boundaries may limit the widespread adoption, application, and sharing of big data, but, as the related technology matures and big data productivity increases, its full impact and significance for the tourism industry will emerge.

Introduction

The tourism industry, by nature, is information-intensive (Poon, 1993): the variety of services it involves, the intangible and perishable nature of its many products, and the inseparable relationship between its production and consumption require the generation, storage, co-ordination, and analysis of information (Nyheim, McFadden, & Connolly, 2004). This characteristic indicates that the tourism industry can benefit greatly from the fast evolution of information technology (IT). Indeed, many adoptions of IT, from Property Management Systems (PMS) and Restaurant Management Systems (RMS), which are designed for optimizing production and increasing efficiency, to the Internet revolution, which disrupted many industries related to tourism, have demonstrated the co-evolution between the tourism industry and IT. In particular, so-called “big data” has become the latest manifestation of this co-evolution and will create more opportunities and challenges for the industry.

“Big data” refers to the large amount of IT data generated every day and that may be beyond the processing capabilities of traditional databases (Mayer-Schonberger & Cukier, 2013). This demands new ways of storing, retrieving, and analyzing the data. Moore’s Law dictates that our computer speed will double every 18 to 24 months (Schaller, 1997). As a result, our capability to capture, store, and process data will keep increasing exponentially, while the cost will keep decreasing. More importantly, the burst of big data symbolizes a paradigm change as industries can develop new business insights that do not come from the sampling and surveying of one’s customers, but from the aggregated digital footprints of their behavior (Mayer-Schonberger & Cukier, 2013).

Big data revolution also indicates that the causal relationship between particular data points and a business' revenue or profit is no longer the central focus; instead, it is the correlation that matters. With only a correlational relationship, one can use the data as benchmark measurements, forecasting performance, and optimizing business operations. Also, when merging a large amount of data from a wide variety of sources, one can gain insight on unexpected patterns that might not be otherwise disclosed by a limited number of conventional sources.

Recent scholars have used search engine traffic, website traffic, and social media content to monitor and predict tourist activities and sentiments. This chapter will review relevant studies in different fields. The authors will provide a conceptual framework on leveraging various sources of big data to monitor and forecast tourist activities and discuss the potential sources for big data forecasting.

Conceptual Framework

A tourist is a person on the move spatially; today's tourists will likely carry many technology gadgets with him or her and use them to interact with IT resources in the tourism industry. Thus, a tourist will generate and contribute a tremendous amount of data: for example, tourism website's analytics data, a hotel mobile app's log data, call center logs, the amount of foot traffic in the city, the sales records of travel services, search engine query volumes, social media mentions, location data from cell phones, GPS and photos, etc. All of these are potential indicators of a tourist's likes and dislikes, motivations, travel planning behavior, and actual travel and stay experience.

Combining these indicators together can make the data even more powerful and telling. Figure 1 shows a behavioral framework connecting the types of big data sources and a tourist's behavior. Different types of data will be available and useful in different stages of traveling: for example, tourists may perform searches before, en route to, or after arriving at a destination, while mobile positions are most useful in determining the location of the visitors while he or she is en route or at the destination. These assorted types of information are useful in monitoring and forecasting tourists' activities in different ways. The following section specifically discusses these information types, their usage, the results from past studies, and the potential for monitoring and forecasting tourist activities.

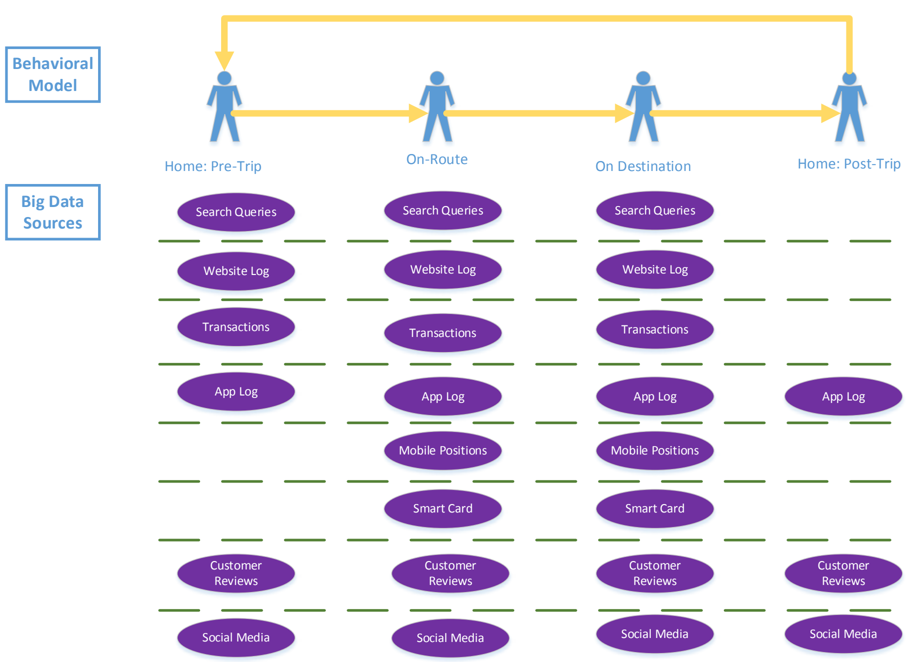

Figure 1. A Behavioral Model of Forecasting Tourist Behavior with Big Data

Types of Big Data for Monitoring, Understanding, and Forecasting Tourists' Behavior

Traditional forecasting methods usually hinge on historical data and a stable economic structure (Pan, Wu, & Song, 2012). Thus, dramatic change in economic structure may decrease the accuracy of these forecasting models. Big data has great potential in the short-term forecasting of dramatic and changing behavior. This section discusses different types of big data and their usefulness in monitoring and forecasting tourist activities.

Search Queries

Searching is the most popular activity on the Internet in the United States (Purcell, 2011). The queries typed in search engines reflect users' interests, informational needs, attitude, and feelings. Most tourists use search engines to look for information; thus, the traces of their search activities could be used to monitor and predict their travel behavior. For example, Pan, Litvin, and O'Donnell (2007) investigated information needs for accommodations, as reflected by search engine queries. Xiang and Pan (2011) studied how users search information for a destination city.

A few researchers have been using search engine queries for forecasting travel demand. Choi and Varian (2012) adopted the Google Trends index for Hong Kong in a time series method. Their model increased the forecasting accuracy for monthly volumes of visitors from the top nine origin countries to Hong Kong. Gawlik, Kabaria, and Kaur (2011) improved Choi and Varian (2012)'s algorithm by considering query-specific data, and they proposed a method for selecting relevant queries, leading to a significant improvement in forecasting accuracy. Similarly, Pan, Wu, and Song (2012) adopted search queries for a U.S. destination and improved the forecasting accuracy for local hotel occupancy rates based on the traditional time series method. Yang, Pan, Evans, and Lv (2015) further demonstrated that Baidu queries are more useful than Google in forecasting visitor volumes to a destination in China. Bangwayo-Skeete and Skeete (2015) used Google Trends data with an Autoregressive Mixed-Data Sample Method and helped increase the forecasting accuracy of five popular tourist destinations in the Caribbean.

Web Analytics Data

When visitors land on a website, their browsers communicate with the Web server continuously. The website owner can use page-tagging and web log analysis to track visitor behavior (Clifton, 2010).

Researchers have used Web traffic data to predict business revenues, including the revenue of Internet companies from 1998-2000 (Trueman, Wong, & Zhang, 2001). Lazer, Lev, and Livnat (2001) correlated Internet traffic data with portfolio returns of publicly-traded Internet companies. Their results showed that higher Web traffic for those companies correlated with higher returns. In the tourism field, Yang, Pan, and Song (2014) used a local Destination Marketing Organization's web traffic to forecast each local hospitality industry's average occupancy. The results highlighted the significant predictive power of DMOs' Web traffic data: these data provided a 7% to 10% increase in accuracy when forecasting hotel occupancy four to eight weeks in advance.

GPS Logs and Mobile Positioning

The understanding of the spatial-temporal pattern of tourist movement reveals vital insights for tourism infrastructure planning, tourist route design, and tourism capacity management (Shoval, Isaacson, & Chhetri, 2013). The widespread adoption of several spatial-temporal digital tracking technologies provides various types of big data to further understand this pattern of tourists at different scales (Shoval & Isaacson, 2007). Even though GPS data sets are fairly popular in studies with a small size of participation-based tourist sample (Shoval, et al., 2013), there are very few large data sets of GPS logs used for tourist tracking. Gang, et al. (2013) recognized the potential to utilize taxi GPS logs to study tourist movement by focusing on traces starting from and/or ending at tourist attractions.

Since modern tourists use mobile phones and smartphones at different stages of their travel (Wang, Xiang, & Fesenmaier, 2014), another type of big data—mobile positioning data—became greatly useful in highlighting hotspots of tourist activities and understanding tourists' data traces (Shoval, et al., 2013). Shoval and Isaacson (2007) compared different methods for tracking tourist movement, such as cell tower tracking, Assisted GPS (A-GPS), and Wi-Fi. Even though non-GPS mobile positioning data have been found to be less accurate than GPS log data (Shoval & Isaacson, 2007), due to the difficulty of data access caused by confidentiality concerns, mobile positioning data offer several notable advantages, such as lower data collection cost, a larger volume of data, functionality in indoor environments, and less sample selection bias due to the non-participatory nature of surveyors (Ahas, Aasa, Mark, Pae, & Kull, 2007).

Asakura and Iryo (2007) designed a route topology index based on mobile positioning in order to understand the topological characteristics of tourist behavior. A group of researchers from Estonia utilized a nationwide roaming mobile dataset of the Estonian GSM network to study the seasonality of tourism hotspots (Ahas, et al., 2007), destination loyalty of visitors (Tiru, Kuusik, Lamp, & Ahas, 2010), space-time flows of tourists (Ahas, Aasa, Roose, Mark, & Silm, 2008), market segmentation of repeat visitors (Vadi, et al., 2011), and travel distance of visitors (Nilbe, Ahas, & Silm, 2014). Based on a rich mobile positioning dataset , Tiru, Saluveer, Ahas, and Aasa (2010) also designed an online tourism monitoring tool for Estonia. Di Lorenzo, Reades, Calabrese, and Ratti (2012) extracted the information from people's past trajectory histories reflected in mobile positioning data, and predicted the location of a person over time.

Bluetooth and Infrared Tracking

In this digital age, with the popularity of Bluetooth-enabled devices (smartphones, laptops, tablets, and headsets), Bluetooth tracking technology has been used to further understand tourists' spatial-temporal movement patterns on a small scale (Versichele, et al., 2014; Versichele, Neutens, Delafontaine, & Van de Weghe, 2012). Versichele, et al. (2012) used this tracking technology to understand visitor movement at the Ghent Festivities, and Versichele, et al. (2014) demonstrated a visit pattern map by mining the big data of citywide Bluetooth tracking in Ghent, Belgium. An Alge-Timing system with infrared technology has also been introduced to monitor the movement of people within a park (O'Connor, Zerger, & Itami, 2005).

Customer Reviews

The Internet has become a major distribution channel for hotel sales (Connolly, Olsen, & Moore, 1998). The growing use of social media allows tourists to post their travel-related information and connect with others on a shared platform (Leung, Law, van Hoof, & Buhalis, 2013). Social media have been widely embraced by tourists through a large variety of social media websites, such as customer reviews, blogs/microblogs, online communities, and media sharing sites.

Electronic word of mouth (eWoM) about hotels is an important source for hotel guests to alleviate information asymmetry disadvantage when making booking decisions, and this source of information is expected to be more convincing and reliable than other information they can obtain on gauging the quality of hotels (Öğüt & Onur Taş, 2012). Sparks and Browning (2011) found that a high level of perceived trust in online reviews is associated with positively framed information and with numerical ratings that focus on interpersonal services. In general, there are two research streams to leverage the big data of customer reviews: the causal model of customer reviews and performance, and data mining of reviews.

For the first stream of research, several studies employed econometric models to decipher the causal relationship between online customer reviews and hotel performance measures. Ye, Law, and Gu (2009) found that a high score in average customer rating from Ctrip.com boosts the sales of Chinese city hotels, whereas a high level of discrepancy in customer reviews (variance of rating) reduces sales. Öğüt and Onur Taş (2012) also discovered the positive relationship between customer rating from the online travel agency (OTA), Booking.com, and online sales of hotels in Paris and London. Andersson (2010) obtained consumer feedback information on Singapore's hotels across six attributes from HotelTravel.com, and an analysis revealed that a higher room price is associated with higher customer numeric ratings on 'standard of room', 'hotel facilities', and 'food and beverage'. Zhang, Ye, and Law (2011) show that, among four types of customer ratings from TripAdvisor.com, the ratings of 'room quality' and 'location' are significantly correlated with room price for hotels in New York. By using the hotel review data from Booking.com, Yacouel and Fleischer (2012) found that hotels with a higher average score from reviewers charge a price premium. Based on the review data from the online meta-booking engine trivago.com, Schamel (2012) also reached a similar finding: consumer rating is positively associated with room rate for both weekend and midweek hotel stays.

For the second stream of research, since the reviews posted online incorporate customers' opinions and attitudes subjectively expressed in natural language text, they are hard to summarize in a single or multiple numeric ratings. Data mining becomes a promising tool to better understand embedded tourists' experiences in an efficient and accurate way. To better analyze the attitudes of customers in their reviews, Pekar and Shiyan (2008) adapted the opinion-mining technique to extract patterns embedded in the customer reviews from Epinions.com. Ye, Zhang, and Law (2009) conducted opinion mining on traveler reviews from Yahoo! Travel. Li, Ye, and Law (2012) text-mined the traveler reviews from a third-party website, Daodao.com, in China, and found six categories of factors influencing customer satisfaction.

In addition, in their efforts to propose a ranking system for hotels based on numerical values, Ghose, Ipeirotis, and Li (2012) parsed customer reviews from Travelocity.com and TripAdvisor.com using text-mining techniques. Liu, Law, Rong, Li, and Hall (2013) used sentiment mining to impute the missing value in the traveler review dataset from TripAdvisor.com, and then they utilized association rule mining to investigate how satisfaction and expectations vary for customers with different trip modes. Capriello, Mason, Davis, and Crotts (2013) compared different methods for mining tourists' sentiment and found that manual content coding, corpus-based semantic methods, and stance-shift analysis provide robust and similar results. Li, Law, Vu, and Rong (2013) used another data-mining technique, the Choquet Integral, to look into the hotel selection preferences of inbound travelers to Hong Kong with customer review data from TripAdvisor. Brejla and Gilbert (2014) text-mined customer reviews from CruiseCritic.com, and recognized patterns of co-creation of cruise value. In addition, Johnson, Sieber, Magnien, and Ariwi (2011) demonstrated the use of automated Web harvesting in extracting review data from TravelReview to better monitor tourists' experience. Zhang, Ye, Song, and Liu (2013) also investigated the structure of customer satisfaction and dissatisfaction with cruise line services using the review data from CruiseCritic.com. Lastly, moving beyond data mining, Korfiatis and

Poulos (2013) designed a demographic recommender system using online reviews from Booking.com as inputs.

Other User-Generated Content

Large amounts of user-generated content (UGC) have become available through social media (Lu & Stepchenkova, 2014), and they provide valuable information to better understand tourists' behavior and experience (Tussyadiah & Fesenmaier, 2009), attitudes and preferences (Magnini, Crotts, & Zehrer, 2011), and public images of tourist destinations (Choi, Lehto, & Morrison, 2007). Akehurst (2009) argued that UGC is more credible and trustworthy than other conventional marketing communications. Sharda and Ponnada (2008) introduced a Blog Visualizer to present the most relevant and useful blogs for tour planning.

Quantitative content analysis has been frequently utilized to analyze UGC by keyword counting and text characteristic measuring (Carson, 2008; Wenger, 2008). Pan, MacLaurin, and Crotts (2007) employed semantic network analysis to understand Charleston, South Carolina's destination image from the UGC on travel blogs. Moreover, several studies introduced netnography and netblography as methods to use available UGC to decipher the interpretation of places, people, and situations by tourists (Hsu, Dehuang, & Woodside, 2009; Woodside, Cruickshank, & Dehuang, 2007). Kwok and Yu (2013) studied restaurant-related social media messages on Facebook and found that, compared to sales and marketing messages, conversational messages are endorsed by Facebook users. Stepchenkova and Zhan (2013) analyzed online user-generated photography, a particular type of UGC, to understand Peru's image as a tourist destination. Pang, et al. (2011) proposed a framework to summarize a tourist destination by mining different aspects of both textual and visual UGC on tourist destinations.

Recently, with the development of reliable and accessible smartphones with built-in GPS antennas, tourists are able to share their UGC on smartphones with high-precision geo-referenced data, such as geo-referenced Twitter sharing and geo-tagged photos. This type of data offers several advantages to track tourists' movement patterns. First, it provides additional data on tourists' travel histories and their profiles (Kádár & Gede, 2013). Second, the data alleviates the sample selection bias of surveyed tourists, which is inherently embedded in the conventional tourist survey (Girardin, Fiore, Ratti, & Blat, 2008). Girardin, et al. (2008) investigated the geo-referenced information of photos taken by tourists from the photo-sharing website Flickr and geo-visualized tourist hotspots and travel trajectories. To better understand visitors' travel patterns in nature protected areas, Orsi and Geneletti (2013) used the visitor flow information embedded in geotagged photographs to estimate a gravity model. Kádár (2014) validated the accuracy of Flickr geo-tagged photos by comparing them with tourism statistical data and found a high level of correlation between them. He argued that tourists are more likely to take multiple photos of complex urban or architectural structures. Vu, Li, Law, and Ye (2015) introduced a framework to understand tourist travel behavior using geo-tagged photos and proposed a Markov chain model for travel pattern mining. On a larger scale, Hawelka, et al. (2014) show the usefulness of geo-located Twitter data as a proxy for country-to-country tourist/visitor flows, and these data provide information that is similar to official international tourism statistics.

Transaction Data

Now, with the development of computer based electronic funds transfer systems, credit cards from different points of origin are widely accepted around the world. As a result, the credit card has become a popular travel companion, and tourists have achieved increased mobility around the world (Weaver, 2005). Morrison, Bose, and O'Leary (1999) retrieved transaction data from a credit card service's marketing database to understand the demographic, socioeconomic, and psychographic characteristics of cardholders who used their cards to engage in hotel transactions. Moreover, credit card transaction data can be geo-coded by the address of card terminals. More importantly, the financial value of transactions provides important information on visitors' expenditures. By using bank card transactions data, Sobolevsky, et al. (2014) investigated the mobility patterns of foreign visitors within Spain by network analysis and gravity models. They concluded that this type of data is particularly useful in understanding large-scale mobility.

After the introduction of computerized reservation systems (CRS) into the tourism and hospitality industry, transaction data from hotel reservations and bookings became important in forecasting hospitality demand and understanding the travel patterns of hotel guests (Sato, 2012). To better understand the pre-purchase comparison behavior of online customers, Chatterjee and Wang (2012) used online transaction data and clickstream data, and they examined the relationship between customers' comparison search dispersion and purchase probability for flights, rental cars, and hotels. Weaver (2008) pointed out the potential use of another type of big data: tourist reward points data. These huge data sets enable airlines, hotels, and casinos to get a more comprehensive picture of preferences and behaviors of their customers.

Different from other types of data sources, transaction data are the results of product purchases and could also be used to monitor and forecast other types of spending. For example, transaction data for airline tickets could be used to predict future hotel purchases and attraction attendance. ForwardKeys (ForwardKeys.com) is a company that mines global distribution system (GDS) transaction data, and a few hotel companies and destinations have adopted its products for analytical and forecasting purposes (FowardKeys, 2014).

App Logs

Large amounts of log data for mobile apps are available through smartphone application software, which runs in the background of mobile operating systems and transmits the records of user activities to the app server. The log data captured by the software, as an alternative to other automated data collection methods, provide detailed records of location, voice calls, SMS messages, data usage, and application usage (Bouwman, de Reuver, Heerschap, & Verkasalo, 2013; Hamka, Bouwman, de Reuver, & Kroesen, 2014).

Schaller, Harvey, and Elsweiler (2014) utilized the log data of an Android app, consisting of all user interactions and positional data from the app, to predict visits to a cultural event in Munich. Hamka, et al. (2014) conducted a psychographic and demographic segmentation of mobile users based on the log data from smartphone measurement software. Vlassenroot, Gillis, Bellens, and Gautama (2014) used several tracking applications installed on Android phones to mine the travel patterns of smartphone users. Passive trip logging keeps the log data from the GPS, the signals from nearby cell towers and Wi-Fi networks, and the data from the accelerometer. Heerschap, Ortega, Priem, and Offermans (2014) demonstrated another example of using app log data for tourism statistics in the Netherlands. The smartphone measurement software registers time and location every five minutes, and this allowed a heat map of travel behavior to be generated.

Smart Cards

Smart cards have been introduced for automated fare collection systems, such as those used to automate ticketing systems for public transportation (Yue, Lan, Yeh, & Li, 2014). More recently, smart card systems have been used to provide payment functions for various businesses like restaurants, grocery stores, and healthcare services. Hotel guests can also swipe smart cards for different activities within a resort. In the field of transportation research, big data from smart cards has been used to understand the transport flow patterns of city residents and predict their future spatial movement trajectories (Pelletier, Trépanier, & Morency, 2011). After conducting a geo-demographic analysis based on transit smart card data, Páez, Trépanier, and Morency (2011) highlighted potential business opportunities for many hospitality business establishments. Moreover, Li, van Heck, and Vervest (2006) demonstrated a method for dynamic pricing strategies based on smart card data from the travel industry. As a type of smart card, the destination card, which offers free/discounted admission to various activities and attractions within a destination, has been found to be particularly useful in understanding the intra-destination movements of tourists (Zoltan & McKercher, 2014).

Conclusions

In conclusion, many types of big data, from a variety of sources, are available to monitor and predict tourist behavior. Travelers generate different types of big data in various travel stages: searches and web visits prior to the trip, GPS locations and transaction during the trip, and social media mentions during and after the trip. The different lag structure determines their distinct utility values: many are useful for real-time tracking at the destination, while others are instrumental in forecasting future tourist activities. For example, search engine queries and website traffic have been used for forecasting purposes, while GPS data and social media content are used more often for real-time position and service quality monitoring.

Future Research

However, many limitations exist for these reviewed studies. For example, search engine queries and web traffic are useful in helping forecast tourist volumes and hotel occupancy for a destination. However, no studies hitherto have adopted these data to forecast the revenue or customer numbers for a specific hospitality or tourism business. Researchers have embraced mobile data for monitoring activities, but no studies have focused on predicting users' spatial behavior based on mobile data. It is the latter that will provide great potential for tourism industry management in order to reduce crowding and to strategically allocate resources. Bluetooth tracking can offer accurate data of visitors' whereabouts because of a unique ID for visitor identification. However, the investment in the required infrastructure will be prohibitive on a large scale, whereas GPS systems use only a few satellites to cover the entire surface of the earth. Very few studies on causal relationship between reviews and business performance have been conducted, so does forecasting the latter from the former. App log data are specific to an application and, thus, wide sharing is limited. Privacy concerns might also prohibit businesses sharing their data. However, individual businesses or organizations may be able to mine the application data to track visitors' behavior and predict their activities.

Thus, there are five future directions of the usage of big data analytics for hotel and tourism industry:

1. Understanding tourist behavior. For example, big data can provide insights on tourists' likes and dislikes; the way tourists plan their stay; the time when they start booking their hotel rooms; the service weaknesses that impact hotel occupancy and revenue;
2. Forecasting tourist activities and the future performance of tourism businesses. The likelihood that one will have an overbooked hotel; whether or not one needs to hire more hourly staff; the amount of increase in occupancy one can expect in the following weeks;
3. Personalizing of service and improving customer experience. Tourism businesses can target each individual guest's likes and dislikes and focus on a market size of one;
4. Optimizing business operations. The discovery of the key predictors of a hotel's occupancy and revenue; the design of marketing and operation strategy to improve performance in those key predictors.
5. Allocating resources and facilities at destinations. By understanding tourists' spatial-temporal movement patterns and their preferences, tourist destinations are better able to propose more specific planning strategies to satisfy tourist needs and, hence, to maximize potential revenue.

First, past studies have been focusing on the first and second directions on understanding and forecasting; to our knowledge, directly connecting big data with personalization, optimization, and resource allocation are still rare, if there is any. Future studies could focus on studying big data along with its direct applications in personalization, optimization, and e resource allocation. It calls for more experimental studies which directly test the insights of big data with changes of business operations. Second, almost all studies have focused on a single source of big data, either search engine queries or web traffic, the combination of different data sources possess great potential in increasing forecasting accuracy. Scholars should focus on more utilization different data sources on monitoring and forecasting. Third, the current studies have focused on the level of a destination, and research on an individual business or organization is still lacking, probably due to a lack of data. This review calls for more collaboration between individual businesses and researchers on fully taking advantages of big data analytics to increase their revenue and profit.

In general, research at the crossroads of the tourism industry and big data is still limited: most studies focus on identifying correlation and causal relationships. Forecasting with big data for specific properties or businesses is rare. Furthermore, the combination of multiple data sources possesses huge potential in dramatically raise the accuracy of forecasting and monitoring. Privacy concerns and the business boundaries may also limit the wide adoption and application and sharing of big data.

Many off-the-shelf tools for these purposes have emerged, including solutions from Datameer, Solace Systems, and Metric Insights, but these are general tools and still in an early stage of development and adoption. These platforms assist researchers in extracting data from diverse sources and analyzing them by using a dashboard-like user interface. Many commercial or free statistical and data mining software solutions, such as SAS, R, Python, and Oracle, provide additional tools, but they are not designed with simplified tourism analysis in mind, and they may be cumbersome for use for this purpose.

From a macro-level perspective, the research and applications of big data in the tourism industry are still at an early stage. Like the adoption and application of other information technologies and byproducts, we expect that big data will likely go through a preliminary phase in which it receives considerable attention and focus before the related technology starts to mature and its productivity starts to increase. Once this occurs, the utilization and application of big data will show its full impact and significance. Hoteliers and tourism professionals who moved quickly and were early adopters of big data will enjoy a competitive advantage. However, with the further development of data and tools, and more research on its application in the hospitality and tourism field, the use of big data will inevitably increase beyond these early adopters, and the market will eventually produce effective tools that are accessible throughout the industry. The evolution of data, tools, and our understanding of this phenomenon will converge, and real increase in productivity and applicable insight, as a result, will occur.

References

Ahas, R., Aasa, A., Mark, Ü., Pae, T., & Kull, A. (2007). Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management, 28*(3), 898-910.

Ahas, R., Aasa, A., Roose, A., Mark, Ü., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management, 29*(3), 469-486.

Akehurst, G. (2009). User generated content: The use of blogs for tourism organisations and tourism consumers. *Service Business, 3*(1), 51-61. English

Andersson, D. E. (2010). Hotel attributes and hedonic prices: An analysis of internet-based transactions in Singapore's market for hotel rooms. *The Annals of Regional Science, 44*(2), 229-240. English

Asakura, Y., & Iryo, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice, 41*(7), 684-690.

Askitas, N., & Zimmermann, K. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly, 55*(2), 107-120.

Bangwayo-Skeete, P. & Skeete, R.W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46: 454-464.

Baram-Tsabari, A., & Segev, E. (2011). Exploring new web-based tools to identify public interest in science. *Public Understanding of Science, 20*(1), 130-143.

Bouwman, H., de Reuver, M., Heerschap, N., & Verkasalo, H. (2013). Opportunities and problems with automated data collection via smartphones. *Mobile Media & Communication, 1*(1), 63-68.

Brejla, P., & Gilbert, D. (2014). An exploratory use of web content analysis to understand cruise tourism services. *International Journal of Tourism Research, 16*(2), 157-168.

Capriello, A., Mason, P. R., Davis, B., & Crotts, J. C. (2013). Farm tourism experiences in travel reviews: A cross-comparison of three alternative methods for data analysis. *Journal of Business Research, 66*(6), 778-785.

Carson, D. (2008). The `blogosphere' as a market research tool for tourism destinations: A case study of Australia's Northern Territory. *Journal of Vacation Marketing, 14*(2), 111-119.

Chatterjee, P., & Wang, Y. (2012). Online comparison shopping behavior of travel consumers. *Journal of Quality Assurance in Hospitality & Tourism, 13*(1), 1-23.

Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management, 28*(1), 118-129.

Clifton, B. (2010). *Advanced Web metrics with Google analytics*: Sybex.

Connolly, D. j., Olsen, M. D., & Moore, R. G. (1998). The internet as a distribution channel. *Cornell Hotel and Restaurant Administration Quarterly, 39*(4), 42-54.

Di Lorenzo, G., Reades, J., Calabrese, F., & Ratti, C. (2012). Predicting personal mobility with individual and group travel histories. *Environment and Planning B: Planning and Design, 39*(5), 838-857.

ForwardKeys. (2014). ForwardKeys - Applied Traveller Intelligence for tactical marketers - DMO, Hoteliers, Merchants, etc. Retrieved from: http://forwardkeys.com/revenue-management/article/quantified-occupancy-forecasts.html

Gang, P., Guande, Q., Zhaohui, W., Daqing, Z., & Shijian, L. (2013). Land-use classification using taxi GPS traces. *Intelligent Transportation Systems, IEEE Transactions on, 14*(1), 113-123.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science, 31*(3), 493-520.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012-1014.

Girardin, F., Fiore, F. D., Ratti, C., & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: A case study. *Journal of Location Based Services, 2*(1), 41-56.

Hamka, F., Bouwman, H., de Reuver, M., & Kroesen, M. (2014). Mobile customer segmentation based on smartphone measurement. *Telematics and Informatics, 31*(2), 220-227.

Hand, C., & Judge, G. (2012). Searching for the picture: forecasting UK cinema admissions using Google Trends data. *Applied Economics Letters, 19*(11), 1051-1055.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science, 41*(3), 260-271.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management, 33*(3), 464-472.

Heerschap, N., Ortega, S., Priem, A., & Offermans, M. (2014, 2014). *Innovation of tourism statistics through the use of new big data sources.* Paper presented at the Global Forum on Tourism Statistics, Prague.

Hsu, S.-Y., Dehuang, N., & Woodside, A. G. (2009). Storytelling research of consumers' self-reports of urban tourism experiences in China. *Journal of Business Research, 62*(12), 1223-1254.

Johnson, P. A., Sieber, R. E., Magnien, N., & Ariwi, J. (2011). Automated web harvesting to collect and analyse user-generated content for tourism. *Current Issues in Tourism, 15*(3), 293-299.

Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. *Tourism Geographies, 16*(1), 88-104.

Kádár, B., & Gede, M. (2013). Where do tourists go? Visualizing and analysing the spatial distribution of geotagged photography. *Cartographica: The International Journal for Geographic Information and Geovisualization, 48*(2), 78-88.

Korfiatis, N., & Poulos, M. (2013). Using online consumer reviews as a source for demographic recommendations: A case study using online travel reviews. *Expert Systems with Applications, 40*(14), 5507-5515.

Kwok, L., & Yu, B. (2013). Spreading social media messages on Facebook: An analysis of restaurant business-to-consumer communications. *Cornell Hospitality Quarterly, 54*(1), 84-94.

Lazer, R., Lev, B., & Livnat, J. (2001). Internet traffic and portfolio returns. *Financial Analysts Journal*, 57(3): 30-40.

Leung, D., Law, R., van Hoof, H., & Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing, 30*(1-2), 3-22.

Li, G., Law, R., Vu, H. Q., & Rong, J. (2013). Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet Integral. *Tourism Management, 36*(0), 321-330.

Li, H., Ye, Q., & Law, R. (2012). Determinants of customer satisfaction in the hotel industry: An application of online review analysis. *Asia Pacific Journal of Tourism Research, 18*(7), 784-802.

Li, T., van Heck, E., & Vervest, P. (2006). Dynamic Pricing Strategies for Yield Improvement with Smart Card Adoption in the Dutch Travel Industry. In M. Hitz, M. Sigala & J. Murphy (Eds.), *Information and Communication Technologies in Tourism 2006* (pp. 234-234): Springer Vienna.

Liu, S., Law, R., Rong, J., Li, G., & Hall, J. (2013). Analyzing changes in hotel customers' expectations by trip mode. *International Journal of Hospitality Management, 34*(0), 359-371.

Lu, W., & Stepchenkova, S. (2014). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. *Journal of Hospitality Marketing & Management*, null-null.

Magnini, V. P., Crotts, J. C., & Zehrer, A. (2011). Understanding customer delight: An application of travel blog analysis. *Journal of Travel Research, 50*(5), 535-545.

Mayer-Schonberger, V., & Cukier, K. (2013). *Big Data: A Revolution that Will Transform how We Live, Work, and Think*: Eamon Dolan/Houghton Mifflin Harcourt.

Morrison, A. M., Bose, G., & O'Leary, J. T. (1999). Can statistical modeling help with data mining. *Journal of Hospitality & Leisure Marketing, 6*(4), 91-110.

Nilbe, K., Ahas, R., & Silm, S. (2014). Evaluating the travel distances of events visitors and regular visitors using mobile positioning data: The case of Estonia. *Journal of Urban Technology, 21*(2), 91-107.

Nyheim, P. D., McFadden, F. M., & Connolly, D. J. (2004). *Technology strategies for the hospitality industry*: Prentice-Hall, Inc.

O'Connor, A., Zerger, A., & Itami, B. (2005). Geo-temporal tracking and analysis of tourist movement. *Mathematics and Computers in Simulation, 69*(1–2), 135-150.

Öğüt, H., & Onur Taş, B. K. (2012). The influence of internet customer reviews on the online sales and prices in hotel industry. *The Service Industries Journal, 32*(2), 197-214.

Orsi, F., & Geneletti, D. (2013). Using geotagged photographs and GIS analysis to estimate visitor flows in natural areas. *Journal for Nature Conservation, 21*(5), 359-368.

Páez, A., Trépanier, M., & Morency, C. (2011). Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards. *Transportation Research Part A: Policy and Practice, 45*(7), 640-652.

Pan, B., Litvin, S. W., & O'Donnell, T. E. (2007). Understanding accommodation search query formulation: The first step in putting 'heads in beds'. *Journal of Vacation Marketing, 13*(4), 371-381.

Pan, B., MacLaurin, T., & Crotts, J. C. (2007). Travel blogs and the implications for destination marketing. *Journal of Travel Research, 46*(1), 35-45.

Pan, B., Wu, D. C., & Song, H. (2012). Forecasting Hotel Room Demand Using Search Engine Data. *Journal of Hospitality and Tourism Technology, 3*(3), 196-210.

Pang, Y., Hao, Q., Yuan, Y., Hu, T., Cai, R., & Zhang, L. (2011). Summarizing tourist destinations by mining user-generated travelogues and photos. *Computer Vision and Image Understanding, 115*(3), 352-363.

Pekar, V., & Shiyan, O. (2008). Discovery of subjective evaluations of product features in hotel reviews. *Journal of Vacation Marketing, 14*(2), 145-155.

Pelletier, M.-P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies, 19*(4), 557-568.

Poon, A. (1993). *Tourism, Technology and Competitive Strategies*. Wallingford, CT: CAB International.

Purcell, K. (2011). Search and email still top the list of most popular online activities. *Pew Research Internet Project: Report*. Retrieved from http://www.pewinternet.org/2011/08/09/search-and-email-still-top-the-list-of-most-popular-online-activities/

Sato, A.-H. (2012). Patterns of regional travel behavior: An analysis of Japanese hotel reservation data. *International Review of Financial Analysis, 23*(0), 55-65.

Schaller, R., Harvey, M., & Elsweiler, D. (2014). Detecting Event Visits in Urban Areas via Smartphone GPS Data. In M. de Rijke, T. Kenter, A. de Vries, C. Zhai, F. de Jong, K. Radinsky & K. Hofmann (Eds.), *Advances in Information Retrieval* (Vol. 8416, pp. 681-686): Springer International Publishing.

Schaller, R. R. (1997). Moore's law: past, present and future. *IEEE spectrum, 34*(6), 52-59.

Schamel, G. (2012). Weekend vs. midweek stays: Modelling hotel room rates in a small market. *International Journal of Hospitality Management, 31*(4), 1113-1118.

Sharda, N., & Ponnada, M. (2008). Tourism Blog Visualizer for better tour planning. *Journal of Vacation Marketing, 14*(2), 157-167.

Shoval, N., & Isaacson, M. (2007). Tracking tourists in the digital age. *Annals of Tourism Research, 34*(1), 141-159.

Shoval, N., Isaacson, M., & Chhetri, P. (2013). GPS, Smartphones, and the Future of Tourism Research. *The Wiley Blackwell Companion to Tourism*, 251-261.

Sobolevsky, S., Sitko, I., Tachet des Combes, R., Hawelka, B., Murillo Arias, J., & Ratti, C. (2014, June 27 2014-July 2 2014). *Money on the Move: Big Data of Bank Card Transactions as the New Proxy for Human Mobility Patterns and Regional Delineation. The Case of Residents and Foreign Visitors in Spain.* Paper presented at the Big Data (BigData Congress), 2014 IEEE International Congress on.

Song, H., Witt, S. F., & Li, G. (2008). *The advanced econometrics of tourism demand*: Routledge.

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management, 32*(6), 1310-1323.

Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism Management, 36*(0), 590-601.

Tiru, M., Kuusik, A., Lamp, M.-L., & Ahas, R. (2010). LBS in marketing and tourism management: Measuring destination loyalty with mobile positioning data. *Journal of Location Based Services, 4*(2), 120-140.

Tiru, M., Saluveer, E., Ahas, R., & Aasa, A. (2010). The Positium Barometer: A web-based tool for monitoring the mobility of tourists. *Journal of Urban Technology, 17*(1), 71-89.

Tnooz. (2014). Google's 2014 travel study: App mania has ebbed, yet booking by smartphone is still hot. Retrieved from http://www.tnooz.com/article/google-research-travel-marketers-traveler-shopping-behavior/

Trueman, B., Wong, M. F., & Zhang, X.-J. (2001). Back to basics: Forecasting the revenues of Internet firms. *Review of Accounting Studies, 6*(2-3), 305-329.

Tussyadiah, I. P., & Fesenmaier, D. R. (2009). Mediating tourist experiences: Access to places via shared videos. *Annals of Tourism Research, 36*(1), 24-40.

Vadi, M., Vedina, R., Karma, K., Kuusik, A., Tiru, M., Ahas, R., et al. (2011). Innovation in destination marketing. *Baltic Journal of Management, 6*(3), 378-399.

Versichele, M., de Groote, L., Claeys Bouuaert, M., Neutens, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium. *Tourism Management, 44*(0), 67-81.

Versichele, M., Neutens, T., Delafontaine, M., & Van de Weghe, N. (2012). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography, 32*(2), 208-220.

Vlassenroot, S., Gillis, D., Bellens, R., & Gautama, S. (2014). The use of smartphone applications in the collection of travel behaviour data. *International Journal of Intelligent Transportation Systems Research*, 1-11. English

Vosen, S., & Schmidt, T. (2012). A monthly consumption indicator for Germany based on Internet search query data. *Applied Economics Letters, 19*(7), 683-687.

Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management, 46*(0), 222-232.

Wang, D., Xiang, Z., & Fesenmaier, D. R. (2014). Adapting to the mobile world: A model of smartphone use. *Annals of Tourism Research, 48*(0), 11-26.

Weaver, A. (2005). ‘Passports to pleasure’: Credit cards and contemporary travel. *International Journal of Tourism Research, 7*(3), 151-159.

Weaver, A. (2008). When tourists become data: Consumption, surveillance and commerce. *Current Issues in Tourism, 11*(1), 1-23.

Wenger, A. (2008). Analysis of travel bloggers' characteristics and their communication about Austria as a tourism destination. *Journal of Vacation Marketing, 14*(2), 169-176.

Woodside, A. G., Cruickshank, B. F., & Dehuang, N. (2007). Stories visitors tell about Italian cities as destination icons. *Tourism Management, 28*(1), 162-174.

Wu, L., & Brynjolfsson, E. (2009). The future of prediction: How Google searches foreshadow housing prices and sales. *Available at SSRN 2022293*.

Yacouel, N., & Fleischer, A. (2012). The role of cybermediaries in reputation building and price premiums in the online hotel market. *Journal of Travel Research, 51*(2), 219-226.

Yang, X., Pan, P., Evans, J.A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, 46: 386-397.

Yang, Y., Pan, B., & Song, H. (2014). Predicting Hotel Demand Using Destination Marketing Organization’s Web Traffic Data. *Journal of Travel Research, 53*(4), 433-447.

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management, 28*(1), 180-182.

Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications, 36*(3, Part 2), 6527-6535.

Yue, Y., Lan, T., Yeh, A. G. O., & Li, Q.-Q. (2014). Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behaviour and Society*, *1*(2), 69-78.

Zhang, Z., Ye, Q., & Law, R. (2011). Determinants of hotel room price: An exploration of travelers' hierarchy of accommodation needs. *International Journal of Contemporary Hospitality Management*, *23*(7), 972-981.

Zhang, Z., Ye, Q., Song, H., & Liu, T. (2013). The structure of customer satisfaction with cruise-line services: An empirical investigation based on online word of mouth. *Current Issues in Tourism*, 1-15.

Zoltan, J., & McKercher, B. (2014). Analysing intra-destination movements and activity participation of tourists through destination card consumption. *Tourism Geographies*, 1-17.