Predicting Tourist Demand Using Big Data

Haiyan Song and Han Liu

1 Introduction

Big data is one of the most popular and most frequently used terms to describe the exponential growth and availability of data in the modern age, which is likely to be maintained or even accelerate in the foreseeable future (Hassani & Silva, 2015). It is a broad term for datasets that are so large in size or complex that traditional data processing applications and software tools are inadequate to capture, curate, manage, and process the data within a reasonable period of time (Snijders, Matzat, & Reips, 2012). There are challenges regarding the analysis, capture, search, sharing, storage, transfer, visualization, and information privacy of big data, and these challenges require new technologies to uncover hidden values from large datasets that are diverse, complex, and massive in scale (Hashem et al., 2015). Big data brings new opportunities to modern society (Fan, Han, & Liu, 2014) since these vast new repositories of information can provide researchers, managers, and policymakers with the data-driven evidence needed to make decisions on the basis of numbers and analysis rather than anecdotes, guesswork, intuition, or past experience (Frederiksen, 2012), and it may lead to more accurate analysis, more confident decision-making, and greater operational efficiencies, cost reductions, and risk reductions (De Mauro, Greco, & Grimaldi, 2015).

Nowadays, people try to use the insights gained from big data to uncover new opportunities for their businesses (Mayer-Schönberger & Cukier, 2013). The process of discovering and determining insights from large, complex, and unstructured datasets attracted our attention. So, what is big data? There is no unified definition

H. Song (✉)
The Hong Kong Polytechnic University, Hung Hom, Hong Kong
e-mail: haiyan.song@polyu.edu.hk

H. Liu
Jilin University, Jilin, People's Republic of China

Z. Xiang, D.R. Fesenmaier (eds.), *Analytics in Smart Tourism Design*, Tourism on the Verge, DOI 10.1007/978-3-319-44263-1_2

of big data. The basic definition is "datasets which could not be captured, managed, and processed by general computers within an acceptable scope" (Chen, Mao, & Liu, 2014). More and more researchers and institutes are exploring the characteristics of big data in order to define it. These definitions always include the characteristics of volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources). Laney (2001), for example, used the above "3V's" model to define big data. In this model, volume means that with the generation and collection of masses of data, the scale of the data becomes increasingly big; velocity means that the collection and analysis of big data must be rapidly and timely conducted so as to maximally utilize its commercial value; and variety indicates the various types of data, which include semi-structured and unstructured data, such as audio, video, web page, and text data, as well as traditional structured data. Beyer and Laney (2012) updated the definition of big data by adding another "V": veracity. Chen, Mao, Zhang, and Leung (2014) added "value" (huge value but very low density) to make the definition perfect. Recently, a consensual definition was produced: "Big data represents the information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value" (De Mauro et al., 2015).

Big data is not simply defined by the 4V's: it is about complexity. Beyond the definition of big data, we should be concerned about the details of it. Hashem et al. (2015) classified big data into five categories: data sources, content format, data stores, data staging, and data processing. In each category, there are numerous subcategories, as shown in Fig. 1. In this chapter, we focus on tourism forecasting using big data, and we will therefore pay special attention to the data sources, data staging, and data processing categories.

2 What Is Tourism Big Data?

The tourism industry thrives on information (Benckendorff, Sheldon, & Fesenmaier, 2014; Poon, 1988). The vast new big data repositories of information—far greater than what is captured in standard databases—can provide researchers, managers, and policymakers with the data-driven evidence needed to make decisions on the basis of numbers and analysis rather than anecdotes, guesswork, intuition, or past experience (Frederiksen, 2012). The bounty of tourism big data has the potential to deliver new and more highly informed inferences about human activity and behavior that will give the tourism industry a big boost and benefit not only customers but also those who participate in the tourism industry (Fuchs, Höpken, & Lexhagen, 2014).

Travelers leave different digital traces behind on the Web when using mobile technologies. Through every traveler, large amounts of data are available about anything that is relevant to any travel stage: prior to, during, and after travel (Hendrik & Perdana, 2014). Most of this data is of an external nature: for example, in the form of Twitter or other social networking feeds. Due to the large amounts of

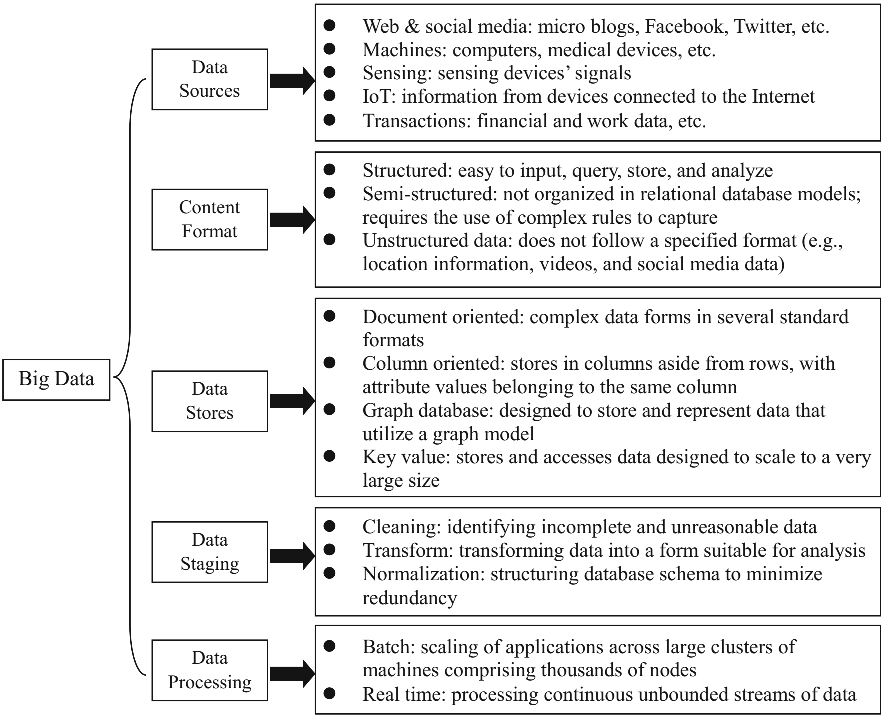

Fig. 1 Big data classification (Hashem et al., 2015)

available data stored in the cloud, analytics are needed in order to make sense of the information within the data. If you are a potential customer planning a trip, you probably get more than a little help from the Internet when you are searching for inspiration, buying tickets, reserving accommodation, or researching attractions. Participants in the tourism industry are increasingly turning to big data to discover new ways to improve decision-making, opportunities, and overall performance (Irudeen & Samaraweera, 2013): for example, big data can be used to interconnect the dispersed information from different systems and then improve decision-making capability.

Big data provides unprecedented insights into customers' decision-making processes by allowing companies to track and analyze shopping patterns, recommendations, purchasing behavior, and other drivers that are known to influence sales. Agencies and merchants involved in tourism can find innovative ways to use a variety of data resources to connect with potential visitors at every stage of a trip and use these big data sources to better and timely understand the fastest growing visitor demographics. They can also remarket to target shoppers who have looked at a specific destination on an online travel agency website (Sust et al., 2014). Through the use of big data, industries become more efficient. More and more companies have started specializing in the storage and evaluation of the large amounts of data on travelers' hotel stays, purchase transactions, and customer information in order to provide more efficient and high quality services.

3 Advantages of Using Big Data in Tourism

We are confident that consumers and tourism product providers will see the benefits of using big data. Personalized marketing and targeted product designs are extremely powerful opportunities for both groups. It is crystal clear that big data can provide better, targeted, and profitable services and products to consumers (Pries & Dunnigan, 2015). For instance, big data analysts can capture information of consumer interests from photos posted on Facebook or other social networks (e.g., a tourism provider could push information about local biking destinations or biking clubs when they obtain a picture of a mountain bike).

Previous studies on tourism have mostly been based on surveys or experts' views, which mean that they have taken samples from the population as a whole and do not have real data about all tourists. In contrast, one study on tourism big data tried to introduce data based on real actions by all users instead of drawing information from survey samples (Irudeen & Samaraweera, 2013). In this chapter, we introduce a framework that incorporates big data produced by tourists themselves (e.g., through mobile phones connecting to the telecom network or bank cards connecting to POS terminals) that increases knowledge of the industry's target market into tourism demand forecasting.

Tourism big data using innovative methods has advantages over traditional methodologies, as discussed below.

(1) Reliability
Big data are based on users' real actions, not on surveys. In other words, real actions have been analyzed rather than stated intentions or answers to questions. Taking all information sources together, it can be stated that big data increases the sample base on which conventional research tends to be based by several orders of magnitude (Meeker & Hong, 2014). The reliability of big data analysis allows us to consider all aspects of the information in order to provide comprehensive results instead of biased conclusions due to information loss in the sample data.

(2) New information flows
Tourism big data is a type of information produced by tourists themselves; it enriches the knowledge of tourism businesses' target market and is very useful for analyzing the consumers' demand for different tourism products and services (Hendrik & Perdana, 2014). Since tourism big data are structured and repositioned data, it is possible to cross-reference them with other sources such as social media and open public data, whether these are sources currently in production or potential information sources that may be created or released in the future. The analysis of tourism big data can be contrasted with internal data

from each tourism business with a view to determining whether the supply of tourism products/ services in each area of a city is in tune with the tourists who demand for these products and services.

(3) Real-time data and nowcasting
One of the innovative uses of big data is nowcasting, that is, the use of real-time data to describe contemporaneous activities before official data sources are made available (Bollier & Firestone, 2010): for example, Varian (2014) argued that real-time Google search queries are a good way to nowcast consumer activities, as the contemporaneous correlation analysis obtained from the Google Correlate data is still a 6-week lead on reported values. A notable example of using Google search queries for nowcasting is Google Flu Trends, which identifies possible flu outbreaks 1–2 weeks earlier than the official health reports by tracking the incidence of flu-related search terms in the Google search engine.

There are many studies that have used structured search-engine data for tourism nowcasting and forecasting (Artola, Pinto, & Pedraza, 2015; Bangwayo-Skeete & Skeete, 2015; Yang, Pan, Evans, & Lv, 2015). Besides search engine queries, there are other types of real-time data streams that can be assembled and analyzed: for example, data on credit card purchases, the trucking and shipping of packages, and mobile phone usage are all useful bodies of information. Much of these data is becoming available on a near real-time basis, which can be used to predict the macro data that will be compiled at some point in the future (Jeng & Fesenmaier, 2002; Yang, Pan, & Song, 2014).

The ultimate objective of using real-time big data is to develop applications that are able to respond as soon as the economic pulse has been taken and provide suggestions; of course, this should be done under controlled conditions and be capable of being switched on and off at any time.

4 Characteristics of Tourism Big Data

Having scoured the literature and found the 4V's characteristics of big data, we used these and added another V (value) to ascertain the unique characteristics of tourism big data.

(1) Volume
Volume always seems to top the list of big data characteristics, and is a key contributor to the problem of why traditional relational database management systems fail to handle big data (Prajapati, 2013). The volume of tourism big data always comes from points of sales or other traditional channels of distribution (i.e., call centers, websites, premises, newsletters, customer relations, etc.). The content of tourism big data is created on a daily, or even hourly, basis, and we are interested in making sense of the information, transforming big data into smart data and then using it for tourism planning.

(2) Variety
Another key characteristic of big data, both in terms of cost and ease of use, is the variety of data that stems from all accessible technologies. Variety describes the different formats of data that do not lend themselves to storage in structured relational database systems. The formats of big data include a long list of data such as documents, e-mails, text messages, images, graphs, videos, and the output from all types of machine-generated data from cell phones, GPS signals, sensors, machine logs, and DNA analysis devices (Li, Jiang, Yang, & Cuzzocrea, 2015). This type of data is characterized as unstructured or semi-structured and has always existed. 80 % of tourism-relevant information originates in unstructured form, and organizations can only count on the 20 % of structured data: for example, property management systems (PMS), Web or blog content management systems (CMS), or customer relationship management (CRM) systems can only deal with structured data, while the data on customer preferences at various points of contact are in the form of unstructured or semi-structured data, which require novel technologies to analyze them in order to develop new or improved products and services.

(3) Velocity
The third key characteristic of big data is velocity, which is referred to as the speed of responsiveness. There are three important aspects of the velocity of tourism big data (Chen, Mao, Zhang, et al., 2014). The first aspect is the consistent and complete capture, storage, and analysis of the fast moving streams of big data: for example, the stream of readings taken from a sensor or the weblog history of page visits and the clicks by each visitor to a website. The second aspect is the characteristics of timeliness or latency. We should capture, store, and use big data within a certain lag time depending on the type of the information since some of the data are permanently valuable while some would no longer be meaningful after a very short period of time. The third aspect is the speed with which big data must be stored and retrieved; the architecture of capture, analysis, and deployment must support real-time turnaround (in this case, fractions of a second); and must do this consistently over thousands of new customers. In tourism, for instance, we are concerned about how to send the right offer to the right person at the right moment when he or she arrives at a destination and what you should do if someone checks in to your hotel and is disappointed with the room and decides to tweet about it rather than call the front desk. Take the airlines in the travel business as an example, the dynamic revenue management could make a timely price change according to complex algorithms based on real-time or near-real-time customer online behaviors.

(4) Veracity
Veracity means the truthfulness and accuracy of data given the context, the variety of communication “touch points”, and the speed at which things happen. Big data veracity refers to the biases, noise, and abnormality in data: Is the data being stored and mined meaningful to the problem being analyzed? Compared with volume and velocity, veracity in data analysis is the biggest

challenge. In developing a big data strategy, you need your team and partners to help you keep your data clean and to have processes to keep "dirty data" from accumulating in your systems.

(5) Value

Value is frequently seen as another important characteristic of big data. The value of tourism big data can be described by its novel application in the tourism industry. First, there is the personalized application of tourism big data. Personalized marketing and targeted product design are extremely powerful opportunities that can be obtained from big data (Jani, Jang, & Hwang, 2014). Using a series of interviews conducted within the travel industry, Radovich (2015) showed how big data can be used to increase impact and reduce friction across disciplines, both within a company and within the industry. Personalization is a key tenet of big data. In order to most effectively win at true personalization, large travel companies must work across information databases to gather the myriad data points created by a consumer at different points. The second valuable application of tourism big data is the customer-centric experience. The customer should be at the center of all big data efforts. If big data gathering is seen as creepy or invasive, the consumer will not be pleased and loyalty will be lost. However, all signs point to consumers being willing to accept big intrusions into their behaviors if the resulting product is more targeted and able to anticipate their needs throughout.

5 Benefits of Big Data to Tourism Businesses

Big data analysis is changing all sorts of industries, not just the usual retail, logistics, and high-tech industries. It is also transforming the worlds of hospitality and travel since hospitality and tourism companies deal with a slew of user data covering all sorts of different information (e.g., flight confirmations or a customer's room preferences), and it creates all sorts of opportunities for correlating data to find otherwise unknown insights (Turner, 2014). In addition, there are some significant changes for big data because the cost of analytics platforms keeps dropping and employees are becoming more familiar with what big data can do. Essentially, big data can be used to tailor marketing campaigns and find business model inefficiencies. Big data analysis can deliver much needed business insights and can be the source of innovation for tourism organizations and the industry in general. The potential for big data in tourism is huge, and tourism organizations should not underestimate its importance (Pries & Dunnigan, 2015).

With the right approach, the tourism industry can learn a lot about consumer preferences and use this information and insight to build connections with individual travelers. Being able to offer travelers the right service or product at the right time is crucial. Without the right information and a very good targeting strategy, advertising will not result in any conversions and there will be no value. Travel is a fast-paced industry, and this drives the need for speedy data analytics and quick decisions. In tourism, any demand needs to be addressed instantly in order to remain relevant to travelers, and this is what makes big data so important. With the vigorous growth of the amount and applications of big data, traditional tourism data and methods are going to be interfacing with the novel data and methodologies: for example, call centers are going to be interfacing with online consumer reviews; loyalty programs are going to be linking with booking histories; and "property preferences" are going to be combined with social media chatter.

5.1 Consumer Behavior

We are in a time of unprecedented flux in consumer behavior, customer expectations, and company business models created by technologies that is simultaneously disrupting established businesses and spawning new ones (Marko, 2015; McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). However, tourism big data show significant changes in the relationship between businesses and their customers. So, we can use big data to provide superior buying and support experiences with a view to enhancing customer choice and expectations. The catalyst of using big data to re-recognize consumer behavior is the pervasive use of mobile devices, apps, and other social media, which play an ever-increasing role in the collection of raw information and easy access to the relevant tasks at hand.

Big data holds many insights into customers' behavior, some of which is already delivering while others yet to be realized. The potentials created by big data is particularly acute in retail since industries and business processes can successfully exploit new communication channels, service delivery options, and unprecedented sources (Marko, 2015). Collecting, correlating, and analyzing tourism big data from customer interactions across channels is the key to transforming the customer experience from a nightmare to nirvana (Chase, 2013). The nexus between big data and machine learning in all its forms, including predictive analytics and even neural network deep learning, is the foundation of well informed, highly efficient, and deeply satisfying interactions that benefit both customers and businesses.

The aim of using tourism big data is to create an authentic emotional connection between customers and partners of the tourism industry in order to make a significant improvement in customer service and support. The exploration of tourism big data has huge implications and provides opportunities for the seamless meshing of consumer experiences across mobile devices, websites, and personal interactions using multiple communication channels (e.g., phone, instant messaging, e-mail, web chat, and social networks). The key to the goal of using tourism big data is to be proactive, not just provide an integrated service. We need to anticipate customer needs and prevent problems. In other words, we can anticipate problems and queries by using statistical modeling and forecasting before we are asked the same question or asked to explain the same situation again and again.

5.2 Feedback Mechanisms

Feedback in the tourism industry is important in the quest to identify customer preferences and deliver positive experiences. Soliciting customer feedback is one of the most important elements in achieving high company growth and building a strategy around better meeting customer needs. Feedback based on tourism big data from customers, employees, partners, suppliers, and communities has also improved the capabilities of big data analytics. Data-driven business and consumer apps are the most common ways to collect feedback anytime and anywhere. A growing set of cloud services gives us the immediate and ubiquitous ability to interact using smart phones, tablets, or even watches (Chen, Mao, Zhang, et al., 2014).

The increase in gathering feedback using modern techniques has led to traditional feedback marketing being progressively replaced by commercial messages which are quick, unique, focused, and personal. One of the applications of the feedback mechanisms applied by the providers of tourism-related goods and services is price adjustment, in which a change in travel demand obtained from big data analysis and forecasting can provide useful information for quick and effective price adjustment.

Machine learning is one of the main technical methods used in the tourism industry to construct the feedback mechanism between customers and tour operators (Bajari, Nekipelov, Ryan, & Yang, 2015): for example, through cooperation between tour services providers, financial institutions, and telecom operators, machine learning can identify whether a person has just changed his/her residential address or travel internationally through checking for unusual charges. Machine learning with big data on customer experience can enable travel businesses and tour operators to proactively send text messages or calls to customers with new offers after they purchased their services. Specifically, machine learning could modify the feedback system by identifying the attempted user tasks and measuring their rates of success. Using this information, tourism businesses could then provide solutions to process inefficiency, customer frustration, and cross-channel breakdowns.

Predictive analytics are often presented as a cure-all for companies and can be incredibly useful. The predictive analytics with tourism big data used in modern feedback mechanisms represent a major improvement over old-fashioned human feedback. Predictive analytics can give marketing professionals more insight into customer preferences, which can be used to understand customers better and improve sales (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). However, the success of predictive analytics depends on both the quality of the big data and the customer feedback mechanisms.

Customer feedback mechanisms must be well designed and comprehensive to deliver actionable data in a timely fashion and acted upon immediately. Timely and reliable tourism big data can provide a rich portrait of customers and potential customers and subsequently lead to marketing efforts with advertising dollars being more precisely targeted toward the most fruitful channels. The framework shown above converts feedback into data that can be incorporated into broader analytics.

6 How to Use Big Data in Tourism Forecasting

We now turn to the key step of using big data in tourism forecasting, since we know that big data could bring many benefits to the tourism industry.

6.1 Capturing Big Data for Tourism Forecasting

Companies that effectively capture and implement big data strategies gain a competitive advantage since the technology required to process big data is a hindrance for many business users because of its complexity and cost. There are several steps in the process of capturing big data before we use it.

(1) Objective
The first step is the objective of using big data, which is to make sure that business benefits are derived from it (Pellegrini, 2013). When we capture big data, we should be able to access it and know what is available and determine where the business value lies. In other words, we should know the capability of big data and exactly what we are looking for and look to see what its values are. It is important to set specific business goals rather than just dealing with the big data itself.

(2) Visualizing big data
The second step is to make the big data visible to users within a company/organization. This will enable tourism forecasters to determine the optimal quantities of a product and to adjust logistical processes to maximize efficiency (Weiler & Black, 2014). The purpose of data visualization is to find the ways in which data could be effectively collected from different sources (visual and non-visual) and presented so that users could easily understand them. This will also help forecasters to better utilize big data in fulfilling their forecasting tasks.

(3) Structuring big data
The third step is to structure the unstructured data. This means to arrange big data according to traditional data length and format so that they can be fitted neatly into rows and columns in the spreadsheet. Structured data generally resides in a relational database and, as a result, is sometimes called relational data (Akerkar, 2013). The unstructured data can be easily mapped into predesigned fields: for example, a call center's structured data include numbers, dates, and groups of words and numbers called strings. It is commonly agreed that this kind of data accounts for about 20 % of the total amount of big data. Unstructured data are very difficult to analyze, since most of the big data is

unstructured or semi-structured data that contains a wealth of valuable information and does not fit into predefined data models. Thus, a number of different software solutions have been designed to search unstructured data and extract important information. In this chapter, we use pre-cleaned structured big data for tourism forecasting.

7 Selecting and Shrinking Big Data

Big data contains lots of information, which creates not only a storage issue but also a massive analysis problem. How to use these large datasets is the biggest problem in tourism forecasting using structured big data. The two most popular methods used in selecting and shrinking large amounts of structured data are the factor and LASSO (least absolute shrinkage and selection operator) modeling approaches.

(1) The factor model
The factor model is the most commonly used method in selecting and shrinking structured big data. A number of recent studies in the economics literature have focused on the usefulness of factor models in the context of forecasting related to the use of large datasets (Bai & Ng, 2006; Bańbura & Rünstler, 2011; Forni, Giannone, Lippi, & Reichlin, 2009; Hallin & Liška, 2011; Schumacher & Breitung, 2008; Stock & Watson, 2002; Stock & Watson, 2006; Teixeira, Klotzle, & Ness, 2008). We particularly analyze the predictive benefits associated with the use of dimension reducing independent component analysis (ICA) and sparse principal component analysis (SPCA), coupled with a variety of other factor estimation and data shrinkage methods, including, amongst others, bagging, boosting, and the elastic net. To assess the success of using big data, we could carry out a forecasting "competition" involving the estimation of different baseline model types, each constructed using a variety of specification approaches, estimation approaches, and benchmark econometric models (Stock & Watson, 2012).

(2) The LASSO method
The LASSO method is a covariates selection method in a linear regression framework (Tibshirani, 1996). It operates by penalizing the optimization problem associated with the regression with a term that involves the L1-norm of coefficients. It belongs to the family of penalized regression models that involve performing least squares with some additional constraints on the coefficients, the L1-norm in the case of LASSO. The literature has shown that LASSO tends to have a lower misspecification risk in forecasting models when compared with the usual information criteria (Ng, 2012). The LARS method (Efron, Hastie, Johnstone, & Tibshirani, 2004) can be combined with the factor model to shrink large datasets and used for forecasting economic series (Bai & Ng, 2008; Bessec, 2013; Schumacher, 2010).

8 A Framework for Predicting Tourism Demand Using Big Data

There is a widespread belief that big data can aid the improvement of forecasts provided we can analyze and discover hidden patterns and that predictions can be improved through data-driven decision-making (Shi, 2014). Some researchers believe that data mining techniques can be exploited to help forecasting with big data (Rey & Wells, 2013; Varian, 2014). However, data mining techniques always use static data as opposed to time series and are seldom used in tourism demand forecasting. When we turn to traditional forecasting methods for tourism demand forecasting with big data, the biggest problem is that the traditional forecasting tools cannot handle the size, speed, and complexity inherent in big data (Madden, 2012) even when it has been structured.

In order to apply a traditional forecasting method to big data, we have to simplify the structured big data (Hassani & Silva, 2015). One of the solutions is to shrink the big data and get the most important information in a suitable format that can be easily applied to the traditional forecasting model. Factor models are the most common and popular statistical and data mining technique used for big data forecasting; neural networks and Bayesian models are two other popular choices. In this chapter, we focus on the factor models.

(1) Mixed frequency model with big data
There has been some research success using big data for tourism forecasting. Choi and Varian (2012) aggregated Google data for Hong Kong's tourism demand forecasting and suggested that Google Trends' data about a destination may be useful in predicting visits to that destination. Yang et al. (2015) used web search query volume to predict visitor numbers for a popular tourist destination in China, and their results showed a significant decrease in forecasting errors when search engine data were used. However, these studies always aggregated or ignored weekly observations in order to make the datasets suitable for the traditional forecasting methods. Choi and Varian (2012), for example, only used the first two weekly observations of the month, discarding information for the latter 2 weeks, to predict total monthly visitors. Yang et al. (2015) aggregated weekly search engine data for forecasting. As a matter of fact, these researches could be improved by using a novel forecasting method, the mixed-data sampling (MIDAS) approach (Ghysels, Santa-Clara, & Valkanov, 2005), to fully utilize the high frequency search engine data (Bangwayo-Skeete & Skeete, 2015). Another mixed frequency model that fulfills the mixed frequency data job is the mixed frequency VAR model (Kuzin, Marcellino, & Schumacher, 2011; Qian, 2010), which treats low frequency data as high frequency data with missing data and then uses the state space model to deal with it.

(2) Factor model and forecasting combination

As a matter of fact, the best way to forecast the low frequency series (such as tourism demand) using high frequency data is to combine the shrinkage method with the mixed frequency models. Some studies used the mixed frequency model with factor high frequency data to forecast the macroeconomic indicators and obtained improved forecasting performance (Frale & Monteforte, 2011; Kuzin et al., 2011; Marcellino & Schumacher, 2010). The existing literature shows that compared with single model forecasts, forecast combination can improve forecasting accuracy in many practical situations (Bates & Granger, 1969; Chu, 1998; Coshall & Charlesworth, 2011; Deutsch, Granger, & Teräsvirta, 1994; Stock & Watson, 2004). In order to reduce the risk of forecasting failure (Wong, Song, Witt, & Wu, 2007), we suggest using forecast combination after obtaining different forecasting results from different methods and data.

Figure 2 displays the framework of tourism forecasting with big data. There are three important steps: (1) data exploration, which is the data processing that prepares the proper data for the model; (2) use modeling techniques to predict user behavior on the basis of their previous business transactions and

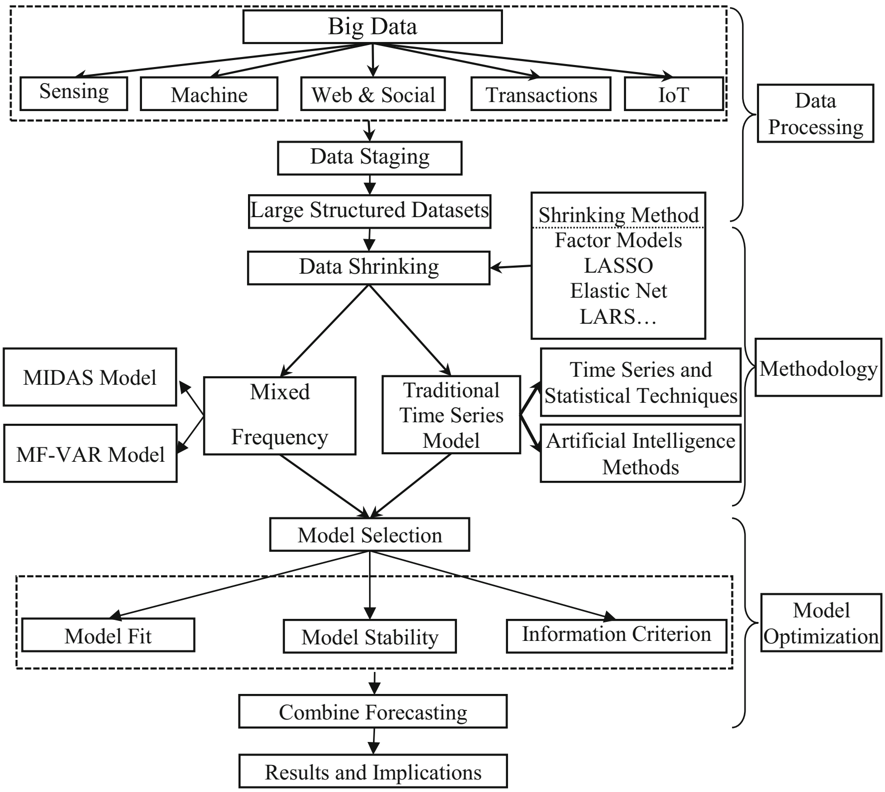

Fig. 2 The framework of tourism forecasting with big data

preferences; (3) optimize the forecast results and decrease the forecast failure risk by model selection and combination forecasting.

9 Conclusions

Big data is a social, cultural, technological, and ethical phenomenon that is not all good, all bad, or consistently neutral. With the proliferation and explosive increase in the application of big data, it has become a common tool in corporate decisions and a number of new social perils have arisen. At the same time, as data technologies become more pervasive, there are also privacy concerns and the potential for the abuse and misuse of big data (Bollier & Firestone, 2010). The use of tourism big data for forecasting has some visible and hidden pitfalls (Chareyron, Da-Rugna, & Raimbault, 2015). There are questions about the stability of the analysis and interpretation when the tools and techniques that we used in analyzing the big data have changed: Can the patterns that emerged from big data analysis or forecasting be generalized? How can information and privacy be controlled when anything and everything is systematically counted and recorded? In other words, there are challenges when tourism big data is used for forecasting. The first challenge is the difficulty of identifying the right data and determining how to best use it. The second challenge is to find the right talent capable of both working with the new technologies and interpreting the data to find meaningful business insights, and the third is to overcome the obstacle of data access and connectivity, which requires the right platforms to aggregate and manage big data. The fourth problem is how to find new ways of leveraging big data. The final concern is the security of big data and how to keep the advantage of using such data.

There are many potential solutions to overcome these challenges. First of all, the results of big data forecasting must promptly meet the need of business decisions. The purpose of tourism forecasting is to find and analyze the relevant data quickly and accurately. Visualization is a good way to present results and help those involved in tourism to make rapid decisions. We can also explore huge data volumes and gain business insights in near real time by improving the hardware and forecasting models. The second solution is to gain an overall understanding of the big data, which is crucial for visualizing and interpreting the data. To be specific, we need to have a deep understanding of where the data come from, what audience will be consuming the data, and how that audience will interpret the information. It is worth noting that outliers are important for tourism; therefore, we should pay more attention to the distribution and pattern of outliers and identify their influence. A third solution is to proactively take advantages of big data, as most of the information contained in big data is real time and huge in volume. Hence, the timely use of big data for forecasting and decision-making using proper approaches and methods is the best way to capitalize the benefits of big data.

All in all, the use of big data in the tourism and hospitality industry is still in its infancy, but the potential growth in application is huge. There is a lot of behind-thescenes work to be done, including sequencing for synchronous and asynchronous events and computing elapsed times of clusters of events, latency, and time between events, before big data results are presented to users. Fortunately, solutions for big data are emerging and the costs are much lower than before. In our opinion, the use of big data by airlines, restaurants, hotels, and other tourism and hospitality related industries enables them to learn a great deal about customers' preferences on the macro level and to benefit a lot with relatively small investment in the near future.

References

Akerkar, R. (2013). *Big data computing*. Boca Raton, FL: CRC Press.

Artola, C., Pinto, F., & Pedraza, P. D. (2015). Can internet searches forecast tourism inflows? *International Journal of Manpower, 36*(1), 103–116.

Bai, J., & Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-Augmented regressions. *Econometrica, 74*(4), 1133–1150.

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics, 146*(2), 304–317.

Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Machine learning methods for demand estimation. *American Economic Review, 105*(5), 481–485.

Bańbura, M., & Rünstler, G. (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting, 27*(2), 333–346.

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

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Or, 20*(4), 451–468.

Benckendorff, P. J., Sheldon, P. J., & Fesenmaier, D. R. (2014). *Tourism information technology*. Wallingford: Cab International.

Bessec, M. (2013). Short-term forecasts of French GDP: A dynamic factor model with targeted predictors. *Journal of Forecasting, 32*(6), 500–511.

Beyer, M. A., & Laney, D. (2012). *The importance of 'big data': A definition*. Stamford, CT: Gartner.

Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data*. Washington, DC: Aspen Institute.

Chareyron, G., Da-Rugna, J., & Raimbault, T. (2015). *Big data: A new challenge for tourism*. Paper presented at the 2nd IEEE International conference on big data.

Chase, C. W. (2013). Using big data to enhance demand-driven forecasting and planning. *Journal of Business Forecasting, 32*(2), 27–32.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19*(2), 171–209.

Chen, M., Mao, S., Zhang, Y., & Leung, V. C. M. (2014). *Big data: Related technologies, challenges and future prospects*. Heidelberg: Springer.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record, 88* (s1), 2–9.

Chu, F.-L. (1998). Forecasting tourism: A combined approach. *Tourism Management, 19*(6), 515–520.

Coshall, J. T., & Charlesworth, R. (2011). A management orientated approach to combination forecasting of tourism demand. *Tourism Management, 32*(4), 759–769.

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. *AIP Conference Proceedings, 1644*(1), 97–104.

Deutsch, M., Granger, C. W., & Teräsvirta, T. (1994). The combination of forecasts using changing weights. *International Journal of Forecasting, 10*(1), 47–57.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics, 32*(2), 407–499.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review, 1*(2), 293–314.

Forni, M., Giannone, D., Lippi, M., & Reichlin, L. (2009). Opening the black box: Structural factor models with large cross-sections. *Econometric Theory, 25*(5), 1319–1347.

Frale, C., & Monteforte, L. (2011). *FaMIDAS: A mixed frequency factor model with MIDAS structure*. Bank of Italy Temi di Discussione (Working Paper) No.788.

Frederiksen, L. (2012). Big data. *Public Services Quarterly, 8*(4), 345–349.

Fuchs, M., Höpken, W., & Lexhagen, M. (2014). Big data analytics for knowledge generation in tourism destinations—A case from Sweden. *Journal of Destination Marketing and Management, 3*(4), 198–209.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics, 76*(3), 509–548.

Hallin, M., & Liška, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics, 163*(1), 29–41.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems, 47*, 98–115.

Hassani, H., & Silva, E. (2015). Forecasting with big data: A review. *Annals of Data Science, 2*(1), 5–19.

Hendrik, H., & Perdana, D. H. F. (2014). Trip guidance: A linked data based mobile tourists guide. *Advanced Science Letters, 20*(1), 75–79.

Irudeen, R., & Samaraweera, S. (2013). *Big data solution for Sri Lankan development: A case study from travel and tourism.* Paper presented at the 2013 International Conference on Advances in ICT for Emerging Regions, ICTer 2, Colombo.

Jani, D., Jang, J. H., & Hwang, Y. H. (2014). Big five factors of personality and tourists' internet search behavior. *Asia Pacific Journal of Tourism Research, 19*(5), 600–615.

Jeng, J., & Fesenmaier, D. R. (2002). Conceptualizing the travel decision-making hierarchy: A review of recent developments. *Tourism Analysis, 7*(1), 15–32.

Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro area. *International Journal of Forecasting, 27*(2), 529–542.

Laney, D. (2001). *3D data management: Controlling data volume, velocity and variety*. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review, 52*(2), 21–31.

Li, K.-C., Jiang, H., Yang, L. T., & Cuzzocrea, A. (2015). *Big data: Algorithms, analytics, and applications*. Boca Raton, FL: CRC Press.

Madden, S. (2012). From databases to big data. *IEEE Internet Computing, 3*, 4–6.

Marcellino, M., & Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics, 72*(4), 518–550.

Marko, K. (2015). Using big data and machine learning to enrich customer experiences. http://www.forbes.com/sites/kurtmarko/2015/04/08/big-data-machine-learning_customer-experience/

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: The management revolution. *Harvard Business Review, 90*(10), 61–67.

Meeker, W. Q., & Hong, Y. (2014). Reliability meets big data: Opportunities and challenges. *Quality Engineering, 26*(1), 102–116.
Ng, E. C. (2012). Forecasting US recessions with various risk factors and dynamic probit models. *Journal of Macroeconomics, 34*(1), 112–125.
Pellegrini, T. (2013). The economics of big data: A value perspective on state of the art and future trends. In R. Akerkar (Ed.), *Big data computing* (pp. 343–371). Boca Raton, FL: CRC Press.
Poon, A. (1988). Tourism and information technologies. *Annals of Tourism Research, 15*(4), 531–549.
Prajapati, V. (2013). *Big data analytics with R and Hadoop*. Birmingham: Packt Publishing.
Pries, K. H., & Dunnigan, R. (2015). *Big data analytics: A practical guide for managers*. Boca Raton, FL: CRC Press.
Qian, H. (2010). *Vector autoregression with varied frequency data*. https://mpra.ub.uni-muenchen.de/34682/
Radovich, A. (2015). *Big data is fundamental in the hospitality and travel industry*. http://cliintel.com/big-data-is-fundamental-in-the-hospitality-and-travel-industry/
Rey, T., & Wells, C. (2013). *Integrating data mining and forecasting*. New York: OR/MS Today.
Schumacher, C. (2010). Factor forecasting using international targeted predictors: The case of German GDP. *Economics Letters, 107*(2), 95–98.
Schumacher, C., & Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting, 24*(3), 386–398.
Shi, Y. (2014). Big data: History, current status, and challenges going forward. *Bridge, 44*(4), 6–11.
Snijders, C., Matzat, U., & Reips, U.-D. (2012). Big data: Big gaps of knowledge in the field of internet science. *International Journal of Internet Science, 7*(1), 1–5.
Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association, 97*(460), 1167–1179.
Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting, 23*(6), 405–430.
Stock, J. H., & Watson, M. W. (2006). Forecasting with many predictors. In C. W. J. G. G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting* (pp. 515–554). North Holland: Elsevier.
Stock, J. H., & Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics, 30*(4), 481–493.
Sust, V. O., Illera, E. G., Berengué, A. S., García, R. G., Alonso, M. V. P., Torres, M. J. T., et al. (2014). *Big data and tourism: New indicators for tourism management*. http://telefonicacatalunya.com/wp-content/uploads/2014/05/BIG-DATA-Y-TURISMO-eng-interactivo.pdf
Teixeira, M. F., Klotzle, M. C., & Ness, W. L. (2008). Determinant factors of Brazilian country risk: An empirical analysis of specific country risk. *Brazilian Review of Finance, 6*(1), 49–67.
Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society Series B (Methodological), 58*(1), 267–288.
Turner, E. (2014). *How big data is changing the travel industry*. http://blog.sumall.com/journal/big-data-changing-travel-industry.html
Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–27.
Weiler, B., & Black, R. (2014). *Tour guiding research: Insights, issues and implications*. Bristol: Channel View Publications.
Wong, K. K. F., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management, 28*(4), 1068–1078.
Yang, X., Pan, B., 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.

Springer

http://www.springer.com/978-3-319-44262-4

Analytics in Smart Tourism Design
Concepts and Methods
Xiang, Z.; Fesenmaier, D.R. (Eds.)
2017, XVI, 307 p. 58 illus., 40 illus. in color., Hardcover
ISBN: 978-3-319-44262-4