PRZEDSIĘBIORCZOŚĆ I ZARZĄDZANIE 2019
Wydawnictwo SAN | ISSN 2543-8190

Tom **XX** | Zeszyt **2** | Część **II** | ss. **39–52**

Magdalena Kachniewska*
Szkoła Główna Handlowa
Kolegium Gospodarki Światowej
ORCID: 0000-0003-3163-0868

Big Data Analysis as a Tool for Predictive Intelligence and Experience Personalization in Tourism

Abstract: The main purpose of the paper is to identify the potential of Big Data Analysis (BDA) as a source of competitive advantage in the tourism market. A review of literature was adapted in order to define and estimate the significance of BDA. Some examples of BD application in travel and tourism were identified (e.g. the creation of tourism experience, relationship management, tourists' involvement and co-creation, personalization of value proposal, effectiveness improvement, promotion enhancement). The results are to be used as an indication for tourism market entities and the new technology industry supporting the tourism market.
Key-words: big data analysis, e-tourism, hospitality, tourism marketing, travel research, predictive marketing
Paper type: Research paper

Introduction

Much of the increase in data quality and application possibilities comes from a mix of new data sources, a smart application of statistical tools and domain knowledge combined with theoretical insights. The possibilities arising from big data analysis (BDA) in tourism concern five major data dimensions: data pertaining to customers, products, time, (geo-spatial) location and channel. Hospitality companies can use predictive analytics to track and send personalized messages by studying the behavior of travelers. A prompt advice makes a huge impact on the customer decision making. Data analytics instill a consumer driven decision-making approach. In the era

* Magdalena.Kachniewska@sgh.waw.pl

of consumerism, shifting the entire focus to the consumers and understanding their buying pattern helps tourism industry in streamlining its diverse operations. The predictive analytics generated by big data (BD) helps service providers and tourism regions in identifying leads, segregating them in real time and focus on the ones with high conversion probability. This helps a company/destination in offering personalized solutions and better understanding of latest market trends.

Although an increasing amount of published materials has focused on tourism practitioners in this domain, the literature remains largely anecdotal and fragmented. The purpose of this research therefore is to identify different conceptual dimensions of big data in travel and tourism and their relevance to business value. This work presents to the data-mining community an overview of research opportunities as well as an approach for improving data studies by considering them as a powerful research tool.

The state-of-the-art of Big Data

Tourism industry has been strongly influenced by the ICT growth, which affected the change in marketing strategies of tourism regions and companies. New technologies have created innovative ways for providing value to clients. The adoption of mobile devices as well as the Internet of things (IoT) support services that are tailored to users and contextual situations. Wearable sensors are feeding user information to social, medical and edutainment networks. The role of big data and predictive analytics in tourism is set to rise in importance, aided by large-scale correlational techniques. The perspective of the research within the field provides tools to achieve a set of objectives: sustainability, fairness, bounded information asymmetry, risk control, gracious failure, the context-aware designing and user centricity.

The term of big data is mainly used to describe enormous datasets that could not be perceived, acquired, managed, and processed by traditional IT and software tools within a tolerable time. Turning big data-call logs, mobile-banking transactions, online user-generated content such as blog posts and tweets, online searches, satellite images, etc. – into actionable information requires using computational techniques to unveil trends and patterns within and between these large socioeconomic datasets.

Gantz and Reinsel [2011] introduced the concept and potential of BD for the first time in 2011. Their report triggered the great interest in both industry and aca-

demia. Nearly all major companies, including EMC, Oracle, IBM, Microsoft, Google, Amazon, and Facebook have started their BD projects, as they enable e-commerce firms to use data more efficiently, drive a higher conversion rate, improve decision making and empower customers [Miller 2013]. The UN Global Pulse [2012] report summarized how governments utilized the data to serve people, while McKinsey announced Big Data as the next frontier for innovation, competition, and productivity [Miller 2013]. Thus, BDA can be estimated from the perspective of transaction cost theory [Williamson 1981] or the resource-based view [Barney 1991]. Drawing on transaction cost, BD can benefit firms by improving market transaction cost efficiency (e.g. buyer-seller interaction online), managerial transaction cost efficiency (e.g. process efficiency) and time-cost efficiency (e.g. searching, bargaining and after sale monitoring). On the ground of the resource-based theory, BD is a distinctive competence of the high-performance businesses (detecting quality problems, determining the optimal price, identifying loyal and profitable customers) [Davenport, Harris 2007].

A number of literature on BD have been completed in the past few years [Sagiroglu, Sinanc 2013, Fosso et al. 2015, Gandomi, Haider 2015, Amiri et al. 2015, Wang et al. 2016], including considerable discussions on the definition of BD [Grobelnik 2012]. The BD research focuses on how to transform "a bunch of data" into "big data" [Gantz, Reinsel 2011]. Laney [2001] defined the BD challenges and opportunities with a "3Vs" model (volume, velocity and variety). According to Manyika et al. [2011] BD characteristics should be completed by the fourth V: value. It highlights the necessity to explore the hidden values and indicates a critical problem: how to discover values from datasets with an enormous scale and rapid generation.

BD applications in science, engineering, medicine, medical care, finance, business, law enforcement, education, transportation, retail, and telecommunication, as well as in small and medium-sized businesses, government departments, human-computer interaction are all important research problems. There is still an ongoing discussion on the present and future of big data source and application. Players on a tourism market (including governmental agendas) should keep an eye on the progress in this field, as the new source of data create potential for innovation: the examples of recent years show that newcomers in travel market constitute an additional challenge for the present entities.

Understanding the sources and tools for Big Data Analytics

The value chain of big data can be divided into four steps: data generation, acquisition, storage and analysis. The last one (BDA) is a production process that utilizes the "raw material" to create new value. The basis for data generation are: Internet (including social media and e-banking), internal data of enterprises, bio-medical data, government agendas, mobile data, internet of things (IoT) and sensor data [Zaharia et al. 2012, Bhatotia et al. 2011]. Structured data focuses on demographic data including name, age, gender, date of birth, address, and preferences, unstructured data includes clicks, likes, links, tweets, voices, etc. Among smart cities and destinations, data may come from industry, agriculture, traffic, transportation, medical care, public departments, families, etc. Due to the way we use things in everyday life, the data represent four important characteristics: large-scale, heterogeneity, strong time and space correlation and low share in an overall data.

The most priceless element of data collection is log files (e.g. web servers record in log files, number of clicks, click rates, visits, and other property records of web users etc.), completed with sensing and mobile devices. All the data stay useless without a proper data mining, which comprises techniques and algorithms for determining interesting patterns and sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems. Specific uses of data mining include: market segmentation, custom churn; fraud detection; direct marketing; interactive marketing (predict what each individual accessing a web site is most likely interested in seeing), market basket analysis (products and services commonly purchased together) and trend analysis [Du 2010].

The importance of BD is determined not only by their size, but also by their diversity and flow rate: any comments, information and „likes" in social media and blogs, complaints or discussions on online forums using the brand name, tourist region or attraction form an uninterrupted, chaotic stream of data that, when analyzed, can contribute to the creation of unique guest experiences. Network data analysis seems to be especially important for travel industry. It has evolved from the initial quantitative analysis [Hirsch 2005] and sociological network analysis [Watts 2004] into the emerging online social network analysis (SNA) in the beginning of 21st century. They include massive linked data and content data. The linked data, the form of graphic structures, describes communications between two entities. The content data con-

tains text, image, and other network multimedia data. Marketing, advertisement and recommendation can benefit from social influence by qualitatively and quantitatively measuring the influence of individuals on others [Tang et al. 2009]. Recently, the progress in wireless sensor, mobile communication and stream processing enable people to build a body area network to have real-time monitoring of people's tourism activity.

The fundamental needs of the proper database marketing is the availability of data and the creation of data mining algorithms (a set of heuristic and calculations that creates a data mining model from data) [Ginsberg et al. 2009]: real-time analytics perform analysis on live data, for example, a service that is currently in progress, detecting issues or problems as they happen, allowing them to be rectified before they negatively impact the service, while batch analytics are used to identify trends over time. They still stay useless without the proper data mining techniques, which include [Han, Kamber 2006]:

- characterization (used to generalize, summarize and possibly different data characteristics)
- classification
- regression
- association (discovers the association between various data bases and the association between the attributes of single database)
- clustering (breaks large data set into smaller groups to make the designing and implementation process to be simple and to reduce the similarity between the classes)
- change and deviation detection
- link analysis (traces the connections between the objects to develop models based on the patterns in the relationships by applying graph theory techniques)
- sequential pattern mining.

An interdisciplinary character of BD research is one of main obstacles in the development of BD applications: data representation, redundancy reduction, data compression; data life cycle management, analytical mechanism, data confidentiality, energy management, expendability and scalability, need advanced cooperation. [Labrinidis, Jagadish 2012]. Despite the problem, some companies have attempted to raise the power of BDA. In hospitality information systems are used to assist the delivery of proper services during the customers stay in a hotel. Some of the key ways

are improved: capacity management and operations efficiency, central room inventory control, last room available information, yield management, sales and operational reports, tracking repeat hotel guests. High-quality database marketing research is used to predict consumer behavior trends, travel choice and willingness to make a reservation.

Big Data applications in tourism predictive marketing

Travel and tourism have been perceived as one of the main users of information technology [Olsen, Connolly 2000]. The amount of undeveloped tourist data results from the frequency of online transactions, plentitude of internet searches, the use of price comparators, and social network activities. Each airline, hotel or car rental reservation could be digitally traced – the data are irrelevant when isolated from the context, but when analyzed along with the whole set of data they constitute the source of information which could become the basis to management decisions [Kachniewska 2014].

The prediction of tourists' future behaviors is not based on typical statistical data, but real personalized data. One can easily assume that people with school-age children plan trips during the school holidays – thanks to data analysis based on web logs or social media communication, marketers can easily get in touch with right people (school children parents). Marketing creation for the group will differ from those addressed to any other potential customers (seniors, singles etc.). Moreover, one can sell properly matched products exactly at the time when parents plan trips. The use of data from various sources enables profiling the customer in such a way that the sales activities are directed to a precisely selected target group. A tourism service provider is able to narrow the audience on the basis of demographics as well as his/her real time decision.

Mobile phones and IoT require larger capacity of supporting location sensing, people-oriented, and context-aware operation. The phenomena are of utmost importance for the tourism market. Contemporary tourists like to change their itinerary during the trip and combine several purposes with travelling, such as business, leisure, entertainment and education [Kachniewska 2014]. They look for the possibility to compose their tourism packages personally and to modify them during the trip according to their personal needs and interests. They need pragmatic information wi-

thin an easy reach (available anytime and anyhow) during their travel. Once having access to BD tools, the marketers gain a unique chance to get information about tourists' needs and habits, to affect travelers buying behavior and provide them with the customized information. On sales planning, enterprises can optimize their commodity prices and improve their operation efficiency, optimize the labor force, accurately forecast personnel allocation requirements, avoid excess production capacity. On supply chain, they conduct inventory optimization, logistic optimization, and supplier coordination. By analyzing customers' transaction records, potential customers can be identified thus enabling up- and cross-selling.

The most classic BD application concerns tourism e-commerce (online reservation systems). A new challenge seems to be the sale of diversified hotel services. Hotels may have profoundly experienced the application of IoT big data, having at their disposal all the real time information on the guest's presence or absence in the room, the way they use lights, water, heating or conditioning as well as electronic devices, different TV channels or music. The research in 25 Polish 4- and 5-star hotels revealed that 67% of guests are not interested in TV and radio station availability – what they really need is internet access – especially VoD and information platforms. The data were not gathered on the basis of any questionnaires or review – but only through the IoT data analysis. BDA has also revealed that consumers are platform hopping, and although the majority of online bookings are made using desktop devices, 20-50% of reservations (depending on the market) are done from mobile devices [Travolution 2018].

90% of travelers share photos and experiences of their travels on social networks. Millions of travel-related reviews are shared on the internet every day. So called "sentiment analysis" permits the estimation of the polarity of these posts in milliseconds [Amadeus 2018]. The hidden patterns unveiled by the process make it possible to make predictions about what tourists buy according to the time of the month or what other items they might purchase. The application of BD from online SNS may help to understand user's behavior on the basis of:

- early warning (to rapidly cope with crisis if any by detecting abnormalities in the usage of electronic devices and services)
- real-time monitoring (to provide accurate information for the formulation of policies and plans by monitoring the current behavior, emotion, and preference of users)
- real-time feedback (acquire groups' feedbacks against some social activities based on real-time monitoring).

In 2008, Microsoft purchased Farecast, a sci-tech venture company, which has an airline ticket forecast system that predicts the trends and rising/dropping ranges of airline ticket price. By 2012, the system has saved nearly USD 50 per ticket per passenger, with the forecasted accuracy as high as 75 % [Bradlow et al. 2017]. This system is accompanied by segmentation and clustering of passengers, a well-known driver to explain customer behavior (e.g., price sensitive vs product quality) and then used to adapt travel products accordingly. Clustering algorithms permit discovery of new kinds of behaviors (menu, accompanies, travel habits etc.).

Predictive analytics play an important role detecting frauds. Security should be understood broadly: this is about weather phenomena, natural disasters, and technical problems (e.g. in aviation). The development of mass tourism has brought disadvantages associated with delays in transport, trouble with lost luggage and longer queues annoying travelers. Large number of online transactions need to be validated in real-time. Amadeus process more than 1 billion transactions per day in one its data centers. New aircraft have close to 6,000 sensors generating more than 2 Tb per day. Using supervised machine learning algorithms, known defects can be anticipated when a combination of factors is observed (like a set of symptoms in medical diagnose) [Amadeus 2018].

The illustration for the process of creating a new market offer on the basis of BDA may be an application that inform travelers on a regular basis about the status of their flight (delays, cancellations, the possibility of choosing an alternative means of transport, etc.). The consequence of using BDA are new types of search engines, including the Amadeus search engine „Featured Results" – suggested to the potential tourist 4 the most suitable options for tourist packages or „Extreme Search" – which is based on quite unusual inquiries about the total budget for the given travel, duration of stay and minimum temperature at the destination, which is willing to accept a tourist [Carat Manchester 2018].

BDA gives the opportunity to link the pricing policy of the enterprise with the management of physical potential (available beds or places on the aircraft) in real time, taking into account passenger profiles, estimating the likelihood of cancellation and the safe size of overbooking. The first works on linking the said information with the employment planning process are carried out (demand fluctuations not only concern the seasons, but also individual days of the week). The learning system is able to plan the necessary number of cabin crew (Lufthansa) or floor service (Marriott).

The most luring application for tourism providers consists in dynamic offers of travel products and offering personalized experiences. In few milliseconds systems need

to predict what the guest wants and then adapt the offer, accordingly, e.g. bringing together a flight, extra baggage, a hotel, while also slightly adapting the price dynamically and considering competitors offers in real-time [Carat Manchester 2018].

Another application of BDA in tourism is a recommender system for travel products, proposing the most valuable and relevant options to users while maximizing revenues of travel providers. Predictive analytics help to better understand user needs and match this knowledge to possible products and services: from inspiration (where to go?) to automation of reservations [Amadeus 2018, Travolution 2018]. A person regularly using certain airlines can expect to be questioned after a certain time about the reason for resignation of their services (if they did not appear on board for some time) or the offer of renting a car at the airport in which they landed regularly. In social networks, we usually receive travel suggestions to go to a place that someone from our friends has just visited; after leaving a positive comment on the restaurant page, we can expect invitations with a coupon for a free dessert etc. A separate, sensitive issue (of both an ethical and legal nature) remains the issue of the urgency of these communications and the right to post them.

BDA can ensure more effective online marketing as it is all about conversion and ability to sell products with minimum exposure. Using the commercial tools offered by the giants of social media market (Facebook, YouTube, SnapChat etc.) tourism service providers can optimize their conversion rates. Attracting users to an ad is not enough if they are not buying. New predictive algorithms could estimate conversion and help better define travel products, better place the ads and finally optimize advertising campaigns. Within the group of 54 hotels testing a new system of conversion rates optimization 50% managed to lower the cost of on-line campaigns by 15%, others by 5–9% [Travolution 2018].

The growing number of transactions concluded via electronic media facilitates the collection of data on tourists' behavior [Zhunge, Shi 2004]. Once knowing the context of the information one can combine the user's demographic data with specific requirements and circumstances of the transaction. Contextual data analysis allows to obtain quantitative (number of transactions, their size, etc.) and qualitative information (e.g. shopping habits of tourists: how often are purchases made in advance and ad hoc, what kind of information ultimately determines the transaction; which of them are browsed at the very beginning, etc.).

Tourist entrepreneurs having Facebook pages do not realize how widespread the cooperation this portal offers. Some of the social media services, such as „Custom Au-

diences" and „Sponsored Stories in the News Feed" allow to appreciate the impact of the promotion based on BDA. In the first case, it is about using your own contact database of clients (without revealing their identity) in order to better personalize the stream of information reaching the user. In the second case, we are talking about an even more sophisticated tool, which causes the Facebook user in the received information stream to get information about the tourist activity of his friends, using the name of the brand.

Last but not least, BDA makes forecasting more reliable (e.g. for revenue management systems and demand analysis). Historically, forecasting engines embedded in revenue management systems considered only past bookings. New generation forecasting systems are based on specific factors (price, schedules, etc.) in addition to simply time-series, thus allowing them to optimize revenue.

Conclusion

Observation of the (potential) tourists' behavior enables the implementation of experience tourism and a new way of thinking about connecting with customers and ensuring their loyalty. Various contextual data that is sought after by tourism service providers, ranging from data about the tourist himself, through information about the location and purchased services, to details on the frequency of logging in and the type of information sought (e.g. about food and hotel services), but also information closely related to the destination (location-aware information), including information on cultural events, history and art, recreation, etc.).

The shaping of the tourism offer should be based on the access to information about tourists' expectations and preferences. The efficiency in meeting these expectations is a derivative of the pace of sharing knowledge and information among the entities responsible for shaping the tourist offer. An indispensable condition for the correct and wide application of BDA to improve the competitiveness of enterprises and tourist regions is an integrated view of the tourism industry. Creating a user-oriented system requires the integration of efforts of various entities involved in providing visitor experiences – that is, adopting a demand-based approach to defining the tourism industry. Despite a clear connection with new technologies and a natural environment of data acquisition and processing, BDA is not intended to build a technocentric vision of the future's travel [Mayer-Schonberger, Cukier 2013, p. 43]. It is

used to predict the behavior of buyers and providers of travel services, in relation to technical infrastructure, business models as well as social values and trends. The potential of BDA is currently used by the largest players in this industry and newly emerging technology companies that treat BDA as a natural business environment. Fragmentation of the tourism industry and the lack of a common platform enabling the collection and sharing of information and the free flow of data limit the chances of using large data sets to improve the competitiveness of individual entities.

References

Amadeus (2018), *Predictive Analytics in Travel Industry,* [online] https://bit.ly/2KQqbpU, access: 12.11.2018.

Amiri M., Moisiadis F., Davarzani H. (2015), *Socio-environmental performance of transportation systems*, "Management of Environmental Quality", Vol. 26, No. 6.

Barney J. (1991), *Firm resources and sustained competitive advantage*, "Journal of Management", No 17.

Bhatotia P., Wieder A., Rodrigues R. et al. (2011), *Incoop: mapreduce for incremental computations*, Proceedings of the 2nd ACM symposium on cloud computing, ACM.

Bradlow E., Gangwarb M., Kopallec P. et al. (2017), *The Role of Big Data and Predictive Analytics in Retailing*, "Journal of Retailing", Vol. 93, No. 1.

Carat Manchester (2018), *How the package holiday is being personalized,* [online] https://bit.ly/2QbqMIw, access: 26.11.2018.

Davenport T., Harris J. (2007), *Competing on analytics: The new science of winning*, Harvard Business School Press, Boston.

Du H. (2010), *Data Mining Techniques and Applications an Introduction*, 1st Edition. Cengage Learning Edition, Boston.

Fosso S., Akter S., Edwards A. et al. (2015), *How 'big data' can make big impact: findings from a systematic review*, "International Journal of Production Economics", Vol. 165.

Gandomi A., Haider M. (2015), *Beyond the hype: big data concepts, methods, and analytics*, "International Journal of Information Management", Vol. 35, No. 2.

Gantz J., Reinsel D. (2011), *Extracting value from chaos*, "IDC iView".

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", Vol. 457, No. 7232.

Grobelnik M. (2012), *Big data tutorial,* [online] https://bit.ly/2DYGob3, access: 13.11.2017.

Han J., Kamber M. (2006), *Data Mining: Concept and Techniques*, 2nd Edition, Morgan Kauffmann, Burlington.

Hirsch J. (2005), *An index to quantify an individual's scientific research output*, "Proc Natl Acad Sci", Vol. 102, No. 46.

Kachniewska M. (2014), *Tourism value added creation through a user-centric context--aware digital system*, "University of Szczecin Scientific Journal", No. 836, "Economic Problems of Tourism", Vol. 4 (28).

Labrinidis A., Jagadish H. (2012), *Challenges and opportunities with big data*, "Proc VLDB Endowment", Vol. 5, No. 12.

Laney D. (2001), *3-d data management: controlling data volume, velocity and variety*, META Group Research Note, 6 February.

Manyika J., Chui M., Brown B. et al. (2011), *Big Data: The Next Frontier for Innovation*, McKinsey Global Institute: Competition and Productivity, San Francisco, CA.

Mayer-Schonberger S., Cukier B. (2013), *Big data: a revolution that will transform how we live, work, and think*, Eamon Dolan/Houghton Mifflin Harcourt.

Miller M. (2013), *6 ways to use 'big data' to increase operating margins by 60%*, [online] http://upstreamcommerce.com/blog/2012/04/11/6-ways-big-data-increase-operating-margins-60-part-2, access: 02.03.2018.

Olsen M., Connolly D. (2000), *Experience-based Travel: How Technology Is Changing the Hospitality Industry*, "Cornell Hospitality Quarterly", Vol. 41, Iss. 1.

Sagiroglu S., Sinanc D. (2013), *Big data: a review, International Conference on Collaboration Technologies and Systems (CTS)*.

Tang J., Sun J., Wang C. et al. (2009), *Social influence analysis in large-scale networks*, The 15th ACM SIGKDD conference on knowledge discovery and data mining, ACM.

Travolution (2018), *Research identifies four key holiday booking trends*, https://bit.ly/2zEAhWB, access: 5.11.2018.

UN Global Pulse (2012), *Big Data for Development: Challenges and Opportunities*.

Wang G., Gunasekaran A., Ngai E. et al. (2016), *Big data analytics in logistics and supply chain management*, "International Journal of Production Economics", Vol. 176.

Watts D. (2004), *Six degrees: the science of a connected age*, WW Norton & Company.

Williamson O.E. (1981), *The economics of organization: the transaction cost approach*, "American Journal of Sociology", No. 87.

Zaharia M., Chowdhury M., Das T. i in. (2012), *Resilient distributed datasets, Proceedings of the 9th USENIX conference on networked systems design*, USENIX Association.

Zhunge H., Shi X. (2004), *Toward the Eco-grid: A Harmoniously Evolved Interconnection Environment*, "Communications of the ACM", Vol. 47, No. 9.